

1

Introducción

En este capítulo se describen los conceptos y definiciones básicas necesarios para poder captar mejor los temas a desarrollar en esta asignatura. La idea es facilitar la comprensión, empleando una terminología común. Así, se establecen una serie de convenciones, para agilizar la lectura del texto. El problema consiste en la diversidad de términos usados para nombrar los mismos conceptos, debido a la gran variedad de autores, modas y tendencias, que existen en la bibliografía actual. Cada uno, aportando su cuota de originalidad, pero complicando la simplicidad. Otro problema que se trata de resolver, es el uso de palabras con un significado claro en el lenguaje diario, pero con uno diferente en la Estadística. Por tanto, conviene empezar precisando algunas ideas generales a modo de convención.

1.1 Definiciones primarias

La palabra **Biometría** proviene del griego, con dos raíces básicas: *Bios* (vida) y *Metron* (medida). Se trata entonces, de una materia que se dedica a la teoría de las mediciones, en las Ciencias Biológicas. Y en un sentido más amplio, a la aplicación de los métodos estadísticos para resolver problemas biológicos, por lo cual la Biometría es llamada también: **Bioestadística**. Para redondear estos conceptos, conviene definir:

Estadística es un conjunto de métodos científicos para la recopilación, representación condensación y análisis de los datos extraídos de un sistema en estudio. Con el objeto de poder hacer estimaciones y sacar conclusiones, necesarias para tomar decisiones.

La Estadística no es una ciencia en sí misma. Se trata de un grupo de métodos con base científica. Los *métodos* son modelos que optimizan matemáticamente los objetivos buscados. De hecho, la Estadística Teórica es una rama de las Matemáticas. *Recopilar* datos significa obtenerlos efectuando mediciones, muestreos, encuestas, censos, etc. La *representación* de datos implica mostrarlos con gráficos, con tablas, en forma de texto, o cualquier combinación de éstas. La *condensación* de los datos implica reducir su número a dos o tres valores representativos de todo el grupo, denominados *estadísticas*, *estadígrafos* o *números índices*, tales como la media, la mediana, la varianza, costo de vida, etc. El *análisis* se hace con las herramientas estadísticas, empleando la información obtenida de los datos, para realizar *estimaciones* o inferencias, testear hipótesis de trabajo y así, poder tomar las *decisiones* más adecuadas en cada caso particular, basadas en la evidencia científica suministrada por estos análisis.

De la definición anterior, surge que la Estadística puede ser usada en cualquier *sistema* en estudio. En la práctica, esto significa una gran cantidad de posibilidades, pues, donde pueda definirse un sistema, allí se podrá emplear la Estadística.

Sistema: es un conjunto de elementos que se aíslan para su estudio en función de sus interrelaciones.

Tienen tres características que permiten identificarlos de entre todos los conjuntos:

- *Objetivos:* Saber lo que se quiere del sistema, o lo que se espera que éste haga.
- *Comportamiento:* Lo que realmente hace el sistema, cómo se comporta.
- *Estructura:* La manera en que se interrelacionan los elementos.

Por ejemplo, el sistema educativo está compuesto por una serie de elementos: escuelas, docentes, alumnos, administrativos, programas de estudio, textos y otros. Su claro objetivo es educar a sus alumnos. Se comporta dando los resultados que están a la vista. Y tiene una estructura formada por un conjunto de leyes, programas de estudio, reglamentos, etc. También son ejemplos de sistema: el aparato circulatorio, el respiratorio, una empresa comercial, el planetario, una persona, etc. De la comparación entre los objetivos buscados y el comportamiento obtenido, surge el concepto de *control*. Tomando el caso de una heladera, tiene un grupo de elementos componentes como el gabinete, la puerta, los estantes, el motor y otros que no están dispuestos de cualquier forma sino que siguen un plano dado (una estructura). El objetivo, es mantener en su interior una cierta temperatura por debajo de la ambiental. Cuando el motor trabaja, comienza a enfriar el interior del gabinete, bajando la temperatura hasta un valor límite que hace funcionar un relee, se corta la corriente eléctrica, y el motor se detiene. Dicho relee, hace las veces de control, pues en todo momento compara la temperatura real con la temperatura deseada; detiene o arranca al motor según convenga. Como puede verse a través de estos ejemplos, sistema es un concepto muy amplio, tan amplio, como el campo de aplicación de la Estadística.

Es importante destacar que el sistema completo bajo estudio estadístico se constituye en la *población o marco de referencia*. Toda la información obtenida adquiere sentido si se sabe respecto de qué o de quiénes, es válida. Por ejemplo, el dato numérico 4,3 solo significa un número real al no tener ninguna referencia concreta. Si se le añade una unidad física, como 4,3 kg se agrega una información extra: se trata de una masa. Finalmente, si se dice que es el peso promedio de doce dorados obtenidos en el último concurso de pesca, entonces se tiene una idea más concreta que el mero significado numérico inicial, pues ahora se tiene un marco de referencia concreto, que es la población de dorados que vivían en el río donde se realizó el concurso, a la fecha del mismo. Análogamente, en Psicología se habla del marco de referencia del individuo, formado por el conjunto de conocimientos, creencias, experiencias vividas, principios éticos y morales del mismo, etc. Todo esto hace que un mismo hecho, pueda ser interpretado de varias maneras por distintas personas, pues cada una, tiene su propio marco de referencia. Ver un desnudo no provoca la misma reacción en un artista plástico que en un censor de cine; el azúcar considerado como un dulce por la mayoría, puede ser visto como un veneno por un diabético. Así también, todo trabajo en Biometría deberá tener su propio sistema referencial: *la población*.

Población es el conjunto de todas las muestras posibles, que pueden obtenerse del sistema en estudio de acuerdo al método de selección empleado.

Población Biológica es la formada por todos los individuos de una misma especie en un estadio dado de su vida histórica o sexual, que habitan un área geográfica definida y en una época determinada.

El marco de referencia de un estudio biométrico, es la población biológica de la cual se obtuvieron los datos. Este marco se necesita, para que adquieran sentido los números que se manejan. Por ejemplo, si se hace un estudio midiendo el hematocrito de un grupo de pacientes, varones adultos y sanos, entonces la población estará compuesta por todos los seres vivos de la especie Homo-Sapiens, que habitan el planeta Tierra, de sexo masculino y con una edad mayor de 18 años, en el momento del estudio. El *tamaño* de la población se saca contando el número de elementos componentes. A veces es un conteo simple, pero otras veces se trata de conteos ordenados y se debe recurrir a las técnicas del Análisis Combinatorio para calcular el tamaño de la población. Si se aplican dos técnicas clínicas diferentes a un mismo paciente para detectar una infección, los resultados posibles no son dos, sino cuatro: (+ +), (- +), (+ -) y (- -). Para aclarar mejor esto, se puede imaginar un experimento en el cual se lanza un dado y se observa el número que sale. Los casos posibles son las seis caras del dado y ese es el tamaño de la población. Sin embargo, si se lanza dos veces ese dado, ahora los casos son (1,1) (1,2) (1,3) ... (6,5) (6,6), o sea, 36 casos posibles en total, y esa cantidad pasa a ser el tamaño de la nueva población, donde cada elemento está formado por un par de números.

Muestra: *es un conjunto de datos obtenidos de una población cualquiera, con el método de recopilación elegido. Se la puede imaginar como un subconjunto del conjunto población.*

Se toman muestras, cuando no se puede o no conviene, tomar la población entera. Si se tiene una población de tamaño infinito, no se podrá nunca tomar todas las muestras posibles, como por ejemplo, las mediciones repetidas de una misma magnitud, que se pueden repetir indefinidamente mientras el ensayo no sea destructivo (repetidas pesadas en una balanza, medir la temperatura de un cuerpo, etc.). Hay ocasiones, donde si bien la población es finita, es tan grande que no resulta práctico tomar todos los casos como por ejemplo, cuando la población es la especie humana. Otras veces, las determinaciones que se deben realizar en las muestras son tan caras que resulta mucho más barato tomar muestras. Pueden haber razones de tiempo que impidan analizar a toda la población. Si el método de ensayo es destructivo, como ver si los fósforos funcionan, o abrir ampollas de medicamento para verificar su contenido, entonces no hay más remedio que tomar muestras. En cambio, si se revisan diamantes para determinar si son falsos, allí se tomará a cada uno de los elementos componentes de la población, y nunca una muestra de los mismos. La idea básica es tomar muestras representativas de la población desconocida, y a través del análisis de las mismas poder hacer deducciones acerca de esa población. La clientela que concurre a una farmacia o a un laboratorio de análisis clínicos proviene de una población relacionada con la cercanía a su ubicación geográfica. Conocerla en sus preferencias, es algo fundamental para toda campaña publicitaria destinada a incrementar las ventas.

Estadística Descriptiva: *Es la parte de la Estadística que se ocupa de recopilar, representar y condensar los datos obtenidos del sistema en estudio.*

Estadística Inferencial: *Es la parte de la Estadística dedicada a la formulación de supuestos y estimaciones, para hacer predicciones y poder sacar conclusiones de los datos obtenidos con el estudio de las muestras. Y así, poder tomar decisiones con base científica.*

La Estadística se emplea en el estudio de los *fenómenos naturales*, tanto los generados en los laboratorios por los científicos como aquellos más allá del control humano. En una gran variedad de disciplinas como economía, sociología, política, ciencias de la salud, en estudios de

mercado, urbanismo, etc. Es una herramienta de uso tan amplio y general que hoy día es difícil imaginar un lugar donde no pueda emplearse. Más aún, en algunas disciplinas es la herramienta básica de medición, como por ejemplo en parapsicología para la determinación de PES (percepciones extra-sensoriales).

1.2 Historia breve de la Estadística

Desde los albores de la civilización, el hombre ha tratado de evaluar de alguna manera los fenómenos que lo rodean cuantificando lo que puede observar. Se cree que en las cavernas prehistóricas sus habitantes contabilizaban sus pertenencias, haciendo muescas en las paredes. A medida que los pueblos evolucionaban en cantidad y conocimientos, se requería cada vez más el uso de recopilaciones de datos, con fines administrativos, religiosos, etc. En el antiguo Egipto existía una administración centralizada destinada a la toma de datos y a la contabilización de todos los bienes del faraón. La diosa de los libros y las cuentas fue Sakhmet y puede decirse que sus adoradores eran los estadísticos de la época. En Asiria se encontraron tablas de la biblioteca de Asurbanipal conteniendo información como las producciones de cada provincia del Imperio. En la Biblia hay un Libro llamado de los Números donde se relata el censo que hizo Moisés con los israelitas antes de cruzar el desierto. En el Libro de los Reyes se habla de un censo hecho por David. Además de otros recuentos como los del Libro de Esdras y Nehemías. En Confucio hay referencias a un rey llamado Yao (3000 AC), quien ordenó un censo comercial y agrícola de toda la China. Sólo un buen manejo de la información estadística explica el desarrollo de los fenicios en el comercio y la navegación épica, como el Periplo cuya traducción griega todavía se conserva. Allí se cuenta sobre un tal Henón, que bordeó la costa de África al mando de sesenta naves, desde Gibraltar hasta Liberia. Así como tantas otras epopeyas que implican la necesidad de contabilizaciones y recuentos para la administración de hombres, armas y suministros. En la Biblia es famosa la sentencia “Con la medida que midáis, así seréis medidos”. Aristóteles, en su libro La República, comienza a asimilar Estadística con el concepto de manejo del Estado. La particular relación que lo unía con Alejandro Magno, quien puso a su disposición grandes sumas para financiar el Liceo, donde los primeros destellos de la ciencia organizada fueron enseñados a sus alumnos, en particular el manejo de la cosa pública.

El relato del nacimiento de Cristo presenta el concepto de empadronamiento. En la Roma antigua, se llevaban registros sistemáticos de datos y aún censos, con fines tributarios. En América, los incas usaban nudos en las cuerdas (quipus) para efectuar recuentos en el manejo de sus almacenes comunitarios. Las majestuosas construcciones de aztecas, mayas e incas, así como otros pueblos constructores, implica un alto grado de sofisticación en el manejo de datos. Durante la Edad Media se multiplican los ejemplos de censos, relevamientos de propiedades, registros demográficos (nacimientos y defunciones), recolección de impuestos, etc. En el Concilio de Trento se establece la obligatoriedad de inscribir matrimonios, nacimientos y muertes. en todas las Iglesias. Todo esto muestra como en la historia, paulatinamente, van apareciendo intentos de usar la Estadística como una herramienta, para el manejo de asuntos civiles, comerciales, religiosos, estatales, etc. Es en Alemania donde aparece la llamada escuela Universitaria o alemana de Estadística; en 1660 se crea la primera cátedra universitaria en la Universidad de Helmstedt, donde Herman Cönnig dicta un curso en el que se pasaba revista a los hechos más notables del Estado y a la sistematización de datos y conocimientos. O sea, un enfoque eminentemente adminis-

trativo de la cosa estatal. Su discípulo, Godofredo de Achewall, la separa de la Sociología y la llama por primera vez Estadística.

En Inglaterra, se desarrollaba paralelamente la escuela Demográfica, llamada también la de los “aritméticos políticos”. Pues sus integrantes trataban de obtener leyes empíricas para la cuantificación de los fenómenos políticos y sociales. Se destacan W. Petty, E. Halley y J. Gaunt (1620/74) quien en una memoria a la Royal Society de Londres, destaca la influencia de las estaciones del año sobre la mortalidad, las migraciones de pobladores del campo a la ciudad, la proporción constante de los sexos, y otros conceptos básicos de la moderna Demografía. Por su parte en Francia, se desarrollaba la escuela Probabilística, concentrando sus estudios en los juegos de azar, tanto de cartas como de dados. En 1657, Christian Huyghens, reputado físico de la época, escribió un tratado sobre las probabilidades de éxito y fracaso en juegos de cartas. Si bien esto tenía lugar en Holanda, la idea sobre chances de ganar en los juegos de azar es tan vieja como el hombre. En la India, en el Rig-Veda, se menciona un juego de dados como intento de medir la probabilidad, se trata de un poema literario escrito unos 1000 años antes de Cristo. Sófocles atribuye a un tal Palamedes la invención del juego de dados durante el sitio de Troya. Cuanto más sofisticados se volvían los juegos, mayor era el deseo de estudiarlos para saber cómo ganar. Tal vez esa búsqueda de una “martingala” haya sido el mejor incentivo para mantener vivo el interés en las probabilidades, a lo largo del tiempo. La primera solución a un juego de dados, el pasadiez (se gana si al lanzar dos dados se sacan más de diez puntos) se debe a Galileo. La anécdota histórica más famosa se encuentra en una carta escrita en 1654 por Blas Pascal a Pierre Fermat. En ella, Pascal menciona al Caballero de Meré como “poseedor de un fino espíritu y una gran habilidad en el juego, a pesar de no ser un geómetra”... El Caballero estaba indignado con la aritmética porque según su experiencia de jugador, si uno acepta sacar un 6 en cuatro jugadas tiene una ventaja de acertar mayor, que si acepta sacar dos 6 (hacer "sones"), usando dos dados. Lo que le escandalizaba, era que la probabilidad de ganar no se mantuviese constante, y eso a sus ojos era una contradicción. Lo notable del caso, es que notó en forma totalmente empírica, una diferencia de apenas el 4% en la probabilidad de ganar, que muestra la correcta resolución teórica del problema.

En 1714 aparece “Ars Conjectandi”, obra póstuma de Juan Jacobo Bernoulli, publicada siete años después de su muerte. En este trabajo se relaciona por primera vez la teoría con los experimentos, se define la probabilidad clásica, y se plantea la posibilidad de introducir lo probable en lo social. Esto es, que el Cálculo de Probabilidades tenía otros usos además de los juegos de azar. Su sobrino Nicolás, quien fue el encargado de la publicación post-mortem, luego publica una colección de problemas resueltos que se vuelve muy popular para su época. Por ejemplo, la solución del problema de Montmort, que consistía en calcular la probabilidad de que un jugador gane si al sacar una carta de un grupo de trece, el número de extracción sea igual al número de la carta. La fórmula propuesta por Montmort fue generalizada al caso de n cartas por uno de sus lectores; quien encontró que la probabilidad tendía a una constante si el número de cartas tendía a infinito ($p=0,6321$). El lector se llamaba Euler, y así se descubrió al número e (2,718...). El incremento que toman los estudios acerca de las probabilidades creció vertiginosamente. El francés De Moivre halla la curva matemática de la probabilidad integral, y a principios del siglo XIX, P.Laplace recopila todo lo publicado hasta entonces acerca del tema, más otros descubrimientos de su propia cosecha, en particular la primera versión del teorema fundamental de la Estadística: el Teorema Central del Límite. Sin embargo, este tema sería popularizado por un físico de su misma época: Gauss. Al estudiar los errores de medición cometidos en los experimentos, Gauss

descubrió que mediciones repetidas, bajo condiciones análogas, arrojan diferentes resultados, si los instrumentos son lo suficientemente sensibles. A falta de una mejor explicación para el fenómeno, atribuyó estas variaciones a la “casualidad”, y obtuvo en su estudio de errores una curva teórica en forma de campana que lleva su nombre. La aplicación de la teoría de errores en mediciones experimentales, les dio un carácter de tipo científico, diferenciándolas de las investigaciones biológicas de ese entonces, que se limitaban a describir y clasificar especies, sin entrar a controlar la repetición de fenómenos en laboratorios.

Un discípulo de Laplace, estableció en 1837 el desarrollo matemático de la Ley de los Grandes Números, a partir de la cual comienza la Inferencia Estadística. Además, estudiando la probabilidad binomial para los sucesos raros, encuentra una función que lleva su nombre: Poisson. Esa función es usada por dos colegas de la Universidad de París, para explicar en forma teórica un nuevo fenómeno que habían encontrado. Y es así, que el matrimonio Curie da forma a su Ley de la Radiactividad Natural, semilla primitiva de la Física Estadística y más adelante de la Física Cuántica. La contribución de la escuela rusa, se hizo en el campo de la teoría. Autores como Chebishev, Markoff, Kintchine y otros, completaron la base matemática de la Estadística y es en Rusia, a fines del siglo XIX, donde se publica por vez primera la versión completa del Teorema Central del Límite, luego de casi un siglo de búsqueda.

Tal vez, el primero en aplicar estos conocimientos a la Biología y Medicina en el campo de lo social, fue el belga Adolfo Quetelet. Su idea del hombre promedio sirvió para estudiar la relación entre las dimensiones humanas. Su obra “Sur L’Homme”, originó la Antropometría y además se ocupó de obtener datos acerca del número de suicidios, delitos, etc. Sin embargo, el llamado Padre de la Bioestadística fue un inglés: Sir Francis Galton (1822-1911). Era primo de Ch. Darwin y estimulado por la publicación del Origen de las Especies intentó hallar teorías de tipo genético para resolver los problemas de herencia en los humanos. Si bien no pudo hallar las respuestas a esas cuestiones, introdujo un método matemático para el ajuste de curvas a puntos experimentales: el de los mínimos cuadrados. Este método lo usó en sus estudios de la herencia de padres a hijos. La propuesta era que hijos de padres más altos que el promedio de la población eran más bajos que sus padres; viceversa, hijos de padres bajos, crecían más que sus progenitores. O sea, la población humana, tiende al promedio de alturas en generaciones sucesivas. Por eso al método se lo llama: Regresión. Hoy se ha verificado esto y se sabe que el promedio de altura de los humanos va en aumento con los siglos, debido principalmente, a una mejor calidad y cantidad de alimentación. El heredero de la cátedra de Galton y seguidor de sus trabajos: K.R. Pearson, descubre la distribución χ^2 y funda la primer revista científica del tema: *Biometrika* Donde los investigadores de la época publicaban sus descubrimientos, como W. Gosset que usaba el seudónimo de Student, y tal vez la figura más grande del siglo en el tema: Ronald A. Fisher (1890-1962) creador del 50% de la Bioestadística actual. Su método de Análisis de Varianza es la herramienta básica para las investigaciones biológicas. Desarrolló la fórmula matemática de la función imaginada por Gosset, a la que llamó distribución *t* de Student, que reemplazó en la práctica al centenario método gaussiano. Fue el creador de muchas cosas más: el método estadístico para Diseñar Experimentos, el método de máxima verosimilitud para los estimadores, y lo buscado por Galton: la *Genética de Poblaciones*. Contribuciones posteriores como las del norteamericano Snedecor con su distribución *F* (por Fisher) completaron esta rama de la Estadística a principios de siglo. Luego de los treinta, en la Universidad de Duke (USA), O. Rhine aplica el método estadístico con singular éxito, en sus estudios de fenómenos paranormales, cuantificando

así tales fenómenos. Sus éxitos desataron imitaciones en todo el mundo. Durante la guerra, grupos multidisciplinarios dedicados a resolver problemas de logística, desarrollan nuevas aplicaciones: la Investigación Operativa. Se van imponiendo los nuevos modelos no-paramétricos para estudios en Psicología, Sociología y tantos otros campos. Se desarrolla la Econometría para los planificadores. Y hoy día siguen apareciendo novedades y mejoras de todo tipo, junto al inestimable aporte de las computadoras, en simulaciones y demás cálculos cuya complejidad, los tornaban inalcanzables para la mayoría hasta hace pocos años.

1.3 Magnitudes y variables

Es frecuente el uso de estas dos palabras como sinónimos. Si bien a veces esto es correcto, conviene puntualizar la diferencia conceptual entre ambas: La *magnitud no depende del sistema de referencia, la variable sí*. Por ejemplo, la masa de un cuerpo es igual en todo el Sistema Solar, pero su peso varía según el planeta donde esté ubicada. Un hombre pesa en la Luna casi un sexto de lo que pesa en la Tierra, pero su masa sigue siendo la misma. Entonces la magnitud es la masa, y la variable el peso. Una característica del hombre, como la de tener sangre caliente es una magnitud. Tomando como población a toda la especie Homo-sapiens, es una constante para todos los seres humanos, por lo tanto no es una variable, sino un *parámetro*. Pero si se toma como población a todos los habitantes del planeta, al comparar con los peces se detecta una diferencia con el hombre, deja de ser un parámetro y esa misma magnitud es ahora una variable. El orden *cordiano* es un parámetro para la familia *mirtáceas* (variable para esa población). A su vez, si se considera a esa familia como la población en estudio, ahora pasa a ser un parámetro y la especie *Eucalyptus* una variable dentro de la misma. Cualquier característica de un elemento de la población que pueda observarse puede ser medida, ya sea con instrumentos como la glucosa, el colesterol, peso, talla, etc.; ya sea con los sentidos como color, olor, sabor, nivel de ruido y textura; o bien, se puede clasificarla en clases para hacer recuentos como sano-enfermo, positivo-negativo, etc. En síntesis:

Magnitud: *Es toda característica, o cualidad, de un elemento integrante de la población, susceptible de ser observada.*

Variable: *Es toda magnitud que permite diferenciar entre sí a los componentes de una misma población. Matemáticamente se expresa como una función.*

Parámetro: *Es toda magnitud que tiene el mismo valor dentro de una población. O sea, no permite diferenciar entre sí a sus elementos componentes.*

Medir: *Es comparar una magnitud con otra de su misma especie, considerada como referencia o patrón. Ya sea usando instrumentos, o bien, por medio de los sentidos.*

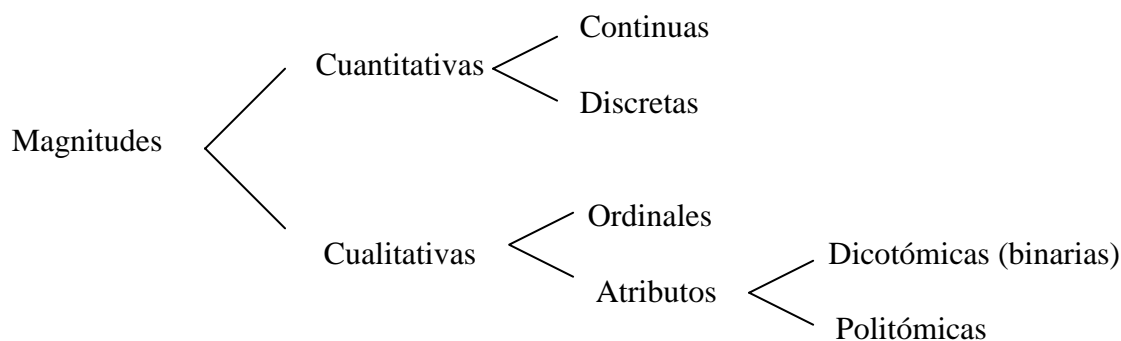
Valor o Dato: *Es el resultado al medir una magnitud en un elemento de la población.*

De las definiciones anteriores surge que una *muestra* puede ser interpretada como un conjunto de valores que puede adoptar una magnitud. Y cuando tal conjunto, está formado por todos los valores posibles de la magnitud, entonces se trata de la *población*. Mientras que si se tiene un único

valor constante para toda la población, se puede tomar como un *parámetro* o como una muestra de un único elemento. En la población masculina, el sexo es un parámetro, o sea, una *magnitud constante*; mientras que si la población es la especie humana, el sexo es una *magnitud variable*. Generalmente, una variable se expresa usando letras mayúsculas como **X**, **Y**, **Z** etc. Mientras que para los datos se usan las minúsculas con subíndice, como x_i , y_j , z_k . Al igual que en Matemática, la nomenclatura técnica para una variable y su conjunto de valores se expresa con: **X** ($x_1, x_2, x_3, \dots, x_n$). Si el peso de un individuo es de 73 kg, entonces ese es un *dato*, o un *valor* que adopta la *variable* al ser medida. Por ejemplo, si ese resultado se tuvo con la pesada undécima de la muestra de individuos escogida, esto se expresa con $x_{11} = 73$ kg.

Cada rama dentro de la Biología presenta un grupo específico de magnitudes propias, como concentraciones de fluidos o de productos químicos en Clínica, medidas morfológicas en Antropometría, frecuencias de factores hereditarios en Genética, las estadísticas vitales en Demografía, los índices económicos en Econometría, y muchas otras más. Es el investigador quien decide cuál magnitud usará en sus estudios y la forma en que usará la herramienta estadística, para analizar los datos obtenidos. La Bioestadística no contribuye al descubrimiento de nuevas magnitudes, sino al tratamiento de la información que hace posible tal descubrimiento. Por ejemplo, un análisis estadístico puede demostrar que el ajuste de los datos de la magnitud elegida, a una curva teórica planteada, no es digna de confianza. Para reemplazarla, el investigador deberá buscar alguna curva nueva usando los conocimientos de su profesión, su intuición, pero nunca podrá usar a la Estadística, para buscar una función que se ajuste mejor a los datos obtenidos, presentándola como si fuese una nueva teoría, pues tan solo es una aproximación empírica.

Las magnitudes pueden ser clasificadas como sigue:



La diferencia entre magnitudes cuantitativas y cualitativas está en la relación que tienen con el patrón o estándar, contra el cual son comparadas al ser medidas. Si esta relación puede ser expresada con números, debido a una proporcionalidad, entonces la magnitud es *cuantitativa*. Por ejemplo, si la altura de un paciente es de 1,7 m, eso significa que es 1,7 veces más largo que el *metro patrón* depositado en París, o sea una proporción numérica real, y por eso la magnitud altura es cuantitativa. En general, todas las magnitudes físicas basadas en el Sistema Internacional de Unidades (metro, kilo, segundo, etc.) son ejemplos clásicos de magnitudes de este tipo. En cambio, una magnitud es *cualitativa* cuando su relación con el patrón no es una proporción numérica. El sexo, el color, el olor, etc. son ejemplos de este tipo de variables. Se pueden codificar con números a los resultados, cuando se miden esas magnitudes, pero eso no quiere decir que la

relación sea numérica. Por ejemplo, si se conviene en codificar con un “2” al sexo masculino y con un “1” al femenino, eso no quiere decir que un hombre valga por dos mujeres.

Las magnitudes *cuantitativas* se clasifican en continuas o discretas. Se diferencian entre sí, porque en la primera la relación numérica con el patrón puede ser cualquiera, mientras que en la segunda hay algunos valores prohibidos. Entonces una magnitud *continua* se expresa mediante números reales. Las mediciones físicas clásicas son el mejor ejemplo de este tipo de magnitudes: peso, talla, densidad, temperatura, concentración, etc. En una magnitud continua hay infinitos puntos posibles dentro de un intervalo cualquiera de la misma, en del dominio de los números reales. En cambio, cuando la magnitud tiene algunos valores que son posibles y otros que no, entonces se trata de una *discreta* (se expresan por lo general con los números enteros positivos). Por ejemplo, las de recuento o enumeración: como el número de hijos, de pacientes, la cantidad de dientes y otras. Allí se usan valores enteros para contar los resultados; no tendría sentido medir 12,75 alumnos en un aula. Existen otros nombres para este tipo de magnitudes: *discontinuas* o *merísticas*.

Las magnitudes *cualitativas* se clasifican en atributos u ordinales. Se diferencian entre sí en que los *atributos* son las cualidades del objeto de la medición, observables sin emplear instrumentos. Como las *organolépticas* (color, sabor, olor, textura y nivel de ruido) donde se usan los sentidos para medir. En cambio, las *ordinales* implican medir el orden de los resultados obtenidos, para luego clasificarlos. Como el resultado de una carrera de autos, un concurso de belleza, un sorteo de lotería, etc. Debe destacarse que una magnitud de tipo continua, puede usarse como magnitud ordinal y aún de atributos. Depende de la convención utilizada para expresar los resultados. Por ejemplo, en un antibiograma se trata de hallar el mejor antibiótico para combatir una infección, el método consiste en extraer del paciente una muestra de la zona infectada y colocarla en un agar gel apropiado, dentro de una caja de Petri; luego de obtener desarrollo microbiano se coloca una arandela conteniendo muestras de los antibióticos a probar, y transcurrido cierto tiempo, la acción antibiótica deja una aureola al atacar al agente infeccioso. Lo que se mide es el diámetro de las diferentes aureolas (cuantitativa), se compara con la tabla de resultados provista por el fabricante, y los resultados se expresan en tres categorías: sensible, muy sensible o resistente al antibiótico (ordinal). La magnitud continua fue agrupada en tres clases, ordenadas de acuerdo con la resistencia al antibiótico. Como esto se repite para todos los antibióticos probados, a cada uno se lo puede considerar como un atributo. Al precisar la talla de una persona, se está midiendo una magnitud de tipo continua, pero si esos datos se usan para clasificarlo en muy alto, alto, normal, bajo o enano, entonces se “ordinaliza” el resultado, expresándolo en forma cualitativa. Es conveniente tener en cuenta, desde el punto de vista de la cantidad de información, que la riqueza contenida en la magnitud continua se va perdiendo al volverla cualitativa, mediante algún tipo de convención.

Las magnitudes de *atributos* se clasifican en dicotómicas o politómicas. La diferencia se basa en si el atributo puede adoptar dos o más valores diferentes. *Dicotómicas* son aquellas que pueden tener sólo dos valores posibles, como: el sexo, el factor sanguíneo Rh, la ausencia o no de cierto microbio, todas las reacciones de tipo (+) o (-), levógiro o dextrógiro, sano o enfermo, éxito o fracaso, etc. Por su parte, las *Politómicas* son aquellas magnitudes que pueden tener más de dos resultados posibles, como: el grupo sanguíneo, la raza, el fenotipo, análisis de sangre en heces, marcas de autos en plaza, tipos de bocas de expendio de mercaderías al público y otras. Nuevamente, conviene resaltar que toda magnitud cualitativa puede ser dicotomizada como las

continuas. Simplemente se toma un resultado y se lo ubica en una de dos categorías, según si verifica o no un atributo dado. Por ejemplo, para clasificar a un varón recién nacido como “robusto” verificamos que pese más de 4 kg al nacer; todo aquel que pesó menos será clasificado en la otra categoría como “no robusto”. Naturalmente, que al dicotomizar una variable se pierde información y eso atenta contra la calidad de las estadísticas obtenidas con esos datos. El hecho de saber que un recién nacido es “robusto” no indica cuál fue su peso al nacer, lo único que se sabe es que pesó más de 4 kg. O bien, si su CPK (creatino-fosfo kinasa) es mayor que 21 UI se conviene en suponerlo afectado por un infarto de miocardio. En grupo sanguíneo se puede considerar al 0 como dador y a los demás como no-dador.

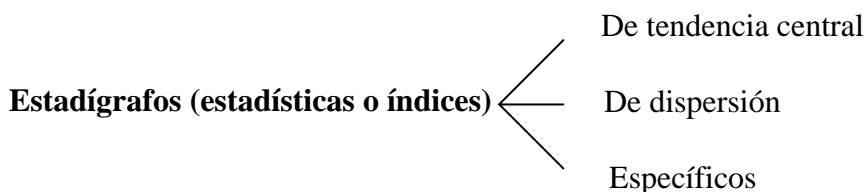
Una magnitud *compuesta* es aquella que relaciona dos o más magnitudes *simples*. Así, dos magnitudes simples como: el peso de una glucosa y un volumen de agua, al mezclarlas se transforman en una compuesta: la concentración de glucosa en agua. Por regla general, las determinaciones de química clínica son ejemplos de este tipo. Por su parte, en farmacia se usan para casos como: ventas mensuales, ganancia diaria, etc. También se puede tener magnitudes compuestas, relacionando magnitudes cualitativas. Por ejemplo, si se está determinando en una población la cantidad de individuos que sufrieron enfermedades infantiles típicas, esta medición tendrá datos del tipo: contrajo varicela, no contrajo paperas, y otros valores del cuestionario, entonces el cruzamiento de esos datos genera varias variables. La idea de magnitudes compuestas está asociada a mediciones multidimensionales; para estudiar sus resultados se suele emplear el análisis estadístico multivariado. A veces se mezclan dos tipos de magnitudes diferentes, como el recuento de rojos, blancos o plaquetas, donde se tiene una magnitud discreta en el numerador y una continua en el denominador. En estos casos, se suele considerar a la resultante como si fuese una magnitud de tipo continuo. Todo depende de la convención adoptada por el profesional.

1.4 Estadígrafos

Se llaman *estadísticas o estadígrafos* a los valores que cuantifican características de un grupo de datos, tales como la dispersión de los mismos, la tendencia a concentrarse alrededor de un cierto punto, etc. Estos *índices* muestran las características más salientes del conjunto de datos con un solo número. Y es a estos números, a los que tradicionalmente se los llama “estadísticas”. Aquí se prefiere el uso de la palabra “estadígrafo” por dos razones: una, para no confundir este concepto con el de toda la materia en estudio; otra, porque en realidad es un valor que ilustra (“gráfica”) una característica saliente del grupo de datos. Por lo general, la información que se extrae del sistema aislado para su estudio es muy voluminosa. A menos de contar con computadoras para su procesamiento en forma electrónica, el manejo de grandes cantidades de números se torna difícil, engorroso y con grandes posibilidades de cometer equivocaciones. Por estos motivos, desde sus comienzos, el problema general de la Estadística fue buscar la manera más conveniente de *condensar* el volumen de datos a unos pocos números (medidas), que fueran indicadores de la conformación de todo el conjunto.

Condensando la “nube” de datos, en unas pocas “gotas”, ricas en información, se puede facilitar enormemente el manejo de las cantidades ingentes de datos. Y además, la transmisión de los resultados obtenidos, a otros investigadores se facilita y resume. Si se piensa en como publicar el resultado de un experimento, del cual se extrajeron cientos de mediciones, la forma más

engorrosa sería hacerlo literalmente. Esto es, escribir cada número y transmitir en forma “cruda” toda esa información. Una manera más práctica sería con una tabla donde los valores fueran agrupados en pocas columnas empleando sus frecuencias de aparición o sus totales. Un gráfico ilustraría mucho mejor, pues de un solo vistazo se abarcaría el grupo en su totalidad, a costa de perder precisión. El problema hasta ahora sigue siendo el de siempre: se da toda la información en detalle, o se agiliza el trámite, resumiéndola a unos pocos indicadores: *los estadígrafos*. Este tipo de indicadores se puede agrupar en tres clases principales:



Los *estadígrafos de tendencia central* indican alrededor de cuál valor se agrupan los datos obtenidos. La media, la mediana, la moda son los más comunes y conocidos indicadores de este tipo. Si se graficaran todos los valores, esa especie de “nube” de puntos se distribuiría alrededor de una zona central, donde se ubican estos índices. La media o promedio suele ser la mejor medida para cuantificar esta zona, una especie de centro de “gravedad” de esos datos. De hecho, en Física, el centro de masa se define usando el promedio de las masas ponderado con la distancia. Dentro de este grupo entran los percentiles: la mediana es el percentil 50, o bien el segundo cuartil, o bien el quinto decil.

Los *estadígrafos de dispersión* indican qué tan esparcidos están los datos obtenidos. El rango, el desvío estándar y la varianza, son los índices más conocidos de este grupo. Se debe definir la dispersión de los datos respecto de algo. Conviene hacerlo respecto de la media aritmética (como lo hace el desvío estándar) porque minimiza los cuadrados de los desvíos, tal como se verá más adelante. Otro estadígrafo asociado a estos dos tipos mencionados es el coeficiente de variación, que es el cociente entre el desvío estándar y la media. En Física, este mismo concepto es el error relativo de una medición, y en bioquímica se lo asocia con la calidad del test clínico: cuanto menor es el coeficiente de variación, mejor es el test.

Los *estadígrafos específicos* son propios de cada especialidad. El índice del costo de vida, el de la construcción y el de la inflación, son los más conocidos en Economía. Para Bioquímica y Medicina se emplean índices como Sensibilidad, Especificidad, Prevalencia, Fecundidad, Rendimiento, etc. Ayudan en la toma de decisiones. En Bioquímica por ejemplo, se los emplea como criterios, para decidir cuál análisis clínico conviene más en la detección de una cierta enfermedad. Hasta la década del setenta, era costumbre seleccionar las técnicas de análisis de acuerdo a los criterios clásicos de precisión y exactitud, tomados como sinónimos de calidad. Desde entonces, en muchos casos se han impuesto los criterios de sensibilidad y especificidad, pues interesa más la capacidad para diagnosticar del método, que su precisión y exactitud. En particular, cuando se miden magnitudes cualitativas donde es imposible usar los criterios clásicos. En Sanitarismo se emplean índices demográficos como las tasas de Natalidad, Mortalidad, Morbilidad, Nupcialidad, etc. También se emplean índices hospitalarios como: Camas ocupadas, Tiempo de per-

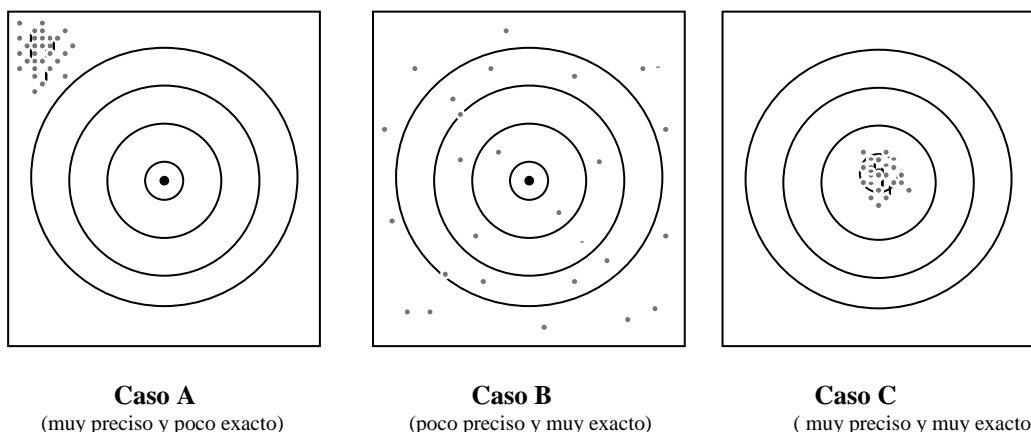
manencia, Prestaciones, etc. En Farmacia se usan los índices de tipo económico y los de tipo industrial, como porcentaje de fallas, rotación de stock, etc.

En el capítulo cuatro se verán estos índices con más detalle.

1.5 Precisión y exactitud

En el lenguaje común se suele usar la palabra *precisión* como sinónimo de *exactitud*. Sin embargo, en estadística son conceptos bien diferentes. Si se mide una magnitud patrón n veces, con un instrumento adecuado, se obtienen n valores que difieren entre sí. La dispersión de estos valores tiene diferentes causas y cuanto menor sea, mayor será la *precisión* del instrumento de medición. Por su parte, cuanto más cercano esté el promedio de los valores al valor del patrón, mayor *exactitud* tendrá el instrumento. Para ilustrar estas ideas, nada mejor que usar un símil, como el del tiro al blanco. En él se supone que: el centro del blanco es el valor patrón, cada impacto es una medición realizada, y la pistola es el sistema de medición. Se puede apreciar en la Figura 1.1 tres ejemplos de una serie de disparos sobre el blanco. En el caso **A**, se puede ver que los impactos están muy cercanos entre sí pero muy lejos del centro, y por eso se dice que hay mucha precisión, pero poca exactitud. En el caso **B**, se ve que los impactos están muy dispersos pero rodeando sistemáticamente al centro, o sea que su promedio resultará muy cercano al mismo. A este caso se lo califica como de poca precisión y mucha exactitud. Y en el caso **C**, el ideal, se tiene un grupo muy compacto de disparos en el centro del blanco, hay mucha exactitud y mucha precisión. Haciendo la analogía entre la serie de disparos y una serie de mediciones, se pueden captar mejor las ideas de precisión y exactitud. Cuanto más cercanos entre sí estén los datos, mayor será la precisión. Cuanto más cerca esté el promedio del centro, valor verdadero de la magnitud, mejor será la exactitud.

Figura 1.1 Símil de tiro al blanco, para ilustrar los conceptos de precisión y exactitud



Para cuantificar estos conceptos en clínica, hay varios métodos. El más común es tomar un suero patrón -un estándar calibrado- y fraccionarlo en n alícuotas. Luego se mide cada alícuota y se tienen n valores, se calcula el promedio y el desvío estándar de esos datos. La diferencia entre el valor medio y el suero patrón da una idea de la exactitud del método y permite cuantifi-

carlo con el llamado error sistemático. Lo mismo se puede hacer para controlar instrumentos; por ejemplo, usando pesas patrón se pueden testear balanzas, comparando el promedio de una serie de pesadas repetidas del patrón con el valor del mismo, se determina si está calibrada. Cuanto menor salga esta diferencia, tanto mejor será la exactitud de la balanza. Usando sueros controles en forma análoga se puede controlar la exactitud de cualquier sistema de medición clínico.

A su vez, la precisión se cuantifica usando como base el desvío estándar del método. En Clínica para saber si la dispersión de un método es razonable, se lo compara con un valor límite descrito en la bibliografía, o empleando algún criterio como el de Thonks o el de Aspen. Así, comparando el desvío estándar del método con un valor aceptable -desde el punto de vista clínico- se puede juzgar la precisión del mismo. Otra forma más elaborada, es diluir el suero control para obtener un grupo de concentraciones diferentes, y luego, repetir para cada una, el método anterior. De esa forma, se tiene un panorama del comportamiento del sistema de medición en un rango de valores, similar al que se tiene en la práctica diaria de la profesión. Más adelante se detallarán mejor estos procedimientos.

En la industria, se han diseñado controles mecánicos de precisión. Como en la fabricación de los cojinetes para rulemanes. Estas esferas de acero, no pueden ser ni más grandes, ni más chicas que la diferencia entre los radios de las arandelas que las contienen; caso contrario, no se podrá armar el rulemán. A la salida de la máquina productora de los cojinetes de acero se los hace circular por una zaranda con agujeros calibrados con el límite mayor. Las que no pasan por estos huecos se descartan, las otras, caen a otra zaranda ubicada debajo, con agujeros calibrados para el límite inferior. Las que quedan dentro de esta zaranda son las buenas y las que pasan por los huecos son desechadas, por ser demasiado pequeñas. De esa manera, se controla en forma permanente, a medida que se fabrican, la precisión de la máquina industrial. Con el mismo criterio, en los laboratorios de análisis clínicos se implementan las Cartas de Control de Calidad para tener un control diario.

En la industria farmacéutica se hace un control de calidad mucho más exhaustivo que en otras industrias. Lo principal es controlar las dosificaciones de los medicamentos producidos. Para ello se seleccionan al azar, cierta cantidad de unidades de todo el lote de producción. El número de unidades elegidas -la muestra- se decide sobre la base del tamaño del lote para poder trabajar con una confianza aceptable. Luego se analiza el contenido de cada uno de sus componentes activos, dentro de ciertos límites de tolerancia. Si el porcentaje de “fallados” es muy alto, no queda más remedio que rechazar todo el lote, para impedir que salga al mercado un producto poco satisfactorio, que podría dañar por lo menos la imagen de la empresa.

El punto débil del uso de estos conceptos, en la práctica, es su inaplicabilidad en el caso de tener que trabajar con variables de tipo cualitativo. En estos casos no hay ni valor promedio, ni desvío estándar. La solución es trabajar con porcentajes. En efecto, si la ausencia de cierto atributo se puede clasificar como “falla”, entonces el porcentaje de fallados pasa a ser una variable de tipo continuo y se subsana el inconveniente. La idea es controlar que el porcentaje de fallados no exceda ciertas normas de calidad definidas por el fabricante, las cuales deben estar en un todos de acuerdo con las normas establecidas por la autoridad sanitaria, para que los mismos puedan ser vendidos al público. Esto implica una serie de estudios muy complejos en poblaciones humanas, realizados de acuerdo a estrictas normas internacionales.

1.6 Cuestiones clínicas

Uno de los principales objetivos tanto de Bioquímica como de Farmacia, es prestar ayuda al médico en el tratamiento de sus pacientes. Entonces parece prudente explicitar los cuatro tipos fundamentales de cuestiones clínicas que ellos necesitan atender en su trabajo diario:

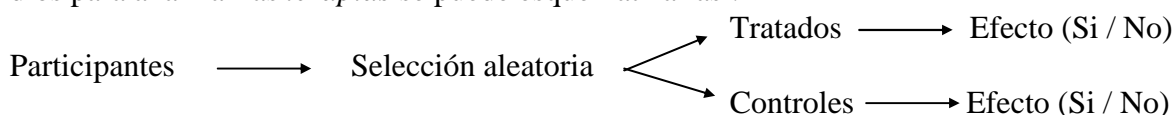
- *Terapia*: Determinar el efecto de diferentes tratamientos para mejorar el estado del paciente, o para evitar los efectos adversos involucrados.

- *Daños*: Discernir los efectos de potenciales agentes dañinos (incluyendo las terapias varias que están interesados en examinar de acuerdo al punto anterior) en el estado del paciente, morbilidad y mortalidad.

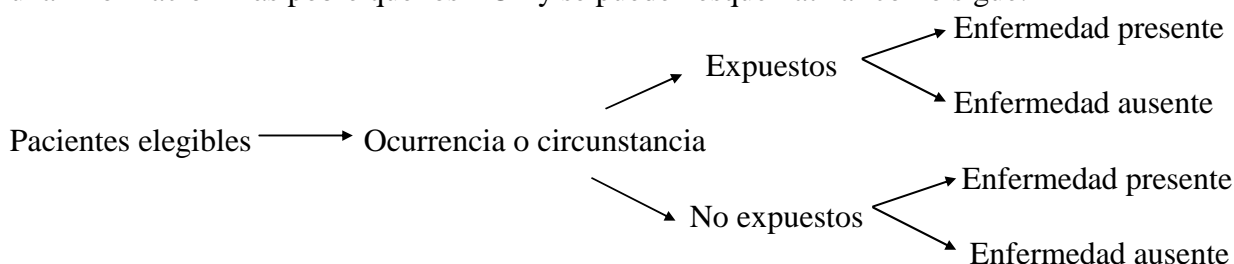
- *Diagnóstico*: Establecer la capacidad de una intervención para poder diferenciar entre aquellas con y sin la condición buscada o enfermedad.

- *Pronósticos*: (prognosis) Estimar el curso futuro de la enfermedad del paciente.

Para contestar la primera cuestión acerca del tema terapéutico se usan estudios denominados *Ensayos aleatorios controlados* (“Randomized Controlled Trials”: RCT). Esta investigación se diseña de la siguiente manera: los participantes se dividen en dos grupos, y para ello se los escoge en forma aleatoria. A uno de los grupos se le aplicará un tratamiento experimental para estudiar el efecto del mismo (*tratados*), mientras que al otro grupo se le aplica el tratamiento estándar o usual (controles o placebos). Luego se sigue la evolución de cada individuo en el tiempo, para observar si aparece el *efecto buscado* o bien no aparece. Por ejemplo, si se le aplica una nueva vacuna contra la gripe al grupo tratado, y el tratamiento usual al grupo control. Este tipo de estudios para analizar las *terapias* se puede esquematizar así:



Idealmente, se podría hacer el mismo tipo de estudios para analizar el daño que puede infligir cierto factor de riesgo. Esto es, en lugar de aplicarle a cada individuo un tratamiento preventivo, se le inocula un agente infeccioso o dañino, para ver que le pasa. Pero la ética impide desarrollar este tipo de estudios, por lo que se encara de otra manera. Por ejemplo, para ver el efecto de fumar con relación a enfermedades pulmonares, no es práctico diseñar un estudio donde se obligue a los individuos a fumar durante los próximos 10 años para ver si desarrollan un cáncer u otro tipo de complicaciones pulmonares. Pero sí, se pueden elegir individuos que han fumado regularmente durante los últimos 10 años para ver si hubo *daños*. Estos estudios observacionales dan una información más pobre que los RCT y se pueden esquematizar como sigue:



Para poder contestar la tercera cuestión sobre los diagnósticos, se trata de establecer que tan bien trabaja un método para diagnosticar, o bien un test clínico. Aquí se necesita otro tipo de estudios basados en el estudio de sus propiedades, o características operacionales. Estos estudios se diseñan para saber que tan buena es la *calidad* del test clínico:

Estudios de calidad: En estos casos, el investigador tiene que identificar a un grupo de individuos que pueden tener una cierta enfermedad o condición de interés, de los otros que no la tienen (por ejemplo: tuberculosis, Chagas, sífilis, etc.). Esta es la *condición buscada*, o bien los *positivos*. Entonces se comienza juntado un grupo de personas que el investigador sospecha que podrían tener esta condición. Luego a cada uno se le efectúa un test clínico y a la vez otro test considerado el “golden standard” o método de referencia (el cual de acuerdo a los conocimientos actuales es el mejor para detectar esa enfermedad). Con todos esos datos el investigador compara los resultados hallados con el test común contra los de referencia, para poder evaluar la capacidad, o *calidad* del test clínico investigado. De esta forma se forman cuatro conjuntos independientes de resultados: los positivos confirmados por el test de referencia considerados como los verdaderos positivos, y los no confirmados considerados como falsos positivos, y análogamente con los negativos que pueden ser verdaderos o falsos. Se agrupan estos datos en una tabla llamada Tabla Diagnóstica o Tabla de la verdad, del tipo doble dicotómica o tabla 2 x 2 de contingencia, como se muestra en el esquema siguiente:

Tabla 1.1. Tabla diagnóstica

Resultados del test	Resultados verificados		Total
	Enfermos	Sanos	
Positivo (+)	<i>vp</i> <i>verdadero positivo</i>	<i>fp</i> <i>falso positivo</i>	TP
Negativo (-)	<i>fn</i> <i>falso negativo</i>	<i>vn</i> <i>verdadero negativo</i>	TN
Total	TE	TS	N

Donde N es el número de sujetos investigados y

TP = $vp + fp$: Total de sujetos diagnosticados positivos

TN = $vn + fn$: Total de sujetos diagnosticados negativos

TE = $vp + fn$: Total de sujetos enfermos, o que poseen la condición buscada

TS = $fp + vn$: Total de sujetos no enfermos (“sanos”), o que no poseen la condición buscada

Este tipo de estudios de calidad se puede esquematizar de la manera siguiente:

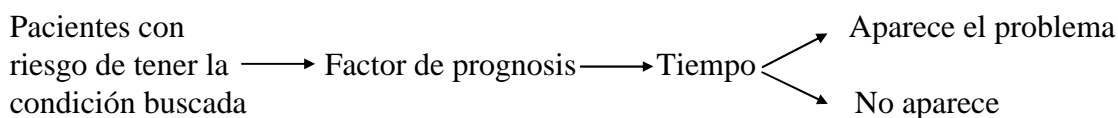
Pacientes bajo

sospecha de tener la condición buscada \longrightarrow Test clínico \longrightarrow Test de referencia \longrightarrow Tabla diagnóstica

El caso anterior es el más sencillo de todos porque se trata de estudios donde los resultados del diagnóstico se dicotomizan en positivos y negativos, pero puede ocurrir otro tipo de situaciones donde los resultados no son binarios como por ejemplo: positivo, dudoso y negativo. Una manera

de salvar esta situación es sencillamente descartar los casos dudosos y trabajar únicamente con los casos de diagnóstico claro para el clínico. O bien esperar el paso del tiempo para confirmar.

Finalmente, para tratar la cuarta cuestión de la prognosis se desarrollan estudios observacionales para poder identificar factores que puedan modificar ese pronóstico. La investigación se basa en identificar pacientes que pertenecen a determinado grupo, tal como: embarazadas, pacientes que serán sometidos a cirugía, o pacientes con cáncer, etc.; con o sin factores que puedan modificar su prognosis (tal como: edad, profilaxis, etc.). El factor de exposición aquí es el tiempo, y el investigador sigue observando a los pacientes para ver la aparición de signos de la condición buscada, tal como: problemas en el recién nacido al término del embarazo, infartos de miocardio luego de la cirugía, o supervivencia al cáncer.



En general, los resultados obtenidos en las cuestiones de terapia, daños y prognosis se pueden agrupar en una tabla de contingencia del tipo 2 x 2, llamada Tabla de riesgo como se muestra a continuación:

Tabla 1.2. Tabla de riesgo

Factor de riesgo	Resultados observados		Total
	Enfermos	Sanos	
Expuesto	<i>a</i>	<i>b</i>	TE _x
No expuesto	<i>c</i>	<i>d</i>	TnEx
Total	TE	TS	N

Donde N es el número de sujetos investigados y

TE_x = a + b : Total de sujetos expuestos al factor de riesgo, inmunizados, o protegidos.

TnEx = c + d : Total de sujetos no expuestos, o bien sin inmunizar o proteger.

TE = a + c : Total de sujetos enfermos, o que poseen la condición buscada

TS = b + d : Total de sujetos no enfermos (“sanos”), o que no poseen la condición buscada

En el caso de *terapias*, un sujeto “expuesto” significa que ha recibido el tratamiento que se está estudiando, mientras que un individuo “no expuesto” es uno del grupo control. A su vez, un individuo clasificado como “enfermo” es uno en el cual se ha observado la presencia del efecto buscado, y “sano” es uno donde no se ha verificado dicha presencia.

En el caso de *daños*, un sujeto “expuesto” es uno que por diversas circunstancias se expuso en su pasado al factor de riesgo investigado (como fumar, radiación, etc.). Y análogamente al anterior, cuando se observa la aparición del efecto buscado se lo clasifica como “enfermo”.

En el caso de *prognosis*, cuando la condición o estado del individuo es el que se quiere analizar se lo califica como “expuesto” (embarazada, operado, etc.). Luego de cierto tiempo se observa si aparece o no el efecto (problema en el recién nacido, infarto, etc.) y se califica como “enfermo”.

Cuando se desean comparar dos tests clínicos para saber cual de ambos es el mejor, la manera adecuada de hacerlo es realizar un estudio de calidad diagnóstica para cada uno, y luego comparar los resultados para poder optar por el de mayor calidad, de acuerdo a la enfermedad involucrada. A este procedimiento para comparar dos métodos clínicos se lo denomina: *comparaciones independientes* para diferenciarlo de otro más económico, donde a cada paciente se le efectúan dos mediciones, una con cada método, y no se necesita de sueros controles ni de métodos de referencia que son más caros que un procedimiento estándar. Se llaman *comparaciones apareadas*, a los estudios realizados aplicando dos tests clínicos a un mismo paciente. Se debe tener muy en cuenta que el objetivo de este tipo de estudios no es determinar cual método es el mejor, sino tan solo ver si uno de ellos (el método viejo) puede ser reemplazado por el otro (el nuevo).

En el caso de medir magnitudes continuas la idea central es ver si ambos métodos tienen la misma exactitud, usada aquí como sinónimo de calidad. Esto es, ver si los pares de mediciones individuales son iguales, o bien si la diferencia entre ambos es cercana a cero. Para poder validar los resultados se emplean modelos estadísticos especiales para muestras apareadas, como se verá más adelante. Para comparar la precisión en muestras apareadas, la única manera es contrastar los resultados de cada método por separado, contra un valor de tablas para ver si pueden ser aceptables. Para ello, se adopta un valor llamado: la dispersión máxima admisible (tabulada al final del capítulo 25). Pero lo que no se puede hacer, es comparar ambos valores de precisión entre sí porque las muestras no son independientes. En cambio en exactitud se hace un truco, porque lo que se estudia es la diferencia de las dos mediciones y así queda un único valor por cada paciente, los cuales sí son independientes.

En el caso de medir otro tipo de magnitudes, que no sean continuas, la idea es transformarlas en un caso dicotómico o binario, adoptando un punto de corte adecuado. Ahora bien, como cualquier magnitud puede ser dicotomizada, el estudio de los casos binarios es el más general de todos, porque abarca todas las mediciones clínicas. Por ejemplo, no se puede determinar la exactitud o precisión en magnitudes cualitativas, mientras que sí se puede hacerlo en continuas; pero así como se puede transformar una magnitud cualitativa en binaria, lo mismo ocurre con las continuas. Por ejemplo, si se está midiendo el colesterol de un paciente se pueden agrupar los resultados hallados en dos grupos: colesterol alto o bajo, si se mide el peso se pueden clasificar en flacos y gordos adoptando un valor de corte apropiado (menos de 70 kg: flacos y 70 kilos o más: gordos, para varones entre 18 y 22 años). Si se mide un antibiograma los resultados se pueden clasificar en microorganismo sensible o resistente al antibiótico. En el resultado de una determinación de sangre en heces se puede adoptar la clasificación: con y sin rastros de sangre. Los casos binarios son muy comunes en microbiología, bacteriología, micología, etc. En toda decisión clínica a ser adoptada después de aplicar un método de diagnóstico, el resultado es binario como por ejemplo: se interna al paciente o no, se lo medica o no, se lo envía a cirugía o no, etc.

El objetivo central de diseñar el estudio clínico en forma apareada para magnitudes binarias, es decidir si se pueden intercambiar dos métodos estudiando la *concordancia* entre ambos. La forma usual de presentar los resultados en estos casos se muestra en la tabla siguiente:

Tabla 1.3. Tabla de concordancia

Método 2	Método 1		Total
	(+)	(-)	
(+)	<i>a</i>	<i>b</i>	<i>a + b</i>
(-)	<i>c</i>	<i>d</i>	<i>c + d</i>
Total	<i>a + c</i>	<i>b + d</i>	<i>N</i>

Donde el número total de concordancias está dado por la cantidad ($a + d$) y el número total de discordancias por la cantidad ($b + c$). Siendo N el número total de casos analizados. Además,

a : es el número de individuos que resultaron positivos con ambos métodos.

d : es el número de individuos que resultaron negativos con ambos métodos.

c : es el número de individuos que resultaron positivos con el método 1 y negativo con el otro.

b : es el número de individuos que resultaron positivos con el método 2 y negativo con el otro.

Las Tablas 1.1, 1.2 y 1.3 esquematizan en forma general los diferentes tipos de estudios clínicos. Más adelante se verá en más detalle a cada una de ellas.

1.7 Problemas propuestos

- | | | |
|---|---|---|
| 1) La estadística es una ciencia para el tratamiento de los datos obtenidos. | V | F |
| 2) El significado de Bioestadística es el mismo que el de Biometría. | V | F |
| 3) El objeto de la Estadística es tomar decisiones con base científica. | V | F |
| 4) Un sistema en estudio es un conjunto de elementos interrelacionados sin restricciones. | V | F |
| 5) La población de la familia <i>rosáceos</i> es la clase <i>angiospermas</i> . | V | F |
| 6) En una población de enfermos de HIV el sexo es un parámetro de la misma. | V | F |
| 7) Las mujeres con síndrome de Turner, no son un parámetro en la población femenina. | V | F |
| 8) Los 5000 casos de gripe analizados, conforman la población en estudio. | V | F |
| 9) Los 20000 enfermos seleccionados constituyen una muestra de la población argentina. | V | F |
| 10) Todo estudio biométrico debe tener un marco de referencia. | V | F |
| 11) Analizando las muestras se pueden inferir los parámetros poblacionales. | V | F |
| 12) Magnitud y variable difieren en que la primera depende de la población. | V | F |
| 13) Medir es comparar una magnitud con otra de su misma especie. | V | F |
| 14) Un elemento de la población que ha sido medido es un dato. | V | F |
| 15) La diferencia entre las cuantitativas y cualitativas es su relación con el patrón. | V | F |
| 16) Una magnitud cualitativa no puede ser expresada con números. | V | F |
| 17) Las magnitudes discretas se expresan con los números enteros positivos. | V | F |
| 18) El peso no es una magnitud continua porque no puede ser negativo. | V | F |
| 19) Dar 10 ejemplos de cada tipo de magnitud. | | |
| 20) La cantidad de grageas de un frasco es una magnitud continua. | V | F |
| 21) El estado civil de una persona es una magnitud dicotómica. | V | F |
| 22) La fragancia de una planta medicinal es una magnitud organoléptica. | V | F |
| 23) La concentración de un jarabe es una magnitud compuesta. | V | F |

36) Clasifique las magnitudes siguientes:

- Número de libros en biblioteca.
- Cantidad de dientes.
- Stock de mercaderías
- Concentración de sustancias
- Droguerías proveedoras de una farmacia
- Laboratorios proveedores de reactivos clínicos
- Raza, religión, color de ojos, color de cabello
- Reacción en bacterias Gram. +
- Resultados en: Carrera de autos, olimpiadas, torneo de esgrima, concurso de belleza
- Factor sanguíneo: Rh
- Grupo sanguíneo
- Ventas en pesos
- Ventas en unidades

37) Unir con flechas los conceptos que se correspondan entre sí:

Medir	Es un conjunto de elementos que se aíslan para su estudio en función de sus interrelaciones
Magnitud cuantitativa	Es comparar un valor con otro de su misma especie considerado como Referencia o patrón. Ya sea usando instrumentos, o bien por medio de los sentidos.
Población	Se pueden clasificar en continuas o discretas.
Sistema	Son los valores que cuantifican características de una serie de datos.
Exactitud	Es el conjunto de todas las muestras posibles que pueden obtenerse del sistema en estudio, de acuerdo con el método de selección usado
Estadígrafos	Es toda magnitud cuyo valor es constante en toda la población.
Estadística Descriptiva	Es el resultado de hacer una medición.
Variable	Está relacionada con el error sistemático de las mediciones.
Parámetro	Es lo que se estudia en este primer tema.
Dato	Es toda magnitud que permite diferenciar entre sí a los componentes de una misma población. Matemáticamente, se expresa como una función.
Verdadero positivo	Es todo resultado de un test que da positivo.
Positivo	El todo resultado de un test verificado como positivo

38) Completar las conceptos siguientes:

- Las magnitudes cualitativas se clasifican en ...
- El error casual se relaciona con los estadígrafos de ...

- Las magnitudes cuantitativas se clasifican en ...
- Las variables dicotómicas tienen solo ... valores posibles.
- Las variables politómicas tienen valor posible.
- Los sirven para condensar una “nube de datos”.
- Las tres características de un sistema son :
- Estadística es un de métodos científicos para la recopilación, , condensación y análisis de los datos extraídos de un en estudio.
- Los estadígrafos de posición son:
- Los estadígrafos de dispersión son:
- Los índices clínicos más comunes son:
- Biometría es la aplicada en Ciencias Biológicas.
- El sistema en estudio se convierte en la o de referencia.
- El de la población se saca contando el número de componentes que tiene.
- La es un subconjunto de la población.

39) Definir las cuatro cuestiones clínicas básicas que necesita atender un médico en su práctica diaria.

40) ¿ Cómo se diseñan las investigaciones clínicas para poder estudiar la respuesta a cada una de las cuestiones del punto anterior ?

41) Explicar como se obtiene una Tabla de la verdad a partir de una investigación clínica, sobre la calidad de un test clínico, o bien un método de diagnóstico.

42) ¿ Cómo se clasifica a un individuo que dio positivo con el test clínico y sin embargo esto no resulta confirmado con el test de referencia o “golden estándar”?

43) Análogamente, para los otros tres casos posibles.

44) Cuando no se dispone de método de referencia para confirmar el resultado del test clínico, las maneras de verificar el resultado para armar una Tabla de la verdad podrían ser: (contestar por sí o no)

- Efectuar análisis complementarios y adicionales.
- Consultar al especialista.
- Hacerle una autopsia al paciente que ha fallecido.
- Esperar el paso del tiempo para verificar el pronóstico hecho con el test.
- Realizar biopsias porque estas nunca fallan.
- Hacer biopsias aunque estas pueden fallar.
- Conseguir los datos de sí estaba enfermo o sano primero, y después analizar en forma retrospectiva el resultado del test clínico que se le había realizado al paciente.
- Lanzar una moneda al aire.
- Usar técnicas más refinadas.
- Estudiar la historia clínica y consultar con otros colegas.

45) ¿ Existe un test clínico perfecto que no arroje ningún resultado falso ?

46) Tratar de definir la calidad de una prueba clínica o de un método de diagnóstico.

47) Explicar la concordancia entre dos métodos de diagnóstico.

48) En una Tabla de riesgo en cual de las cuatro celdas ubicaría los casos siguientes:

- Un fumador que contrajo cáncer de pulmón.
- Un individuo vacunado que no contrajo la enfermedad.

- Un no fumador con problemas pulmonares.
- Un enfermo que no fue inmunizado.
- Una infección bacteriana luego de someterse a cirugía.
- Un paciente que no fue sometido a tratamiento preventivo antes de la cirugía y no se infectó.
- Un paciente de SIDA muy liberal en sus costumbres sexuales.
- Un recién nacido muerto de madre drogadicta.
- Un muerto de infarto de miocardio con un carácter apacible.
- Una persona sana a pesar de vivir en una zona epidémica.
- Un recién nacido sano luego de nacer por cesárea.

49) Explique como determinar el daño que puede provocar un factor de riesgo en forma ética.

50) De las razones por las cuales puede fallar una prognosis de contraer enfermedades venéreas, si no se usan protectores adecuados en las relaciones sexuales.

51) En los estudios del efecto de un tratamiento, como se hace para separar los controles de los demás casos y explique la forma en que Ud. lo haría de manera práctica.

52) ¿ Pueden alguna vez coincidir una Tabla de la verdad con una de riesgo ? Dar ejemplos.

53) ¿ Pueden alguna vez coincidir una Tabla de la verdad con una de concordancia ? Dar ejemplos.

54) Explicar la diferencia entre muestras apareadas e independientes.

55) Marcar por sí o por no las siguientes afirmaciones:

- Se pueden comparar la exactitud de dos muestras apareadas entre sí.
- Se puede comparar la precisión de dos muestras independientes entre sí.
- Se puede comparar la exactitud de dos muestras independientes entre sí.
- Se puede comparar la precisión de dos muestras apareadas entre sí.
- La concordancia se puede estudiar en magnitudes continuas.
- La concordancia se puede estudiar en magnitudes cualitativas.

56) Explicar la diferencia entre una Tabla diagnóstica, de riesgo y de concordancia.

57) ¿Cuál es el objetivo principal en un estudio de concordancia ?

58) ¿Cuál es el objetivo principal en un estudio de calidad diagnóstica ?

59) ¿Cuál es el objetivo principal en un estudio de riesgo ?

60) ¿Cuál es el objetivo principal en un estudio del tipo RCT ?

61) Explicar la diferencia entre un estudio de tipo observacional y otro de tipo experimental.

62) Explicar la diferencia entre un estudio prospectivo y uno retrospectivo.

63) Explicar uno de los objetivos básicos de Bioquímica y de Farmacia.

2

Recopilación de datos

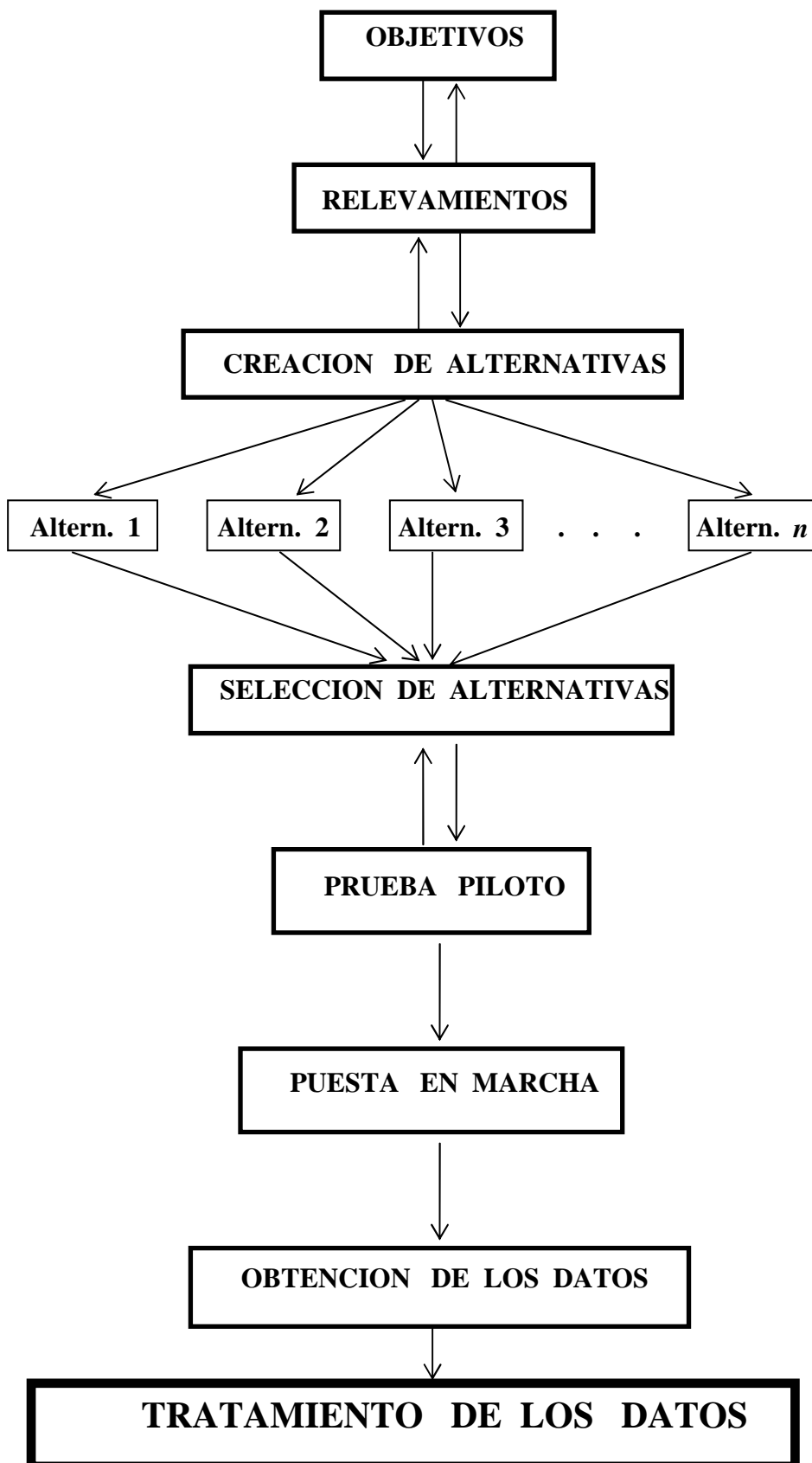
En este capítulo se describen las diferentes técnicas de recopilación de datos, y se da un procedimiento para diseñar las diferentes etapas del proceso de obtención de datos. Se hace hincapié en los métodos de medición. O sea, en la recopilación de datos por medio de instrumentos, como es la práctica diaria en un laboratorio de análisis clínicos y en las dosificaciones de medicamentos en la industria farmacéutica. Otras formas de recopilar como encuestas, censos, etc. se tocan muy someramente pues se espera que el lector interesado use los textos específicos para profundizar estos conceptos.

2.1 Etapas de la recopilación

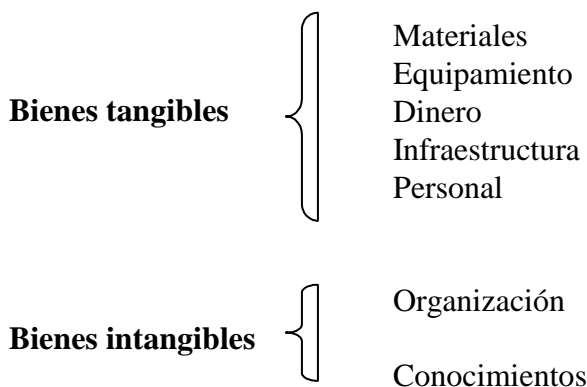
Haciendo abstracción de la manera de recopilar, se puede generalizar todo el proceso de obtención de datos usando un diagrama de flujo como es habitual en programación de sistemas. El objeto de este diagrama es ilustrar las diferentes etapas lógicas a seguir en el proceso de recopilación, y a la vez mostrar cómo se puede diseñar cualquier otra tarea. La Figura 2.1 muestra el ejemplo adoptado para este caso particular.

Etapas 1 - Objetivos de la Recopilación: esta primera etapa consiste en determinar con claridad qué es lo que se quiere lograr con la recopilación. No siempre es fácil saber lo que se quiere y menos determinarlo en detalle. Por eso, se deben definir primero los objetivos generales del trabajo estadístico. Y a partir de ellos se conocerán las variables a medir y así saber cuáles elementos se necesitarán. Con esto se tiene una primera idea de los alcances y limitaciones de la tarea a realizar, según sea el tipo de información a obtener de la población en estudio. Los objetivos deben redactarse como si fueran un telegrama, esto es, concisos, breves y claros. Una buena delimitación temprana de la tarea ahorra tiempo, dinero y trabajo, que se desperdiciaría en futuras revisiones, correcciones, ampliaciones, etc. Otra cosa a determinar con claridad en esta etapa es la población en estudio y las hipótesis necesarias para simplificar el trabajo, desde el punto de vista práctico. Normalmente, la persona a cargo de la investigación es la responsable de esta etapa pues tiene una visión más completa y actualizada del tema en estudio. Este paso y el siguiente están íntimamente relacionados. La doble flecha en el diagrama indica esta situación. Lo práctico es ajustar lo que se quiere, en función de lo que se tiene o se puede conseguir, para no caer en utopías. Por ejemplo, si se necesita la distribución de la población por edades y sexo, no es lo mismo disponer de la información del último censo realizado que hacerlo uno mismo.

Figura 2.1 - Etapas de la Recopilación de Datos.



Etapa 2 - Relevamientos: esta etapa consiste en *determinar lo que se tiene* para alcanzar los objetivos definidos en la etapa anterior. Se trata de listar los bienes necesarios para poder hacer el trabajo, y el listado de los disponibles. Conviene tener en cuenta la siguiente clasificación de los bienes:



Los *materiales* incluyen los de vidrio, de limpieza, drogas, reactivos, analitos, etc. Por *equipamiento* se entiende no sólo los aparatos de medición, sino los accesorios como muebles y útiles de laboratorio y para oficina. El *dinero* o los recursos monetarios deben ser determinados con mucho detalle para afrontar gastos e inversiones durante la investigación. Además, hay que determinar los fondos disponibles y las posibles fuentes financieras adonde poder recurrir. La *infraestructura* incluye a los edificios, laboratorios, electricidad, agua, etc. El *personal* es todo el necesario en sus diferentes niveles, como ser: profesionales, técnicos, ayudantes, consultores externos, de servicio, etc. Este relevamiento de los bienes tangibles disponibles y de los necesarios para la recopilación condiciona de alguna manera los objetivos. Puede ser que se disponga de bienes sobrados para alcanzar los objetivos, por lo que se pueden plantear metas más ambiciosas. Por otra parte, puede ocurrir que los bienes disponibles estén lejos de cubrir los necesarios, y por lo tanto se deberán resignar los objetivos planteados por otros más modestos. Por su parte, los bienes intangibles son dos: la *organización* de los bienes tangibles, de manera tal de alcanzar los objetivos, y los *conocimientos* para saber cómo usarlos. Esto es el “*know how*” de cada profesión. Y también lo es la búsqueda bibliográfica de trabajos similares en revistas especializadas, textos y otras fuentes de información. Una vez terminada esta etapa, que seguramente habrá ayudado a depurar la anterior, se debe comenzar a pensar en las diferentes maneras de hacerlo.

Etapa 3 - Creación de alternativas: esta etapa consiste en *saber cómo hacerlo*. O sea, generar distintas alternativas de sistemas de recopilación de datos, de acuerdo con los objetivos adoptados y los bienes disponibles. Se debe hacer un listado con todas las formas posibles de efectuar la recopilación a fin de tener un panorama completo. Esto se ilustra en el diagrama de flujo con el planteo de n alternativas. Básicamente, hay dos tipos de alternativas: una es lograr los datos a través de publicaciones existentes llamadas *fuentes*; la otra forma es con mediciones propias. A su vez, dentro de cada tipo, se pueden plantear varias alternativas, por ejemplo: si las determinaciones se harán en laboratorio propio o por encargo a terceros, si se controlará la calidad de los instrumentos a usar, etc. En síntesis, se habla de *fuentes propias* cuando se decide extraer los datos mediante mediciones. *Fuente Primaria* es cuando se toman los datos de otros investigadores que publican los resultados de sus propias mediciones. *Fuente Secundaria* es cuando los datos se extraen de publicaciones que usan como referencia a fuentes primarias. Y así sucesivamente. Siempre que sea posible se aconseja el empleo de fuente propia en razón de la confiabi-

alidad de los datos a manejar. Cuando esto no sea posible, entonces conviene usar una fuente primaria para evitar errores de imprenta, de transcripción, etc. En Argentina existe el Instituto Nacional de Estadísticas y Censos que es la fuente primaria por excelencia de los investigadores. El INDEC, como se lo conoce popularmente, publica boletines periódicos con información obtenida de fuentes propias. O en forma directa, a través de los Censos Nacionales de Población y Vivienda, encuestas etc. Ya sea, a través de los organismos provinciales similares con los que está asociado en esta tarea.

Etapas 4 - Selección de alternativas: esta cuarta etapa consiste en determinar cuál es la mejor entre las n alternativas planteadas en la etapa anterior. Se necesita de un método para la adopción de un criterio de selección. El método más común es el “grading” que se muestra en el Cuadro 2.1 siguiente:

CUADRO 2.1 Método de “grading” para seleccionar alternativas

Se tienen cuatro alternativas para realizar una determinación de los valores de referencia de glucosa, en una población dada. Estas son las Alternativas A, B, C y D.

Procedimiento:

1. Se establece una lista de criterios a tener en cuenta en la selección:
 - (1) Mínimo costo. (2) Máxima cantidad de datos. (3) Máxima confiabilidad.
 - (4) Mínimo tiempo. (5) Mínimo trabajo.
2. A cada criterio se le asigna un puntaje de 0 a 100 de acuerdo con la importancia que se le asigne a cada uno:
 Costo: 90 N° de datos: 50 Confiabilidad: 80 Tiempo: 40 Trabajo: 30
3. Dentro de cada criterio se hace un *ranking* de las alternativas disponibles de peor a mejor y se le asigna un puntaje, de acuerdo con ese orden. Cuando hay empate se promedia el puntaje. Luego se multiplica por el grado de importancia definido en el paso anterior.

Orden	Criterios				
	Trabajo	Confiabilidad	N° de pruebas	Tiempo	Costo
1ª	A= 1x30= 30	D=1,5x80=120	C=1x50= 50	D=1 x 40= 40	B=1x90= 90
2ª	D= 2x30= 60	C=1,5x80=120	B=2x50=100	A=2,5x40=60	C=2x90=180
3ª	B= 3x30= 90	A=2,5x80=200	D=3x50=150	B=2,5x40=60	D=3x90=270
4ª	C= 4x30=120	B=2,5x80=200	A=4x50=200	C=4x40= 160	A=4x90=360

En caso de empate se promedia y se asigna el orden promedio (Tiempo: hubo empate entre A y B, o sea entre el 2° y 3° puesto, se asigna 2,5 puntos a cada uno). En Trabajo, N° y Costo no hay empates, en Confiabilidad hay empate entre el puesto 1 y 2, y entre el 3 y el 4; en Tiempo entre el segundo y el tercer puesto.

4. Ahora se suma el total de puntos para cada alternativa:

$$\begin{aligned}
 \mathbf{A} &= 30+200+200+60+360 = 850 & \mathbf{C} &= 120+120+50+160+180 = 630 \\
 \mathbf{B} &= 90+200+100+60+90 = 540 & \mathbf{D} &= 60+120+150+40+270 = 640
 \end{aligned}$$

5. Finalmente, se escoge como mejor alternativa aquella que obtuvo el puntaje mayor. En este caso se escoge la Alternativa **A**, y como segunda opción la Alternativa **D**.

La única forma de sumar peras con manzanas es transformando cada una a pesos. El “grading” hace algo similar porque se trata de valorizar criterios totalmente subjetivos para poder sumar los puntajes y así decidir. En el ejemplo, se debe efectuar una determinación clínica de glucosa en una población dada para obtener sus valores de referencia. Se determinaron cuatro formas de hacer este trabajo y son las Alternativas A, B, C y D. El problema es decidirse por alguna de ellas. Los criterios a tener en cuenta para esta selección son: 1) menores costos; 2) obtener la mayor cantidad de datos posibles porque se trata de los valores de referencia; 3) optimizar la calidad del sistema de medición para tener confiabilidad en los datos a obtener; 4) no demorar mucho con todo el trabajo; y 5) hacerlo en la forma más sencilla posible. Una vez determinados los criterios a usar, se debe valorizarlos. Para ello, se le asigna un puntaje a cada uno, en forma subjetiva, comparándolos entre sí. En el ejemplo, el tema del costo es muy importante y recibe 90 puntos sobre 100. Si hubiese sido decisivo, entonces tendría los 100 puntos. La cantidad de datos es más o menos importante y por eso se le otorgan 50 puntos. La confiabilidad está a medio camino entre los dos anteriores y recibe 80 puntos. El tiempo que se tarde en hacer todo el trabajo no es crítico, pero sí conveniente, y por ello se le dan 40 puntos. La laboriosidad de cada método incide poco en la decisión a tomar y por lo tanto se le dan 30 puntos.

Con esto ya se tienen valorizados los criterios entre sí, y lo que falta es aplicarlos en cada una de las cuatro alternativas dadas. Para simplificar la tarea, se emplea dentro de cada criterio un “ranking” de las alternativas. La escala se puede hacer de mejor a peor. Como hay cuatro alternativas, habrá cuatro puestos en el ranking. A la alternativa más conveniente se le asigna el mayor valor posible, esto es 4. A la peor el menor valor posible: 1. Luego, entre las dos restantes se elige la mejor de ambas para ponerle un 3 y a la remanente un 2. Por ejemplo, en el criterio del costo, se sabe que la alternativa más cara es la B y así le toca el primer puesto con un punto, mientras que la más barata es la A y le tocan 4 puntos. En el medio quedan las alternativas C y D, pero como la C es un poco más cara tiene 2 y la D queda con 3. Análogamente, para los demás criterios. Debe notarse que en caso de empate se deben promediar los puntos. Por ejemplo, en el criterio del tiempo, lo único que se sabe es que la alternativa D es la peor (por eso sale con un punto) y la C es la mejor (4 puntos). Pero, de las otras dos lo único que se sabe es que demandarán un tiempo similar, una especie de empate entre ellas. Por lo tanto, quedan empatados los puestos 2º y 3º, el puntaje promedio resulta 2,5 y es el que se le asigna a las alternativas A y B.

Ahora se pueden volcar todos estos puntajes a una tabla, como la que se muestra en el Paso 3 del Cuadro 2.1. En cada casilla se multiplica el puntaje obtenido en el “ranking” por los puntos obtenidos en el “grading” de cada criterio (Paso 2). Así, cada combinación posible de criterio y alternativa, resulta valorizada en una de las casillas de esa tabla. Se tiene una ponderación de los valores subjetivos presentados. El último paso es sumar el total de puntos logrado por cada alternativa para saber cuál gana. En el ejemplo, la alternativa A resulta seleccionada por tener el valor máximo de 850 puntos. Y queda como segunda opción la alternativa D que le sigue con 640 puntos. Puede verse que la alternativa ganadora obtuvo el 75% del puntaje máximo posible. Esto es, una buena selección.

Si la alternativa seleccionada es el empleo de fuente propia, entonces hay que continuar con las etapas siguientes de la Figura 2.1. En cambio, si se decide usar otras fuentes, de publicaciones de otros investigadores, entonces hay que pasar al Tratamiento de los datos como se verá en el próximo capítulo.

En el Cuadro 2.2 se muestra otro ejemplo de aplicación práctica. El problema ahora es seleccionar una de tres farmacias para firmar un convenio de prestaciones con una Obra Social. El responsable de tomar tal decisión toma en cuenta cinco factores y visita a cada una para poder definir el ranking entre ellas.

CUADRO 2.2 Aplicación del “grading”

El gerente de una Obra Social quiere elegir la mejor farmacia de Posadas para firmar un convenio con ella para la atención de sus afiliados. Los criterios que emplea son: reparto a domicilio, al cual le da una importancia del 90%. Buen surtido de medicamentos con 80%. Atención personalizada 50%. Muchas bocas de expendio repartidas en la ciudad 60% y buen trato a los clientes 40%. Las candidatas son tres: las farmacias A, B y C. Realizando un *ranking* entre ellas para cada criterio resulta :

Criterios	FARMACIA A	FARMACIA B	FARMACIA C
Reparto a domicilio	90 x 3 = 270	90 x 2 = 180	90 x 4 = 360
Surtido de medicamentos	80 x 2 = 160	80 x 3 = 240	80 x 3 = 240
Atención personalizada	50 x 1 = 50	50 x 1 = 50	50 x 5 = 250
Bocas de expendio	60 x 4 = 240	60 x 5 = 300	60 x 1 = 60
Buen trato	40 x 5 = <u>200</u>	40 x 4 = <u>160</u>	40 x 2 = <u>80</u>
TOTALES	920	930	990

No hay empates y en conclusión, se decide negociar en primer término con la Farmacia C.

Este tipo de análisis puede emplearse en todos los casos donde resulte difícil tomar una decisión, porque los factores a tener en cuenta son muy diversos. Desde un estudio biométrico hasta una decisión de tipo personal.

Etapa 5 - Prueba piloto: existe una diferencia entre el diseño en los papeles y la realidad. Es por eso que siempre es aconsejable hacer una prueba piloto antes de la puesta en marcha para poder juzgar cómo trabaja el sistema de recopilación de datos. Se sacan unos pocos datos y se analizan las dificultades no previstas, junto con los resultados. Comparando los valores obtenidos con los que se esperaba tener, se hace una especie de *control del sistema*.

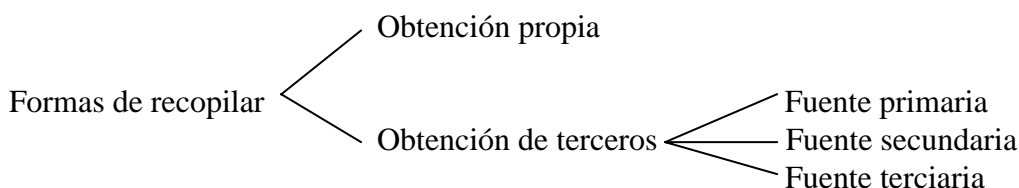
Etapa 6 - Ajustes: sería demasiado optimista esperar que todo salga bien al primer intento y que no sea necesario efectuar correcciones. Lo normal es tener que hacer pequeños ajustes que permitan optimizar al sistema. De las diferencias detectadas en el control de la etapa anterior se sacan indicios. Estos muestran qué tópicos retocar y surgen nuevas ideas de cómo hacer mejor las cosas. Básicamente, usando el sentido común se corrigen los principales defectos, como ser: mejorar el entrenamiento y conocimientos del personal, rediseñar formularios, calibrar equipos de medición, estimación de la magnitud del error de medición, etc. Pero también hay técnicas de optimización especiales como son los distintos modelos de la *Investigación Operativa*. Esta es una disciplina muy emparentada con estadística y sus modelos más conocidos son: Teoría de Líneas de Espera, Programación por Camino Crítico (PERT), Programación Dinámica y Lineal, Reemplazos, Simulaciones, etc. Una vez hechos los ajustes, se vuelve a la etapa anterior y se efectúa una nueva prueba piloto. Este ensayo permite decidir si se continúa adelante, o si son ne-

cesarios más ajustes. Hay que continuar hasta que todo sea satisfactorio y recién entonces pasar a la etapa siguiente. Pero puede ocurrir que la prueba piloto demuestre que la alternativa escogida no sea satisfactoria para los objetivos previstos. Entonces, hay que escoger una nueva alternativa y recomenzar con los ensayos. En el Cuadro 2.1, esto implicaría desechar la alternativa A y optar por la B para hacerle una nueva prueba piloto. Y así sucesivamente hasta lograr una alternativa aceptable. Si ocurre que ninguna de las alternativas planteadas es satisfactoria, entonces hay que generar más alternativas para que sean probadas. Si todavía esto no alcanza, lo que queda es volver atrás hasta definir nuevos objetivos y empezar todo de nuevo.

Etapa 7 - Puesta en marcha: una vez optimizado y ajustado el método de obtención de datos solo resta ponerlo en marcha. De esa manera, se logra la cantidad de datos necesarios para alcanzar los objetivos previstos. El resultado final es la obtención de un volumen grande de información que debe ser presentada en forma más resumida y comprensible usando tablas, gráficos y otras formas, como se verá más adelante.

2.2 Formas de recopilación

Existen varias maneras de recopilar datos. En términos generales pueden ser agrupadas de diferentes formas. Una, es la que se vio en la Etapa 3 del punto anterior; otra es:

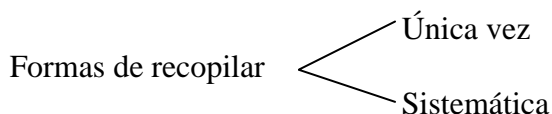


Esta primera clasificación se basa en la fuente de donde se toman los datos. La idea es que siempre conviene la fuente propia por razones de seguridad y confiabilidad. Mientras que si es de terceros, entonces conviene más la fuente primaria que las otras por la posibilidad de errores en la transcripción, publicación, etc. Pero hay otros factores a tener en cuenta como las *definiciones* de los términos y de los conceptos usados en el trabajo, las *unidades* usadas, etc.

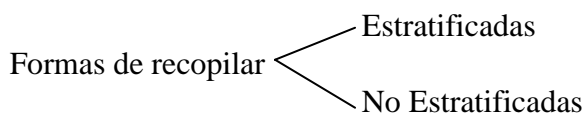
Cuando es una fuente propia se conocen todas las *definiciones*. Pero cuando se trata de una fuente primaria no siempre están escritas en la publicación, y en el caso de una fuente secundaria no se acostumbra ponerlas. El hecho de no tener muy claras las definiciones de términos implica no tener claro el ámbito de aplicación del trabajo. Por ejemplo, si se toman datos de un estudio de las *familias* de Misiones, conviene saber qué cosa entiende el autor del trabajo por familia. Una cosa es la *familia censal* constituida por la cabeza de familia, su esposa y los hijos solteros que habitan bajo el mismo techo. Otra cosa es la *familia biológica* formada por el padre, la madre y todos los hijos de ambos, de acuerdo al Diccionario Demográfico de Naciones Unidas. Otra cosa es la *familia* como se entiende vulgarmente, constituida por padres, hijos y parientes. Por lo tanto, si no se tiene claro a qué tipo de familia se refiere el autor del trabajo, no se puede saber a qué población aplicar las conclusiones. Ni tampoco si tales datos pueden ser tomados como fuente para un nuevo trabajo.

Cuando no se conocen muy bien las *unidades* que se usaron para medir los datos, estos son directamente inútiles. Por ejemplo, si se poseen datos de la producción maderera de Misiones

expresados en m^3 . Una cosa es el m^3 *estéreo* usado en la cubicación de los rollos de madera en planchada de camión, otra el m^3 *Alto Paraná* usado para cubicar madera en el monte, y otra muy diferente el m^3 *lineal* del Sistema Internacional de Unidades ($1 m^3 = 18,25 m^3$ Alto Paraná).



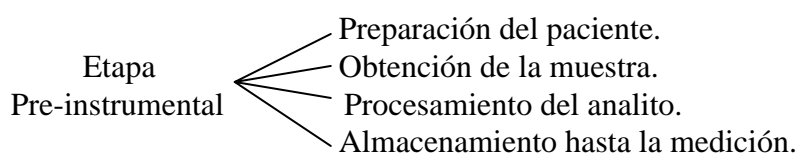
Otra forma de clasificar, es tomar en cuenta la cantidad de veces que se hará el mismo trabajo de recopilación para una investigación determinada. Hay veces en que se repite la tarea en períodos fijos de tiempo, caso típico los Censos Nacionales de Población y Vivienda que se realizan cada diez años. También se pueden efectuar Censos sobre otras actividades, como por ejemplo: Censo Industrial, Censo de Establecimientos Educativos, Censo Agropecuario etc. Por su parte existen Encuestas o trabajos similares de recopilación hechas en forma *sistemática*. Un ejemplo, son las mediciones necesarias para calcular los Índices Económicos mensuales que publica el INDEC (Índice de Costo de vida, de la Construcción, de Precios, etc.). En Misiones se realiza la Encuesta Permanente de Hogares, para controlar la situación socioeconómica de sus habitantes y ver su evolución en el tiempo.

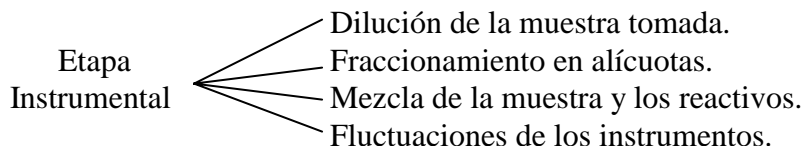


Se puede clasificar la forma de recopilar de acuerdo con la manera en que se tomaron las muestras. Muchas veces la población se divide en subconjuntos llamados *estratos* para lograr mayor calidad en las estimaciones que si se toma toda la población en su conjunto. En Clínica, la estratificación más común es la clasificación de la población por Edad y Sexo. Pero es usual el empleo de otros estratos como: Nivel de Educación o de Ingresos, Raza, Religión, etc. Para un Laboratorio de Análisis Clínicos, la manera diaria de *recopilación* de datos es *medir*, y su problema principal es la variabilidad (*error*) que afecta toda determinación clínica. En una Farmacia, el movimiento diario generado por las ventas afecta el stock. La manera de *medir* es haciendo el recuento de mercadería - un error en esto debe ser considerado aquí, como una equivocación.

2.3 Mediciones de laboratorio

La *variabilidad* en las mediciones no sólo se debe al azar. Además, hay una serie de factores, en el complejo proceso de una determinación clínica, que producen variaciones. La contribución de cada una de estas causas al total de la variabilidad, depende de los cuidados y controles que se tengan. Pueden ser acotadas, disminuidas en algo, pero nunca pueden eliminarse del todo. Los orígenes de estas fluctuaciones, se detallan en un trabajo de Statland y Winkel (1985) y se pueden abreviar, como se muestra a continuación:





Las variaciones debidas a la *preparación del paciente*, no solo incluyen a los factores biológicos que operan antes de la recogida de la muestra, sino también a los factores analíticos y psicológicos que actúan durante la toma. Dentro de todos ellos, los que tienen un efecto más acentuado son: el ejercicio, la dieta y la ingestión de fármacos. El *ejercicio* genera una actividad muscular que tiene efectos transitorios, a veces prolongados, sobre diversas cantidades clínicas. Las variaciones transitorias de los constituyentes del plasma inducidas por ejercicio incluyen una disminución de ácidos grasos libres, un aumento fuerte (del 180%) en la concentración del aminoácido alanina, y uno mayor (300%) en la de lactato. Esto se observa durante la primera hora después del ejercicio por razones de tipo metabólico, y vuelven a sus niveles anteriores en las horas siguientes. Los efectos duraderos del ejercicio se notan en el aumento de la actividad de las enzimas musculares medidas en suero como fosfatasa-alcalina (AP), creatino-fosfo-cinasa (CPK), lactato-deshidrogenasa (LDH), etc. Por ejemplo, se midieron variaciones de AP del 120%, once horas después de realizar el ejercicio.

Las variaciones debidas a la *dieta* son muy conocidas. Aunque se recomienda que antes de someterse a la extracción el paciente esté en ayunas, un ayuno prolongado (de 24 horas) puede conducir a resultados inesperados. La bilirrubina se incrementa en 240% después de ayunar 48 horas. Así también se detectaron incrementos en la concentración plasmática de leucina, valina e isoleucina, como disminuciones de concentraciones de glucosa y proteínas. El tipo de ingesta influye también en las determinaciones. Una dieta rica en proteínas produce un incremento en la concentración sérica de urea, amoníaco y urato. La ingesta de alta proporción de ácidos grasos no saturados disminuye el colesterol. Las bebidas con cafeína, mateína o teína, pueden llegar a triplicar la concentración de ácidos grasos. La supresión de hidratos de carbono aumenta la excreción de acetona en orina. La ingestión de etanol induce incrementos en la concentración de lactatos, uratos, triglicéridos y metabolitos del etanol, como disminuciones de la CPK y de la AP.

Las variaciones producidas por los efectos fisiológicos de los fármacos son muy grandes. Sobre todo, por la gran disponibilidad de los mismos usadas en el tratamiento de los pacientes. Existen dos categorías básicas: los efectos fisiológicos *in vivo* del fármaco o de alguno de sus metabolitos, sobre la variable clínica a determinar, y los efectos *in vitro* debido a propiedades físico-químicas del fármaco que interfieren con el material de ensayo. Young (1975), compiló un archivo de más de 2.000 referencias con 15.000 interacciones de fármacos, donde puede verse la gran cantidad de posibilidades existentes. Pudo observar que la respuesta farmacológica de un paciente difiere mucho de la de un individuo sano y que la respuesta *in vivo* depende de una serie de factores como: dosis, estado físico y otras medicaciones suministradas en conjunto.

En la *obtención de la muestra*, la postura del paciente es la que produce las variaciones más grandes. También inciden otros factores como la aplicación del torniquete y el ejercicio del puño. Por ejemplo, para la determinación del lactato no se debe aplicar un torniquete, mientras que si se prolonga por mucho tiempo la aplicación de este antes de la extracción en sí, se producen variaciones en proteínas, colesterol, hierro y bilirrubina. Statland (1974), encontró diferencias por variaciones en la postura de los pacientes antes de la venipunción. Determinó 17 componentes de suero en individuos sanos que estuvieron de pie hasta un minuto antes de la extracción,

y de los mismos individuos luego de estar media hora acostados. Aparecieron variaciones de más del 10% en Fosfatasa alcalina, Albúmina y ALT; más del 5% en Proteínas totales, Hierro, Lípidos totales, Colesterol, AST y Fosfatasa ácida.; más del 3% en Fosfato Potasio y Calcio; mientras que el ácido úrico disminuyó en casi un 4%. La selección de sangre capilar frente a sangre venosa no suele ser de mucha importancia, excepto en las pruebas de tolerancia de glucosa donde la capilar tiene concentraciones mayores del 10% al 30% respecto de la venosa. En la recogida de especímenes de orina de 24 horas, se deben seguir cuidadosamente las instrucciones sobre el método a seguir; caso contrario se produce mucha confusión en los resultados.

En el *procesamiento del analito* se pueden producir variaciones por contaminación de la muestra si los residuos de detergentes en los tubos empleados no se lavan correctamente. El tapón de corcho, al igual que algunos de vidrio, pueden liberar calcio al ser expuestos a la sangre; entonces la concentración sérica de calcio se incrementa falsamente. La recogida de orina para análisis de metales exige lavar el recipiente con ácidos sin contaminantes. Otro problema común, es la *hemólisis* que puede producirse tanto *in vivo* como *in vitro*. Cuando se colocan las muestras de sangre en tubos cerrados al vacío, la fuerte expansión de la sangre en el tubo provoca la lisis de las células. Lo mismo usando un anticoagulante de oxalato. Durante los pasos de centrifugación y de separación, se produce una fuga de células adicional de dos formas: una es la liberación de los constituyentes del eritrocito, y la otra es la interferencia directa de la hemoglobina con diferentes ensayos, como albúmina, bilirrubina, etc. Por su lado, la lisis de plaquetas puede aumentar la concentración sérica de Potasio y Magnesio, tanto como la actividad sérica de aldolasa y Fosfatasa ácida. El oxalato de potasio provoca una dilución del plasma porque el agua de las células se escapa. El EDTA o heparina pueden producir disminución en Calcio cuando se lo determina con un método de colorante, aunque no afecta en el método de absorción atómica. La turbiedad en suero originada por altas concentraciones de triglicéridos, provocará resultados altos para todos los ensayos basados en absorbancia a las mismas longitudes de onda. Las partículas de lípido también absorben luz y falsean el resultado. Los pasos de centrifugación y decantación pueden producir variaciones si no son efectuados correctamente. La transferencia de contaminantes, equivocaciones en el marcado de los tubos, fuga del potasio, suelen ocurrir en estos pasos si no se tiene cuidado. La congelación y descongelación pueden producir la desnaturalización de las proteínas.

En el *almacenamiento antes de medir*, pueden ocurrir variaciones por evaporación del agua que contiene el suero y por la inestabilidad de los constituyentes químicos. Una elevada temperatura ambiente, fuerte corrientes de aire en el laboratorio, mucho tiempo de exposición al entorno, pueden producir evaporación. Inversamente, si hay mucha humedad relativa ambiente. Esta evaporación puede verse ayudada por la forma y el tamaño de los recipientes usados. Todo lo cual produce lecturas con mayor concentración de las sustancias a medir en el suero, que la correcta. Muchos constituyentes químicos de la sangre no son estables durante el almacenamiento, lo que ocasiona falsas lecturas. Por ejemplo, una pérdida de la actividad enzimática, pérdida de dióxido de carbono en almacenamientos abiertos, glucólisis en sangre por esperar demasiado tiempo entre la toma de sangre y la separación del suero de la misma, la pérdida de potasio si se conserva la sangre en una heladera, etc.

Durante la *Etapa Instrumental*, es decir, mientras se están midiendo las muestras, pueden ocurrir variaciones debidas al manejo. Una cosa común es hacer las diluciones con pipetas en lugar de usar balanzas de precisión. Aunque se trate de pipetas calibradas, las determinaciones volumétricas son de menor calidad que las efectuadas en balanzas analíticas. Estas pueden detec-

tar variaciones de hasta 10^{-4} g para cantidades de hasta 200 g, precisión que no logra una pipeta. Análogamente, cuando se fracciona la muestra total en alícuotas, el trasvasamiento y la determinación volumétrica pueden dar lugar a variaciones por falta de precisión instrumental. La mezcla de las alícuotas con los reactivos para efectuar las determinaciones, es otro paso de combinación de volúmenes susceptible de lo expresado antes. Pero no sólo pueden producirse variaciones por emplear pipetas manualmente, sino que pueden ocurrir contaminaciones tanto en los materiales empleados como en los reactivos. Por ejemplo, una contaminación bacteriana conduce a diferencias detectables en una Carta de Control de Calidad. Finalmente, las fluctuaciones de los instrumentos son una de las causas clásicas del llamado *error de medición*.

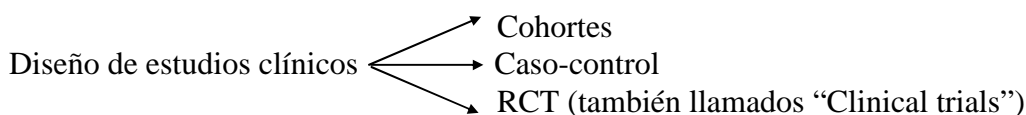
El sentido común indica que si se efectúan mediciones repetidas del mismo material bajo condiciones estables y con el mismo instrumento, los resultados obtenidos deberían ser los mismos. Sin embargo, todo investigador que haya operado con instrumentos sensibles sabe que *esto no es así*. Hay variaciones debidas a causas desconocidas y a esa ignorancia se le puso un nombre: “error”, y una causa: “el azar”.

Error: *Es la indeterminación que afecta a toda medición con instrumentos.*

En un conteo la medición es exacta si hay algún “error”, entonces se trata de una equivocación al contar mal. Cuando al efectuar una serie de mediciones se obtienen valores iguales, se piensa que el instrumento no es lo suficientemente sensible como para detectar el error. Nunca se puede creer que son iguales porque no hay errores. Las sutiles variaciones ocasionadas por el azar no son detectadas en el instrumento. Ejemplos: detectar en un reloj común una variación del orden de unas centésimas de segundo, o el peso de un grano de arena en una balanza de almacén, etc. Si bien los errores no pueden ser evitados, por lo menos pueden ser cuantificados. Como se muestra en el punto 2.5.

2.4. Recopilación en investigaciones clínicas

Tanto en Medicina, como en Bioquímica y Farmacia las investigaciones más comunes sobre poblaciones se centran en los estudios del *riesgo*. O sea, del daño potencial que pueden ocasionar en los pacientes, ciertos factores que se sospecha son causa de enfermedades. La disciplina que abarca estos estudios es la Epidemiología. Se pueden clasificar los diseños de este tipo de estudios en tres fundamentales:



En un *estudio por cohortes* el punto de partida es el estado de la exposición sufrida con el factor de riesgo (a veces de inmunización). El investigador identifica y clasifica a los que intervendrán en el estudio en dos grupos (o cohortes): Expuestos y no expuestos. Entonces se sigue en el tiempo a los sujetos, para detectar si aparece la enfermedad o signo esperado y se mide la cantidad de casos en que aparece el estado buscado (enfermos) y los casos de ausencia del mismo (sanos). Con estos datos se puede armar una Tabla de riesgo como la vista en el capítulo anterior.

Este caso es de un estudio prospectivo, aunque también se podría hacer uno de tipo retrospectivo donde se busca en la historia del paciente si alguna vez estuvo expuesto al factor de riesgo. Uno de los problemas de este tipo de diseño es la gran cantidad de datos requerida cuando la enfermedad es poco frecuente, por ejemplo si una hemorragia gastrointestinal en personas que usan cierto medicamento aparece 1,5 veces cada mil personas, en comparación con aquellos que no usan el medicamento en los cuales aparece 1 cada mil, el tamaño muestral estimado es de 75.000 sujetos para cada grupo. El otro problema de los estudios observacionales es la aparición de ciertos *factores de confusión*, que enmascaran otras causas que provocan la enfermedad. Por ejemplo, si en el caso de las hemorragias el medicamento se prescribe para pacientes de avanzada edad, y se sabe que los pacientes viejos tienen más tendencia a sangrar que los jóvenes; entonces el factor edad puede incrementar el riesgo de hemorragias que se presume debido al medicamento. Por todo esto, los investigadores deben documentar muy cuidadosamente las características de los sujetos investigados, tanto expuestos como no expuestos, ya sea para demostrar su compatibilidad o para corregir las diferencias usando técnicas estadísticas. Se supone que ambos grupos deberían tener características similares para poder ser comparados como por ejemplo tener aproximadamente la misma edad, sexo, condiciones socio-económicas etc. Cuando los investigadores no consigan un buen balance en edad entre ambos grupos, se deberán usar métodos estadísticos para ajustar el desbalance de los grupos. Pero aunque se tomen todas las precauciones posibles al respecto como: la comparación entre los factores de confusión potenciales, la documentación cuidadosa para evitar un desbalance, grandes tamaños muestrales, etc.; ni aún así, pueden estar seguros que factores importantes para el diagnóstico no se hayan tenido en cuenta, porque sencillamente los desconocen, o no los han medido y por lo tanto no saben si ocasionan un desbalance entre los grupos. Todo esto muestra que las inferencias obtenidas con los estudios por cohortes nunca será tan confiables como las que se pueden obtener con un riguroso estudio del tipo RCT.

En un estudio diseñado como *caso-control* los pacientes que han desarrollado la enfermedad, o el síntoma buscado, se clasifican como *casos*, y luego se eligen como *controles* a individuos que son razonablemente similares a los casos, con relación a aspectos relevantes como pueden ser sexo, edad, condiciones médicas concurrentes, etc., pero que no han desarrollado la enfermedad. Este tipo de estudios es más conveniente que el anterior cuando las enfermedades son raras, o cuando un estudio por cohortes toma demasiado tiempo para que sea factible. Usando un diseño del tipo caso-control los investigadores pueden determinar la frecuencia relativa de exposición al agente que se cree origina la enfermedad, ajustando los resultados con las otras variables conocidas en el pronóstico. Este diseño permite también el estudio simultáneo de múltiples factores de exposición que podrían tener una asociación con la enfermedad resultante. Por ejemplo, los investigadores podrían hacer un estudio del tipo caso-control para demostrar la relación entre la ingesta de “dietilstibestrol” (DST) en mujeres embarazadas y el desarrollo de adenocarcinoma vaginal en sus hijas muchos años más tarde. Un estudio del tipo RCT o por cohorte requerirían más de 20 años para poder ser completados. Peor aún, dado que no es muy frecuente esta enfermedad la cantidad de casos requerida sería de cientos de miles. Por el contrario, usando un diseño del tipo caso-control los investigadores separaron dos grupos: en el primero colocaron a las jóvenes que habían desarrollado la enfermedad como casos (8 en total) y luego buscaron como controles jóvenes con similares características que no la habían desarrollado como controles (32 en total). Luego retrocediendo en el tiempo determinaron si existió la ingesta de DST durante el embarazo de sus madres. Se encontró una fuerte asociación ($p < 0,00001$) entre la ingesta de DST y el desarrollo del adenocarcinoma vaginal, sin tener que esperar 20 años y con tan solo 40 mujeres en lugar de cientos de miles.

En los estudios diseñados como *RCT* se pueden estudiar la exposición a agentes dañinos tanto como a agentes beneficiosos para el paciente como inmunizaciones, protección antibiótica previo a una cirugía, etc. Como la aleatoriedad de la muestra es la mejor manera de balancear los grupos de expuestos y no expuestos, las estimaciones resultarán menos sesgadas que con los otros tipos de estudios. Se ha encontrado que los resultados inesperados que a veces aparecen en estos estudios, tienen gran potencial en la determinación del daño potencia de un factor en estudio (por ejemplo, drogas que los investigadores suponen beneficiosas a veces aparecen asociadas con un incremento de la mortalidad).

Cuadro 2.3. Cuadro comparativo de los tipos de estudios clínicos

<i>Diseño</i>	<i>Comienza en</i>	<i>Se estudia el</i>	<i>Ventajas</i>	<i>Desventajas</i>
Cohortes	Estado de la exposición.	Resultado de la enfermedad.	Posible cuando no se puede hacer aleatoria la selección de lo individuos expuestos.	Susceptible a sesgo y validez limitada.
Caso-control	Estado de la enfermedad	Resultado de la exposición	No insumen largos períodos de tiempo, y se pueden usar muestras pequeñas	Susceptible a sesgo y validez limitada
RCT	Estado de la Exposición.	Resultado de casos adversos	Baja susceptibilidad al sesgo	Posibilidad limitada, generalizaciones

Hay otro tipo de estudios tales como las *series de casos* (descripción de una serie de pacientes) y el *informe de un caso* (descripciones de pacientes individuales). Estos dos tipos de estudios no suministran comparaciones entre grupos, ni requieren que los grupos tratados y de control tengan similar prognosis. Sin embargo, ocasionalmente se encuentran dramáticos hallazgos que demandan un inmediato cambio en el trabajo de los médicos, como por ejemplo el caso del uso de la talidomina en embarazadas y sus consecuencias en el recién nacido. O bien, la serie de casos que se fueron reportando hasta que se identificó al Sida como una nueva enfermedad. Pero también pueden haber consecuencias indeseables cuando se toman acciones en respuesta a cierta evidencia de enfermedad, por ejemplo el caso de la droga Bendectin usada como antiemético en el embarazo, donde el fabricante tuvo que sacarla de inmediato del mercado, como resultado de informes de series de casos que sugerían que era peligrosa. Aunque estudios posteriores demostraron que la droga era relativamente segura, la mala fama adquirida no pudo ser erradicada, como tampoco la posibilidad de juicios en contra del fabricante al reintroducir la droga. En consecuencia, las posibilidades benéficas de esa droga se perdieron por esos motivos. Por eso los médicos no deben dejarse llevar tanto por los informes de casos, donde aparecen ciertas relaciones del tipo causa-efecto, sino más bien reconocer que esos resultados pueden conducir a las investigaciones pertinentes por los entes reguladores, o los investigadores clínicos.

2.5. Mediciones industriales en Farmacia

En este caso, las mediciones habituales se orientan hacia el control de producción. Se usan instrumentos en cada etapa del proceso productivo, generalmente de lectura continua, como es el caso de una termocupla insertada en un horno para controlar temperatura. A la que puede

adosarse una graficadora de pluma, que registra sobre un papel continuo el valor de temperatura en forma permanente (al estilo de un electrocardiograma). De esa forma se pueden controlar las variaciones indeseadas y saber el momento en que ocurrieron. Todavía, para automatizar este proceso, se puede incorporar un timbre de alarma que suene cuando la temperatura sobrepase ciertos límites prefijados; o bien, accionar el encendido y apagado del horno como un releé.

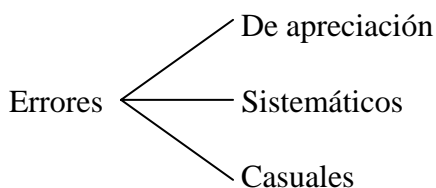
Las mediciones típicas de Laboratorio forman parte del control de calidad en la industria. Se debe controlar tanto la materia prima ingresante como el producto final terminado. En la industria farmacéutica, alimenticia y de cosmética, los procesos son similares. La materia prima se inspecciona en el laboratorio para determinar la concentración del reactivo comprado, de los solventes y otros integrantes de la fórmula del producto a fabricar. Es decir, su grado de "pureza". Pero también deben controlarse los demás insumos como envases, envolturas, frascos, ampollas, etc. Aquí, el objetivo es asegurarse que cumplan los requisitos pedidos, como por ejemplo los colores e impresión de los envases de cartón, de las etiquetas a pegar, el estampado en los tapones, etc. Luego, al final del proceso productivo se deben hacer más controles de calidad. En este caso, el objeto es asegurarse que el producto terminado cumpla las normas de calidad adoptadas por la empresa. Estas normas van desde la forma de empaquetamiento hasta el grado de pureza de los principios activos del medicamento. La presentación del producto final hace a la imagen empresaria entre los consumidores. En un perfume, por ejemplo, la estética del envase a veces es más importante que el contenido. En un medicamento es al revés, por lo tanto el mayor cuidado en esos casos se centra en el contenido. (Las acotaciones hechas en el punto anterior, en términos generales, son aplicables al laboratorio de control en la industria. Pero las mediciones efectuadas a lo largo del proceso productivo son otra cosa).

El problema general en la industria química es la mezcla en una o más fases del proceso de varios componentes dentro de aparatos adecuados, ya sea en forma continua o en forma de "batch" (lotes de producción). Cada componente puede ser considerado transferible a otras fases posteriores, teniendo en cuenta los cambios producidos por las reacciones químicas, con liberación o absorción de calor, reacciones catalíticas, oxidaciones, humidificaciones, etc. Dadas las condiciones iniciales necesarias para que comiencen las reacciones o para poner al sistema en cierto punto, luego de cierto tiempo de contacto comienza el proceso que debe ser controlado en todas las etapas subsiguientes, en cada punto o momento donde se produzcan los efectos buscados. En términos generales, se debe prestar atención a los puntos siguientes:

- *Balance de materiales*: la masa de la entrada, menos la de salida, debe ser igual a la acumulada durante el proceso. Esto permite tener un balance diferencial para cada componente y se puede disponer de relaciones estequiométricas disponibles para el control.
- *Primera Ley de la Termodinámica*: tomando la etapa como un todo, se puede aplicar la manera diferencial de la primera ley para tener una ecuación de control. Esto se reduce a un balance de entalpía si no hay flujo de calor desde el contorno hacia dentro o fuera, ni trabajo hecho por el sistema.
- *Equilibrios*: en los límites de las regiones interfaciales entre las etapas, se producen condiciones de equilibrios que prevalecen en cada fase, y que se pueden considerar como partes del equilibrio final alcanzado por el proceso como un todo.
- *Ecuaciones porcentuales*: se aplican a la transferencia de calor a través de todas las interfaces entre las etapas del proceso, a la transferencia de masa de cada componente individual y de las reacciones químicas ocurridas.

2.6 Cuantificación de errores de medición

En el punto anterior se detallaron las muchas causas que pueden producir variaciones en las mediciones del Laboratorio de Análisis Clínicos. La variación total que afecta una medición de laboratorio debe ser considerada como la suma de muchas variaciones muy pequeñas, que agregan su efecto de manera tal que puede ser detectada por instrumentos. Esto se expresa así: el *error total* en una medición de una *magnitud cuantitativa* es ocasionado por un gran número de causas, conocidas y desconocidas, que pueden producir efectos infinitesimales individualmente, pero que en conjunto, pueden ser cuantificados y clasificados como sigue:

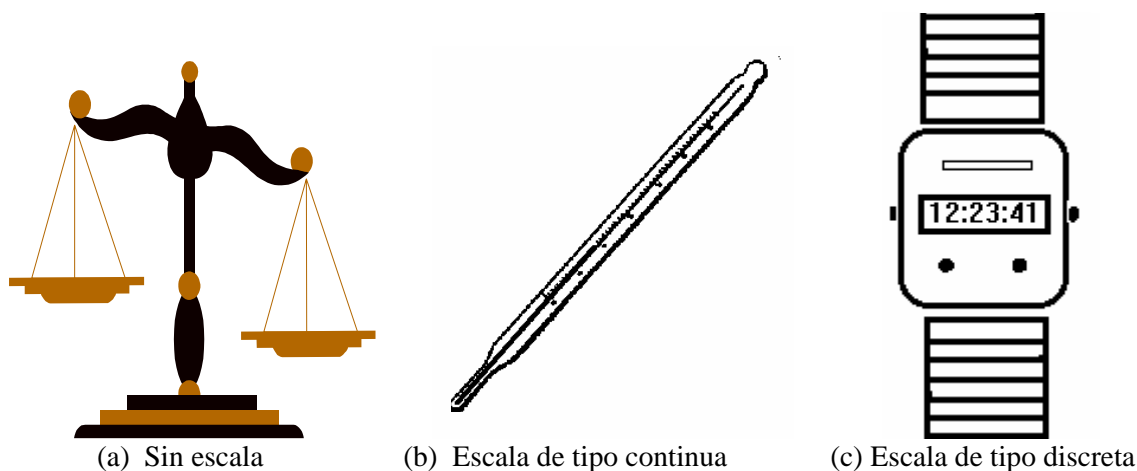


La clasificación anterior no es estricta. La aparición de un tipo de error no excluye la existencia simultánea de los demás. Si no se los detecta es porque el sistema de medición no es lo suficientemente sensible para ello. En la literatura sobre este tema es habitual encontrar a los dos últimos tipos solamente. Aquí se agregó el primero por su utilidad en el diseño previo de las mediciones y como criterio en la selección de instrumentos. Sirve para cuantificar el error cuando se efectúa una sola medición. Los otros dos necesitan más de una medición para poder ser cuantificados.

Error de apreciación: es aquel originado en la indeterminación de la escala del instrumento de lectura utilizado.

Básicamente, hay tres tipos de escala que usan los instrumentos: lectura directa, digital y de alineación. Estos casos se esquematizan en el Gráfico 2.1 siguiente:

Gráfico 2.1: diferentes tipos de escalas en instrumentos de medición



Los instrumentos de *lectura directa* son todos aquellos cuya escala es continua. Donde a simple vista se puede ver hasta la menor unidad de la escala. El ejemplo clásico es una regla milimetrada o un termómetro de mercurio, como el caso (b) del Gráfico 2.1. Para efectuar una medición de temperatura en un líquido, se sumerge el bulbo del termómetro en el mismo y se compara la altura de la columna de mercurio con la regla grabada en el vidrio del mismo. Si el termómetro está graduado en grados centígrados, significa que la distancia entre dos marcas sucesivas en el vidrio tiene ese valor. Si se puede ver que el menisco de mercurio está entre la marca de 56°C y la de 57°C, entonces, el resultado x de esta medición se debe expresar:

$$x \in (56; 57)^\circ\text{C}$$

o lo que es lo mismo:

$$x \in (56,5 \pm 0,5)^\circ\text{C}$$

La segunda manera de expresar el resultado de la medición significa tomar el valor más probable dado por el promedio, y aceptar una incertidumbre de medio grado llamada *error de apreciación*. Cuando la lectura es directa, se cuantifica a ese error con un valor del 50% de la menor unidad de la escala. Por ejemplo, en una regla milimetrada el error de apreciación sería de 0,5 mm; en un reloj de agujas con segundero sería de medio segundo. La indeterminación, en estos casos, está dada por la capacidad del ojo humano para discernir entre las dos marcas de la regla. No sería razonable decir que el ojo puede ver las centésimas de milímetro en una regla común.

Los instrumentos de *lectura indirecta* pueden ser de dos tipos: los de escala discreta y los de comparación. Cuando el instrumento es digital se tiene una escala discreta, pues los valores van de uno en uno y es imposible “apreciar” la mitad de la menor magnitud de la escala. En el Gráfico 2.1 puede verse en el caso (c) un reloj digital que informa hasta el segundo. Indica una lectura de 12 hs. 23' y 41". La siguiente indicación podrá verse como 12:23:42. O sea, el error de apreciación en estos casos se cuantifica con la menor magnitud de la escala. Así, una lectura cualquiera se expresará como:

$$x \in (\bar{x} \pm \Delta x)$$

El valor x obtenido hace pensar que el *valor verdadero* de la magnitud medida pertenece a un intervalo definido por: el *valor más probable* más o menos el error de apreciación Δx . Para el ejemplo anterior, si se mide un intervalo que empieza a las 12:23:42 y termina a las 12:24:05, el tiempo transcurrido se expresa como: $x \in (23 \pm 1)$ segundos.

Los instrumentos con una escala de comparación son algo más complicados. En el caso (a) del Gráfico 2.1 se muestra el esquema de una balanza de platillos. Aquí se tiene una aguja, llamada el fiel, que oscila alrededor de la posición central de equilibrio. Cuando se coloca la masa a pesar en un plato de la balanza esta se desequilibra. Se van agregando pesas hasta volver a lograr el equilibrio. Sumando el valor de las pesas se obtiene la medición. En este caso no hay una escala en el sentido clásico del término. Por lo tanto, para cuantificar el error de apreciación se procede como sigue: se toma la menor pesa disponible y se la coloca con mucho cuidado en el plato; si la aguja no se mueve, entonces se toma la pesa inmediata superior. Y así sucesivamente hasta poner una que altere el equilibrio, lo más chica posible. El valor de tal pesa será el módulo del error de apreciación para esa lectura.

Todo lo anterior muestra que siempre habrá una manera de poder cuantificar la indeterminación que afecta a toda lectura en un instrumento. Cualquiera sea el tipo de escala del instru-

mento, cualquiera sea el instrumento, *toda medición puede y debe ser expresada* usando el concepto de error de apreciación. Es mucho más prudente estimar el valor desconocido de la magnitud, con un intervalo, en vez de un único valor. Como ejemplo, se puede pensar en un paciente que concurre a un bioquímico a medirse la glucosa y este le informa (90 ± 15) mg/l, luego para confirmar el valor va a otro laboratorio a repetirse la medición y en éste último le informan (103 ± 7) mg/l. Al comparar ambos informes no ve incongruencias. Muy distinto hubiera sido el caso si ambos laboratorios le hubiesen informado con un único valor 90 vs. 103.

Cuando se mide más de una vez la misma magnitud, entonces cambia todo el procedimiento para determinar la incertidumbre de las medidas. Ahora se pueden obtener los otros dos tipos de errores.

Error sistemático: *es aquel que se caracteriza por tener siempre el mismo signo y aproximadamente el mismo módulo, en una serie de mediciones.*

Un reloj que atrasa, una regla dilatada, un espectrofotómetro sin el ajuste del cero, son los ejemplos típicos de esta clase de errores. Son los más difíciles y costosos de detectar. Para poder corregirlos, generalmente se deben calibrar los instrumentos con personal especializado. Cuando se los acota, se pueden corregir los resultados finales en alguna medida, pero nunca se los puede eliminar totalmente. Hay una serie de causas que los originan debidos al:

- . *Observador:* tendencia personal a ubicarse a un costado de la aguja indicadora (paralaje). Mal lavado de los materiales a utilizar. Falta de control en los procedimientos. Mala aplicación de los cubreobjetos. Mezclas deficientes. Descuido en recuentos, etc.
- . *Instrumento:* lentes o filtros sucios. Electrodo viejos. Agujas indicadoras dobladas. Cubetas de caras no paralelas en el espectrofotómetro. Mala calibración del aparato, etc.
- . *Materiales:* contaminación de los reactivos por falta de precauciones. Drogas vencidas. *Kits* mal refrigerados. Mala extracción de la muestra al paciente. Conservación de las muestras deficiente, etc.
- . *Método:* tiempo de centrifugación excesivo o escaso. Elección inadecuada del diluyente. No emplear EDTA o similar en recuentos en sangre. Variaciones bruscas de temperatura o energía eléctrica sin tener previsto su control, etc.

La responsabilidad por este tipo de causas recae en el profesional, pues si toma las adecuadas precauciones podría evitar la mayoría de ellas. Este es uno de los motivos por los cuales toda técnica de laboratorio tiene un *protocolo* a respetar estrictamente. Además, todo laboratorio debería tener procedimientos de control de calidad para sus prácticas, junto con los planes de mantenimiento preventivo para los aparatos, materiales y drogas. Sin embargo, como el factor humano es el principal responsable de su prevención, lo mejor es mantener a todo el personal motivado y entrenado en esta problemática. Las técnicas estadísticas sólo informan de la aparición de estos errores *después* que han ocurrido, cuando lo ideal sería *evitar* que ocurran. Si todo el personal conoce el tema de errores en mediciones a fondo y si ponen su empeño en evitarlos, casi todo el trabajo está hecho. No debe perderse de vista la tendencia humana a evitarse problemas. Cada vez que se detectan desvíos, eso implica una serie de trabajos extras fuera de la rutina habitual del laboratorio y lo más difícil: descubrir las causas desconocidas que los provocan. El ser humano reacciona instintivamente ante lo desconocido con temor. El abandono de una estereotipia de conducta cómoda y habitual, le genera ansiedad. Le es difícil aceptar el hecho de que tiene fallas ante sí mismo y más aún ante los demás. Así, lo natural es la resistencia al cambio.

Por eso, la parte más difícil en la implementación de un programa de control de calidad no es su diseño ni su implementación, sino lograr que el personal lo adopte de buen grado en su fuero íntimo. Un cambio de actitud. Lo más difícil.

Error casual: es aquel provocado por una gran serie de causas desconocidas y variables llamadas casualidad, en una serie de mediciones. Su módulo no es constante y tampoco, tiene el mismo signo.

Toda vez que se efectúan varias mediciones y los resultados obtenidos son diferentes, se puede tener una medida del mismo. Para poder cuantificarlo se usa la teoría estadística de los errores casuales, que se verá más adelante. Por ahora, basta decir que se los usa en técnicas de calibración para poder cuantificar la precisión y exactitud de las técnicas de laboratorio. Ambos criterios se usan también para seleccionar prácticas entre sí. Cuando hay varios métodos posibles para hacer la misma determinación clínica conviene elegir el de mejor calidad, y hasta hace poco eso significaba el de mejor precisión y exactitud. Actualmente se emplean otros criterios de selección. Sobre todo en las magnitudes cualitativas donde precisión y exactitud no tienen sentido.

En síntesis, todo resultado de una práctica clínica debe ser informado a los usuarios usando un intervalo formado por un valor más probable y una estimación del error de medición. Este error acota y cuantifica la *variabilidad* de las mediciones de laboratorio ocasionadas en la forma explicada en el punto 2.3 anterior. Se supone que el método empleado ha sido testeado, calibrado, y está controlado para mantener una calidad aceptable (más de un 95% de confiabilidad en los resultados).

2.7 Problemas propuestos

1) Ordenar las siguientes etapas de una recopilación de datos: () Puesta a punto; () Relevamientos; () Creación de alternativas; () Objetivos; () Selección de alternativas; () Puesta en marcha.

2) Empleando el método de grating determinar la puntuación de 3 alternativas: A, B y C donde al Costo se lo valoriza en 90 puntos, siendo A la mejor y C la peor. El Tiempo con 50 puntos es el mismo para las tres. En Confiabilidad con 80 puntos la mejor es B, luego C y A:

A () B () C ()

- | | | |
|---|---|---|
| 3) La mejor forma de hacer una recopilación es usando fuente primaria. | V | F |
| 4) En la etapa instrumental de las mediciones se hace el procesamiento del analito. | V | F |
| 5) El error de apreciación de un reloj común digital es 0,5 seg. | V | F |
| 6) El error sistemático puede ser corregido por el bioquímico. | V | F |
| 7) El error casual no puede ser corregido por el profesional. | V | F |
| 8) Nombrar los bienes tangibles e intangibles en la etapa de relevamiento..... | | |
| 9) Un censo es una forma de recopilar por única vez. | V | F |
| 10) Una encuesta puede ser estratificada para tener mayor riqueza de información. | V | F |
| 11) La variabilidad en las mediciones sólo es producto del azar. | V | F |
| 12) Para evitar variaciones por la dieta se debe preparar al paciente con un ayuno. | V | F |
| 13) El ejercicio no influye en la medición de los parámetros clínicos. | V | F |
| 14) La postura del paciente produce variaciones en la obtención de la muestra. | V | F |
| 15) Nombrar los cuatros puntos a tener en cuenta en la etapa pre-instrumental. | | |

- 16) Ídem anterior, pero para la etapa instrumental.
17) Nombrar los tres tipos de errores de medición.
18) El error de apreciación de un termómetro común es $0,1^{\circ}\text{C}$. V F
19) El error casual se puede cuantificar sólo si se hacen varias mediciones. V F
20) El error sistemático de una balanza puede ser producido por una mala puesta a cero. V F
21) En un reloj digital un intervalo de 45 seg. se informa como $(45,0 \pm 0,5)$ seg. V F
22) Los instrumentos de lectura indirecta son de comparación únicamente. V F
23) El error sistemático puede ser de distinto signo. V F
24) Una serie de causas desconocidas producen el error casual. V F
25) Toda medición está afectada por una cierta indeterminación, llamada error. V F
26) Conviene informar los valores medidos con un intervalo, antes que con un punto. V F
27) Usando el método del *grading* encontrar la mejor de las alternativas propuestas:

En cuatro laboratorios de la ciudad se prestan especialidades relacionadas con hormonas y radio-inmuno ensayo. Un sanatorio se apresta a firmar un convenio con uno de ellos y debe decidir con cuál. Para esto toma en cuenta los criterios siguientes:

- Cercanía al sanatorio (los laboratorios A y C están a pocas cuadras y los otros dos muy lejos, siendo el B más cercano que el D).
- Control de calidad (los laboratorios C y D llevan un programa de control y los otros no).
- Rapidez en la entrega de resultados (A es el más veloz, luego D, C y por último B).
- Comisiones (B da la máxima, C una intermedia y A y B la mínima).
- Experiencia (D es el más antiguo, A el más novato y los otros dos parecidos entre sí).

Se le asigna 40 puntos al primer criterio, 80 puntos al segundo, 60 al tercero, 100 puntos al cuarto y 50 puntos a la experiencia.

¿Por qué se decide finalmente por el laboratorio B?

28) Completar los siguientes conceptos:

- La recopilación de datos a través de formularios se hace en una
- La primera etapa de la recopilación de datos es definir claramente los del mismo.
- Para seleccionar alternativas se puede usar cualquier método, pero el más recomendable es el llamado “.....”.
- El relevamiento se hace usando libros, informes, revistas científicas, etc.
- Conviene hacer una prueba antes de la puesta en marcha.
- Cuando se sacan los datos del informe de otro investigador, se usa una fuente
- Cuando se miden los datos a usar, se está usando una fuente
- Cuando los datos se obtienen de terceros, se usan fuentes, o terciarias.
- Se pueden recopilar datos por única vez, o en forma
- Si la población se divide en, la forma de recopilar se llama
- Las variaciones de las mediciones en un laboratorio se deben al azar y a una serie de factores como ser:
- Hay dos etapas básicas en las mediciones: la y la
- La mezcla de la muestra del paciente con los reactivos se hace en la etapa
- Las variaciones producidas por los efectos fisiológicos de los fármacos son muy Hay dos tipos de efectos los *in vivo* y los *in*

- La dieta que deben hacer los pacientes para una toma de muestra de sangre es un de 12 horas, porque un ayuno de 24 horas no es
- Para que no ocurran evaporaciones antes de medir la muestra, conviene que
- Error es la que afecta a toda medición con instrumentos.
- Hay que prestar atención en las mediciones industriales a los siguientes puntos:
- Los errores de apreciación se emplean cuando se hace una sola
- Los errores de apreciación se originan en la indeterminación de la del instrumento.
- El error sistemático es la diferencia entre
- El error casual está relacionado con una serie de causas
- El error siempre es del mismo signo y de magnitud en una serie de mediciones de laboratorio. En cambio el error , puede tener diferente y diferente
- La precisión se relaciona con el y la exactitud con el
- La calidad en un sistema de medición se relaciona con y
- El error sistemático puede ser originado por el observador, el, los y el

29) Definir los siguientes conceptos:

- Recopilación de datos
- Fuente propia
- Fuente primaria
- Bienes tangibles e intangibles
- Formas de recopilación

30) Clasificar los pasos a seguir en las etapas pre-instrumental e instrumental.

31) Definir y clasificar los tipos de errores de medición.

32) Diferencias entre mediciones de laboratorio e industriales.

33) Formas de cuantificar los tres tipos de errores de medición.

34) Explicar los tipos de diseños de estudios clínicos más comunes.

35) Ventajas y desventajas entre los diferentes diseños clínicos.

36) Explicar las ventajas y desventajas de los informes de serie de casos.

37) Que tipo de diseño se debe usar cuando se quiere investigar:

- la exposición a un cierto factor de riesgo
- el desarrollo de la enfermedad.
- la inmunización o protección aplicada
- una enfermedad muy rara
- el efecto de un factor dañino al paciente.

3

Presentación de datos

En este capítulo se describen las diferentes técnicas de presentación de datos, y se dan varios procedimientos para efectuarlas. Primero, se muestra una manera de proceder en el Laboratorio de Análisis Clínicos cuando este debe informar sus resultados al usuario. Luego, se muestran las diversas técnicas para armar tablas y gráficos cuando se tienen muchos datos. Se repasan los gráficos más usuales y se hace hincapié, en los de uso más frecuente o importante, como los histogramas y las pirámides de población.

3.1 Informes de laboratorio

Diariamente, el laboratorio de análisis clínicos prepara una serie de informes escritos. El más usual es el reporte al médico solicitante (Gráfico 3.1) de los resultados obtenidos en el paciente analizado. Cada laboratorio puede confeccionar un formulario preimpreso con copias para que una de ellas quede en su poder como archivo de todo lo actuado. En este formulario basta volcar los datos medidos, colocar las observaciones si las hubiera y firmar para tener el informe listo. Estos formularios normalmente contienen la siguiente información:

. *Datos de identificación:* en el borde superior del formulario se coloca el logotipo del laboratorio (si lo tiene), el nombre, la dirección, el horario de atención y los teléfonos.

. *Título:* en forma breve y concisa se explica el tipo de informe.

. *Paciente:* los datos principales del paciente son: nombre y apellido, número de afiliado y la obra social (o análogo) a la que pertenece, porque con esos datos se puede facturar. Además el sexo y la edad son dos cosas necesarias para el médico.


. *Solicitante:* es para especificar el nombre del médico que envió al paciente.

. *Cuerpo:* el cuerpo es el informe propiamente dicho en una columna se codifican las prestaciones realizadas; en la columna de al lado la denominación técnica de la prestación; y en la siguiente se coloca cada uno de los resultados. A veces se agrega, y conviene hacerlo, otra columna con los valores del intervalo, dentro del cual se estima esté el valor verdadero de la medición. Otras veces, en esa última columna se colocan los valores de referencia para poblaciones humanas, de la edad y sexo del paciente. Esto último le facilita al médico saber si su paciente tiene los valores “normales”, o bien, algún exceso o defecto que le ayudan en el diagnóstico y tratamiento. Finalmente, algunos profesionales dejan en blanco esa última columna para poder agregar algo si lo consideran oportuno.

. *Notas:* al final del informe se dejan dos o tres renglones libres para observaciones.
 . *Firma:* todo informe debe ser rubricado por el profesional responsable con aclaración del nombre y apellido, título profesional y matrícula pública con la que está habilitado oficialmente.

En el Gráfico 3.1 se muestra un ejemplo de este tipo de formularios, usando datos imaginarios. Lo que se presenta es la copia que debería guardar el profesional para su archivo personal, donde se colocan todos los datos relevantes del trabajo realizado. El original que se envía al médico no tiene porqué tener toda esa información; por ejemplo, los códigos de prestaciones pueden ser obviados. El resguardo de la copia es importante para prevenir cualquier inconveniente que pudiese surgir en el futuro, tal como un litigio, un reclamo por una cobranza mal liquidada, etc.


Gráfico 3.1: Ejemplo de formulario para informar resultados

LABORATORIO DE ANÁLISIS CLÍNICOS			
de la Clínica La Dulce Espera		INFORME DE RESULTADOS	
Félix de Azara 3789 - Posadas (3300)			
Paciente: Ana S. de Fernández		Solicitante:	FECHA: 12 08 01
Nº Afiliado: 56788948412-9		Dr. Claborde	Horario: <i>Lunes a Viernes 8 a 20 hs</i> <i>Sábados de 8 a 14 hs.</i>
Obra Social: Madereros		Clínica el Sol -Caingúas	
Edad : 25 años			
Código	Práctica	Valor medido	Valores de referencia
174	Colesterol Total	170 mg/dl (± 20)	hasta 220 mg/dl
876	Triglicéridos	102 mg/dl (± 25)	hasta 170 mg/dl
192	Creatinina	9 mg/l (± 2)	entre 8 y 14 mg/l
902	Uremia	0,2 g/l (± 0,05)	hasta 0,4 g/l
Observaciones:			 Juan Sebastián Mastropiero BIOQUIMICO M. P. Nº 4567

A veces se acostumbra usar *separatas*, o sea, una especie de volantes preimpresos para los informes especiales como ser: Hemograma completo, Proteinograma, Inmunoglobulinas y algunos otros. Estas hojas separadas se completan con los datos medidos y se abrochan a la hoja del informe principal, vista más arriba. En el Gráfico 3.2 se muestra un ejemplo. Conviene destacar que en toda separata se vuelven a colocar fecha, paciente y firma, aunque parezca redundante. Esto conviene hacerlo pues, si las hojas se desglosan, el médico puede saber de quién se trata y cuándo se hizo el informe. Además, toda hoja de un informe debe estar firmada, para darle visos de seriedad.

Se recomienda el uso de sobres para poner los informes que se entregan a los pacientes aunque estos no sean cerrados. Sin embargo, si la información es confidencial, es mejor cerrar el sobre de manera tal que nadie lea el informe, salvo el médico solicitante. Es usual el uso de sobres con membrete de tamaño oficio, para estos menesteres.

Gráfico 3.2: Ejemplo de informes profesionales usando separatas.

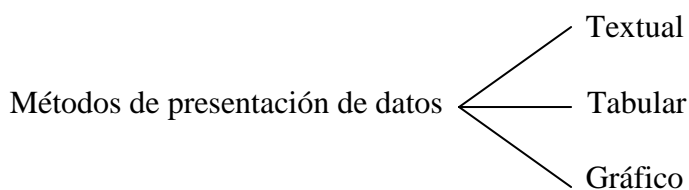
LABORATORIO DE ANÁLISIS CLÍNICOS de la Clínica La Dulce Espera PROTEINOGRAMA Félix de Azara 3789 - Posadas (3300)						
PRÁCTICA	Unidad	Valor Medido	Valores de referencia			
Proteínas totales.....	g/dl	7,66 (\pm 0,5)	6,1 a 7,9			
Albumina.....	g/dl	4,46 (\pm 0,6)	4,0 a 5,2			
Globulinas.....	g/dl	3,20 (\pm 0,3)	1,9 a 2,7			
α 1 globulinas.....	g/dl	0,23 (\pm 0,1)	0,2 a 0,4			
α 2 globulinas.....	g/dl	0,75 (\pm 0,1)	0,4 a 0,7			
β globulinas.....	g/dl	0,85 (\pm 0,1)	0,7 a 0,9			
γ globulinas.....	g/dl	1,15 (\pm 0,1)	0,7 a 1,4			
Relación A/G.....	--	1,39	1,5 a 2,7			
FECHA <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td style="width: 20px; text-align: center;">12</td> <td style="width: 20px; text-align: center;">08</td> <td style="width: 20px; text-align: center;">01</td> </tr> </table>	12	08	01	Paciente: Ana S. de Fernández	 Juan Sebastián Mastropiero BIOQUIMICO M. P. N° 4567	
12	08	01				

La otra presentación de datos importante de los laboratorios, es el listado de todas las prácticas realizadas en el mes (o quincenal) que se presenta a modo de factura para cobrar.

En lugar de presentar una factura a cada una de las Obras Sociales con las cuales se trabajó durante el mes, es mucho más práctico elevar una sola, completa, al Círculo o al Colegio de Bioquímicos, como es el caso de la provincia de Misiones. Entonces, la entidad que los agrupa se encarga de cobrarle a cada Obra Social, y la suma total la deposita en la cuenta bancaria del profesional, con los descuentos de práctica. Este servicio lo prestan a cada uno de sus afiliados, cobrando una comisión a descontar de los depósitos. A veces, además se le hace la liquidación de impuestos correspondientes a esas operaciones. Naturalmente, el profesional debe contar con Facturas reglamentarias para poder cobrar a los pacientes particulares. Guardar una copia de todo trabajo realizado y firmado es tan importante como lo es en cualquier negocio el guardar una copia de toda factura emitida. No solo para cumplir ante la DGI, sino para resguardar su firma profesional ante cualquier requerimiento legal que pudiera enfrentar.

3.2 Informes estadísticos

Los métodos para la presentación de datos en Estadística son de tres tipos. Casi siempre se los combina con el objeto de lograr una mayor claridad y transparencia en la información que se transmite al lector. Sería muy engorroso leer un informe donde se detallan uno por uno los cientos de valores obtenidos en la recopilación. Por eso, toda esta *nube de datos*, como se la llama en la jerga estadística, debe ser presentada de manera simple y en lo posible amena. El director del trabajo es el responsable de esta parte y le conviene aprovechar las características más salientes de cada método.



Cualquiera sea el método de presentación adoptado, es obligatorio hacer mención específica de la fuente desde donde se obtuvieron los datos. Se aclara *fuentes propia* cuando el autor es quien hizo las mediciones. Cuando se trata de una fuente secundaria, terciaria o alguna otra, entonces se deben aclarar los datos de la misma, de manera tal, que el lector pueda hallarla, en caso de ser necesario. Para esto, hay una costumbre casi universal de detallar los datos identificatorios de la fuente, ya sea un libro, ya sea un artículo de una revista científica. Como se muestra a continuación:

1) Libros:

Lison, L. ; *Estadística aplicada a la Biología experimental*; EUDEBA, 1ª Ed., Bs.As. 1976

Primero se coloca el apellido del autor principal y separadas con una coma, sus iniciales. Luego en letra cursiva se pone el título entre dos punto y coma. A continuación la editorial que lo publicó, el número de la edición y el año. A veces se coloca la ciudad donde se editó.

2) Artículos:

Davis, P; Russell, AS; Percy, JS: A comparative study of techniques for the detection of antibodies to native deoxyribonucleic acid. *Am J Clin Pathol* , **67**, 374-378, 1977

En este caso, se colocan los apellidos y las iniciales de los autores, luego el título del artículo en letra común. El nombre de la revista se coloca en letra cursiva, pero con la abreviatura oficial aprobada de la misma. En el ejemplo es la *American Journal of Clinical Pathologist*. En negrita se coloca el número de tomo: **67**, luego las páginas de inicio y fin del artículo, más el año de publicación. En los textos, se acostumbra colocar una llamada usando un número escrito en superíndice, correspondiente a la bibliografía detallada al final del libro o del capítulo. Así, el lector puede profundizar un tema específico de su interés sabiendo con certeza a qué fuente recurrir en cada tema.

3.3. Método textual

Este método de presentación de la información consiste en el empleo de palabras y cifras combinadas en un texto, para informar los datos obtenidos. En el Cuadro 3.1 se puede ver un ejemplo. Hay un pequeño párrafo que explica el origen de los datos, junto con 400 números desordenados, para que el lector los analice y trate de sacar sus propias conclusiones.

CUADRO 3.1: Datos obtenidos de un Servicio de Unidad Coronaria.

De los archivos del Servicio se tomaron todas las fichas de pacientes ingresantes con diagnóstico presunto de *Infarto de Miocardio*. Estas fichas se separaron en dos grupos: en uno se ubicaron los que no tuvieron infarto como se verificó después, y en el otro aquellos con diagnóstico confirmado. Se desecharon los casos dudosos. De cada grupo se escogieron al azar 200 casos y se tomó el valor de CPK expresado en unidades enzimáticas. Los valores obtenidos fueron:

SIN INFARTO: 10 - 33 - 20 - 2 - 76 - 38 - 8 - 63 - 76 - 80 - 15 - 18 - 37 - 23 - 22 - 7 - 3 - 2 - 30 - 35 - 39 - 6 - 11 - 17 - 20 - 23 - 19 - 18 - 26 - 28 - 21 - 77 - 1 - 14 - 4 - 6 - 33 - 30 - 39 - 40 - 8 - 22 - 2 - 18 - 13 - 12 - 23 - 33 - 16 - 18 - 67 - 70 - 34 - 19 - 98 - 38 - 11 - 5 - 9 - 30 - 23 - 28 - 8 - 12 - 13 - 24 - 13 - 14 - 19 - 22 - 32 - 12 - 79 - 25 - 63 - 5 - 19 - 16 - 3 - 7 - 9 - 10 - 13 - 17 - 2 - 4 - 5 - 6 - 6 - 6 - 18 - 91 - 3 - 11 - 10 - 88 - 15 - 19 - 20 - 11 - 12 - 18 - 2 - 6 - 7 - 8 - 9 - 11 - 10 - 10 - 17 - 3 - 99 - 5 - 2 - 8 - 1 - 40 - 115 - 9 - 11 - 12 - 32 - 72 - 14 - 24 - 23 - 18 - 19 - 3 - 5 - 22 - 29 - 53 - 55 - 1 - 12 - 13 - 105 - 10 - 3 - 5 - 41 - 55 - 56 - 44 - 7 - 79 - 9 - 6 - 59 - 1 - 66 - 43 - 2 - 65 - 48 - 44 - 5 - 70 - 77 - 9 - 7 - 5 - 59 - 6 - 12 - 11 - 13 - 14 - 21 - 66 - 23 - 61 - 80 - 55 - 45 - 42 - 60 - 58 - 47 - 43 - 56 - 59 - 54 - 49 - 48 - 45 - 44 - 56 - 45 - 49 - 81 - 88 - 100 - 110 - 97 - 120 - 92 - 90.

CON INFARTO: 23 - 44 - 55 - 81 - 100 - 88 - 93 - 99 - 40 - 34 - 57 - 21 - 35 - 122 - 56 - 37 - 154 - 45 - 39 - 56 - 41 - 60 - 49 - 43 - 56 - 51 - 52 - 141 - 56 - 57 - 59 - 150 - 29 - 61 - 128 - 80 - 85 - 87 - 90 - 91 - 97 - 66 - 65 - 150 - 144 - 61 - 69 - 81 - 100 - 88 - 93 - 99 - 77 - 76 - 155 - 74 - 79 - 68 - 80 - 62 - 64 - 73 - 75 - 132 - 66 - 77 - 68 - 132 - 70 - 71 - 73 - 77 - 61 - 63 - 72 - 71 - 74 - 76 - 61 - 167 - 66 - 180 - 179 - 69 - 121 - 140 - 123 - 122 - 130 - 134 - 138 - 140 - 124 - 85 - 87 - 90 - 91 - 97 - 127 - 129 - 132 - 131 - 139 - 135 - 134 - 123 - 133 - 139 - 101 - 120 - 102 - 100 - 85 - 87 - 90 - 91 - 97 - 111 - 113 - 115 - 103 - 104 - 106 - 108 - 110 - 118 - 116 - 81 - 100 - 88 - 93 - 99 - 114 - 112 - 110 - 105 - 103 - 107 - 106 - 109 - 107 - 112 - 115 - 119 - 120 - 102 - 104 - 103 - 105 - 81 - 100 - 88 - 93 - 99 - 109 - 107 - 105 - 85 - 87 - 90 - 91 - 97 - 103 - 104 - 106 - 107 - 112 - 115 - 118 - 111 - 102 - 111 - 113 - 115 - 103 - 104 - 106 - 108 - 110 - 118 - 116 - 114 - 112 - 110 - 105 - 103 - 107 - 106 - 109 - 102 - 81 - 100 - 88 - 93 - 99 - 115 - 103 - 104 - 106 - 108 -

Desde el punto de vista del lector, este método es el más engorroso de los tres para poder captar el significado de los valores obtenidos y sacar alguna conclusión. En el ejemplo del Cuadro 3.1 se aprecia que los que no tuvieron infarto, en general tienen un valor menor de CPK que los infartados. Pero como hay muchos valores coincidentes en ambos casos, se hace confuso el uso de esta técnica clínica para diagnosticar un infarto. Sin embargo, un análisis más cuidadoso permitiría ver que las cosas no son así. En forma textual se informaría: "... el primer valor obtenido con los infartados fue 10, el segundo 33, el tercero 20...", y así sucesivamente.

Desde el punto de vista del redactor, el método textual tiene una ventaja importante con respecto a los otros: se puede influenciar al lector. El autor puede resaltar ciertas cifras de su interés, puede remarcar conceptos apropiados para sus fines y hacer pasar desapercibidos a los otros. Se puede focalizar la atención del lector, de tal manera que pase por alto ciertos datos evitando que saque sus propias conclusiones. Todo lector prevenido debe desconfiar cuando encuentra juicios de valores en un trabajo científico. Se ha preparado un pequeño ejemplo ilustrativo de estos conceptos en el Cuadro 3.2. Para entenderlo mejor, es conveniente seguir los pasos indicados: primero se debe leer el texto de la Parte A. Luego, en un papel, el lector escribe las cifras que recuerde y las conclusiones que saca. Por último, lee la Parte B y la compara con sus propias conclusiones:

CUADRO 3.2: Ejemplo de un método textual.

Paso 1) Leer una sola vez el párrafo identificado como Parte A.

Parte A: "La actual comisión directiva de nuestro club de canotaje, que hoy finaliza su tarea, tiene el gusto de presentar a vuestra consideración los exitosos logros alcanzados durante su gestión, tales como el incremento de un 400% en la cantidad de remos extras, y todavía un bote tipo doble más. Se *mejoró* la seguridad y vigilancia de los bienes de nuestro club de dos años de vida, en un 200%. Con gran previsión de futuro y preocupación por el equilibrio ecológico, se *incrementó en un 500%* el capital arbóreo. Los *ingresos* por cuota societaria *aumentaron* en un 50%, acrecentando a su vez el número de socios. Con esta exitosa gestión financiera de este período, se construyó un nuevo vestuario para mayor comodidad de nuestros asociados mejorando substancialmente el aspecto sanitario..."

Paso 2) Escribir en un papel lo que se recuerda de la primer lectura y opinar sobre la gestión realizada por la comisión directiva saliente.

Paso 3) Leer el Parte B, comparar con la anterior comisión directiva para ver si se mantiene la opinión escrita en el Paso 2.

Parte B: "...Del Resumen y Balance de los dos primeros períodos del club se han extraído las siguientes cifras:

	1ª Gestión	2ª Gestión	
- Cantidad de Botes	20	21	
- Cantidad de Remos	42	50	
- Árboles plantados	5	25	
- Número de serenos	1	3	
- Cuota mensual del socio	30\$	40\$	
- Número de socios	110	112	
- Superficie construida	600 m ²	606 m ²	(los 6 m ² se destinaron a vestuario)

Puede notarse del Cuadro 3.2, que mientras las primeras autoridades del club pudieron comprar 20 botes, la segunda solo uno, pero al comprar 8 remos de más, incrementaron en un 400% la cantidad de remos sobrantes. Colocaron 20 plantines de árboles en el terreno y con eso dicen que incrementaron un 500% el “capital arbóreo”. El alza de un 50% en los ingresos del club no fue aumentando el número de socios, sino subiendo la cuota. Contrataron 2 serenos, cuando antes con uno alcanzaba. Construyeron 6 m², cuando los anteriores lograron centuplicar ese valor. Sin embargo, todas estas cosas no surgen del texto del discurso de la comisión saliente, sino de una tabla comparativa que presenta sólo cifras. Esto ilustra la conveniencia de adjuntar a los textos las tablas correspondientes, para hacer más transparentes los conceptos que se puedan verter. Una tabla no puede influir al lector tanto como lo puede hacer un texto.

3.4. Método tabular

Este método de presentación de la información consiste en presentar los datos por medio de una tabla o cuadro. En esta, se debe colocar al principio un *título* identificatorio que en forma clara, breve y completa, explique el contenido de la tabla. Luego viene el *cuerpo* y al final se debe colocar la *fuentes* de donde se tomaron los valores mostrados. Una tabla simple consta de un cuerpo dividido en columnas, cada una con su encabezado y una aclaración de las unidades que se están usando. Una tabla de doble entrada consta de un cuerpo dividido en filas y columnas, con sus respectivos encabezados, como la del Cuadro 3.3. Si se diera el caso de usar las mismas unidades en toda la tabla, entonces conviene especificar las mismas en el título, para simplificar. El cuerpo de la tabla propiamente dicho muestra los datos numéricos. Conviene emplear recuadros para facilitar la lectura y ubicación de los valores.

CUADRO 3.3: Ejemplo del método tabular.

Valores de CPK extraídos de fichas de pacientes de una Unidad Coronaria (medidos en UI.)

Valores de CPK medidos			Diagnósticos	Confirmados
			Sin Infarto	Con infarto
1	a	20	104	0
21	a	40	36	8
41	a	60	28	16
61	a	80	18	32
81	a	100	10	46
101	a	120	4	66
121	a	140	0	23
141	a	160	0	6
161	a	180	0	3

FUENTE: Cuadro 3.1 anterior.

Para armar la tabla del Cuadro 3.3, se usaron los valores del Cuadro 3.1, agrupándolos en *clases* de igual tamaño como se ve en la columna *matriz* o principal donde van los encabezados de las filas. En este caso, el encabezado de la cuarta fila es: “61 a 80”. Para cada fila (*intervalo*), se cuenta el número de veces (*frecuencia*) que un valor cae dentro del mismo, tanto para el caso de infartados como no-infartados. Por ejemplo, el valor “23” de la séptima fila y tercera columna significa que se han encontrado 23 casos de pacientes infartados, cuya CPK está comprendida entre los valores 121 y 140 UI y ningún caso entre los no infartados por lo que su valor es “0”.

Una tabla es el método más imparcial para presentar la información, muestra los datos crudos, dejando al lector la tarea de interpretarlos sin hacer ni sugerencias ni comentarios. Al contrario de un gráfico, no se puede captar la idea general de un simple vistazo salvo a trazos gruesos. Como por ejemplo decir usando el cuadro anterior, que cuanto mayor sea la CPK mayor la posibilidad de tener un infarto.

3.5. Método gráfico

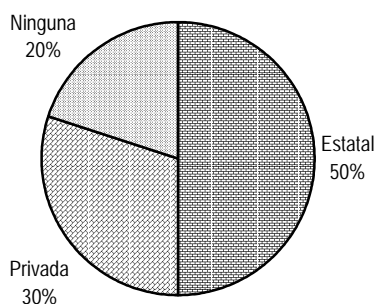
Este método de presentación de la información es el más simple para el lector porque puede captar el panorama general, o la tendencia de los datos, de un solo vistazo. Es mucho más fácil de comprender que una tabla o un texto. La sencillez de líneas, una atractiva manera de presentación, la posibilidad de usar las tres dimensiones (3D), junto con colores, hacen de los gráficos una de las herramientas más poderosas para transmitir ideas en forma rápida y simple al lector. Su desventaja más notoria es la pérdida de precisión y exactitud, si se lo compara con una tabla. Pero siempre se puede sacrificar algo en aras de la sencillez. Por estas razones, la mayoría de los autores usan las tres formas vistas (métodos mixtos) para la redacción de sus obras. La idea es aprovechar las ventajas de cada uno de ellos. Cada gráfico es un “original”, quedando a criterio de su autor sus dimensiones y formas, pero sometidas a las “reglas del buen arte”. Esta libertad de creación provoca un gran número de posibilidades. Aquí, sólo se muestran los tipos más usuales en clínica, usando una clasificación muy general. Para quien necesite crear sus propios gráficos, conviene repasar algunos consejos de expertos en el tema:

- antes de efectuar el gráfico definitivo, conviene hacer varios modelos diferentes en borrador para poder elegir entre ellos al más adecuado;
- la disposición del gráfico debe hacerse de izquierda a derecha;
- debe colocarse siempre el cero de la escala cuando alguna es usada en el mismo;
- para las comparaciones conviene emplear una sola dimensión, antes que dos o tres;
- en los gráficos de porcentaje acumulativo además del nivel cero se debe colocar el 100%;
- la línea más gruesa de todas debe ser la del gráfico o curva que se muestre, para subordinar las demás a la principal
- la línea de ayuda visual en el gráfico debe ser la más fina de todas;
- debe tener un título claro, conciso y completo;
- debe indicar la fuente de donde se extrajo la información;
- deben colocarse siempre la escala empleada y las unidades de las magnitudes mostradas;
- la escala de un gráfico debe adaptarse para que entre toda la información en el mismo;
- las normas IRAM de dibujo técnico deben emplearse en la medida de lo posible;
- no se debe usar un gráfico para mostrar la información en modo tendencioso;
- si un gráfico no resulta claro para el autor, mucho menos lo será para el lector.

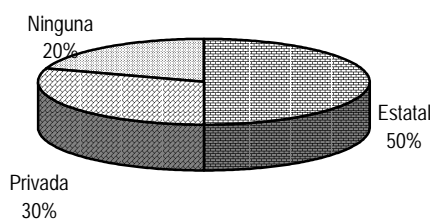
3.5.1 Gráfico circular

A este tipo de gráfico se lo conoce con su nombre más popular como el *diagrama de la torta* (*pie-chart*) o *reparto de la torta*. Consiste en repartir el área del círculo en trozos proporcionales a las cantidades que se están representando. En el Gráfico 3.3 se muestra un ejemplo, usando estos valores inventados. Se muestra el diagrama en dos y tres dimensiones; para una mejor ilustración se le pueden agregar colores.

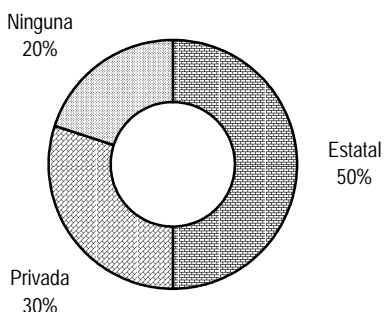
Gráfico 3.3: Pacientes por tipo de Obra Social atendidos en el L.A.C. durante julio de 1999



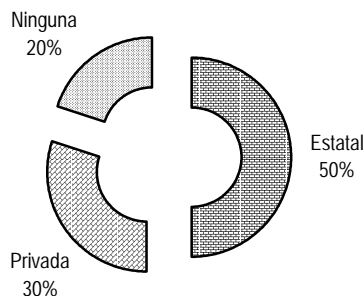
2 D



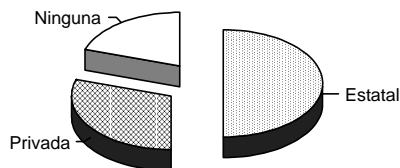
3 D



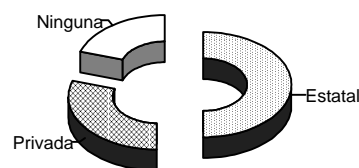
2D anillo



2D anillo separado



3 D separado



3 D anillo

Fuente: Propia.

Por ejemplo, si se deben presentar tres tipos diferentes cuyos porcentajes del total sean respectivamente: 50%, 30% y 20%, entonces se divide el círculo en tres sectores de superficie proporcional. El sector circular de la primera deberá tener un ángulo que sea el 50% de los 360° del ángulo total, o sea, abarcará la mitad del círculo con un ángulo de 180°. El segundo ocupará el 30% del total con un sector de ángulo de 108°. Y el tercero tendrá un ángulo de 72°. En el ejemplo anterior se presentan seis formas de ilustrar la misma información. En la primera se ha trazado un gráfico en 3D al modo tradicional. El semicírculo más grande corresponde a los pacientes que llegan al laboratorio con una orden de prestación autorizada en una Obra Social de tipo estatal, en este caso la mitad de los pacientes totales. El sector más chico corresponde al 20% de los pacientes atendidos durante el mes de julio de 1999 y son todos aquellos que no están cubiertos por una Obra Social, o sea, los pacientes “particulares”. El restante sector, de un 30%, corresponde a los que provienen de Obras Sociales de tipo privada.

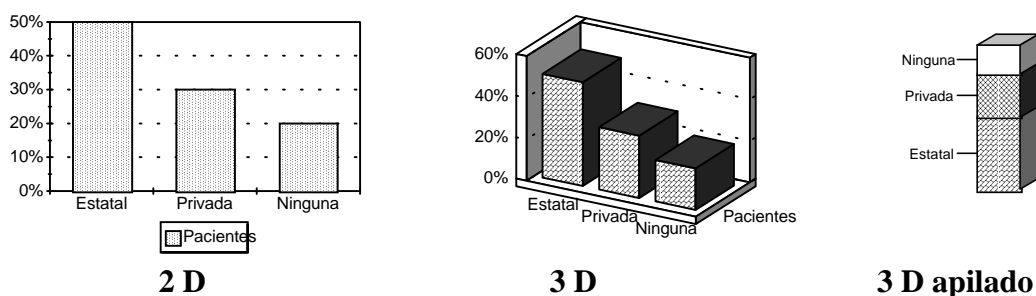
Todo esto se capta más rápido dando una ojeada al gráfico que leyendo todo el párrafo anterior, e ilustra sobre las ventajas de este método con respecto al textual. Se presentan también otros tres círculos usando el modo tridimensional, pero con técnicas diferentes, para ilustrar al lector sobre algunas de las posibilidades disponibles para poder elegir entre ellas.

3.5.2 Gráfico de barras

Este tipo de gráfico comparativo es uno de los más usados en clínica por la sencillez de su construcción. Consiste en representar las cantidades con rectángulos de igual base, y de altura proporcional a los valores respectivos. Los intervalos libres entre barras también deben ser del mismo tamaño, aunque a veces algunos autores las muestran pegadas.

Gráfico 3.4: Gráfico de barras en 2D y en 3D.

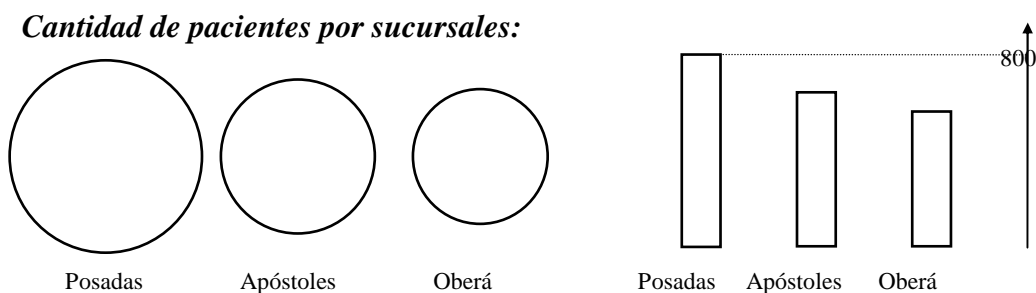
Pacientes por tipo de Obra Social atendidos en el L.A.C. durante julio de 1999:



En el Gráfico 3.4 se muestra un ejemplo clásico del diagrama de barras en dos dimensiones, usando los mismos datos del punto anterior. La línea punteada ayuda a ver mejor los datos numéricos, y un ligero sombreado destaca mejor las barras. Al gráfico tridimensional se lo gira en cierto ángulo para obtener una mejor apreciación por parte del lector. Además, se han engrosado las escalas para obtener una mejor sensación de profundidad. Por último, se muestra un gráfico tridimensional apilando las barras una encima de otra. Todo esto se puede hacer pues solo se representan tres categorías; cuando estas son muchas más no conviene apilar ni recargar tanto de información al eje de abscisas. Las grandes ventajas del gráfico de barras son: sencillez de cons-

trucción y facilidad de captación por parte del lector. Esto último se debe a que visualmente se compara una sola dimensión: la altura, pues como las bases son todas iguales no se les presta atención. Muy distinto sería si hay que comparar más “dimensiones”, por ejemplo, si en lugar de comparar barras se comparasen círculos (ver Gráfico 3.5), cuadrados o rombos. En estos casos, es imposible hacer abstracción mental de una de las dos. En tales casos, las diferencias de tamaño son sutiles y muy engañosas.

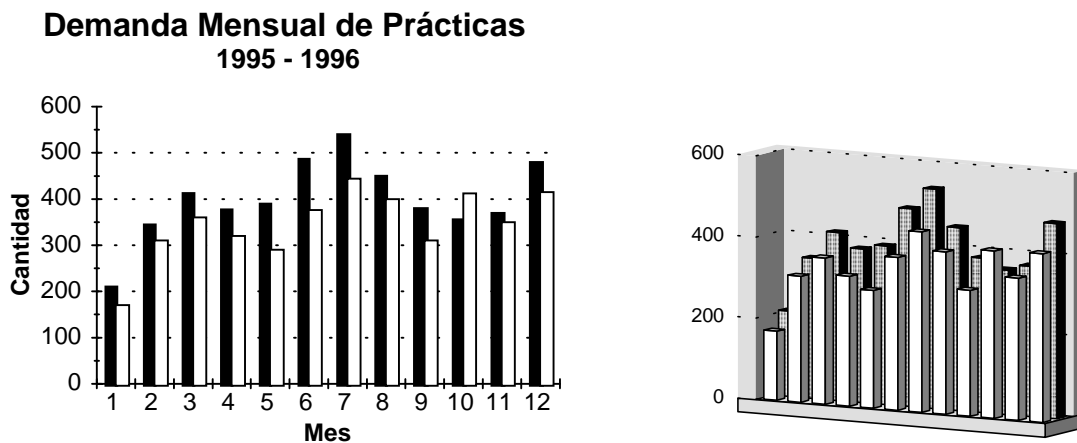
Gráfico 3.5: Gráficos comparativos de círculos versus de barras.



NOTA: las cantidades representadas son Posadas: 800, Apóstoles: 584 y Oberá: 496

Se pueden superponer dos o tres diagramas de barras para hacer comparaciones, tal como se estila en las series temporales donde el foco de atención se concentra en las variaciones anuales o semestrales. En el Gráfico 3.6 siguiente se ilustra esta situación.

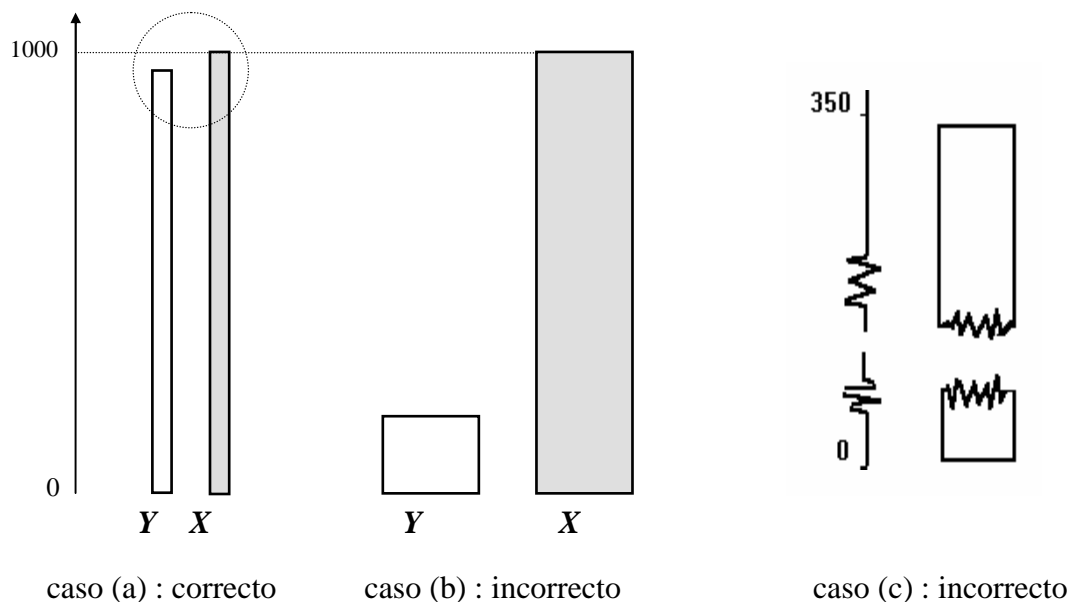
Gráfico 3.6: Gráfico de barras comparativas en 2D y en 3D. (1995-1996).



Un mal uso de las reglas puede inducir al lector a una mala interpretación. Por ejemplo, si no se indica el nivel de cero, se pueden magnificar diferencias muy pequeñas para que el lector de un vistazo crea que son mucho más grandes. En el Gráfico 3.7 se presentan dos situaciones: en el caso (a) se muestran dos barras, la sombreada representa una venta para el producto X de mil unidades, la otra barra en blanco representa al producto Y con 980 unidades. O sea, la situación es que X se vende nada más que un 2% más que Y. Si se mira el caso (b) se ve una gran diferencia entre ambas barras, y da la impresión que X se vende mucho más que Y. Este engaño es posible pues en el caso (b) no se muestra el nivel cero. Es como obligar al lector a ver las diferencias con una “lupa” (el círculo punteado) y eso no es correcto. Algunas veces se quiebra el

dibujo mostrando la parte de arriba como el caso (b), pero agregando la parte de abajo con el nivel cero, y una escala cortada. Esto engaña mejor al lector - caso (c).

Gráfico 3.7: Cómo influenciar al lector con un gráfico de barras.



3.5.3 Pictogramas

Son gráficas comparativas de imágenes donde se usan símbolos para representar las magnitudes que se están usando. Por ejemplo, si el tema tratando es el número de nacimientos, se hace un esquema de un bebé. De esa forma, el lector puede captar más rápido la idea. Para formar los pictogramas deben tenerse en cuenta los puntos siguientes:

- . Una vez adoptado el símbolo a usar, se le asigna un tamaño fijo y una cantidad de unidades.
- . Para representar una cantidad dos veces mayor no se aumenta el tamaño del símbolo al doble, se deben usar dos símbolos y análogamente para múltiplos enteros.
- . Para representar fracciones de unidades “pictográficas” se puede dividir al símbolo hasta su cuarta parte. O sea, se deben aproximar las fracciones hasta el 25% de la unidad definida.
- . Para cantidades mucho mayores conviene combinar el pictograma con otro diagrama.
- . El símbolo adoptado como pictograma debe explicarse por sí mismo.

Si bien el pictograma es el gráfico más llamativo a su vez es el más engañoso, por la propensión que se tiene a comparar áreas y volúmenes, antes que líneas. En el Gráfico 3.8 se ilustran estos puntos. Se muestran una serie de símbolos y en cada uno se coloca la escala correspondiente. Por ejemplo, para mostrar 50 insectos se debe dibujar la mitad del pictograma, pero no es co-

recto reducirlo a la mitad. Para graficar 100 tractores se ponen dos figuras completas y no uno solo, de doble tamaño.

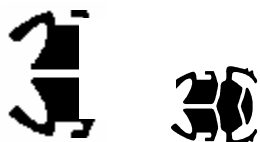
Gráfico 3.8: Pictogramas

1. Ejemplos de pictogramas con los valores que representan



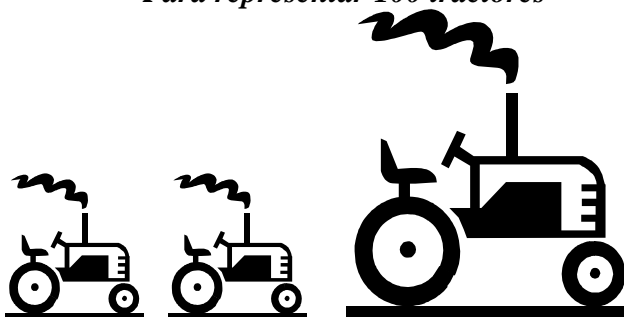
2. Ejemplos de usos de los pictogramas para mostrar cifras

Para representar 50 Insectos



Forma correcta Forma incorrecta

Para representar 100 tractores



Forma correcta

Forma incorrecta

3.5.4 Gráficos cronológicos

Se denomina así a todo gráfico de dos o más variables, una de las cuales es el tiempo. Los más difundidos son los de dos variables, con el tiempo en el eje de abscisas. Se usan para mostrar la evolución en el tiempo de una magnitud cualquiera. Se clasifican en dos tipos:

- . Simples: se usan para representar una sola serie de datos homogéneos.
- . Compuestos: son utilizados para comparar dos o más series de datos cronológicos.

En el Gráfico 3.9 se pueden ver los mismos datos del Gráfico 3.6, pero ahora mostrados como dos series cronológicas anuales. Utilizando diferentes escalas se puede resaltar una tendencia sobre la otra, tanto en un sentido como en el otro. Si se busca hacer más notorio el hecho de que 1996 fue un mejor año que 1995 para el laboratorio del ejemplo, entonces se agrandan las ordenadas y se comprimen las abscisas. O viceversa, si se busca el efecto contrario.

Gráfico 3.9: Comparación de dos series cronológicas.

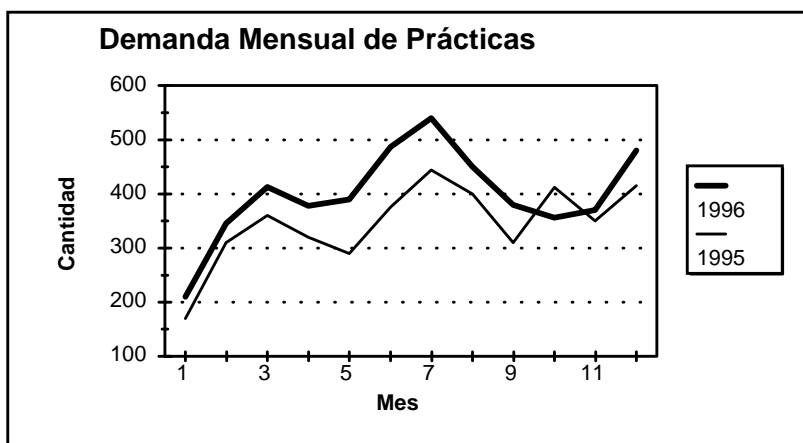
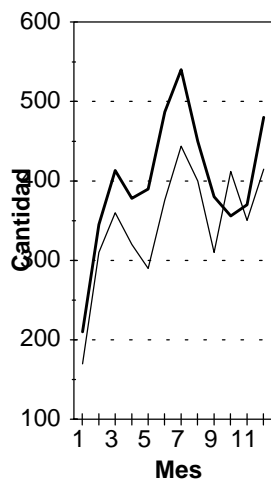
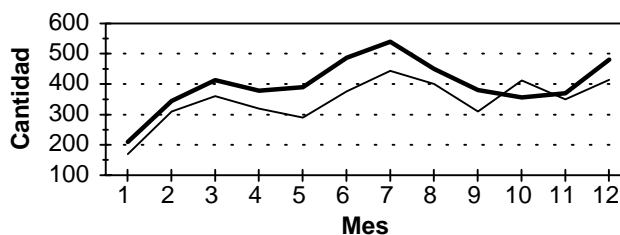


Gráfico 3.10: Dos formas incorrectas de presentar la misma información



Para resaltar al año 1996 sobre 1995

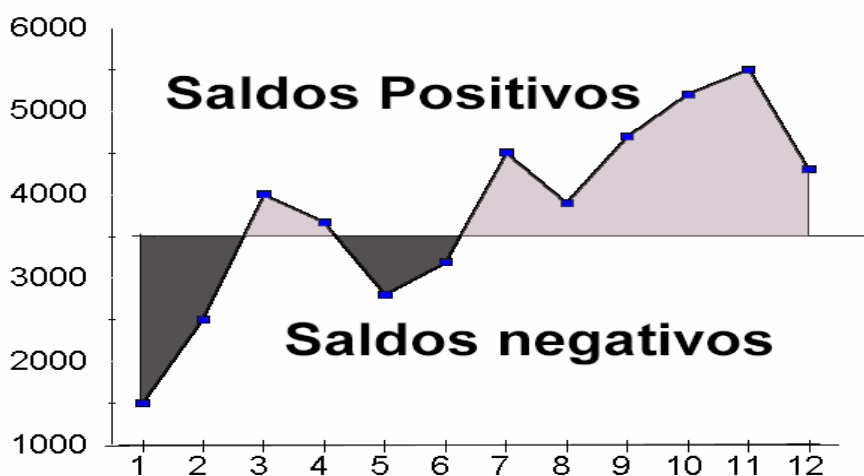


Para atenuar las diferencias entre las series

3.5.5 Diagrama de saldos

Es una variante de las series cronológicas donde lo que interesa es mostrar las fluctuaciones de la magnitud, respecto de un nivel de referencia. El ejemplo clásico es el diagrama que muestra las variaciones de las ganancias en el tiempo, como se aprecia en el Gráfico 3.11. Por regla general se somborean las áreas ubicadas entre la curva y el nivel de referencia, como ayuda visual al lector. De esa manera se destacan los saldos positivos, ubicados por encima del nivel, de los negativos.

Gráfico 3.11: Ingresos (\$) de un laboratorio en un año, con gastos fijos de \$ 3.500/mes.



3.6 Histogramas

Este es un método especial de presentar los datos de mediciones de un laboratorio. Es el más utilizado cuando se tienen muchos datos obtenidos al medir una magnitud clínica, algunos de los cuales pueden repetirse varias veces. Para armarlo, se deben colocar en el eje de abscisas los valores que adopta la magnitud, y en ordenadas se muestra el número de veces que se repite el dato, o sea, su *frecuencia*.

El procedimiento es simple. Primero se buscan el valor máximo y el mínimo de todo el grupo de datos. Luego se ordenan los valores en forma creciente y se hace el recuento, anotando el número de veces que se repite cada uno. La forma más sencilla para hacer el recuento es colocar un palote al lado de cada número en forma vertical, y cruzarlos con el quinto (///). Aunque algunos prefieren usar el símbolo del cuadrado cruzado:

Cuando se grafica cada valor individual, con su frecuencia correspondiente, se tiene el llamado: *Diagrama de bastones*. En los histogramas, en cambio, se agrupan los datos en *clases* para evitar un número muy grande de valores en las abscisas, que dificultaría su representación. Además, esto simplifica el recuento. Cuando la cantidad de datos supera el medio millar, se acostumbra usar entre 10 y 20 clases. Cuando son menos, se usan de 5 a 10 clases, según convenga, por estética y por practicidad. Para determinar la cantidad de clases conviene definir primero el

ancho de clase (a). Por ejemplo, tomando los datos del Cuadro 3.1 para los pacientes sin infarto, se puede determinar que el valor mínimo de la CPK fue de 1 UI/l, mientras que el máximo fue 120 UI/l, luego el *rango* será 119 UI/l. Se define al rango, como la diferencia entre el valor máximo y el mínimo del grupo de datos. Con ese valor se puede usar una fórmula empírica para determinar el ancho de clase:

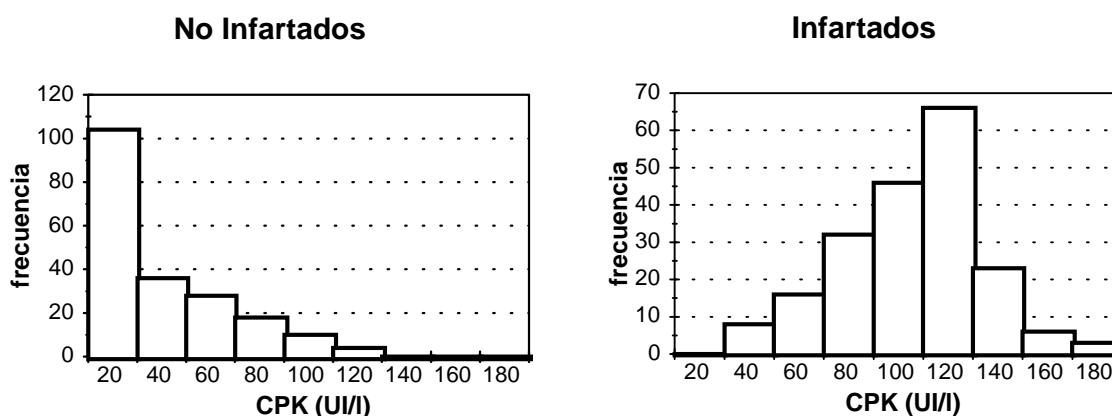
$$\frac{\text{rango}}{10} > a > \frac{\text{rango}}{20} \quad (\text{más de 500 datos}) \quad \text{o} \quad \frac{\text{rango}}{5} > a > \frac{\text{rango}}{10} \quad (\text{para menos de 500})$$

En el ejemplo, como son 200 datos, se usa la fórmula de la derecha, y se obtiene un ancho de clase comprendido entre:

$$23,8 > a > 11,9$$

por comodidad conviene usar un número entero y redondo, como ser $a = 20$. Una vez adoptado el ancho de clase, se arman las clases empezando del menor valor. Esto se hizo en el Cuadro 3.3 utilizando los datos mostrados en el Cuadro 3.1. El resultado se muestra en el Gráfico 3.12 donde se presentan los dos histogramas.

Gráfico 3.12: Histograma de los datos del Cuadro 3.1, agrupados en el Cuadro 3.3



Límites de Clase (LC): son los dos valores extremos del intervalo de cada clase. Por ejemplo, el límite de clase inferior de la primera clase es 1, mientras que 20 es el superior.

Límites Reales de Clase (LRC): son los valores obtenidos al sumarle la mitad de la menor unidad de la escala empleada al límite de clase superior, y restarle al inferior. En el ejemplo anterior serían: 0,5 y 20,5. Estos se usan para evitar discontinuidades en el gráfico cuando la magnitud clínica es continua. Así, el límite real superior de una clase siempre debe ser igual al límite real inferior de la clase siguiente, y así sucesivamente.

Marcas de clase: es el punto medio del intervalo de clase. Se obtiene como la semisuma de los dos límites reales de clase. En el ejemplo sería: $(0,5 + 20,5) / 2 = 10,5$.

Ancho de clase: es la diferencia entre los dos límites reales de una clase cualquiera. Por ejemplo, en la primer clase será: $a = 20,5 - 0,5 = 20$ y análogamente en las demás.

Frecuencia: es la cantidad de datos que contiene cada clase. En la primera clase de los no infartados la frecuencia es de 104, eso significa que hay 104 datos cuyos valores son menores que el límite real superior de esa clase.

Frecuencia relativa: es la frecuencia de cada clase dividido el número total de datos.

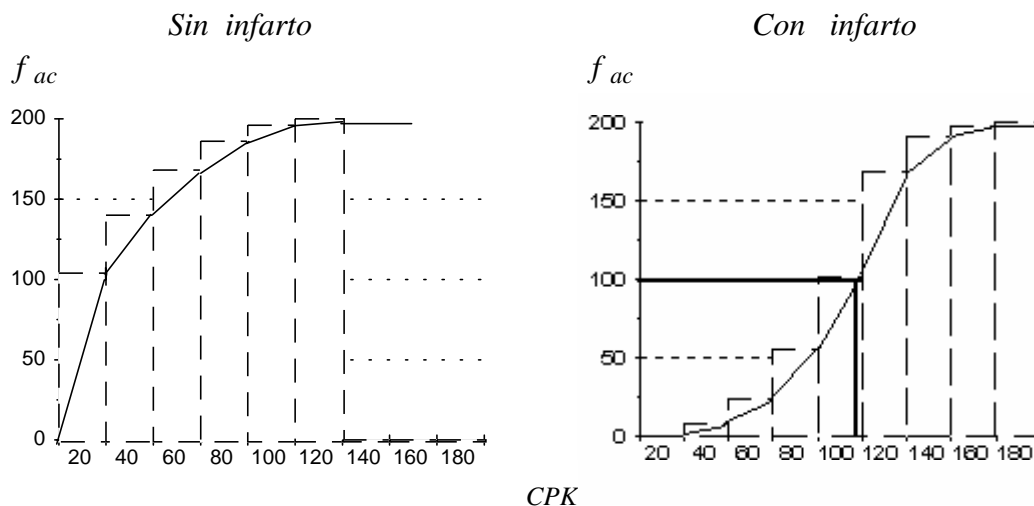
Frecuencia acumulada: de una clase cualquiera, es la suma de la frecuencia de esa clase más la sumatoria de las frecuencias de todas las clases anteriores.

Al agrupar en clases se pierde información. Ya no se sabe la frecuencia de cada dato individual, sino la de todo el grupo incluido en esa clase. Pero por otra parte, se simplifican los recuentos y los gráficos. Se pierde precisión en aras de la simpleza.

3.7 Polígono de frecuencias acumuladas

Otra manera de presentar los histogramas, es con el polígono de frecuencias acumuladas, que se muestra en el Gráfico 3.13. Para trazarlo, se coloca la magnitud clínica en el eje de abscisas, subdividida en clases. En el eje de ordenadas se pone la frecuencia acumulada de cada clase así resulta que cada barra se forma superponiéndola con la de su izquierda (línea punteada gruesa del gráfico). El polígono se traza comenzando en el origen y trazando una recta que lo une con el extremo superior derecho de la primera barra. Desde ese punto, se traza otra recta que lo una con el mismo extremo de la segunda barra y así sucesivamente.

Gráfico 3.13: Polígono de frecuencias acumuladas (datos del Gráfico 3.12).

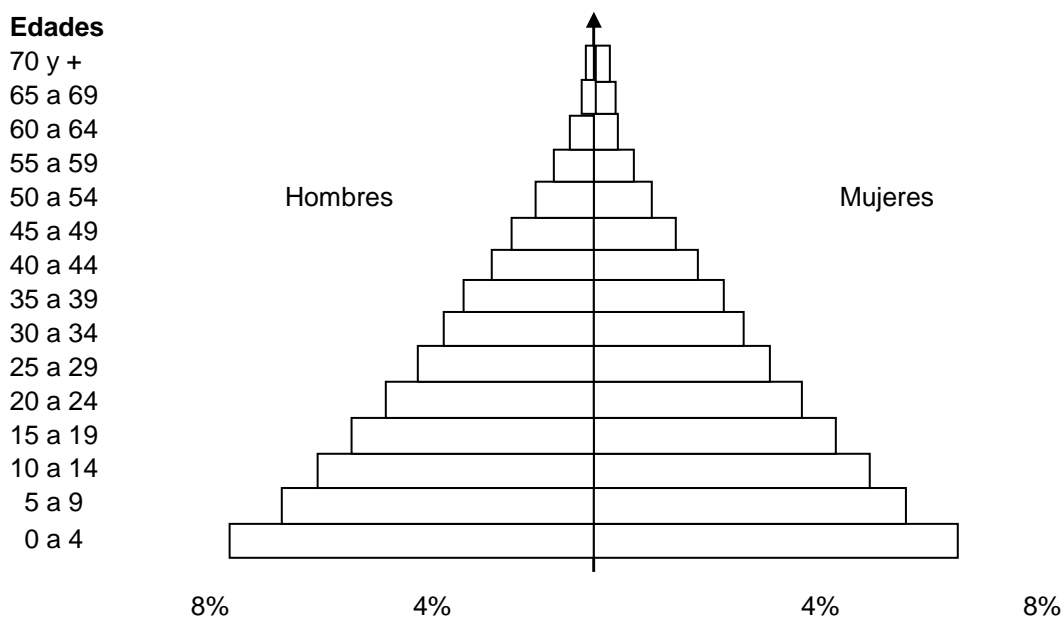


El uso práctico del polígono de frecuencias acumuladas, es para determinar en forma gráfica y rápida el porcentaje de observaciones que quedan por debajo de cierto valor, o viceversa. Por ejemplo, si se deseara saber qué valor deja el 50% de los valores por debajo de él -la mediana- se toma la frecuencia acumulada 100 y se traza una paralela al eje de abscisas hasta tocar la curva, de allí se baja una vertical, y el valor que se lee de la CPK es la mediana. En el Gráfico 3.10, para el caso de los infartados, el valor estimado de la mediana es 115 UI/l.

3.8 Pirámides de población

Este es un caso especial de histograma. Se usan dos histogramas, uno para hombres y otro para mujeres, que se ubican en forma vertical y enfrentados uno con otro como se puede ver en el Gráfico 3.14 siguiente. Las edades se agrupan en intervalos de 5 años de edad; el primero de 0 a 4 años, el segundo de 5 a 9 años y así sucesivamente (eje de ordenadas). Mientras que en el eje horizontal se representa el porcentaje de ese grupo de edades en la población total. Este diagrama, llamado también “pirámide de edades”, se emplea mucho en Demografía. En Análisis Clínicos se usan sus datos para preparar los valores de referencia poblacionales de las magnitudes clínicas (como se verá más adelante), los cuales se vuelcan en una tabla del tipo Peso-talla. En Farmacia se lo emplea para tener una idea de la composición de la población circundante, a los efectos de diseñar técnicas de mercadeo.

Gráfico 3.14: Pirámide de población



Es interesante destacar la forma de la pirámide. Cuando adopta la forma de un pino se tiene una población del tipo “rural”. En cambio, si la población es “urbana” toma la forma de un ataúd, angosta en su base (pocos nacimientos), ancha en su parte media (muchos jóvenes en edad de trabajar) y truncada en su punta (mayor promedio de edades). Es la forma típica de las grandes ciudades como Buenos Aires, New York, Tokio, Los Ángeles, y otras.

De sus datos se pueden sacar conclusiones de la tasa de mortalidad, tasa de fecundidad, movimientos de la masa laboral y hasta de la historia de una población.

3.9 Recetas en Farmacia

El tema de presentación de datos en Farmacia, a diferencia de Bioquímica, tiene dos fases. En la primera se trata de verificar toda la información requerida en una receta para poder efectuar la venta de medicamentos. En la segunda se trata de presentar toda la información a cada Obra Social para poder recuperar los descuentos realizados a sus afiliados. La diferencia principal con Bioquímica es que, mientras en esta hay que elaborar el informe de las prácticas realizadas a cada paciente, en Farmacia hay que revisar lo recetado por el médico para ver si se ajusta a los requisitos vigentes. En ambas, estas tareas se realizan diariamente, y mensualmente se presentan las facturas de sus servicios a cada Obra Social.

Conviene, entonces, repasar el trabajo diario del farmacéutico con las recetas como puede verse a continuación. Básicamente se verán los requisitos para obtener los descuentos establecidos con las Obras Sociales y todos los pasos a seguir para ello. Se debe recordar que los edulcorantes, los productos dietéticos, los sueros o medicamentos sin troquel y todo los medicamentos de venta libre, no son reconocidos por las Obras Sociales. Estas, solo reconocen aquellos medicamentos incluidos en el *Manual Farmacéutico*, con descuentos que oscilan entre un 30% y un 100% de su costo.

Datos requeridos en la receta

1. Las recetas deben estar escritas de puño y letra del médico actuante y *con la misma tinta*.
2. Nombre y apellido del paciente.
3. Tipo y número de Documento de Identidad.
4. Nombre de la Obra Social (que a veces viene preimpreso en la receta).
5. Número de afiliado.
6. Lista de medicamentos recetados, cantidad de envases y contenido de los mismos.
7. Firma y sello del profesional con su número de matrícula.
8. Completar los casilleros de la receta preimpresa y anular los no utilizados para los medicamentos. En caso de tachaduras, salvar en el anverso con la firma respectiva.
9. Fecha de emisión de la prescripción (por la validez de 15 a 30 días a partir de ella).
10. Firma y aclaración del comprador, su número de documento y su domicilio.
11. La aclaración cuando se trate de *Tratamiento Prolongado*.

Datos a completar en la receta por la Farmacia

1. Las recetas deben tener adheridos los troqueles de los medicamentos vendidos, cubiertos por la Obra Social respectiva.
2. El precio unitario de cada medicamento vendido, la cantidad y su precio total.
3. El precio total a cobrar al afiliado.
4. El precio total a cobrar a la Obra Social por los medicamentos cubiertos por la misma.
5. El total de la receta cumplimentada.
6. La fecha de la venta, con el sello de la farmacia y firma del profesional habilitado.
7. Por conveniencia, se suelen registrar en el anverso de la receta los datos de la persona que retira el producto.

Documentación que debe aportar el paciente

Cada paciente deberá presentarse en la Farmacia, ya sea en persona o a través de un tercero. Deberá exhibir la credencial que lo identifica como afiliado a la Obra Social y su documento de identidad. Para casos como el PAMI, donde no hay credencial, deberá exhibir el último recibo de sueldo y su documento.

Requisitos a verificar dentro de la prescripción médica

1. Hasta 4 envases por receta.
2. Hasta 3 especialidades por receta.
3. Hasta 1 unidad por renglón.
4. Hasta 2 unidades por especialidad del tamaño menor, intermedio o único.
5. Hasta 1 unidad por especialidad del tamaño mayor.
6. Hasta 5 unidades de antibióticos inyectables monodosis.
7. Hasta 2 unidades de antibióticos inyectables multidosis.
8. Cuando la farmacia no disponga del tamaño grande del medicamento y el *afiliado estuviere de acuerdo*, se podrán entregar dos (2) envases chicos en su reemplazo. Pero siempre que la cantidad no supere la indicada en la receta.
9. Cuando por carencia en plaza del medicamento recetado no fuera posible entregar este y en la receta *vengan indicados* medicamentos alternativos o substitutos, entonces se expenderá el de menor precio disponible.
10. Cuando en la receta no venga especificada la cantidad de medicamentos, se deberá expender el de *menor* cantidad.
11. Cuando en la receta venga aclarado el *tratamiento prolongado* se puede vender el tamaño mayor del medicamento. Caso contrario, se debe vender el de menor tamaño.

Caso de psicofármacos (receta archivada)

Para la provisión de estos productos, además de lo anterior, se deberá exigir el duplicado de la receta expedida por el profesional. El cual quedará en poder de la Farmacia y archivado de manera tal de poder exhibirlo en el caso de inspecciones.

Caso de falta de existencias

Cuando un medicamento prescripto faltase, la Farmacia está autorizada a emitir un comprobante donde se compromete a obtenerlo dentro de las 48 horas hábiles, para completar su servicio. Pero si la falta en plaza de dicho medicamento fuese notoria, la Farmacia queda eximida de esta obligación y devolver el dinero, si el medicamento ya fue pago.

Cobro de facturas a las Obras Sociales

Para poder cobrar los descuentos realizados a los clientes por Obra Social, la Farmacia deberá presentar el último día hábil del mes lo siguiente:

1. Cada una de las recetas cumplimentadas en el mes. Estas deberán tener adheridos los troqueles de cada uno de los medicamentos vendidos que estén dentro del *Manual Farmacéutico*, y el sello de la Farmacia, con la firma del profesional responsable.

2. Un resumen de las ventas donde debe figurar:

- Nombre y apellido de afiliado.
- Número de afiliado.
- Medicamentos vendidos con sus cantidades y precios de venta.
- Fecha de venta.

3) Total facturado.

Esta factura se debe confeccionar para cada Obra Social todos los meses. De acuerdo con convenio anual renovable, firmado por ambas partes. Las diferencias que pudiesen aparecer se transforman en Notas de Débito o de Crédito, según corresponda.

3.10 Problemas propuestos

1. Mencionar los principales datos que se deben presentar en un informe de laboratorio.

2. Ídem anterior para una separata como por ejemplo un Proteinograma.

3. El mejor método para presentar los datos es aquel que emplea:

4. Dibujar algunos gráficos para estos datos: (A = 200Kg B = 400Kg C = 200Kg).

5. ¿ Cuántas magnitudes se pueden representar en un histograma () y en uno circular () ?

6. Mencionar las ventajas y desventajas de los métodos de presentación de datos.

- | | | |
|--|---|---|
| 7. El mejor método para influenciar al lector es el método tabular. | V | F |
| 8. En el método gráfico se pierde precisión en aras de la sencillez. | V | F |
| 9. Se debe siempre indicar la fuente de donde se tomaron los datos. | V | F |
| 10. Para las comparaciones conviene usar muchas dimensiones. | V | F |
| 11. Cuando se usa escala en un gráfico, siempre debe colocarse el cero de la escala. | V | F |
| 12. Las líneas con las cuales se trazan los gráficos deben tener todas el mismo grosor. | V | F |
| 13. Los gráficos circulares y de barras se pueden presentar en dos y tres dimensiones. | V | F |
| 14. No se debe cortar la línea de la escala en un diagrama comparativo de barras. | V | F |
| 15. En un pictograma hay que usar un icono de tamaño doble, para mostrar una cantidad doble. | V | F |
| 16. El símbolo adoptado por el pictograma debe explicarse por sí mismo. | V | F |
| 17. Un gráfico cronológico sólo puede presentar una serie de datos homogéneos. | V | F |
| 18. Una pirámide de población muestra cómo se distribuye por edad y sexo. | V | F |

19. Dibujar el histograma de los datos agrupados siguientes:

Frecuencia	10	20	30	15	8
Valor	2	4	6	8	10

20. Dibujar el polígono de frecuencias acumuladas del punto anterior.

21. Calcular frecuencias de clases para los siguientes datos:

100 - 330 - 120 - 342 - 176 - 238 - 98 - 163 - 176 - 280 - 115 - 189 - 237 - 230 - 220 - 170 - 83 - 102 - 300 - 350 - 339 - 168 - 111 - 173 - 209 - 253 - 190 - 181 - 261 - 228 - 121 - 177 - 105 - 214 - 304 - 156 - 330 - 302 - 239 - 240 - 188 - 222 - 125 - 218 - 139 - 212 - 123 - 133 - 116 - 148 - 267 - 170 - 234 - 190 - 88 - 115 - 220 - 129 - 153 - 126 - 138 - 124 - 132 - 214 - 225 - 163 - 156 - 192 - 216 - 143 - 149 - 202 - 133 - 250 - 239 - 138 - 117 - 202 - 165 - 148 - 159 - 132 - 128 - 123 - 146 - 147 - 129 - 153 - 126 - 138 - 124
--

22) Obtener las frecuencias acumuladas de los datos anteriores.

23) Con los mismos datos trazar el histograma y polígono de frecuencias acumuladas.

24) Listar los principales datos que debe contener un Informe de Laboratorio Bioquímico.

25) Listar los datos requeridos en una receta, que debe verificar el farmacéutico.

26) Listar los datos que debe completar el farmacéutico en una receta.

27) Listar los requisitos a verificar dentro de una prescripción médica.

28) ¿Cuáles datos se pueden obviar en una separata de un informe de laboratorio?

29) Completar los siguientes conceptos:

- Los tipos de informes estadísticos son:
- Las ventajas de usar el método textual son:
- Las desventajas de usar el método textual son:
- Las ventajas y desventajas de una tabla son:
- La mejor ventaja de un gráfico es su
- Los diagramas comparativos son:
- Cuando se usan símbolos para representar magnitudes se trata de un
- Para mostrar las variaciones de ingresos y egresos de dinero en un negocio se usa el
- Para representar a una población humana por edad y sexo se usa el
- En informe de laboratorio se necesita la firma del Bioquímico responsable.
- En receta presentada en una Farmacia se necesita la firma del médico responsable.
- Para graficar la variación de los ingresos en el tiempo se usa un
- El método es mejor que una tabla para captar de un vistazo los datos.
- Una tabla es mucho más que un gráfico.
- La forma correcta de presentar un gráfico de barras es siempre colocando el en el mismo.

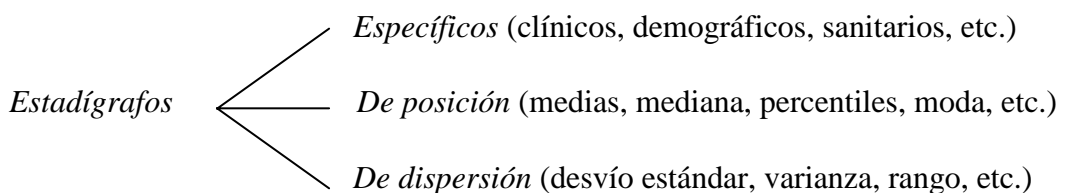
4

Índices clínicos y estadígrafos

En este capítulo se describen diferentes índices clínicos que se usan para juzgar el comportamiento de un método de diagnóstico en Medicina y los estadígrafos clásicos. Un diagnóstico es una predicción acerca del estado del paciente. Al mismo, se puede llegar a través de una serie de análisis tales como: estudio de los síntomas, tests clínicos, radiografías, electrocardiogramas, etc. El resultado final de todo este proceso es binario (sano /enfermo), al igual que la decisión final del médico como: se lo interna al paciente o no, con o sin medicación, cirugía, quimioterapia, etc. Si el resultado es dudoso, esto significa que serán necesarios más estudios y la decisión final solo se pospone para un futuro cercano; pero esta es siempre dicotómica o binaria. En los comienzos, el concepto de calidad se asociaba estrechamente a las ideas de *precisión y exactitud*. La idea clásica era: cuanto más exacto y preciso sea un método, mejor es la calidad del mismo. El problema es que estas ideas solo pueden aplicarse a magnitudes de tipo continuo, donde pueden determinarse los errores de tipo sistemático y casuales asociados a estos conceptos. Por lo que hubo que buscar una nueva manera de entender la calidad en los análisis clínicos y para abarcar a todo tipo de magnitud que se pueda presentar. Lo que conduce al caso dicotómico, pues toda magnitud se puede transformar en una binaria, adoptando un *punto de corte* apropiado en la magnitud que se está analizando. Por ejemplo, la temperatura es una magnitud continua, si se adopta un punto de corte como 37° C para separar un caso normal de un caso de fiebre, entonces se la ha transformado en binaria: con o sin fiebre. La solución a este problema fue propuesta por Galen y Gambino en 1975, en su obra “Beyond Normality”, y desde entonces se ha ido refinando esta forma de encarar la cuestión, tal como se verá a continuación. Básicamente las objeciones son dos: (a) Los análisis clínicos se usan mayormente para detectar enfermedades. Se necesita que ayuden a efectuar un buen diagnóstico. De nada sirve que una técnica, sea muy precisa y exacta si diagnostica mal. (b) Cuando las magnitudes son cualitativas, el criterio de calidad tradicional de precisión y exactitud es inaplicable.

4.1. Clasificación de estadígrafos.

Los estadígrafos se clasifican como se muestra a continuación:



Los estadígrafos de posición y dispersión están estrechamente vinculados con los conceptos de precisión y exactitud. Si se mide muchas veces, en forma repetida, un valor patrón cualquiera de una magnitud como un suero control de concentración conocida o *estándar*, se tiene un conjunto de datos para juzgar el comportamiento de la técnica clínica de medición. Una manera de saber la exactitud de la misma es tomar un valor índice de tendencia central, como el promedio, y compararlo con el valor de referencia. Si el estándar está muy alejado del valor promedio, entonces se puede pensar que la técnica no es muy exacta. Caso contrario, se piensa que la técnica está *calibrada en exactitud*. Para juzgar si la diferencia observada entre ambos valores es por causalidad, o bien se trata de un error de tipo sistemático, no se puede usar un criterio *cualitativo* sino las técnicas estadísticas apropiadas, que se verán más adelante, para poder tener un criterio *cuantitativo*. Esto significa no juzgar las diferencias con criterio humano, subjetivo, sino con un criterio matemático de tipo estadístico: es el adoptado en la ciencia como un *criterio objetivo*. Análogamente, en el caso de la precisión, si la técnica clínica tiene un índice de dispersión muy pequeño respecto de los valores aceptables de referencia en la bibliografía, se puede pensar que es bastante precisa. Nuevamente, sólo un método estadístico permite juzgar imparcialmente si la diferencia entre el índice hallado y el valor de referencia se debe al azar.

Se debe resaltar: *Toda medición científica, de magnitudes continuas, debe ir siempre acompañada con el error de medición, que cuantifica y engloba ambos conceptos*. Un informe presentado sin la aclaración de error de método empleado, no es serio. Cuando se trata de un recuento de unidades enteras, no hay “error” sino equivocaciones. En cambio, en las magnitudes cualitativas, al no ser numéricas, no tienen sentido los conceptos de precisión y exactitud usados tradicionalmente para hablar de *calidad*. Es costumbre en los laboratorios de física, química y relacionados, decir que si un método tiene buena precisión y exactitud, entonces es de buena calidad. Esa costumbre se extendió a los laboratorios de análisis clínicos, donde para optar entre dos técnicas se empleaba ese criterio. Sin embargo, a mediados de la década del cuarenta algunas críticas empezaron a aparecer en la bibliografía. El primer trabajo precursor fue el de Yerushalmi y otros en 1947, junto con los de Galen y Gambino en 1975. Básicamente las objeciones son dos:

1. Los análisis clínicos se usan mayormente para detectar enfermedades. Se necesita que ayuden a efectuar un buen diagnóstico. De nada sirve que una técnica, sea muy precisa y exacta si diagnóstica mal.
2. Cuando las magnitudes son cualitativas, el criterio de calidad tradicional es inaplicable.

La respuesta para ambas críticas, fueron los llamados *índices clínicos*.

4.1.1. Clasificación de enfermedades por diagnosis

Debe tenerse muy en cuenta el tipo de enfermedad que se está analizando y el grado de avance de la misma en el paciente, porque de acuerdo a ello se deberá puntualizar en ciertos aspectos, más que en otros, como se explicará más adelante. Por ahora, basta con efectuar una primera clasificación de las enfermedades en tres tipos básicos de acuerdo al siguiente esquema:

Tipo I : Son todas aquellas enfermedades donde la peor equivocación que se puede cometer en el diagnóstico es un falso negativo.

Tipo II : Son todas aquellas enfermedades donde la peor equivocación que se puede cometer en el diagnóstico es un falso positivo.

Tipo III : Son las restantes, donde no se puede clasificar claramente como una de las anteriores.

Las enfermedades que se pueden clasificar como Tipo I son:

- Una enfermedad que es curable si se detecta a tiempo.
- Una enfermedad que es grave y no puede dejarse pasar inadvertida.
- Los falsos positivos no trauman ni psicológicamente, ni económicamente al paciente, pero los diagnósticos falsos negativos sí lo hacen.

Ejemplos: infarto de miocardio, enfermedades venéreas y otras enfermedades infecciosas curables, la fenilcetonuria, el feocromocitoma (curable en el 100% de los casos si es detectado a tiempo), cáncer de mama o de útero, etc.

En cambio, una enfermedad donde para el paciente, un diagnóstico falso positivo es más peligroso que un falso negativo, se considera como del Tipo II, como ser:

- La enfermedad es grave, pero difícilmente curable o sin remisión.
- Es muy importante para el paciente o para la población él saberse un verdadero negativo.
- Los falsos positivos trauman seriamente al paciente, en cambio los falsos negativos no.
- El tratamiento de los falsos positivos ocasiona serios perjuicios al paciente.

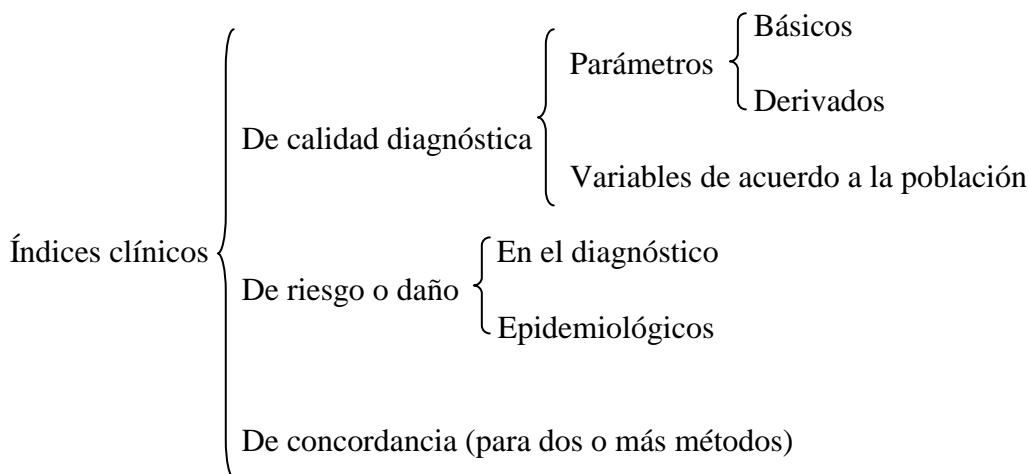
Ejemplos: estado terminal o sin remisión de la enfermedad, cáncer oculto, esclerosis en placas, tratamientos como radioterapia, quimioterapia, lobectomía, cirugía innecesaria, estado IV de la sífilis cuando ya no tiene remedio, etc.

Cuando la enfermedad no puede ser encasillada en ninguno de los dos casos anteriores, entonces se la considera como del Tipo III, como por ejemplo: lupus eritomatoso, ciertas formas de leucemia o linfoma, diabetes, etc. En el caso del SIDA un falso positivo puede ser muy dañino para el paciente, pero un falso negativo sería muy peligroso para la sociedad y como ambos peligros son graves, lo mejor es clasificarla como del Tipo III desde un punto de vista ético. Notar que una misma enfermedad puede ser clasificada de maneras diferentes, de acuerdo al estado de avance que tiene en el paciente. Un cáncer que pueda ser curado si se detecta a tiempo es un claro caso de Tipo I, pero una vez que se produjo una metástasis en el paciente pasa a ser del Tipo II. Lo mismo ocurre con los diferentes estados de la Sífilis, que se puede curar si es detectada a tiempo, pero se torna incurable cuando está muy avanzada.

4.2. Índices clínicos

*El objetivo principal de una técnica clínica es que sirva para diagnosticar bien. Si para efectuar el mismo trabajo hay muchas técnicas disponibles, entonces se debe optar por aquella que mejor sirva a ese propósito, y si además es precisa y exacta, mejor. Al realizar un diagnóstico se está “dicotomizando” a la magnitud clínica medida, no importa del tipo que sea. Se puede postular que siempre habrá dos resultados posibles: *Sano* o *Enfermo*. En principio, se pueden descartar a los casos intermedios, donde no está muy claro el estado real del paciente, para simplificar*

el problema. La práctica clínica ideal sería una que discrimine perfectamente a los sanos de los enfermos, una que no cometa equivocaciones, una utopía. La realidad actual es bastante diferente; para poder diagnosticar bien se necesita, además de la revisión personal, varias determinaciones de laboratorio que confirmen la diagnosis. Aún así, todavía habrá un cierto grado de incertidumbre, mayor cuanto menos avanzada esté la enfermedad. Y entonces, se puede sospechar que la calidad del método, estará relacionada con la distribución en la población del grado de avance de la enfermedad, con la forma de seleccionar las muestras, además de la calidad intrínseca de la técnica clínica. Por otra parte, de acuerdo a los objetivos del estudio clínico los índices clínicos derivados se pueden clasificar con:



Los índices clínicos para un solo método se dividen en dos partes, los relacionados con el método de diagnóstico (la *calidad* del diagnóstico hecho y el *riesgo* que implica hacer un diagnóstico) y los relacionados con el *riesgo o daño* potencial de un factor, que se cree relacionado con la causa de la enfermedad que se está estudiando. Pero cuando se trata de comparar dos o más métodos entre sí, se pueden comparar los índices anteriores si las muestras son independientes, o bien se puede analizar la *concordancia* entre ellos cuando las muestras son apareadas. Los tipos de investigaciones clínicas se detallaron en el capítulo 1, donde se mostraron las tres maneras más sencillas de presentar los datos en tablas de 2 x 2 (Tabla 1.1, 1.2 y 1.3). Ahora, se mostrará como calcular los índices relacionados a cada una de ellas:

4.3. Índices de calidad diagnóstica.

En su forma más simplificada el problema de hacer una predicción o diagnóstico se puede esquematizar con la Tabla de diagnóstico tal como la vista en el primer capítulo Tabla 1.1. Esta forma de organizar los datos se llama tablas doble dicotómicas, donde los cuatro casos posibles de la tabla son: dos verdaderos (vp y vn) y dos falsos (fp y fn) en diagnósticos donde se verifica la presencia de la enfermedad o síntoma buscado (positivo) o la ausencia del mismo (negativo). Los totales marginales de la tabla son:

- TE = $vp + fn$: Total de Enfermos = verdaderos positivos + falsos negativos
- TS = $vn + fp$: Total de Sanos = verdaderos negativos + falsos positivos
- TP = $vp + fp$: Total de positivos = verdaderos positivos + falsos positivos
- TN = $vn + fn$: Total de negativos = verdaderos negativos + falsos negativos

Resultados del test	Resultados verificados		Total
	Enfermos	Sanos	
Positivo (+)	<i>vp</i> <i>verdadero positivo</i>	<i>fp</i> <i>falso positivo</i>	TP
Negativo (-)	<i>fn</i> <i>falso negativo</i>	<i>vn</i> <i>verdadero negativo</i>	TN
Total	TE	TS	N

La técnica clínica ideal sería aquella que no cometa equivocaciones en sus pronósticos, o sea una con $fn = fp = 0$. Las consecuencias de las equivocaciones no son igualmente graves para el paciente, todo depende del tipo de enfermedad de que se trate. Y por ese motivo se realizó la clasificación de las enfermedades en el punto anterior, teniendo en cuenta el tipo de equivocación cometida en el diagnóstico. Una manera de analizar la *calidad* de un método de diagnóstico es tener en cuenta siempre los dos tipos de equivocaciones, o sea se trata de un problema de tipo dual. Y cuantificar su calidad de acuerdo a la capacidad intrínseca del mismo en equivocarse.

4.3.1 Sensibilidad, Especificidad e Índice de Youden.

El comportamiento de un método de diagnóstico al pronosticar, puede ser analizado en forma dual de la manera siguiente:

- *Sensibilidad*: Es la capacidad del método para diagnosticar como positivo a un enfermo.
- *Especificidad*: Es la capacidad del método para diagnosticar como negativo a un sano.

Por ejemplo, si el método es capaz de detectar a tres casos de positivos cada cuatro enfermos que analiza, entonces su capacidad para identificar correctamente a los que pronostica como positivo es del 75%. Aquí se dice que su sensibilidad $S = 0,75$. Por otro lado, si el método es capaz de detectar a ocho casos negativos, cada diez individuos sanos, entonces su capacidad para identificar correctamente como negativo es del 80%. Se dice aquí que su especificidad $E = 0,8$. Aplicando estos conceptos al caso general presentado en la Tabla de diagnóstico, se pueden deducir las fórmulas que siguen:

Sensibilidad: $S = vp / TE = (\text{verdaderos positivos}) / (\text{total de enfermos})$

Especificidad: $E = vn / TS = (\text{verdaderos negativos}) / (\text{total de sanos})$

Ambos valores están acotados entre cero y uno. Cuando la sensibilidad es máxima $S = 1$, quiere decir que no hay falsos negativos, ya que $TE = vp + fn$, y esa sería la cualidad buscada para una enfermedad del Tipo I. En cambio, para una enfermedad del Tipo II, lo deseable es que $fp = 0$ y para ello deber ser máxima la especificidad: $E = 1$, pues $TS = vn + fp$.

La calidad puede ser expresada con los criterios de sensibilidad y especificidad, por lo tanto una prueba clínica es ideal cuando tiene una capacidad para diagnosticar $S = E = 1$, o sea cuando no comete equivocaciones en sus pronósticos.

A las técnicas clínicas de laboratorio, o bien a los métodos de diagnóstico que se parecen mucho a este ideal se los denomina: métodos de referencia (o “golden standard”). La realidad es bastante diferente a esta utopía, son muy raros los casos donde ambos índices se acercan al máximo a la vez, aún en una biopsia se pueden cometer equivocaciones. Entonces el problema se reduce a elegir aquella prueba que tenga máxima sensibilidad cuando hay disponibles varias técnicas para las enfermedades del Tipo I, y máxima especificidad para las del Tipo II. Para el caso de las enfermedades del Tipo III, se necesita otro índice de calidad que mezcle ambos conceptos.

Este índice clínico fue propuesto por Youden para analizar la capacidad del método de diagnóstico, usando un único valor en reemplazo de la forma dual de hacerlo. La idea es mezclar los dos índices anteriores para hacer el estudio de calidad. Se define como:

Índice de Youden: $Y = S + E - 1 = (\text{sensibilidad}) - (\text{especificidad}) - 1$

En el caso ideal, este índice vale 1 ya que $S = E = 1$. En el caso de lanzar una moneda al aire para efectuar el diagnóstico, es decir puro azar, será $S = E = 0,5$ y por lo tanto este índice es nulo. La condición evidente para tener una calidad aceptable es que no sea negativo. La idea para seleccionar un método de diagnóstico en las enfermedades del Tipo III, es elegir aquel que tenga el máximo valor de este índice, pues eso equivale a tener un mínimo de resultados falsos, prorrateados entre el total de los casos analizados. Estos tres índices dan un valor que cuantifica los conceptos básicos de calidad diagnóstica, de acuerdo al tipo de enfermedad analizada. Sin embargo, puede verse que el último es una consecuencia de los dos primeros, a los que se considera los índices básicos o *parámetros* de calidad, de los cuales se derivan todos los demás.

4.3.2 Eficiencia diagnóstica (“diagnostic accuracy”)

La otra manera de ver la misma idea es usando otro índice, sugerido por muchos autores como la manera de evaluar el “poder discriminatorio” del método clínico. Una especie de índice general de la calidad, porque lo que se analiza aquí es el porcentaje de éxitos logrados por el método en sus pronósticos. Es como el clásico concepto del rendimiento, ver que tan bien trabaja el método de diagnóstico o test clínico. Por ejemplo, si se imagina al método como una “máquina de hacer predicciones”, el rendimiento del mismo se puede expresar con:

Eficiencia: $A = (vp + vn) / N = (\text{verdaderos positivos} + \text{verdaderos negativos}) / (\text{total de casos})$

La idea de eficiencia es simplemente la del porcentaje de éxitos obtenidos al diagnosticar, se puede descomponer en dos tipos de éxitos: el término (vp / N) es el porcentaje de éxitos en positivos y el término (vn / N) es el porcentaje de éxitos en negativos. Cuanto mayor sea la eficiencia de una prueba clínica, mejor será su poder discriminatorio. La prueba clínica ideal tendrá una eficiencia máxima (100%) $A = 1$, mientras que cuando $A = 0$ significa que no tuvo ningún acierto de sus pronósticos. Notar que si se lanza una moneda al aire para hacer pronósticos, la eficiencia sería del 50%, o sea $A = 0,5$. También se puede usar para seleccionar el mejor método en el caso de las enfermedades del Tipo III. De hecho, es el recomendado en la bibliografía. Sin embargo, hay una diferencia conceptual importante entre estos dos índices, y es que el índice de Youden no depende de la población en donde se aplica el método, sino que es una característica intrínseca propia del método, en cambio la eficiencia depende del lugar donde se está usando el

método, o bien del momento en que se lo aplica. Esto es, la eficiencia depende de la *prevalencia* de la población mientras que el índice de Youden no, como se puede ver a continuación:

Prevalencia: $p = TE / N = (\text{total de enfermos}) / (\text{total de individuos de la población estudiada})$

Teniendo en cuenta que la eficiencia se puede expresar como:

$$A = (vp / N) + (vn / N) = [(S \cdot TE) / N] + [(E \cdot TS) / N]$$

Reemplazando los valores verdaderos con los índices sensibilidad y especificidad respectivamente, y luego haciendo las simplificaciones necesarias resulta:

$$A = S \cdot p + E (1 - p) = (\text{sensibilidad} \times \text{prevalencia}) + [\text{especificidad} \times (1 - \text{prevalencia})]$$

Por ejemplo, en una población donde la enfermedad es muy rara o casi inexistente, la prevalencia sería prácticamente nula ($p = 0$) y entonces la eficiencia resulta igual a la especificidad. En cambio en un caso extremo de epidemia, donde casi toda la población está enferma, la prevalencia es prácticamente del 100% ($p = 1$) y entonces la eficiencia resulta igual a la sensibilidad. Notar que, con el transcurso del tiempo una enfermedad puede propagarse y empezar a aumentar la prevalencia de la misma, lo que originaría un cambio en la eficiencia. Pero esto no ocurre con el índice de Youden, porque este no depende de la prevalencia. En forma matemática esto sería $A = f(p)$ mientras que $Y \neq f(p)$. Otro punto importante a tener en cuenta es que, en una misma población geográfica, se pueden tener distintos valores de prevalencia. Por ejemplo, el número de pacientes con problemas cardíacos que concurren a hacerse atender en un centro asistencial de la zona rural de Posadas, es mucho menor que los que concurren al hospital, los que a su vez son menores que los concurrentes a la unidad coronaria del mismo, derivados de otros centros asistenciales. Por lo tanto, el mismo método de diagnóstico aplicado en estos tres puntos tendrá un valor diferente de eficiencia (A) a pesar de que todos están en una misma zona geográfica. Debe remarcar aquí, que la población de referencia para una investigación clínica, no es la de la zona geográfica circundante, sino de aquellos pacientes que concurren a hacerse atender en el centro asistencial.

Para explicar mejor esta diferencia entre parámetros y variables de la calidad del método de diagnóstico, se puede usar un símil de la Física. Calor y frío son dos conceptos opuestos que aparecen como una dualidad. Sin embargo se trata de la misma magnitud física: la temperatura, o sea de la misma cosa. En un extremo de la escala cuando la temperatura es baja se asocia con uno de los extremos, una sensación: el frío. En el otro extremo de la escala, cuando la temperatura es alta se la asocia con la sensación opuesta: el calor. Aparecen como cosas contrarias, cuando en el fondo se trata de lo mismo. Análogamente en Medicina, la *eficiencia* de una prueba diagnóstica es la magnitud clínica variable, en un extremo de la escala cuando la prevalencia es muy baja se la asocia con el concepto de *especificidad*. Mientras que en el otro extremo de la escala, cuando la prevalencia es muy alta se la asocia con el concepto de *sensibilidad*. Es decir, que sensibilidad y especificidad no son otra cosa que la eficiencia de una prueba clínica, pero expresada en sus dos casos extremos.

Para entender estas ideas se puede usar la noción de *relatividad* tal como se estudia en Física. Por ejemplo: la diferencia conceptual entre masa y peso, es que la primera es una cualidad *intrínseca* del cuerpo que no depende del sistema de referencia en el cual se halla. En cambio el peso varía con la influencia de la atracción de la gravedad del planeta en donde se encuentre el

cuerpo. Se dice por ese motivo que el peso es una cualidad *extrínseca* del cuerpo. Expresado en otros términos, a la *magnitud* masa le corresponde una variable llamada peso. Este es un *parámetro* en el planeta Tierra porque todo individuo está afectado por la misma fuerza de gravedad, pero es una *variable* dentro del sistema solar.

Análogamente, si se supone que la capacidad de un método clínico es constante en toda la población humana, entonces la sensibilidad y especificidad se mantendrán fijas para todos los individuos, por lo que pueden ser considerados como cualidades *intrínsecas* del método. Y lo mismo ocurre con el Índice de Youden. En cambio la eficiencia varía (cualidad extrínseca) según el sistema de referencia o población en donde se usa el método, de acuerdo a la *prevalencia* de la enfermedad existente en la población. Es decir, sensibilidad y especificidad son los principales parámetros de un método clínico mientras que la eficiencia es una variable.

Si se considera a un método de diagnóstico, o bien a un test clínico, como si fuese una máquina para hacer predicciones, entonces se puede explicar este concepto con otro símil: Un Ferrari es un excelente auto para correr en un premio de Fórmula I, pero a pesar de su alta calidad puede que no sea la mejor opción para correr en Indianápolis, como también puede llegar a ser un desastre para competir en las arenas del desierto en el Rally de Dakar.

4.4 Relatividad de los tests clínicos.

Todo esto muestra que la eficiencia debe ser informada al lector con una curva, en lugar de usar un solo valor, como se acostumbra hacer en la práctica actual. Para obtener estas curvas fácilmente se puede usar un algoritmo desarrollado para computadoras en Excel, disponible en la página web de la cátedra. Pero también se lo puede obtener con un procedimiento que se muestra a continuación:

Ejemplo: Sea un test clínico que tiene una calidad cuantificada por $S = 0,9$ y $E = 0,75$. Se desea averiguar el efecto de la relatividad en la eficiencia.

Como S y E son los parámetros del método, se puede obtener el otro parámetro con $IY = 0,65$. Y para ver como afecta la prevalencia de la enfermedad a la eficiencia, se pueden simular diferentes valores de la misma tomando varios valores de p , como ser $p = 0,1; 0,2; 0,3; \dots; 0,9$ y un tamaño muestral cualquiera como por ejemplo $N = 1000$.

Paso 1: Para la primera simulación se adopta el valor $p = 0,1$, es decir 1 enfermo cada 10 casos, y se calcula el número de verdaderos positivos y negativos:

Como $S = vp / TE$ resulta $vp = S \cdot TE = S \cdot N \cdot p = 0,9 \cdot 1000 \cdot 0,1 = 90$

Como $E = vn / TN$ resulta $vn = E \cdot TN = E (N - TE) = E (1 - p) N = 0,75 (1 - 0,1) 1000 = 675$

Paso 2: Se calculan ahora los valores de falsos negativos y positivos con:

$TE = N \cdot p = 100 = vp + fn$; o sea $fn = 100 - vp = 100 - 90 = 10$

$TS = N (1 - p) = 900 = vn + fp$; o sea $fp = 900 - vn = 900 - 675 = 225$

Paso 3: Con los cuatro valores obtenidos en las dos etapas anteriores se puede armar la Tabla de diagnóstico para el caso de una prevalencia del 10% y así calcular el total de positivos y negativos como se muestra a continuación:

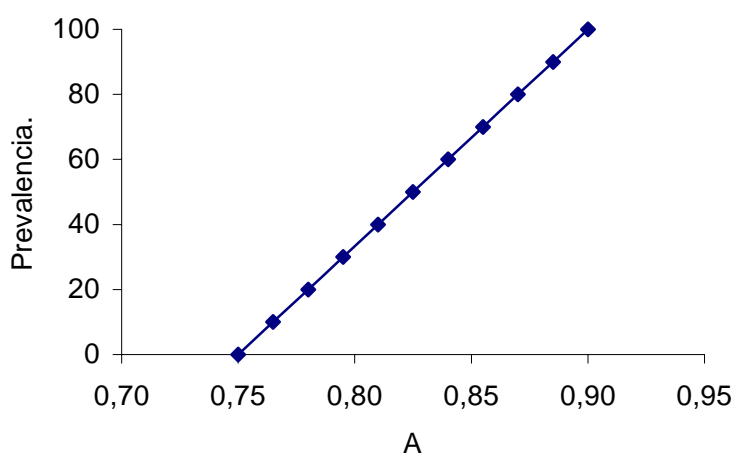
Test	Resultado verificado		Total
	Enfermo	Sano	
(+)	90	225	TP = 315
(-)	10	675	TN = 685
Total	TE = 100	TS = 900	1000

Paso 4: Se calcula la eficiencia para estos valores de la Tabla completa:

$$A = (vp + vn) / N = (90 + 675) / 1000 = 0,765$$

Y se tiene estimado el primer valor de la eficiencia para una prevalencia del 10%. Análogamente, se toman diferentes valores de la prevalencia y se repiten los 4 pasos anteriores para obtener los respectivos valores de la eficiencia, resultando la curva siguiente:

Gráfico 4.1: Variabilidad de la eficiencia con la prevalencia (S = 0,9 y E = 0,75).



En el Gráfico 4.1 se muestra un método de diagnóstico con los parámetros S = 0,9 y E = 0,75. En una población donde la enfermedad es muy rara, la prevalencia será muy baja tal como es el caso de la meningitis, y la eficiencia será muy parecida al valor de la especificidad (si $p \rightarrow 0$ entonces $A \rightarrow E$). En cambio, para otra población donde la enfermedad es muy frecuente, como puede ser una epidemia, la eficiencia del mismo método clínico cambia bastante y se acerca a la sensibilidad (si $p \rightarrow 1$ entonces $A \rightarrow S$). Este ejemplo muestra que si se mide la capacidad de un método de diagnóstico, con el índice de la eficiencia (A), este valor será diferente en poblaciones diferentes. Y por lo tanto, no es aconsejable seguir ese criterio para informar la calidad o poder discriminatorio del método como se recomienda usualmente en alguna bibliografía. O por lo menos, no usar un único valor para presentar el dato, sino que se debe mostrar la curva para que el lector sepa como trabajará ese método en su propia población de referencia. Notar que esta curva no depende del tamaño de la muestra (N) usada para determinarla.

4.5 Valores Predictivos.

Otra manera de analizar la capacidad de un método es analizando el porcentaje de éxitos logrados sobre la base de sus pronósticos. La idea es estudiar que tan bien predice el método. Otra vez se trata de una dualidad, porque hay que ver los dos aspectos principales: como trabaja en los casos positivos y como lo hace en los casos negativos. Se puede definir entonces:

- *Valor Predictivo de Positivos:* Es la capacidad del método para identificar correctamente a un caso positivo, de entre todos los resultados positivos obtenidos.
- *Valor Predictivo de Negativos:* Es la capacidad del método para identificar correctamente a un caso negativo, de entre todos los resultados negativos obtenidos.

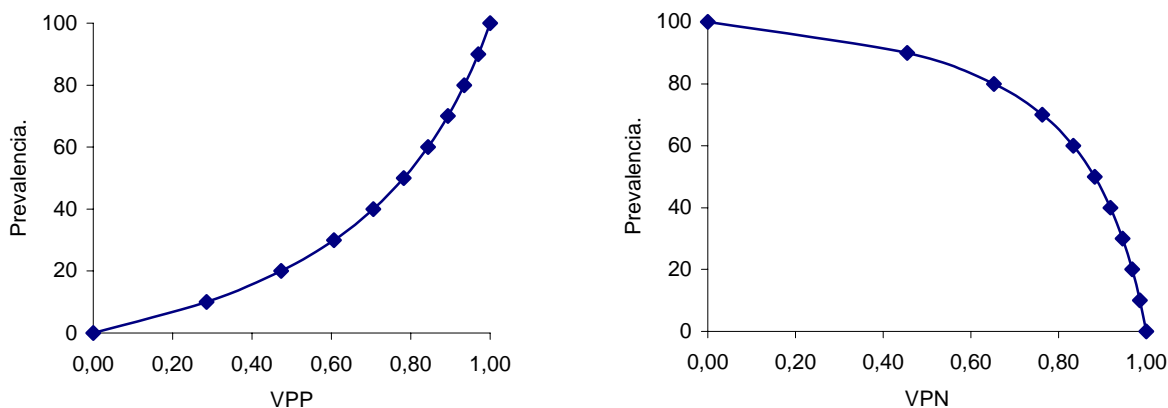
Por ejemplo, si el método es capaz de detectar a cuatro casos verdaderos de positivos cada cinco resultados positivos encontrados, entonces su capacidad para identificar correctamente a los que pronostica como positivo es del 80%. Aquí se dice que su Valor Predictivo de Positivos es $VPP = 0,8$. Por otro lado, si el método es capaz de detectar a nueve casos verdaderos de negativos, cada diez casos que resultaron negativos, entonces su capacidad para identificar correctamente como negativo es del 90%. Se dice aquí que su Valor Predictivo de Negativos resulta: $VPN = 0,9$. Aplicando estos conceptos al caso general presentado en la Tabla de diagnóstico, se pueden deducir las fórmulas que siguen:

Valor Predictivo de Positivos: $VPP = vp / TP = (\text{verdaderos positivos}) / (\text{total de positivos})$

Valor Predictivo de Negativos: $VPN = vn / TN = (\text{verdaderos negativos}) / (\text{total de negativos})$

La relación matemática entre estos índices relativos, con los dos índices básicos S y E se expresa con el Teorema de Bayes, que se explicará más adelante. Por ahora, se puede visualizar esta idea con el gráfico siguiente, obtenido de la misma manera que la explicada para la eficiencia, pero ahora se calculan además los valores predictivos:

Gráfico 4.2: Variabilidad de los valores predictivos con la prevalencia ($S = 0,9$ y $E = 0,75$)



En este ejemplo se analiza la variabilidad de los valores predictivos con la prevalencia, para un método clínico de parámetros $S = 0,9$ y $E = 0,75$. Por ejemplo, para una prevalencia del 10% resultan: $VPP = vp / TP = 90 / 315 = 0,286$ y $VPN = vn / TN = 675 / 685 = 0,985$

Mientras los VPP varían desde cero a uno, a medida que aumenta la prevalencia, los VPN lo hacen de forma inversa. Si la prevalencia aumenta, aumentan los valores predictivos de positivos y bajan los negativos. Cuando el $VPP = 1$, significa que $vp = TP$, es decir no hay falsos positivos, la condición ideal para una enfermedad del Tipo II. En cambio, si $VPP = 0$, esto implica que no hay aciertos en positivos y todos los pronósticos de esa índole tienden a ser falsos. Notar que esto ocurre cuando la enfermedad es rara y su prevalencia muy baja. Por otra parte, cuando $VPN = 1$, significa que $vn = TN$, o sea que no hay falsos negativos, la condición buscada para las enfermedades del Tipo I. Viendo la gráfica anterior, se puede observar que cuando la prevalencia de la enfermedad sea muy alta (epidemias, unidades coronarias, etc.), el $VPN = 0$ y que casi no van a existir aciertos en negativos. En cambio, si la enfermedad es muy rara, casi todos los resultados negativos serán falsos.

Ejemplo 1: Se ha realizado un estudio para ver la capacidad de un test clínico que se usa habitualmente en el laboratorio. Las historias clínicas de los pacientes se separan en dos grupos. En el primero se agrupan los casos verificados de enfermedad “a posteriori”, es decir se hace un tipo de estudio retrospectivo una vez que se está seguro del estado real del paciente. Así se obtiene el número de casos identificados correctamente como positivos (vp) y las falsas predicciones (fp). En el segundo conjunto se agrupan los casos verificados de ausencia de la enfermedad, y con el estudio retrospectivo se obtienen los números de casos verdaderos negativos y falsos negativos (vn y fn). Los casos dudosos se descartan para eliminar variabilidad indeseada. Se seleccionaron al azar 400 casos de enfermedad y otros tantos de no-enfermedad. Los resultados obtenidos se muestran en la tabla. Se desean conocer los índices clínicos correspondientes a este test.

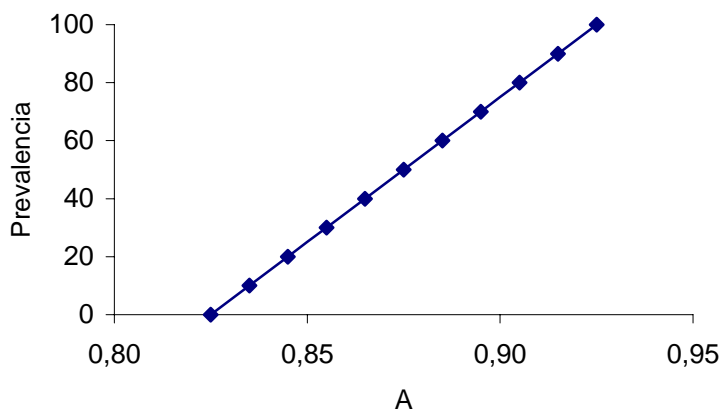
Test	Resultado verificado		Total
	Enfermo	Sano	
(+)	370	70	440
(-)	30	330	360
Total	400	400	800

Lo primero es calcular los parámetros principales del método:

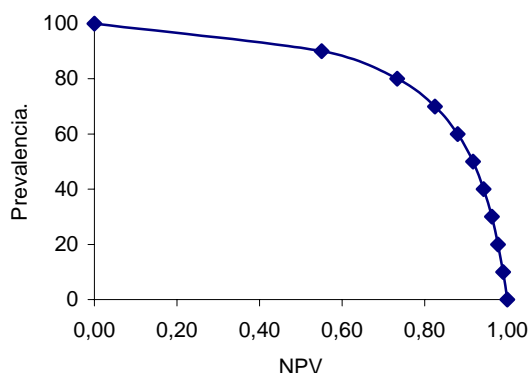
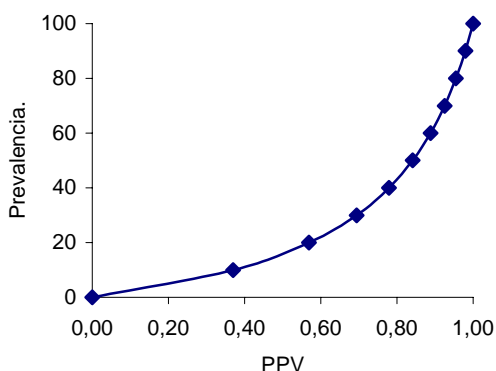
$$S = 370 / 400 = 0,925 \quad \text{y} \quad E = 330 / 400 = 0,825$$

$$\text{Con estos valores se obtiene } Y = 0,925 + 0,825 - 1 = 0,75$$

Y para informar la eficiencia y los valores predictivos se muestran las curvas calculadas con el método que se vio en el ejemplo anterior. La eficiencia (A) varía como se grafica a continuación:



Los valores predictivos varían con la prevalencia como se ve más abajo:



Ejemplo 2: De la sumatoria total de casos encontrados por los investigadores del ejemplo anterior se determina que la prevalencia de la enfermedad para ese laboratorio es $p = 0,4$ (solo se usan los casos totales verificados). Se pide calcular como son los valores relativos para ese laboratorio en particular, de acuerdo a su población de referencia.

De las curvas anteriores se obtiene: $A = 0,865$; $VPP = 0,779$ y $VPN = 0,943$. Esto significa que el porcentaje de éxitos de ese test, en esa población, será del 86,5%. Además predecirá correctamente a los positivos en un 77,9% de los casos, y a los negativos en el 94,3% de los casos.

4.6 Likelihood Ratios.

Habitualmente, estos índices suelen informarse en los trabajos de investigación junto a los parámetros básicos. Se trata de un par de valores que no varían con la prevalencia de la población, sino que son características propias del método clínico. Su uso se relaciona con las probabilidades “a posteriori” del test como se explicará más adelante. Por ahora, basta saber como se los define:

- *Likelihood Ratio de Positivos:* $LR+ = S / (1 - E)$

- *Likelihood Ratio de Negativos:* $LR- = (1 - S) / E$

Por ejemplo, para el caso visto más arriba donde $S = 0,925$ y $E = 0,825$ se pueden obtener:

$$LR+ = 5,286 \text{ y } LR- = 0,091$$

A mediados de la década del ochenta, se los comienza a emplear para facilitar el cálculo de la compleja fórmula de Bayes en los trabajos de Medicina. Un ejemplo desarrollado por Simel y publicado por The Lancet en 1985 muestra como determinar si una prueba es o no necesaria para mejorar el diagnóstico. Los LR no son una probabilidad, ni dependen de la prevalencia, es decir *son magnitudes o valores intrínsecos* de la prueba clínica. Sin embargo, se calculan como un cociente de probabilidades. El LR+ es el cociente entre la probabilidad de predecir correctamente la enfermedad y la probabilidad de predecir incorrectamente a dicha enfermedad. Da una indicación de cuanto puede aumentar o bajar la probabilidad del resultado del método de diagnóstico, antes de efectuarlo (probabilidad pre-test). Cuando el LR+ = 1 indica que si se le realiza el test al paciente, no se obtendrá información adicional y por lo tanto lo mejor es no someter al paciente a gastos inútiles. En cambio cuando LR+ > 1 indica que es conveniente efectuar el test porque se obtendrá información adicional. Pero si LR+ < 1 entonces el test no incrementa la probabilidad de asesorar bien al clínico, sino que disminuye. Para profundizar estos conceptos hay que tener una noción de probabilidades, cosa que se hará más adelante.

4.7 Influencia de los puntos de corte.

Hasta ahora se vio el caso de resultados dicotómicos del método de diagnóstico que es el caso más general. Sin embargo, muchas mediciones clínicas se hacen con otro tipo de magnitudes como las continuas. La solución para estos casos es transformar una magnitud cualquiera en una binaria adoptando un *punto de corte* adecuado. Por ejemplo, el peso es una magnitud continua y si se toma el valor 70 kg para diferenciar a gordos de flacos, entonces cuando un individuo pesa hasta 70 kg se lo considera flaco y en caso contrario gordo. Este valor de 70 kg es el punto de corte adoptado para separar dos zonas bien definidas en la variable continua. El problema es que no existe en los casos reales, un valor que permita discriminar claramente a los sanos de los enfermos en magnitudes continuas. Se encuentran tres zonas: una para los sanos, una intermedia donde los casos son dudosos y la tercera donde los resultados indican claramente la presencia de la enfermedad. Para ilustrar estos conceptos se analizará el caso de infartos de miocardio del capítulo anterior, en los Cuadros 3.1 y su resumen en el Cuadro 3.3. Pero antes de proseguir hay que hacer notar una diferencia conceptual importante:

Sensibilidad, especificidad y sus índices derivados no varían con la prevalencia, pero sí lo hacen con el punto de corte que se adopta en magnitudes continuas. Los otros índices variables como los valores predictivos y la eficiencia, pueden variar por ambas cosas.

Por ejemplo, tomando el caso de los pacientes con supuesto infarto de miocardio visto en el Cuadro 3.3, y definiendo como punto de corte un valor de CPK de 20 UI/l, entonces cuando un paciente tenga un valor menor o igual de CPK se lo supondrá sano (-), o sea una falsa alarma. En cambio, si el valor medido es mayor se lo internará de inmediato en la unidad coronaria, pues se lo supondrá infartado, es decir como enfermo (+). Así, de los infinitos valores posibles que puede arrojar la medición de una variable continua, al adoptar un punto de corte, este delimitará dos

zonas, “dicotomizando” la variable en cuestión. Analizando los histogramas del Gráfico 3.12, se puede ver que muchos puntos de corte arrojarán falsos diagnósticos. Puede verse esta superposición entre 20 y 120 UI/l, creando una zona de incertidumbre muy amplia. Adoptando como punto de corte 20 no habrá ningún caso (-) que sea falso. Viceversa, si se adopta 120 como punto de corte entonces no habrá ningún falso (+). En cambio, una técnica clínica ideal, separaría ambas zonas sin problemas y así no se cometerían equivocaciones. Pero, en la actualidad es muy raro el disponer de una prueba que de certidumbre total. Aún una biopsia puede fallar. El desafío es encontrar una prueba lo más cercana posible a la ideal, llamada el “golden standard”. Los investigadores están probando combinaciones de dos o más pruebas usando diferentes modelos en serie o paralelo, el Análisis Multivariado, etc.

El problema es encontrar un criterio, basado en la ética, que le permita al clínico elegir el mejor punto de corte posible. Luego de adoptarlo, la variable continua se transforma en una binaria y se le puede aplicar todo lo visto en los puntos anteriores. Para definir un criterio conviene analizar *la variabilidad de los índices de acuerdo al punto de corte* adoptado en la magnitud clínica. En el Gráfico 3.12 se observa que, tomando a 120 UI/l como punto de corte, no habrá falsos positivos. En cambio, tomando a 20 UI/l como valor referente, entonces $fn = 0$. Adoptando diferentes puntos de corte se puede deducir que si se aumentan los fp, bajan los fn y viceversa. La pregunta más difícil de contestar es: *¿Cuál será el valor óptimo a elegir para poder discriminar entre sanos y enfermos?* Hay varias respuestas. Si se trata de una enfermedad mortal, en caso de no ser atendida a tiempo, salta a la vista que lo más importante es tener $fn = 0$, como los infartos, ciertos tipos de tumores, etc. En cambio, si se trata de una enfermedad terminal, sin cura posible, puede resultar muy grave informarle al paciente que la padece cuando en realidad está sano. Aquí se requiere de $fp = 0$, como el caso un cáncer terminal, metástasis de ciertos tumores, etc. Por otro lado, hay muchos casos donde esta decisión no es tan sencilla como por ejemplo en el caso del SIDA: desde el punto de vista del paciente lo peor es un fp por los temores y angustias que le producirá creerse enfermo cuando en realidad está sano, pero desde el punto de vista de la sociedad lo peor es un fn porque al no tomar las prevenciones del caso la enfermedad se propagará en más personas. Como antes, para *decidir* no hay que perder de vista el tipo de enfermedad que se está analizando.

En el caso visto de mediciones de CPK se tienen los archivos de pacientes que llegan a una unidad coronaria con un diagnóstico presunto de infarto, el cual fue confirmado o rechazado a posteriori. Es decir, de todas las historias clínicas de esa unidad se descartan los casos dudosos y solo se usan aquellos confirmados en dos clases: los que tuvieron un infarto y los que no lo tuvieron pues fue una falsa alarma. De entre todos los casos disponibles se eligen al azar una cierta cantidad N lo más grande posible (por lo menos 50 casos), clasificadas un 50% para los de enfermedad confirmada y el otro 50% para los sanos. En el ejemplo visto se tomaron 400 pacientes, 200 con infarto confirmado y 200 sin infarto. Para poder estudiar la influencia del punto de corte, que se adopte para la medición de la CPK se emplearán los datos del Cuadro 3.1 y las frecuencias del Cuadro 3.3. Como se puede clasificar a los infartos como una enfermedad del Tipo I, se deduce que el principal criterio a tener en cuenta va a ser la sensibilidad. La estrategia será adoptar un punto de corte que maximice dicho índice, para evitar la mayor cantidad de falsos negativos posibles. Sin embargo, si se tratase de otra enfermedad habría que maximizar la especificidad para las enfermedades del Tipo II, o el índice de Youden para las del Tipo III. Para este ejemplo, en la Tabla 4.1 siguiente, se toman a los límites superiores de cada intervalo de clase como puntos de corte y así se pueden efectuar los cálculos como se muestra a continuación:

Tabla 4.1: Obtención de los diagnósticos posibles con datos del Cuadro 3.1 y 3.3

Clases	Sin Infarto		Con Infarto		Puntos de Corte	TN	TP	vn	fn	vp	fp
	Fr.	Fr. ac.	Fr.	Fr. ac.							
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
0-20	104	104	0	0	20	104	296	104	0	200	96
21-40	36	140	8	8	40	148	252	140	8	192	60
41-60	28	168	16	24	60	192	208	168	24	176	32
61-80	18	186	32	56	80	242	158	186	56	144	14
81-100	10	196	46	102	100	298	102	196	102	98	4
101-120	4	200	66	168	120	368	32	200	168	32	0
121-140	0	200	23	191	140	391	9	200	191	9	0
141-160	0	200	6	197	160	397	3	200	197	3	0
161-180	0	200	3	200	180	400	0	200	200	0	0

El procedimiento para obtener los valores de la Tabla 4.1 es como sigue:

Paso 1 : En la primer columna (1) se colocan las clases adoptadas con un ancho de 20 UI/l.

Paso 2 : En las columnas (2) y (4) se ponen las frecuencias obtenidas del Cuadro 3.3.

Paso 3 : Se calculan las frecuencias acumuladas para ambos casos y se colocan en (3) y (5).

Paso 4 : En la columna (6) se colocan los diferentes puntos de corte elegidos.

Paso 5 : En la columna (7) se calcula el TN como la suma de todos los datos menores que el punto de corte adoptado. O sea, la suma de las frecuencias acumuladas correspondientes. En este caso resulta la suma de las columnas (3) y (5).

Paso 6 : En la columna (8) se obtiene el TP como la diferencia entre el total de datos y TN de la columna (7). O sea $TP = N - TN$.

Paso 7 : En la columna (9) se calculan los vn como el número de pacientes cuyo diagnóstico resulta (-) por el punto de corte adoptado, y que realmente son no infartados. Entonces, será el valor de la frecuencia acumulada de los no infartados. Así resulta: columna (3) = columna (9).

Paso 8 : En la columna (10) se obtiene el valor de los fn con los datos de la columna (5). Pues todos los valores menores que el punto de corte corresponden a la frecuencia acumulada de los infartados, estos serán diagnosticados como (-) cuando en realidad tienen un infarto.

Paso 9 : En la columna (11) se usa la relación $vp = TE - fn$. O sea, $tp = 200 - fn$. Esto es: columna (11) = 200 - columna (10).

Paso 10 : En la columna (12) se usa la relación $fp = TS - vn = 200 - tn$. Otra forma de obtener estos valores es usando: $fp = TP - vp$. Así, columna (12) = 200 - columna (9).

Se puede notar que, para cada punto de corte adoptado, se puede formar una Tabla de diagnóstico. Analizando los resultados de la Tabla 4.1, puede verse que los fn son nulos sólo cuando el punto de corte se adopta en 20 UI/l. Esto es, a todo paciente que ingrese al servicio con

una CPK mayor que 20 UI/l se lo internará porque se sospecha que está infartado. Naturalmente, esto arrojará el número más alto de $fp = 96$ pacientes internados sin que sea necesario. A medida que se aumenta el valor del punto de corte, la situación se invierte. Así van aumentando los fn mientras disminuyen los fp , hasta el punto 120 UI/l; allí resulta $fp = 0$. O sea, no habrá pacientes internados sin necesidad, pero la contrapartida, son los $fn = 32$ pacientes diagnosticados como sanos, o peor aún, dados de alta, a pesar de su infarto. A la luz de esta información, se puede contestar la pregunta planteada anteriormente: ¿Cuál es el mejor punto de corte a ser adoptado desde el punto de vista ético ?

Para el caso de infartos de miocardio (Tipo I) el valor óptimo es aquel que reduce al mínimo la posibilidad de un fn . (para este ejemplo se debe tomar como punto de corte 20 UI/l).

Claro que si al responsable del servicio, no le interesa minimizar el número posible de muertos por falta de atención, sino reducir sus costos evitando las falsas alarmas, entonces para una persona así el valor óptimo sería de 120 UI/l, y sólo internaría a pacientes cuyas CPK sean superiores a tal valor. Pero, si al responsable lo que más le interesa es no cometer errores, por una cuestión de imagen, mejorará su rendimiento adoptando como punto de corte 60 UI/l, pues allí la suma de fp y fn es mínima. La adopción de un punto de corte está influenciada por la subjetividad de las personas y por los condicionantes externos producto de las presiones laborales y económicas de la cuestión. *El deber es siempre usar la ética como criterio básico de decisión.*

De todos los índices propuestos hay dos básicos de los cuales derivan todos los demás, que son la Sensibilidad y la Especificidad. *Conviene tener la máxima Sensibilidad en aquellas enfermedades del Tipo I.* Es decir, en enfermedades de consecuencias graves si no se las detecta a tiempo. Además del infarto de miocardio, el caso más común son las enfermedades venéreas. La sífilis en su fase IV ya no tiene remisión, pero en una etapa más temprana igual puede tener secuelas nefastas, como la sífilis congénita que transmite la embarazada, la neurosífilis y la sífilis cardiovascular. En las fases tempranas hay mucha dificultad para detectarla cuando aún no se observan reacciones epidérmicas y es fundamental usar una técnica con el mínimo fn posible. Por lo tanto, si se está usando la VDRL para detectarla y aparece una nueva técnica, el profesional debe optar por aquella con máxima sensibilidad entre ambas. Hay otros casos como el feocromocitoma, fatal si no se lo detecta a tiempo, pero curable con tratamiento; el hipotiroidismo, la toxoplasmosis en especial entre las embarazadas, el cáncer de mama o de útero, etc.

Conviene tener máxima especificidad en las enfermedades del Tipo II. Este caso se busca cuando es muy grave informarle al paciente que está enfermo, si en realidad está sano. Esto le podría ocasionar serios daños en lo psicológico, moral o económico. Por ejemplo, el caso de un cáncer incurable, la esclerosis múltiple, etc. Cuando una enfermedad es muy grave y entró en una etapa de no-remisión, se requiere mucha especificidad para minimizar los falsos positivos.

Cuando la magnitud clínica es continua, la especificidad varía con el punto de corte adoptado por el clínico. En la Tabla 4.2 se ilustra esta situación para el caso de los pacientes con diagnóstico presunto de infarto de miocardio. La especificidad es mínima cuando el corte adoptado es 20 UI/l y máxima luego de los 120 UI/l. Análogamente para el índice de Youden, este será máximo cuando se igualen sensibilidad y especificidad, como es en el punto de corte 60 UI/l. Los demás índices pueden obtenerse a partir de los anteriores, y de acuerdo al punto de corte adoptado, como se puede apreciar en la tabla siguiente:

Tabla 4.2: Obtención de los Índices Clínicos (continuación de la Tabla 4.1).

Clases	Puntos de corte	S %	E %	Y	A	VPP %	VPN %	LR+	LR-
1	6	13	14	15	16	17	18	19	20
0-20	20	100	52	0,52	0,76	67,57	100	2,08	0
21-40	40	96	70	0,66	0,83	76,19	94,59	3,2	0,057
41-60	60	88	84	0,72	0,86	84,62	87,5	5,5	0,143
61-80	80	72	93	0,65	0,825	91,14	76,86	10,29	0,301
81-100	100	49	98	0,47	0,735	96,08	65,77	24,5	0,52
101-120	120	16	100	0,16	0,58	100	54,35	∞	0,84
121-140	140	4,5	100	0,045	0,523	100	51,15	∞	0,955
141-160	160	1,5	100	0,015	0,508	100	50,38	∞	0,985
161-180	180	0	100	0	0,5	***	50	∞	1

Paso 11 : Para calcular la sensibilidad S en forma porcentual de la columna (13), se emplea la relación: $S = (vp / TE) 100$ y en este caso resulta ser $S = vp/2$, pues $TE = 200$.

Paso 12 : Para calcular la especificidad E en forma porcentual de la columna (14), se aplica la relación: $E = (vn / TS) 100$ y en este caso resulta ser $E = vn/2$, pues $TS = 200$.

Paso 13 : El Índice de Youden de la columna (15) se obtiene con: $Y \% = [S + E - 1] \cdot 100$.
 O sea se suman las columnas (13) y (14) y se le resta 100.

Paso 14 : La Eficiencia de la columna (16) se la calcula con la relación $A = (vp + vn) / N$
 O sea, el número total de diagnósticos acertados sobre el número total de casos.

Paso 15 : En la columna (17) se calcula el Valor Predictivo de Positivos en forma porcentual con: $VPP = (vp / TP) 100$

Paso 16 : En la columna (18) se saca el Valor Predictivo de Negativos en forma porcentual con: $VPN = (vn / TN) 100$

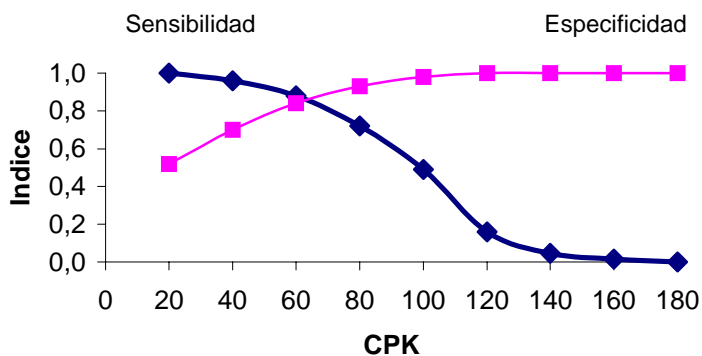
Paso 17 : En la columna (19) se saca el Likelihood Ratio de la enfermedad dado un resultado positivo con: $LR+ = S / (1 - E)$ usando los valores respectivos de las columnas (13) y (14).

Paso 18 : En la columna (20) se saca el Likelihood Ratio de la enfermedad dado un resultado negativo con: $LR- = (1 - S) / E$ usando los valores respectivos de las columnas (13) y (14).

Se pueden graficar los valores obtenidos, para ilustrar acerca de la variabilidad de los índices clínicos con el punto de corte. Notar que, todos los índices varían con el punto de corte adoptado, no así con la prevalencia de la enfermedad que es otra de las causas de variabilidad. Todo esto indica la dificultad de tener situaciones estables en la práctica. El supuesto principal es que la sensibilidad y la especificidad no varían con la prevalencia. Sin embargo se puede dar el caso de encontrar una variación en ambas y las explicaciones más comunes son: (a) el método no es estable. (b) el avance de la enfermedad en los pacientes (se supone que cuanto más avanzada esté la enfermedad, hay menores posibilidades de equivocarse en el diagnóstico).

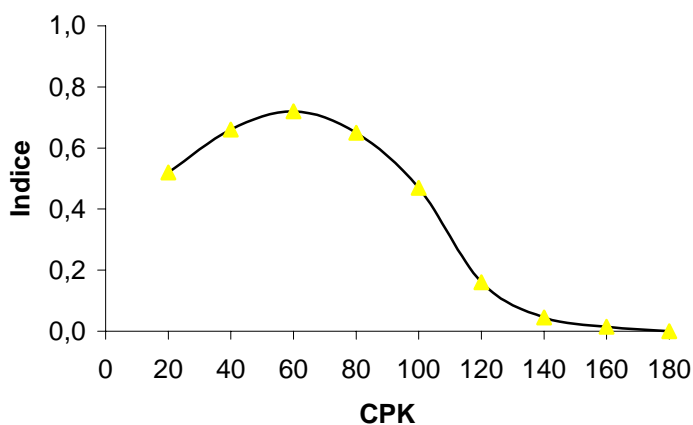
Gráfico 4.3: Variación de Sensibilidad y Especificidad con el punto de corte en la CPK.

El hecho más notable es que *si la sensibilidad aumenta, la especificidad disminuye y viceversa*.



Por eso, el problema real es decidir cual de ellas debe ser optimizada en cada caso, adoptando un punto de corte apropiado al tipo de enfermedad estudiada. Si la enfermedad es del Tipo I conviene elegir el punto de corte que maximiza la Sensibilidad ($X = 20$ UI/l). Pero si la enfermedad es del Tipo II conviene elegir al punto que maximiza la Especificidad ($X \geq 120$ UI/l). Cuando la enfermedad sea del Tipo III conviene maximizar el índice de Youden, como se puede ver en el Gráfico 4.5 siguiente. Notar que esto ocurre para el valor $X \approx 60$ UI/l de la magnitud clínica, que es el punto donde se cortan las curvas de sensibilidad y especificidad aproximadamente.

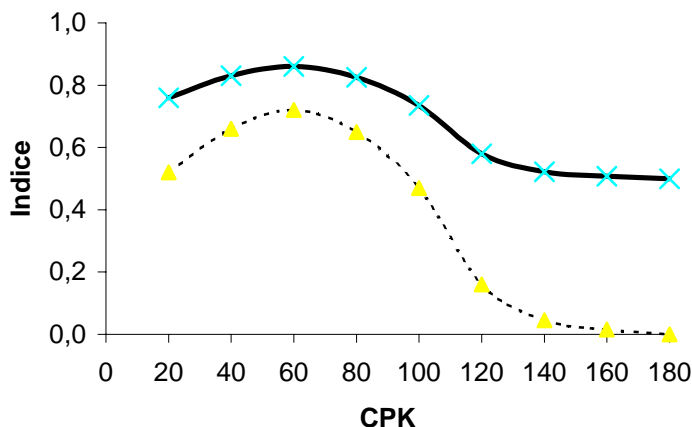
Gráfico 4.4: Variación del Índice de Youden con el punto de corte en la CPK



Esto muestra otra de las ventajas del índice de Youden. Advierte el peligro de tomar puntos de corte tales que el diagnóstico no se diferencie del azar. Por ejemplo, si se toma como punto de corte el valor 180 UI/l, en la Tabla 4.2, se obtiene $E = 1$ y $S = 0$, o sea un $Y = 0$ y el diagnóstico resultará como al azar, esto es como si se utilizase el lanzar una moneda al aire para efectuar el diagnóstico. En resumen, este índice debe acompañar siempre el análisis de sensibilidad y especificidad de los diagnósticos basados en una prueba clínica de laboratorio.

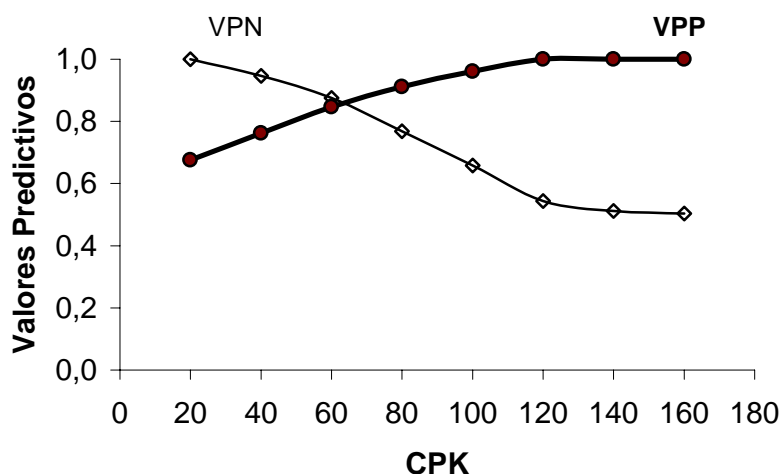
En el Gráfico 4.5 se muestra la variabilidad de la eficiencia con el punto de corte. Además se ha vuelto a mostrar la variabilidad del Índice de Youden en línea punteada, para visualizar lo similar de ambas curvas en su forma. En ambos casos el punto más aconsejable para una enfermedad del Tipo III es donde se maximizan ambas curvas ($X \approx 60$ UI/l).

Gráfico 4.5: Variación de la eficiencia (A) con el punto de corte en la CPK



Para ilustrar la variabilidad de los valores predictivos de acuerdo al punto de corte adoptado por el clínico, se presenta el Gráfico 4.6 donde se puede ver que: (a) el punto de corte ideal para una enfermedad del Tipo I corresponde al valor máximo del VPN ($X = 20$ UI/l), al igual que ocurrió en el caso de la sensibilidad. (b) cuando $X \geq 120$ UI/l el VPP se maximiza y corresponde al punto ideal para las enfermedades del Tipo II, como en el caso de la especificidad. (c) ambas curvas se cortan en $X \approx 60$ UI/l que es el punto ideal para las del Tipo III.

Gráfico 4.6: Variación de los valores predictivos con el punto de corte en la CPK



En resumen:

- 1) Sensibilidad, Especificidad, Índice de Youden y Likelihood Ratios varían con el punto de corte, pero no lo hacen con la prevalencia de la población (son parámetros del método clínico).
- 2) Eficiencia y Valores Predictivos varían con el punto de corte y con la prevalencia (no son parámetros sino variables del método clínico).

4.8 Otros Índices diagnósticos

Además de los índices ya vistos, en la bibliografía aparecen algunos otros que no agregan mucho en el ámbito conceptual; son variantes lógicas derivadas de los anteriores. Por ejemplo, en lugar de calcular el porcentaje de aciertos con la eficiencia, también se puede definir su complemento como:

$$\text{Porcentaje de errores} : \quad PE \% = \{ (fn + fp) / N \} \cdot 100$$

$$PE \% + A \% = 100\%$$

Se busca minimizar a este porcentaje para acercarse a la técnica ideal para diagnosticar. A veces, también se lo llama *error combinado*. Otro índice es el llamado *Positivo presunto*: se calcula como el porcentaje de pruebas, cuyo resultado dio positivo, respecto del total N de pruebas realizadas, es decir:

$$\text{Positivo presunto} : \quad PP\% = (TP / N) \cdot 100$$

Otra forma parecida es calcular el porcentaje de aciertos en positividad definido con :

$$\text{Índice de detección} : \quad ID = (vp / N) \cdot 100$$

La proporción de error se define como el cociente entre el número total de errores y los diagnosticados como enfermos correctamente, esto es :

$$\text{Proporción de error} : \quad P\%E = \{ (fp + fn) / TN \} \cdot 100$$

Hay confusiones en la bibliografía, acerca de los términos: *Incidencia*, *Predominancia* y *Morbilidad*. Para acotar este problema se toma el aporte de Galen definiendo:

Incidencia : Es la cantidad de enfermos que aparecen en un período dado de tiempo (un año), en la población, sobre el tamaño de la población expresado cada 100.000 habitantes. Es una medida epidemiológica tomada en períodos fijos de tiempo.

Predominancia : Es la cantidad de enfermos, en un momento dado de la población, dividido el número total de habitantes y expresado cada 100.000. O sea, es una medida de tipo censal.

Morbilidad : Se usa en Demografía, y puede adaptarse a los dos casos anteriores, pero sin usar un nombre específico que los distinga.

La predominancia o prevalencia de una enfermedad en la población afecta grandemente a los Valores Predictivos y sus índices relacionados, pero no lo hacen con la Sensibilidad, la Especificidad y el Índice de Youden. Por lo tanto, es muy importante tener en cuenta este valor a la hora de seleccionar un punto de corte buscando optimizar el diagnóstico. Se podrían seguir agregando otras proporciones e incrementando la lista de índices clínicos, más allá de los vistos. Sin embargo, a modo de resumen se presentan los cuatro postulados de Galen y Gambino, para la elección práctica de las pruebas clínicas en el diagnóstico:

(1) *Se elige alta Sensibilidad cuando las enfermedades son del Tipo I*

- La enfermedad puede curarse si es detectada a tiempo.
- La enfermedad es grave y no puede dejarse pasar inadvertida.
- La enfermedad es tratable.
- Los fp no trauman ni psicológicamente, ni económicamente al paciente.
- Ejemplos: Cáncer de útero o de mama, enfermedades venéreas y otras enfermedades infecciosas curables, la fenilcetonuria, el feocromocitoma

(2) *Se elige alta Especificidad cuando las enfermedades son del Tipo II*

- La enfermedad es grave, pero difícilmente curable o sin remisión.
- Es muy importante para el paciente o para la población él saberse un vn.
- Los fp trauman seriamente al paciente.
- Ejemplos: Cáncer oculto, esclerosis en placas, etc.

(3) *Se elige un alto Índice de Youden cuando las enfermedades son del Tipo III :*

- La enfermedad es importante pero curable.
- Tanto los fp como los fn, sean igualmente graves para el paciente o para la población.
- Ejemplo: SIDA, lupus eritomatoso, ciertas formas de leucemia o linfoma, diabetes, etc.

4.9 Índices de riesgo o daño.

En el primer capítulo se explicaron las cuatro cuestiones clínicas fundamentales que son: terapia, daños, diagnóstico y prognosis. Para tratar la primera cuestión de la terapia, se mostró el diseño de los RCT para hacer investigaciones. Para tratar la segunda cuestión se usan las tablas diagnóstica y los índices detallados en los puntos anteriores. Para la tercera y cuarta cuestión se presentaron los diferentes diseños de investigación del riesgo o daño, que se explicaron más detalladamente en el segundo capítulo (estudios por cohortes, caso-control y RCT). La forma más sencilla de presentar los datos es con una tabla doble dicotómica llamada tabla de riesgo (ver Tabla 1.2) donde se generaliza el cuadro para los diferentes tipos de diseño experimental para el estudio del riesgo:

Factor de riesgo	Resultados observados		Total
	Enfermos	Sanos	
Expuesto	<i>a</i>	<i>b</i>	TE _x
No expuesto	<i>c</i>	<i>d</i>	TnEx
Total	TE	TS	N

Usando la información de esta tabla se pueden calcular los dos índices clínicos básicos para poder cuantificar el concepto de riesgo o daño:

Odds Ratio: $OR = (a \cdot d) / (b \cdot c)$

Riesgo Relativo: $RR = [a (c + d)] / [c (a + b)]$

El Odds Ratio se define como el cociente entre dos Odds, el numerador es el Odds de un evento en el grupo de individuos expuestos, dividido el Odds del mismo evento entre los individuos no expuestos. El evento en este caso es la ocurrencia de la enfermedad o la aparición del síntoma (efecto) buscado en la investigación. Un Odds se define como el cociente entre la probabilidad de que ocurra un evento, dividida la probabilidad de que no ocurra. Por ejemplo, si la probabilidad de contraer una cierta enfermedad es del 20%, el Odds de ese evento será el cociente entre ese 20% y el 80% de que no ocurra. Esto es, $Odds = 20\% / 80\% = 1 / 4$; lo que significa que en 5 casos hay una sola posibilidad de contraer la enfermedad y cuatro de no hacerlo. El significado del OR en Medicina representa la proporción de pacientes que presentan la condición buscada, dividida por la proporción que no la presentan. Muchas veces se toma el concepto de Odds y de riesgo como si fuera la misma cosa, o aproximadamente iguales. Muchos autores calculan los Odds relativos (otra forma de denominar los OR) y entonces informan sus resultados como si hubiesen calculado el riesgo relativo. Por lo tanto conviene mirar esta cuestión más de cerca, por ejemplo si se sabe que 1 / 5 de los pacientes estudiados sufren un golpe de presión, los Odds de que sufran el golpe es de 1 / 4 como se vio antes, o bien 0,25. Para convertir los Odds en riesgo se puede aplicar la relación siguiente:

$$Riesgo = (Odds) / (1 + Odds)$$

En este caso sería, $Riesgo = 0,25 / 1,25 = 0,2$, que es lo mismo que 1 / 5 o 20% que fue el punto de partida. Análogamente se puede expresar la relación inversa con:

$$Odds = Riesgo / (1 - Riesgo)$$

Notar que el riesgo siempre es un valor entre 0 y 1, en cambio los Odds siempre son positivos entre dos casos extremos 0 e infinito. La siguiente es una tabla con los valores más comunes:

Tabla 4.3: Relación entre el riesgo y los Odds

<i>Riesgo</i>	<i>Odds</i>
80%	4
60%	1,5
50%	1
40%	0,67
33%	0,5
25%	0,33
20%	0,25
10%	0,11
5%	0,053

Ejemplo: A 360 pacientes que tuvieron cesáreas en sus partos, 170 de ellas tuvieron una cateterización umbilical durante la cirugía y las demás no. Entre ellas se observaron 60 que desarrollaron una bacteremia y 300 que no enfermaron. Los datos se presentan en la tabla siguiente:

Factor de riesgo	Resultados observados		Total
	Infección	No-infección	
Cateterización	50	120	170
No cateterización	10	180	190
Total	60	300	360

$$OR = (50 \cdot 180) / (120 \cdot 10) = 7,5 \text{ mientras que } RR = (50 \cdot 190) / (10 \cdot 170) = 5,6$$

En este caso el Odds de enfermarse en las cateterizadas es de 50 versus las 120 que no se infectaron, o sea $50 / 120 = 0,42$. Mientras que el Odds de infectarse en las no cateterizadas es de 10 versus las 180 que no enfermaron, esto es $10 / 180 = 0,056$. Por lo tanto, el cociente entre estos dos Odds resulta $(50/120) / (10/180) = OR = 7,5$. Si se hablase en términos de riesgo o daño, se podría denominar al cociente de Odds como Odds relativos, pero el término impuesto por la Epidemiología es OR. Un médico tiene una buena intuición acerca del riesgo y aún del concepto de un cociente de riesgos como RR. Un jugador tiene buena intuición sobre los Odds, pero nadie (excepto tal vez algunos estadísticos) puede intuitivamente captar el concepto de OR. A pesar de esto, el OR es la medida de asociación más empleada en Medicina, porque tiene ventajas estadísticas en cuanto al hecho de que es independiente de la selección arbitraria entre una comparación de riesgos entre enfermos y sanos, lo cual no es verdadero con el RR. Los clínicos tienen una tendencia natural a sustituir el concepto intuitivo de RR por el concepto poco intuitivo del OR, lo cual puede ser correcto bajo ciertas condiciones: Por ejemplo si $a \lll b$ y $c \lll d$. En otras palabras, la enfermedad tiene que ser mucho menos frecuente que la no-enfermedad, tanto entre los expuestos como entre los no expuestos.

Por su parte el RR muestra la proporción del riesgo original (en este ejemplo el riesgo de enfermarse al ser cateterizadas), que está aún presente cuando las pacientes no fueron expuestas (en este caso sin cateterización). En el ejemplo, el riesgo de enfermarse al ser cateterizadas es de 50 casos entre las 170 en total ($50/170$), mientras que el riesgo de enfermarse sin haber sufrido una cateterización es de 10 casos entre las 190 en total ($10/190$). Por lo tanto el cociente entre ambos riesgos es: $RR = (50/170) / (10/190) = 5,6$. Esto significa que el hecho de sufrir una cateterización durante la cesárea incrementa el riesgo de infectarse 5,6 veces. Intuitivamente, uno puede pensar que no conviene la cateterización en cesáreas, pero clínicamente se sabe que hay otros factores que pueden producir una infección en una cirugía, los que deben ser analizados cuidadosamente si lo que se busca es reducir el número de infectadas.

4.9.1 Índices de riesgo derivados.

Se suelen usar otros índices de riesgo conceptualmente derivados de los anteriores como por ejemplo los siguientes:

Riesgo absoluto: Es el riesgo de contraer la enfermedad habiendo estado expuesto, o el de enfermarse sin haber estado expuesto.

En el ejemplo anterior sería 29,4% (50/170) el riesgo de infectarse habiendo sufrido la cateterización, mientras que el riesgo absoluto de enfermarse sin haber sido cateterizada es de un 5,3% (10/190). A este último valor se lo denomina también proporción del evento en el grupo control (CER) o bien *riesgo básico*. En cambio al riesgo absoluto de enfermarse habiendo estado expuesto se lo denomina (EER) proporción del evento experimental en el grupo expuesto. En otras palabras, se puede considerar al riesgo relativo como el cociente entre los dos riesgos absolutos:

$$RR = EER / CER = [a / (a + b)] / [c / (c + d)]$$

Reducción del riesgo absoluto (ARR): Es la diferencia entre el riesgo básico y el experimental tomado en valor absoluto.

$$ARR = | CER - EER |$$

En el ejemplo visto se calcula como: $ARR = | 5,3\% - 29,4\% | = 24,1\%$. Cuando la diferencia da negativa como en este caso, significa que el exponerse a una cateterización aumenta el riesgo de infectarse en un 24,1% en términos relativos. En cambio cuando la diferencia da positiva significa que el riesgo disminuye, y corresponde al caso de aplicarse una terapia de cura, o al efecto de un medicamento que en efecto funcionan.

Reducción del riesgo relativo (RRR): Es una estimación de la proporción de riesgo básico, que es removida por la terapia, o bien por la exposición. Se calcula como el cociente entre:

$$RRR = ARR / CER = | 1 - RR |$$

En el ejemplo es $RRR = 24,1\% / 5,3\% = 4,6$; o bien $| 1 - 5,6 | = 4,6$. Y significa, usando un lenguaje no técnico que la cateterización incrementa el riesgo relativo de enfermarse 4,6 veces, en comparación a la cirugía sin la cateterización.

4.10 Concordancia clínica

Además de la validez interna de un método, dada por su prestación medida por los índices clínicos antes descriptos, se necesita otra cosa más: *fidelidad*. Esto es, no debe ser influenciado por la persona que lo ejecuta. No debe ocurrir que el mismo test, hecho por dos bioquímicos diferentes al mismo paciente, en el mismo tiempo y lugar, arrojen resultados opuestos o muy diferentes. Significa, que se necesita la reproducibilidad de los resultados de la prueba, independientemente de quien la ejecuta. Se define un índice para cuantificar esta situación, con el nombre de grado o *Nivel de Concordancia* (λ %):

$$\lambda \% = (NC / N) \cdot 100$$

Donde NC: es el número de diagnósticos que concuerdan
N : número total de casos

Un test perfectamente reproducible tendría un índice $C\% = 100\%$. Para estudiar la concordancia entre dos métodos clínicos, se arma la llamada Tabla de Concordancia que consta de 2 filas y 2 columnas (Tablas 2×2). A su vez cada fila o columna puede estar subdividida en clases o categorías. Se habla de una tabla de $k \times k$ cuando hay k clases en filas y columnas. Esto se esquematiza en el Cuadro 4.1, donde por ejemplo hay dos bioquímicos A y B, que pueden diagnosticar k casos diferentes, a partir de la misma prueba clínica. Entonces, se puede armar una tabla como la siguiente:

Cuadro 4.1: Tabulación cruzada en un diagnóstico.

		Bioquímico A					Σ
		1	2	3	...	k	
Bioquímico B	1	n_{11}	n_{12}	n_{13}	n_{1k}	$n_{1.}$	
	2	n_{21}	n_{22}	n_{23}	n_{2k}	$n_{2.}$	
	3	n_{31}	n_{32}	n_{33}	n_{3k}	$n_{3.}$	
		
	k	n_{k1}	n_{k2}	n_{k3}	n_{kk}	$n_{k.}$	
Σ	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.k}$	N		

$$\lambda \% = \{ (n_{11} + n_{22} + n_{33} + \dots + n_{kk}) \} \cdot 100 / N$$

$$\lambda \% = (100 \cdot \Sigma_k n_{kk}) / N$$

La concordancia observada, es la sumatoria de la cantidad de diagnósticos coincidentes para cada caso n_{kk} , dividida por el número total de casos, expresada porcentualmente. En el gráfico, se suman filas y columnas, colocando un punto como subíndice para subrayar este hecho. Así, el total de la 2ª columna es $n_{.2}$ y el total de la primera fila $n_{1.}$. En el ejemplo anterior se obtuvo un índice del 50% lo que significa una mala concordancia entre ambos bioquímicos.

La manera más usual de aplicar este concepto es para comparar dos test clínicos entre sí, cuando se aplican al mismo paciente (es decir muestras apareadas), como fuera presentado en el primer capítulo (Tabla 1.3) como la siguiente:

Método 2	Método 1		Total
	(+)	(-)	
(+)	a	b	$a + b$
(-)	c	d	$c + d$
Total	$a + c$	$b + d$	N

Donde el número total de concordancias está dado por la cantidad (a + d) y el número total de discordancias por la cantidad (b + c). Siendo N el número total de casos analizados. Los índices de concordancia más usados son dos:

Nivel de concordancia: $\lambda = (a + d) / N$ (expresado en forma porcentual: $\lambda\% = 100 / [1 + DO]\%$)

Odds de discordancia: $DO = (b + c) / (a + d)$ y su relación con $\lambda\%$ es: $DO = (100 / \lambda\%) - 1$

El nivel de concordancia indica el porcentaje de éxitos (concordancias) obtenido por la comparación de las pruebas. La idea es compararlo contra un valor recomendado en la literatura médica para que la concordancia sea considerada aceptable. Por ejemplo, un nivel de concordancia del 90% es suficiente para muchas enfermedades, pero no alcanza para casos como el Sida o ciertos tipos de cáncer.

Los Odds de discordancia son otra manera de expresar el mismo concepto y es usado por aquellos que prefieren el concepto de Odds al de probabilidad. Por ejemplo, un $DO = 1/9$ significa que habrá 1 discordancia y 9 concordancias en 10 casos. O que se tendrá un desacierto cada 9 aciertos entre ambas pruebas. La concordancia ideal se da cuando $DO = 0$ ($\lambda = 100\%$), en cambio cuando la concordancia sea $DO = 1$ significa que habrá el mismo número de concordancias y de discordancias, lo cual clínicamente está en el límite de lo aceptable ($\lambda = 50\%$). Cuando DO sea mayor que 1, o bien $\lambda < 50\%$, la discordancia entre ambos métodos será inaceptable desde un punto de vista clínico.

Ejemplo: Comparación de diagnósticos hechos por dos individuos diferentes.

A 200 pacientes internados en un sanatorio, se les efectuó una análisis de sangre en heces. Dos bioquímicos proceden a efectuar el test al mismo grupo de individuos. Los resultados fueron:

Diagnósticos:	Sangre en heces			Σ
	0	+	++	
Bioquímico A	120	56	24	200
Bioquímico B	120	56	24	200

De acuerdo al cuadro anterior, la coincidencia entre ambos parece perfecta. Sin embargo, si se efectúa una tabulación cruzada para evitar que el mismo paciente aparezca dos veces en la tabla y lograr así independencia. Se obtiene:

		Bioquímico A			Total
		0	+	++	
Bioquímico B	0	80	40	0	120
	+	28	12	16	56
	++	12	4	8	24
Total		120	56	24	200

De donde se puede calcular:

$$\lambda \% = 100 \cdot (80 + 12 + 8) / 200 = 50 \%$$

4.11 Estadígrafos clásicos

En este punto se describen los diferentes estadígrafos, llamados también *números índices*, *estadísticas*, *estadísticos*, etc. Clásicamente, se trata de valores que se calculan para magnitudes clínicas cuantitativas, y no cualitativas. Se trata de valores que aportan mucha más información que la propia, como datos en sí mismos. Así como un histograma puede ser considerado como una manera clara y práctica de resumir la información contenida en la “nube” de datos, los estadígrafos la condensan aún más, reduciéndola a unos pocos valores que toma el lector para darse una idea de todo el conjunto. El trabajo de analizar un gran número de datos siempre es engorroso; en la actualidad, con las computadoras la tarea está mucho más aliviada, pero en los inicios era una tarea muy pesada. Por eso, desde siempre se buscó la manera de unificar criterios y técnicas para homogeneizar la manera de presentar datos en los informes científicos. Eso evitó la anarquía que se hubiese tenido, si cada investigador los tratase a su peculiar manera, para luego presentarlos a consideración de los demás en congresos, simposios, publicaciones, etc.

Todo esto llevó a encontrar formas simples y concisas para describir el conjunto de datos hallados en experimentos. Formas conocidas y aceptadas por todos los científicos, a manera de convención. Maneras cuantitativas que permitan efectuar comparaciones entre trabajos y caracterizar la población de donde fueron extraídos los datos. Por esa necesidad, aparecieron los estadígrafos, como por ejemplo números para dar una idea al lector de la posición de la nube de datos, o un valor alrededor del cual los datos parecen agruparse, o un número que cuantifica la dispersión de los datos alrededor del anterior. Además, cada profesión aporta más índices propios de su campo específico. En Economía se usan en forma habitual, los índices de costo de vida, de la construcción, de precios mayoristas no agropecuarios, etc. En Demografía se usan las tasas de mortalidad, morbilidad, fecundidad, casamientos, etc. En Sanitarismo se emplean la prevalencia, la predominancia, y otros muchos casos. En Clínica, los hay específicos para el diagnóstico de enfermedades, a partir de los resultados de los análisis del laboratorio. Los estadígrafos clásicos se dividen en dos tipos generales: los llamados estadígrafos de *posición* que dan una idea de alrededor de que valor parecen agruparse los datos, y los de *dispersión* que muestran como se distribuyen los datos alrededor de los de posición.

4.12 Estadígrafos de posición

Se denominan así, a los estadígrafos que indican un valor, alrededor del cual los datos parecen agruparse de cierta manera, como si fuese el “centro de gravedad de los datos”, tal es el caso de los estadígrafos de tendencia central (media, mediana y moda). O bien, dejando una cierta cantidad de valores por debajo o por encima de él (percentiles).

4.12.1 Medias

$$\text{Media aritmética : } \bar{x} = \sum_{i=1}^n (x_i / n) \quad (\text{para datos sueltos})$$

$$\text{Media aritmética ponderada : } \bar{x} = \sum_{k=1}^m (f_k \cdot x'_k / n) \quad (\text{para datos agrupados})$$

Cuando se conocen los datos en forma individual, se puede calcular la *media aritmética* tradicional en forma directa aplicando la primera de las fórmulas de arriba. En cambio, si la información se suministra a través de n datos agrupados en m clases, se usa la *media aritmética ponderada* usando la otra fórmula. Donde la frecuencia de cada clase se expresa con f_k , siendo x'_k la marca de cada clase (es decir el punto medio del ancho de clase, obtenido como la semi-suma de sus límites reales de clase). La expresión de la media aritmética usando las frecuencias de clase no es otra cosa que un promedio ponderado, donde se le da un peso porcentual a cada marca de clase, de acuerdo al tamaño que tiene cada una. Como cada valor de frecuencia está dividido por la frecuencia total, la fórmula resulta la sumatoria de las marcas de clase (o valores individuales) multiplicada por la frecuencia relativa respectiva. Esta fórmula de media ponderada se aplica de muchas maneras en la práctica: es un concepto similar al centro de masa visto en Física (o centro de gravedad), la velocidad promedio en la teoría cinética de los gases, etc.

CUADRO 4.2: Cálculo de la media aritmética.

Caso 1: Datos individuales. (usando los datos para los no infartados del Cuadro 3.1).

$$\bar{x} = \sum_{i=1}^n (x_i / n) = \frac{10 + 33 + 20 + \dots + 120 + 92 + 90}{200} = 30,735 \text{ UI/l}$$

Caso 2 : Datos agrupados en clases (usando los mismos datos, pero tomados del Cuadro 3.3).

CPK	f_k	x'_k	$f_k \cdot x'_k$	
1 - 20	104	10,5	1092	La media aritmética aproximada se calcula con : $\bar{x} = 6220 / 200 = 31,1 \text{ UI/l}$
21 - 40	36	30,5	1098	
41 - 60	28	50,5	1414	
61 - 80	18	70,5	1269	
81 - 100	10	90,5	905	
101- 120	4	110,5	442	
121- 140	0	130,5	0	
141- 160	0	150,5	0	
161- 180	0	170,5	0	
Σ	200		6220	

El valor exacto de la media aritmética, está dado por el primer caso, donde se tienen identificados los datos, uno por uno. El segundo valor de la media, obtenido a partir de su agrupación en clases, es solo una aproximación, pues al perder información se pierde exactitud (aumentó un 2,7%).

Caso 3 : Datos y sus frecuencias (valores tomados de un conteo celular - Student, 1907).

Nº de células x cuadrícula	0	1	2	3	4	5
Frecuencias Observadas	213	128	37	18	3	1
$\bar{x} =$	$\frac{\{0(213) + 1(128) + 2(37) + 3(18) + 4(3) + 5(1)\}}{213 + 128 + 37 + 18 + 3 + 1} = \frac{273}{400} = 0,6825$					

En este caso las marcas de clase coinciden con la cantidad de células por cuadrícula de la cámara que tiene en total 400.

Caso 4. Una farmacia maneja 4 familias de productos en su trabajo. Los márgenes de utilidad de cada uno de ellos, durante el período fiscal pasado fueron: Familia A: 8,4%; la B: 11,0%; la C: 14,8% y la familia D: 20,2%. Se pide calcular el margen de utilidad promedio.

Producto	Ventas (miles de \$) (V)	Márgenes (%) (M)	Participación (V.M)
A	300	8,4	25,20
B	200	11,0	22,00
C	50	14,8	7,40
D	<u>30</u>	20,2	<u>6,06</u>
Totales	580		60,66

Si se calculase el promedio de los márgenes en forma directa, sería de: 13,6%, lo que no es correcto. Pero esto *sólo es correcto si las 4 familias tuvieron igual participación en las ventas*. Lo cual es muy raro que ocurra. Por lo tanto, el farmacéutico debe estudiar como se le distribuyeron las ventas y hacer la *ponderación* respectiva como se calcula más arriba. Entonces, el margen promedio ponderado buscado es: $M = \frac{\sum (V_k \cdot M_k)}{\sum V_k} = 60,66 / 580 = 10,46\%$

Caso 5. El tiempo de atención al cliente en una farmacia con varios vendedores fue medido durante un día de trabajo donde se atendieron 47 clientes. El encargado de medir los tiempos informó lo siguiente. El farmacéutico a cargo quiere calcular el tiempo promedio que tiene su negocio a fin de mejorarlo.

Informe recibido		Cálculos efectuados por el farmacéutico			
Tiempo en minutos	Nº clientes	Clases	Medios	Frec. ponderada	Frec. acumulada
Entre 5 y 8	10	5- 7,99	6,5	65,0	10
Más de 8 y menos de 11	17	8-10,99	9,5	161,5	27
Más de 11 y menos de 14	12	11-13,99	12,5	150,0	39
Más de 14 y menos de 17	6	14-16,99	15,5	93,0	45
Más de 17 y menos de 20	2	17-19,99	18,5	37,0	47
Más de 20	<u>0</u>	20 y más	---	<u>---</u>	
Totales	47			506,5	

El tiempo promedio de atención al cliente es: 10,8 minutos = 506,5 / 47

Caso 6. En una industria farmacéutica se fabrican tres tipos de ampollas de un mismo medicamento. Se desea averiguar el porcentaje de fallas de la máquina envasadora. Luego de un mes de pruebas, Control de Calidad obtuvo los datos siguientes:

Ampollas	% defectuosas	Nº de ampollas	Amp. defectuosas
Tipo A	1,1	210.000	2310
Tipo B	1,5	120.000	1800
Tipo C	2,3	<u>50.000</u>	<u>1150</u>
Totales		380.000	5260

El promedio ponderado de ampollas defectuosas es de 1,38 % = 5260/380.000

Caso 7. Se observan 10 pacientes adultos de sexo masculino, que llegan a hacerse a atender al laboratorio Otros 15 pacientes jóvenes y 25 niños. En cambio, de sexo femenino llegan 20 adultas, 5 jóvenes y 15 niñas. Calcular el promedio de cada grupo de edades (grupo etáreo).

Resulta que en promedio llegan: 15 adultos, 10 jóvenes y 20 niños.

En el Cuadro 4.2 se han desarrollado ejemplos de aplicación. En el primer caso se muestra un ejemplo de media aritmética simple aplicado a los 200 casos de los no infartados del Cuadro 3.1. La media es 30,735 UI/l. En cambio, si se tuviesen los mismos datos pero agrupados en clases, solo se podría calcular una estimación de la media a través de una ponderación. Usando los datos del Cuadro 3.3 para los no infartados y la fórmula de media ponderada, se obtiene 31,1 UI/l, o sea un aumento del 2,7% respecto del valor exacto. En el tercer caso se calcula la media ponderada del conteo celular pues se carecen de las mediciones individuales hechas por Student en 1907. Los tres casos siguientes son aplicaciones en Farmacia del concepto de la media. En el Caso 4 se desea calcular el margen promedio de ganancias del negocio, utilizando los márgenes de 4 líneas de productos. En el Caso 5 el farmacéutico desea mejorar la atención al público reduciendo los tiempos de espera de sus clientes. El primer paso para ello es calcular el tiempo promedio de atención al cliente en el mostrador. Toma los tiempos agrupados en clases y sus frecuencias para poder hacer la ponderación del promedio. Descubre que en su farmacia se tarda casi 11 minutos en despachar a un cliente. En el último caso se desea saber el porcentaje de ampollas falladas de una máquina envasadora. Durante un período de un mes se calculan los porcentajes de fallas para tres productos diferentes A, B y C que se envasan en ese equipo, y efectuando la ponderación resulta 1,38% de fallas. En el último ejemplo, se muestran los promedios por grupo etéreo de pacientes que concurren al laboratorio.

La media sólo se usa en variables cuantitativas, pues en las cualitativas carece de sentido. Las principales propiedades de la media se muestran en el Cuadro 4.3., y se explican a continuación por su importancia en los capítulos posteriores:

Cuando a una media se le suma o resta una constante, la nueva media queda sumada (o restada) por la misma constante. Esto es el efecto que tiene un error de tipo sistemático en una serie de mediciones, porque el efecto del mismo es siempre del mismo signo y aproximadamente de la misma magnitud. Todo ocurre como si se trasladase el eje de coordenadas a lo largo del eje de abscisas (a la derecha cuando se le suma una constante y a la izquierda cuando se la resta).

Cuando a la media se la multiplica o divide por una constante, la nueva media queda multiplicada o dividida por dicha constante. Esto es el efecto de cambiar de escala en un gráfico. Otra propiedad de mucha importancia es: La sumatoria de los desvíos de valores medidos en un sistema respecto de su media es nula. Esto significa que de todos los puntos posibles para referir las mediciones, la media aritmética equilibra esos desvíos. Lo que explica que se puede considerar a la media como el “centro de gravedad” de los datos. Una aplicación práctica de la media aritmética ponderada es el cálculo del centro de masa, o el de inercia en Física.

El método de los mínimos cuadrados se basa en la propiedad: Si se suman los cuadrados de los desvíos de todas las mediciones respecto a un punto cualquiera, se demuestra que tal suma se hace mínima cuando el punto elegido es la media aritmética. Por eso, si se divide tal suma por los grados de libertad se tiene un desvío de referencia llamado: *desvío estándar*.

Existen otros tipos de medias como la *media geométrica* y la *media armónica*, usadas en otras disciplinas. Por eso, se deja al lector la inquietud de verlas en la bibliografía. La media geométrica se usa para interpolar datos en funciones exponenciales. El uso clásico de la media armónica es el cálculo de la velocidad promedio de un móvil que recorre una distancia a diferentes velocidades.

CUADRO 4.3: Propiedades de la media aritmética.

Sean n observaciones de una variable clínica cualquiera $(x_1, x_2, x_3, \dots, x_n)$, cuya media aritmética es: \bar{x} , entonces :

1. Si a cada valor medido se le suma (o resta) una constante a , la nueva media aritmética es igual a la anterior, más (o menos) dicha constante. Esto se demuestra así:

$$\text{Si } z_i = x_i + a, \text{ entonces } \bar{z} = \frac{\sum_{i=1}^n z_i}{n} = \frac{\sum (x_i + a)}{n} = \frac{\sum x_i}{n} + \frac{n \cdot a}{n}$$

O sea

$$\bar{z} = \bar{x} + a$$

Por comodidad, se le han sacado los subíndices a las sumatorias y así será de ahora en más.

2. Si a cada valor medido se le multiplica (o divide) por una constante a , la nueva media aritmética es igual a la anterior, multiplicada (o dividida) por dicha constante. Esto se demuestra :

$$\text{Si } z_i = x_i \cdot a; \text{ entonces } \bar{z} = \frac{\sum z_i}{n} = \frac{\sum (x_i \cdot a)}{n} = a \cdot \frac{\sum x_i}{n} = a \cdot \bar{x}$$

3. La sumatoria de todos los *desvíos* es nula. Esto se demuestra así :

Por desvío, se entiende a la diferencia entre cada valor medido y su promedio, es decir, para el dato número i su desvío es $d_i = x_i - \bar{x}$

$$\sum d_i = \sum (x_i - \bar{x}) = \sum x_i - \sum \bar{x} = n \cdot \bar{x} - n \cdot \bar{x} = 0$$

4. La sumatoria de los cuadrados de los desvíos es un mínimo. Más conocida como el principio de los mínimos cuadrados. Esto se demuestra así :

Sea un valor cualquiera a , para saber que valor de a hace mínima la sumatoria, se saca la primer derivada y se iguala a cero:

$$\frac{\delta}{\delta a} (\sum d_i^2) = \frac{\delta}{\delta a} [\sum (x_i - a)^2] = \sum [2(x_i - a)(-1)] = -2 \sum (x_i - a) = 0$$

El único valor que cumple con la propiedad anterior, es el promedio. O sea, $\sum (x_i - a) = 0$ sólo sí $a = \bar{x}$, de acuerdo a la propiedad 3 vista más arriba. Luego,

$$\frac{\delta^2}{\delta a^2} (\sum d_i^2) = \frac{\delta}{\delta a} [(-2) \sum (x_i - a)] = (-2) \sum (-1) = 2 \cdot n \cdot a$$

Como $(2 \cdot n \cdot a)$ es una expresión siempre positiva, cuando $a = \bar{x}$ hay un mínimo.

4.12.2 Mediana

La mediana de un grupo de datos se define como el valor que es superior al 50% de los mismos, e inferior del 50% restante. Si se ordenan los datos en forma creciente, la mediana debe quedar en el medio. Por ejemplo, la mediana del grupo (1 , 4 , 5 , 6 , 8) es el valor 5, pues deja dos valores por encima y dos por debajo. Cuando la cantidad total de datos es par, se toma como mediana a la semisuma de los dos valores centrales. Si se tiene (1 , 2 , 4 , 8 , 9 , 10), la mediana es el valor 6, que deja igual cantidad de valores a ambos lados y está en el centro de 4 y 8. Cuando los datos vienen agrupados en clases, se usa la fórmula de los percentiles para encontrar más fácilmente el valor de la mediana.

CUADRO 4.4 : Cálculo de la mediana.

Caso 1: Datos individuales. (Usando los datos para los no infartados del Cuadro 3.1).

El valor de la mediana es 19, porque deja la misma cantidad de datos a ambos lados. En este caso, donde hay una cantidad par (200), hay que calcular el promedio entre los dos valores centrales, y como ambos son 19, entonces la mediana adopta dicho valor.

Caso 2 : Datos agrupados en clases (usando los mismos datos, pero tomados del Cuadro 3.3).

La mediana cae dentro de la primer clase, donde los límites reales de clase son 0,5 – 20,5.

En este caso hay que aplicar la fórmula de la mediana para datos agrupados, que se verá en el punto siguiente: los fractiles donde el percentil 50 deja el 50% de los datos por encima. O sea, dicho percentil es lo mismo que la mediana.

Caso 3 : Hacer una estimación de cuánto ganan, en promedio, los empleados de una farmacia, con los siguientes datos:

Gerente (1)	\$ 2000	Si se calcula el promedio directo es: $\frac{2000 + 5 \cdot (400) + 350}{7} = \$ 621,5$
Vendedores (5)	\$ 400	
Repartidor (1)	\$ 350	

Se puede decir entonces que el sueldo promedio, o “ingreso per cápita” es de \$ 621,5.

Sin embargo, este valor no es muy representativo de la realidad, pues de los siete empleados sólo uno de ellos supera dicho monto. En cambio, si se usa la mediana $Md = \$ 400$; este valor se acerca mejor a la idea de “ingreso per cápita”. Esto muestra la conveniencia de usar la mediana, antes que la media aritmética, para estimar los ingresos promedios en un ámbito económico.

La mediana se emplea en lugar de la media aritmética cuando ésta se ve influenciada por valores extremos. Cuando en el grupo de datos obtenidos existen unos pocos valores muy altos o muy bajos, o ambos a la vez. La debilidad de la media aritmética, es esa sensibilidad a los valores extremos, que la corren mucho de la posición centralista que debería tener. Por ejemplo, sea el grupo de datos (1 , 2 , 4 , 5 , 6 , 6 , 7 , 7 , 9 , 9 , 10 , 174) , su valor medio es 20, mientras que su mediana es 6,5 , el cual da una mejor idea acerca del centro de los datos. El promedio está sesgado hacia arriba por el valor extremo de 174. Notar que si este se elimina, entonces la mediana pasa a ser 6 y el promedio pasa a ser también 6, algo más coherente con la noción de centro, o posición central de los datos.

4.12.3 Fractiles

Hay diferentes tipos de fractiles, pero en general son valores que dejan por debajo de él una cierta fracción de los datos ordenados en forma creciente y el resto por encima. Cuando la fracción es la mitad, se trata de la mediana. Los más comunes son:

- *Quartiles* : Las fracciones son cuartas partes del total. El primer cuartil Q_1 deja el 25% de los valores por debajo. El segundo cuartil es igual a la mediana y el tercero Q_3 deja el 75%.
- *Deciles* : Las fracciones son décimas partes del total. El primer decil D_1 deja el 10% de los valores por debajo y el resto por encima. El quinto decil es la mediana.
- *Percentiles* : Las fracciones son centésimas partes del total. Así, el percentil catorce (P_{14}) deja el 14% de los valores por debajo. El percentil cincuenta es la mediana, o el quinto decil.

El método para calcular los percentiles es análogo al de la mediana: primero se ordenan los datos en forma creciente, se calcula la fracción buscada y se lo ubica entre ellos. Pero si los datos vienen agrupados en clases, la cosa se complica y hay que emplear la fórmula siguiente:

$$F_x = LRI_i + (a / f_i) [N (x / 100) - f_{ac\ i-1}] \quad \text{donde}$$

F_x : Fractil x-avo LRI_i : límite real inferior de la clase i donde se encuentra el fractil
 a : ancho de clase f_i : frecuencia de la clase i
 N : total de datos $f_{ac\ i-1}$: frecuencia acumulada de la clase anterior a la del fractil

Cuadro 4.5: Obtención de diversos fractiles (con datos de la Tabla 3.3).

Caso 1 : Obtención de la mediana (segundo cuartil, quinto decil o cincuentavo percentil)

Límites reales	f_i	x'_i	$f_{ac\ i}$	
0,5 - 20,5	104	10,5	104	La mediana debe dejar 100 datos por debajo, como en la primer clase ya hay 104, significa que ahí se halla el valor buscado. Por lo tanto es : $F_{50} = 0,5 + (20 / 104) [200 (50/100) - 0]$ $F_{50} = 19,73 \text{ UI/I} = Q_2 = D_5 = P_{50}$
20,5 - 40,5	36	30,5	140	
40,5 - 60,5	28	50,5	168	
60,5 - 80,5	18	70,5	186	
80,5 - 100,5	10	90,5	196	
100,5- 120,5	<u>4</u>	110,5	200	
Σ	200			

Caso 2 : Obtención del tercer cuartil Q_3 .

El 75% del total de valores es 150. Luego el valor buscado está en la tercer clase .

$Q_3 = F_{75} = 40,5 + (20 / 28) [200 (75/100) - 140] = 47,64 \text{ UI/I}$

Caso 3 : Obtención del percentil cinco y del noventa y cinco : P_5 y P_{95} .

El 5% del total de valores es 10. Luego el valor buscado está en la primer clase.

$$P_5 = F_5 = 0,5 + (20/104) [200 (5 /100) - 0] = 2,42 \text{ UI/l}$$

El 95% del total de valores es 190. Luego el valor buscado está en la quinta clase.

$$P_{95} = F_{95} = 80,5 + (20/10) [200 (95/100) - 186] = 88,5 \text{ UI/l}$$

De esto último, se puede concluir que el 90% de los pacientes sin infarto, pero con diagnóstico presunto de infarto de miocardio, sacó una CPK comprendida entre 2,42 y 98,5 UI/l.

Caso 4 : Calcular la mediana para el Caso 5 del Cuadro 4.2.1 y los percentiles 5 y 95.

Mirando la columna de frecuencias acumuladas de este caso, se nota que la mediana cae en el segundo intervalo de clases. Siendo $N = 47$; $Fac_1 = 10$; $LRI_2 = 8$; $F_2 = 17$ y $a = 2,99$. Entonces,

$$P_{50} = F_{50} = 8 + (2,99/17) [(47/2) - 10] = 10,4 \text{ minutos para atender a un cliente.}$$

Notar que en este caso, la mediana se parece al valor medio 10,8 minutos calculado más arriba. Esto porque, al no haber valores extremos, como en el caso de los salarios del Caso 3 del Cuadro 4.4; los dos valores se pueden emplear indistintamente en los cálculos econométricos.

Análogamente, el percentil 5 cae en el primer intervalo.

$$P_5 = F_5 = 5 + (2,99/10) [(0,05 \cdot 47 - 0)] = 5,7 \text{ minutos}$$

Por su parte, el percentil 95 cae en el anteúltimo intervalo, muy cerca del LRS_4 .

$$P_{95} = F_{95} = 14 + (2,99/6) [0,95 \cdot 47 - 39] = 16,8 \text{ minutos}$$

Esto significa, que en el 90% de los casos, se tardará entre 5,7 y 16,9 minutos en atender a un cliente. Este dato, es de mayor utilidad para el farmacéutico que el tiempo promedio de atención (10,8). Le permitirá calcular mejor el número de vendedores que necesita a medida que vaya variando el número de clientes que entra a su farmacia, para mantener el mismo ritmo de atención.

4.12.4 Moda

Es el valor de la variable de mayor frecuencia. El más salidor, el que se repite más veces en el grupo de datos. Se lo usa en lugar de la media o mediana, cuando se desea señalar el valor más típico o común del total de observaciones efectuadas. En términos matemáticos es un máximo relativo. Es más ambiguo que la mediana, muy inestable y no permite tratamientos algebraicos sencillos. La moda no queda influenciada por valores extremos como la media y es el más sencillo de comprender. A su antítesis, se lo llama *antimoda*, o sea el valor menos frecuente. Cuando la distribución de datos presenta más de un máximo relativo, se la llama bimodal por ese motivo. Para los datos del Cuadro 4.4 anterior, la moda pertenece a la primer clase que tiene el valor máximo de frecuencia: 104; para definirlo se toma el centro de la clase, o sea, la marca de clase que en este caso es:

$$x'_1 = 10,5 \text{ UI/l} = \text{Moda}$$

4.13 Estadígrafos de dispersión

Se denominan así a los estadígrafos que cuantifican la separación de los datos entre sí, respecto de un punto de referencia central, como la media aritmética. Es una medida de la dispersión de los valores obtenidos al medir una variable clínica. Cuanto menor es la dispersión, tanto mayor será la precisión del sistema de medición. Por eso se usan a estos estadígrafos como índices de la calidad de las técnicas clínicas. Si los estadígrafos de posición vistos en la sección 4.3 se relacionan con el concepto de exactitud, los de dispersión se relacionan con la precisión de las técnicas de laboratorio. Aquellos con el error sistemático, estos con los errores casuales. Hay varias formas de cuantificar esta dispersión.

La más simple de todas es el *rango*, definido como la diferencia entre el valor máximo y mínimo del grupo de datos. De fácil cálculo y comprensión, tiene la desventaja de ser la medida más grosera de la dispersión. Dos grupos de datos, con muy distinta dispersión pueden llegar a tener rangos similares. Uno de ellos puede tener el 99% de los valores junto al mínimo y el otro el 99% junto al máximo, pero al tener extremos iguales, sus rangos resultarían iguales a pesar de ser tan disímiles intrínsecamente.

Una medida un poco más refinada estaría dada por los desvíos, o sea la diferencia entre cada valor y el promedio. La sumatoria de todos estos desvíos sería una medida de la dispersión, pero resulta ser nula porque las desviaciones se compensan, tal como se vio en las propiedades de la media. Ahora bien, si se toman en valor absoluto el problema de los signos se resuelve, pero aún así no es suficiente. Tal sumatoria tendría el defecto de ser mayor, cuantos más datos se tomen en cuenta, utilizando el mismo sistema de medición en condiciones similares. Por tal motivo, se define al *desvío medio absoluto* como la sumatoria del valor absoluto de todos los desvíos, dividida el número total de mediciones. Este índice es mucho mejor que el rango, pero tiene un defecto: es muy difícil de manejar algebraicamente.

Otra forma de eliminar el signo de las desviaciones es elevándolas al cuadrado. O sea, la sumatoria de los cuadrados de los desvíos, que es un mínimo, ponderada con el número total de mediciones. Esta es la idea de la *varianza* de un grupo de datos. Como las unidades de semejante índice quedarían al cuadrado, conviene tomar la raíz cuadrada para poder compararla con la media y otras medidas análogas. Esta es la idea del *desvío estándar* o desviación típica. También es la fórmula planteada por Gauss en su teoría de errores casuales. Como se verá más adelante, la curva de Gauss tiene dos puntos de inflexión simétricos ubicados a una distancia del centro igual al desvío estándar.

Una medida de posición y otra de dispersión, son suficientes para dar una idea de la forma en que se agrupan los datos. En particular, si se supone que los datos fueron tomados de una población “normal” o “gaussiana”, entonces la media y el desvío estándar son los valores necesarios y suficientes para definir la curva de Gauss involucrada en forma unívoca. Por regla general, los investigadores saben que mediciones repetidas de la misma magnitud, en condiciones análogas, tienen una distribución gaussiana. Separando a los pacientes por sexo y edad, las distribuciones de las magnitudes clínicas continuas en dichos segmentos suele ser gaussiana. Si bien algunos autores disienten en algunos aspectos, es indudable que separando a los pacientes sanos, agrupados por misma edad y sexo, con un número muy grande de casos, las distribuciones obtenidas se parecen mucho a la normal. Esto se discutirá más adelante con más detención.

4.13.1 Desvío estándar y Varianza

Se definen a estos estadígrafos con las fórmulas siguientes,

$$\text{Varianza poblacional: } \sigma^2 = \sum (d_i^2) / N = \sum (x_i - \mu)^2 / N$$

$$\text{Varianza muestral: } DS^2 = \sum (d_i^2) / (n - 1) = \sum (x_i - \bar{x})^2 / (n - 1)$$

Donde :

\bar{x} es la media de la muestra	n es el tamaño de la muestra
μ es la media de la población	N es el tamaño de la población
x_i es la i -ésima medición o valor	d_i desvío de un valor respecto de su media

A su vez, el *desvío estándar poblacional* (σ) y el *muestral* (DS) se obtienen con la raíz cuadrada de las respectivas varianzas. Para poder conocer o calcular tanto la media como la varianza poblacionales, se necesita conocer la población completa. Esto es imposible en el caso de mediciones repetidas porque estas son infinitas desde el punto de vista teórico. Análogamente, la cantidad total de muestras que se le puede extraer a un paciente, para hacerle una determinación en el laboratorio, es tan grande que puede ser considerada infinita. Por lo tanto, para todos los efectos prácticos se consideran a las mediciones de magnitudes clínicas continuas, como provenientes de una población de tamaño infinito. Es decir, que nunca se podrán conocer con certeza sus parámetros poblacionales μ y σ . Necesariamente, habrá que estimarlos a partir de sus respectivas medidas muestrales obtenidas con \bar{x} y DS .

Cuando los datos vienen agrupados en clases se emplea la fórmula:

$$DS^2 = \sum (x'_j - \bar{x})^2 f_j / (n - 1)$$

Donde f_j es la frecuencia de la clase j
 x'_j su respectiva marca de clase

Debe considerarse al desvío estándar como un estadígrafo y a la varianza como a otro. Aquí se desarrollan en conjunto para simplificar el texto, aprovechando la gran similitud de sus fórmulas. Pero, mientras el primero se puede comparar directamente con cualquiera de los valores medidos o de sus estadígrafos de posición, la varianza no. Así, la media y el desvío se emplearan para construir intervalos donde se espera encontrar adentro al verdadero valor de la población, llamado a veces *parámetro*. Eso se estudiará en la estadística inferencial. En cambio el *análisis de la varianza*, llamado ANOVA, es uno de los capítulos más importantes de la Estadística aplicada a experimentos. Se usa fundamentalmente para testear supuestos que se hacen acerca de poblaciones, como detectar si un medicamento sirve, o si un fertilizante es útil a la tierra, o si un alimento balanceado es mejor que otros y muchos usos más.

La fórmula del desvío estándar muestral dada más arriba, debe ser corregida por un factor presentado por Gurland y Tripathi. Si bien DS^2 es un estimador insesgado de σ^2 , no ocurre lo mismo con su raíz cuadrada DS que en forma consistente subestima a σ . El factor de corrección presentado por estos dos autores corrige ese defecto. Cuando $n \geq 30$ se calcula con la relación $G_n \approx 1 + 1 / [4(n-1)]$. Para los demás valores ver en Cuadro 4.6.

CUADRO 4.6: Cálculo de desvío estándar y varianza.

Caso 1: Datos individuales. (usando los datos para los no infartados del Cuadro 3.1).

$$DS^2 = \sum (x_i - \bar{x})^2 / (n-1) = \frac{(10 - 30,3)^2 + (33 - 30,3)^2 + \dots + (90 - 30,3)^2}{200 - 1}$$

$$DS = 27,77 \text{ UI/I}$$

Caso 2 : Datos agrupados en clases (usando los mismos datos, pero tomados del Cuadro 3.3)

CPK	f _j	x' _j	(x' _j - \bar{x}) ²	f _j .(x' _j - \bar{x}) ²	
1 - 20	104	10,5	392,04	40772,16	DS ² = 143256 / 199
21 - 40	36	30,5	0,04	1,44	
41 - 60	28	50,5	408,04	11425,12	DS ² = 719,88 (UI/I) ²
61 - 80	18	70,5	1616,04	29088,72	
81 - 100	10	90,5	3624,04	36240,40	DS = 26,83 UI/I
101- 120	4	110,5	6432,04	25728,16	
Σ	200			143256,00	

El valor exacto del desvío y de la varianza está dado por el primer caso, donde se tienen identificados los datos uno por uno. A partir de su agrupación en clases, los valores aproximados resultan ser un 3,5% menores, pues al perder información se pierde exactitud en el cálculo.

Caso 3 : Datos y sus frecuencias (valores tomados de un conteo celular - Student, 1907).

Nº de células x cuadrícula	0	1	2	3	4	5
Frecuencias Observadas	213	128	37	18	3	1

$$DS^2 = \{213.(0-0,68)^2 + 128.(1-0,68)^2 + 37.(2-0,68)^2 + 18.(3-0,68)^2 + 3.(4-0,68)^2 + (5-0,68)^2\} / 399$$

$$DS^2 = 324,68 / 399 = 0,8137 \quad \text{y el desvío estándar resulta} \quad DS = 0,902$$

En este caso las marcas de clase coinciden con la cantidad de células por cuadrícula de la cámara que tiene en total 400.

Caso 4 : Factor de corrección de Gurland y Tripathi (1971).

Este factor puede ser importante cuando el tamaño muestral es pequeño (n < 30).

$$G_n = [(n-1/2)^{1/2} \Gamma (n-1/2)] / \Gamma (n/2)$$

Y para n ≥ 30 se reduce a :

$$G_n \approx 1 + 1 / 4 (n-1)$$

Para el caso de Student, una mejor estimación de σ es el valor DS_{corregido}.

$$DS_{\text{corregido}} = G_n . DS = [1 + 1 / 4.(400-1)] . 0,902 = 0,9026$$

CUADRO 4.7: Propiedades de la varianza:

Sean n observaciones de una variable clínica cualquiera $(x_1, x_2, x_3, \dots, x_n)$, cuya media aritmética es: \bar{x} y su desvío estándar DS , entonces:

1. Si a cada valor medido se le suma (o resta) una constante a , la nueva varianza no se altera.

$$\text{Si } z_i = x_i + a \Rightarrow DS^2_z = \sum (z_i - \bar{z})^2 / (n-1) = \sum [(x_i + a) - (\bar{x} + a)]^2 / (n-1)$$

O sea, $DS^2_z = DS^2_x$

Esto significa que la presencia de un error sistemático en el sistema de medición no altera la varianza, es decir, el error casual no se afecta.

2. Si a cada valor medido se le multiplica (o divide) por una constante a , la nueva varianza es igual a la anterior, multiplicada (o dividida) por el cuadrado de dicha constante.

$$\text{Si } z_i = x_i \cdot a \Rightarrow DS^2_z = \sum (z_i - \bar{z})^2 / (n-1) = \sum [(x_i \cdot a) - (\bar{x} \cdot a)]^2 / (n-1)$$

O sea, $DS^2_z = a^2 \cdot DS^2_x$

3. La varianza de la suma (o resta) de dos variables se obtiene con :

Sea la variable $z = x + y$

$$DS^2_z = \frac{(n_x - 1) DS^2_x + (n_y - 1) DS^2_y}{(n_x + n_y - 2)} \Leftrightarrow \mu_x \approx \mu_y$$

Las propiedades más importantes de la varianza se dan en el Cuadro 4.7, recordando que al ser mínima la suma del cuadrado de los desvíos, como se vio para la media, entonces la varianza respecto de la media aritmética, será un mínimo. Es decir, la medida de la dispersión de los valores alrededor de la media usando la varianza, será menor que si se toma respecto de otra medida de posición cualquiera.

4.13.2 Coeficiente de Variación

Es una relación entre la media y el desvío estándar de una población, o de una muestra extraída de ella. Da una idea del error relativo que se comete en el sistema de medición. No es un estadígrafo de dispersión en un sentido absoluto, sino una relación entre dos de ellos: el de dispersión respecto al de posición. Una especie de coeficiente, que mide la dispersión en términos de la media. El CV no da una buena idea de la “bondad” o “calidad” de dos series de mediciones entre sí, sino una primera aproximación. Sin embargo, en la bibliografía clínica es común encontrarlo usado como índice para decidir entre dos métodos. Al respecto, se debe remarcar que si bien es condición necesaria, no es suficiente para optar entre técnicas clínicas; la forma correcta de hacerlo se verá más adelante en la sección dedicada a ello.

Coefficiente de variación poblacional $v = \sigma / \mu$

Coefficiente de variación muestral $CV = DS / \bar{x}$

Se suele expresar en forma porcentual (CV %). Mientras en la Física, un sistema de medición con coeficientes superiores al 10% es inaceptable, en Clínica, donde las variabilidades inherentes al material de trabajo son mucho más grandes, no se puede ser tan estricto. Por eso se sugiere usar :

10% < CV% < 50% poco aceptables

1% < CV% < 10% aceptable

CV% < 1% muy aceptable

CUADRO 4.8: Cálculo del CV%.

Caso 1: Datos individuales. (usando los datos para los no infartados del Cuadro 3.1).
Del Cuadro 4.6 anterior es $DS^2 = 771 (UI/l)^2$ y el $DS = 27,77 UI/l$ con una media $\bar{x} = 30,3 UI/l$. Por lo tanto, se puede obtener el CV% con la relación:
 $CV\% = (27,77 / 30,3) \cdot 100 = 91,65 \%$

Caso 2 : Datos y sus frecuencias (valores tomados de un conteo celular - Student, 1907).
 $DS = 0,902$ y $\bar{x} = 0,68$ Entonces: $CV\% = (0,902/0,68)100 = 132,6\%$

Caso 3 : El precio de 2 acciones comunes de un Laboratorio en la Bolsa de valores tuvo un promedio, durante un mes, de \$ 1500 para las del tipo A y de \$ 5000 para las B. Sus respectivos desvíos estándar fueron de \$ 500 y \$ 300 respectivamente. Comparando en valores absolutos, la variabilidad de las acciones tipo A fue mayor que las del B por tener mayor desvío estándar. Pero, con respecto al nivel de precios, hay que comparar ambos CV. Resulta, $CV_A = 500/1500=0,033$ y $CV_B = 300/5000= 0,06$. Se deduce que el precio de la acción del tipo B fue casi dos veces más variable que las del tipo A.

Sea el caso de un experimentador que mide tres longitudes diferentes, usando la misma regla milimetrada: A = 2 mm, B = 2 cm y C = 20 cm. Suponiendo que hace varias mediciones, siguiendo el mismo método en condiciones análogas; es de esperar que los valores promedios se acerquen a los valores reales y que el desvío estándar se mantenga constante en los tres casos, con un valor no mayor de 1 mm. Entonces:

Caso A $CV = 1 \text{ mm} / 2 \text{ mm} = 0,5$ es decir un $CV\% = 50\%$ (inaceptable)

Caso B $CV = 1 \text{ mm} / 20 \text{ mm} = 0,05$ es decir un $CV\% = 5 \%$ (aceptable)

Caso C $CV = 1 \text{ mm} / 200 \text{ mm} = 0,005$ es decir un $CV\% = 0,5\%$ (muy aceptable)

Análogamente en análisis clínicos, si la misma técnica se usa para medir concentraciones muy diferentes, puede ocurrir que el CV% varíe demasiado y ya no sea confiable como índice para la calidad de la técnica.

4.14 Problemas propuestos

1) Decidir si son Verdaderas o Falsas las siguientes afirmaciones:

- | | | |
|--|---|---|
| 1) Todos los índices diagnósticos se pueden deducir con sensibilidad y especificidad. | V | F |
| 2) Los índices clínicos forman parte de los estadígrafos de dispersión. | V | F |
| 3) Lo ideal para diagnosticar es tener mucha exactitud y precisión. | V | F |
| 4) Sensibilidad y especificidad varían con el punto de corte y con la prevalencia. | V | F |
| 5) La eficiencia varía con la prevalencia, lo mismo que el Índice de Youden | V | F |
| 6) Si aumenta la sensibilidad con los puntos de corte, también lo hace la especificidad. | V | F |
| 7) Punto de corte significa lo mismo que valor referente. | V | F |
| 8) Los pacientes Sanos y con resultado (+) son los verdaderos positivos. | V | F |
| 9) Las características: Sano, enfermo, (+) y (-) son una partición del conjunto pacientes. | V | F |
| 10) Donde se cortan las curvas de especificidad y sensibilidad, el Youden es máximo. | V | F |
| 11) Ídem anterior para la eficiencia. | V | F |
| 12) La eficiencia de una prueba clínica es el cociente entre sus aciertos y el total de casos. | V | F |
| 13) Si el VP de positivos aumenta, el VPN disminuye con los puntos de corte. | V | F |
| 14) El concepto de calidad puede ser interpretado como una dualidad | V | F |
| 15) Las enfermedades del Tipo I requieren el menor número posible de fn | V | F |
| 16) Las enfermedades del Tipo II requieren el menor número posible de fp | V | F |
| 17) Las enfermedades del Tipo III son las que se curan si se detectan a tiempo. | V | F |
| 18) A veces, la clasificación de una enfermedad depende del grado de avance de la misma | V | F |
| 19) Clasificar las enfermedades en tres tipos y explicar en que consiste cada tipo | | |
| 20) Los Likelihood Ratios varían con la prevalencia de la población. | V | F |
| 21) Para una enfermedad del tipo I hay que maximizar la sensibilidad | V | F |
| 22) Para una del tipo II la especificidad y del tipo III el Youden | V | F |
| 23) Clasificar los índices clínicos y explicar cuales varían con la prevalencia y cuales no..... | | |
| 24) Incidencia y prevalencia son lo mismo. | V | F |
| 25) Los índices de riesgo y daño son los mismo. | V | F |
| 26) El Odds Ratio es el cociente entre dos tipos de Odds, el de expuestos y no expuestos. | V | F |
| 27) Hay una relación matemática entre los Odds y l riesgo, y viceversa. | V | F |
| 28) El RR es la proporción de riesgo original presente en pacientes no expuestos. | V | F |
| 29) Los riesgos absolutos son la proporción de enfermos expuestos y de no expuestos. | V | F |
| 30) El riesgo básico es la proporción de enfermos en los expuestos. | V | F |
| 31) El CER es la proporción de enfermos no expuestos. | V | F |
| 32) El EER es la proporción de enfermos expuestos. | V | F |
| 33) El RR se calcula dividiendo EER por CER. | V | F |
| 34) La reducción del riesgo absoluto (ARR) es la diferencia entre CER y EER (sin el signo). | V | F |
| 35) La reducción del riesgo relativo (RRR) es la diferencia entre 1 y RR (sin el signo). | V | F |
| 36) El RRR también se puede calcular como el cociente entre ARR y CER. | V | F |
| 37) Explicar el concepto de fidelidad o reproducibilidad..... | | |
| 38) Explicar los conceptos de nivel de concordancia y Odds de discordancia | | |
| 39) La concordancia se usa para ver si se puede reemplazar un método por otro. | V | F |
| 40) El análisis de concordancia se hace para saber cual método clínico es el mejor. | V | F |
| 41) Se puede analiza la concordancia de dos factores, y cada uno con muchos subgrupos. | V | F |
| 42) El error de apreciación de una regla milimetrada es de 0,5 mm. | V | F |
| 43) El error casual no puede ser corregido por el investigador. | V | F |
| 44) Clasificar los índices clínicos y a los estadígrafos en general | | |
| 45) Los estadígrafos de tendencia central son: | | |
| 46) La suma de los desvíos es nula. | V | F |
| 47) La mediana es más inmune a valores extremos que la media aritmética. | V | F |

- | | | |
|---|-------|-------|
| 48) La suma de los cuadrados de los desvíos respecto a la media es un máximo. | V | F |
| 49) La media ponderada se usa en Física para calcular el centro de masa. | V | F |
| 50) Si a un grupo de datos se le suma una constante la varianza no cambia, pero la media sí | V | F |
| 51) El segundo cuartil, es lo mismo que el 50vo percentil y que la mediana. | V | F |
| 52) La moda es el valor de la variable que tiene mayor frecuencia. | V | F |
| 53) Si se multiplica a un grupo de datos por una constante, la varianza no se altera. | V | F |
| 54) Explicar las propiedades de la media aritmética..... | | |
| 55) Explicar las propiedades de la varianza..... | | |
| 56) El valor de la mediana coincide con uno de los datos cuando el número total es impar. | V | F |
| 57) El coeficiente de variación puede medir la calidad de un test clínico. | V | F |
| 58) El coeficiente de variación muestral es lo mismo que el poblacional. | V | F |
| 59) El rango es un índice muy pobre en comparación a la varianza. | V | F |
| 60) Explicar el concepto de fractiles | | |

2) Para los siguientes datos calcular:

	(+)	(-)	Sensibilidad =	Especificidad =
			Eficiencia =	VPN =
Sano	50	100	Índice de Youden =	Likelihood ratio (+) =
Enfermo	120	30	VPP =	Likelihood ratio (-) =

¿ En cual tipo de enfermedad conviene usar este método clínico ?

3) Para los siguientes datos calcular:

	Protegido	No-protegido	OR =	CER =
Sano	46	36	RR =	ERR =
Enfermo	18	29	ARR =	RRR =

4) Para los siguientes datos calcular:

	(+)	(-)	
(+)	400	30	Nivel de concordancia =
(-)	10	200	Odds de discordancia =

5) Para los datos siguientes, encontrar el coeficiente de variación la media, mediana y el desvío:

Frecuencia	10	20	30	15	8
Valor	2	4	6	8	10

6) Los salarios semanales de los operarios de una industria medicinal son:

Salario (\$)	50-59,9	60-69,9	70-79,9	80-89,9	90-99,9	100-109,9	110-119,9
Nº operarios	8	10	16	14	10	5	2

- Hallar: a) Los límites reales de cada clase y sus frecuencias acumuladas (ancho = \$10).
 b) La marca de cada clase.
 c) La media, mediana y moda.
 d) El desvío estándar, la varianza y el Coeficiente de variación.
 e) El percentil 5 y el 95.
 f) Representar con un histograma y un polígono de frecuencias acumuladas, para calcular en forma gráfica, todos los valores anteriormente pedidos.

7) Se lanza una moneda al aire 1000 series, de 5 veces cada una, para medir el número de caras. Los resultados obtenidos se muestran en la tabla siguiente:

Número de caras (X)	0	1	2	3	4	5
Número de series (frecuencia)	38	144	342	287	164	25

Se pide: a) Graficar los datos con un diagrama de bastones y con un histograma.
b) Calcular los índices de tendencia central.
c) Calcular los índices de dispersión.
d) Dibujar el polígono de frecuencias acumuladas porcentual.

8) Para hacer un test de Sífilis se emplea la técnica 1 en 400 pacientes y se encuentran 220 pacientes que dieron un resultado (+). Los aciertos en (+) fueron 180 y 160 en (-). Con estos datos calcular la Sensibilidad y Especificidad de esta técnica. A las mismas pacientes se les aplicó la técnica 2 y resultaron 240 pacientes con (+), mientras el número de aciertos en (+) fue de 190 y en (-) 150. Decidir de manera cualitativa cual de ambas técnica es la mejor para adoptar en nuestro Laboratorio y explicar el porqué.

9) Se midieron 50 alícuotas de un suero con el Kit de Glucosa marca A, y otros 50 con la marca B. Los resultados encontrados fueron:

Marca A : Entre (80-80,9) 8 veces, (81-81,9) 22 veces, (82-82,9) 15 veces y el resto (83-83,9)

Marca B : Entre (80-80,9) 6 veces, (81-81,9) 30 veces, (82-82,9) 10 veces y el resto (83-83,9)

Decidir en forma cualitativa, si hay diferencia en la exactitud y en la precisión entre las marcas. Si además se sabe que el suero utilizado tenía una concentración de 83 hallar en forma cualitativa el error sistemático de cada marca. Estudiar la precisión si en la bibliografía se dice aceptable la dispersión de la glucosa cuando llega hasta un 8%.

10) Se sabe que la concentración de coloides en el agua de un río es del 5,4 /ml. Se tomó una muestra del mismo a la que se le agregó un reactivo. Luego de un rato, se extrajo una pequeña cantidad para analizarla en una cámara de recuento. Los resultados fueron:

Nº de coloides:	2	3	4	5	6	7 (Nº /ml)
Frecuencia:	10	20	170	100	90	10

Decidir en forma cualitativa si hubo algún cambio significativo en la concentración del agua, por la acción del reactivo (Sulfato de Aluminio).

11) Los clientes de una farmacia gastaron en remedios y son otros artículos los datos que se muestran en el cuadro siguiente expresados en \$:

Remedios: Entre (10-14,9) 5 veces, (15-15,9) 25 veces, (20-25,9) 10 veces.

Otros: Entre (10-14,9) 3 veces, (15-15,9) 20 veces, (20-25,9) 12 veces, (26-30) 5 veces.

Decidir en forma cualitativa si hay diferencia entre los promedios de ventas. Determinar el desvío estándar de cada grupo y su moda.

5

Probabilidad

En este capítulo se introduce el concepto de la probabilidad, necesaria para la comprensión de temas a desarrollar en capítulos posteriores. Se comienza con el enfoque probabilístico clásico, se sigue con desarrollos algebraicos, para llegar a los teoremas básicos, vinculantes de los índices clínicos entre sí. La mayoría de los conceptos científicos tienen un significado exacto como las cantidades físicas; en cambio, la probabilidad es a menudo asociada con un concepto vago de casualidad, incertidumbre o azar. Muchas veces esta ignorancia oculta la verdadera naturaleza matemática de la teoría de las probabilidades, como una disciplina científica más. Una ciencia exacta con conclusiones obtenidas por la lógica, partiendo de axiomas y principios básicos y con muchas ramificaciones de uso extendido en la práctica. Sin embargo debe dejarse muy en claro que la incertidumbre, es una palabra usada para disimular la ignorancia que se tiene acerca de las múltiples causas que pueden ocasionar un efecto. Y que el azar, o la probabilidad es una forma de cuantificar dicha incertidumbre.

5.1 Introducción

J. J. Bernoulli fue el primero en estudiar este tema en forma sistemática con un enfoque científico. Observando los resultados del lanzamiento de una moneda un número grande de veces, notó que el número de caras y secas tendía a igualarse. Es decir, que la frecuencia relativa de la obtención de caras se acercaba más al número de secas, cuanto mayor era el número de lanzamientos. O bien, ambas frecuencias relativas se parecían cada vez más a 0,5. Otro tanto le ocurría en el lanzamiento de dados: la frecuencia relativa de un as tendía a 1/6. Repitió una y otra vez este tipo de experimentos con monedas, dados y cartas, y siempre llegaba a la misma conclusión. Imaginó haber encontrado un fenómeno más general y así dio comienzo la teoría de probabilidades. Sus resultados teóricos se corresponden razonablemente con la realidad. Sin embargo, debe marcarse siempre una clara distinción entre los resultados empíricos y los teóricos. El uso comenzó en la teoría de juegos de azar, en los albores del siglo XVII, y gracias a estos se hizo popular entre los “geómetras” de aquel entonces. Hoy se la emplea en el campo de los seguros, control de calidad, genética, investigación operativa, mecánica estadística y muchos más.

Probabilidad teórica: La probabilidad de ocurrencia de un suceso A se define como el cociente entre el número esperado de veces que ocurra un suceso N_A y el número total de casos posibles N

$$P(A) = N_A / N$$

Esta definición primaria de probabilidad se calcula en forma teórica. Por ejemplo, una moneda tiene dos caras ($N = 2$) y una sola de ellas es cara ($N \text{ cara} = 1$), o sea que la probabilidad de sacar una cara en un lanzamiento, teóricamente es $P(\text{sacar cara}) = 1/2$. Por otra parte, se puede calcular la probabilidad a partir de experimentos, como por ejemplo tirar la moneda para ver si en efecto hay una proporción 0,5 de obtener cara. Esta noción se define como:

Probabilidad empírica: La probabilidad empírica de ocurrencia de un suceso A es igual a su frecuencia relativa (Fr_A). O sea, el cociente entre el número de veces en que ocurrió el suceso A (F_A) y el número total de experimentos (F_t).

$$Fr_A = F_A / F_t$$

La relación entre ambas, fue descripta por Bernoulli en su “Ley empírica del azar” de su famoso libro *Ars Conjectandi* publicado en 1713 donde se dice: “Si la probabilidad de un suceso es p y se realiza un número muy grande N de pruebas, el cociente entre el número de casos favorables obtenidos y el número total de pruebas llega a diferir de p en tan poco como se quiera, con tal de tomar a N lo suficientemente grande.” Esto es:

$$Fr_A \longrightarrow p \quad (\text{si } N \text{ es grande})$$

Sin embargo, fue Laplace en su “Teoría Analítica de Probabilidades” (1812) quien le dio una forma más rigurosa definiendo la diferencia en valor absoluto entre ambas probabilidades como menor o igual a un cierto valor positivo ϵ , haciendo a este tan pequeño como se quiera.

$$| Fr_A - P(A) | \leq \epsilon$$

o sea:

$$\begin{array}{ccc} Fr_A & \longrightarrow & P(A) \\ N & \longrightarrow & \infty \end{array}$$

Naturalmente, como un experimento real nunca puede llegar a tener lugar infinitas veces, no se usa la definición clásica de límites sino una aproximación del tipo

$$Fr_A \approx P(A) \quad (\text{si } N \text{ es lo suficientemente grande})$$

Laplace introdujo dos conceptos nuevos en la definición de la probabilidad teórica:

Principio de equiprobabilidad: todos los casos deben tener igual probabilidad de ocurrencia.

Principio de razón suficiente: mientras nada haga sospechar lo contrario, se debe suponer que todos los casos son igualmente probables.

Agregando estos dos principios se completa la definición clásica de probabilidad. En los casos de un dado cargado, una moneda mal balanceada, y otros análogos, no se puede aplicar la definición porque los casos ya no tienen la misma probabilidad. Hay otros casos: por ejemplo, suponiendo un dado tal que tiene dos ases pintados en sus caras pues le falta el número 4, allí habría cinco casos posibles en lugar de seis. Pero si el dado está bien construido la definición se sigue cum-

pliendo, pues en tal dado todavía hay seis caras con igual probabilidad de ocurrencia y la probabilidad de sacar un as sería $2/6$ (2 caras que tienen un as, sobre 6 caras posibles). Mientras que la de sacar 4 sería nula y las demás de $1/6$.

CUADRO 5.1 Ejemplos de aplicación

1) De la población de pacientes de un laboratorio se eligieron $N = 1000$ personas al azar, y se encontraron que 38 de ellas padecían de hipoglucemia. Calcular la probabilidad de que si se escoge un paciente al azar este padezca la enfermedad.

Frecuencia empírica = $TE / N = 0,038 \rightarrow$ probabilidad teórica = prevalencia

2) Al realizar una comprobación de los valores predichos por la CPK, se encontró que 64 de ellos fueron mal clasificados como positivos y 36 como falsos negativos, del total de 400 historias clínicas analizadas. Calcular la probabilidad de acertar en el pronóstico.

Frecuencia empírica de aciertos = $(vp + vn) / N = (400 - 64 - 36) / 400 = 0,75 \rightarrow$ eficiencia

3) En un curso de Bioestadística hay 30 alumnos, 10 de la carrera de Bioquímica, 15 de Farmacia y 5 de Enfermería. El 70% del total son mujeres y el resto varones. Se escoge uno al azar; hallar la probabilidad de que sea de la carrera de:

- a) Bioquímica $P(B) = 10 / 30 = 1 / 3 = 0,333$
- b) Farmacia $P(F) = 15 / 30 = 1 / 2 = 0,5$
- c) Enfermería $P(E) = 5 / 30 = 1 / 6 = 0,167$
- d) ¿Cuál será la cantidad de mujeres M ?

$P(\text{Mujer}) = 0,7 = M / N = M / 30$ Por lo tanto $M = 0,7 \cdot 30 = 21$ mujeres

2) En una farmacia ingresan en promedio 60 clientes por día y hacen compras por \$ 420. Si se sabe que el 30% se van sin comprar nada. ¿Cuánto vale el gasto per cápita promedio?

La probabilidad de que un cliente compre es del 70%. O sea $0,7 = \text{compradores} / \text{total}$. Luego el número de compradores diarios en promedio es: $42 \text{ clientes} = 0,7 \cdot 60$, entonces si el ingreso diario es de \$ 420, el gasto per cápita será: $\$ 420 / 42 \text{ clientes} = \10 por cliente.

Aparecieron más enfoques: Von Mises propuso la noción de “colectivo” para evitar el defecto de la noción clásica: ¿cómo se puede estar seguro en la realidad que todos los casos son equiprobables? Muchos piensan que el principio de razón suficiente es una forma de cerrar los ojos, porque lo correcto sería hacer un experimento para asegurarse de la equiprobabilidad. Este autor propuso considerar una sucesión infinita de sucesos: $S_1, S_2 \dots$ dentro de los cuales se verifica cierto atributo A , tomando el cociente entre el número de veces que se da A y el total, se tiene un valor que tiende a $P(A)$ dentro de un *colectivo* de sucesos. Esta propuesta no tuvo demasiado éxito, comparada con la de Kolgomoroff (1933), que planteó un acercamiento *axiomático* a la noción abstracta de probabilidad y el desarrollo de la teoría como una disciplina matemática basada en la teoría de mediciones.

Hay otro tipo de probabilidades como la llamada *probabilidad subjetiva o intuitiva* que es una cierta evaluación personal de la probabilidad, en lugar de ser teórica o experimental. Por ejemplo, si se desea opinar acerca de la probabilidad de lluvia el día de mañana, cada uno puede emitir una opinión, como 0,9 si hoy está nublado y no aparece el sol, o bien 0,1 si se está en el medio de una magnífica temporada estival. En Medicina, el uso de esta clase de probabilidades se hace cuando el clínico asigna un valor subjetivo de probabilidad a un paciente que está evaluando, la llamada probabilidad Pre-test, y mediante un cálculo obtiene la probabilidad Post-test para evaluar si le conviene hacerle el test al paciente o no. Notar que la misma pregunta puede ser contestada de manera diferente por una misma persona, de acuerdo al momento en que se le haga. Además, diferentes individuos darán diferentes valores subjetivos de probabilidad. Las encuestas de opinión calculan la frecuencia de las respuestas obtenidas, a preguntas subjetivas, para conformar técnicas de mercadeo, propaganda, etc.

Hasta ahora, se han planteado casos con N de tamaño finito, pero si se trabaja con las probabilidades de encontrar un punto en una recta donde cualquier segmento de la misma tendrá infinitos puntos posibles, las definiciones vistas ya no sirven. Si por ejemplo el suceso A (unos pocos centímetros) ocurre dentro de una recta de 3 metros, el cociente dado por la fórmula clásica sería una razón irresoluble de infinitos. Por eso se desarrolló el concepto de la *probabilidad geométrica*, como la razón entre el espacio o tiempo, ocupado por los sucesos A , y el total. Como ser, si se sabe que el suceso dura 20 s, la probabilidad que ocurra en la hora siguiente se puede obtener con: $20 \text{ s} / 3600 \text{ s} = 1/180$.

En síntesis, la noción de probabilidad tiene cuatro acepciones básicas:

- 1) *Clásica* (teórica con casos equiprobables)
- 2) *Empírica* (práctica con la frecuencia relativa)
- 3) *Intuitiva* (creencias o mediciones subjetivas)
- 4) *Axiomática* (modelo matemático)

5.2 Concepto de partición.

Para desarrollar un modelo matemático conviene empezar con la teoría de conjuntos, repasar los conceptos básicos, y luego formular axiomas que permitan definir la noción de probabilidad en forma más precisa. Con los axiomas y ciertas reglas de comportamiento interno se puede definir un álgebra, dentro de un espacio de probabilidades (Álgebra de Boole). A partir de allí se deducen ciertas propiedades interesantes para poder demostrar teoremas como el de Bayes, de uso extendido en el campo del diagnóstico clínico. En cambio, cuando los resultados posibles no son finitos se requerirá de un álgebra especial (σ -álgebra), pero este caso escapa a los contenidos planeados para el presente trabajo.

Las nociones básicas de la teoría de conjuntos fueron vistas en asignaturas anteriores dentro de las matemáticas. Sin embargo, para quien desee repasar estos conceptos se le sugiere ver el Apéndice 1 al final de este capítulo. Análogamente en el Apéndice 2, para repasar la teoría del cálculo combinatorio que se necesita para la obtención del número total de casos, o bien para calcular el número de casos favorables de ocurrencia del suceso A . Lo que interesa por ahora es recordar como se llega al concepto de *partición*, cuyo uso se vio en las Tablas 1.1 a 1.3

Sean n conjuntos $C_1, C_2, C_3, \dots, C_n$ pertenecientes al universo \varnothing , se dice que:

1. La unión de todos ellos pertenece al universo y en el caso particular que resulten ser iguales, se los llama colectivamente exhaustivos.

$$\bigcup_{i=1}^n C_i \subseteq \varnothing$$

Por ejemplo, si cada marca comercial que se vende en una farmacia es un conjunto C_i , entonces el conjunto de todas las marcas existentes en la misma será el conjunto universo, detallado en el listado completo del *stock* actualizado.

2. Son mutuamente excluyentes, cuando se excluyen entre sí todos los pares posibles, o sea

$$C_i \cap C_j = \emptyset \quad \forall i \neq j$$

El conjunto de medicamentos fabricado por los Laboratorios Roché: C_i , no tiene ningún elemento en común con el conjunto de los fabricados por Bayer: C_j . Por lo que son excluyentes entre sí.

3. Forman una partición del universo cuando estos conjuntos son a la vez: mutuamente excluyentes y colectivamente exhaustivos. Esto se puede denotar así: los C_i particionan a \varnothing si se cumplen las dos condiciones siguientes:

$$\left. \begin{array}{l} C_i \cap C_j = \emptyset \quad \forall i \neq j \\ \bigcup_{i=1}^n C_i = \varnothing \end{array} \right\} \text{Partición de } \varnothing$$

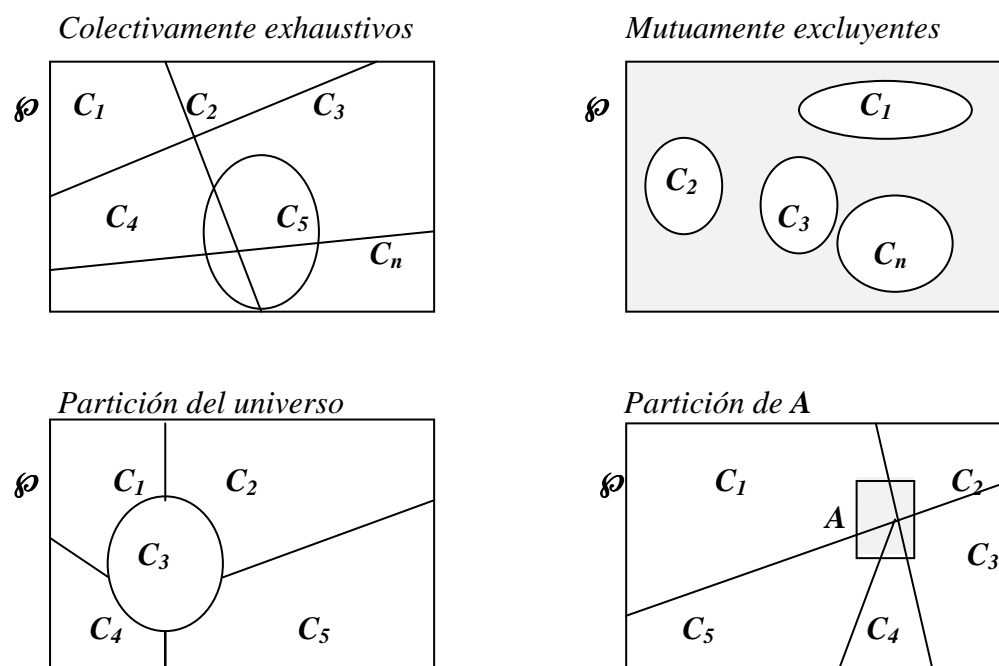
Si se clasifica a la población en: Sanos y Enfermos de una cierta dolencia, se tiene una partición de la misma, pues cada individuo de esta puede estar sano o enfermo únicamente y, además, la unión de sanos y enfermos conforman todos los casos posibles. Lo mismo ocurre si se clasifica a la población en Positivos y Negativos. Así los cuatro casos posibles en la Tabla de la verdad conforman una partición del universo de pacientes.

4. Particionan al conjunto A cuando a la vez son mutuamente excluyentes y colectivamente exhaustivos, dentro de A . Esto es, los C_i particionan a A si se cumplen:

$$\left. \begin{array}{l} C_i \cap C_j \cap A = \emptyset \quad \forall i \neq j \\ \bigcup_{y=1}^n C_i \supset A \end{array} \right\} \text{Partición de } A$$

Si se toma el subconjunto: enfermos de gripe de la población de Posadas. Este se puede particionar con la condición de: Medicado – No medicado, pues ambos casos son mutuamente excluyentes entre sí y colectivamente exhaustivos para el conjunto de engripados. El conjunto de resultados positivos está particionado por los dos conjuntos: Sano y Enfermo, lo mismo que el conjunto de resultados negativos. Estos cuatro casos se visualizan mejor en el Gráfico 5.1 siguiente:

Gráfico 5.1: Diagramas de Venn para el concepto de partición



5.3 Modelo Axiomático

Para desarrollar un modelo axiomático de probabilidad conviene definir primero una serie de conceptos, para poder usar un idioma común.

. *Elemento:* es uno de los resultados posibles de la prueba clínica o experimento. A veces se lo llama dato o resultado. Tiene el mismo significado que el visto en teoría de conjuntos.

. *Suceso:* es un conjunto de resultados obtenidos en una serie de pruebas. Se denomina también acontecimiento, evento o muestra (en el sentido estadístico del término).

- . *Suceso imposible:* cuando se sabe que no puede ocurrir, se lo denota con \emptyset .
- . *Suceso seguro:* cuando se tiene la certeza que ocurrirá, se lo denota con S .
- . *Suceso elemental:* está compuesto por un solo elemento, o sea, una muestra simple.

. *Tamaño:* el tamaño de un suceso es la cantidad de sus elementos componentes. Equivale al concepto ya visto de tamaño muestral.

. *Espacio Muestral:* está formado por el conjunto de todos los resultados posibles del experimento. Se lo llama también: espacio de probabilidad. Se corresponde con el suceso seguro.

La tarea fundamental a la que se dedica el bioquímico es efectuar análisis de laboratorio para ayudar a efectuar un diagnóstico correcto en los pacientes. Sea el caso de una enfermedad cualquiera, como ser el SIDA, empleando diversas técnicas de laboratorio, junto con revisiones corporales del paciente, se puede llegar a uno de dos diagnósticos:

A: luego de efectuarle todos los estudios se concluye que el paciente está infectado;

B: luego de efectuarle todos los estudios se concluye que el paciente no está infectado.

Esos dos serían los resultados posibles o *sucesos*, no importa el camino seguido para realizar tal diagnóstico. Habrá un conjunto de análisis diferentes, en la batería de análisis, los resultados finales serán **A** o **B**. Es evidente que si se hace el estudio y se observa paso a paso el camino seguido, se podrá determinar la serie de evidencias que condujeron al diagnóstico final. O sea, se podrá verificar uno de los dos sucesos. Ahora bien, es imposible que ocurran ambos sucesos a la vez, esto es **A** y **B** debe ser igual al suceso **0**. Pero es seguro que ocurrirá uno de ambos, entonces **A** o **B** coincide con **S**. Por su parte, el complemento de **A** es **B** y viceversa. Análogamente, un cliente que entra a la farmacia puede hacer dos cosas: comprar algo (**A**) o no comprar (**B**). En resumen:

Un conjunto de sucesos observables que presenta una estructura de Álgebra de Boole para las leyes de composición interna: conjunción, intersección y complemento, se la denomina Álgebra de Sucesos.

A su vez, un Álgebra de Sucesos siempre puede ser representada por un Álgebra de Conjuntos que es isomorfa a ella. En particular, tomando el conjunto de los números reales \mathfrak{R} , se puede definir una aplicación **P** del Álgebra de Sucesos Ψ sobre tal conjunto, $\mathbf{P}: \Psi \rightarrow \mathfrak{R}$ como una *probabilidad* si cumple los tres axiomas siguientes:

1. La probabilidad de que ocurra un suceso **A** es siempre positiva o nula

$$P(A) \geq 0$$

2. La probabilidad de que ocurra el suceso seguro es una certeza y vale 1

$$P(S) = 1$$

3. Si dos sucesos son excluyentes, la probabilidad de su unión es igual a la suma de sus probabilidades individuales

$$P(A \cup B) = P(A) + P(B) \Leftrightarrow A \cap B = 0$$

*Definición: sea un conjunto de sucesos que definen un espacio muestral **S**; la probabilidad de ocurrencia de un suceso **A** perteneciente al mismo es un número adimensional que cumple con los tres axiomas anteriores.*

Debe destacarse que no se ha hecho mención en la definición de la probabilidad axiomática a la manera en que se obtendrá la cantidad $P(A)$. Se puede usar una forma intuitiva de valorarla, o realizar un experimento para medir la empírica, o bien, hacer deducciones teóricas para su cálculo. Todas las maneras son igualmente válidas para este enfoque algebraico. La mejor aproximación a la realidad que se obtenga dependerá del modo de calcularla, y no del modelo matemático usado. Mediante este modelo axiomático o algebraico se pueden traducir a símbolos las relaciones entre acontecimientos para ver la aplicabilidad de los conceptos teóricos en la vida real. La idea central es que todo modelo estadístico debe ceñirse a la realidad lo más posible y no

a la inversa. El peligro fue advertido por Einstein cuando afirmó "... los científicos suelen ver la realidad a través de los anteojos que les pone la teoría en la que creen...".

Para analizar la consistencia entre probabilidad y frecuencia relativa se pueden ver:

. Cuando se tenga la certeza que ocurrirá un resultado al hacer una prueba, entonces en todos los casos se verificará la ocurrencia del suceso en cuestión. Por lo tanto, la frecuencia de casos favorables coincidirá con la frecuencia total, y así la frecuencia relativa será la unidad.

$$Fr_A = F_A / Ft = 1 \approx P(A) = P(S)$$

. Cuando se tenga la certeza que nunca ocurrirá un resultado al hacer una prueba, entonces el número de casos favorables será nulo, y así anulará la frecuencia relativa.

$$Fr_A = F_A / Ft = 0 \approx P(A) = P(\emptyset)$$

De los dos puntos anteriores se deducen de inmediato los dos valores extremos de frecuencia relativa o probabilidad.

$$0 \leq Fr_A \approx P(A) \leq 1$$

Cuando dos sucesos se excluyan mutuamente, la frecuencia relativa de ambos ocurriendo simultáneamente será nula ($F_{AB} = 0$), y por lo tanto:

$$Fr_{A \circ B} = (F_A + F_B + F_{AB}) / Ft = (F_A / Ft) + (F_B / Ft) = Fr_A + Fr_B \approx P(A \cup B)$$

Con los tres puntos anteriores se puede denotar la *consistencia* entre el modelo teórico de la probabilidad axiomática y las probabilidades empíricas relacionadas. El símbolo \approx fue usado en el sentido de la ley empírica de Bernoulli.

5.3.1 Propiedades derivadas

Al desarrollar los tres axiomas anteriores, deduciendo propiedades a partir de ellos, se va armando la teoría de probabilidades.

1) *Inclusión de sucesos entre sí:*

$$\text{si } A \subset B \text{ entonces } P(A) \leq P(B)$$

Lo que se demuestra con $B = A \cup (B \nabla A)$ y como estos dos conjuntos no tienen elementos comunes, se les puede aplicar el axioma 3, así: $P(B) = P(A) + P(B \nabla A)$. Y como toda probabilidad es siempre positiva o nula, se deduce que la $P(B)$ es siempre mayor, o a lo sumo igual que la $P(A)$. Por ejemplo, el conjunto A (10 personas) de los internados por fracturas en un pabellón del hospital, es un subconjunto (B) de todos los internados que hay en el mismo (20 personas), y por lo tanto $P(A) \leq P(B)$. Si en el hospital hay 100 internados en total se pueden estimar: $P(A) = 10/100 = 0,1 \leq P(B) = 20/100 = 0,2$ con lo que se muestra esta propiedad de inclusión que tienen los subconjuntos.

2) Unión de varios sucesos excluyentes:

$$P\left(\bigcup_{i=1}^n C_i\right) = P(C_1) + P(C_2) + P(C_3) + \dots + P(C_n) \Leftrightarrow C_i \cap C_j = \emptyset \quad \forall i \neq j$$

Para demostrarlo basta hacer substituciones, se toman los n-1 conjuntos $C_2 \dots C_n$ y a su unión se la llama B , así quedan dos conjuntos el C_1 y el B a los que se le puede aplicar el axioma 3, resultando la suma de sus respectivas probabilidades. Luego a B se le descompone en C_2 y el resto R , y se hace lo mismo. Así queda la suma de las probabilidades de los dos primeros conjuntos más la de la unión del resto. Procediendo análogamente se llega a demostrar la propiedad anterior. Por ejemplo, sean los cuatro sucesos posibles al efectuar un diagnóstico:

VP: verdadero positivo **FP**: falso positivo **VN**: verdadero negativo **FN**: falso negativo

Como los sucesos son mutuamente excluyentes entre sí, su unión equivale al universo, entonces:

$$P\left(\bigcup C_i\right) = P(\text{VP}) + P(\text{FP}) + P(\text{VN}) + P(\text{FN}) = P(S) = 1$$

3) Sucesos complementarios:

$$P(A) = 1 - P(\bar{A})$$

Sean dos sucesos complementarios A y \bar{A} , su unión es el universo y su intersección es el conjunto vacío, o sea, son excluyentes entre sí, por lo que se le puede aplicar los axiomas 1 y 3

$$P(A \cup \bar{A}) = P(S) = 1 = P(A) + P(\bar{A})$$

De donde se deduce de inmediato la propiedad anterior. Por ejemplo, la probabilidad de estar sano será igual a uno menos la prevalencia, y la probabilidad de resultar positivo será igual a uno menos la probabilidad de resultar negativo. Por ejemplo, la prevalencia de una enfermedad en la población se calcula como el cociente entre el número de enfermos y el total de individuos que componen la población: $p = TE / N$. El complemento de esta probabilidad será $q = 1 - p$, la cual se calcula con: $q = 1 - (TE / N) = (N - TE) / N = TS / N$. O sea la probabilidad de estar sano.

En el caso de complementos relativos hay otra propiedad expresada por:

$$P(B \setminus A) = P(B \cap \bar{A}) = P(B) - P(B \cap A)$$

De la figura se puede ver que el conjunto B está formado por dos conjuntos excluyentes, tales que:

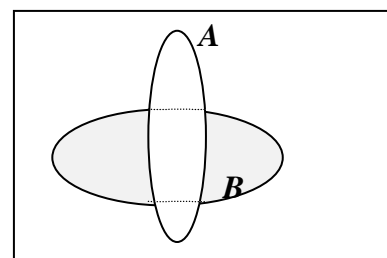
Gráfico 5.2

$$(B \cap \bar{A}) \cap (B \cap A) = \emptyset$$

El complemento relativo y la intersección a su vez son

$$(B \cap \bar{A}) \cup (B \cap A) = B$$

Se aplica el axioma 3 y resulta



$$B \cap \bar{A} \quad \square$$

$$P(B \cap \bar{A}) + P(B \cap A) = P(B) \quad \text{luego es : } P(B \cap \bar{A}) = P(B) - P(B \cap A)$$

Por ejemplo, sea A el conjunto de todas las embarazadas que concurren al hospital, luego su complemento será el conjunto de todas las mujeres que no están embarazadas que concurren al hospital. Sea B el conjunto de todas las personas que vienen al hospital a tratarse de enfermedades venéreas. Entonces, el complemento relativo $P(B \setminus A) = P(B \cap \bar{A})$ es la probabilidad de encontrar un mujer que concurra por una enfermedad venérea y que no esté embarazada.

4) *Unión de dos sucesos cualesquiera:*

Sean dos sucesos A y B , pertenecientes al espacio de probabilidades S ; la probabilidad de la unión de ambos es la suma de sus probabilidades individuales, menos la de su intersección.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Esto es una generalización del axioma 3, al caso general, cuando los conjuntos no son excluyentes, como en la figura del Gráfico 5.2 anterior. Puede verse que el suceso $(A \cup B)$ es la unión de tres sucesos mutuamente excluyentes entre sí, a saber, la intersección entre ambos y los dos complementos relativos. O sea,

$$(A \cap B) \cap (B \cap \bar{A}) \cap (A \cap \bar{B}) = \emptyset$$

$$(A \cap B) \cup (B \cap \bar{A}) \cup (A \cap \bar{B}) = (A \cup B)$$

Aplicado el axioma 3 generalizado, resulta:

$$P(A \cap B) + P(B \cap \bar{A}) + P(A \cap \bar{B}) = P(A \cup B) \quad \text{Reemplazando resulta:}$$

$$P(A \cap B) + \{ P(B) - P(A \cap B) \} + \{ P(A) - P(A \cap B) \} = P(A \cup B) \quad \text{O sea,}$$

$$P(B) + P(A) - P(A \cap B) = P(A \cup B)$$

Usando el mismo ejemplo anterior sería: $P(A \cup B)$ la probabilidad del conjunto formado por mujeres embarazadas o que tengan enfermedades venéreas, la que será igual a la suma de la probabilidad de que una mujer está embarazada $P(A)$, más la probabilidad de que una mujer tenga una enfermedad venérea $P(B)$, menos la probabilidad del conjunto formado por las mujeres embarazadas y con enfermedades venéreas $P(A \cap B)$.

5) *Particiones de sucesos:*

Sean n sucesos pertenecientes al mismo espacio muestral S , tales que particionan al suceso A (ver Gráfico 5.2), entonces se cumple que:

$$P(A) = \sum_{i=1}^n P(A \cap C_i) \quad \Leftrightarrow \begin{cases} (C_i \cap C_j) \cap A = \emptyset \quad \forall \quad i \neq j \\ \bigcup_{i=1}^n C_i \supset A \end{cases}$$

Si se piensa que el suceso A está particionado por los C_i , entonces queda dividido en n pedazos del tipo: $(A \cap C_i)$, que son sucesos mutuamente excluyentes entre sí y, a su vez, su unión total debe ser igual al suceso A ; entonces aplicando el axioma 3 generalizado se demuestra la expresión anterior. Por ejemplo, el suceso A resultar positivo está particionado por los sucesos C_1 : sano y C_2 : enfermo. Entonces:

$$P(A) = \sum P(A \cap C_i) = P(A \cap C_1) + P(A \cap C_2) = P(FP) + P(VP) = (fp/N) + (vp/N) = TP/N$$

6) *Conteo de resultados:*

Sea una prueba clínica con un total de N resultados posibles $C_1, C_2 \dots C_n$ que forman el espacio muestral S . Entonces cada experimento tendrá una probabilidad individual dada por $P(C_i)$. Sea R un suceso cualquiera formado por r resultados, entonces puede ser escrito con:

$$R = \bigcup_{i=1}^r C_i \quad \text{Y de allí resulta} \quad P(R) = \sum_{i=1}^r P(C_i) \quad \text{por la propiedad anterior.}$$

Extendiendo el caso al universo S al tomar todos los resultados posibles será:

$$P(S) = \sum_{i=1}^N P(C_i) = 1$$

Para el caso particular de *resultados igualmente probables* la expresión se reduce a:

$$P(C_1) = P(C_2) = P(C_3) = \dots = P(C_n) = 1/N$$

7) *Conclusión:*

De las expresiones del punto anterior se sabe que la $P(R)$ es la suma de las probabilidades de cada uno de los resultados que verifican R , y si estos a su vez son equiprobables resulta:

$$P(R) = \sum_{i=1}^r P(C_i) = \underbrace{(1/N) + (1/N) + \dots + (1/N)}_{r \text{ veces}} = N_R / N = r / N$$

o sea,

$$P(R) = N_R / N$$

Entonces, la probabilidad $P(R)$ de un evento cualquiera R es igual al número N_R , de elementos de R , dividido por el número total de elementos N . Lo que se parece mucho a la definición clásica de probabilidad; sin embargo existe una diferencia fundamental:

. En el enfoque de la probabilidad axiomática, la equiprobabilidad es un *supuesto* usado para establecer las probabilidades de los resultados de un experimento cualquiera.

. En el enfoque de la probabilidad clásica, la equiprobabilidad es una *conclusión lógica* y es usada de hecho para *definir* a la misma.

5.3.2 Aplicaciones

a) Si se lanza una moneda al aire, los resultados posibles son cara y seca, ambos con una probabilidad igual a 0,5. O sea, cumplen las condiciones anteriores. Ahora, si se lanza la misma moneda dos veces al aire, los casos posibles son $S \{ cc, cs, sc, ss \}$ o sea, 4 casos igualmente probables, cada uno con un valor de probabilidad de $\frac{1}{4}$. Un evento cualquiera como por ejemplo, $R \{ \text{obtener una seca en el primer lanzamiento} \} = \{ sc, ss \}$ que tendrá un valor de probabilidad dado por $P(R) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} = \frac{N_R}{N} = \frac{2}{4}$.

b) Si se lanza un dado al aire, hay seis resultados posibles, equiprobables ($P = 1/6$). Pero si se lanza el mismo dado dos veces, ahora el espacio muestral estará formado por 36 elementos (resultados posibles) $S \{ (1,1) (1,2) (1,3) \dots (6,5) (6,6) \}$ todos igualmente probables ($P = 1/36$) Un evento cualquiera $R \{ (3,3) (3,4) (4,3) \}$ tendrá una probabilidad $P(R) = 3/36 = 1/12$. Otro suceso $M \{ \text{obtener dos ases} \} = \{ (1,1) \}$ tendrá $P(M) = 1/36$.

c) Se lanza una moneda cargada al aire donde la probabilidad de cara $p = 1/3$ y la de seca $q = 2/3$ entonces se cumple $p + q = 1$. Ahora se la lanza de nuevo y se genera un espacio muestral dado por $S \{ cc, cs, sc, ss \}$, con $P(cc) = p^2$; $P(sc) = P(cs) = p \cdot q$ y $P(ss) = q^2$. Se cumple que $P(S) = p^2 + p \cdot q + q \cdot p + q^2 = (p+q)^2 = 1$. Lo anterior se cumple pues la ocurrencia de un resultado no incide en el siguiente ni es influenciado por el anterior.

d) Se efectúa un diagnóstico a partir de los análisis clínicos del paciente los resultados posibles son (+) con una probabilidad p y (-) con una probabilidad q . Se cumple la relación $p + q = 1$ pues $p = TP/N$ y $q = TN/N$, de allí $p + q = (TP + TN) / N = 1$. Se efectúa otro diagnóstico a otro paciente, las probabilidades p y q se mantienen iguales y constantes y el espacio muestral formado por este segundo caso es $S \{ (+ +) (+ -) (- +) (- -) \}$ con probabilidades análogas a las del caso anterior. (TP: total de Positivos ; TN : Total de Negativos)

e) Ocurre un nacimiento, la probabilidad que sea varón es igual a la de mujer y vale 0,5. Si son mellizos hay cuatro casos posibles, en forma análoga al ejemplo (a) de más arriba.

f) En un concurso para auxiliar de la cátedra de Bioestadística se presentan 3 candidatas. La primera: A por ser profesional tiene dos veces más posibilidades de ganar que una alumna avanzada B. Y esta a su vez el triple de ganarle a C que es de otra carrera. Se pide calcular las respectivas posibilidades de ganar y la probabilidad de que A no gane.

Si $P(C) = p$ entonces $P(B) = 3 P(C) = 3 p$ y $P(A) = 2 P(B) = 6 p$

El suceso seguro es que alguna de las tres va a ganar: $1 = P(A) + P(B) + P(C) = 6p + 2p + p = 9p$
Y no es posible que gane más de una, por lo tanto los tres sucesos son mutuamente excluyentes.

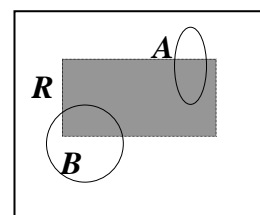
De donde se obtiene $p = 1/9$ y con eso se pueden obtener: $P(A) = 6/9$; $P(B) = 2/9$ y $P(C) = 1/9$

La probabilidad de que no gane A es igual a la de que ganen B o C. Esto es:

$$P(\bar{A}) = 1 - P(A) = 1 - 6/9 = 3/9 = 1/3 = P(B \cup C) = P(B) + P(C) = 2/9 + 1/9$$

g) Se fabrica un medicamento con 11,9 g de sulfato ferroso ($\text{SO}_4\text{Fe} \cdot 7 \text{H}_2\text{O}$) en 100 ml. y el resto excipientes como sacarosa, sorbitol y ácido cítrico con agua triple destilada. Las normas establecen una tolerancia para el sulfato de $\pm 2,1$ g/100ml, junto con otras especificaciones para los demás componentes. El farmacéutico a cargo del control de calidad sabe, por su larga experiencia, que el 1% de lo producido se rechaza por superar el límite y el 2% por no llegar al valor mínimo en el caso del sulfato. Además, un 4% se rechaza por no cumplir las restantes especificaciones. Si tiene que inspeccionar 10.000 unidades: ¿cuál será el número total de unidades que espera aceptar? Sea el esquema siguiente:

- S: lote de 10.000 unidades fabricadas
- A: unidades rechazadas por sobrepasar límite de sulfato
- B: unidades rechazadas por no llegar al límite de sulfato
- R: unidades rechazadas
- X: unidades aceptadas = \bar{R}
- C: unidades rechazadas por otras causas



$$R = C \cup (B \cap R) \cup (A \cap R) \text{ (sombreado)}$$

$$P(\bar{X}) = 1 - P(X) = 1 - P(A \cup B \cup C) = 1 - \{ P(A) + P(B) + P(C) - P(A \cap C) - P(B \cap C) \}$$

$$P(X) = 1 - \{ P(A) - P(A \cap C) \} - \{ P(B) - P(B \cap C) \} - P(C)$$

$$P(X) = 1 - \{ P(A \cap C) \} - \{ P(B) - P(B \cap C) \} - P(C)$$

De los datos del problema surgen:

$$P(C) = 4\% = 0,04 \quad P(A \cap \bar{C}) = 1\% = 0,01 \quad P(B \cap \bar{C}) = 2\% = 0,02$$

De allí es $P(X) = 1 - 0,01 - 0,02 - 0,04 = 0,93$. O sea un 93% de unidades aceptadas, lo que significa $N \cdot X = 9.300$ unidades aceptadas.

5.3.3 Índices clínicos como probabilidades

Sean los cuatro sucesos posibles al efectuar un diagnóstico:

Tabla 5.1: Eventos posibles al realizar un diagnóstico o Tabla de diagnóstico

Test clínico	Diagnóstico verificado		Total
	Enfermo: C_1	Sano: C_2	
(+)	VP: verdadero positivo	FP: falso positivo	TP: de Positivos
(-)	FN: falso negativo	VN: verdadero negativo	TN: de Negativos
Total	TE: Total Enfermos	TS: Total Sanos	N: Total pacientes

La Tabla 5.1 es un Diagrama de Venn, donde los sucesos sano (C_2) y enfermo (C_1) particionan al universo, igual que positivo y negativo. Y a su vez, los cuatro sucesos posibles VP , FP , VN , FN también particionan al universo. Notar que si el suceso es VP su valor observado o medido es la frecuencia (vp) usada en la Tabla 1.1, y lo mismo con los tres casos restantes.

$$FP = (C_2 \cap TP) \quad VP = (C_1 \cap TP) \quad FN = (C_1 \cap TN) \quad VN = (C_2 \cap TN)$$

$P(VP) = \text{número de casos posibles} / \text{número total de casos} = vp / N$ – es una probabilidad

$P(C_1) = TE / N = p$ es una probabilidad llamada la *prevalencia* de la enfermedad.

Por lo tanto el cociente entre ambas probabilidades será: $Sensibilidad = vp / TP = P(VP) / P(C_1)$
De donde se deduce que la sensibilidad es un cociente de probabilidades. Análogamente para la especificidad $E = vn / TS$ pues:

$P(VN) = vn / N$ y $P(C_2) = TS / N$ es igual al complemento de la prevalencia; o sea que el cociente entre ambas probabilidades será la *Especificidad* = $P(VN) / P(C_2)$

Puede verse que los índices clínicos principales S y E son en realidad un cociente de probabilidades, cuyo significado se explicará mejor en el capítulo siguiente. Lo mismo ocurre con los valores predictivos. En efecto:

$$P(TP) = TP / N \quad \text{y por su lado} \quad P(TN) = TN / N$$

$$VPP = vp / TP = P(VP) / P(TP) = (vp / N) / (TP / N) = vp / TP$$

$$VPN = vn / TN = P(VN) / P(TN) = (vn / N) / (TN / N) = vn / TN$$

Por su parte la Eficiencia (A) se puede obtener con la unión de los dos tipos de éxitos:

$$P(\text{n}^\circ \text{ de éxitos}) = A = P(VP \cup VN) = P(VP) + P(VN) = (vp / N) + (vn / N) = (vp + vn) / N$$

En resumen, mientras que la prevalencia y eficiencia son probabilidades simples y directas, los índices tales como la sensibilidad, especificidad y valores predictivos son un cociente de probabilidades directas.

5.4. Odds

Este concepto, que no tiene aún una traducción del idioma inglés, se puede definir como un cociente de probabilidades de la manera siguiente:

Odds: es el cociente entre la probabilidad de ocurrencia de un suceso A y la de su complemento.

$$\text{Odds} = P(A) / P(\bar{A}) = P(A) / [1 - P(A)]$$

Se pueden definir diferente tipos de Odds como los de Enfermos o Sanos con:

$$\text{Odds de enfermos} = P(\text{Enfermo}) / P(\text{Sano})$$

Si se considera a la población total, entonces la Prevalencia de la enfermedad es la cantidad de enfermos que esta tiene y la relación anterior se puede expresar con:

$$\text{Odds de enfermos} = \text{Prevalencia} / 1 - \text{Prevalencia}$$

Proporción de enfermos: es la probabilidad de contraer la enfermedad (Prevalencia).

Se puede encontrar la relación siguiente:

$$\text{Prevalencia} = P(E) = TE / N = TE / (TS + TE) = (TE / TS) / [1 + (TE / TS)]$$

$$\text{Prevalencia} = \text{Odds} / (1 + \text{Odds})$$

Esto ya fue presentado en la Tabla 4.3 para el concepto de riesgo, pero ahora se lo explica desde el punto de vista de las probabilidades para el caso particular de que el riesgo sea contraer una enfermedad cualquiera. Para entender el significado clínico de estos conceptos se dan algunos ejemplos como sigue:

Caso 1) Si la probabilidad de enfermarse (Prevalencia) es del 80%, entonces la probabilidad de no enfermarse (1 – Prevalencia) será del 20%. Por lo tanto, el Odds de enfermos es 4 a 1, lo que significa que hay 4 chances entre 5 de enfermarse.

Caso 2) Si se sabe que el Odds de enfermedad es de 5 a 1, entonces la prevalencia (o riesgo) se puede calcular como $p = 5 / (1 + 5) = 5/6 = 0,83$

Caso 3) Si se tiene una sensibilidad del 60% y una especificidad del 80%, se pueden calcular sus respectivos Odds con:

$$\text{Odds de sensibilidad} = S / (1 - S) = 0,6 / 0,4 = 1,5$$

$$\text{Odds de especificidad} = E / (1 - E) = 0,8 / 0,2 = 4$$

El producto de ambos Odds será igual a 6. Por su parte, se pueden calcular los Likelihood ratios:

$$\text{LR+} = S / (1 - E) = 0,6 / (1 - 0,8) = 0,6 / 0,2 = 3$$

$$\text{LR-} = (1 - S) / E = (1 - 0,6) / 0,8 = 0,4 / 0,8 = 0,5$$

Entonces, el cociente de ambos Likelihood Ratios es igual al producto de los Odds de sensibilidad y Especificidad. En efecto,

$$\text{LR+} / \text{LR-} = S \cdot E / [(1 - S) (1 - E)] = \text{Odds sensibilidad} \times \text{Odds especificidad} = 6 = \text{Odds Ratio}$$

A este índice se lo denomina Odds Ratio y se lo presentó en el Capítulo 4, se lo puede entender como un cociente entre el producto de los éxitos, dividido el producto de los fracasos.

Caso 4) Si el suceso analizado es la cantidad de fallas cometidas en el diagnóstico se tiene:

$$\text{Odds de fallas} = P(\text{fracasos}) / [1 - P(\text{fracasos})] = P(\text{fracasos}) / P(\text{éxitos})$$

Pero la $P(\text{fracasos}) = (fn + fp) / N$ y la $P(\text{éxitos}) = (tn + tp) / N$. Luego reemplazando y simplificando en la relación anterior se obtiene:

$$\text{Odds de fallas} = \text{Failure Odds} = (fp + fn) / (tp + tn)$$

Significa que si se tiene 1 falla en 10 pruebas el FO = 1/9, o bien cuando el FO = 2/8 entonces habrá 2 fallas cada 10 pruebas, o cada 8 éxitos. Se puede tomar a este número como otro índice para calificar la calidad de un diagnóstico.

5.5 Cuadro resumen de los índices clínicos

En el cuadro siguiente se muestra las relaciones teóricas de todos los índices clínicos con los dos índices básicos: sensibilidad, especificidad. Además se clasifica entre aquellos que varían con la prevalencia y los demás:

Cuadro 5.2 Relaciones entre índices clínicos de diagnóstico

Índices clínicos básicos: Parámetros o características principales de un test clínico

$$\text{Sensibilidad} = S = vp / TE$$

$$\text{Especificidad} = E = vn / TS$$

Índices clínicos derivados: No dependen de la prevalencia, son otros parámetros del test clínico.

$$\text{Índice de Youden} = Y = S + E - 1$$

$$\text{Likelihood Ratio de positivos} : LR+ = S / (1 - E)$$

$$\text{Likelihood Ratio de negativos} : LR- = (1 - S) / E$$

Índices clínicos extrínsecos: Dependen de la prevalencia, son variables del test clínico

$$\text{Valor Predictivo de Positivos} : VPP = vp / TP = S \cdot p / \{[(1-p)(1-E)] + [S \cdot p]\}$$

$$\text{Valor Predictivo de negativos} : NPP = vn / TN = E(1-p) / \{[E(1-p)] + [p(1-S)]\}$$

$$\text{Eficiencia} : A = (vp + vn) / N = E(1-p) + S \cdot p$$

Efecto de la población en los tests clínicos:

$$\text{Prevalencia} : p = TE / N$$

$$\text{Odds de enfermedad} : O = p / (1-p) = TE / TS$$

$$\text{Odds} / (1 + \text{Odds}) = \text{Prevalencia} ; \text{ o bien: } p = O / (1 + O)$$

5.6 Problemas propuestos

- 1) La probabilidad teórica es exactamente igual a la empírica. V F
- 2) A medida que aumenta N la probabilidad empírica se aproxima mejor a la teórica. V F
- 3) La noción de probabilidad tiene 4 acepciones básicas. V F
- 4) El conjunto universo está formado por todos los casos posibles. V F
- 5) El complemento del conjunto universo es el conjunto elemental. V F
- 6) El conjunto vacío tiene el cero como único elemento. V F
- 7) Un elemento pertenece a un conjunto es lo mismo que decir: está incluido en él. V F
- 8) Si A está incluido en B , entonces son iguales. V F
- 9) Si A está incluido en B , y este se incluye en C , entonces A está incluido en C . V F
- 10) El complemento de A está formado por los del universo que no son A . V F
- 11) La unión de un conjunto y su complemento forman al universo. V F
- 12) Dos conjuntos se dicen excluyentes si no tienen elementos comunes. V F
- 13) La unión del conjunto universo con otro cualquiera es un nuevo universo. V F
- 14) La intersección de un conjunto con el vacío es el mismo conjunto. V F
- 15) La unión de un conjunto con el conjunto vacío es el conjunto elemental. V F
- 16) La intersección de un conjunto con el universo es el mismo conjunto. V F
- 17) La unión de un conjunto con otro más grande es igual al más grande. V F
- 18) Un grupo de conjuntos es colectivamente exhaustivo si es igual al universo. V F
- 19) Un grupo de conjuntos es mutuamente excluyente si lo son todos los pares posibles. V F
- 20) Una partición se forma con un grupo de conjuntos mutuamente excluyentes. V F
- 21) Las combinaciones son el cociente entre variaciones y permutaciones. V F
- 22) Escribir las fórmulas de cálculo de combinaciones, variaciones y permutaciones.
- 23) Es lo mismo realizar extracciones con y sin reposición en el cálculo combinatorio. V F
- 24) Definir los sucesos: Universo, Vacío y Elemental.
- 25) Para que un conjunto de sucesos observables sea un Álgebra de Boole debe cumplir
- 26) Los tres axiomas básicos del Álgebra de Probabilidades son:
- 27) Las probabilidades del conjunto vacío y del universo son complementarias. V F
- 28) La probabilidad de la unión de sucesos excluyentes, es igual a la suma directa de:
- 29) Para la partición del universo por un grupo de conjuntos se requiere:
- 30) La relación entre los índices clínicos está dada por la relación:

- 31) Once estudiantes de la Facultad son condenados a muerte y se les concede pedir un último deseo. Cuando el que aprobó Bioestadística nota que tardan 5 minutos en formarlos contra la pared, pide que los ejecuten luego de que los coloquen en el paredón de todas las formas posibles. Explicar porqué salvó a sus compañeros.

- 32) Una bolsa contiene 4 bolas blancas y 2 negras, otra contiene 3 blancas y 5 negras; si se extrae una bola de cada bolsa. Hallar la probabilidad de que:
 - a) ambas sean blancas;
 - b) ambas sean negras;
 - c) una blanca y la otra negra.

- 33) De cuántas formas pueden ordenarse 7 libros en un estante si:
 - a) es posible cualquier ordenación;
 - b) tres libros determinados deben estar juntos;

c) dos libros determinados deben estar en los extremos.

34) Una farmacia que hace reparto a domicilio tiene 4 cadetes con moto y llegan 3 pedidos a la vez. ¿De cuántas formas puede el encargado asignar la tarea a los cadetes?

35) Doce estudiantes deben rendir un examen parcial y el profesor tiene 4 temas diferentes para interrogarlos. ¿De cuántas formas diferentes puede hacerlo?

36) Cinco pacientes vienen a efectuarse cinco análisis diferentes al laboratorio de análisis clínicos. ¿De cuántas maneras diferentes pueden colocarse en la fila de espera si se les da libertad para hacerlo? Ver si cambian las cosas si se les da un número por orden de arriba para ser atendidos por el bioquímico.

37) ¿De cuántas maneras diferentes pueden sentarse en una fila 3 bioquímicos, 5 farmacéuticos y 4 enfermeros? Ver si cambia el resultado si se sientan en una mesa redonda.

38) Se tiene un mazo de 40 cartas españolas y se pide calcular la probabilidad de que salga:
a) un as; b) una carta de oro; c) una figura; d) un número cualquiera; e) un dos o un tres.

39) En una farmacia hay 5.000 remedios distintos, 1.200 artículos medicinales y 600 productos de perfumería. Si se toma un ítem al azar calcular la probabilidad de que sea: a) un remedio; b) un perfume; c) un artículo medicinal o un perfume.

40) Se fabrica una vacuna contra la gripe que falla en un 2% de los casos si la persona está tomando antibióticos, en un 4% si está haciendo un régimen adelgazante y en un 8% por otras causas diferentes. Si se analizan 5.000 casos: ¿cuál será el número total de curados?

41) Sopa de letras: buscar y marcar las palabras que se definen a continuación:

B	R	N	O	R	A	L	I	N	A	C	O	N	H	D	O	X	A	G	U	E	S	W	O	U	M	T	R	A	X
A	E	I	N	G	R	A	G	C	I	L	N	T	A	Z	T	O	L	Z	E	G	H	X	B	F	A	U	N	A	S
L	E	R	I	C	I	E	N	C	I	A	X	I	X	P	A	R	T	I	C	I	O	N	I	L	A	W	C	N	O
O	E	A	N	E	S	T	A	N	D	P	R	T	T	Z	O	U	G	L	T	E	Z	W	C	Q	X	R	O	W	G
X	I	S	C	O	Y	N	A	S	V	L	C	X	Q	I	R	F	O	I	T	A	L	Q	U	P	Y	O	M	X	S
E	M	Y	T	J	U	C	H	T	M	A	B	C	W	A	N	O	I	S	U	L	C	X	E	M	O	B	P	L	A
R	P	B	Q	I	V	L	U	I	Y	C	A	I	C	N	E	U	C	E	R	F	R	E	S	A	X	P	L	S	R
O	O	R	G	D	M	H	L	J	N	E	T	R	V	A	C	I	O	L	A	X	V	N	R	O	E	E	E	Y	T
J	S	O	Q	E	L	P	N	I	R	B	E	H	I	S	R	I	J	U	T	A	K	K	D	P	M	V	M	S	A
I	I	E	M	Q	C	X	L	T	W	D	F	Q	B	F	J	L	S	H	L	J	A	F	I	I	P	Z	E	I	V
O	B	S	D	S	U	N	I	V	E	R	S	O	T	N	V	A	C	E	C	Z	X	I	E	C	I	S	N	L	S
U	L	T	N	O	I	S	R	E	P	S	E	D	Y	Q	I	N	C	L	U	S	I	O	N	T	R	E	T	S	Ñ
G	E	A	Z	B	N	V	U	O	I	O	G	T	K	R	X	C	Z	A	F	G	O	T	T	O	I	J	O	K	N
L	L	D	E	L	A	R	I	P	U	D	U	M	P	Q	I	Z	Q	W	E	C	M	E	O	G	C	A	C	A	X
A	Q	I	M	W	T	E	B	O	T	I	R	C	L	A	C	E	K	R	V	J	A	R	M	R	A	A	I	A	S
B	D	P	E	R	M	U	T	A	C	I	O	N	L	E	M	A	L	W	D	U	S	R	I	A	F	N	F	U	O
V	Q	E	L	F	U	C	R	W	Ñ	Ñ	C	A	R	T	N	I	W	G	O	I	C	I	C	M	U	D	I	T	P
T	Ñ	S	S	U	F	I	C	I	E	N	T	E	A	N	T	Y	R	S	L	L	N	V	M	A	T	N	C	X	H
K	I	B	W	D	A	D	I	L	I	B	I	S	N	E	S	N	G	H	D	A	Ñ	O	S	C	A	R	I	O	D

1. Primer matemático en estudiar probabilidades. 2. Teorema de relaciona probabilidades empíricas con las teóricas. 3. Un grupo de conjuntos colectivamente exhaustivos y mutuamente excluyentes forman una del universo. 4. Una está formada cuando dos conjuntos no tienen ningún elemento en común. 5. Lo inverso del punto anterior, cuando un conjunto tiene todos los elementos de otro. 6. Suceso cuando no puede ocurrir. 7. Lo contrario del anterior. 8. El conjunto que tiene todos los elementos posibles. 9. El que no tiene ningún elemento. 10. Un conjunto es ... de otro cuando contiene a todos los elementos del universo que no pertenecen al primero. 11. Los tres ... básicos del Álgebra de Probabilidades. 12. El principio de razón se agregó a la definición clásica de probabilidad. 13. Probabilidad ... se calcula como cociente de frecuencia experimentales. 14. Una ... de n elementos se calcula con el factorial de tal número.

42) Conociendo los siguientes valores de Prevalencia de una enfermedad, calcular los correspondientes Odds y la Probabilidad de estar sano:

Prevalencia	Odds enfermos	(1 – Prevalencia)
10		
20		
33		
43		
50		
66		
80		
88		
94		

43) Para los valores de las Tablas 4.1 y 4.2 del capítulo anterior, calcular los correspondientes LR+ y LR- para cada uno de los puntos de corte adoptados. ¿Cuánto vale el Odds?

44) Encontrar la relación entre el VP+ y LR+ reemplazando en las ecuaciones dadas en el punto 5.5.3 anterior. ¿Qué relación se encuentra con los Odds de enfermedad?

45) Calcular todos los índices clínicos conociendo los datos de la tabla siguiente:

En Realidad Está	Diagnóstico		Total
	D+	D-	
Sano	45	98	143
Enfermo	105	102	207
Total	150	200	350

46) Con la tabla de datos obtenidas en un experimento hecho sobre 400 sujetos, se pide calcular los índices clínicos y compararlos con los del ejemplo anterior. Explicar cual de los dos es el mejor de acuerdo a los tres Tipos de enfermedades posibles, efectuando una comparación cualitativa entre los mismos. Es decir una comparación “a ojo” sin tener la correspondiente validación estadística de estas conclusiones.

Verdad	Diagnóstico		Total
	+	-	
Sano	180	98	278
Enfermo	40	82	122
Total	220	180	400

47) Analizar la calidad de el test clínico desde un punto de vista dual, es decir teniendo en cuenta Sensibilidad, Especificidad si es una enfermedad del Tipo II. Analizar que ocurre con el Delta Predictor simulando 10 valores de prevalencia entre 0,1 y 0,9.

Verdad	Diagnóstico		Total
	+	-	
Sano	100	250	350
Enfermo	400	250	650
Total	500	500	1000

48) Lo mismo que en el ejemplo anterior pero si se trata de una enfermedad del Tipo I

Verdad	Diagnóstico		Total
	+	-	
Sano	90	350	440
Enfermo	310	50	360
Total	400	400	800

49) Ídem anterior pero para una enfermedad del Tipo III

Verdad	Diagnóstico		Total
	+	-	
Sano	40	460	500
Enfermo	360	40	400
Total	400	500	900

50) Conociendo la sensibilidad (90%) y la especificidad (75%) de una prueba clínica se pide calcular todos los siguientes índices clínicos que conozca. Representar la variabilidad de ciertos índices con la prevalencia, y la obtención de sus valores para $p = 0,4$.

Apéndice 1: Teoría de conjuntos

Se define como *conjunto* a una serie de objetos llamados *elementos* o *entes*. Como en Filosofía, se entiende por ente a *todo lo que es*. Los elementos pueden ser reales o imaginarios, tangibles o intangibles; mientras se pueda decir algo de ellos, son entes. Al caso contrario se lo denomina la *nada*, o sea *lo que no es*. Como la nada no puede tener elementos, conviene asimilar este concepto al denominado *conjunto vacío* que se denota con:

\emptyset : es el conjunto que no tiene elementos

Si los pacientes se clasifican en (+) y (-), no puede haber ninguno que a la vez sea (+) y (-). Su opuesto, o complemento, es el conjunto que incluye a todos los elementos posibles, al que se llama *conjunto universo* denotado con:

\mathcal{U} : es el conjunto de todos los elementos posibles

Por ejemplo, todos los pacientes que concurren a hacerse atender por el clínico, configuran su *universo* de pacientes o su población de referencia.

Cualquier subconjunto del conjunto universo estará formado por una serie de elementos pertenecientes al mismo. Este conjunto será una parte del universo porque no puede definirse fuera del mismo. Por ejemplo, el conjunto de pacientes sanos, el conjunto de enfermos, etc. Normalmente, en biología el universo será la población en estudio.

A : un conjunto formado por varios elementos e_i

Hay dos casos extremos: cuando el suceso A esté formado por todos los elementos será el universo, y cuando tenga un solo elemento será el *conjunto unitario* o *elemental*. Se pueden usar como ejemplos de conjuntos a los casos siguientes:

A : { centrífuga, jeringa, pipeta, balanza } = { e_1, e_2, e_3, e_4 }

B : { sangre, suero, kit de glucosa, remedio } = { e_5, e_6, e_7, e_8 }

C : { bioquímico, farmacéutico, paciente } = { e_9, e_{10}, e_{11} }

Donde con el símbolo e_i se ha denotado a cada uno de los diferentes elementos usados como ejemplos. A partir de allí se pueden definir una serie de *relaciones* entre los conjuntos y sus elementos, como la *pertenencia*:

Se dice que un elemento *pertenece* a un conjunto cualquiera cuando está en el mismo, en caso contrario se dice que *no pertenece* al mismo. Se los expresa con:

$e_3 \in A$ pertenece

$e_{23} \notin A$ no pertenece (con e_{23} : heladera)

Esta idea es algo diferente a la de *inclusión* que se utiliza para relaciones entre conjuntos. Por ejemplo: sea el conjunto K : { pipeta, balanza } = { e_3, e_4 }, K está incluido en A , pues *todos* sus

elementos pertenecen también a A . En cambio, el conjunto $L: \{ e_3, e_5, e_7, e_{23} \} = \{ \text{pipeta, sangre, kit de glucosa, heladera} \}$ no lo está:

$$K \subset A \quad (K \text{ está incluido en } A)$$

o bien,

$$L \not\subset A \quad (L \text{ no está incluido en } A)$$

El concepto de inclusión, significa que K es un subconjunto de A . En cambio, si algunos o todos los elementos de uno no pertenecen al otro, entonces no se trata de un subconjunto y no se puede hablar de inclusión. Para el caso particular donde ambos tienen los mismos elementos, la inclusión se transforma en **igualdad**. Así, dos conjuntos son iguales cuando:

$$A = X \Leftrightarrow X: \{ e_i \forall i / e_i \in A \}$$

o sea: $A = X \{ \text{centrífuga, jeringa, pipeta, balanza} \} = \{ e_1, e_2, e_3, e_4 \}$

La **unión** de dos conjuntos es otro conjunto formado por los elementos de ambos. O sea:

$$B \cup C: \{ e_i \forall i / e_i \in B \text{ o } e_i \in C \} = \{ e_5, e_6, e_7, e_8, e_9, e_{10}, e_{11} \}$$

La **intersección** de dos conjuntos es otro conjunto formado por todos los elementos comunes a ambos, y se denota con:

$$B \cap L: \{ e_i \forall i / e_i \in B \text{ y } e_i \in L \} = \{ e_5, e_7 \}$$

El **complemento** de un conjunto A cualquiera es otro conjunto: \bar{A} formado por los elementos del universo, que no pertenecen a A . O sea:

$$\bar{A}: \{ e_i \forall i / e_i \in \emptyset \text{ y } e_i \notin A \}$$

Si para los ejemplos anteriores se define un universo formado por cien elementos, tal como $\emptyset: \{ e_i \forall i = 1, 2, 3, \dots, 100 \}$, entonces el complemento del conjunto A será:

$$\bar{A}: \{ e_5, e_6, e_7, \dots, e_{99}, e_{100} \}$$

Se dice que el conjunto $B \nabla L$ es el **complemento relativo** del conjunto L cuando está formado por todos los elementos de B que no pertenecen a L . Esto es:

$$B \nabla L = B \cap \bar{L}: \{ e_i \forall i / e_i \in B \text{ y } e_i \notin L \} = \{ e_6, e_8 \}$$

o viceversa:

$$L \nabla B = L \cap \bar{B}: \{ e_i \forall i / e_i \in L \text{ y } e_i \notin B \} = \{ e_3, e_{23} \}$$

Se dice que dos conjuntos son **excluyentes** cuando no tienen ningún elemento en común. A veces se los llama **disjuntos**. Esto significa que la intersección de ambos conjuntos no tiene elementos.

O sea, debe ser igual al conjunto vacío. Por eso el concepto de exclusión no necesita un símbolo especial como los anteriores. Se define con:

Si $A \cap C = \emptyset$, entonces son excluyentes o disjuntos.

También se pueden ilustrar estos conceptos con la Tabla de la Verdad vista en el capítulo anterior

El conjunto de pacientes enfermos se expresa con **TE** y tiene una cantidad de elementos conformado por todos los individuos enfermos (TE) que concurren al laboratorio, donde se verificó que padecían una cierta enfermedad. Análogamente con los sanos, entonces:

$TE \cap TS = \emptyset$ (no hay ninguno que a la vez sea sano y enfermo).

$TE \cup TS = \wp$ (todos los casos posibles, el universo)

Además se puede hacer lo mismo con el conjunto de los diagnosticados positivos (**TP**) y los negativos (**TN**), entonces queda:

$TP \cap TN = \emptyset$ (no hay ninguno que a la vez sea sano y enfermo).

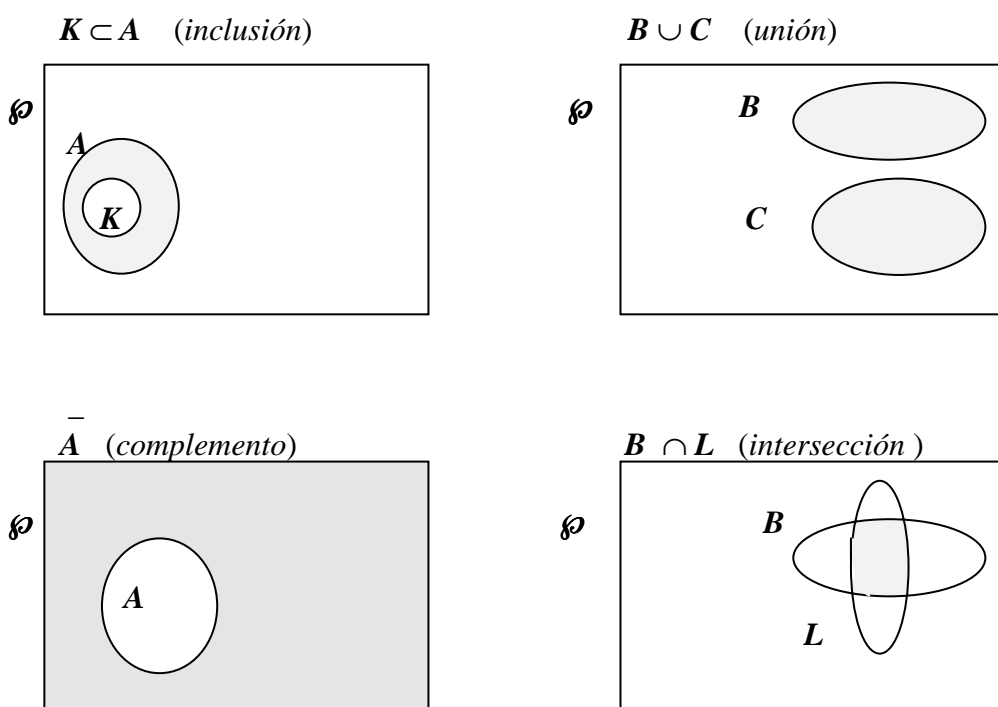
$TP \cup TN = \wp$ (todos los casos posibles, el universo).

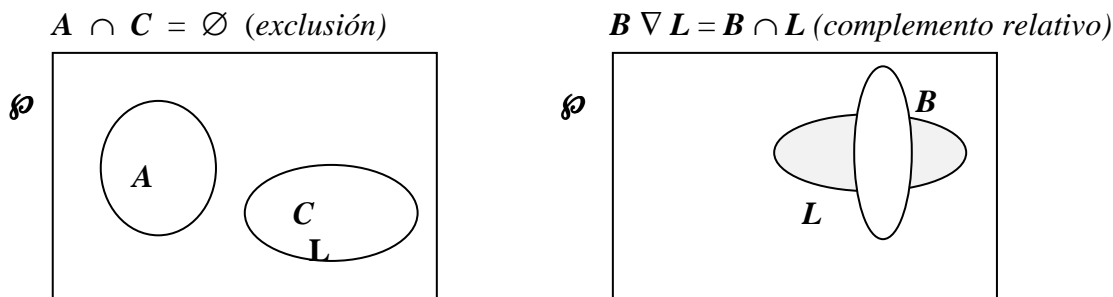
Los complementos serían:

$\overline{TS} = TE$ o bien $\overline{TE} = TS$ y para el diagnóstico $\overline{TP} = TN$ o bien $\overline{TN} = TP$

Para una mejor visualización de los conceptos de la teoría de conjuntos, conviene utilizar un modo gráfico al mostrar sus relaciones mutuas:

Gráfico A5.1: Diagramas de Venn en la teoría de conjuntos





Los Diagramas de Venn suelen usarse para esta tarea. Con un rectángulo se representa al universo \emptyset . Dentro del mismo se usan entornos cerrados de forma ligeramente ovalada para mostrar a los conjuntos. Adecuados sombreados resaltan las relaciones que se quieren mostrar. Todo esto se muestra en el Gráfico 5.1 siguiente, donde se han representado las relaciones vistas más arriba entre los conjuntos definidos como A , B , C , K y L para los conceptos de inclusión, unión, intersección, complemento común y relativo y para la exclusión.

Propiedades

El siguiente conjunto de propiedades se puede deducir de las definiciones vistas más arriba; aquí se generaliza para no repetir las mismas propiedades en cada una de las relaciones entre conjuntos descriptas.

Propiedad transitiva: Si $K \subset A$ y $A \subset \emptyset$ entonces $K \subset \emptyset$

Propiedad conmutativa: $B \cap L = L \cap B$ o bien $B \cup C = C \cup B$

Propiedad asociativa: $A \cup (B \cup C) = (A \cup B) \cup C$ o bien $A \cap (B \cap L) = (A \cap L) \cap B$

Propiedad distributiva: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Idempotencia: $B \cap B = B$ y también $A \cup A = A$

Identidad: $A \cup \emptyset = \emptyset$ y también $A \cup \emptyset = A$

$$A \cap \emptyset = A \quad \text{y también} \quad A \cap \emptyset = \emptyset$$

Por ejemplo, la unión de tres sucesos A (formado por 5 bolillas numeradas de 0 a 4) y B (formados por otras 3 bolillas numeradas de 5 a 7), y C (formado por 2 bolillas la 8 y la 9), va a resultar igual al conjunto de las diez bolillas que conforman un bolillero común. Es lo mismo unir A con B primero, y luego unir C , que cualquier otra manera de unir a esos tres conjuntos, el resultado siempre será el mismo y eso ilustra la propiedad transitiva. Análogamente para el caso de dos conjuntos en lugar de tres. La idempotencia significa que todos conjunto unido a sí mismo, da el conjunto original, al igual que su intersección. La identidad significa que todos conjunto unido con el universo resulta el universo, mientras que si se hace la intersección el resultado es el mismo conjunto.

Ejemplos de aplicación

1) Decidir si los siguientes conjuntos son vacíos:

a) $X : \{x \mid x^2 = 36 \text{ y } 2x = 8\} = \emptyset$ porque no hay ningún x que satisfaga ambas relaciones.

b) $X : \{x \mid x \neq x\} = \emptyset$ porque no hay ningún x que cumpla el requisito.

c) $X : \{x \mid x + 23 = 23\} \neq \emptyset$ porque cero la satisface. O sea, $X = \{0\}$

2) Si un conjunto A de pacientes resultó positivo, y entre ellos un conjunto K estaba enfermo y el resto L estaba sano: ¿ Como se pueden usar las propiedades anteriores ? Lo mismo si se tiene un conjunto de pacientes B con resultado negativo y de entre ellos hay otro conjunto X que estaban sanos y otro Y que estaban enfermos.

Hallar el conjunto de sanos y el de enfermos, si : $A \cup B = \emptyset$

3) Los 4 conjuntos de figuras en el plano son:

$A : \{x \mid x \text{ es un cuadrilátero}\}$

$C : \{x \mid x \text{ es un rombo}\}$

$B : \{x \mid x \text{ es un rectángulo}\}$

$D : \{x \mid x \text{ es un cuadrado}\}$

Determinar cuáles son subconjuntos de los otros.

Un cuadrado es un rectángulo por tener sus ángulos rectos, es un rombo por tener sus 4 lados iguales y además es un cuadrilátero por tener cuatro lados. Esto es: $D \subset B$; $D \subset C$; $D \subset A$
Como hay cuadriláteros, rectángulos y rombos que no son cuadrados, D es un subconjunto propio de los otros tres.

Por su parte, como un rombo y un rectángulo son casos particulares de cuadriláteros resulta: $B \subset A$ y $C \subset A$

4) Determinar cuáles de los siguientes conjuntos son iguales: \emptyset ; $\{\emptyset\}$; $\{0\}$:

Ninguno. El conjunto $\{0\}$ tiene un elemento: el número cero. El conjunto $\{\emptyset\}$ tiene un elemento: el conjunto vacío. Y el conjunto vacío \emptyset : no tiene ningún elemento.

5) Determinar cuál es la principal diferencia entre los conjuntos siguientes:

A : Los días del año.

C : Los bioquímicos argentinos.

B : La cantidad de farmacias en Posadas.

D : El conjunto de los números reales.

La principal diferencia es que el último es un conjunto de tamaño infinito, mientras que los tres primeros son finitos.

6) Determinar los cuatro conjuntos posibles al diagnosticar a muchos pacientes con la misma técnica clínica:

Son cuatro: vp es el conjunto de los verdaderos positivos, vn es el conjunto de los verdaderos negativos, fp los falsos positivos y fn los falsos negativos.

Apéndice 2: Cálculo combinatorio

Antes de entrar al campo algebraico conviene repasar los conceptos del cálculo combinatorio, necesarios para calcular el número de casos en los problemas de probabilidad.

. *Conteos ordenados*: si un suceso puede ocurrir de r maneras diferentes, y si, continuando el proceso, otro puede ocurrir de s maneras distintas, y luego un tercer evento de t formas, y así sucesivamente, entonces existirán N maneras de realizar todos los sucesos en el orden indicado.

$$N = r \cdot s \cdot t \dots$$

Por ejemplo, para calcular la cantidad de Códigos Postales del país, que tiene cuatro cifras, donde la primera no puede ser cero. Entonces, en el primer puesto habrá 9 casos posibles (los números de 1 a 9), en el segundo, tercer y cuarto puesto habrá 10 casos posibles (los números de 0 a 9), por lo tanto resulta $N = 9 \cdot 10 \cdot 10 \cdot 10 = 9000$ códigos posibles en total.

. *Variaciones*: si se tienen n objetos diferentes en un conjunto, y de ellos se extraen r , se llama *variación de n objetos tomados de a r* al número de maneras diferentes que hay de hacerlo en forma ordenada.

$$V(n, r) = n! / (n-r)!$$

Por ejemplo, se tienen $n = 4$ objetos diferentes a, b, c y d. Tomado de a uno ($r = 1$), resultan 4 casos posibles a b c d; $V(4,1) = 4! / (4-1)! = 4$. En cambio, si se toman de a dos, entonces es $r = 2$ y los casos son: ab ac ad ba bc bd ca cb cd da db dc, o sea, 12 casos diferentes en total; $V(4,2) = 4! / (4-2)! = 12$. Para las variaciones, no es lo mismo el subconjunto “ab” que el “ba” pues interesa el orden. Además, como son extracciones sin reposición, no se pueden repetir los objetos, es decir el caso “aa” es imposible. Todo ocurre como si se colocara la mano en una “bolsa” de n objetos, y se sacan r de ellos, “de a uno y por orden”.

Otro ejemplo: una secretaria debe distribuir a 5 pacientes en 3 salas de atención diferentes con turnos disponibles; entonces dispone de $V(5,3) = 5! / (5-3)! = 20$ variaciones diferentes.

. *Permutaciones*: si se tienen n objetos diferentes, se llama permutación al número de formas diferentes que hay de ordenarlos entre sí. Este concepto es equivalente al de una variación, pero con $r = n$. O sea,

$$P_n = V(n, n) = n!$$

Como ejemplo, se puede imaginar a un bioquímico que debe realizar 3 determinaciones a, b y c en una muestra de un paciente. Las posibilidades son $P_3 = 3! = 6$. Que se pueden visualizar con abc acb bac bca cab cba.

. *Combinaciones*: si se tienen n objetos diferentes en un conjunto, y de ellos se extraen r a la vez, se llama *combinación de n objetos tomados de a r* , al número total de maneras posibles que hay de hacerlo (sin tomar en cuenta el orden).

$$C(n, r) = n! / (n-r)! \cdot r!$$

La relación entre variaciones, permutaciones y combinaciones es:

$$C(n, r) = V(n, r) / r! = V(n, r) / Pr$$

Por ejemplo, un visitador médico de un laboratorio farmacéutico concurre a un sanatorio a entrevistar a 5 médicos, pero dispone de tiempo para solo 3 de ellos, entonces tiene 10 maneras distintas de hacerlo: $10 = C(5, 3) = 5! / (5-3)! \cdot 3! = 10$ o sea hay 10 combinaciones posibles. Se trata de combinaciones pues no interesa el orden de las visitas.

Otro ejemplo: un bioquímico quiere comprar 5 kits diferentes, pero solo le alcanza el dinero para 3 de ellos. Entonces, tiene 10 combinaciones posibles para hacer su compra. Vuelve a su laboratorio con los tres que pudo adquirir, y ahora se pregunta en qué orden los utilizará; resulta que puede hacerlo de 6 formas diferentes o permutaciones posibles. Finalmente, decide entregar uno a cada uno de sus dos ayudantes y guardar el tercero. Así, tiene 3 variaciones posibles de cómo poder hacerlo.

. *Propiedades de la combinatoria:*

$$C(n, 0) = 1 = C(n, n) = n! / 0! n! = n! / n!$$

$$C(n, 1) = n = C(n, n-1) = n! / 1! (n-1)! = n! / (n-1)!$$

$$C(n, r) = C(n, n-r) = n! / r! (n-r)!$$

Las combinaciones de 10 productos tomados de a 1 son 10; lo mismo que si son tomados de a 9. A su vez, si se toman de a 10 no hay ninguna, lo que es lo mismo que si no se toma ninguno. Pero si se toman de a 2 entonces hay: $10! / 8! 2! = 45$ combinaciones posibles.

. *Permutaciones con repetición:* puede ocurrir que muchos objetos se repitan en el conjunto total, entonces, el número de permutaciones se calcula con :

$$PR(n/s, t, u... z) = n! / s! \cdot t! \cdot u! \dots z!$$

Por ejemplo, si se tienen 10 clientes en una farmacia de los cuales 8 son hombres y 2 son mujeres, el número de permutaciones posibles es $PR = 10! / 8! \cdot 2! = 45$. Eso significa que hay 45 maneras de formar dos grupos ordenados en la cola de espera, uno de a 8 y el otro de 2.

. *Pruebas repetidas:* en muchos casos, el experimento consiste en escoger un elemento cualquiera de un conjunto de tamaño n , y repetir esto r veces. En el caso de los juegos de azar, el ejemplo más común es cuando se extraen cartas de un mazo, o en un sorteo cuando se sacan bolillas del bolillero. El detalle a tener muy en cuenta en estos casos es ver si se reponen o no los elementos escogidos del conjunto. De acuerdo con ello se diferencian así:

.. *Pruebas con substitución:* cuando se repone el objeto escogido antes de seguir con la extracción siguiente. En tal caso, cada vez que se escoge un elemento hay siempre un total de n elementos. Así, el conteo de casos es :

$$n \cdot n \cdot n \dots n = n^r$$

Por ejemplo, en un mazo de barajas españolas de 40 cartas se pueden sacar 3 de ellas en forma sucesiva una cantidad $40 \cdot 40 \cdot 40 = 64.000$ maneras diferentes, si se hace con reposición.

.. *Pruebas sin substitución*: cuando no se repone el objeto extraído, antes de seguir con la siguiente extracción. En tal caso, cada vez que se escoge un elemento, el total de n elementos disminuye en una unidad. Así, el conteo para r casos es:

$$n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-r+1) = n! / (n-r)! = V(n, r)$$

Por ejemplo, de un mazo de barajas españolas de 40 cartas se sacan 3 de ellas en forma sucesiva y hay un número de: $40 \cdot 39 \cdot 38 = 59.280$ maneras diferentes de hacerlo, cuando no hay reposición. Un bioquímico toma 2 carpetas de su fichero de 150 pacientes, entonces tiene unas 22.350 formas diferentes de hacerlo. Un farmacéutico saca 2 calmantes de su stock de 10 para mostrarle al cliente, entonces tiene 90 maneras distintas de hacerlo.

Producto cartesiano: sean dos conjuntos A y B cuyos respectivos elementos son A_i y B_i ; se denomina producto cartesiano a un nuevo conjunto formado con todos los pares posibles $A_i B_i$

$$C = A \times B$$

Por ejemplo, si A (calmante, laxante, antibiótico) y B (agua, sangre) el producto cartesiano de ambos será C (calmante-agua, calmante-sangre, laxante-agua, laxante-sangre, antibiótico-agua, antibiótico-sangre).

6

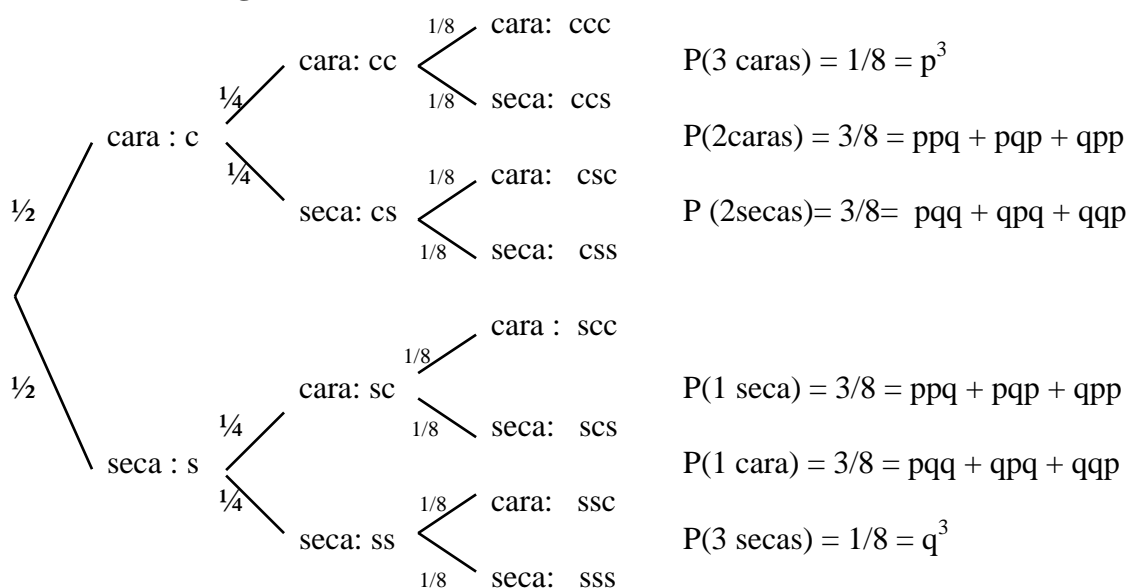
Probabilidad Condicional

En este capítulo se introduce el concepto de la probabilidad condicional a la problemática probabilística. Se comienza con el enfoque clásico, para luego ir avanzando hacia las relaciones bayesianas, en particular entre índices clínicos. Se discuten los conceptos de independencia y condicionalidad para lograr un enfoque más general que el visto hasta ahora.

6.1 Introducción

La Teoría de Probabilidades comenzó con el estudio de los juegos de azar. El caso más simple es el lanzamiento de una moneda, donde pueden ocurrir dos casos posibles. (c: cara y s: seca). Al repetir la misma prueba, los casos posibles ahora serán cuatro, y al hacerlo una tercer vez serán ocho. Un diagrama de árbol ayudara a visualizar mejor la cuestión (Gráfico 6.1). Es de esperar que la probabilidad $p = 1/2$ de obtener una cara, o $q = 1/2$ de tener una seca, siempre sea la misma y no se altere por más veces que se lance la moneda al aire. Es decir, se trata de una probabilidad constante a lo largo de una serie de pruebas. Lo mismo ocurre con la mayoría de los juegos de azar, incluso con barajas cuando las extracciones se hacen con reposición. Resultó natural entonces, calcular la probabilidad de los lanzamientos sucesivos, como el producto de las probabilidades básicas p y q . Así, la probabilidad de sacar 3 caras seguidas es de $1/8$ como se aprecia en el diagrama y se obtiene con p^3 .

Gráfico 6.1: Diagrama del árbol.



Como puede verse más arriba, el producto de las probabilidades de cada rama, siguiendo el camino adecuado, permite obtener los valores de cada uno de los resultados posibles al hacer los tres lanzamientos. Ese método se conoció con el nombre de *regla del producto*. A su vez, cuando se busca obtener un resultado indirecto que se da en más de una rama, entonces se aplica la *regla de la suma*, esto es, se suman las probabilidades de cada una de las ramas involucradas. Los ejemplos de todo se han presentan en el Gráfico 6.1.

Sin embargo, esta manera de resolver problemas no sirve para todos los casos. Hay un ejemplo histórico de ello llamado “La paradoja de Marte”:

“... Muchos años atrás, se planteaba como problema la posible existencia de vida en el planeta Marte. Como no se sabía mucho del planeta, salvo que tenía atmósfera como la Tierra, se especulaba acerca de la existencia de alguna especie animal o vegetal con vida. Si se parte de la base de no saber nada, se puede asumir como $\frac{1}{2}$ la probabilidad p de que haya vida, o q de que no la haya, para una especie cualquiera como la del Homo-Sapiens. Ahora, tomando otra especie como los peces, por un razonamiento análogo se llega al mismo valor, $p = q$. Y así sucesivamente, para n especies se tendrá que la probabilidad total de que haya especímenes vivos en Marte será:

$$P(\exists \text{ vida}) = P(\text{hombres y vacas y aves y peces vivos})$$

$$P(\exists \text{ vida}) = P(\text{hombres vivos}) \cap P(\text{vacas vivas}) \cap P(\text{aves vivas}) \cap \dots \cap P(\text{peces vivos})$$

$$P(\exists \text{ vida}) = p \cdot p \cdot p \dots p = p^n$$

Lo que conduce a una paradoja, pues
$$P(\exists \text{ vida}) = p^n \xrightarrow[n \rightarrow \infty]{} 0$$

O sea, se parte de una ignorancia estimada en $p = q = \frac{1}{2}$ y se llega a la certeza de que no hay vida. Y ocurre lo mismo tomado otros valores de p y q . Si se estima que la probabilidad de vida es muy baja $p = 0,001$, también tiende a cero con n muy grande. Si se estima lo contrario, una alta probabilidad de vida $p = 0,9$, entonces tiende más rápido”.

Otra paradoja: tomando una cepa de *Drosophila* que difiera de la cepa salvaje en dos caracteres: el color del cuerpo es negro ébano (E), mientras el color normal es ($E+$) y las alas son vestigiales (A), mientras las normales ($A+$). Si se cruza una hembra salvaje ($e a$) con un macho normal ($e+ a+$), en la primera generación F1 se obtienen únicamente moscas del tipo ($e+ a+$) que son los caracteres dominantes. Luego, se realiza un segundo cruzamiento retrógrado, de los obtenidos en F1 con normales. En la segunda generación filial F2 se obtendrán cuatro tipos de casos posibles, todos igualmente probable $p = \frac{1}{4}$ de acuerdo a las leyes de Mendel: “los gametos sólo transportan uno de los genes alelos presentes en los progenitores”, habiendo igualdad de frecuencias para ambos gametos. Los cuatro tipos de *Drosophilas* son:

		Mendel	Frec. obtenida
cuerpo normal y alas normales :	$E+ \cap A+$ o sea ($e+ a+$)	25%	240
cuerpo normal y alas vestigiales:	$E+ \cap A$ o sea ($e+ a$)	25%	266
cuerpo negro y alas normales :	$E \cap A+$ o sea ($e a+$)	25%	278
cuerpo negro y alas vestigiales:	$E \cap A$ o sea ($e a$)	25%	252

Habría igual número de machos y hembras. Hecho el experimento se obtuvieron valores muy parecidos a los mendelianos, como era de esperar. $p = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

Si se hubieran cruzado machos y hembras de la F1, entonces en la F2 se hubiesen obtenido también cuatro tipos de moscas, pero con diferentes proporciones, de acuerdo a la segunda ley de Mendel : 9/16, 3/16, 3/16 y 1/16. O sea (9:3:3:1), válida para la segregación de caracteres independientes cuando se cruzan entre sí dos individuos de la F1.

Si se analizan ahora otros caracteres, como por ejemplo una cepa salvaje con alas vestigiales y ojos pardos (**B**), en el primer cruzamiento F1, resultaran todos con los caracteres dominantes (**b+ a+**) Un cruzamiento retrógrado entre una hembra F1 y un macho salvaje dará lugar a cuatro tipos de moscas:

	Mendel	Frec. obtenida
ojos normales y alas normales : $B+ \cap A+$ o sea (b+ a+)	25%	370
ojos normales y alas vestigiales: $B+ \cap A$ o sea (b+ a)	25%	138
ojos pardos y alas normales : $B \cap A+$ o sea (b a+)	25%	155
ojos pardos y alas vestigiales: $B \cap A$ o sea (b a)	25%	350

Se da una paradoja, pues los valores obtenidos, distan de parecerse a los mendelianos esperados. Tomando esa evidencia, la conclusión sería que para esos caracteres dejan de cumplirse las leyes de Mendel, cosa que ningún genetista está dispuesto a aceptar.

En síntesis, se muestran dos paradojas que despertaron polémicas en su momento, consecuencia de aplicar la llamada *regla del producto*, lejos de su ámbito natural de los juegos de azar. Se puede pensar lo siguiente: si la probabilidad de encontrar una vaca viva en Marte es p, seguramente la probabilidad de encontrar la misma vaca conviviendo con tigre vivo no será la misma. Aumentará para el tigre y disminuirá para la vaca. Esto es, hay hechos cuya ocurrencia *condicionan* la ocurrencia de otros. En cambio, si al tirar una moneda sale cara, este hecho *no condicionará* lo que salga en la tirada siguiente. Entonces, había que replantearse el concepto de las probabilidades teniendo esto en cuenta. Y así, se llegó al concepto de *Probabilidad Condicional*. En la segunda paradoja, los genetistas lejos de sacrificar las leyes de Mendel, pensaron que los caracteres que no la cumplían no eran *independientes*, y así definieron un nuevo concepto el de los *caracteres ligados*.

6.2 Independencia

Se dice entonces que dos sucesos son *independientes* cuando la ocurrencia de uno de ellos, no modifica la ocurrencia del otro, ni esta influenciado por este. Si se realiza una serie de pruebas repetidas, las pruebas son *independientes*, cuando el resultado de una de ellas no está influenciada por el resultado de la prueba anterior, ni tampoco influenciará el resultado de la prueba siguiente.

Es importante no confundir independencia con exclusión. Pues en este último caso, la ocurrencia de un suceso hace imposible la ocurrencia del otro. O sea, se “obligan” entre sí. Esto lleva a pensar que dos sucesos son excluyentes cuando su intersección es el conjunto vacío, o sea, la probabilidad de que ocurra uno y el otro, será nula. En cambio, para que sean independientes deberán tener por lo menos un elemento en común, para que la probabilidad de su ocurrencia conjunta no sea nula. Por otro lado, para que un suceso condicione a otro también debe

haber algún elemento en común, pues si no hay ninguna relación entre ellos no podrá haber influencia de uno sobre el otro. Se pueden resumir estos casos con:

. *Sucesos excluyentes* : No tienen ningún elemento en común. O sea, $A \cap B = \emptyset$

. *Sucesos independientes*: Deben tener algún elemento en común. O sea, $A \cap B \neq \emptyset$

. *Sucesos condicionales* : Deben tener algún elemento en común. O sea, $A \cap B \neq \emptyset$

Se dice que dos sucesos A y B son *independientes* cuando la probabilidad de que ambos ocurran es igual al producto de sus probabilidades mutuas (independencia estadística). O sea:

Si se cumple, $P(A \cap B) = P(A) \cdot P(B)$ entonces son independientes.

Esta es la vieja regla del producto. Se cumple en las reglas de la herencia humana, en los juegos de azar, en la teoría cinética de gases (Maxwell), y en un sinnúmero de casos de la vida real. Además, toda sucesión finita de experimentos en las cuales cada experimento tiene un número finito de resultados posibles, se llama *proceso estocástico* y es el caso típico de aplicación de la regla del producto donde las probabilidades en cada rama son independientes paso a paso; en cambio, si ocurre la secuencia de una rama, es imposible que ocurra la secuencia de otra. O sea, las ramas son excluyentes entre sí, y se les puede aplicar el axioma 3 o la regla de la suma para obtener la probabilidad de más de una rama.

Se puede generalizar con: Sean n sucesos C_i todos independientes entre sí, entonces se cumple que :

$$P(C_1 \cap C_2 \cap C_3 \cap \dots \cap C_n) = P(C_1) \cdot P(C_2) \cdot P(C_3) \dots P(C_n)$$

Si se realiza una prueba n veces, donde cada resultado es un suceso independiente de los demás, la probabilidad de ocurrencia de un resultado común a todos ellos es el producto de sus probabilidades. Por ejemplo, en el diagrama del árbol de Gráfico 6.1, la probabilidad que salgan 3 caras es igual al producto $p \cdot p \cdot p = 1/8$, y generalizando al caso de n lanzamientos, la probabilidad de obtener todas caras será $(\frac{1}{2})^n$.

Como corolario de la propiedad anterior, se cumple que si los n sucesos son independientes entre sí, entonces también lo serán para una cantidad menor de sucesos.

$$P(C_j \cap C_{j+1} \cap C_{j+2} \cap \dots \cap C_k) = P(C_j) \cdot P(C_{j+1}) \cdot P(C_{j+2}) \dots P(C_k)$$

Pero debe destacarse que *lo contrario no se cumple*, es decir si n sucesos son independientes de a pares, en todos los pares posibles, eso no permite afirmar que lo serán de a tres, de a cuatro, etc. Lo mismo para el caso inverso. Todo esto se muestra en los siguientes ejemplos:

Ejemplo 1) Una clínica tiene nueve pacientes en la sala de espera. De los cuales 3 son adultos, 3 son niños y 3 son ancianos. Cada paciente, de cada grupo, tiene un número de turno que va de 1 a 3. Además, el paciente adulto con el turno 1, el niño con el turno 2 y el anciano con el 3 son del sexo masculino y los demás del sexo femenino.

Sean los sucesos:
 A : un adulto

$$\begin{aligned}P(A) &= P(AC \cup AG) = P(AC) + P(AG) = 1/9 + 2/9 = 1/3 \text{ (por el axioma 3)} \\P(C) &= P(AC \cup GC \cup CG) = P(AC) + P(GC) + P(CG) = 2/3 \text{ (generalización de axioma 3)} \\P(G) &= P(GC \cup AG \cup CG \cup GG) = 1/9 + 2/9 + 1/9 + 2/9 = 2/3 \text{ (idem anterior)} \\P(U) &= P(UU \cup UU) = P(UU) = 2/9 \\P(A \cap C) &= P(AC) = 1/9 = P(A) \cdot P(C) \text{ entonces } A \text{ y } C \text{ son independientes} \\P(A \cap G) &= P(AG) = 2/9 = P(A) \cdot P(G) \text{ entonces } A \text{ y } G \text{ son independientes} \\P(C \cap G) &= P(CG) = 2/9 = P(C) \cdot P(G) \text{ entonces } C \text{ y } G \text{ son independientes}\end{aligned}$$

En cambio $P(A \cap C \cap G) = P(\emptyset) = 0 \neq P(A) \cdot P(C) \cdot P(G)$ no son independientes. Se puede interpretar esto diciendo: Si un fragmento contiene adenina, la probabilidad de que contenga citosina o uracilo se mantiene constante; en cambio, si contiene adenina y guanina, es imposible que contenga citosina.

Los tres ejemplos anteriores ilustran la cuestión de independencia de a pares y de a triples. Puede verse que, si los pares lo son, no hay garantía que los tripletes lo sea, y viceversa, que lo sean los tripletes, no implica que lo sean de a pares. La única forma de asegurarse, es aplicando la definición a cada caso particular en estudio. *No se puede ver la independencia de un vistazo*

6.3 Condicionalidad

Recordando las anécdotas de la paradoja de Marte, o el concepto genético de carácter ligado, como así también la extracción de cartas de un mazo sin reposición, resulta clara la existencia de sucesos que condicionan la ocurrencia de otros. Las relaciones ya vistas de probabilidad resultan insuficientes para tratar estos casos. Se necesita entonces, definir un nuevo concepto: la *probabilidad condicional* que ocurra un suceso **A**, luego de que ha sucedido un suceso **B**, puede ser calculada con:

$$P(A / B) = P(A \cap B) / P(B)$$

O lo que es lo mismo,

$$P(A \cap B) = P(A / B) \cdot P(B)$$

Se dice entonces, la ocurrencia de **B** incide en la ocurrencia de **A**. Por ejemplo, la probabilidad de sacar dos cartas sucesivas de espadas en un mazo de 40 cartas españolas, se obtiene con:

$$P(\text{sacar } 1^{\text{a}} \text{ y } 2^{\text{a}} \text{ de espadas}) = P(2^{\text{a}} \text{ espada} / 1^{\text{a}} \text{ espada}) \cdot P(1^{\text{a}} \text{ espada}) = 9/39 \cdot 10/40 = 0,058$$

Esto permite encontrar otra manera de definir dos sucesos independientes. Si sucede que se cumple: $P(A / B) = P(A)$, entonces **B** no condiciona la ocurrencia de **A** y resultan independientes. Resumiendo :

$$\begin{aligned}\text{Condicionalidad} &: P(A \cap B) = P(A / B) \cdot P(B) \\ \text{Independencia} &: P(A \cap B) = P(A) \cdot P(B) \\ \text{Exclusión} &: P(A \cap B) = 0\end{aligned}$$

Como se ve, la condición más general es la condicionalidad, y como casos muy particulares de la misma, se da la condición de independencia o la de exclusión. La consistencia entre esta probabilidad de tipo teórica y la empírica se da con el lema de Bernoulli. Si se sabe que en una serie de

N experimentos hay resultados que verifican la condición **A**, acompañado por la ocurrencia de la condición **B**, entonces habrá un cierto número de pruebas N_{AB} , donde se verifican juntas ambas condiciones; entonces su frecuencia relativa se calcula con :

$$Fr_{AB} = N_{AB} / N_B$$

Pues ahora el número total de casos posibles no es más N sino N_B , porque se sabe que debe ocurrir primero **B**, para que ocurra **A**, de acuerdo a la condición impuesta "a priori". De todos los casos N_B , solo algunos N_{AB} se verificarán junto con **A**. Entonces,

$$\begin{array}{ccc} Fr_{AB} & \longrightarrow & P(A / B) \\ N & \longrightarrow & \infty \end{array}$$

Además,

$$Fr_{AB} = N_{AB} / N_B = (N_{AB} / N) / (N_B / N) = Fr_{(A \cap B)} / Fr_B$$

$$\begin{array}{ccc} Fr_{AB} = Fr_{(A \cap B)} / Fr_B & \longrightarrow & P(A \cap B) / P(B) = P(A / B) \\ N & \longrightarrow & \infty \end{array}$$

Se puede generalizar la definición de probabilidad condicional de la manera siguiente: Sean n conjuntos $C_1, C_2, C_3, \dots, C_n$ pertenecientes al universo ϕ , tales que no son mutuamente excluyentes, o sea sí $C_1 \cap C_2 \cap C_3 \cap \dots \cap C_n \neq \emptyset$ entonces,

$$P(C_1 \cap C_2 \cap C_3 \cap \dots \cap C_n) = P(C_1) \cdot P(C_2 / C_1) \cdot P(C_3 / C_1 \text{ y } C_2) \dots$$

Por ejemplo, para obtener la probabilidad de sacar flor de espadas en un juego de truco, dando las tres primeras cartas seguidas: $P(\text{flor espadas}) = (10/40) \cdot (9/39) \cdot (8/38) = 0,0012$

6.3.1 Simulación para un test clínico

Se puede realizar una simulación con probabilidades condicionales siguiendo los lineamientos del experimento de Bernoulli. Imaginando que se tienen 20.000 fichas, la mitad rojas y la otra mitad blancas, cada una numerada consecutivamente de 1 a 20.000. Esto simula a 10.000 individuos enfermos (rojas) y el resto sanos (blancas), como si fueran la población general. El supuesto principal es que la capacidad de la prueba clínica para detectar a los enfermos y a los sanos es constante en toda la población. Y puede ser expresada con por ejemplo:

- 1) El método es capaz de detectar correctamente a 3 enfermos de cada 4 enfermos.
- 2) El método es capaz de detectar correctamente a 8 sanos entre 10 de ellos.

Esto significa que la sensibilidad es $S = 3/4 = 0,75$ y la especificidad es $E = 8/10 = 0,8$. Esta capacidad es la misma para todo individuo de la población, y esos valores son los teóricos para dicha población, que tiene una prevalencia $\pi = 0,5$. Luego se marcan al azar a las fichas para que cumplan con este requisito: 3 de cada 4 rojas se las identifica como verdadero positivo (vp), y 8 de cada 10 blancas se las identifica como verdadero negativo (vn). Luego cada ficha tendrá un número propio (como si fuese nombre y apellido del individuo) y su verdadero estado (uno de estos cuatro: vp, vn, fp y fn).

Caso 1: Se mezclan las fichas rojas y blancas, y de entre ellas se eligen al azar a $N = 4.000$. Luego se cuentan los casos encontrados y se tendrán las cuatro frecuencias experimentales:

$N^{\circ} vp$: número de casos encontrados de fichas rojas con diagnóstico verdadero positivo.

$N^{\circ} vn$: lo mismo que arriba pero de las fichas blancas (verdadero negativo).

$N^{\circ} fp$: la frecuencia encontrada de fichas rojas con diagnóstico falso.

$N^{\circ} fn$: lo mismo que arriba pero para las fichas blancas.

Con estas cuatro frecuencias se puede armar una Tabla de diagnóstico experimental. Pero, por el lema de Bernoulli, como $N = 4.000$ es grande se sabe estas frecuencias experimentales tenderán a sus respectivos valores teóricos. Y entonces, con los valores experimentales se pueden obtener los índices clínicos experimentales, para ver que tanto se acercan a los valores teóricos. Por ejemplo, se puede deducir que:

$S = P(VP/TE)$ es la probabilidad condicional de encontrar un verdadero positivo, sabiendo que el individuo está enfermo. Lo que se puede expresar como:

$S = P(VP \cap TE) / P(TE)$ y en otros términos sería:

$S = \text{número de fichas rojas y positivas} / \text{número total de fichas rojas}$

Análogamente, $E = \text{número total de fichas blancas y negativas} / \text{número total de fichas blancas}$.

De esta forma se pueden obtener experimentalmente los valores de S , E , Y , $LR+$ y $LR-$ muy parecidos a los esperados. Además la prevalencia experimental p es aproximadamente igual a la teórica $p \approx \pi = 0,5$. Mientras que los valores predictivos y la eficiencia obtenidos son similares a los esperados cuando la prevalencia $p = 0,5$.

Se puede repetir muchas veces esto hasta convencerse de que el lema de Bernoulli funciona bien.

Caso 2: Se separan las fichas rojas y blancas en dos conjuntos, y se eligen al azar 2.000 fichas rojas y otras tantas blancas, para tener un total de $N = 4.000$ muestras en total. La diferencia con el caso anterior es que la muestra ha sido estratificada en mitad enfermos y mitad sanos, *antes de hacer la selección al azar*. Luego se cuentan los casos seleccionados en sus cuatro resultados posibles y se puede armar otra Tabla de la Verdad experimental. Los resultados son muy similares a los del caso anterior.

Para estudiar que pasa si se cambia el valor de la prevalencia en la muestra, se puede comenzar por ejemplo eligiendo 400 fichas rojas y 3.600 blancas. Esto simula el llamado diseño de estudio del tipo Caso-control, donde la relación es de 1 caso cada 9 controles. O sea, una prevalencia simulada en esta muestra $p = 0,1$. Se cuentan las frecuencias halladas y se arma la Tabla de la Verdad experimental, con lo que se calculan todos los índices clínicos. Los resultados son que:

- a) S , E , Y , $LR+$ y $LR-$ son iguales a los anteriores.
- b) A y valores predictivos cambian.

Si se van simulando distintas prevalencias $p = 0,2; 0,3; \dots; 0,9$ las consecuencias son iguales a las dos anteriores (a) y (b). Se puede repetir esto tantas veces como se quiera hasta convencerse que:

Hay un grupo de índices que no varían no importa la manera en que se tome la muestra (los llamados *parámetros* o características internas del método), y hay otros que varían con la prevalencia de una manera muy similar a la explicada en el capítulo 4. Si se grafican los valores variables, se pueden encontrar las mismas curvas detalladas en los Gráficos 4.2 y 4.3. Esto demuestra experimentalmente el razonamiento efectuado antes.

Caso 3: Se separan las fichas en dos grupos, uno con los resultados positivos (15.500 fichas) y otro con los negativos (4.500 fichas) sin que importe el color de las mismas. Luego se eligen al azar 2.000 casos de positivos y otros tantos de negativos, de manera tal que el tamaño de la muestra sea $N = 4.000$. Es decir, que acá se ha estratificado la muestra en positivos y negativos, para simular el llamado estudio por cohortes. En este caso, la relación entre positivos y negativos elegida es $P = 50\%$. Se cuentan los resultados obtenidos para obtener las cuatro frecuencias, con las cuales se arma la Tabla de la Verdad experimental para este caso. Los resultados obtenidos son $p \approx \pi = 0,5$, y los mismos resultados obtenidos en el *Caso 1*.

Luego para ver la influencia del porcentaje de positivos y negativos seleccionados, se eligen una serie de valores $P = 10\%, 20\%, \dots, 90\%$ y se vuelven a obtener los mismos resultados anteriores. De donde se puede concluir que: *No importa el porcentaje de positivos y negativos, los índices clínicos no se alteran*. En otras palabras, no interesa la manera de estratificar la muestra basada en los resultados del diagnóstico, nada cambia (Ver Cuadro 5.2 del capítulo anterior).

En conclusión, *la manera de seleccionar las muestras no afecta los índices clínicos paramétricos como S, E, IY y Likelihood Ratios. Los únicos afectados son la eficiencia y los valores predictivos por el efecto de la prevalencia simulada en los estudios del tipo caso-control*. Todo esto desde el punto de vista experimental, la explicación desde un punto de vista teórico se alcanza a través del teorema de Bayes como se explica a continuación.

6.4 Teoremas de Probabilidad Total y de Bayes

En el apartado 5.2. se puede ver el Diagrama de Venn para la partición de un suceso cualquiera A por un grupo de n sucesos C_i . y en la propiedad (5) vista en 5.3.1, se puede ver la probabilidad del suceso $P(A)$, expresada como la sumatoria de las probabilidades de sus trozos componentes $P(A \cap C_i)$, formados por la partición del suceso. Ahora bien, reemplazando estas probabilidades compuestas con la definición de probabilidad, se tiene la expresión del *Teorema de la Probabilidad Total*.

Sean n sucesos pertenecientes al mismo espacio muestral S , tales que particionan al suceso A (ver Gráfico 5.2) entonces se cumple que :

$$P(A) = \sum_{i=1}^n P(A \cap C_i) \quad \Leftrightarrow \quad \left\{ \begin{array}{l} (C_i \cap C_j) \cap A = \emptyset \quad \forall i \neq j \\ \bigcup_{i=1}^n C_i \supseteq A \end{array} \right.$$

Reemplazando $P(A \cap C_i)$ por la definición resulta:

$P(A) = \sum_{i=1}^n P(C_i) \cdot P(A / C_i)$	<i>Teorema de Probabilidad Total</i>
---	--------------------------------------

Como una consecuencia del teorema anterior, haciendo substituciones adecuadas, se obtiene el otro teorema. En efecto, de la definición resulta :

$$P(C_i) \cdot P(A / C_i) = P(A \cap C_i) = P(C_i \cap A) = P(A) \cdot P(C_i / A)$$

Despejando,

$$P(C_i / A) = \{ P(C_i) \cdot P(A / C_i) \} / P(A)$$

Y reemplazando $P(A)$ con el Teorema de Probabilidad Total, se obtiene el *Teorema de Bayes*

$$P(C_i / A) = \frac{P(C_i) \cdot P(A / C_i)}{\sum_{i=1}^n P(C_i) \cdot P(A / C_i)}$$

Donde :

$P(C_i)$: son las llamadas probabilidades “a priori” por ser las que tienen los sucesos C_i antes de saber que ha ocurrido el suceso A .

$P(A \cap C_i)$: son las probabilidades conjuntas

$P(A / C_i)$: son las probabilidades condicionales

$P(C_i / A)$: son las llamadas probabilidades “a posteriori” porque son las que tienen los sucesos C_i , luego de saber que ha ocurrido el suceso A .

Para ilustrar estos conceptos se desarrollan los ejemplos siguientes:

Ejemplo 1) En una sala de una clínica especializada solo se tratan tres tipos de enfermedades. Se sabe que en promedio ingresan un 50% de pacientes con la afección K , 30% con la enfermedad L y el resto con la afección M (datos obtenidos con las estadísticas de los últimos dos años). Realizando un relevamiento de historias clínicas se dedujo que un 70% de los ingresados con la enfermedad K se curan, mientras que para L y M , se obtuvieron 80% y 90% respectivamente. En la fecha, se dio de alta a un paciente: ¿Cuál es la probabilidad que se haya internado por la enfermedad K ?

Sean C_1 : pacientes con la enfermedad K . ∴ $P(C_1) = 0,5$ y la $P(C / C_1) = 0,7$
 C_2 : pacientes con la enfermedad L . ∴ $P(C_2) = 0,3$ y la $P(C / C_2) = 0,8$
 C_3 : pacientes con la enfermedad M . ∴ $P(C_3) = 0,2$ y la $P(C / C_3) = 0,9$
 C : pacientes curados y dados de alta.
 C / C_i : pacientes curados y dados de alta que se internaron con la afección i

Con los datos anteriores se puede armar la tabla del Tabla 6.1

Tabla 6.1 : Cuadro bayesiano para resolución de problemas.

Sucesos	A priori	Condicional	Conjunta	A posteriori
	(a)	(b)	(c)=(a) . (b)	(d)=(c)/ $\sum(c)$
C_1	0,5	0,7	0,35	0,4546
C_2	0,3	0,8	0,24	0,3117
C_3	0,2	0,9	0,18	0,2337
Totales	1		0,77	1

Las probabilidades *a priori*, se colocan en la primer columna de datos (a). Como verificación, el total de las mismas debe ser 1. Las probabilidades condicionales se colocan en la segunda (b). Luego se calculan las probabilidades conjuntas, multiplicando las dos anteriores entre sí. O sea, la columna (c) es igual a (a) .(b) La sumatoria de esta columna, es la probabilidad total de que el paciente se cure $P(C) = 0,77$. Finalmente, se pueden calcular las probabilidades *a posteriori*, dividiendo cada término de la columna (c) por la probabilidad total igual a 0,77. La respuesta a la pregunta del problema, se halla en el primer casillero de la columna (d). Esto es, la probabilidad de que el paciente dado de alta haya ingresado con la enfermedad **K** es de 45,46%.

Ejemplo 2) En una farmacia se sabe por experiencia que el 60% de su clientela es de sexo femenino. La estadística histórica muestra que un 4% de los hombres gasta más de \$100 en una compra, mientras que solo el 1% de las mujeres lo hace. Se elige al azar una compra mayor de \$100 y se desea saber cuál es la probabilidad de que el cliente sea mujer.

Definiendo los sucesos: **H** : hombre es $P(H) = 0,4$ **C** : compra mayor a \$100
M : mujer es $P(M) = 0,6$

$$\text{La probabilidad pedida es: } P(M/C) = \frac{P(M) P(C/M)}{P(M) P(C/M) + P(H) P(C/H)}$$

$$P(M/C) = \frac{(0,6) (0,01)}{(0,6) (0,01) + (0,4) (0,04)} = \frac{3}{11} = 0,273$$

Hay un 27,3% de probabilidad que el cliente sea una mujer.

6.5. Diagnóstico y el Teorema de Bayes

Los cuatro casos posibles al efectuar un diagnóstico, basándose en la evidencia suministrada por el resultado de un análisis clínico, fueron presentados en el Tabla 5.1. (apartado 5.3.3). A la luz de los nuevos conceptos, conviene volver a revisar tal deducción, pero tratando de entender el significado de los índices clínicos, usando el concepto de la probabilidad condicional.

VPP : Se calcula con la fórmula del Teorema de Bayes. O sea, es una probabilidad bayesiana *a posteriori* y se puede imaginar como la probabilidad de diagnosticar al paciente en forma correcta un resultado positivo, sabiendo que el resultado de su análisis dio + Permite discriminar de entre todos los resultados positivos, la fracción de resultados correctos.

$$VPP = \frac{P(C_1 \cap TP)}{\sum_{i=1}^2 P(C_i \cap TP)} = P(C_1 / TP)$$

$$VPP = \frac{P(C_1) \cdot P(TP / C_1)}{P(C_1) \cdot P(TP / C_1) + P(C_2) \cdot P(TP / C_2)} = P(C_1 / TP)$$

Sensibilidad : es la probabilidad condicional $P(TP / C_1)$ de que un paciente sea diagnosticado como enfermo, cuando se sabe que en realidad está enfermo.

Prevalencia : es la probabilidad de enfermarse $P(C_1)$. O sea, es la probabilidad de encontrar la enfermedad en la población.

Sensibilidad . Prevalencia : El producto de ambos índices es $P(C_1 \cap TP)$, una probabilidad conjunta que significa, la probabilidad de que un paciente este enfermo y al mismo tiempo su diagnóstico haya dado positivo. Difiere del concepto anterior, pues no se parte de un hecho condicionante conocido. Es la probabilidad de obtener un verdadero positivo $P(VP)$.

$$P(VP) = \text{Sensibilidad} \cdot \text{Prevalencia} = P(C_1 \cap TP) = P(TP / C_1) \cdot P(C_1)$$

Especificidad : Es la probabilidad condicional $P(TS / C_2)$ de que un paciente sea diagnosticado como sano, cuando se sabe que en realidad está sano.

Probabilidad de estar sano : Es el complemento de la Prevalencia $P(C_2) = 1 - P(C_1)$. Además:

$$P(TP / C_2) = P(C_2 \cap TP) / P(C_2) = P(FP) / P(C_2) = (fp / N) / (TS / N) = fp / TS = (TS - vn) / TS$$

$$P(TP / C_2) = 1 - (vn / TS) = (1 - \text{especificidad}) : \text{Complemento de Especificidad}$$

$$\text{Entonces, } P(C_2) \cdot P(TP / C_2) = (1 - \text{especificidad}) (1 - \text{prevalencia})$$

$$\text{O bien, } P(C_1 \cap TP) = (1 - \text{Prevalencia}) (1 - \text{Especificidad}) = P(FP)$$

El Teorema de Bayes para los índices clínicos se puede escribir con:

$$VPP = \frac{\text{Prevalencia} \cdot \text{Sensibilidad}}{(1 - \text{Prevalencia})(1 - \text{Especificidad}) + \text{Prevalencia} \cdot \text{Sensibilidad}} = P(C_1 / TP)$$

$$VPP = Sp / [(1-p)(1-E) + pS] = 1 / [1 + (O \cdot LR+)^{-1}] \text{ donde } O = p / (1 - p) \text{ y análogamente:}$$

$$VPN = 1 / [1 + (O \cdot LR-)] \text{ fórmulas más sencillas de recordar el Teorema de Bayes.}$$

Notar que los valores predictivos son función de la prevalencia, y reemplazando los valores del problema visto en el Gráfico 4.2 anterior, con $S = 0,9$ y $S = 0,75$, se verifican teóricamente las dos curvas mostradas. Además, se verifican las relaciones experimentales halladas en 6.3.1, con este modelo teórico. O sea que el Teorema de Bayes puede ser comprobado en forma empírica.

6.5.1 Odds a posteriori

Aplicando el Teorema de Bayes a los nuevos índices clínicos, se pueden calcular los Odds a posteriori. Esto es, el Odds de la enfermedad luego de hacer el test clínico. Ya sea que los resultados encontrados modifiquen o no los Odds a priori disponibles. Esto se usa cuando hay un dilema en el diagnóstico a efectuarle al paciente y no se está seguro de cómo proceder. Es virtualmente imposible recordar la Sensibilidad y Especificidad de cada test clínico efectuado para hacer un diagnóstico. Sin embargo, cuando hay un dilema con el diagnóstico a efectuar, es muy útil usar la información de S; E; LR+; LR- y los Odds a priori como guía a seguir. En especial, cuando se hacen varios tests clínicos en secuencia y el Odds a posteriori del primero, se usa como a Odds a priori para la segunda prueba clínica y así sucesivamente. Un cuidadoso análisis de la información contenida en un Odds a priori, arroja luz sobre cual debería ser el paso siguiente hacia el diagnóstico. O sea, cuál debiera ser la prueba clínica a ordenar.

Como se vio en el capítulo anterior, el Odds de la enfermedad se calcula como el cociente entre la probabilidad de enfermedad y su complementaria. Como esta información se dispone al armar la tabla de decisión para el diagnóstico, se define como:

$$\text{Odds a priori} = TE / TS \text{ (para la enfermedad)}$$

El Odds a posteriori se calcula con el Odds del valor predictivo de positivos:

$$\text{Odds a posteriori} = VPP / (1 - VPP) = (p \cdot S / X) / [(1 - (p \cdot S / X))]$$

Donde $X = p \cdot S + (1 - E) \cdot (1 - p)$, simplificando queda

$$\text{Odds a posteriori} = p \cdot S / [X - (p \cdot S)] = p \cdot S / [p \cdot S + (1 - E) \cdot (1 - p) - p \cdot S]$$

Odds a posteriori = $p \cdot S / [(1 - E) \cdot (1 - p)]$ y reagrupando resulta:

$$\text{Odds a posteriori} = \frac{\text{Sensibilidad} \times \text{Prevalencia}}{(1 - \text{Especificidad})(1 - \text{Prevalencia})} = \frac{\text{Sensibilidad}}{(1 - \text{Especificidad})} \times \frac{\text{Prevalencia}}{(1 - \text{Prevalencia})}$$

$$\text{Odds a posteriori} = LR+ \cdot \text{Odds a priori}$$

Concepto mucho más sencillo de recordar y usar que el Teorema de Bayes. El uso de los Odds a posteriori tanto en el caso de positivos como de negativos, se ejemplifica a continuación:

Ejemplo 1) Para diagnosticar una enfermedad en la arteria coronaria, un cardiólogo realiza un test de tolerancia a un ejercicio físico programado. De sus historias clínicas con resultados verificados escoge a 147 pacientes con problemas coronarios y los clasifica de acuerdo a su diagnóstico, junto con el estado real que se pudo verificar después. De esta forma, arma una Tabla de diagnóstico para ver que tan bien trabaja (datos tomados del trabajo de Simel). Los resultados fueron:

Test de tolerancia al ejercicio físico	Enfermedad Coronaria		
	Si	No	Total
(+)	73	9	82
(-)	28	37	65
Total	101	46	147

Se pueden calcular los principales índices clínicos con:

$$\begin{aligned} \text{Sensibilidad} &= \text{tp} / \text{TE} = 73 / 101 = 0,72 & \text{Prevalencia} &= \text{TE} / \text{N} = 101/147 = 0,69 \\ \text{Especificidad} &= \text{tn} / \text{TS} = 37 / 46 = 0,80 & \text{Eficiencia} &= (\text{tp} + \text{tn}) / \text{N} = (73+37)/147= 0,75 \\ \text{LR+} &= \text{Sens.} / (1 - \text{Espec}) = 0,72 / 0,2 = 3,6 & \text{PPV} &= \text{tp} / \text{T+} = 73 / 82 = 0,89 \\ \text{LR-} &= (1 - \text{Sens.}) / \text{Espec.} = 0,28/0,80 = 0,35 & \text{NPV} &= \text{tn} / \text{T-} = 37 / 65 = 0,57 \end{aligned}$$

Ejemplo 2) Luego de este trabajo, viene un paciente de 52 años de edad que tiene antecedentes familiares importantes de enfermedades coronarias y presenta los síntomas de una típica angina inducida por esfuerzo. Se desea saber si el test de tolerancia al ejercicio físico influirá en el diagnóstico. El cardiólogo estima en un 80% la sospecha de que tenga esta enfermedad y usando la información obtenida en el ejemplo anterior, calcula:

Su estimación de un 80% implica un Odds de 4 a 1 de que tenga la enfermedad. Y usando sus antecedentes vistos en el ejemplo anterior como referencia poblacional, deduce que:

. Si el paciente da (+) con el test, el Odds a posteriori será de $4 \times 3,6 = 14,4$ su chance de estar enfermo. Esto implica que el Odds sube de 4 a 1 hasta 14,4 a 1 (o sea, de un 80% a un 93,5%).

. Si el paciente da (-) con el test, el Odds a posteriori será de $4 \times 0,35 = 1,4$. Esto implica que el Odds baja de 4 a 1 hasta 1,4 a 1 (o sea, de un 80% a un 58%).

Se concluye que un resultado (+) no influye mucho en el diagnóstico de enfermedad coronaria y además, un resultado (-) tampoco añade mucha más información. Por lo tanto, no conviene indicarle que se haga un test de tolerancia al ejercicio físico porque no ayuda mucho al diagnóstico. La probabilidad pre-test no varía demasiado y no agregará información importante.

Ejemplo 3) Otro hombre de 52 años de edad, robusto en apariencia tiene dolores en el pecho desde hace unos meses. No encuentra relación entre dolor y esfuerzo físico, además carece de los síntomas asociados. ¿ Puede un test de tolerancia al ejercicio físico ser de utilidad para diagnosticar una enfermedad coronaria, al sospechar que el dolor es una señal de una angina no típica ? Se sabe de los libros de texto que la prevalencia para una angina no típica es del 48%.

La estimación de un 48% implica a priori un Odds de 0,92 a 1 de que tenga una angina no típica. Los LR no varían porque Sensibilidad y Especificidad no dependen del paciente sino de la población tomada como referencia. Usando los antecedentes del ejemplo (1) anterior como referencia, se deduce que:

. Si el paciente da (+) con el test, el Odds a posteriori será de $0,92 \times 3,6 = 3,3$. Esto implica que el Odds sube de 0,92 a 1 hasta 3,3 a 1 (o sea, de un 48% a un 77%).

. Si el paciente da (-) con el test, el Odds a posteriori será de $0,92 \times 0,35 = 0,32$. Esto implica que el Odds baja de 0,92 a 1 hasta 0,32 a 1 (o sea, de un 48% a un 24%).

Se concluye que un resultado (+) o (-) influye mucho en el diagnóstico de enfermedad coronaria. Por lo tanto, conviene indicarle que se haga un test de tolerancia al ejercicio físico porque sería de gran ayuda para el diagnóstico. En este caso la información pre-test resultará enriquecida y se ve la conveniencia de efectuar el ejercicio a la tolerancia física en ese paciente.

6.5.2 Simplificación del Teorema de Bayes

Una forma más sencilla de recordar el Teorema de Bayes es usando las relaciones:

$$NPV = 1 / \{1 + [(LR- \cdot O)]\} \quad \text{y} \quad PPV = 1 / \{1 + [1 / (LR+ \cdot O)]\}$$

En la bibliografía clínica (JAMA) la recomendación actual es usar la metodología siguiente para calcular las probabilidades Post-test (a posteriori) a partir de las probabilidades Pre-test (a priori):

Paso 1: En forma subjetiva el clínico asigna una probabilidad a priori, luego de estudiar al paciente, de acuerdo a su experiencia en este tipo de casos. Esto es, la probabilidad sospechada que el mismo esté enfermo (Pre-test), que en el Ejemplo 2 anterior era 80%.

Paso 2: Luego transforma esta probabilidad en Odds ($0,8/0,2 = 4$). Entonces tiene el llamado Pre-test Odds, que en el ejemplo vale 4 : 1

Paso 3: A continuación aplica el Teorema de Bayes para obtener los Odds a posteriori, denominados Post-test Odds, que para el ejemplo será: 14,4

Paso 4: Transforma este Odds en probabilidad con la relación $p = O / (O + 1)$ resultando el valor de la llamada probabilidad Post-Test, que en el ejemplo es: 93%.

Paso 5: De acuerdo al resultado obtenido en el paso anterior, decide si le conviene, o no, hacerle el test al paciente y obtiene su primera conclusión. En el ejemplo, como la probabilidad sube de 80% a 93,5% decide que no ayudaría en mucho el hacerle la ergometría al paciente.

Paso 6: Para el caso en que el paciente de (-) en el test, aplica nuevamente el Teorema de Bayes para obtener el Odds a posteriori en el caso de los negativos. Es decir, multiplica el Pre-Test Odds de estar enfermo (4 : 1), por el Likelihood ratio de negativo (0,35) y obtiene $1,4 = 4 \times 0,35$.

Paso 7: Transforma este Odds en probabilidad y obtiene la probabilidad a posteriori de no estar enfermo (Post-test para el caso de negativo) de un 53%.

Paso 8: De acuerdo al resultado anterior decide si le conviene hacerle el test al paciente, en el caso de que este estuviera sano. Como la probabilidad antes de hacerle el test es del 80% y luego de hacerle la ergometría baja al 53%. Le puede ayudar mucho el hacerla, pero por otro lado sería riesgosa si el paciente estuviera enfermo como él sospecha (80%). Concluye finalmente, que no es conveniente efectuar la ergometría.

Como puede verse no es un procedimiento tan sencillo de seguir, a pesar de la simplificación hecha con el Teorema de Bayes. Un estudio realizado en Suiza en marzo del 2002 a más de 600 concurrentes a un curso de especialización, mostró que solo 30% de ellos había oído de este procedimiento y tan solo un 10% lo aplicaba en su práctica diaria. No parece lógico pedirle al clínico que tenga que hacer estos cálculos cada vez que estudie a un paciente. Por ese motivo, la cátedra propuso un método gráfico, que consta de un solo paso, y que no requiere del conocimiento de conceptos tales como Pre y Post-Odds de sanos y enfermos, ni del Teorema de Bayes.

Esta idea se basa en dos gráficos, donde se muestra la variabilidad de las probabilidades a posteriori, partiendo de la probabilidad a priori (un gráfico para el caso de positivos y otro para el de negativos). El concepto se basa en las siguientes relaciones:

Caso positivo: La probabilidad Pre-test y Post-test se relacionan con:

$$\text{Odds a posteriori (+)} = \text{LR+} \times \text{Odds a priori (+)} = [S / (1-E)] \times [p / (1-p)] = v_p / f_p \quad \text{O sea,}$$

$$\text{Odds a posteriori (+)} = \text{PPV} / (1 - \text{PPV})$$

De donde se puede deducir que la Post-test probabilidad de estar enfermo cuando el resultado del test fue positivo es simplemente el VPP.

Caso negativo: La probabilidad pre-test y Post test se relacionan con:

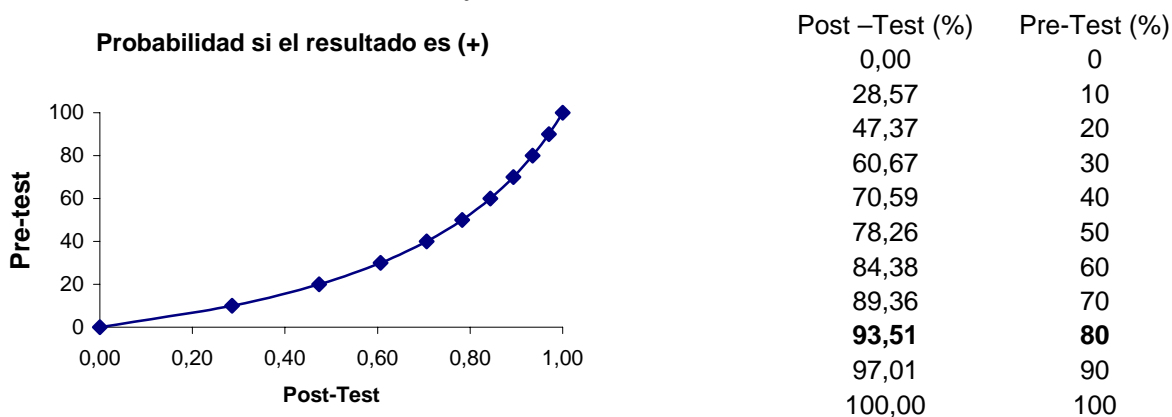
$$\text{Odds a posteriori (-)} = \text{LR-} \times \text{Odds a priori (+)} = [(1 - S) / E] \times [p / (1-p)] = f_n / v_n$$

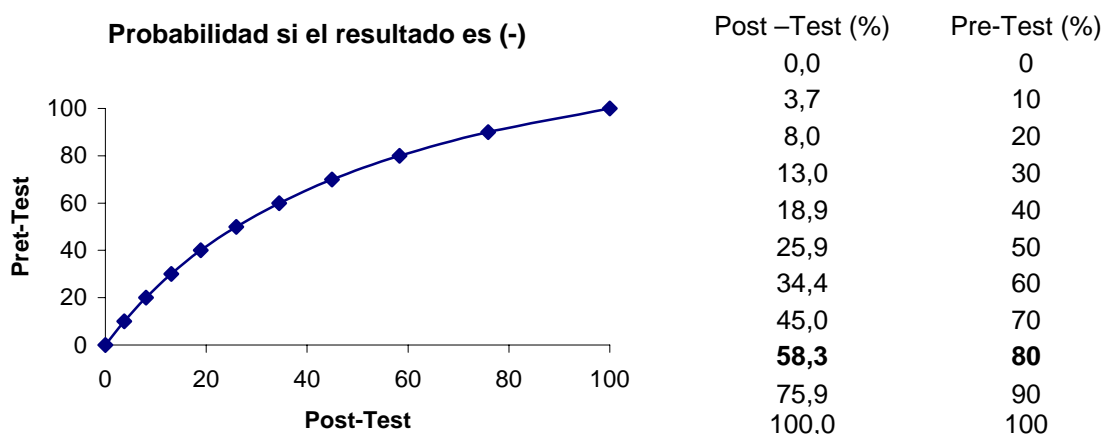
$$\text{Odds a posteriori (-)} = (1 - \text{NPV}) / \text{NPV}$$

De donde se puede deducir que la Post-test probabilidad de estar sano cuando el resultado del test fue negativo es simplemente (1 - NPV).

Entonces, recordando la simulación para verificar experimentalmente el Teorema de Bayes, se pueden graficar las variabilidades de los valores predictivos con una probabilidad cualquiera (tomada como la prevalencia del estudio antes, y que ahora es sencillamente la probabilidad asignada por el clínico al paciente en forma subjetiva)

Gráfico 6.2. Probabilidades Pre-test y Post-test





Se puede observar de estos gráficos, que entrando con la sospecha del clínico del ejemplo anterior (80%) se obtienen directo ambas probabilidades Post-test, sin hacer ninguna cuenta. Para obtener este par de gráficos en forma rápida, la cátedra ha desarrollado un programa en Excel para computadoras, donde entrando con los valores de sensibilidad y especificidad se pueden obtener directamente ambos gráficos, y luego imprimirlos para tenerlos disponibles.

Sin embargo, a pesar de esta nueva simplificación del problema, no es razonable esperar que cada clínico tenga a mano el par de gráficos para decidir si le efectúa la paciente el test. Por lo tanto, se presenta a continuación una manera simplificada de hacer todo esto de manera sencilla, rápida y fácil de recordar:

La idea es jugar con el par de gráficos para derivar criterios clínicos generales, que a modo de guía faciliten la tarea de ver si conviene hacer un test clínico para un caso individual. A modo de ejemplo se puede usar el caso de la ergometría anterior:

Paso 1: Se toma una probabilidad subjetiva alta de que el paciente esté enfermo cuando viene a la consulta, como ser del 80%. Mirando en los gráficos se obtienen las dos probabilidades posteriores al test asociadas de un 93,5% si en realidad está enfermo, y de un 58,3% si estaba sano.

Paso 2: Se toma una probabilidad subjetiva baja respecto a que el paciente esté enfermo, como ser del 20%. Ahora las probabilidades posteriores al test son del 47,4% si estaba enfermo y del 8% si estaba sano.

Paso 3: Se toma una probabilidad subjetiva intermedia, como ser del 50% (esto significa en el fondo que el clínico no tiene una sospecha clara en ningún sentido, sería como el tirar una moneda al aire, o bien usar estos casos para cuando no se tiene mucha idea acerca del estado real del paciente). Mirando en los gráficos se puede obtener una probabilidad Post-test del 78,3% si en realidad estaba enfermo, y del 26% si estaba sano.

De estos 3 pasos se pueden deducir unas reglas generales, a modo de protocolo, para saber como proceder en este tipo de casos. Por ejemplo, las reglas generales pueden ser:

- 1) Si la sospecha de que el paciente está enfermo es alta, no conviene hacerle la ergometría.
- 2) Si la sospecha de que esté enfermo es baja, conviene efectuar la ergometría.
- 3) Si no se tiene una idea clara del estado del paciente conviene hacer el test.

6.6 Estudio de la independendencia en tablas de riesgo

En el primer capítulo se vieron los estudios de riesgo, donde el objeto era ver el efecto de cierto factor (de exposición, de protección, etc.) sobre la condición buscada. Es el típico estudio epidemiológico que se clasificaba en: Caso-control, Cohortes o RCT. En todos ellos se usaba una tabla de 2 x 2 para volcar los datos del estudio como la siguiente:

Tabla 6.2. Tabla de riesgo: datos observados

Factor	Condición observada		Total
	Enfermos	Sanos	
Expuesto	a	b	$a + b$
No expuesto	c	d	$c + d$
Total	$a + c$	$b + d$	N

Como puede verse, se trata de una partición de la población usada en el estudio en cuatro conjuntos básicos, mutuamente independientes y colectivamente exhaustivos. Esta tabla está construida con los datos observados. Haciendo el siguiente supuesto:

- . El factor de riesgo es independiente de la condición observada. O bien,
- . El factor y la condición no están asociados.

Aquí se habla de *asociación estadística*, que no es una asociación del tipo *causa-efecto*, sino que podría llegar a serlo. Es muy importante tener muy en claro esta diferencia, porque es común ver que se confunde una cosa con la otra.

Entonces, bajo ese supuesto se pueden calcular las frecuencias esperadas, usando la teoría de que hay independencia estadística, como se muestra en la tabla siguiente:

Tabla 6.3. Tabla de riesgo: datos esperados teóricamente si hay independencia

Factor	Condición observada		Total
	Enfermos	Sanos	
Expuesto	$\frac{(a+c)(a+b)}{N}$	$\frac{(b+d)(a+b)}{N}$	$a + b$
No expuesto	$\frac{(a+c)(c+d)}{N}$	$\frac{(b+d)(c+d)}{N}$	$c + d$
Total	$a + c$	$b + d$	N

Notar que cada la frecuencia esperada de cada celda, se saca con el producto de sus respectivos totales marginales, dividido el tamaño muestral total N .

La manera de demostrar esto es como sigue, sean los sucesos siguientes:

E : la condición de enfermo.
 S : la condición de sano.

X : la condición de expuesto.
 Y : la condición de no-expuesto

Los conjuntos siguientes son los cuatro casos posibles:

$(E \cap X)$: el conjunto de los enfermos que estuvieron expuestos, cuya frecuencia observada es a
 $(S \cap X)$: el conjunto de los sanos que estuvieron expuestos, cuya frecuencia observada es b
 $(E \cap Y)$: el conjunto de los enfermos que no estuvieron expuestos, con frecuencia observada c
 $(S \cap Y)$: el conjunto de los sanos que no estuvieron expuestos, cuya frecuencia observada es d

Sus probabilidades se pueden estimar con:

$$P(E \cap X) = a^* / N \qquad P(S \cap X) = b^* / N$$

$$P(E \cap Y) = c^* / N \qquad P(S \cap Y) = d^* / N$$

Donde el asterisco denota que son valores teóricos y no observados. Luego, para que haya independencia se tiene que cumplir la condición:

$$P(E \cap X) = P(E) \cdot P(X) = [(a + c) / N] \cdot [(a + b) / N] = a^* / N$$

$$P(E \cap Y) = P(E) \cdot P(Y) = [(a + c) / N] \cdot [(c + d) / N] = c^* / N$$

$$P(S \cap X) = P(S) \cdot P(X) = [(b + d) / N] \cdot [(a + b) / N] = b^* / N$$

$$P(S \cap Y) = P(S) \cdot P(Y) = [(b + d) / N] \cdot [(c + d) / N] = d^* / N$$

Despejando y simplificando se obtienen las frecuencias esperadas de la Tabla 6.3 anterior. Por ejemplo, si las frecuencias observadas fueron:

Factor	Condición observada		Total
	Enfermos	Sanos	
Expuesto	20	80	100
No exp.	50	150	200
Total	70	230	300

Las frecuencias esperadas si hay independencia serán:

Factor	Condición observada		Total
	Enfermos	Sanos	
Expuesto	23,3	76,7	100
No exp.	46,7	153,3	200
Total	70	230	300

6.7 Probabilidad hipergeométrica

Se trata de un caso muy frecuente en los muestreos de aceptación donde se debe decidir si el comprador acepta comprar un lote de mercadería, a partir del análisis de una pequeña muestra del lote. Los profesionales que trabajan en el sector Control de Calidad de la industria farmacéutica, química, cosméticos, alimentos, etc. suelen aplicarla desde otro punto de vista. Allí interesa saber si se da por bueno un lote de producción, a partir del control de una muestra del mismo. O sea, el control no lo hace el consumidor sino el fabricante. En ambos casos, la idea central es la misma, no se pueden revisar todas las piezas fabricadas, ya sea porque llevaría mucho tiempo, o porque el ensayo es destructivo, o porque es muy caro, etc.

Existen Normas IRAM para el control de calidad, que se usan como una solución de compromiso entre los intereses encontrados de comprador y vendedor. Por un lado, el comprador quiere minimizar su riesgo de comprar un lote malo aumentando el tamaño de la muestra a revisar. Y por el otro lado, el vendedor desea minimizar el costo del control revisando la menor cantidad posible de piezas. Entonces, la norma ayuda a definir el tamaño de la muestra a controlar, que es el cuello de botella de toda negociación de esa naturaleza, a partir de riesgos de compra y venta aceptables para ambas partes. Tal riesgo se calcula a partir de la probabilidad hipergeométrica.

Sea un lote de producción de tamaño N que se lleva al laboratorio de Control de Calidad para su revisión, antes de lanzarlo a la venta. El profesional a cargo, sabe por su larga experiencia, el porcentaje de piezas falladas que encontrará al efectuarle una serie de análisis que permitan dos conclusiones: pieza aceptada o pieza rechazada. Multiplicando tal porcentaje de fallas p por el tamaño del lote N , obtiene el número esperado R de piezas rechazadas. De acuerdo a las normas vigentes, se elige una muestra control de tamaño n . ¿ Cuanto vale la probabilidad de encontrar exactamente r piezas falladas en la muestra control ? Tal pregunta se responde con la *probabilidad hipergeométrica* denotada con $P_H(r/n, N, R)$. Como toda probabilidad, esta también resulta el cociente entre el número de casos favorables y el número total de casos posibles.

El numerador son todos los casos donde al revisar una pieza se la rechaza. Hay que hacer un conteo ordenado, pues habrá $C(R, r)$ combinaciones diferentes de encontrar las r piezas falladas de la muestra, dentro de las R piezas totales del lote de producción. Pero a su vez, esto ocurre con cada una de las formas posibles de encontrar la $(n-r)$ piezas buenas, restantes de la muestra, de entre todas las piezas buenas del lote $(N-R)$. O sea, $C(N-R, n-r)$. En síntesis, el número total de casos donde hay rechazo se obtendrá con el producto de ambos valores, esto es: $C(R, r) \cdot C(N-R, n-r)$. El denominador esta dado por todas las combinaciones posibles que existen al tomar n elementos, de un total de N elementos del lote: $C(N, n)$

$$P_H(r/n, N, R) = \frac{C(R, r) \cdot C(N-R, n-r)}{C(N, n)}$$

Ejemplo 1) Se fabricó un lote de 1000 ampollas conteniendo un medicamento liofilizado. Se sabe que un 0,2 % tiene problemas con el cierre hermético del vidrio. Se elige al azar una muestra control de 10 ampollas. ¿ Cuánto vale la probabilidad de encontrar una fallada ?

$$P_H (r=1 / n=10, N=1000, R=2) = C(2, 1) C (998 , 9) / C (1000 , 10)$$

$$P_H (r=1 / n=10, N=1000, R=2) = \frac{2 (998)! (990)! 10!}{(989)! 9! 1000!} = \frac{2 \cdot 998! \cdot 990 \cdot 989! \cdot 10 \cdot 9!}{989! \cdot 9! \cdot 1000 \cdot 999 \cdot 998!}$$

$$P_H (r=1 / n=10, N=1000, R=2) = 0,0198$$

La probabilidad de no encontrar ninguna fallada será

$$P_H (r=0 / n=10, N=1000, R=2) = C(2, 0) \cdot C (998 , 10) / C (1000 , 10) = 0,98$$

En síntesis, hay un 98% de probabilidades de no encontrar ampollas falladas y 1,98% de encontrar una sola fallada. Entonces habrá: (100-98-1,98)% =0,02% de encontrar dos o más falladas.

Ejemplo 2) El mismo caso de sacar tres barajas de espadas del mazo, visto antes y resuelto con el concepto de probabilidad condicional, ahora se puede resolver con:

$$P_H (r=3 / n=3, N=40, R=10) = C(10, 3) \cdot C (30 , 0) / C (40 , 3) = \frac{10! 37! 3!}{40! 7! 3!}$$

$$P_H (r=3 / n=3, N=40, R=10) = (10 \cdot 9 \cdot 8) / (40 \cdot 39 \cdot 38) = 0,012$$

Ejemplos 3) Un médico suministra un fármaco a siete enfermos de similares características. Seis de ellos experimentan una gran mejoría, en cambio el restante permanece igual. Por otra parte, durante todo el tiempo transcurrido se ha controlado a un grupo de cinco pacientes, de características similares a los otros sometidos al tratamiento. La diferencia es que, a este grupo control se le ha suministrado un placebo. Resultó que sólo uno mejoró, mientras los otros cuatro permanecieron iguales. El problema es calcular si la mejoría observada en el grupo tratado, se debe al azar o a la acción del medicamento.

La probabilidad hipergeométrica puede emplearse en estos casos para contestar tal pregunta. En efecto, tomando en conjunto total de pacientes (N=12), de los cuales (n=7) fueron tratados; los resultados finales fueron (R=7) han mejorado y entre éstos (r=6) fueron los medicados, mientras que al restante (R-r =1) se le había suministrado un placebo inocuo para tratar la enfermedad. Se puede presentar lo mismo en forma esquemática con el diagrama del Gráfico 6.3. Asumiendo que los pacientes fueron distribuidos al azar entre ambos grupos, para evitar cualquier tipo de tendencia, y tratando de mantener todas sus características lo más parecidas posibles para no introducir fluctuaciones extrañas. Se puede asumir, que se trata de un problema de “ n extracciones al azar sin reemplazamiento, de un grupo de tamaño N”.

Gráfico 6.2 Diagrama esquemático para el problema 3.

Con mejoría	Sin mejoría	Total
-------------	-------------	-------

Con fármaco (medicados)	6 = r	1	7 = n
Sin fármaco (placebo)	1	4	5
Total	7 = R	5	12 = N

Entonces, la probabilidad de ocurrencia por azar en una tabla del tipo 2 x 2 como las vistas en índices clínicos, de los sucesos esquematizados en el diagrama, se calcula con:

$$P_H (r=6 / n=7, N=12, R=7) = C(7, 6) \cdot C(5, 1) / C(12, 7) = 0,0442$$

La probabilidad de que no ocurra por azar, sino por la acción del medicamento será el complemento de la anterior, o sea 0,9558. Se tiene entonces un 95,58% de probabilidad de que el medicamento (o tratamiento) sea eficaz y un 4,42% que no lo sea. Se puede afirmar que es muy improbable, que la mejoría de los pacientes sea casual. Sin embargo, todavía esto no es una *prueba científica* con evidencia estadística. Para que así ocurra, se deben plantear tests de hipótesis como se verá en capítulos posteriores.

6.8 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|--|-------|-------|
| 1) La regla del producto se puede aplicar en todos los casos. | V | F |
| 2) La probabilidad de sacar tres caras al lanzar al aire tres monedas es 1/8. | V | F |
| 3) Dos sucesos son independientes si la ocurrencia de uno no afecta al otro. | V | F |
| 4) Dos sucesos independientes no tienen ningún elemento en común. | V | F |
| 5) Si tres sucesos son independientes, también lo son todos los pares posibles. | V | F |
| 6) Si todos los pares posibles son independientes, también lo son si se toman de a tres. | V | F |
| 7) Dos sucesos son excluyentes si no tienen ningún elemento en común. | V | F |
| 8) Es lo mismo el concepto de independencia que el de exclusión. | V | F |
| 9) En sucesos condicionales, la probabilidad de uno se afecta si se sabe que ocurrió el otro. | V | F |
| 10) Exclusión e independencia, son casos particulares de la Condicionalidad. | V | F |
| 11) El Teorema de Bayes se puede aplicar si un suceso, es particionado por otros. | V | F |
| 12) Ídem para el Teorema de Probabilidad Total. | V | F |
| 13) Los Índices Clínicos se relacionan entre sí, a través del Teorema de Bayes. | V | F |
| 14) Algunos Índices Clínicos se pueden interpretar como probabilidades condicionales. | V | F |
| 15) En los muestreos de aceptación se emplea la probabilidad Hipergeométrica. | V | F |
| 16) La Normas IRAM el Control de Calidad aplican la probabilidad Hipergeométrica. | V | F |
| 17) Si el producto de las probabilidades de dos sucesos es igual a la conjunta, son | | |
| 18) Si la probabilidad condicional de dos sucesos es igual a la del producto de ambas, hay.... | | |
| 19) El diagrama del árbol se usa para visualizar | | |
| 20) La sensibilidad es la capacidad de un test de detectar como positivos a los enfermos | V | F |
| 21) La especificidad es la capacidad de un test de detectar como negativos a los sanos | V | F |
| 22) Sensibilidad, especificidad, Youden y los Likelihood ratio no varían con la prevalencia. | V | F |
| 23) Se puede mostrar experimentalmente lo anterior. | V | F |

- 24) La eficiencia y los valores predictivos varían con la prevalencia. **V F**
 25) La forma de extraer las muestras de la población altera a los índices clínicos. **V F**
 26) Para un estudio por cohortes se hace un muestreo basándose en la exposición. **V F**
 27) Para un estudio caso-control se hace un muestreo basándose en la condición buscada. **V F**
 28) Explicar como hacer un experimento, el estilo Bernoulli, para verificar lo anterior
 29) Expresar el Teorema de Bayes en forma más simplificada.
 30) No hay relación entre el teorema de probabilidad total y el de Bayes. **V F**
 31) Las probabilidades Pre-test se relacionan con las Post-test. **V F**
 32) Los Odds a priori multiplicados por el LR+ dan los Odds a posteriori en sanos. **V F**
 33) Se pueden transformar las probabilidades en Odds y viceversa. **V F**
 34) La probabilidad a posteriori en el caso de enfermos es igual al VPP. **V F**
 35) La probabilidad a posteriori en el caso de sanos es igual al VPN. **V F**
 36) Con dos gráficos se pueden resumir los 8 pasos recomendados por JAMA para Post-Test. **V F**
 37) Es mejor diseñar un protocolo general que usar los gráficos en cada paciente. **V F**
 38) La independencia estadística significa:
 39) En una tabla de riesgo la asociación estadística significa:
 40) Si hay asociación estadística, hay una relación del tipo causa-efecto. **V F**
 41) Si hay una relación causa-efecto, tiene que haber una asociación estadística. **V F**
 42) Los valores de frecuencias esperadas, bajo el supuesto de independencia, se calculan con:
 43) El producto de los totales marginales dividido N es la frecuencia esperada de una celda. **V F**
 44) Las tablas de independencia 2 x 2 son un caso general de la tabla de riesgo. **V F**
 45) Las tablas de riesgo pueden llegar a considerarse como tablas diagnósticas **V F**

2) La clasificación por edades de los profesionales de una zona de salud es:

Servicios	Menor de 25	Entre 26 y 30	Entre 31 y 35	Mayor de 35	Total
Médicos	0	5	25	75	105
Bioquímicos	20	30	35	35	120
Alimentación	3	6	6	10	25
Registración	7	15	8	12	42
Enfermeros	200	375	442	203	1220
Farmacéuticos	1	12	8	3	24
Radiología	4	10	19	12	45
Terapeutas	5	25	15	10	55
Otros servicios	20	35	50	25	130
Total	260	513	608	385	1766

a) Calcular las probabilidades para cada uno de los cuatro grupos de edades.

b) Lo mismo para los nueve tipos de servicios.

c) Si se escoge una persona al azar. Averiguar la probabilidad de que:

- Sea un médico de más de 35 años
- Sea un enfermero menor de 25 años.
- Sea un farmacéutico de 28 años.

c1) Sabiendo que tiene 33 años, cuanto vale la probabilidad de que sea:

- de Radiología
- de otros servicios

c2) Sabiendo que es un bioquímico, cuanto vale la probabilidad de que tenga:

- más de 35 años.
- 30 años.
- 21 años

d) Si se escogen dos personas al azar, cuanto vale la probabilidad de que:

- Uno sea médico y el otro bioquímico.

- Los dos sean farmacéuticos

e) Verificar el Teorema de Probabilidad Total en los puntos (a) y (b)

3) Una compra en una farmacia fue de 7 tabletas de aspirinas y 3 paquetes de algodón. Si se saca de la bolsa de a uno, averiguar la probabilidad de que se saquen dos tabletas y la tercera sea un algodón.

4) Calcular los índices clínicos para una prueba de laboratorio, que en un estudio de 400 personas arrojó los resultados siguientes:

Test Clínico	Enfermedad		Total
	Si	No	
(+)	30	190	220
(-)	20	160	180
Total	50	350	400

Con los valores obtenidos de S y E, simular una prevalencia variable entre $p = 0,1; 0,2; \dots; 0,9$ para obtener las curvas que muestren la variabilidad de la eficiencia y los valores predictivos con la prevalencia. Verificar si se cumplen las relaciones teóricas expuestas en el cuadro resumen 5.2.

5) Resolver la siguiente sopa de letras:

B	R	N	O	R	A	L	I	N	A	C	O	N	H	D	O	X	A	G	U	E	S	W	O	U	M	T	R	A	X
X	B	S	C	H	Y	N	A	S	V	A	C	X	Q	I	R	F	O	I	T	A	L	Q	U	P	Y	O	L	X	S
E	F	Y	O	J	L	C	H	T	M	V	B	C	W	A	E	X	C	L	U	Y	E	N	T	E	S	X	Ñ	L	A
R	C	B	N	I	V	N	U	I	Y	R	A	I	C	N	E	U	C	E	R	F	R	E	S	A	X	P	L	S	R
O	V	R	D	D	M	H	R	J	N	M	T	R	W	A	D	E	S	L	A	X	V	N	R	O	D	E	D	Y	T
J	W	O	I	E	L	P	S	U	R	E	S	T	O	C	A	S	T	I	C	O	Ñ	K	D	P	I	S	M	S	A
I	T	E	C	Q	C	X	O	T	W	D	F	Q	B	F	J	L	S	H	L	J	P	F	I	I	Q	E	D	I	V
O	S	S	I	S	U	C	C	R	F	N	T	O	T	N	V	A	C	E	C	Z	O	D	E	C	A	T	A	L	S
U	I	T	O	O	I	S	I	E	P	S	I	P	Y	Q	C	C	N	L	I	C	A	S	O	T	T	N	D	S	Ñ
G	S	A	N	B	N	V	T	O	I	O	M	R	K	R	X	C	Z	A	F	T	I	T	T	O	P	E	I	K	N
N	D	L	A	D	R	E	S	T	R	U	T	O	R	E	S	N	I	O	I	I	N	S	I	O	A	I	S	L	E
A	O	A	L	C	U	S	O	I	T	R	I	D	Z	F	R	E	D	M	U	R	R	M	D	A	D	D	E	V	S
M	S	A	S	M	Z	X	N	R	D	I	N	U	L	E	S	S	C	V	I	G	O	I	A	L	O	N	C	O	T
B	K	U	S	S	R	O	G	T	W	S	W	C	L	R	A	L	A	S	O	N	O	L	L	A	R	E	S	C	G
E	A	J	Ñ	M	G	P	A	O	B	R	I	T	E	O	R	E	M	A	R	I	B	L	O	S	I	P	E	S	U
I	G	Y	I	N	A	N	I	A	S	A	D	O	O	U	M	I	L	A	A	D	M	I	S	T	R	E	U	I	Z
D	U	G	E	K	A	Ñ	D	C	I	N	A	A	U	L	A	D	I	M	U	A	U	B	Ñ	T	A	D	M	V	I
A	S	R	I	S	B	D	G	O	J	C	E	Ñ	F	U	L	S	I	R	A	R	S	O	T	A	R	N	S	E	R
N	Q	U	A	R	T	H	I	P	E	R	G	E	O	M	E	T	R	I	C	A	Z	T	O	R	T	I	P	L	A
K	I	B	W	D	A	D	I	L	I	B	I	S	N	E	S	N	G	H	D	A	Ñ	O	S	C	A	R	I	O	D

a) El tipo de probabilidad más general de todas.

b) Teorema de es el que permite relacionar los índices clínicos entre sí.

c) La probabilidad usada en los muestreos de aceptación.

- d) Dos sucesos cuya probabilidad conjunta es igual al producto de sus probabilidades individuales.
- e) Dos sucesos cuya probabilidad conjunta es nula.
- f) Proceso conformado por una sucesión finita de experimentos, cada uno con probabilidad finita.
- g) La regla del permite calcular la probabilidad conjunta de dos sucesos independientes.
- h) de Probabilidad Total se aplica cuando un grupo de sucesos particiona a otro.
- i) Conclusiones de los médicos cuando examinan a un paciente.
- j) La probabilidad de sacar cara es la de la probabilidad del suceso seguro.

6) Usando los datos del ejercicio (4) anterior, e imaginándolos como las frecuencias observadas en el experimento, se pide:

- a) Calcular las 4 frecuencias esperadas teóricamente si hay independencia total entre el test clínico y el hecho de que exista o no la enfermedad.
- b) Recalcular los índices clínicos obtenidos en el ejercicio 5 con los datos teóricos bajo el supuesto de independencia.
- c) Discutir los resultados obtenidos y explicar como se puede relacionar la independencia estadística con la “independencia clínica”.

7) En un estudio epidemiológico se determinaron los datos siguientes:

		Enfermedad		
		Si	No	Total
Exposición a una fuente de contagio	Si	50	150	200
	No	50	150	200
Total		100	300	400

- a) Haciendo el supuesto de que la enfermedad no tiene nada que ver con la fuente de contagio (también llamado factor de riesgo), esto es que son estadísticamente independientes. Calcular las cuatro frecuencias esperadas.
- b) Calcular la Prevalencia en el caso de frecuencias esperadas y también en el caso de las frecuencias observadas. ¿ Qué conclusión puede sacarse ?
- c) Calcular el odd de los enfermos y de los que no enfermaron entre los pacientes expuestos.
- d) Calcular el odd de los enfermos y de los no enfermos entre los pacientes no expuestos.

8) Se pide repetir el ejemplo anterior pero con estos nuevos datos:

Enfermedad		
Si	No	Total

Exposición a una fuente de contagio	Si	150	50	200
	No	150	50	200
	Total	300	100	400

9) Comparar los resultados de los ejercicios 7 y 8 para sacar algunas conclusiones entre la independencia estadística y la realidad.

10) Explicar la diferencia entre variabilidad de los índices clínicos debidas a la prevalencia y la debida a la elección de un punto de corte en un método clínico.

11) Se efectuaron dos estudios epidemiológicos con un total de 400 individuos. De entre todos los casos disponibles, se seleccionaron 200 enfermos al azar y otros tantos sanos. Efectuando un estudio retrospectivo se separaron los casos que estuvieron expuestos a la fuente de contagio de los otros, y las frecuencias encontradas fueron las siguientes:

		Enfermedad		
		Si	No	Total
Exposición a una fuente de contagio	Si	50	50	100
	No	150	150	300
	Total	200	200	400

Se pide calcular:

- a) El Odds Ratio
- b) El riesgo relativo

12) En un estudio similar al anterior se encontraron los datos siguientes:

		Enfermedad		
		Si	No	Total
Exposición a una fuente de contagio	Si	150	150	300
	No	50	50	100
	Total	200	200	400

Se pide calcular los índices OR y RR y compararlos con el caso anterior.

7

Pruebas repetidas

En este capítulo se tratan dos tipos de procesos básicos en pruebas repetidas. El primero es el estudiado por Bernoulli en Ginebra, a fines del siglo XVI, aplicado a los juegos de azar. Condujo al modelo de la Probabilidad Binomial, Binomial Negativa, Probabilidad de Pascal y la Probabilidad Multinomial. El segundo fue desarrollado por Poisson en París, en los inicios del siglo XVIII, como un caso particular de la binomial aplicado a los casos raros. La Probabilidad de Poisson se aplica en las técnicas de recuentos celulares, fenómenos de contagio, virología, como así también en la Ley de Radiactividad General, seguros de vida, teoría de las líneas de espera o teoría de colas, etc.

7.1 Procesos tipo Bernoulli

Se denominan procesos de tipo Bernoulli, a todo experimento consistente en una serie de pruebas repetidas, caracterizadas por tener resultados que se pueden clasificar en si verifican o no cierta propiedad o atributo, siendo aleatorios e independientes. La relación con un proceso de tipo hipergeométrico, visto en el capítulo anterior, es directa: Si las extracciones se hacen *con reposición* se transforma en un proceso del tipo Bernoulli.

Para identificar un proceso Bernoulli en una serie de pruebas repetidas, se deben verificar tres condiciones:

1) *Resultados dicotómicos*: Los resultados de cada prueba se pueden clasificar en “éxito” si verifican cierta condición, o “fracaso” en el caso contrario. No importa el tipo de magnitud clínica que se trate; al efectuarle la determinación en el laboratorio, siempre se podrá ver si el resultado obtenido en la muestra del paciente verifica o no cierto atributo. Y entonces, siempre se podrán clasificar los resultados en forma dicotómica. En el caso de la determinación de la CPK -que es una magnitud de tipo continua- al tomar un punto de corte se clasifican los infinitos resultados posibles en dos: Infartado - No infartado. En una determinación del fenómeno Psi, el recuento de resultados -que es una magnitud de tipo discreta- siempre se podrá clasificar en: hay evidencia de percepción extrasensorial, o no la hay. En un concurso de belleza, o el resultado de una carrera (magnitudes ordinales), se puede subdividir en: ganó-perdió.

2) *Independencia de las pruebas*: El resultado de una prueba cualquiera es independiente del resultado obtenido en la prueba anterior, y no incide en el resultado de la prueba siguiente. Esto se cumple en los juegos de azar como al lanzar una moneda o un dado, y en la extracción de barajas, si se realizan con reposición luego de cada extracción. Es uno de los supuestos básicos de la Teoría Genética de Mendel y en las leyes de la herencia humana, como el sexo de un recién nacido. La independencia deja de cumplirse cuando se da un fenómeno de contagio o repulsión.

3) *Estabilidad de las pruebas*: La probabilidad p de obtener un resultado considerado como un éxito se mantiene constante a lo largo de toda la serie de pruebas. Análogamente, para la probabilidad q de obtener un fracaso. Como los resultados son dicotómicos, el universo de resultados aparece particionado por ambos sucesos, entonces $p + q = 1 = P(S)$.

La tirada de una moneda o de un dado es el ejemplo clásico de este proceso. Se debe notar que eso no cambia si la moneda o el dado están viciados. Si un dado cargado tiene una probabilidad de $2/6$ de sacar un as, como esa probabilidad se mantiene constante e independiente a lo largo de la serie de pruebas, se cumplen las condiciones anteriores.

En cambio, si se da un fenómeno de reacción en cadena, o en cascada, o contagio, las probabilidades dejan de ser constantes. Por ejemplo, si la probabilidad de enfermarse de hepatitis en Misiones es p , y se presenta una epidemia, entonces el contagio hace crecer mucho el número de enfermos, p cambia y deja de ser estable. El corrimiento de los puntos de una media de nylon femenina es un ejemplo de fenómeno en cascada. Para los últimos casos, se podrá emplear otro modelo, el de Poisson, que se verá más adelante. Notar que el Odds de éxitos en el modelo Bernoulli es $p / q = p / (1-p)$.

7.2 Probabilidad Binomial

Cuando en un proceso del tipo Bernoulli se desea saber la probabilidad de obtener exactamente r éxitos, en una serie de n pruebas, con una probabilidad de éxito p , se puede aplicar la fórmula de la probabilidad binomial:

$$P_B (r / n , p) = C (n , r) \cdot p^r (1 - p)^{n-r}$$

Sea S_1 la primera serie de n pruebas realizadas donde se verifica la ocurrencia de r éxitos, con una probabilidad p y la de fracaso q :

S_1 : $E F E F F \dots E$ (donde E indica la ocurrencia de un éxito y F la de un fracaso)

$P(S_1) = P(E \cap F \cap E \cap F \cap F \cap \dots \cap E)$ y como los sucesos son independientes

$$P(S_1) = P(E) \cdot P(F) \cdot P(E) \cdot P(F) \cdot P(F) \dots P(E) = p \cdot q \cdot p \cdot q \cdot q \dots p = p^r (1 - p)^{n-r}$$

Pero esta secuencia, es una de todas las k posibles series donde se verifican r éxitos en n pruebas. Otra serie cualquiera como la S_2 : $E F E F F \dots F$

$$P(S_2) = P(E \cap F \cap E \cap F \cap F \cap \dots \cap F) = P(E) \cdot P(F) \cdot P(E) \cdot P(F) \cdot P(F) \dots P(F)$$

$$P(S_2) = p \cdot q \cdot p \cdot q \cdot q \dots q = p^r (1 - p)^{n-r}$$

En esta segunda serie, diferente de la primera en el orden de ocurrencia de los sucesos, se obtiene la misma probabilidad. Generalizando, para una serie cualquiera i es: $P(S_i) = p^r (1 - p)^{n-r}$

El número k total de pruebas posibles se obtiene con la combinatoria de n elementos tomados de a r . O sea: $C(n, r)$. La probabilidad binomial pedida será la probabilidad de la unión de todos los sucesos posibles, esto es :

$$P_B (r / n , p) = P (S_1 \cup S_2 \cup S_3 \cup \dots \cup S_k) = P (\cup_k S_k) = \sum_k P (S_k) = k \cdot P (S_k)$$

Y reemplazando por las relaciones anteriores, se llega a la fórmula de la Binomial de más arriba

Ejemplo 1) Sea el caso de una droga X, con una dosis mortal de 1g/100 ml para cobayos experimentales, en el 25% de los casos. Aplicando esta dosis a cien cobayos se desea saber cuanto vale la probabilidad de que mueran veinte de ellos.

Esto se calcula con la binomial, siempre y cuando se cumplan los supuestos básicos, a saber:

- 1) Los cobayos mueren (éxito) o sobreviven (fracaso).
- 2) Que un cobayo muera con la dosis, no significa que lo hará el siguiente (independencia) pues no se trata de una epidemia.
- 3) La probabilidad de que mueran se mantiene constante a lo largo de la serie de pruebas ($p = 0,25$). Entonces, aplicando la fórmula será:

$$P_B (r=20 / n=100 , p=0,25) = C (100 , 20) \cdot (0,25)^{20} (1 - 0,25)^{80} = 0,04902$$

Ejemplo 2) En el Cuadro 7.1 se presenta un ejemplo de proporción de sexos la descendencia de parejas humanas; los datos provienen de un estudio de Geissler (1889), quien analizó 6.115 familias de doce hijos cada una en Sajonia. La proporción de sexos es 1 : 1 aproximadamente. Sin embargo, como se sabe que en poblaciones humanas diferentes estos pueden variar un poco, para este trabajo se tomó la proporción obtenida en el mismo. La proporción de 519 : 481 entre hombres y mujeres. Si el número de hijos es $n = 12$, la cantidad r de mujeres, puede ser imaginada como una variable que va de 0 a 12, ya para cada valor de r habrá un valor de la binomial que se muestra en la columna $P_B (r)$ del Cuadro 7.1, calculado con :

$$P_B (r=0 / n=12 , p=0,481) = C (12 , 0) \cdot (0,481)^0 (0,519)^{12} = 0,00038$$

Eso implica que en 6.115 familias de 12 hijos, la probabilidad de encontrar familias sin hijas mujeres es de 0,00038, esto es, una frecuencia esperada de 2,3 familias del total. Los valores esperados se obtienen multiplicando el número total de familias estudiadas por la frecuencia esperada. Esto es, $6.115 P_B (r = 12) = 2,3$. Geissler obtuvo 7 casos reales, más del doble de lo esperado. En el caso inverso, es decir familias de 12 hijas, da : $6.115 P_B (r = 0) = 0,9$ familias es el número de casos esperado. Pero hubo 3 casos reales, casi el triple. Si se comparan los valores esperados con los observados, se nota una buena concordancia en los valores del centro de la tabla, pero en las puntas eso deja de verse. Por ejemplo, para el caso de 6 mujeres y 6 varones, se observan 1.343 casos y se esperaban 1.367,3 o sea una diferencia del orden del 1,8%; en cambio, en ambas puntas ronda el 100%. La base genética de este fenómeno no está muy clara. Pero hay familias que “tienden” a tener todos varones y otras a mujeres. Todo ocurre como si hubiese una especie de “contagio”. Para tener una prueba científica de esto, hay que realizar un *test estadístico* para hacer una prueba de hipótesis, como se verá en capítulos posteriores. En muchos casos de estudios mendelianos, la proporción de sexos se calcula con la relación 1:1; en este caso no se hizo así, sino que se usaron los *datos experimentales* para su obtención, en lugar de los *teóricos*. Más adelante, se verá la importancia de esta diferencia.

Cuadro 7.1 Aplicaciones de la Probabilidad Binomial.

Caso 1 : Proporción de sexos en 6.115 familias con doce hijos de Sajonia -datos tomados de Sokal(*)

Cantidad de mujeres en familias de 12 hijos r	Probabilidad Binomial $P_B (r)$	Número de casos esperados $E_r = N \cdot P_B (r)$	Número de casos observados Or	Diferencias relativas (Or - Er)/Er
0	0,000384	2,3	7	+ 2,04
1	0,004264	26,1	45	+ 0,724
2	0,021725	132,8	181	+ 0,363
3	0,067041	410,0	478	+ 0,166
4	0,139703	854,3	829	- 0,030
5	0,206973	1265,6	1112	- 0,121
6	0,223590	1367,3	1343	- 0,018
7	0,177459	1085,2	1033	- 0,048
8	0,102708	628,1	670	+ 0,066
9	0,042280	258,5	286	+ 0,106
10	0,011743	71,8	104	+ 0,448
11	0,001975	12,1	24	+ 0,164
12	0,000153	0,9	3	+ 2,333

Ejemplo 3) Se toman al azar 1000 muestras de $n = 5$ individuos cada una, de la especie Homo Sapiens. Se formula la hipótesis de un 40% de ellos presenta cierta enfermedad, no contagiosa. Con computadora se simulan los casos observados usando la tabla de números al azar - (*)

Cantidad de enfermos por muestra r	Probabilidad Binomial $P_B (r)$	Número de casos esperados $E_r = N \cdot P_B (r)$	Número de casos simulados Or	Diferencias relativas (Or - Er)/Er
0	0,07776	388,8	416	+ 0,0694
1	0,25920	1296,0	1327	+ 0,0238
2	0,34560	1728,0	1686	- 0,0243
3	0,23040	1152,0	1110	- 0,0365
4	0,07680	384,0	406	+ 0,0586
5	0,01024	51,2	55	+ 0,0703

En este ejemplo, se simula la población humana, con la tabla de números al azar, haciendo $N=1000$ extracciones de 5 números cada una; como no hay límite para hacer esto, el tamaño teórico de la población es infinito. Cuando en la muestra aparecen los números 1,2,3 o 4 se toma como infectada. Cuando aparecen los números 0,5,6,7,8 o 9 se la toma como no infectada. Por ejemplo, si en una de las muestras simuladas resulta (3,8,9,6,0) se la cuenta como una muestra con 1 infectado. Estos datos se vuelcan en la columna correspondiente a frecuencias observadas. En la primer columna se calcula la probabilidad binomial con un número de éxitos que varía de 0 a 5, un tamaño de muestra de $n = 5$, y una probabilidad $p = 0,4$. Esta porque hay diez dígitos posibles en la tabla de números al azar, de los cuales 4 se usan para los éxitos. Como puede verse hay una buena concordancia (ninguna diferencia relativa supera el 7%) entre los valores esperados y los

observados. Se cumple que cada muestra es independiente de la otra y que la probabilidad es constante a lo largo de la serie de pruebas.

Ejemplo 4) En una farmacia se ha calculado la probabilidad de venderle a un cliente con obra social es del 20%. Se eligen al azar 15 clientes de ese tipo que ingresan al negocio y se desea calcular la probabilidad de concretar menos de tres ventas.

Para analizar este problema conviene determinar todos los casos posibles. Menos de 3 ventas significa que se pueden concretar dos, una o ninguna venta, es decir los casos posibles son:

A : no se vendió nada ($r = 0$)

B : solo se hizo una venta ($r = 1$)

C : se concretaron dos ventas ($r = 2$)

Estos sucesos son excluyentes entre sí, por lo tanto aplicando el Axioma 3 generalizado será:

$$P(\text{menos de 3 ventas}) = P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

$$P(A) = P_B (r=0 / n=15, p=0,2) = C(15, 0) \cdot (0,2)^0 (0,8)^{15} = 0,0352$$

$$P(B) = P_B (r=1 / n=15, p=0,2) = C(15, 1) \cdot (0,2)^1 (0,8)^{14} = 0,1319$$

$$P(C) = P_B (r=2 / n=15, p=0,2) = C(15, 2) \cdot (0,2)^2 (0,8)^{13} = 0,2309$$

Sumando los tres valores anteriores se obtiene $P(\text{menos de 3 ventas}) = 0,398$. Es decir, hay casi un 61% de probabilidades de que se concreten 3 ventas o más.

7.2.1 Contagio y repulsión

Para aplicar la fórmula de la función binomial se deben verificar los tres supuestos de un proceso Bernoulli. El ejemplo clásico es en los juegos de azar. Recordando el problema de lanzar tres veces una moneda, resuelto con el diagrama del árbol en el punto 6.1, puede verse que es mucho más elegante y simple resolverlo con la binomial. En efecto, la probabilidad de sacar tres caras (o secas) seguidas será:

$$P_B (r=3 / n=3, p=0,5) = C(3, 3) \cdot (0,5)^3 (0,5)^0 = (0,5)^3 = 0,125$$

Naturalmente, si en lugar de ser tres, fueran cien lanzamientos, la resolución usando el Diagrama del árbol sería sumamente engorrosa y es más práctico usar este modelo.

En los lanzamientos de una moneda, es muy claro el cumplimiento de los tres supuestos, pero en los casos de la vida real, ya no lo son tanto. Si se piensa en el ejemplo visto cuando se simuló por computadora el número de casos observados usando una tabla de números al azar, los valores necesariamente son aleatorios e independientes. Pero cuando dejan de cumplirse uno o más de estos supuestos, los resultados varían mucho. Sea el mismo caso anterior, pero ahora se le obliga a la computadora a seguir buscando en la tabla, cada vez que salga el valor 3, hasta que encuentre otro de los cuatro números correspondientes al suceso: infectado. Entonces, las proba-

bilidades de cada número dejan de ser iguales, no son más constantes. Luego de efectuar un cierto número de sorteos aumentará la frecuencia de las clases con dos o más enfermos en la muestra y disminuirá el número de clases con un solo enfermo. Todo ocurre como si, cuanto más enfermos hay en la muestra, más fácil se enferman sus vecinos. O sea, se ha simulado un fenómeno de *contagio*. En cambio, si se arregla el programa de la computadora, para que sean más frecuentes los casos heterogéneos, es decir, con similar proporción de infectados, disminuyen los casos más homogéneos (todos sanos, o todos infectados) y se simula un fenómeno de *repulsión*.

Cuadro 7.2 Contagio y repulsión

Cantidad de enfermos por muestra r	Probabilidad Binomial $P_B (r)$	Número de casos esperados $E_r = N \cdot P_B (r)$	Número de casos simulados para <i>contagio</i>	Número de casos simulados para <i>repulsión</i>
0	0,07776	388,8	464	295
1	0,25920	1296,0	1450	1275
2	0,34560	1728,0	1368	1946
3	0,23040	1152,0	1160	1131
4	0,07680	384,0	488	324
5	0,01024	51,2	70	29

En el Cuadro 7.2 se muestran los resultados de la simulación. Las frecuencias de *contagio* aumentan mucho para los valores extremos de r y disminuyen para r=2. Lo contrario ocurre en la *repulsión*, con dos enfermos por muestra la frecuencia sube y baja en los dos extremos. La forma correcta de verificar si estas diferencias son debidas al efecto de contagio o repulsión, es con una prueba o test de hipótesis estadístico, que se verá más adelante.

Si en vez de ser un caso simulado, fuese real, el significado de un *contagio* es que cada vez se encuentra más muestras totalmente infectadas o totalmente sanas, y menos muestras con igualdad de sanos y enfermos. Todo esto, con respecto a lo que cabría esperar si la probabilidad de contraer la enfermedad fuese *independiente*. El contagio es un fenómeno entre los integrantes de cada muestra. O sea, si uno de ellos está enfermo, la probabilidad de enfermarse aumenta en sus “vecinos” de la muestra. Cabe la posibilidad de encontrarse con este fenómeno en la práctica debido a una mala elección de las muestras, porque el investigador tiene tendencia a tomar muestras infectadas. Es decir, no hay un contagio sino una selección mal hecha. Pero si se hace correctamente, con un muestreo al azar, entonces los resultados muestran contagio. Por ejemplo, puede haber ocurrido que se hayan expuesto a un mismo foco infeccioso, o que se hayan contagiado en el intervalo de tiempo entre que fueron extraídos y revisados.

El fenómeno opuesto, ya es más complicado para explicarlo clínicamente. El número de grupos homogéneos ha disminuido sensiblemente, aumentando aquellos grupos con proporciones similares de enfermos y no enfermos. Todo ocurre como si, al estar dos o tres individuos de la muestra enfermos, los restantes son menos propensos a estarlo. Como si los enfermos de alguna manera inmunizaran a sus compañeros muestrales. Se da una especie de *repulsión*. Una interpretación es que hay un número limitado de gérmenes patógenos y una vez que se ubicaron en sus respectivos huéspedes, se agotó la fuente de infección. En infecciones del tipo microbianas, esto es menos plausible, pero muy razonable en el caso de parásitos. Otra cosa que puede haber ocurrido es que con los primeros enfermos se haya vacunado a sus vecinos y de allí los resultados.

7.3 Probabilidad Pascal

Cuando en un proceso Bernoulli, se busca la probabilidad de que sean necesarias exactamente n pruebas, para lograr r éxitos, con una probabilidad de éxito p , entonces se recurre a la fórmula de Pascal, para resolver el problema.

$$P_{Pa} (n / r , p) = C (n-1 , r-1) \cdot p^r (1 - p)^{n-r}$$

Sea S_1 la primera serie de n pruebas realizadas donde se verifica la ocurrencia del r -avo éxito, en la prueba número n , con una probabilidad p y la de fracaso q :

S_1 : $E F E F F \dots F E$ (donde el último debe ser un E)

$$P (S_1) = P (E \cap F \cap E \cap F \cap F \cap \dots \cap F \cap E) = P (R_1 \cap E) = P (R_1) \cdot P (E)$$

Donde R_1 es el suceso : Obtención de $r-1$ éxitos en las primeras $n-1$ pruebas realizadas dentro de la primer serie de pruebas S_1 . Como los sucesos son independientes se obtiene

$$P (S_1) = P (E) \cdot P (F) \cdot P (E) \cdot P (F) \cdot P (F) \dots P (F) \cdot P (E) = p \cdot q \cdot p \cdot q \cdot q \dots p = p^r (1 - p)^{n-r}$$

Pero esta secuencia es una de todas las m posibles series donde se verifican r éxitos en n pruebas. Otra serie cualquiera como la S_2 : $F F E F F \dots E$

$$P (S_2) = P (F \cap F \cap E \cap F \cap F \cap \dots \cap E) = P (F) \cdot P (F) \cdot P (E) \cdot P (F) \cdot P (F) \dots P (E)$$

$$P (S_2) = P (R_2 \cap E) = P (R_2) \cdot P (E) = p \cdot q \cdot p \cdot q \cdot q \dots q = p^r (1 - p)^{n-r}$$

En esta segunda serie, diferente de la primera en el orden de ocurrencia de los sucesos, se obtiene la misma probabilidad. Generalizando, para una serie i es :

$$P (S_i) = p^r (1 - p)^{n-r}$$

El número m total de pruebas posibles se obtiene, con la combinatoria de $n-1$ elementos tomados de a $r-1$: $C (n-1 , r-1)$. La probabilidad Pascal pedida, será la probabilidad de la unión de todos los sucesos posibles, esto es:

$$P_{Pa} (r / n , p) = P (S_1 \cup S_2 \cup S_3 \cup \dots \cup S_m) = P (\cup_m S_m) = \sum_m P (S_m) = m \cdot P (S_m)$$

Reemplazando m con el valor de la combinatoria, se llega a la fórmula general de la P_{Pa} .

Otra forma de deducir lo mismo es la siguiente:

Sea R el suceso: Obtención de $r-1$ éxitos en las primeras $n-1$ pruebas realizadas
 y X el suceso: Obtención de un éxito en la última prueba realizada

Entonces la probabilidad buscada será

$$P_{Pa} (n / r , p) = P (\mathbf{R} \cap \mathbf{X}) = P (\mathbf{R}) \cdot P (\mathbf{X})$$

Donde $P (\mathbf{R})$ es una binomial de parámetros $r-1, n-1$ y $P (\mathbf{X})$ es la probabilidad p de éxito

$$P_{Pa} (n / r , p) = P_B (r-1 / n-1 , p) \cdot p = C (n-1 , r-1) \cdot p^{r-1} (1 - p)^{n-1-r+1} \cdot p$$

$$P_{Pa} (n / r , p) = C (n-1 , r-1) \cdot p^r (1 - p)^{n-r}$$

La relación entre la binomial y la Pascal se deduce con:

$$C(n-1 , r-1) = (r/n) \cdot C(n , r)$$

De donde,

$$P_{Pa} (n / r , p) = (r/n) \cdot P_B (r / n , p)$$

Lo que significa que todos los problemas del tipo Pascal se pueden resolver con la Binomial. Desde el punto de vista del cálculo si se tienen tablas para los valores binomiales (ver la bibliografía), no hace falta tener otra tabla para las probabilidades tipo Pascal.

Ejemplo 1) Se debe calcular la probabilidad de que sean necesarios 100 ensayos para lograr la vigésima muerte de los cobayos experimentales, del primer ejemplo del tema 7.2.

Aplicando la fórmula general es:

$$P_{Pa} (n=100 / r=20 , p=0,25) = C (99 , 19) \cdot (0,25)^{20} (0,75)^{80} = 0,00986$$

Usando los valores calculados antes con la binomial resulta lo mismo:

$$P_{Pa} (100 / 20 , 0,25) = (20 / 100) P_B (r=20 / n=100 , p=0,25) = 0,2 \cdot 0,0493 = 0,00986$$

Ejemplo 2) Usando los datos del cuarto ejemplo del punto 7.2 calcular la probabilidad de que sean necesarios 15 intentos antes de concretar la primera venta.

$$P_{Pa} (n=15 / r=1 , p=0,2) = (1 / 15) P_B (r=1 / n=15 , p=0,2) = (0,067) \cdot C (15 , 1) \cdot (0,2)^1 \cdot (0,8)^{14}$$

$$P_{Pa} (n=15 / r=1 , p=0,2) = (0,067) (0,1319) = 0,0088$$

Ejemplo 3) La probabilidad de que un alumno elegido al azar apruebe un examen es de 0,6. Si se eligen 5 alumnos al azar calcular la probabilidad de que haya que corregir 3 exámenes antes de tener al primer aprobado.

$$P_{Pa} (n=5 / r=3 , p=0,6) = (3 / 5) P_B (r=3 / n=5 , p=0,6) = (0,6) \cdot C (5,3) (0,6)^3 (0,4)^2 = 0,21$$

7.4 Probabilidad Binomial Negativa

Cuando en un proceso de tipo Bernoulli se desea saber la probabilidad de que sean necesarias efectuar exactamente g pruebas adicionales a fin de obtener r éxitos, en un total de n pruebas, se puede aplicar la fórmula de la binomial negativa para calcularla:

$$P_{BN} (g / n , r , p) = C (g+r-1 , r-1) \cdot p^r (1 - p)^g$$

Donde $g = n-r$ pues las pruebas adicionales a r éxitos solo pueden ser fracasos. Entonces, el número de pruebas será $n = g + r$. La probabilidad pedida es del tipo Pascal, pero para g fracasos; por lo tanto con un simple reemplazo se puede llegar a lo buscado:

$$P_{BN} (g / n , r , q) = P_{Pa} (n=g+r / r , q) = C (g+r-1 , r-1) \cdot (1-q)^r (q)^{g+r-r} = C (g+r-1 , r-1) \cdot p^r (1 - p)^g$$

Otra manera de expresarla más tradicional es la siguiente. Sea $q = 1 - p$, la probabilidad de un fracaso al hacer el experimento. Sean además $Q = (1/p)$ y $P = (q/p)$, entonces se pueden deducir las relaciones siguientes: $(Q-P) = 1$ y $(1-p) = (P/Q)$ reemplazando en la fórmula de Pascal será:

$$P_{Pa} (n / r , q) = C (g+r-1 , r-1) \cdot Q^{-r} (P/Q)^g = P_{BN} (g / n , r , q)$$

Siendo esta última la manera convencional de expresar la probabilidad binomial negativa. Sin embargo, se debe destacar que se trata de otra manera de ver la probabilidad de Pascal.

Ejemplo 1) Sea un sanatorio al cual arriban pacientes a tratarse diversas afecciones. Un 40% de los mismos concurren a efectuarse análisis clínicos. ¿ Cuánto vale la probabilidad de que veinte pacientes sigan de largo, antes de que el quinto paciente entre al laboratorio a tratarse, si se supone independencia en la llegada?

De acuerdo a los datos es : $p = 0,4$; $q = 0,6$; $r = 5$ y $g = 20$ luego el valor pedido será :

$$P_{BN} (g / n , r , p) = C (20+5-1 , 5-1) \cdot 0,4^5 (0,6)^{20} = 0,004$$

Ejemplo 2) Sea una farmacia donde el 40% de los clientes compran artículos de perfumería. Calcular la probabilidad de que 20 clientes no se acerquen a dicho sector, antes de que el cliente número 25 pida un artículo de perfumería.

Los datos son los mismos del problema anterior, por lo que la respuesta es:

$$P_{BN} (20 / 25 , 5 , 0,4) = 0,004$$

7.5 Probabilidad Geométrica

Cuando en un proceso de tipo Bernoulli se desea saber la probabilidad de que sean necesarias efectuar exactamente g pruebas antes de obtener el primer éxito ($r = 1$), en un total de n pruebas, donde $n = g + 1$. Se puede aplicar la fórmula de la probabilidad geométrica:

$$P_G (g / p) = p \cdot q^g$$

Ejemplo 1) Si en una farmacia la probabilidad de que un cliente compre un artículo de perfumería es del 30%, un remedio del 60%, y cualquier otra cosa un 10%. Calcular la probabilidad de tener que hacer 4 intentos antes de vender el primer: a) remedio; b) un artículo de perfumería.

a) Para remedio es : $P_G (g / p) = p \cdot q^g = (0,6) \cdot (0,4)^4 = 0,01536$

b) Para perfumería es: $P_G (g / p) = (0,3) \cdot (0,7)^4 = 0,07203$

Ejemplo 2) Si en un laboratorio de análisis clínicos especializado en HIV la probabilidad de que un paciente de positivo es del 20%. Calcular la probabilidad de que entren 8 pacientes antes de encontrar al primer positivo.

$$P_G (g / p) = p \cdot q^g = (0,2) \cdot (0,8)^8 = 0,034$$

7.6 Probabilidad Multinomial

Sean los sucesos $C_1, C_2, C_3, \dots, C_k$ todos mutuamente excluyentes entre sí, cuyas respectivas probabilidades de ocurrencia son $P(C_1), P(C_2) \dots P(C_k)$. Entonces, la probabilidad de que en n pruebas sucesivas el suceso C_1 ocurra r_1 veces, el C_2 suceda r_2 veces, y el suceso C_k ocurra r_k veces, se puede calcular con la fórmula de la *probabilidad multinomial*:

$$P_M = \frac{n!}{r_1! \cdot r_2! \cdot \dots \cdot r_k!} \cdot P(C_1)^{r_1} \cdot P(C_2)^{r_2} \cdot \dots \cdot P(C_k)^{r_k}$$

Esta probabilidad es una generalización de la probabilidad binomial, donde $r_1 + r_2 + \dots + r_n = n$. Y en el caso particular que $k=2$, la multinomial se reduce a la binomial.

Ejemplo 1) Un dado se lanza ocho veces; la probabilidad de obtener un 5 y un 6 dos veces es :

$$P_M = \frac{8!}{2! 2! 1! 1! 1! 1!} (1/6)^2 (1/6)^2 (1/6)(1/6)(1/6)(1/6) = 0,006$$

Ejemplo 2) En una urna se tienen 5 bolillas rojas, 4 blancas y 3 azules. Se saca una bola al azar de la urna y se anota su color, se la devuelve antes de hacer la siguiente extracción. Para calcular la probabilidad que de 6 pruebas realizadas, 3 sean rojas, 2 blancas y 1 azul

$$P_M = \frac{6!}{3! 2! 1!} (5/12)^3 (4/12)^2 (3/12)^1 = 0,12$$

Ejemplo 3) En una investigación epidemiológica retrospectiva respecto al cáncer de pulmón, se escogieron 75 pacientes de cáncer y 75 como control. Luego se averiguó entre ellos quienes fumaban y quienes no. Se encontraron los siguientes datos:

	Enfermos		Total
	SI	NO	
Fumador	50	60	110
No fumador	25	15	40
Total	75	75	150

La probabilidad exacta de que ocurran esos sucesos se calcula con la Probabilidad Multinomial

$$P_M = \frac{150!}{50! 25! 60! 15!} (50/150)^{50} (25/150)^{25} (60/150)^{60} (15/150)^{15} = 0,00072$$

Odds de enfermedad: $p / (1-p) = (75/150) / [1 - (75/150)] = TE / TS = 75 / 75 = 1$

Odds del factor de riesgo: $(110/150) / [(1 - (110/150))] = 0,73 / 0,27 = 2,7 \cong 3$

Riesgo Relativo: $RR = [(50/110)] / [(25/40)] = 0,73$

Eso se interpreta como que hay 73 chances en 100 de contraer cáncer si se fuma.

7.7 Procesos de tipo Poisson

Se denominan procesos de tipo Poisson, o poissonianos, a todo experimento consistente en una serie de pruebas repetidas dentro de un *continuo*, caracterizadas por tener resultados que se pueden clasificar en si verifican o no, cierta propiedad o atributo, siendo aleatorios e independientes del lugar que ocurren dentro del continuo.

Para identificar un proceso Poisson en una serie de pruebas repetidas, se deben verificar tres condiciones:

1) *Sucesos puntuales*: Los sucesos ocurren dentro de un continuo (espacio o tiempo) y ocupan una parte infinitesimal del mismo. Es decir, en el espacio un suceso es puntual y en el tiempo es instantáneo. En términos prácticos, los sucesos no ocupan una parte apreciable del continuo.

2) *Sucesos independientes*: La ocurrencia de un suceso en un lugar del continuo no condiciona la ocurrencia del anterior (o del siguiente) en otra parte del mismo. Es decir, no hay reacción en ca-

dena, ni en cascada, ni hay apilamientos de células, ni agrupamientos en colonias, etc. Los sucesos son independientes del lugar que ocupan en el continuo.

3) *Probabilidad constante*: La probabilidad de ocurrencia de un suceso en un lugar del continuo es la misma en todo punto del mismo. Es decir, no hay lugares donde los sucesos ocurran con mayor frecuencia que en otros. Los puntos del continuo son equiprobables.

Son ejemplos de este tipo de proceso, un recuento celular en cámara de Neubauer o Hemocitómetro, la llegada de pacientes a una cola o línea de espera, los accidentes en una ruta, el desgaste de una pieza en una máquina de producción produce una reacción en cascada de productos defectuosos, y muchos otros casos. Esta probabilidad se *aproxima* a la binomial cuando la probabilidad de éxito es muy pequeña, por eso muchos la llaman: la "binomial de los sucesos raros".

Para imaginar casos de la vida real donde se la puede encontrar, se debe considerar un medio continuo como la sangre, orina, agua de mar o de río, el aire, el tiempo, hasta pasta de hornear de tamaño muy grande, dentro del cual ocurre un número total elevado de pequeñas cantidades discretas tales como los glóbulos blancos, rojos o plaquetas en la sangre, residuos en orina, plancton en el agua de mar, partículas de tierra en la de río, emisión de una partícula en sales de uranio, ralladura de limón en la pasta de una torta, etc. Si nada hace que tales partículas se agrupen (como la presencia de un coagulante), se cumple la condición de densidad uniforme de las partículas en el medio (tipo coloide). En caso contrario, se debe agitar o mezclar bien para lograrlo, tal como ocurre naturalmente en el torrente sanguíneo o en las aguas de un río, y artificialmente en la amasadora de una panadería, el vibrador de los laboratorios o el agregado de EDTA para evitar la coagulación en las muestras de sangre de los pacientes.

Otro modelo que produce casos Poisson son los que ocurren en el tiempo, como las llamadas que entran a una central telefónica, emisión de partículas radioactivas, pacientes o clientes en general esperando ser atendidos, autos que llegan a una estación de combustible, etc. En estos casos, se debe suponer: (a) los sucesos ocurren en forma independiente; (b) la probabilidad que un suceso ocurra en un corto intervalo sea proporcional a la longitud del intervalo; y (c) la duración del suceso es tan pequeña que hace insignificante a su probabilidad de ocurrencia. Entonces, si se cumplen estos tres supuestos la probabilidad de que ocurran exactamente r sucesos en un intervalo de tiempo finito es la de Poisson. En estos principios se basaron los Curie para su ley de Radiactividad Natural, que es la integral del valor medio de la función de Poisson variando en función del tiempo.

En términos generales, cuando un proceso del tipo Poisson tiene una "intensidad" promedio μ (es una especie de densidad de las partículas en el medio continuo), la probabilidad de que ocurran exactamente r sucesos o éxitos, se obtiene con:

$$P_{PO}(r) = (e^{-\mu} \cdot \mu^r) / r!$$

Un proceso es de tipo Poisson, cuando los sucesos puntuales se producen individual y colectivamente al azar dentro de un continuo.

7.7.1 Aplicaciones del Modelo Poisson

El primer ejemplo fue tomado del clásico estudio de Student sobre los recuentos celulares y publicado en la revista Biometrika en 1907. Se usa una cámara de recuento llamada en aquel entonces hemocitómetro, hoy más conocida como cámara de Neubauer. Para el recuento de glóbulos rojos se emplea la zona dividida en 400 cuadrados, como se conoce la cantidad de sangre y el grado de dilución, una vez contados los glóbulos se obtiene la concentración de los mismos por ml de sangre. Student (F. Gosset) la usaba para el conteo de células de levaduras de cerveza y observó al aplicar la teoría de errores gaussianas, que la función Poisson se adaptaba mejor que la de Gauss a los valores observados. Las implicancias históricas de este trabajo son grandes, pero en el ejemplo aquí desarrollado solo interesa ver la manera de calcular las frecuencias esperadas. Si se conoce la probabilidad de encontrar una cuadrícula vacía, mediante la fórmula Poisson se puede sacar el número esperado de cuadrículas vacías al multiplicarla por el número total de cuadrículas: $O(r=0) = N \cdot P_{PO}(r=0)$. En este caso se muestra la *manera iterativa* de resolver este tipo de problemas. La probabilidad Poisson para r sucesos es igual a la probabilidad anterior ($r-1$) multiplicada por μ y dividida por r . En su famoso informe, Student (1907) mostró que el ajuste usando Poisson era mejor que el usado hasta entonces, cosa que se verá más adelante en el tema Bondad de Ajuste estadística.

Ejemplo 1) Recuentos celulares.

Se trata de un recuento de células de levadura de cerveza en las 400 cuadrículas de un hemocitómetro. (Cuadro 4.2: Caso 3). La media aritmética allí calculada es una especie de densidad o concentración de células de levadura por unidad de volumen de agua. Es la intensidad en la fórmula de Poisson, o sea: $\mu = 0,6825$ células por cuadrícula. Entonces:

$r = 0, 1, \dots, 5$ son los valores esperados de la cantidad de células por cuadrícula, se puede aplicar la fórmula y calcular las probabilidades para cada caso.

Nº de células /cuadrícula (r) :	0	1	2	3	4	5	Σ
Frecuencias Observadas (Oi) :	213	128	37	18	3	1	400
Probabilidad Poisson : $P_{PO}(r)$:	0,505	0,345	0,118	0,027	0,0045	0,0005	1,00
Frecuencias Esperadas (Ei) :	202	138	47,2	10,8	1,8	0,2	400

Paso 1 : Se calcula la probabilidad de cuadrícula vacía:

$$P_{PO}(r=0) = e^{-\mu} = e^{-0,6825} = 0,505$$

Luego la frecuencia esperada será :

$$E(r=0) = N \cdot P_{PO}(r=0) = 400 \cdot 0,505 = 202$$

Paso 2 : Se calcula la probabilidad de encontrar una célula por cuadrícula

$$P_{PO}(r=1) = (e^{-\mu} \cdot \mu^1) / 1! = P_{PO}(r=0) \cdot \mu / 1! = 0,505 \cdot 0,6825 = 0,345$$

$$\text{Y así, es } E(r=1) = N \cdot P_{PO}(r=1) = 400 \cdot 0,345 = 138$$

Paso 3 : Se calcula la probabilidad de encontrar dos células por cuadrícula

$$P_{PO}(r=2) = (e^{-\mu} \cdot \mu^2) / 2! = P_{PO}(r=1) \cdot \mu / 2 = 0,345 \cdot 0,6825 / 2 = 0,118$$

$$\text{Y así, es } E(r=2) = N \cdot P_{PO}(r=2) = 400 \cdot 0,118 = 47,2$$

Paso 4 : Se calcula la probabilidad de encontrar tres células por cuadrícula

$$P_{PO}(r=3) = (e^{-\mu} \cdot \mu^3) / 3! = P_{PO}(r=2) \cdot \mu / 3 = 0,118 \cdot 0,6825 / 3 = 0,027$$

Y así, es $E(r=3) = N \cdot P_{PO}(r=3) = 400 \cdot 0,027 = 10,8$

Paso 5 : Se calcula la probabilidad de encontrar cuatro células por cuadrícula

$$P_{PO}(r=4) = (e^{-\mu} \cdot \mu^4) / 4! = P_{PO}(r=3) \cdot \mu / 4 = 0,027 \cdot 0,6825 / 4 = 0,0045$$

Y así, es $E(r=4) = N \cdot P_{PO}(r=4) = 400 \cdot 0,0045 = 1,8$

Paso 6 : Se calcula la probabilidad de encontrar cinco células por cuadrícula

$$P_{PO}(r=5) = (e^{-\mu} \cdot \mu^5) / 5! = P_{PO}(r=4) \cdot \mu / 5 = 0,0055 \cdot 0,6825 / 5 = 0,0005$$

Y así, es $E(r=5) = N \cdot P_{PO}(r=5) = 400 \cdot 0,0005 = 0,2$

De la comparación entre las frecuencias observadas (O_i) y las esperadas (E_i), se puede deducir que tan bien se ajusta la teoría (E_i) a la realidad (O_i). Student demostró que si las E_i se las obtenía con el modelo de Poisson en lugar del gaussiano tradicional, el ajuste era mejor.

En el segundo ejemplo se toma un conteo radiológico de 100 cuentas de 0,1 minuto de una sola fuente, y también se puede apreciar un buen ajuste entre valores esperados y observados. Los datos reales se tomaron de la obra de Remington-Schorck (ver Bibliografía).

Ejemplo 2) Conteos radiológicos.

Las frecuencias observadas de 100 cuentas radiológicas de 0,1 minuto, fueron medidas por J. Nehemías en un trabajo publicado en la obra antes citada.

Conteo radiológico (r) :	0	1	2	3	4	5	6 y más	Σ
Frec. Observadas (O _i) :	11	20	28	24	12	5	0	100
Probab. Poisson P _{PO} (r) :	0,111	0,244	0,268	0,197	0,108	0,048	0,004	1
Frec. Esperadas (E _i) :	11,1	24,4	26,8	19,7	10,8	4,8	2,4	100

Repitiendo los pasos del ejemplo anterior, se vuelve a aplicar la manera iterativa para resolver estos casos.

Se ha comentado que la función de Poisson se ha utilizado reiteradamente para las teorías de fenómenos raros, tanto como para aplicaciones más prácticas. Para ilustrar estos puntos se dan a continuación dos ejemplos más:

Ejemplo 3) Teoría de extinción de los dinosaurios.

Budyko en 1974, usó la fórmula de Poisson para responder una pregunta clásica :

¿ La extinción de los dinosaurios en el Cretáceo fue por causas naturales o por algunas causas de tipo ambiental ?

Al final del período Cretáceo, de las diez órdenes de reptiles existentes cinco desaparecieron, a saber: dos órdenes de dinosaurios terrestres, una de dinosaurios de tipo volador, los pterosaurios y dos órdenes de reptiles acuáticos: el ictosaurio y el pleiosaurio. Estas cinco órdenes, incluyen treinta y cinco familias cada una de las cuales a su vez, estaba compuesta por una numerosa cantidad de especies. Se ha estimado el tiempo promedio de existencia de un orden de repti-

les en el planeta, en 100 millones de años, y se ha calculado que la extinción se produjo en un período de 5 millones de años, al finalizar el Cretáceo.

Si se supone que la desaparición de un orden de dinosaurios es independiente de las otras, y que no hubo ningún desastre que haya desencadenado el proceso, se puede concluir que: "las extinciones se produjeron individual y colectivamente al azar". Entonces, el número promedio de extinciones será de $\mu = 5$ millones de años / 100 millones de años = 0,05. Por lo tanto, la probabilidad que se produzcan 5 extinciones en 5 millones de años será de acuerdo a la fórmula de Poisson:

$$P_{PO}(r = 5) = (e^{-0,05} \cdot 0,05^5) / 5! = 2,5 \cdot 10^{-9}$$

Un valor demasiado bajo como para suponer que se debió a causas naturales. Mas bien se puede pensar que una serie de circunstancias geológicas y climatológicas, unidas a grandes desequilibrios ecológicos, fueron determinantes para la desaparición de los grandes reptiles.

Ejemplo 4) Técnica rápida de recuento en hemocitómetro.

El efectuar un recuento de glóbulos rojos en cámara es una tarea larga y puntillosa. Es fácil equivocarse y por seguridad conviene hacerla más de una vez. Como demanda mucho tiempo y los honorarios son muy bajos, comparado con otras, esta técnica se hizo impopular. En su lugar, ahora se usa la técnica del microhematocrito (con una inversión baja), o mejor aún las cámaras de recuento automático que son aparatos de alto costo pero de alta producción. Sin embargo, hay lugares y situaciones donde se deben hacer las determinaciones y no se cuenta con los medios suficientes para tales aparatos mecánicos. Para esos casos especiales, se presenta esta técnica de conteo rápido. Aquí, lo que se deben contar no son las células sino las cuadrículas vacías en toda la cámara.

Sea una cámara típica de Neubauer, con una concavidad de 0,1 mm de profundidad y de 1 mm² de superficie, sobre un porta objeto, la misma está dividida en 20 x 20 cuadrículas; entonces el volumen de líquido sobre ellas será de: 0,1 mm . 1 mm² / 400 = (1/4000) mm³. Se toma una muestra de sangre y se diluye por ejemplo al 25%, se homogeniza y se agita para evitar el efecto de apilamiento celular (o se agrega EDTA como anticoagulante). De esta forma se puede aceptar que los hematíes quedan distribuidos individual y colectivamente al azar dentro de la cuadrícula. Se dan las condiciones de un proceso del tipo Poisson. Sea H el número de glóbulos rojos en una cuadrícula, los que siguen una distribución en el continuo de tipo Poisson, con una intensidad μ (concentración de hematíes por unidad de volumen) desconocida y objeto de la medición. Entonces, la probabilidad de encontrar un número r de hematíes en una cuadrícula se obtiene con :

$$P_{PO}(H = r) = (e^{-\mu} \cdot \mu^r) / r!$$

En particular si $r = 0$

$$P_{PO}(H = 0) = e^{-\mu}$$

De donde

$$-\ln \{ P_{PO}(H = 0) \} = \mu$$

Y para estimar la probabilidad de encontrar una cuadrícula vacía se pueden usar las frecuencias observadas. Basta contar la cantidad de cuadrículas vacías V y dividirla por el número total de cuadrículas N = 400. O sea,

$$P_{PO}(H = 0) \approx V / N$$

Reemplazando es :

$$-\ln(V/400) = \mu = \ln(400/V)$$

Como se había hecho una dilución de 400 volúmenes (al 25%), para obtener el número de glóbulos rojos por mm^3 : R , se deberá multiplicar el valor promedio hallado para una cuadrícula por un factor : f . Donde f es el volumen de líquido por cuadrícula ($1/4000 \text{ mm}^3$) multiplicado por el factor de dilución (1/400). Esto es: $f = 1 / 1.600.000$

Entonces resultará :

$$f \cdot R = \mu = \ln(400/V)$$

De donde :

$$R = 1,6 \cdot 10^6 \cdot \ln(400/V)$$

Por ejemplo, si se detectaron 20 cuadrículas vacías al hacer la medición el número de glóbulos rojos por mm^3 de sangre será: $R = 1,6 \cdot 10^6 \cdot \ln(400/40) = 3,68 \cdot 10^6$ hematíes / mm^3

7.8 Aproximación de la binomial a la Poisson

Una binomial de muy baja probabilidad de éxito p y con un gran número de pruebas repetidas tales que su producto es finito, resulta ser muy parecida a una Poisson.

Si $p \rightarrow 0$ y $n \rightarrow \infty$ tales que $n \cdot p \rightarrow \mu$, entonces $P_B(r/n, p) \rightarrow P_{PO}(r/\mu)$

En la práctica, muchas veces aparecen casos con un número de pruebas lo suficientemente grande, como para ser considerado infinito. Con $n > 30$ y $p < 0,1$ el error que se comete al aproximar la binomial con la Poisson no tiene importancia. La ventaja, es que resulta mucho más cómodo calcular los valores de P_{PO} que los de la P_B .

Ejemplo 1) Se sabe que el 1% de las jeringas descartables fabricadas tienen defectos en el cierre hermético. Se compra un lote y se prueban 30 de ellas. ¿Cuánto vale la probabilidad de que dos o más de ellas tengan defectos?

$$\begin{aligned} P_B(r \geq 2/n=30, p=0,01) &= P_B(r=2) + P_B(r=3) + \dots + P_B(r=30) = 0,0328 + 0,031 + \dots + 0 = 0,0361 \\ &= 1 - P_B(r=0) - P_B(r=1) = 1 - 0,7397 - 0,2242 = 0,0361 \end{aligned}$$

$$\begin{aligned} P_{PO}(r \geq 2/\mu=0,3) &= P_{PO}(r=2) + P_{PO}(r=3) + \dots + P_{PO}(r=30) = 0,0333 + 0,0033 + 0,0002 = 0,0368 \\ &= 1 - P_{PO}(r=0) - P_{PO}(r=1) = 1 - 0,7408 - 0,2222 = 0,037 \end{aligned}$$

Esta aproximación puede llegar a ahorrar mucho trabajo y tiempo en los cálculos. En el ejemplo anterior se puede apreciar que es bastante aceptable usar la Poisson, para calcular una Binomial, con importantes ventajas en cuanto a la simplicidad. A pesar de que estas aproximaciones dejaron de tener tanta importancia como antaño cuando no se disponía de computadoras para los cálculos, sigue siendo interesante ver el comportamiento de estas funciones de probabilidad en condiciones de borde para entender el hecho fundamental que se verá en el Teorema Central del Límite más adelante.

7.9 Aproximación de binomial a hipergeométrica

Cuando se cumplan las condiciones (1) o (2) siguientes:

- (1) $r \ll R$ y $n \ll N-R$
 o Entonces $P_B(r/n, p=R/N) \rightarrow P_H(r/n, N, R)$
 (2) $n \ll R$ y $n-r \ll N-R$

Ejemplo 1) De 20 kits de glucosa guardados en el depósito, 8 de ellos tienen 5 o más semanas dentro del mismo y los 17 restantes son nuevos. Si se toman 4 kits al azar ¿ Cuánto vale la probabilidad que no más de tres de ellos sean viejos ?. Acá es $N=20, R=8, r=2$ y $n=4$.

$$P_H(r \leq 2 / 4, 20, 8) = 1 - P_H(0 / 4, 20, 8) - P_H(1 / 4, 20, 8) = 1 - 0,1532 = 0,8468$$

Pero como $r = 2 \ll R = 8$ y $8-4 = 4 \ll 12 = 20 - 8$ se puede aplicar la aproximación de la binomial con :

$$P_B(r \leq 2 / 4 ; 0,4) = 1 - P_B(1 / 4 ; 0,4) - P_B(2 / 4 ; 0,4) = 1 - 0,1792 = 0,8208$$

Se cometió un error de estimación del orden del 3% aproximadamente.

7.10 Problemas propuestos

- | | | |
|--|---|---|
| 1) Para identificar un proceso del tipo Bernoulli se deben verificar 3 condiciones. | V | F |
| 2) Cualquier variable puede ser dicotomizada. | V | F |
| 3) Si el resultado de una prueba incide en la siguiente se dice que son independientes. | V | F |
| 4) La independencia de las pruebas es un supuesto básico de la teoría mendeliana. | V | F |
| 5) La probabilidad de éxito se mantiene constante en un proceso Bernoulli. | V | F |
| 6) Si los fenómenos ocurren en cascada, o en cadena, no es un caso de Bernoulli. | V | F |
| 7) La probabilidad Binomial solo sirve para calcular la probabilidad de tener r éxitos. | V | F |
| 8) La probabilidad Binomial se puede aplicar en Genética y en los nacimientos. | V | F |
| 9) En un contagio la probabilidad de éxito permanece constante. | V | F |
| 10) En un caso de repulsión la probabilidad de fracaso es una constante. | V | F |
| 11) Las Probabilidades Binomial y Pascal están relacionadas con una fórmula. | V | F |
| 12) En Pascal se considera que se necesitan n pruebas para tener el último éxito. | V | F |
| 13) Si se trata de las pruebas adicionales, para obtener r éxitos, es una hipergeométrica. | V | F |
| 14) La Multinomial es una generalización de la Pascal. | V | F |
| 15) Los sucesos puntuales ocurren individual y colectivamente al azar en una Poisson. | V | F |
| 16) Si hay apilamiento celular en un recuento se siguen cumpliendo los supuestos Poisson. | V | F |
| 17) En un modelo Poisson los sucesos ocurren en cadena o se agrupan en colonias. | V | F |
| 18) En un modelo Poisson los sucesos ocurren en un continuo. | V | F |
| 19) Los supuestos básicos del Modelo Poisson, son | | |
| 20) Los supuestos básicos del Modelo Bernoulli, son | | |
| 21) Un recuento celular es un ejemplo de aplicación del Modelo | | |
| 22) La probabilidad Binomial se puede aproximar con un Modelo | | |
| 23) La probabilidad Binomial se aproxima a una Hipergeométrica, bajo ciertas condiciones. | V | F |
| 24) A la probabilidad Poisson se la llama la binomial de los casos raros. | V | F |

2) En un apunte de cátedra de 200 páginas hay 220 errores de tipeo distribuidos al azar a lo largo del texto. Se pide calcular la probabilidad de que una página cualquiera tenga:

- a) Ningún error. b) Dos errores. c) Dos o más errores.

3) En una fábrica de medicamentos el dosificador de droga en polvo produce un 1% de remedios defectuosos por mal funcionamiento. Si el farmacéutico a cargo del sector Control de Calidad elige al azar 100 medicamentos de un lote de producción, se pide calcular la probabilidad de encontrar: a) Ningún defectuoso. b) Exactamente 2 defectuosos. c) Más de 3 defectuosos.

4) En la población de alumnos de la Facultad de Ciencias Exactas hay un 5% de zurdos. Si se eligen al azar 100 estudiantes calcular la probabilidad de que: a) 97 o más usen su mano derecha. b) 2 o más sean zurdos. c) Ninguno sea zurdo. d) Todos usen la mano izquierda.

5) Si en la población de la Provincia de Misiones estimada en 800.000 personas ocurren en promedio 80 suicidios al año, calcular la probabilidad de que el año que viene ocurran: a) 90 suicidios. b) 80 o más suicidios. c) Menos de 80 suicidios.

6) De un mazo de barajas españolas de 40 cartas se sacan 3 cartas, hallar la probabilidad de que: a) Sean las 3 del mismo palo si se reponen a medida que se sacan. b) Ídem sin reposición. c) Por lo menos una sea de oros. d) Dos sean de copas. e) Ninguna sea de espadas. f) Las 3 sean figuras. g) Salgan 3 ases. h) Sacar el 7 de oros, el 1 de espadas y el 1 de bastos.

7) En la cátedra de Bioestadística hay 8 auxiliares que pueden trabajar por igual en su casa tanto como en la Facultad. Si todos se destinan a un mismo gabinete, se desea averiguar cuantos escritorios hay que colocar para que cada uno tenga en el 90% de las veces un escritorio disponible.

8) En una farmacia se reciben en promedio 5 pedidos por TE para entrega a domicilio por día. En una hora cualquiera elegida al azar, calcular la probabilidad de que: a) Se reciban 3 pedidos. b) Se reciban más de 5 pedidos. c) Se reciban entre 2 y 4 pedidos.

9) En un hospital se mide cuantos enfermos con infecciones venéreas solicitan atención en un período de 50 días. Los datos se muestran en la tabla siguiente. En la primer columna se colocan las frecuencias observadas diarias. En la segunda la cantidad de días en que ocurrieron. Y en la última columna se calcularon las frecuencias relativas (probabilidades empíricas) observadas. Se pide calcular en una cuarta columna las probabilidades teóricas respectivas usando un modelo de Poisson con una media de 5,66 pacientes diarios.

Pacientes diarios	Nº de días	Frec. Relativa Empírica
3	3	0,06
4	7	0,14
5	12	0,24
6	14	0,28
7	10	0,20
8	4	0,08
Total	50	1,00

8

Funciones de probabilidad

En este capítulo se tratan las bases mínimas para comprender las distribuciones de probabilidad. Es un tratamiento matemático un poco más formal que el visto hasta ahora. Si bien no se requiere un nivel matemático avanzado para leer este capítulo, se recomienda a los lectores que no tengan aprobados cursos de análisis matemático, que eludan los desarrollos y se concentren solo en los conceptos tratados.

8.1 Fenómenos aleatorios en Bioquímica y Farmacia

Tanto en Bioquímica como en Farmacia y en otras ciencias se busca en forma sistemática la adquisición de conocimientos, cada una aplicada a su campo específico. Esta búsqueda de la verdad la efectúan a través del intelecto humano con la realización de experimentos que ayuden en la tarea. Los científicos reúnen sus hechos mediante una cuidadosa observación de los fenómenos que ocurren, los agrupan, los clasifican y los tratan de explicar a través de otros hechos que conozcan. Con las teorías ya adquiridas se intenta explicar un nuevo fenómeno analizado. Cuando no se puede, entonces se trata de elaborar una nueva teoría que se ajuste a las nuevas observaciones realizadas. Pero en ambos casos siempre deben verificar las teorías propuestas a través de experimentos, cuidadosamente planeados y controlados. Esto implica un trabajo permanente de comparación entre los valores observados y los valores esperados teóricamente. De eso se ocupan los tests de hipótesis estadísticos. Y para llegar allí, lo primero es desarrollar modelos matemáticos como se presentan en este apartado.

Con todo ese trabajo se puede elaborar una imagen que explique los fenómenos observados. Luego con la información proveniente de los experimentos se puede ir corrigiendo y ajustando la teoría, o bien, su campo de aplicación. Este ciclo permanente de teoría a experimento y de experimento a teoría -para estudiar los fenómenos- es buena parte del accionar científico. Se debe hacer así pues nunca se tiene una certeza total acerca del fenómeno estudiado.

Fenómeno aleatorio: es todo fenómeno sobre el cual no se tiene la certeza absoluta de poder explicarlo, en por lo menos algún ámbito o sistema de referencia.

Definido así, se deduce que todo fenómeno conocido es o fue aleatorio alguna vez. Por ejemplo, la determinación del sexo de un recién nacido fue aleatoria hasta antes del alumbramiento, momento en que se alcanza la certeza. No hay ciencia sin experimentos y tampoco hay experimentos sin ciencia.

Experimento aleatorio: es todo experimento sobre cuyo resultado no se tiene "a priori" la certeza de su resultado.

Por ejemplo, si el experimento es lanzar un dado, se tiene la certeza que hay seis resultados posibles pero nunca se sabe cuál cara saldrá si se trata de un dado normal. Existe un cierto grado de incertidumbre asociado a cada cara posible. En cambio, si se trata de contar los dedos de una mano, solo hay certeza si se trata de la mano del investigador; pero cuando se va a estudiar a un grupo de desconocidos aparece la incertidumbre.

A la indeterminación asociada al resultado de un experimento se la llama *error* y se la puede cuantificar con una probabilidad. Esa variabilidad de los resultados se produce por causas desconocidas, en la gran mayoría de los casos, a las que se les puso un nombre: *azar*. Como no se puede eliminar la variabilidad en los experimentos lo único que se puede hacer es acotarla a márgenes razonables que permitan mantenerlo controlado. Cuanto menor sea la indeterminación, menor será el error y mayor la precisión. La manera clásica es fijar todas las variables externas que se cree, pueden causar variaciones, menos una, la que se va a medir. Entonces, la repetición del experimento arrojará resultados lo más uniformes posibles ganando homogeneidad. Las mediciones diarias del laboratorio de análisis clínicos son experimentos aleatorios, tanto como la cantidad de clientes que llegan a una farmacia. El problema adicional en un laboratorio, a diferencia de la Física o la Química, reside en la poca estabilidad del material de trabajo. Una sangre entera se altera mucho más fácilmente que un trozo de metal. Por ende, sus resultados estarán asociados a una incertidumbre mayor. Lo mismo en una farmacia donde la llegada de clientes está influenciada por factores exógenos como la propaganda, su ubicación geográfica, etc. Esta falta de precisión no ayuda a la hora de efectuar diagnósticos o pronósticos y mucho menos para elaborar teorías. Sin embargo, se puede trabajar científicamente en la práctica diaria pero dentro de márgenes más amplios.

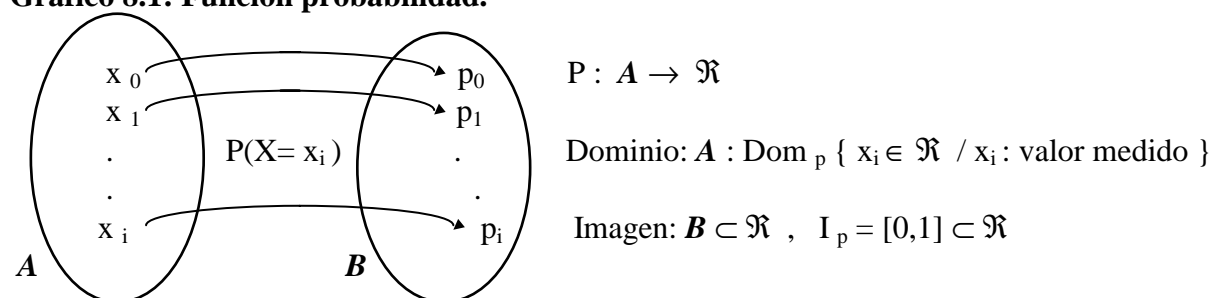
Si se imagina un resultado obtenido en una medición biológica (dato) como un valor de una variable matemática, entonces se puede asimilar la magnitud biológica medida a una:

Variable aleatoria: son todas aquellas magnitudes donde cada uno de los valores que pueda tomar, en un sistema de referencia o población, tiene asociada una cierta probabilidad de ocurrencia.

Así, toda variable biológica es una variable aleatoria. Cuando una magnitud cualquiera no varía en todo su ámbito de referencia, deja de ser variable aleatoria y se transforma en un *parámetro poblacional*. Como puede ser, la magnitud sexo en toda la población femenina actual de Posadas.

Sea el conjunto A de todos los resultados posibles de una medición de la magnitud biológica X , la cual es una variable aleatoria; entonces, cada valor medido x_i tiene asociada una cierta probabilidad de ocurrencia p_i , a través de una *función probabilidad* $P(X = x_i) = p_i$ con un dominio Dom_p definido como todos los valores obtenidos dentro de los números reales \mathfrak{R} y con una imagen B perteneciente a los reales entre 0 y 1.

Gráfico 8.1: Función probabilidad.



Entonces, una *función de probabilidad* es una función tal que cumple dos condiciones:

- 1) $1 \geq P(X = x_i) \geq 0$
- 2) $\sum_i P(x_i) = 1$

La segunda condición refleja el hecho que el conjunto de sucesos correspondientes a los diferentes x_i es una partición. Como por ejemplo el caso de los cuatro casos posibles al efectuar un diagnóstico.

$$\sum_i P(x_i) = vp/N + fp/N + vn/N + fn/N = (vp + vn + fp + fn)/N = 1$$

8.2 Función de distribución

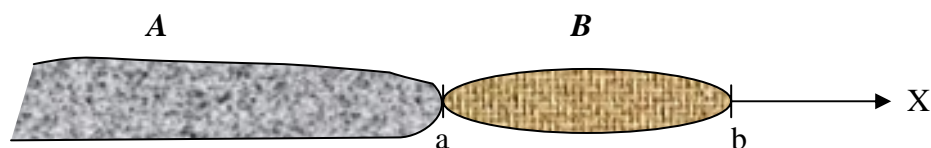
Un experimento aleatorio puede ser la determinación de una glucosa, un colesterol, la temperatura del cuerpo del paciente, su altura etc., donde se está midiendo una sola magnitud clínica aunque esta sea compuesta como una concentración o densidad. En estos casos se dice que se está trabajando con una variable simple o en un caso *unidimensional*. Al repetir sistemáticamente el experimento, bajo condiciones controladas por el observador, se dispone de una serie de valores para calcular la frecuencia relativa de los sucesos. O sea, una medida experimental de su probabilidad de ocurrencia. Otra forma es conociendo a través de la teoría todas las probabilidades asociadas a los valores posibles que toma la variable aleatoria, como la relación mendeliana 9:3:3:1 para la segunda generación filial.

Cuando se conoce la función probabilidad de una variable aleatoria, se puede obtener su acumulada: la *función distribución* de probabilidad. Sea r un valor cualquiera que puede adoptar una variable aleatoria X , y sea $P(X \leq r)$ la probabilidad que X adopte un valor menor o igual que r , entonces se llama:

$$F(r) = P(X \leq r) \text{ función distribución}$$

$$P(r) = P(X = r) \text{ función probabilidad}$$

Sean dos sucesos $A: X \leq a$ (siendo a un valor cualquiera de la variable aleatoria X)
 $B: a < X \leq b$ (donde $b > a$)



Entonces $P(A) = P(X \leq a) = F(a)$ (incluye al valor a)
 $P(B) = P(a < X \leq b)$ (no incluye al valor a)

Como ambos sucesos A y B no pueden ocurrir a la vez, al no tener ningún punto en común son mutuamente excluyentes ($A \cap B = \emptyset$), luego se puede aplicar el Axioma 3 de probabilidad, para la unión de ambos sucesos:

$$P(A \cup B) = P(A) + P(B) = P(X \leq a) + P(a < X \leq b) = P\{(X \leq a) \cup (a < X \leq b)\}$$

$$P(A \cup B) = P(X \leq b) = F(b) = P(X \leq a) + P(a < X \leq b) = F(a) + P(a < X \leq b)$$

De donde
$$P(a < X \leq b) = F(b) - F(a)$$

La igualdad anterior expresa lo siguiente: *si se conoce la función distribución $F(r)$ para todos los valores posibles de r , de la variable aleatoria X , entonces la función probabilidad queda completamente determinada.*

8.2.1 Distribuciones discretas unidimensionales

Si la variable aleatoria estudiada es de tipo discreto, es decir, solo puede tomar algunos valores $x_1, x_2, x_3 \dots x_r$ dentro de un intervalo $[a, b]$ cualquiera en los números reales, entonces se puede definir la probabilidad que ocurra un evento cualquiera $p_i = P(x_i)$ y su función distribución con la relación :

$$F(k) = \sum_i P(X \leq x_k) = P(x_1) + P(x_2) + \dots + P(x_k)$$

Luego, si hay r eventos posibles será:
$$\sum_{i=1}^r P(X = x_i) = p_1 + p_2 + \dots + p_r = 1$$

Cuando r sea muy grande, los valores de probabilidad se hacen muy pequeños, y en el caso límite, cuando $r \rightarrow \infty$ los valores $p_i \rightarrow 0$. Pero aún así las expresiones anteriores siguen siendo válidas. Por ejemplo, sea el caso de lanzar tres veces una moneda al aire, representado en el Gráfico 6.1 mediante el diagrama del árbol. Se pueden graficar las funciones probabilidad y distribución para ese caso (Gráfico 8.2), tomando como variable aleatoria X al número de caras obtenidos al lanzar tres veces la moneda. Los casos posibles serán: $X = 0$ (no salió ninguna cara); $X = 1$ (salió una sola cara); $X = 2$ (salieron dos caras) y $X = 3$ (las tres fueron caras). O sea,

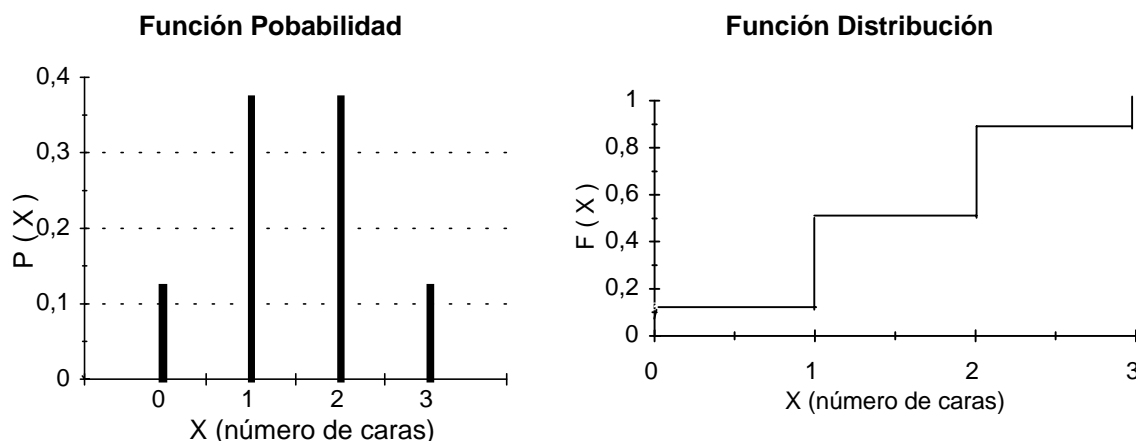
$$X = 0, 1, 2, 3 \Rightarrow P(X=0) = P(X=3) = 1/8 \quad ; \quad P(X = 1) = P(X = 2) = 3/8$$

En el caso de variable discreta, la función probabilidad es un diagrama de bastones parecido al diagrama de frecuencias relativas (histograma). Por su parte, la función distribución se parece al polígono de frecuencias acumuladas. En un intervalo cualquiera $[1, 2]$ resulta

$$F(1) = 0,125 \quad y \quad F(2) = 0,5 \quad \text{por lo tanto} \quad P(1 < X \leq 2) = 0,5 - 0,125 = 0,375 = P(2)$$

Se pueden calcular así los valores de las respectivas funciones para cada valor de X . Porque si se conocen todos los valores posibles de la función probabilidad, quedan completamente determinados todos los valores de la función distribución.

Gráfico 8.2: Función probabilidad y distribución en el lanzamiento triple de una moneda.



En el caso de la Tabla de la Verdad para los resultados de un diagnóstico en N pruebas, aún sin conocer la función de probabilidad, se sabe que los cuatro resultados posibles tienen una probabilidad calculada con:

$$\sum_{i=1}^r P(X = x_i) = P(X = vp) + P(X = fp) + P(X = vn) + P(X = fn) = 1$$

8.2.2 Distribuciones continuas unidimensionales

Si ahora la variable aleatoria puede tomar cualquier valor en un intervalo cualquiera, entonces se trata de una variable continua. Pero, en un intervalo como $[a, b]$ existen infinitos puntos de la variable aleatoria eso significa infinitos resultados posibles de experimento. Entonces los valores de la función probabilidad para un punto cualquiera, solo pueden ser infinitesimales para que la acumulada de todas ellas sea la unidad. Por ello, en variable continua la función de probabilidad *siempre se define para un intervalo y no para un punto*.

Sea un intervalo cualquiera $[a, b]$ de una variable biológica X continua y aleatoria, definida en los números reales \mathfrak{R} , y sea x un punto dentro de ese intervalo, tal que $a < x < b$, la función probabilidad asociada $P(X = x)$ puede definirse para un pequeño intervalo Δx de manera tal de hacer: $a = x - \Delta x$ y $b = x + \Delta x$; o sea: $x - \Delta x < x < x + \Delta x$ entonces:

$$P(a < x < b) = F(b) - F(a) = P(x \pm \Delta x)$$

Tomando límites será

$$\lim_{\Delta x \rightarrow 0} P(x \pm \Delta x) = \int_a^b f(x) dx = F(b) - F(a)$$

Esa es la función distribución para el caso de variable continua. Por su parte, ahora se puede definir una *función densidad de probabilidad* o *función frecuencia* con:

$$f(x) = \frac{dF(x)}{dx} \quad \text{donde:}$$

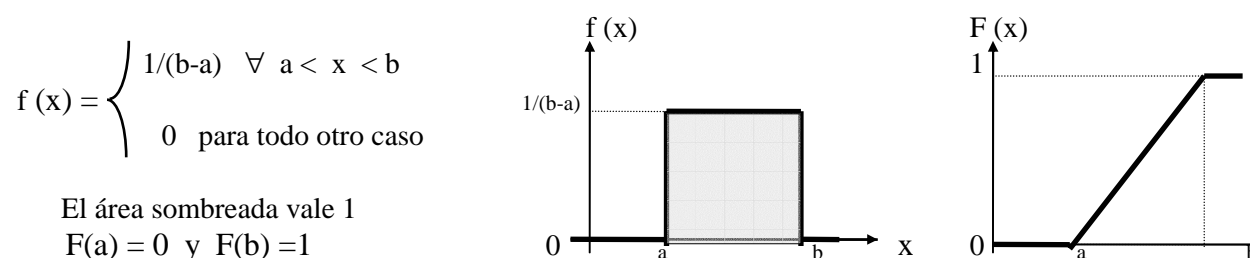
$dF(x)$: es un elemento de probabilidad
 dx : es un pequeño incremento de la variable

Integrando la función frecuencia para todo el campo real resulta ser:

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

Análogamente, si $a \rightarrow \infty$ es $\int_{-\infty}^b f(x) dx = F(b)$

Gráfico 8.3: Función frecuencia y función distribución para la función “diente de sierra”.



8.2.3 Distribuciones conjuntas e independencia

Cuando se trabaja con dos variables aleatorias X e Y se trata de un espacio bidimensional, donde si las respectivas funciones de distribución $P(X \leq x)$ y $P(Y \leq y)$ existen para todo x y para todo y , entonces la función distribución *conjunta* de X e Y se define con:

$$F_{x,y}(x, y) = P[(X \leq x; Y \leq y)] = \iint f_{x,y}(x, y) dx dy \text{ (continua)}$$

$$F_{x,y}(x, y) = P[(X \leq x; Y \leq y)] = \sum_i \sum_j P(X \leq x_k) P(Y \leq y_k) \text{ (discreta)}$$

Dos variables aleatorias X e Y son *independientes* (o mutuamente) si los eventos $X \leq x$ e $Y \leq y$ son mutuamente independientes para cada par de valores x e y . Esto significa que la distribución de valores de X no es afectada por los valores de Y , y viceversa. El concepto de independencia se puede expresar por la relación siguiente:

$$F_{x,y}(x, y) = P[(X \leq x; Y \leq y)] = P[X \leq x] \cdot P[Y \leq y]$$

Para variables discretas la función distribución es la sumatoria de los valores (la acumulada) y para las variables continuas se expresa con las integrales como se ilustró antes Generalizando si se tienen n variables aleatorias X_1, X_2, \dots, X_n la función distribución conjunta de todas ellas será:

$$F_{x_1, x_2, \dots, x_n}(x_1, x_2, \dots, x_n) = P[(X_1 \leq x_1; X_2 \leq x_2, \dots; X_n \leq x_n)] = P[X_1 \leq x_1] \cdot P[X_2 \leq x_2] \dots P[X_n \leq x_n]$$

Siempre y cuando la función exista o sea convergente.

8.3 Valor esperado

Se llama también *esperanza matemática* o *momento de primer orden*. Se trata de un operador matemático que al ser aplicado a la función probabilidad permite el cálculo de ese valor en el caso discreto, mientras que en el caso continuo se lo aplica a la función frecuencia:

$$\mathbb{E}(x) = \sum_i (x) P(X=x) \quad \text{en el caso discreto}$$

$$\mathbb{E}(x) = \int_{\mathbb{R}} (x) f(x) dx \quad \text{en el caso continuo}$$

Por ejemplo, si el experimento consiste en arrojar un dado normal, la variable aleatoria X podrá tomar seis valores equiprobables, y aplicando el operador anterior resultará:

$$\mathbb{E}(x) = \sum_1^6 (x) P(X=x) = 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) = 3,5$$

Notar que el valor esperado coincide con el promedio de los dos datos centrales 3 y 4, o sea que este no tiene por qué coincidir con uno de los valores de la variable. Observando atentamente la definición anterior, puede verse que el *valor esperado es un valor promedio, ponderado* con las probabilidades de ocurrencia de cada caso y una medida de estas probabilidades es la frecuencia relativa de cada una.

Sea ahora el caso del ejemplo visto en el Gráfico 8.3 anterior

$$\mathbb{E}(x) = \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_a^b x / (b-a) dx = 1/(b-a) \left[x^2/2 \right]_a^b = 1/(b-a) \{ b^2/2 - a^2/2 \} = (b+a)/2$$

El valor esperado resulta ser la semisuma de los límites del intervalo (una especie de centro de “gravedad” sobre el eje de abscisas, en términos probabilísticos).

Se demuestran fácilmente las propiedades siguientes:

1) El valor esperado de una constante es la constante:

$$\mathbb{E}(a) = a$$

2) El valor esperado del producto de una variable por una constante es igual al producto de la constante por el valor esperado de la variable.

$$\mathbb{E}(b \cdot x) = b \mathbb{E}(x)$$

3) El valor esperado de una variable aleatoria, definida por una función indirecta: $y = g(x)$, es distributivo respecto de la misma si el valor converge.

$$\mathbb{E}(y) = \mathbb{E}(g(x)) = g \{ \mathbb{E}(x) \}$$

Por ejemplo, si $y = a + bx$, entonces:

$$E(y) = E(g(x)) = E(a + bx) = E(a) + E(b \cdot x) = a + b E(x)$$

4) El valor esperado de dos variables independientes X e Y es igual al producto de los valores esperados de cada una de ellas:

$$E(X, Y) = E(X) \cdot E(Y)$$

Generalizando para n variables aleatorias independientes:

$$E(X_1, X_2, \dots, X_n) = \prod_1^n E(X_i) = E(X_1) \cdot E(X_2) \dots E(X_n)$$

Esta propiedad es equivalente a la ya vista como *regla del producto* en el caso del lanzamiento de una moneda al aire tres veces y esquematizada con el diagrama del árbol en el capítulo 5.

Tabla 8.1: Valores esperados en Procesos Bernoulli, Poisson e Hipergeométricos.

Probabilidad $P(x)$	Valor esperado $\mu = E_B(x)$
Binomial	$n \cdot p$
Pascal	$r \cdot p$
Binomial Negativa	$r \cdot q / p$
Geométrica	q / p (Odds de fracaso)
Hipergeométrica	$n \cdot R / N$
Poisson	μ

8.3.1 Aplicaciones de valor esperado

A continuación se presentan algunas aplicaciones muy comunes en este tema.

Ejemplo 1) Un bioquímico sabe por experiencia que el 20% de sus pacientes son enviadas a efectuarse un análisis de tipo Gravindex, los cuales le dejan una ganancia de \$ 2,7 según sus cálculos. Si en su primer día del mes tuvo 20 pacientes:

- ¿Cuánto espera ganar diariamente por los Gravindex?
- ¿Cuál es la probabilidad de tratar 5 pacientes antes de tener primer Gravindex?
- ¿Cuál es el número esperado de pacientes antes de tener el primer Gravindex?

a) Se estima que se trata de un proceso tipo Bernoulli donde hay una probabilidad de éxito $p=0,2$ y la cantidad diaria de pacientes es $n = 20$ por día, luego:

El valor esperado de prácticas hematológicas será $\mu = n \cdot p = 20 \cdot 0,2 = 4$ Gravindex diarios; y su ganancia promedio esperada será $G = 2,7 \cdot 4 = 10,8$ \$ diarios por esa práctica.

b) Se trata de un caso de probabilidad geométrica con $r = 1$, $g = 5$ y $q = 0,8$, entonces
 $P_G (g=5 / p=0,2) = 0,2 (0,8)^5 = 0,066$

c) El valor geométrico esperado es: $\mu = q / p = 0,8 / 0,2 = 4$ pacientes
O sea, espera 4 pacientes antes del quinto Gravindex del día.

Ejemplo 2) A un hospital llegan pacientes por la mañana a efectuarse extracciones de sangre. Se ha medido la frecuencia de llegada de los mismos, en intervalos elegidos al azar en intervalos de 10 minutos. La distribución de probabilidad empírica se muestra en la tabla siguiente. Para definir cuántos puntos de atención deben prepararse. Se desea saber:

- El número esperado de pacientes en un tiempo de 10 minutos.
- Si se puede suponer que los arribos se producen según un proceso del tipo Poisson.

Número de pacientes:	X	0	1	2	3	4	5	6 ó más
Probabilidad empírica:	P(X)	0,15	0,25	0,25	0,20	0,10	0,05	0,0

a) El número esperado de pacientes es la media aritmética ponderada

$$\mu = 0 (0,15) + 1 (0,25) + 2 (0,25) + 3 (0,20) + 4 (0,10) + 5 (0,05) = 2 \text{ pacientes cada 10 minutos}$$

Suponiendo que se trata de un modelo Poisson, el valor esperado de la frecuencia de arribos será de 2 pacientes cada 10 minutos. Entonces, midiendo cuánto se demora en atender a cada paciente en promedio, se puede saber el tiempo de espera de cada uno y colocar cierto número de puntos de atención para: 1) lograr que nunca haya más de x pacientes en la cola de espera; o 2) lograr que nunca se espere más de n minutos en la cola. Los modelos matemáticos para resolver estas dos cuestiones se ven en la Teoría de Colas (Investigación Operativa).

b) Una vez obtenido el valor esperado, se puede aplicar la fórmula para calcular la probabilidad Poisson, usando la manera iterativa, y comparar con la empírica.

Número de pacientes:	X	0	1	2	3	4	5	6 ó más
Probabilidad Poisson:	$P_{PO}(X)$	0,135	0,27	0,27	0,18	0,09	0,036	0,01

La concordancia entre la probabilidad empírica (O_i : observada) y la teórica (E_i : esperada) parece ser buena a simple vista. Sin embargo, se debe efectuar una *validación estadística* para comprobarlo como se mostrará más adelante en el tema Pruebas de Bondad de Ajuste.

Ejemplo 3) Se recibe un lote de mil ampollas con un cierto medicamento en una compra realizada. Para revisar la compra el farmacéutico decide controlar a 20 de ellos tomados al azar. Si el proveedor dice que el porcentaje de fallas no supera el 0,5% ¿cuántas ampollas falladas se espera encontrar en la muestra a revisar?

Es un caso de probabilidad Hipergeométrica, por lo tanto su valor esperado se obtiene con:

$$\mu = n \cdot R / N = 20 \cdot 5 / 1000 = 0,1 \quad \text{significa que no debería aparecer ninguna fallada.}$$

8.3.2 Aplicaciones en Epidemiología: OR y RR

En Epidemiología se emplean dos índices básicos llamados Odds Ratio (OR) y Riesgo Relativo (RR) que se usan de acuerdo al tipo de estudio que se está efectuando. En Medicina el *riesgo* puede ser definido como la probabilidad de que ocurra un determinado suceso que implique un peligro para la salud de un paciente. Tal como el riesgo de contraer una septicemia luego de ser operado, etc. El *factor de riesgo* es una magnitud cualitativa, de tipo dicotómica, donde sus dos resultados posibles indican si el factor de riesgo (o exposición) está presente o no. Usualmente esto se denota con: SÍ-NO. Son ejemplos de factor de riesgo características tales como: edad, sexo, tratamiento profiláctico antes de una cirugía, inmunización, medicación preventiva, etc. Cuando un grupo de pacientes recibe una droga que se está probando, será el caso SÍ del factor de riesgo, y el otro grupo usado como control o placebo será el caso NO. En cambio, cuando el primer grupo es cateterizado en su cirugía se piensa que está expuesto a una infección (SÍ) y otro grupo donde no fue necesaria (NO) será el de control.

La comparación del riesgo es uno de los objetivos principales de los estudios epidemiológicos, y su cuantificación implica la medición experimental de sus dos índices asociados: *Riesgo Relativo* (RR) y "*Odds-Ratio*" (OR).

El diseño experimental para medir ambos índices es el planteo de una Tabla de Continencia del tipo 2x2. Cuando el investigador es quien decide el tamaño muestral y cuántos serán expuestos al factor de riesgo, es un caso de *muestreo basado en la exposición*. En cambio, cuando el investigador decide cuantos sujetos enfermos y sanos entran en su estudio se trata de un *muestreo basado en la enfermedad*. Por ejemplo, suponiendo que toda la población estudiada es homogénea en todas sus características excepto en una, luego una parte de esta se expone a un factor de riesgo que se cree es de importancia en causar una determinada enfermedad, y el diseño experimental usado generalmente se muestra esquemáticamente en la Tabla 15.2 siguiente:

Tabla 8.2: Frecuencias observadas entre la exposición a un factor de riesgo y la contracción de una enfermedad cualquiera.

Factor de Riesgo	Enfermedad		Total
	SÍ	NO	
O	a	b	$n_1 = a+b$
Característica	c	d	$n_2 = c+d$
Grupal	Total $n_3 = a+c$	$n_4 = b+d$	n

$$RR = (a \cdot n_2) / (c \cdot n_1)$$

$$OR = (a \cdot d) / (b \cdot c)$$

Riesgo relativo (RR) se define como el cociente entre la probabilidad de contraer la enfermedad de la población expuesta y la probabilidad de contraerla de los no expuestos.

$$RR = P(E) / P(noE) = (a/n_1) / (c/n_2) = (a \cdot n_2) / (c \cdot n_1)$$

que es la expresión de más arriba.

Cuando no hay pacientes de la población expuesta que contraigan la enfermedad ($a = 0$) el RR se anula. En cambio, si ningún paciente de los no expuestos contrajo la enfermedad ($c = 0$) entonces el RR se vuelve infinito. Si los factores son *independientes* el $RR = 1$. El RR solo puede ser estimado con estudios prospectivos. La interpretación se puede ver mejor con un ejemplo: se

sabe que entre los que sufrieron infarto de miocardio el nivel de colesterol medio es de 300 mg/dl. En la población general el nivel es de 200 mg/dl. Se conoce la distribución de colesterol en ambas poblaciones, la de los infartados y la de la población humana usadas como referencia. Desde el punto de vista médico se toma como valor referencial 240 mg/dl, se considera a un paciente con “colesterol alto” cuando supera dicho valor. Luego se cuentan los casos de encontrados con colesterol alto en los infartados y entre los no infartados. Sea por caso, una relación de $120/180 = 2/3$ entre los infartados y de $1000/9000 = 1/9$ entre los no infartados, entonces se calcula el $RR = (2/3) / (1/9) = 6$. Eso se interpreta así: “Si un paciente tiene un colesterol mayor que 240, su chance de infarto es 6 veces mayor que si tuviese un valor de 200 o menos”. Así se construyen las simplificaciones en los diarios y revistas de divulgación.

Si se trata de un estudio del tipo caso-control, generalmente el RR no puede calcularse y se necesita del OR como una medida de asociación entre ambos factores analizados. Entonces, el *Odds Ratio* (OR) se define como el cociente entre dos “odds” posibles. Un “*odd*” es la relación entre la cantidad de “enfermos” y los “no enfermos” de una población dada. Como hay dos poblaciones, la expuesta y la no expuesta al factor de riesgo, hay dos tipos de “*odd*” posibles y la tasa entre ambos es el valor de OR. Sea:

$P_1 = a / n_1$: la probabilidad de contraer la enfermedad de la población expuesta.

$(1 - P_1) = b / n_1$: la probabilidad de no contraer la enfermedad de la población expuesta

El “*odd*” de la población expuesta es $O_1 = P_1 / (1 - P_1) = a / b$. A su vez,

$P_2 = c / n_2$: la probabilidad de contraer la enfermedad de la población no expuesta.

$(1 - P_2) = d / n_2$: la probabilidad de no contraer la enfermedad de la población no expuesta.

El “*odd*” de la población no expuesta es $O_2 = P_2 / (1 - P_2) = c / d$.

Por lo tanto, su cociente es $OR = O_1 / O_2 = (a \cdot d) / (c \cdot b)$ que es la expresión vista en la Tabla 8.2 de más arriba. Cuando el $OR = 1$ significa que ambos factores estudiados son independientes entre sí. Si a o d se anulan entonces el $OR = 0$; en cambio, si se anulan c o b , se hace infinito. Para el caso usado como ejemplo para el RR, se puede calcular:

$$OR = (120 / 1000) / (60 / 8000) = 16$$

Y se interpreta así: “cuanto más alto el OR, peor”. Otra forma de ver este índice es como el cociente de dos pares de probabilidades. Una es el RR entre los expuestos y el otro es el RR entre los no expuestos:

$$RR1 = \frac{Pe}{(1-Pe)} = \frac{a/(a+b)}{b/(a+b)} = a/b \quad \text{y} \quad RR2 = \frac{Pno\ e}{(1 - Pno\ e)} = \frac{c/(c+d)}{d/(c+d)} = c/d$$

Luego es: $OR = RR1 / RR2$

La diferencia en el uso de uno u otro índice reside en el tipo de investigación que se está realizando. *RR* se emplea en un estudio de *incidencia*, donde la frecuencia de algún resultado se cal-

cula entre dos grupos determinados por la presencia o ausencia de alguna característica. En cambio, *OR* se usa en un *estudio de caso-control* donde los resultados se obtienen en dos grupos, uno expuesto a un factor de riesgo y el otro usado como placebo o control.

El valor esperado de una variable aleatoria discreta X se denota con $E(X)$. Cuando la sumatoria abarca todos los valores posibles de X , la definición será

$$E(X) = \sum_j X_j P(X = X_j)$$

Y cuando se aplica a n variables aleatorias mutuamente independientes, el valor esperado es

$$E\left(\prod_{j=1}^n X_j\right) = \prod_{j=1}^n E(X_j) \text{ Para este caso donde } n = 4 \text{ será:}$$

$$E(X_1 X_2 X_3 X_4) = E(X_1) E(X_2) E(X_3) E(X_4)$$

Por ejemplo si es : $X_1 = a$; $X_2 = d$; $X_3 = 1/b$; $X_4 = 1/c$ entonces el valor esperado de *OR* será

$$E(OR = X_1 X_2 X_3 X_4) = E(a) E(d) E(1/b) E(1/c) = [E(a) E(d)] / [E(b) E(c)]$$

Por otra parte se puede obtener el valor esperado de acuerdo a la definición de independencia:

$P(A \cap B) = P(A) \cdot P(B)$ donde el evento A son los individuos expuestos y B los sanos

Donde $P(A) = n_1 / n$ y $P(B) = n_3 / n$ de acuerdo a la notación de la Tabla 8.2

$P(A \cap B) = E(a) / n$ de acuerdo a la definición clásica de probabilidad, entonces reemplazando

$$E(a) / n = (n_1 / n) \cdot (n_3 / n) \quad \text{O sea} \quad E(a) = (n_1 \cdot n_3) / n$$

Si hay independencia el valor esperado de cada celda de la Tabla se puede calcular como el producto de sus totales marginales dividido el total muestral, como ya se vio en 6.6. Análogamente:

$$E(b) = (n_1 \cdot n_4) / n \quad E(c) = (n_2 \cdot n_3) / n \quad E(a) = (n_2 \cdot n_4) / n$$

Reemplazando en el valor esperado de *OR* resulta:

$$E(OR) = [E(a) E(d)] / [E(b) E(c)] = 1$$

Y la conclusión es: Si el factor es independiente (matemáticamente hablando) de la enfermedad el valor esperado de *OR* es igual a 1, como ya se vio en el punto 6.6. Análogamente:

$$E(RR) = [E(a / n_1)] / [E(c / n_2)] = [E(a) / n_1] / [E(c) / n_2] = 1$$

El factor de riesgo, o de protección en el caso de una vacuna, puede ser cualquiera. Sin embargo, cuando se considera a un método de diagnóstico (o a un test clínico) como si fuese el factor de riesgo, entonces la tabla epidemiológica usual, se transforma en una Tabla de la Verdad y pueden hallarse relaciones entre ambos índices.

8.3.3 Caso especial: riesgo en los test clínicos

Cuando un test clínico es considerado como el factor de riesgo, entonces el factor Enfermedad (presente o ausente) divide a la muestra en dos conjuntos: sanos y enfermos. Por su parte el factor de riesgo ahora es el método diagnóstico y cuando está presente equivale al caso positivo; en cambio si está ausente equivale a un caso negativo. En el gráfico siguiente se muestra la equivalencia entre ambas tablas

Gráfico 8.4: Tabla epidemiológica y Tabla diagnóstica

Tabla epidemiológica

		Enfermedad		Total
		SÍ	NO	
Factor de Riesgo	SÍ	a	b	$n_1 = a+b$
	NO	c	d	$n_2 = c+d$
Total		$n_3 = a+c$	$n_4 = b+d$	n

Tabla diagnóstica

		Enfermedad		Total
		SÍ	NO	
Factor de Riesgo Test Clínico	(+)	vp	fp	TP
	(-)	fn	vn	TN
Total		TE	TS	N

En el apartado 6.3.1 se mostró una simulación donde siempre se obtenían dos índices básicos (sensibilidad y especificidad) no importa de que manera se seleccionaban las muestras. Esto significa que es lo mismo una selección basada en la enfermedad (Caso control) que una selección basada en la exposición (estudio por Cohorte o Clinical trial), para el caso especial donde el factor de riesgo es el método del diagnóstico o test clínico. Por ejemplo, para determinar si un individuo estuvo expuesto alguna vez a la toxoplasmosis, se debe efectuar un análisis clínico por inmuno fluorescencia. Y si estuvo expuesto en su pasado, el resultado da positivo. Se define:

Diagnostic Ratio (DR): Es el OR cuando el factor de riesgo es un método de diagnóstico.

Se cambia la denominación para recordar que el DR es un caso especial del OR. Se puede calcular de tres maneras diferentes, según el significado que se busca, pero siempre se encuentra el mismo resultado:

$$DR = (vp \cdot vn) / (fp \cdot fn)$$

Significado estadístico 1) Es el odd de una enfermedad cuando se la predice (o diagnostica) dividido el odd de la enfermedad cuando no es predicha (o diagnosticada).

Para este caso se pueden suponer dos eventos opuestos: **A** (un enfermo con un diagnóstico positivo) y **B** (un enfermo con un diagnóstico negativo). Y sus probabilidades son:

$$P(A) = P(TE \cap VP) / P(VP) = vp / (vp + fp)$$

$$P(B) = P(TD \cap VN) / P(VN) = fn / (vn + fn)$$

$$\text{Odds (A)} = P(A) / [1 - P(A)] = vp / fp$$

$$\text{Odds (B)} = P(B) / [1 - P(B)] = fn / vn$$

Entonces el cociente de ambos términos es:

$$\text{DR} = \text{Odds (A)} / \text{Odds (B)} = (vp \cdot vn) / (fp \cdot fn)$$

Significado estadístico 2) Es el odd de no-enfermedad cuando se la diagnostica, dividido el odd de la no-enfermedad cuando no es diagnosticada.

Para este caso se pueden suponer dos eventos opuestos: **C** (un sano con un diagnóstico negativo) y **D** (un sano con un diagnóstico positivo). Y sus probabilidades son:

$$P(C) = P(TS \cap TN) / P(TN) = vn / (vn + fn)$$

$$P(D) = P(TS \cap TP) / P(TP) = fp / (vp + fp)$$

$$\text{Odds (C)} = P(C) / [1 - P(C)] = vn / fn$$

$$\text{Odds (D)} = P(D) / [1 - P(D)] = fp / vp$$

Nuevamente el cociente de estos dos términos es:

$$\text{DR} = \text{Odds (C)} / \text{Odds (D)} = (vp \cdot vn) / (fp \cdot fn)$$

Significado clínico: Es el cociente entre los dos Likelihood Ratios (LR+ dividido LR-)

$$\text{LR+} = S / (1 - E) = (vp / TE) / [1 - (vn / TS)] = (vp / TE) / (fp / TS)$$

$$\text{LR-} = (1 - S) / E = [1 - (vp / TE)] / (vn / TS) = (fn / TE) / (vn / TS)$$

$$\text{DR} = \text{LR+} / \text{LR-} = S \cdot E / [(1-S)(1-E)] = (tp \cdot tn) / (fp \cdot fn)$$

Por ejemplo, si DR = 9 significa que los odd de enfermedad del método cuando diagnostica la enfermedad es nueve veces mayor que cuando no la diagnostica. O más simple, el LR+ es nueve veces mayor que el LR-. Por lo tanto, el DR puede ser considerado como un índice de calidad intrínseco al método que no varía con la prevalencia, ya que es función solo de S y E.

Por su parte, el riesgo relativo puede ser definido desde un punto de vista estadístico como: *La probabilidad de una enfermedad cuando es diagnosticada, dividida por la probabilidad de la enfermedad cuando no es diagnosticada:*

$$\text{RR} = [vp / (vp + fp)] / [fn / (vn + fn)] = (vp \cdot TN) / (fn \cdot TP)$$

Por ejemplo, si RR = 10 indica que la probabilidad de predecir la enfermedad es diez veces más grande que la de predecir la no-enfermedad. Por su parte, clínicamente se lo puede definir como:

$$RR = VPP / (1 - VPN)$$

Por lo tanto, este índice variará de acuerdo a la prevalencia de la población de referencia en la cual es usado, y puede ser considerado como una variable de la calidad del método clínico. Es conveniente informar su valor con una curva en lugar de un número. Puede verse, que para el caso de una enfermedad muy rara, como la meningitis, los valores vp y fp serán muy chicos y su producto será casi nulo. Por eso, los epidemiólogos lo aproximan con OR. Pero esto es solo una aproximación matemática porque conceptualmente son conceptos muy diferentes. Se puede encontrar la relación entre ellos, para este caso en particular con:

$$DR = RR + [(RR - 1) (LR+) \cdot O] \quad \text{donde el odd de enfermedad es: } O = p / (1-p), \text{ o bien:}$$

$$RR = [DR + (LR+) \cdot O] / [1 + (LR+) \cdot O]$$

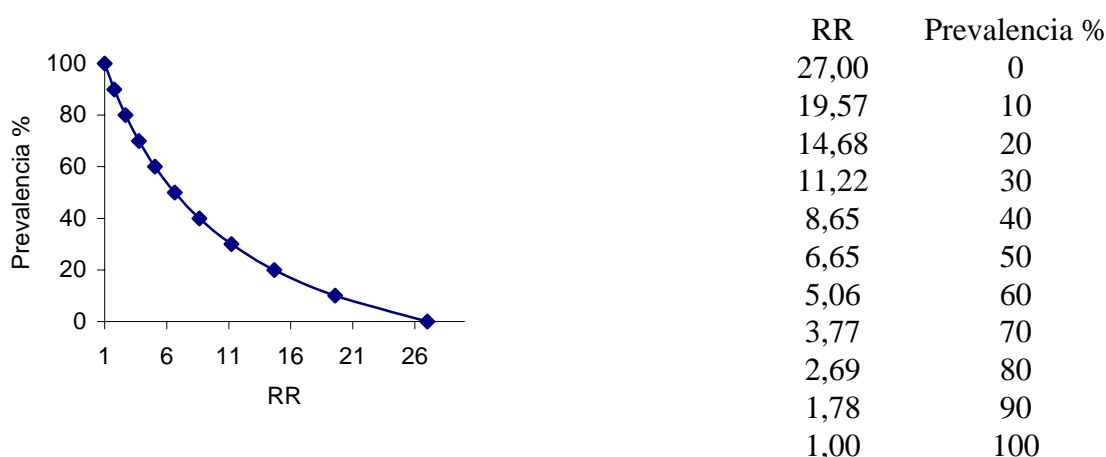
Como DR es una constante, entonces se puede deducir de la última ecuación que:

- Cuando p tiende a cero, también O tiende a cero y así $RR \approx DR$
- Cuando p tiende a uno, entonces O tiene a infinito y así RR tiene a uno.

Esto significa que, cuando la enfermedad es muy rara p tiene a cero, y cuando hay una epidemia muy grande entonces p tiene a uno. Para mostrar la variabilidad de RR con la prevalencia, se puede ver un caso donde $S = 0,9$ y $E = 0,75$. En el Gráfico 8.5 se puede ver que cuando la prevalencia es nula el valor de RR es uno, a medida que la prevalencia disminuye el valor de RR aumenta hasta un máximo de $RR = DR = 27$ cuando la prevalencia es nula.

Gráfico 8.5: Variabilidad del Riesgo Relativo con la prevalencia en la población

Método clínico: $S = 0,9$ y $E = 0,75$



La simulación para obtener el gráfico anterior se comienza usando los valores de sensibilidad y especificidad del método, que son conocidos: $S = 0,9$ y $E = 0,75$. Con ellos quedan determinados los valores: $DR = 27$ y $LR+ = 3,6$. Luego se toma un valor $p = 0,1$ y se calcula el RR respectivo con la ecuación de más arriba y se obtiene $RR = 19,57$. Y así sucesivamente, para valores $p = 0,2; 0,3; 0,4; \dots; 0,9$ con los cuales se calculan los respectivos valores de RR. Tal como se muestra en la tabla de datos del gráfico anterior. Todos los demás índices clínicos se pueden obtener a partir de S y E como se vio antes.

8.4 Momentos de orden k

Se llama *momento de orden k, respecto de un punto c* de una variable aleatoria, a un operador matemático que al ser aplicado a la función probabilidad, permite el cálculo de ese valor en el caso discreto, mientras que para el caso continuo se lo aplica a la función frecuencia:

$$\mu_k = \Xi \{ (x - c)^k \}$$

Por ejemplo, si $k = 0$ es $\mu_0 = 1$
 si $k = 1$ es $\mu_1 = \Xi (x) - c = \mu - c$
 si $k = 2$ es $\mu_2 = \Xi \{ (x - c)^2 \} = \Xi \{ (x^2) \} - 2 \mu c + c^2$

Si se toman momentos *respecto al origen*, esto es $c = 0$, entonces resulta

Por ejemplo, si $k = 0$ es $\mu'_k = \Xi \{ (x)^k \}$
 $\mu'_0 = 1$
 si $k = 1$ es $\mu'_1 = \Xi (x) - 0 = \mu$
 si $k = 2$ es $\mu'_2 = \Xi \{ (x - 0)^2 \} = \Xi \{ (x^2) \}$

Si se toman momentos respecto al valor esperado, se llaman *momentos centrales* de orden k, como $\mu = c$ resultan ser

$$\mu''_k = \Xi \{ (x - \mu)^k \}$$

Por ejemplo, si $k = 0$ es $\mu''_0 = 1$
 si $k = 1$ es $\mu''_1 = \Xi (x) - \mu = 0$
 si $k = 2$ es $\mu''_2 = \Xi \{ (x - \mu)^2 \} = \Xi \{ (x^2) \} - 2 \mu \mu + \mu^2 = \Xi \{ (x^2) \} - \mu^2$

En particular, el más importante es este último caso; se define como *varianza* de una función de probabilidad al momento central de segundo orden de la misma

$$\sigma^2 = \Xi \{ (x^2) \} - \mu^2$$

$$\sigma^2 = \sum_i (x - \mu)^2 P(X=x) \quad \text{en el caso discreto}$$

$$\sigma^2 = \int_{\mathfrak{R}} (x - \mu)^2 \cdot f(x) dx \quad \text{en el caso continuo}$$

Otra forma de escribirlo es:

$$\sigma^2 = \Xi \{ (x^2) \} - \Xi^2 \{ (x) \}$$

Resulta sencillo verificar las propiedades siguientes:

1) La varianza de una constante cualquiera es nula:

$$\sigma^2 = \Xi \{(a^2)\} - \Xi^2 \{(a)\} = a^2 - a^2 = 0$$

2) La varianza del producto de una variable por una constante es igual al producto del cuadrado de la constante por la varianza de la variable. Si $x = a \cdot y$

$$\sigma_x^2 = \Xi \{(ay)^2\} - \Xi^2 \{(ay)\} = a^2 \Xi \{(y^2)\} - a^2 \Xi^2 \{(y)\} = a^2 \sigma_y^2$$

3) La varianza de una variable aleatoria, definida por una función indirecta: $y = g(x)$. Es distributivo respecto de la misma si el valor converge. Por ejemplo, si $y = a + bx$, entonces:

$$\sigma_y^2 = \sigma^2(g(x)) = \sigma^2(a + bx) = \sigma^2(a) + \sigma^2(b \cdot x) = b^2 \sigma_x^2$$

Los momentos de órdenes superiores suelen usarse como índices de dispersión y asimetría. La varianza es el índice de dispersión por excelencia, mientras que los momentos de tercer orden indican la desviación en simetría de la función de probabilidad respecto de un eje cualquiera. Esto se verá más adelante un poco mejor.

Se puede calcular la varianza para las funciones de probabilidad ya vistas mediante un simple cálculo. En la Tabla 8.3 siguiente se presenta un cuadro con las varianzas de las probabilidades de los procesos Bernoulli, Poisson e Hipergeométrico.

Tabla 8.3 : Varianza en Procesos Bernoulli, Poisson e Hipergeométrico.

Probabilidad $P(x)$	Varianza $\sigma^2(x)$
Binomial	$n \cdot p \cdot q$
Pascal	$r \cdot q / p^2$
Binomial Negativa	$r \cdot q / p^2$
Geométrica	q / p^2
Hipergeométrica	$(N-n) n \cdot p \cdot q / (N-1)$
Poisson	μ

Se puede aplicar esta tabla a los tres ejemplos del punto anterior:

Ejemplo 1) Caso de los Gravindex era Binomial: $\sigma^2(x) = n \cdot p \cdot q = 20 \cdot 0,2 \cdot 0,8 = 3,2$

Ejemplo 2) Caso de arribo de pacientes era Poisson: $\sigma^2(x) = \mu = 2$

Ejemplo 3) Caso de muestreo de aceptación era Hipergeométrico: $\sigma^2(x) = (N-n) n \cdot p \cdot q / (N-1)$
 $\sigma^2(x) = (1000-20) 20 (0,995)(0,005) / (1000-1) = 0,098$

8.4.1 Variables aleatorias tipificadas

Las propiedades vistas más arriba pueden usarse para transformar una variable, de manera tal de facilitar la operatoria y los cálculos. Es un artificio matemático para lograr que una variable aleatoria cualquiera sea tipificada se usa la relación:

$$z = (x - \mu) / \sigma$$

Sea $a = 1/\sigma$ y sea $b = -\mu/\sigma$, entonces la variable $z = ax + b$, valdrá $z = (x - \mu) / \sigma$ y aplicando los operadores valor esperado y varianza se calcula:

$$\mathbb{E}(z) = \mathbb{E}\{(x - \mu) / \sigma\} = 1/\sigma (\mathbb{E}(z) - \mu) = 1/\sigma(\mu - \mu) = 0$$

$$\sigma_z^2 = \mathbb{E}\{(z^2)\} - \mathbb{E}^2\{(z)\} = \mathbb{E}\{((x - \mu)/\sigma)\} - \mathbb{E}^2\{((x - \mu)/\sigma)\} = 1/\sigma_x^2 \mathbb{E}\{(z^2)\} - (\mu - \mu)^2 / \sigma$$

$$\sigma_z^2 = (1/\sigma_x^2) \sigma_x^2 = 1$$

Entonces, una variable tipificada se caracteriza por tener:

$$\left. \begin{array}{l} \text{Variable tipificada} \end{array} \right\} \begin{array}{l} \cdot \text{ Valor esperado nulo} \\ \cdot \text{ Varianza unitaria} \end{array}$$

8.5 Aplicaciones en Bioquímica

Existen muchas aplicaciones prácticas de los momentos estadísticos. La mayoría de ellas se refieren a los valores económicos esperados o a los productos fabricados en serie. De entre ellos, se seleccionaron los siguientes.

8.5.1 Índice de Agregación

Es un índice de muchas aplicaciones en campos diversos tales como biología, botánica, etología, ecología, etc. Mide el grado de agrupamiento de los individuos en una determinada región del espacio. Cuando este índice tiene un valor cercano a la unidad, significa que el número de individuos se distribuye al azar, en la unidad de espacio que ocupan (ver Gráfico 8.6); es decir, se ubican en el espacio siguiendo una distribución del tipo Poisson. En cambio, cuando es mucho mayor que uno significa que los individuos tienden a presentarse agrupados, como las colonias de bacterias, los cardúmenes de peces, grupos de hematíes, manadas, una especie de árbol en el bosque, etc. Puede ocurrir que, a su vez, el conjunto de individuos se distribuya al azar dentro del continuo en forma grupal. Esto es, los grupos se distribuyen en el continuo según una distribución de Poisson, pero los individuos dentro del grupo no lo hacen sino que muestran una especie de “distribución contagiosa”. Tal como el fenómeno de apilamiento celular en las cámaras de recuento, o una aglutinación, etc. Por regla general, se trata de una distribución del tipo binomial negativa. Finalmente, cuando el índice de agregación es muy pequeño, próximo a cero, se trata de una distribución uniforme dentro de la ubicación espacial considerada. El

se trata de una distribución uniforme dentro de la ubicación espacial considerada. El número de individuos por unidad de espacio es una constante, la varianza es casi nula. Esto muestra claramente una distribución artificial de los individuos y no una natural, como el caso de reforestación plantando árboles alineados, un campo sembrado, etc.

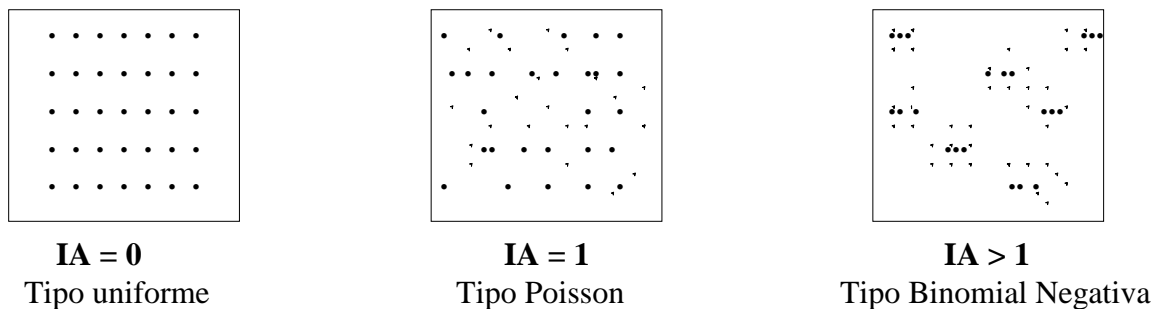
Se define como *Índice de Agregación* al cociente entre la varianza y el valor esperado de una variable aleatoria x cualquiera, o sea:

$$IA = \sigma_x^2 / \mu_x$$

Distribución de Poisson: $\sigma_x^2 = \mu_x \Rightarrow IA = 1$
 Distribución Binomial Negativa: $\sigma_x^2 = r \cdot q / p^2$ y $\mu_x = r \cdot q / p \Rightarrow IA = 1 / p$
 Distribución Uniforme: $\sigma_x^2 \approx 0 \Rightarrow IA \approx 0$

En el gráfico siguiente se esquematizan las tres situaciones planteadas, una plantación artificial muestra una distribución espacial uniforme y resulta nulo el Índice de agregación; en cambio, al mirar por el microscopio una dilución de glóbulos de sangre en el Hemocitómetro se puede ver una figura como la del cuadro del medio, donde $IA = 1$ pues se trata de una distribución poissoniana. Pero si no se le agrega EDTA en forma adecuada, las células se atraen por carga eléctrica, produciéndose el efecto *Rolleaux*; se agrupan como en un apilamiento y entonces el índice crece a valores mayores que la unidad. En Ecología, este agrupamiento de individuos se ve en los cardúmenes de peces, o en las manadas de animales, bandadas de pájaros, etc., según el continuo sea el agua, la tierra o el aire. Hay una especie de “contagio” entre los integrantes del grupo.

Gráfico 8.6 Índice de Agregación en diferentes casos.



8.5.2 Muestreo de aceptación

En una relación de compraventa es usual tener un comprador que recibe un lote de piezas o elementos de un vendedor que se las provee. El problema básico es decidir si el comprador le acepta o no la mercadería enviada por el vendedor, de acuerdo con el resultado que obtenga al efectuarle una inspección para determinar la calidad de la misma. Cuando se trata de piezas muy valiosas como los diamantes, el comprador revisará con mucho cuidado cada pieza, una por una. Pero cuando se trate de una cantidad demasiado grande de elementos, como arroz a granel, o cuando el ensayo para probarla es destructivo como en el caso de un fósforo, o cuando el costo de la inspección sea muy elevado, entonces no revisará todo el lote sino una muestra del mismo.

El procedimiento usual se denomina *muestreo de aceptación* y consiste en tomar una muestra al azar de n unidades del lote comprado de tamaño N , revisar las piezas elegidas, y si se encuentran r o más piezas defectuosas se rechaza todo el lote. Caso contrario, se lo acepta.

Existen dos riesgos: el primero es *rechazar algo que se debería haber aceptado*; es el riesgo que corre el vendedor. Esto puede ocurrir porque en el muestreo de aceptación le encontraron las únicas piezas falladas del lote. A su probabilidad de ocurrencia se la suele llamar nivel de significación y se la simboliza con α . En Estadística, a ese tipo de equivocación se le designa con el nombre de Error de Tipo I. Por otra parte, la segunda manera de equivocarse es aceptar un lote con muchas piezas falladas porque en el muestreo de aceptación no apareció ninguna de ellas, esto es, *aceptar algo que se debería haber rechazado*. Ese es el riesgo del comprador y en estadística se lo llama Error de tipo II, y a su probabilidad de ocurrencia se la designa con el símbolo β (llamada la potencia del ensayo).

Cuando un profesional trabaja en un laboratorio de Control de Calidad en la industria farmacéutica, química, alimenticia, etc., se le presentará esta problemática en forma diaria. Por un lado, deberá hacer muestreos de aceptación para las compras de su industria, pero por otro lado deberá revisar la producción realizada para controlar la buena calidad del producto final que hace a la imagen de la empresa. Independientemente de los diferentes tipos de revisiones que le efectúe, el diagnóstico final será dicotómico: pieza aceptada o rechazada. Y el problema es que nunca se podrá conocer con certeza la cantidad total R de piezas falladas de todo el lote; solo una estimación de las mismas. Además, los riesgos de comprador y vendedor deben ser aceptables para ambas partes para que la operación de compraventa se concrete.

El procedimiento es simular el número de piezas falladas, para determinar los riesgos, como se muestra a continuación con un ejemplo numérico: un comprador recibe lotes de 250 unidades, de las cuales selecciona 50 al azar. Cuando encuentra 2 o más defectuosas rechaza el lote y se lo devuelve al vendedor. ¿Cómo se puede juzgar la bondad de esta regla de aceptación para ambas partes?

El método consiste en estimar la probabilidad de aceptar lotes con R defectuosos, tomando un valor esperado de $p = R/N$ (donde p es el porcentaje de defectuosos). De esta forma, dando diferentes valores a R se van obteniendo los porcentajes, y se puede graficar esta situación en una gráfica denominada *Curva Característica de Operación* del plan de muestreo específico. En este caso será una probabilidad hipergeométrica, en función del valor p , o sea:

$$P_H (r < 2 / N=250; R=250 p, n=50) = P_H(r = 0) + P_H(r = 1)$$

$$P_H(p) = \frac{C(250-250 p ; 50)}{C(250 ; 50)} + \frac{C(250 p ; 1) C (250-250 p ; 49)}{C(250 ; 50)}$$

Esta es la ecuación general de la probabilidad hipergeométrica en función del porcentaje de defectuosos. Ahora, dándole valores a p se obtienen los valores respectivos de $P_H(p)$ con los cuales se puede trazar la curva.

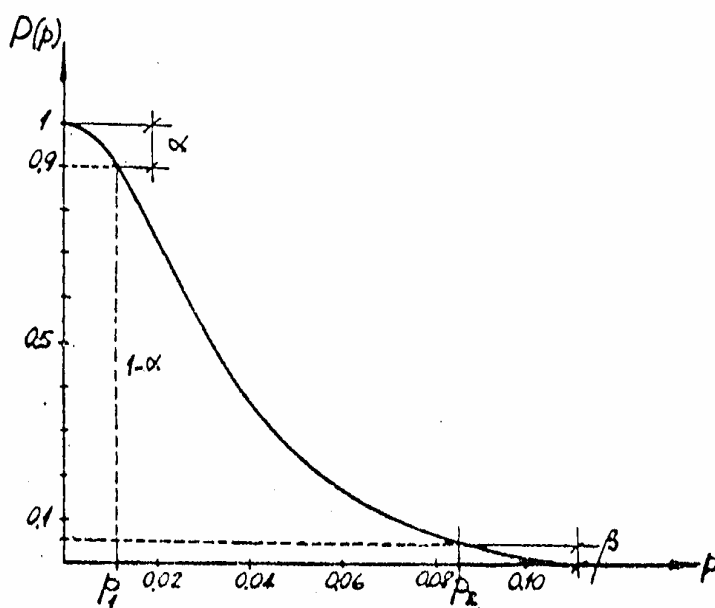
Por ejemplo, si $p = 2\%$ resulta:

$$P_H(p=0,02) = \frac{C(245; 50)}{C(250; 50)} + \frac{C(5; 1) C(245; 49)}{C(250; 50)} = \frac{245! 200!}{250! 195!} (1 + 250/196)$$

$$P_H(p=0,02) = \frac{245! 200! 446}{250! 195! 196} = 0,738 \quad \text{O sea, un 73,8\%}$$

Análogamente, para valores de p variando de 0,4% cada vez, se logra la tabla del Gráfico 8.7. Para juzgar un plan de muestreo es necesario definir el porcentaje de defectuosos que el comprador está dispuesto a aceptar con mucha frecuencia. Por ejemplo, si el comprador está dispuesto a tolerar un total de $R=3$ piezas defectuosas por lote, entonces será $p_1 = 0,012$ el llamado: *Límite de Calidad Aceptable* (LCA). Para este problema, con un LCA del 1,2% se corre un riesgo $\alpha = 0,10$ de rechazar lotes, mientras que $P_H(0,012) = 0,9$. Lo cual es bastante razonable para el vendedor que en el 10% de los casos le rechazarán injustamente su venta.

Gráfico 8.7 Curva característica de operación



P	$P_H(p)$
0	1
0,004	1
0,008	0,96
0,012	0,897
0,016	0,821
0,02	0,738
0,024	0,655
0,028	0,575
0,032	0,501
0,036	0,432
0,04	0,371
0,044	0,316
0,048	0,268
0,052	0,226
0,056	0,19
0,06	0,159
0,064	0,132
0,068	0,11
0,072	0,091
0,076	0,075
0,08	0,062
0,084	0,05
0,088	0,041
0,092	0,034
0,096	0,027
0,1	0,022

Además, para que sea razonable para la otra parte, debe ocurrir que el plan de muestreo rechace con mucha frecuencia lotes con un porcentaje p_2 de defectuosos muy elevado. Este es un porcentaje que se está dispuesto a tolerar solo en raras ocasiones y se lo denomina: *Tolerancia porcentual de defectuosos en el lote* (TPD). Para este caso, si se toma $p_2 = 0,084$, la probabilidad de aceptar un lote que debería haber sido rechazado, o sea el riesgo del comprador, entonces será $\beta = 0,05$. En el 5% de las veces aceptará lotes con más de 3 defectuosos en total.

8.6 Teorema Central del Límite

Llamado el teorema fundamental de la estadística. Es el pilar que tomó casi un siglo hasta llegar a su demostración formal completa. Laplace lo publicó en su primera versión en 1812 y recién en 1901 Liapounoff los demostró bajo condiciones más generales, para todos los casos posibles. Su enunciado resumido dice:

“Sean n variables aleatorias independientes x_i con $i = 1, 2, \dots, n$; cada una con distribuciones de probabilidad conocidas, en *condiciones muy generales*, la función distribución de la variable suma ($X = x_1 + x_2 + x_3 + \dots + x_n$), tipificada $\{Z = (X - \mu) / \sigma\}$ será aproximadamente igual a una función de Gauss tipificada $P_{\text{Gauss}}(z)$ si la secuencia de los x_i es ilimitada, esto es:

$$\lim_{n \rightarrow \infty} P \{((X - \mu) / \sigma) \leq a\} \approx F_{\text{Gauss}}(a)$$

En tal caso, se dice que la variable suma X es asintóticamente normal de parámetros $N(\mu; \sigma)$.

¿Qué significa en condiciones muy generales? ... Por ejemplo, se cumple en los casos:

- Es condición suficiente de validez, que todas las x_i tengan la misma función distribución.
- Si ocurre que no tienen la misma función distribución, entonces es suficiente que cada una de ellas contribuya al total, con un aporte insignificante.
- Cuando todas las variables tengan distribuciones de tipo continuo.
- Si no son independientes, el teorema se sigue cumpliendo bajo ciertas condiciones.

Como se aprecia, esto es tan amplio que la mayoría de los casos reales se tomaban como teniendo distribuciones gaussianas con tal de tener un número muy grande de experimentos. Eso constituye la llamada Teoría de Grandes Muestras. Se deduce que si el número de casos n es lo suficientemente grande, las distribuciones vistas (Binomial, Poisson, Hipergométrica, Pascal, etc.) todas tienden a la función de Gauss. Lo que permite hacer aproximaciones para simplificar los cálculos de probabilidad.

Cuando una variable no parece distribuirse normalmente se pueden hacer transformaciones que ayuden a solucionar el problema, como por ejemplo tomar el logaritmo, el arco seno, etc. Cuando X es una variable aleatoria, tal que una función no lineal del tipo $g(X)$ se distribuya normalmente, se dice que X tiene una *distribución normal transformada*.

Por ejemplo, si se supone que la indeterminación total (*error*) en una medición es producida por la suma de un gran número de causas aleatorias e independientes, cada una de ellas contribuyendo con un aporte despreciable frente al total (Hipótesis de Haeguen-Bessel), se está en el caso (b) descrito más arriba; por lo tanto, se puede asumir que la función distribución para los *errores casuales* es la función de Gauss o Normal, como se anunció al principio de este capítulo.

Desde su descubrimiento por Gauss en la primer década del siglo XIX, la curva se usó ampliamente en todas las mediciones físicas y químicas de aquel entonces. Prácticamente era la solución para cualquier caso práctico. Hasta 1907 donde Student muestra que el error de medición en el conteo con hemocitómetro (cámara de Neubauer), se parecía más a una distribución Poisson que a una de Gauss; fue el comienzo de la Estadística moderna.

8.7 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase.

- | | | |
|--|-------|-------|
| 1) Todo fenómeno que no se puede explicar con certeza es de tipo aleatorio. | V | F |
| 2) Un experimento es aleatorio cuando cada valor posible tiene una cierta probabilidad. | V | F |
| 3) Una variable es aleatoria cuando algunos valores tienen asociadas ciertas probabilidades. | V | F |
| 4) Si se conoce la función probabilidad, se conoce la función distribución. | V | F |
| 5) Si se conoce la función distribución, se puede determinar la función de probabilidad. | V | F |
| 6) La función distribución siempre es una sumatoria. | V | F |
| 7) La función frecuencia es una especie de densidad de probabilidad. | V | F |
| 8) La función densidad es una integral. | V | F |
| 9) La esperanza matemática es un momento de segundo orden. | V | F |
| 10) El valor esperado es un operador matemático. | V | F |
| 11) La varianza es el momento central de segundo orden de una función probabilidad. | V | F |
| 12) El riesgo relativo es lo mismo que el odd ratio. | V | F |
| 13) Explicar la diferencia entre ambos conceptos desde un punto de vista epidemiológico..... | | |
| 14) ¿ Que ocurre con los índices de riesgo OR y RR cuando el factor es un test clínico ? | | |
| 15) Explicar la variabilidad de los índices de riesgo con la prevalencia de la población | | |
| 16) Una variable tipificada tiene un valor esperado nulo. | V | F |
| 17) La varianza de una variable aleatoria tipificada es 2. | V | F |
| 18) Si la varianza es igual al valor esperado, se trata de un caso de Poisson. | V | F |
| 19) Cuando el índice de agregación es la unidad se trata de una distribución uniforme. | V | F |
| 20) Cuando se acepta algo que debía ser rechazado es un error de tipo I. | V | F |
| 21) Cuando se rechaza algo que debía ser aceptado es un error del tipo II. | V | F |
| 22) El error del tipo I es el riesgo del comprador. | V | F |
| 23) El error del tipo II es el riesgo del vendedor. | V | F |
| 24) La tolerancia porcentual de defectuosos en el lote es | | |
| 25) El Límite de Calidad Aceptables es | | |
| 26) Los valores esperados y las varianzas de un proceso Bernoulli y de Poisson son | | |
| 27) Una variable se tipifica si se le resta el valor esperado y se la divide por su varianza. | V | F |
| 28) El dominio de la función distribución son todos los números reales. | V | F |

2) Calcular las varianzas y los valores esperados para todos los ejemplos desarrollados en el presente capítulo.

3) A una Farmacia llega cada hora una cantidad de pedidos que se muestran en la tabla siguiente.

Se pide calcular para cada caso:

- El número esperado de llegadas por hora.
- La varianza de esta distribución de probabilidad.

Nº de pedidos	Probabilidad
0	0,05
1	0,10
2	0,15
3	0,25
4	0,30
5	0,10
6	0,05

NOTA: Usar la relación $\sigma^2 = \Xi \{ (x^2) \} - \Xi^2 \{ (x) \}$

4) Calcular el Riesgo Relativo y el Odds Ratio de las hipotéticas tablas siguientes:

Factor de Riesgo	Enfermedad		Total
	SÍ	NO	
SÍ	10	90	100
NO	10	90	100
Total	20	180	200

Factor de Riesgo	Enfermedad		Total
	SÍ	NO	
SÍ	10	10	20
NO	90	90	180
Total	100	100	200

Factor de Riesgo	Enfermedad		Total
	SÍ	NO	
SÍ	90	10	20
NO	90	10	180
Total	180	20	200

Factor de Riesgo	Enfermedad		Total
	SÍ	NO	
SÍ	90	90	180
NO	10	10	20
Total	100	100	200

A su vez calcular para cada caso la Prevalencia (probabilidad de estar enfermo), la probabilidad de exposición, el Odds de enfermos y el Odds de expuestos. Explicar clínicamente el significado de cada uno de los valores hallados para todos estos índices y extraer alguna conclusión.

5) Para los mismos datos de las cuatro tablas anteriores, suponer que el factor de riesgo es un método de diagnóstico (o un test clínico) y calcular:

- Los índices básicos de calidad S y E
- Los índices de calidad que no varían con la prevalencia (IY, LR+ y LR-)
- Los índices de calidad que varían con la prevalencia (A, VPP y VPN)
- Los índices de riesgo DR y RR.
- Mostrar la variabilidad de RR con la prevalencia en forma gráfica.
- Repetir los mismos cálculos que los del problema anterior para los datos siguientes:

Test Clínico	Enfermedad		Total
	SÍ	NO	
(+)	180	50	230
(-)	20	150	170
Total	200	200	400

9

La Normalidad

En este capítulo se trata el concepto de normalidad en Estadística. Se comienza con un primer cuadro comparativo del concepto, en diversas especialidades o campos, y se propone un método para evitar las confusiones comunes con la normalidad. Luego se analiza con cuidado la idea de “valores normales” en clínica y se da un método para su obtención de acuerdo con las recomendaciones de la bibliografía. Finalmente se trata a la función normal o de Gauss, sus propiedades, usos y manejo de tablas para el cálculo de probabilidades.

9.1 ¿Qué es lo normal?

Cada especialidad parece tener su definición particular del concepto de normalidad. Una de las más difundidas en medicina es considerar al paciente sano como normal. Sin embargo, la Organización Mundial de la Salud define a la *salud* como un estado e bienestar físico, psíquico y social, y no solamente como la ausencia de enfermedad o deterioro. Lo toma como un estado que permite al organismo adaptarse y funcionar en forma adecuada, a las condiciones endógenas y a los factores ambientales a los que está sometido. Por otra parte, el concepto de *enfermedad* puede definirse como una respuesta inadecuada del organismo, o un agotamiento de sus mecanismos de defensa y adaptación, una falta de respuesta a los estímulos a los que está expuesto. El proceso puede ser una perturbación de la estructura y/o función de un órgano, de un sistema o de la totalidad del organismo.

Según Murphy, E.A., del Departamento de Medicina de la J.Hopkins University, con muchos trabajos dedicados a este tipo de clasificaciones, se pueden identificar cinco categorías de manifestaciones que pueden considerarse como una enfermedad, y siete significados diferentes del concepto de normalidad. Además de estas siete definiciones teóricas, todavía se pueden agregar cinco más desde el punto de vista pragmático, como se puede apreciar el Cuadro 9.1 Es necesario detenerse para analizar estos aspectos.

Las manifestaciones que pueden considerarse como enfermedades son cinco:

- 1) Determinadas respuestas homeostáticas (inflamación no supurativa)
- 2) Determinados mecanismos homeostáticos aberrantes (hipertensión u obesidad)
- 3) Respuestas que impiden una adaptación, o bien, respuestas inhabituales frente a los factores que las provocan (golpe de calor, desfallecimiento, hipertensión de origen renal)
- 4) Una respuesta inadecuada de origen endógeno (agammaglobulinemia)
- 5) Una respuesta totalmente anárquica (cáncer o tireotoxicosis).

A toda esta diversidad de aspectos se pueden agregar otros aspectos operativos que permiten identificar individuos, en los cuales es aconsejable una acción médica o sanitaria en forma preventiva. Entonces, a cualquier paciente en alguna de estas situaciones se lo cataloga como “enfermo o no normal”.

Cuadro 9.1 Diferentes aspectos del concepto de normalidad.

Noción	Dónde se usa	Término preferido
Clasificación teórica de Murphy		
1. Probabilística: a) Población total b) Individuos sanos	Estadística	Gaussiano
2. Más representativa en su clase	Ciencias descriptivas	Promedio, modal
3. Comúnmente observable en su clase	Ciencias descriptivas	Habitual
4. La más apta para la supervivencia	Genética, Invest. Operativa	Optimo, más apto
5. Que no causa daños ni implica castigos	Medicina Clínica	Inofensivo, no nocivo
6. A lo que se aspira en general	Política, Sociología	Convencional
7. Más perfecto en su clase	Matemática, estética, moral	Ideal
Clasificación de acuerdo con los aspectos prácticos		
8. Que engloba a los individuos que pueden ser atendidos	Sociología, política, economía administración.	Aceptable
9. Efecto más provechoso que nocivo	Terapéutica, Org. de salud	Máximo efecto
10. Ausencia de signos externos raros	Diagnóstico Profesional	Sano
11. Idem anterior, pero percibidos por uno mismo.	Anamnesis	Saludable
12. Ausencia de signos inaceptables para el prójimo	Vida comunitaria	Admisible

Siguiendo el cuadro de Murphy se pueden tener siete respuestas diferentes a una simple pregunta como: *¿Cuál es el nivel normal de colesterol?*

- 1) Una distribución de probabilidades, en forma de campana, descrita por la función de Gauss con los valores de colesterol en la población de referencia (un estadístico).
- 2) El valor más representativo del colesterol en la población definido por un valor esperado, o promedio general (un demógrafo).
- 3) Los valores de colesterol más comunes definidos por una gamma de límites, o sea los habituales y conocidos “límites de normalidad de los laboratorios” (un bioquímico).

- 4) Los valores de colesterol más adecuados para la reproducción y para la supervivencia de la especie estudiada (un genetista).
- 5) Los valores de colesterol inocuos que causan poco o ningún daño al organismo (un médico).
- 6) La opinión conjunta de una comisión, o sea los valores “aprobados de colesterol” por las autoridades respectivas (un político).
- 7) El valor “ideal” de colesterol (un utópico).

El primer uso de la palabra correspondiente a la Estadística causa el mayor problema porque tiene poco que ver con los demás significados. No hay ninguna razón científica para que un grupo de personas “normales” tenga una distribución gaussiana en alguna magnitud clínica que se les esté midiendo. Lamentablemente, las estadísticas “gaussianas” se emplean frecuentemente para determinar los “límites normales” aplicados a las pruebas del laboratorio de análisis clínicos. Al respecto, Galen hace mención a las conclusiones de Benson, director del laboratorio médico de la Universidad de Minnesota:

“Los límites normales han tenido un papel indefinido, pero tranquilizante en el laboratorio clínico. Asoman en el horizonte de nuestra conciencia, perfectamente simétricos como el Monte Fujiyama, un poco nebulosos en su significado, pero aceptados y reverenciados con gratitud. Sin embargo, lejos de ser puros y simples, como una querida ilusión infantil, examinados de cerca resultan ser insoportablemente complicados, siendo en realidad uno de los problemas más difíciles y que más limitan la utilidad de los datos de un laboratorio clínico”.

Como alternativa al problema de los “límites normales”, varios investigadores han adoptado los *límites de referencia*, de uso cada vez más frecuente en la bibliografía y laboratorios, que se analizarán en el próximo punto.

En uno de los trabajos recomendados por la Clinical Chemistry para el análisis estadístico de los datos del laboratorio, Wu, Towmey y Thiers examinaron publicaciones previas, revelan que en virtud de la *costumbre* las técnicas estadísticas usadas casi invariablemente se basan en el supuesto de normalidad (distribución gaussiana de probabilidad) en la distribución de los datos obtenidos. Sin embargo, estos investigadores encontraron que a menudo los datos clínicos no adhieren a una función de Gauss. Entre los estadísticos hay dos corrientes de pensamiento. Un grupo, con el enfoque clásico, piensa que el supuesto gaussiano es lo suficientemente *robusto* como para aguantar discrepancias usuales, es decir que los modelos *paramétricos* para análisis de datos, basados en un supuesto de normalidad, sirven igual aun cuando tal supuesto no se cumpla estrictamente. Esto, por las condiciones generales del Teorema Central del Límite. Pero hay otra forma de pensar; entre los que no creen en tamaña robustez de los métodos clásicos, piensan que es mejor utilizar los modelos *no paramétricos*, los cuales no requieren el supuesto de normalidad aunque sean menos sensibles que los primeros. Depende de cada caso particular, y del área de estudio específica analizada, para saber cuál de las maneras es mejor. Estos investigadores estudiaron doce magnitudes clínicas (constituyentes bioquímicos) en suero, obtenidos de más de 1400 individuos en una muestra controlada de hombres y mujeres, y ninguno de ellos se ajustaba bien a una distribución normal.

A pesar de todas estas observaciones, es indudable que por mucho tiempo el supuesto de normalidad estadístico se seguirá usando en los laboratorios. Y por ello debe ser estudiado con cuidado, además del uso intensivo que tiene en la materia Estadística.

9.2 Criterios de normalidad

En principio existen tres criterios de normalidad diferentes, a tener en cuenta:

1. *El criterio matemático o estadístico puro.*
2. *El criterio de referencia a la población de donde se extrajeron los datos.*
3. *El criterio operativo y lógico.*

Los individuos identificados como anormales o enfermos son aquellos que deben recibir asistencia porque son susceptibles de ser tratados. Esta anormalidad se asocia a una patología definida por otros estudios. Por lo tanto, un valor “normal”, en el sentido de “sano”, de una función fisiológica o de una morfología determinada, puede tener esos tres significados:

1. Desde el punto de vista *estadístico* un valor puede ser *normal* por la ubicación que él tiene dentro de un intervalo, donde están la mayoría de las observaciones realizadas. Por ejemplo, se ubica en el intervalo que comprende al 95% de las personas consideradas con colesterol normal. Esta definición no es tan aceptable desde el punto de vista epidemiológico o médico. Tomando por caso una población que viva en un medio determinado, como ser el Altiplano boliviano, a miles de metros de altura respecto al nivel del mar, y que posea un patrimonio genético definido. La medida de una función (tensión arterial, número de glóbulos rojos, etc.) variará de un individuo a otro, pero se puede definir un intervalo que englobe al 95% y 99% de la población. Si se adopta en forma mecánica a esos valores como umbrales de normalidad, se corre el riesgo de que al examinar a un habitante costero se lo tome por anémico y se lo trate inútilmente. En las alturas, el efecto de apunamiento no le ocurre a los nativos porque su sangre tiene un número mayor de glóbulos rojos que el promedio de otros habitantes del globo. Al tener mayor capacidad de transportar oxígeno, el efecto de enrarecimiento del aire por la altura se amortigua grandemente, cosa que no le ocurre al habitante de las llanuras. Esta capacidad se transmite genéticamente de generación en generación, y pasa a ser una característica de la población del Altiplano. Comparados con ellos, todos los demás tendrán un hematocrito y un recuento de rojos mucho menor. Entonces, si se toma a los del Altiplano como referencia de normalidad, todos los otros parecerían anémicos. Esto conduce al siguiente punto.

2. La *normalidad* puede variar según las condiciones de vida y desarrollo de los individuos; por ejemplo, es un hecho probado que los descendientes de japoneses que habitan California son mucho más altos en promedio que los que habitan el Japón. La explicación básica parece ser la manera de alimentarse de unos y de otros. En USA, la cantidad ingerida de productos lácteos, cereales y otros alimentos vitaminizados artificialmente, es mucho más grande que la consumida en Japón, por lo que el desarrollo de los huesos no es tan grande. Entonces:

- Cuando una población vive en condiciones adecuadas los valores obtenidos pueden ser considerados óptimos, típicos o ideales, en ausencia de factores nocivos para la magnitud clínica analizada. Se tratará de valores normales en condiciones de vida normales.

- Cuando una población está expuesta a factores ambientales, sanitarios y alimenticios inadecuados, mostrará valores típicos que pueden llegar a diferir bastante de aquellos obtenidos entre los que poseen mejores condiciones de vida. Para estos casos, la normalidad se corresponde con lo

ordinario o habitual, en tal tipo de poblaciones. Entonces, una minoría de la población considerada como ordinaria está sana en realidad.

- Cuando se ajustan los datos obtenidos, según edad y sexo, o cualquier otro criterio adoptado, es necesario considerar el hecho de que se establezcan categorías por la acumulación de situaciones de vida desfavorables. Por ejemplo, si al envejecer un individuo urbano no sigue un régimen alimenticio favorable, ni hace la cantidad de ejercicio físico necesaria, y vive sometido a tensiones psicológicas mucho mayores que las de un individuo en un medio rural, entonces cabe esperar: obesidad y tensión arterial elevada. Ahora, el deterioro que se puede llegar a observar en tales individuos es por el deterioro progresivo de los órganos por envejecimiento, o más bien, es el resultado acumulado de condiciones desfavorables de vida.

3. Un valor es *normal* cuando forma parte de la proporción de observaciones aceptables para toda la comunidad desde un punto de vista individual, social y económico. Habida cuenta de las posibilidades terapéuticas o de otro tipo disponibles.

9.3 Valores de referencia o normales

En una industria farmacéutica dedicada a la fabricación de medicamentos, productos de belleza y derivados, los valores de referencia están referidos a los *límites de tolerancia* de los productos terminados. Cada producto tiene una fórmula con la composición de los productos activos y los complementos. En particular, el certificado de aptitud del medicamento emitido por la autoridad sanitaria es otorgado bajo ciertas condiciones que establecen las cantidades mínimas y máximas de los principales componentes. Dichas cantidades son los límites de tolerancia de fabricación, dentro de los cuales debe encuadrarse el producto para ser considerado “normal” y no ser desechado.

El producto final de un método de análisis clínicos usado en el laboratorio es un resultado relacionado con una muestra de un determinado paciente. El médico interpreta este resultado como indicativo de variación, ausencia de esta, o posible cambio en el estado de salud de su paciente. Para dar sentido a esta interpretación, el médico compara el resultado obtenido con algún *sistema de referencia* indicativo de la “normalidad”. Este ámbito de referencia de un cuadro clínico generalmente se construye con los valores de un intervalo en la magnitud clínica, que incluya al 95% de los individuos considerados sanos de la población. Por desgracia, muchas veces el médico interpreta estos extremos como indicativos de los límites superior e inferior dentro de los que se debe ubicar el paciente para ser considerado sano, según hacen mención Winkel y Statland en sus trabajos. Ellos explican que en la década del setenta el término valores normales se asimilaba al concepto de individuos sanos, lo cual condujo a confusiones en el contexto de la química clínica. Se ha propuesto usar los términos de individuos sanos en lugar de normales, y distribuciones de datos gaussianas en lugar de normales, para evitar este tipo de confusiones. Galen y Gambino (1975) analizaron el tema con detenimiento, y Grasbeck (1969) introdujo el concepto de *valores de referencia* para abolir el término de valores normales. Hay dos tipos de valores de referencia: los basados en una población y los basados en un individuo. Aquí se tratará el primer caso y se definirán estos valores de la manera siguiente:

Los valores de referencia de una magnitud clínica son un grupo de datos de la magnitud medida de un grupo de individuos (o de uno solo de ellos), en un definido estado de salud.

Para utilizar este concepto, el médico evalúa un resultado del laboratorio dentro de un cierto ámbito, para calcular la probabilidad de que el paciente del que se tomó la muestra analizada pertenezca al grupo de individuos sanos. En cambio, cuando el médico compara con los datos del mismo individuo cuando este se encontraba en un estado de buena salud, lo que hace es ver si todavía se mantiene dentro del mismo estado de salud.

No se deben usar los límites de referencia como rígidas fronteras dentro de las cuales se encuentran los pacientes sanos, y los anormales en su exterior.

Por ejemplo, un paciente con un valor sérico de colesterol situado por debajo del límite inferior podría ser considerado como un signo de buena salud, en lugar de lo contrario. Otro caso, no es un buen síntoma para un diabético el tener sus valores de glucosa dentro del recinto de "normalidad". La idea central, es nunca usar el sistema de referencia en forma aislada sin tener en cuenta todo el cuadro clínico, para no llegar a conclusiones erróneas.

Para poder construir los intervalos de referencia de una magnitud clínica y poder usarlos convenientemente, siempre se deben explicitar los siguientes cinco puntos:

- 1) *La población de referencia y la forma en que se eligió la muestra estudiada.*
- 2) *Las condiciones ambientales y fisiológicas bajo las cuales se extrajeron las muestras.*
- 3) *La técnica de extracción, el momento de recolección, el transporte, la preparación y la conservación de las muestras tomadas.*
- 4) *El método analítico usado, con datos acerca de su exactitud, precisión y control de calidad*
- 5) *La cantidad de datos observados y los intervalos de referencia derivados.*

Como ejemplo de aplicación se usa un trabajo de Galen (1983) donde el autor evaluó los datos históricos de pacientes externos ambulatorios que buscaban cuidado médico primario de tipo privado. Estos pacientes recibieron un perfil bioquímico de 26 pruebas como parte de una evaluación sanitaria de rutina o una investigación diagnóstica en este marco. Para cada constituyente se identificaron 1000 pacientes para cada grupo de edades y para cada sexo. No hubo subselección por otros factores como ser: origen étnico, ocupación, peso corporal, ubicación geográfica, etc. Los datos se tomaron simplemente por su disponibilidad en cuanto a edad y sexo dentro de los pacientes ambulatorios. No se tuvieron en cuenta adicionales diagnósticas, ambientales y fisiológicas, ni se dispuso de ellas. Las muestras venosas se tomaron después de aplicar torniquete, el suero se separó a los 30 a 45 minutos después de la recolección. La muestras de suero se guardaron a temperatura ambiente y se analizaron entre las horas 12.00 y 18.00. después de su extracción. La documentación estadística de precisión y exactitud de cada procedimiento realizado se acumula diariamente. Se agruparon los datos por sexo, y dentro de cada uno por grupo de edades, comenzando de 10 a 14 años y así sucesivamente con el mismo ancho hasta los de 99 años. Se muestra el cuadro de valores (Cuadro 9.2) de calcio en mg/dl, medido con el método de auto químico (equipo automático) de la reacción de complejación de sulfonato de alizarina. Te-

niendo en cuenta la gran cantidad de datos (1000 en cada celda), los autores supusieron una distribución gaussiana porque razonaron "...Como la definición de una población sana o normal, como la población de referencia para estos estudios es larga, costosa y quizás imposible, el análisis de *grandes muestras* evita estos requerimientos...". La idea básica es desarrollar en cada celda un histograma con los 1000 datos disponibles, y obtener los percentiles 0,5; 2,5; 50; 97,5 y 99,5 como los límites de referencia para cada celda.

Cuadro 9.2 Valores de referencia del calcio sérico en mg/dl.

EDAD	HOMBRES				MUJERES			
	Percentiles				Percentiles			
años	0,5	2,5	97,5	99,5	0,5	2,5	97,5	99,5
10 a 14	8,1	9,0	10,7	11,3	7,9	8,9	10,7	11,8
15 a 19	8,2	8,8	10,7	11,4	8,5	8,8	10,7	10,9
20 a 24	8,6	8,9	10,8	11,4	8,2	8,7	10,5	10,8
25 a 29	8,8	8,9	10,6	10,8	8,4	8,8	10,4	10,9
30 a 34	8,7	8,9	10,8	11,3	8,3	8,7	10,4	10,7
35 a 39	8,5	8,9	10,6	11,0	8,5	8,8	10,4	10,8
40 a 44	8,8	8,9	10,6	10,8	8,0	8,5	10,5	10,7
45 a 49	8,7	8,8	10,6	10,9	8,4	8,8	10,6	10,8
50 a 54	8,5	8,8	10,6	11,1	8,1	8,8	10,6	10,9
55 a 59	8,4	8,7	10,5	11,0	8,1	8,8	10,6	10,9
60 a 64	8,2	8,7	10,6	10,9	8,4	8,8	10,7	11,3
65 a 69	8,2	8,6	10,6	10,8	8,5	8,8	10,7	11,3
70 a 74	8,3	8,8	10,7	11,3	8,3	8,8	10,6	11,3
75 a 79	8,0	8,5	10,5	11,0	8,2	8,7	10,6	11,0
80 a 84	7,6	8,5	10,4	10,7	8,3	8,7	10,7	10,9
85 a 89	8,1	8,5	10,6	11,7	7,6	8,6	10,5	11,1

FUENTE: Sonnenwirth AC y Jarret L, *Métodos y diagnósticos del laboratorio clínico*. Pág. 38.

A veces se confunde el término "valor referente", que es un sinónimo de punto de corte, valor crítico o valor de discriminación, con los valores límites del intervalo de referencia. Un punto de corte no puede determinarse en la misma forma que se han determinado los valores de referencia, estudiando individuos sanos o "normales". Se debe determinar estudiando el constituyente en personas sanas y con distintos estados de avance de la enfermedad. De esta forma, se pueden determinar valores referentes o puntos de corte que permitan un máximo de discriminación entre sanos y enfermos; o entre una enfermedad en particular y otras incluidas en el diagnóstico diferencial.

Por ejemplo, si un médico está tratando a una paciente de 52 años, los valores de referencia que debería utilizar están entre 8,8 y 10,6 mg/dl, con una probabilidad asociada del 95% o entre 8,1 y 10,9 mg/dl con una probabilidad del 99%. Para resumir, si se quiere prescindir de la edad o del sexo, se buscan los valores extremos en la tabla para generalizar, al estilo usual de la vieja bibliografía. Se puede ver que entre los valores 7,6 y 11,8 mg/dl se encuentran todos los datos tabulados, luego se podría decir que el Calcio tiene ese rango de valores de referencia en la población humana de donde se extrajo la muestra.

Debido a la forma clásica de definir la "normalidad" en los laboratorios de análisis clínicos, muchos resultados anormales no son identificados como tales, no anuncian estados de en-

fermedad ni de riesgo para la salud. La respuesta no es tan solo una recolección de datos elaborados con intervalos de referencia. Para un buen diagnóstico de laboratorio se deben considerar varios factores no patológicos, responsables de dar señales de falsos positivos, además de valores referenciales para pruebas individuales de laboratorio que son buenos anunciadores de la enfermedad. Cuando el perfil bioquímico del paciente se evalúe así se habrá avanzado un buen paso en patología clínica.

Para asegurarse de tener una distribución normal en un experimento cualquiera de un laboratorio, se pueden medir alícuotas de una misma muestra en condiciones estables teniendo en cuenta que: *Mediciones repetidas de la misma magnitud tienen una distribución gaussiana.*

9.4 La función de Gauss o “la normal”

Las distribuciones de probabilidad analizadas hasta ahora eran del tipo de variable discreta. Pero la gran mayoría de las magnitudes clínicas son del tipo continuo, cuya distribución teórica de la variable se llama *función de densidad* como se vió en el capítulo anterior. Recordando el Teorema Central del Límite cuando las muestras son grandes, las funciones de probabilidad discretas tienden a una continua: la función de Gauss. Esta función tiene forma de campana y se expresa con:

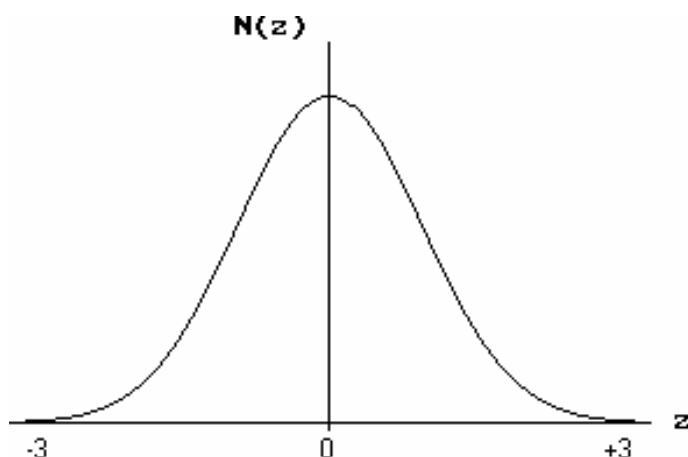
$$N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2}$$

donde $N(x)$ indica la altura de la ordenada de la curva acampanada, que representa la función densidad de Gauss. Como se puede apreciar en la fórmula, existen infinitas funciones normales posibles. La manera de identificar en forma unívoca a una de ellas es dando los valores a sus dos parámetros μ y σ . Luego, para destacar este hecho se puede escribir $N_x(\mu, \sigma)$, para denotar a una función normal de la magnitud clínica x , con sus dos parámetros identificatorios.

De toda la familia gaussiana hay una curva en particular que se toma como referencia o patrón de las demás. Es aquella que se obtiene tipificando la variable x , o sea: $N_z(0, 1)$. Esta curva se muestra en el Gráfico 9.1 y sus valores están tabulados en el Tabla 1. Se la denomina curva normal *estandarizada o tipificada*. Se la emplea para resolver cualquier problema de obtención de la probabilidad normal pues no importa la curva normal que se tenga, con el simple recurso de tipificarle la variable se la transforma en la de tablas, y con eso se simplifica la tarea de calcular la integral correspondiente. Se aprecia que la curva es simétrica respecto del eje que pasa por su centro (el origen), y tiene dos puntos de inflexión también simétricos situados a un desvío estándar del eje. El área total bajo la curva es uno, como se puede calcular con:

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2} = 1$$

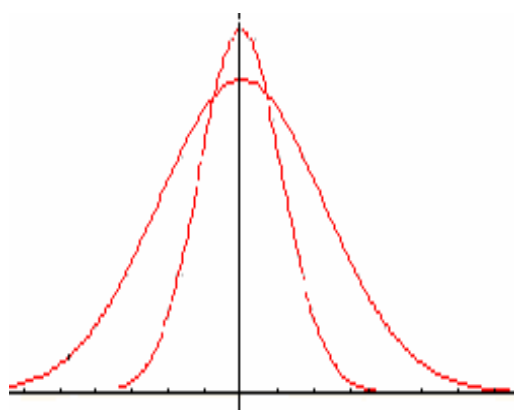
Gráfico 9.1 : Campana de Gauss tipificada.



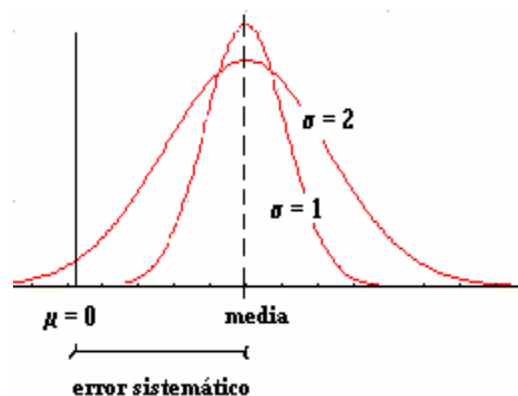
En el Gráfico 9.2 se presentan funciones normales con diferentes valores en sus parámetros. Se aprecia que a medida que se disminuye el valor σ , dejando fijo el valor de μ , la curva se angosta y se hace más alta (leptocúrtica), mientras que si aumenta el valor de su desvío estándar se achata, tiende a pasar de su forma acampanada a otra forma como de plato invertido (platocúrtica). Eso significa que si la dispersión de los valores disminuye, el error casual es menor, la precisión del sistema de medición es mayor, y la curva se afina. Viceversa, cuanto peor es la precisión de la técnica clínica empleada, más aplanada resultará la curva. En la parte superior del gráfico se muestran dos curvas con valor esperado nulo y desvío estándar 1 y 1,5. En la parte derecha del gráfico, se ilustra el caso de los errores sistemáticos. El efecto de la aparición de un error de esta naturaleza en el sistema de medición, hace que la curva gaussiana se corra a derecha o izquierda, *sin cambiar de forma*. En efecto, si aparece un valor positivo μ la curva de referencia $\sigma = 1$ se desplaza a la derecha. En cambio, si la precisión disminuye porque la dispersión se duplicó $\sigma = 2$ la curva se aplanan (cuando la dispersión se achica a la mitad $\sigma = 0,5$ la curva se afina). O sea, si cambia la dispersión de los valores, *cambia la forma de la curva*. Los errores sistemáticos la desplazan sin cambiarla, mientras que los errores casuales la transforman sin desplazarla. El efecto conjunto es el ilustrado más abajo.

Gráfico 9.2 : Efecto de los errores sistemáticos y casuales en la curva de errores de Gauss.

a) Efecto de los casuales



b) Efecto de los casuales y sistemáticos



9.5 Propiedades

Las propiedades siguientes se enuncian sin demostración, las que pueden hallarse en los libros de teoría como el de Cramer.

- 1) La curva normal es simétrica respecto de un eje vertical que pase por su centro.
- 2) El área total bajo la curva es igual a uno, por lo tanto a cada lado del eje de simetría es igual a 0,5 la mitad del área total. El área bajo la curva es igual a la probabilidad.
- 3) Cada curva normal se identifica en forma unívoca con los valores de sus 2 parámetros (μ y σ).
- 4) La función normal tipificada tiene parámetros $\mu = 0$ y $\sigma = 1$, es simétrica respecto al origen y la probabilidad a cada lado del eje de ordenadas vale 0,5 (ver Tabla en el Anexo).
- 5) Las funciones de Gauss tienen dos puntos de inflexión simétricos en $x = \mu \pm \sigma$
- 6) En toda función normal la media coincide con la mediana y con la moda.
- 7) Si x es una variable aleatoria de parámetros $N_x(\mu, \sigma)$ entonces la variable $y = ax + b$, donde a y b son constantes, también tendrá una distribución normal $N_y(a\mu + b, a\sigma)$.
- 8) Si x e y son dos variables aleatorias normales de parámetros $N_x(\mu_x, \sigma_x)$; $N_y(\mu_y, \sigma_y)$, entonces la variable suma de ambas $z = x + y$, también será normal de parámetros dados por la suma de sus valores esperados y de sus varianzas $N_z\{(\mu_x + \mu_y), (\sigma_x^2 + \sigma_y^2)^{1/2}\}$.
- 9) Si x_1 y x_2 son dos variables aleatorias normales independientes, tienen los mismos parámetros (μ, σ) , entonces la variable aleatoria $w = (x_1 + x_2)/2$, promedio de ambas, tendrá una distribución normal $N_w(\mu, \sigma/\sqrt{2})$.
- 10) Generalizando, sean n variables aleatorias normales independientes, cada una de ellas proveniente de la misma población, es decir con los mismos parámetros (μ, σ) , entonces la variable aleatoria promedio de las mismas también será una función normal con los parámetros:
Si $y = \bar{x} = \sum_{i=1}^{i=n} \frac{x_i}{n}$ entonces y tendrá una distribución normal: $N_y(\mu, \sigma/\sqrt{n})$
- 11) El área bajo la curva normal, correspondiente a los intervalos siguientes, es igual a la probabilidad de que un valor de la variable aleatoria caiga dentro de esos intervalos:
 $P\{x \in (\mu \pm \sigma)\} = 0,6828 \Rightarrow$ la probabilidad que el valor caiga en $(\mu \pm \sigma)$ es del 68,3%
 $P\{x \in (\mu \pm 2\sigma)\} = 0,9546 \Rightarrow$ la probabilidad que el valor caiga en $(\mu \pm 2\sigma)$ es del 95,5%
 $P\{x \in (\mu \pm 3\sigma)\} = 0,9973 \Rightarrow$ la probabilidad que el valor caiga en $(\mu \pm 3\sigma)$ es del 99,7%

9.6 Cálculo de probabilidades con Gauss

A continuación, se muestran algunos ejemplos de aplicación del cálculo de probabilidad usando la función de Gauss con la Tablas 1 y 2 del Anexo de Tablas Estadísticas que corresponden a la Curva de Gauss Estandarizada (con valor esperado nulo y desvío estándar igual a uno).

Ejemplo 1) En una población gaussiana de parámetros (114 ; 23) se extrae un individuo al azar y se desea calcular las probabilidades de que: a) sea mayor de 100; b) esté comprendido entre 56 y 74.

Para resolver estos casos se deben calcular primero las tipificaciones de los valores dados. Luego hacer un pequeño esquema para entender el problema y por último usar la tabla de valores tipificados para calcular las probabilidades. Tal como se procedió más arriba. Pero ahora, se detallará un poco más:

Los valores tipificados son:

$$Z_1 = (100 - 114)/23 = -0,6087$$
$$Z_2 = (56 - 114)/23 = -2,5217$$
$$Z_3 = (74 - 114)/23 = -1,7391$$

a) Para encontrar las probabilidades respectivas hay dos caminos: el primero es ir a la Tabla de Gauss tipificada e interpolar para calcular el valor pedido; el otro es obtener el valor directamente con una computadora y un programa estadístico como el Excel, el Statistica, y otros.

De tablas para $z = 0,60$ es $p' = 0,2257$ y para $z = 0,61$ es $p'' = 0,2291$, el valor buscado está entre ambos. Para interpolar se razona que, si para una diferencia de $0,01 = 0,61 - 0,60$ corresponde una diferencia de probabilidades $P = 0,2291 - 0,2257 = 0,0034$, para la diferencia buscada $(0,61 - 0,6087) = 0,0013$ corresponderá una probabilidad $0,000442 = (0,0034)(0,0013)/0,01$.

Entonces será:

$$P(Z_1) = P(0,6087) = P(0,61) - 0,000442 = 0,2291 - 0,000442 = 0,228658$$

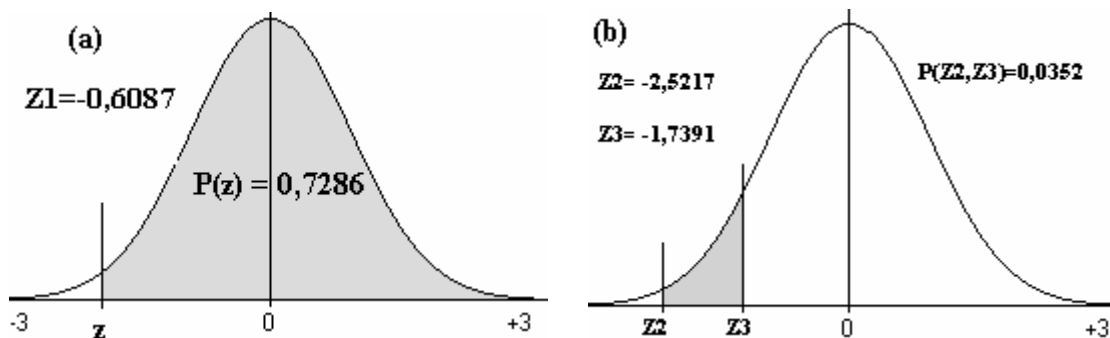
Análogamente, si se hubiese calculado la otra diferencia $0,0087$ el valor correspondiente, de probabilidad es $0,002958$, que debería sumarse al valor p' y el resultado sería el mismo.

Con el programa de Microsoft, Excel 2000, se puede calcular directamente el valor de probabilidad usando las funciones estadísticas que tiene incorporadas. En este caso: Distr. Norm. Estand. Que da un valor $P(Z_1) = P(0,6087) = 0,2286384$ el cual coincide en las cinco primeras cifras con la interpolación de tablas.

Ahora bien, $P(Z_1) = P(-Z_1) = 0,2286$ por la simetría de la función de Gauss. Entonces, el valor pedido en el problema resultará:

$$P(x > 100) = P(Z > -Z_1) = P(-0,6087 < Z < 0) + P(Z > 0) = 0,2286 + 0,5 = 0,7286$$

$$b) P(56 < x < 74) = P(Z_2 < Z < Z_3) = P(-2,5217 < Z < -1,7391) = 0,49416 - 0,45899$$
$$P(56 < x < 74) = 0,0352$$



Ejemplo 2) Se determinó la Lipemia con un método colorimétrico de un grupo de 200 individuos que padecen hipotiroidismo, obteniéndose un promedio de 13 g/l, con un desvío estándar de 0,1 g/l. Se pide calcular la probabilidad de encontrar un paciente elegido al azar cuyo valor:
 a) esté entre 13 y 13,2 g/l; b) sea mayor de 13,25 g/l; c) esté comprendido entre 13,1 y 13,2 g/l

Método 1) Se supone que los valores poblacionales son $\mu = 13$ g/l y $\sigma = 0,1$ g/l

a) Esté entre 13 y 13,2 g/l

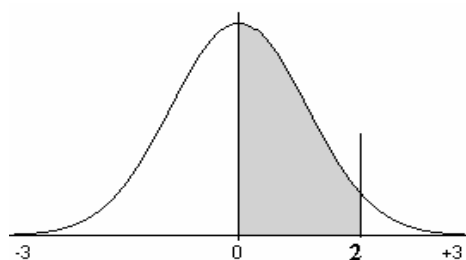
Tipificando la variable resulta $z = \frac{13,2 - 13,0}{0,1} = 2$

Y $z = 0$ para el valor 13, entonces los valores pedidos son:

$$P(13,0 < x < 13,2) = P(0 < z < 2)$$

Buscando en la Tabla 1 se halla: $P(z = 2) = 0,4772$

La probabilidad es casi del 48%.



b) Sea mayor de 13,25 g/l

Tipificando la variable resulta $z = \frac{13,25 - 13,0}{0,1} = 2,5$

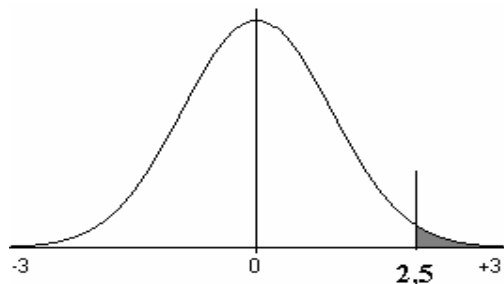
Lo pedido es $P(z > 2,5) = P(x > 13,25)$

Buscando en la Tabla 1 se halla: $P(z = 2,5) = 0,4938$

La mitad del área bajo la curva es 0,5; lo buscado es:

$$P(x > 13,25) = P(z > 2,5) = 0,5 - 0,4938 = 0,0062$$

La probabilidad es del 0,62%.



c) Esté comprendido entre 13,1 y 13,2 g/l

Tipificando resulta $z_1 = \frac{13,1 - 13,0}{0,1} = 1$

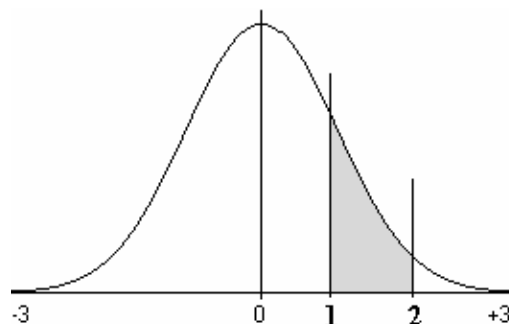
y $z_2 = \frac{13,2 - 13,0}{0,1} = 2$

$$P(13,1 < x < 13,2) = P(1 < z < 2) = P(x < 2) - P(x < 1)$$

O sea,

$$P(1 < z < 2) = 0,4772 - 0,3413 = 0,1359$$

La probabilidad es de 13,6%



Método 2) No se conocen los parámetros poblacionales y se los estima con los valores muestrales. O sea, se toman $\mu \approx \bar{x} = 13,0 \text{ g/l}$ y $\sigma \approx DS = 0,1 \text{ g/l}$

Averiguar: ¿cuántos pacientes tendrán un valor comprendido entre 13,1 y 13,2 g/l?

Tomando una aproximación de 0,1 g/l en el sistema de medición, los valores medidos entre 13,1 y 13,2 pueden realmente tener cualquier valor comprendido entre los límites reales, o sea, entre 13,05 y 13,25. Por lo tanto será

$$z_1 = \frac{13,05 - 13,0}{0,1} = 0,5$$

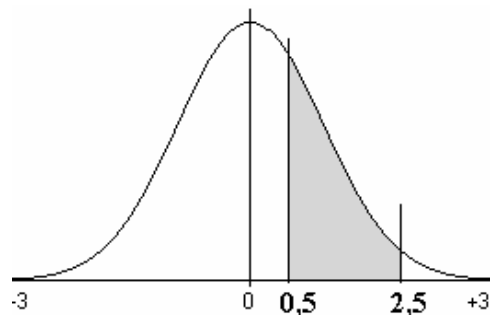
y por el cálculo anterior $z_2 = 2,5$

Luego la probabilidad es:

$$P(13,05 < x < 13,25) = P(0,5 < z < 2,5) = 0,4938 - 0,1915 = 0,3023$$

Y el valor esperado será

$$E = N \cdot P(13,05 < x < 13,25) = 200 \cdot 0,3023 \approx 60 \text{ pacientes}$$



Nota: Cuando se usan los valores de la muestra para estimar los valores desconocidos de la población, entonces se debe incrementar (o disminuir) en la mitad de la menor unidad de la escala de medición, a los valores buscados.

Ejemplo 3) Los límites de tolerancia del componente Sulfato de Neomicina en un antibiótico fueron fijados en (40 ; 60) mg. Se sabe que el valor promedio histórico de proceso de fabricación es 50 mg con un desvío de 4 mg. Calcular el número esperado de defectuosos en un lote de 9000.

Los valores de tolerancia tipificados son:

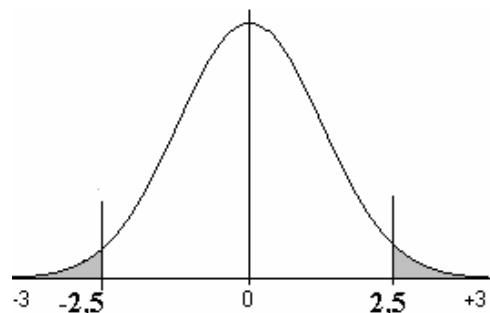
$$Z_1 = (40 - 50)/4 = -2,5 \quad \text{y} \quad Z_2 = (60 - 50)/4 = 2,5$$

La zona de rechazo es por fuera de estos límites, entonces la probabilidad de rechazo es:

$$P(\text{rechazo}) = 1 - P(-2,5 < z < 2,5) = 1 - 2 P(0 < z < 2,5) = 1 - 2(0,4938) = 1 - 0,9876 = 0,0124$$

El número esperado de defectuosos será:

$$D = N \cdot P(\text{rechazo}) = 9000 (0,0124) = 111,6$$



O sea, se esperan rechazar 112 comprimidos de los 9000.

9.7 Aproximaciones con la función de Gauss

Se puede usar la función de Gauss para ajustarla a una serie de datos agrupados en una tabla de frecuencia de una magnitud clínica de tipo asumida como gaussiana. El supuesto básico es que la muestra fue extraída de una población normal. Para ilustrar el proceso se tomarán los datos de peso de 195 estudiantes varones, que se muestran en el Cuadro 9.4. En la primera columna figuran los límites reales de clase, y en la segunda las frecuencias respectivas. Con esos datos se calcula el promedio de 69,47 kg con un desvío típico de 2,82 kg. Para determinar la frecuencia esperada teóricamente, asumiendo una distribución normal de la variable, se debe determinar

primero la probabilidad de cada clase y luego multiplicarla por el número total de casos (N=195). Debe tenerse cuidado con el primero y el último de los intervalos, pues si no se tiene en cuenta las colas de la distribución que tienden asintóticamente a infinito, la suma de la probabilidad total no será la unidad. Para esquivar ese problema se toma toda el área a la izquierda del límite real superior de la primera clase ($-\infty, 63,5$), para el primer caso. Luego, para el segundo intervalo se calcula la probabilidad entre ($-\infty, 65,5$) y se le resta el valor anterior, con lo que se determina la probabilidad del segundo intervalo, y así sucesivamente hasta el último, donde se debe incluir el área a la derecha, obteniéndola por diferencia con la unidad.

Cuadro 9.4: Ajuste a una distribución de pesos de 195 varones.

Intervalos	Frecuencia Observada	Valor de z para LRS	Área acumulada en ($-\infty, LRS$)	Probabilidad en cada clase	Frecuencia Esperada
hasta 63,5	3	-2,12	0,0170	0,0170	3,3
63,5 - 65,5	14	-1,41	0,0793	0,0623	12,2
65,5 - 67,5	30	-0,70	0,2420	0,1627	31,7
67,5 - 69,5	48	+0,01	0,5040	0,2620	51,1
69,5 - 71,5	48	+0,072	0,7642	0,2602	50,7
71,5 - 73,5	39	+1,43	0,9236	0,1594	31,1
73,5 - 75,5	11	+2,14	0,9838	0,0602	11,7
75,5 y más	2		1,0000	0,0162	3,2

La aproximación de una distribución de frecuencias reales a una teórica puede hacerse, siempre y cuando se asuma que las muestras provienen de una población normal.

Los procesos del tipo Bernoulli, representados mayormente por la probabilidad binomial, también pueden aproximarse con una función normal de parámetros: $\mu = np$ y $\sigma = \sqrt{npq}$. Siempre que las muestras sean lo suficientemente grandes, basta que $np > 5$ y $nq > 5$ para considerar buena a la aproximación. Lo curioso es que una variable de tipo discreta se puede aproximar con una de tipo continua. Por ello, se debe efectuar una corrección por *continuidad* que consiste en sumarle o restarle 0,5 a los valores enteros en los intervalos buscados. Esto es, los valores 0 y n son los máximos y mínimos de una binomial al aproximar las correspondientes probabilidades se tomará: toda el área a la izquierda de 0,5 para el valor $x = 0$, y toda el área a la derecha de $(n-0,5)$ para aproximar el valor de $x = n$. Muy similar a lo visto en el punto anterior como *Método 2*. En el Cuadro 9.5 se toma un caso extremo para ver cómo resulta la peor aproximación en el límite: 10 lanzamientos de una moneda ($np = nq = 5$).

Cuadro 9.5: Aproximación entre la normal y la binomial.

<i>Caso 1: Se lanza una moneda 10 veces, obtener las probabilidades exactas del número de caras con la binomial: $P_{bi} = P(x < 11 / n = 5; p = 1/2)$, para comparar con las aproximadas $N (\mu = 5 ; \sigma = 1,58)$</i>											
x	0	1	2	3	4	5	6	7	8	9	10
Binomial:	0,001	0,010	0,044	0,117	0,205	0,246	0,205	0,117	0,044	0,010	0,001
Normal :	0,002	0,011	0,044	0,114	0,203	0,251	0,203	0,114	0,044	0,011	0,002

Caso 2: Se deben adivinar los símbolos que se sacan al hacer una extracción al azar en las cartas Zener: como hay 5 símbolos, $p = 1/5 = 0,2$. Se efectúan 25 extracciones, calcular las probabilidades de tener hasta 9 éxitos, en forma exacta y aproximando con Gauss. $N (\mu = 5 ; \sigma = 2)$

x	0	1	2	3	4	5	6	7	8	9
Binomial:	0,004	0,024	0,071	0,136	0,187	0,196	0,163	0,111	0,062	0,029
Normal :	0,009	0,027	0,065	0,121	0,176	0,199	0,176	0,121	0,065	0,027

Cuando sea $np > 5$ y $nq > 5$, se puede aproximar la binomial con una normal.

En un proceso del tipo Poisson, como se puede aproximarla con la binomial es de esperar que se pueda usar la normal para aproximar una Poisson. En efecto, si el valor esperado es mucho mayor que 1, la aproximación es aceptable. En la binomial sería $\mu = np$, entonces si es $n \gg 1$ y $p \ll 1$, su producto resultará $np > 1$, así se tendrá una Poisson con $\sigma = \mu = np$.

Cuadro 9.6: Aproximaciones entre la Binomial, Poisson y la de Gauss.

Caso 1: La probabilidad de éxito en un proceso del tipo Bernoulli es muy pequeña ($p = 0,01$). Si se repite 100 veces el experimento, calcular la probabilidad de obtener 0,1,2,3, 4 y 5 éxitos. Donde las probabilidades se sacan con : $P_{PO} (x / \sigma^2 = \mu = 1)$, $P_{bi} = P(x / n = 100; p = 0,01)$, $N (\mu = 1 ; \sigma = 1)$

x	0	1	2	3	4	5
Binomial:	0,366	0,370	0,185	0,061	0,015	0,003
Poisson:	0,368	0,368	0,184	0,061	0,015	0,003
Normal :	0,309	0,383	0,242	0,061	0,005	0,000

La aproximación entre la Poisson y la normal podría ser mejor si np hubiese sido mucho mayor que 1.

Caso 2: La probabilidad de que un individuo sufra una reacción alérgica por una inyección es de 0,01. Calcular la probabilidad de que entre 8 y 14 pacientes tengan una reacción, sobre un total de 1000.

Para resolver, se usa una Poisson de parámetro $\mu = np = 10$

x	8	9	10	11	12	13	14	Total
Poisson:	0,1126	0,1251	0,1251	0,1137	0,0948	0,0729	0,0521	0,6963

Luego la probabilidad de acuerdo con Poisson es de 69,63%

Usando la función normal será $z_1 = (7,5 - 10) / \sqrt{10} = - 0,79$ y $z_2 = (14,5 - 10) / \sqrt{10} = 1,42$

$P(8 < x < 14) = 0,2852 + 0,4222 = 0,7072$. O sea un 70,7% (la diferencia es del 1,5%).

La función de Gauss descubierta alrededor de 1810, no solo es la más tradicional en Estadística sino la de uso más difundido por las posibilidades de simplificar los cálculos a través de las aproximaciones. Como se verá más adelante, casi todos los modelos estadísticos tienen una probabilidad teórica propia que se aproxima a la función de Gauss en condiciones bastante generales y cuando las muestras son lo *suficientemente grandes*. Esto significa para el estudiante, que esta función la usará en forma reiterada y por lo tanto debe asegurarse de entender muy bien su empleo, propiedades y aplicaciones.

9.8. Correcciones por continuidad

Como se verá más adelante la función normal se usa como aproximación asintótica en muchos modelos de estadística, tanto la paramétrica como la no-paramétrica. Y como se vio más arriba se usa como aproximación a las distribuciones binomiales y de Poisson. En resumen, cuando el valor del tamaño muestral n crece indefinidamente, la forma de la función binomial se parece cada vez más a la forma de la normal. Y esto ocurre para cualquier valor de la probabilidad p . Cuando los valores de la probabilidad son más cercanos a $p = 0,5$ tiende más rápido, que cuando los valores de p están cercanos a 0 o 1, pues todas las distribuciones binomiales con $p = 0,5$ tienen la gran ventaja de ser simétricas. Así, si n es lo suficientemente grande se puede imaginar la binomial como muy parecida a la normal.

Cuando la variable que se está usando es una frecuencia r (0, 1, 2,..., n) la magnitud se considera discreta, pero las tablas de la función Normal son para magnitudes de tipo continuo, entonces existe una cierta corrección que se le puede hacer, para lograr un mejor ajuste entre ambas distribuciones. Esta se llama: *corrección por continuidad*. Se basa en la idea que el valor de la probabilidad binomial para un valor r en realidad es aproximadamente la probabilidad del intervalo entre $(r - 1/2)$ y $(r + 1/2)$ porque la integral de la función normal para un punto es nula, y siempre se debe tomar la integral de algún intervalo para tener un resultado diferente de cero. Por consiguiente, de acuerdo a la hipótesis nula efectuada la corrección por continuidad se realiza así:

1) *Corrección de Yates*: Al valor tipificado $z = (r - \mu) / \sigma$ se lo corrige con

$$Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma \quad \text{donde } \sigma = \sqrt{npq}$$

Caso 1: Si se estima que la probabilidad de nacer varón es $p = 0,5$. Se pide calcular la probabilidad de que en 40 nacimientos en una maternidad nazcan por lo menos 14 varones:

$H_0 : r \leq n \cdot p$ donde $\mu = n \cdot p = 20$ y $\sigma = \sqrt{npq} = 3,162$ entonces

$$z = (r - \mu) / \sigma = (14 - 20) / 3,162 = - 1,8975 \quad \text{y entonces la probabilidad es } 0,02888$$

Pero si se hace la corrección por continuidad el resultado sería:

$$Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma = [6 - 0,5] / 3,162 = 1,739 \quad \text{y entonces la probabilidad es } 0,0410$$

Comparando con la probabilidad exacta calculada por la fórmula binomial $P_{\text{Bi}} = 0,40345234$. De donde se puede ver que sin la corrección de Yates la aproximación no es muy aceptable. Si en cambio se pide calcular la probabilidad de que nazcan 26 o más varones es:

$H_0 : r \leq n \cdot p$ entonces y resulta igual a la anterior pues $P_{\text{Bi}}(r \leq 14) = P_{\text{Bi}}(r \geq 26)$

$$z = (r - \mu) / \sigma = (26 - 20) / 3,162 = 1,8975 \quad \text{y así la probabilidad es la misma } 0,0289.$$

Aplicando nuevamente la corrección resulta $Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma = 1,739$ y entonces la probabilidad es 0,0410 muy parecida a la exacta $P_{\text{Bi}} = 0,0403$.

2) *Correcciones de Molenaar*: Al valor tipificado $z = (r - \mu) / \sigma$ se lo corrige con

$$Z_{\text{corr.}} = [(4r + 3) \cdot (1-p)]^{1/2} - [(4n - 4r - 1) p]^{1/2}$$

Para el caso anterior donde $r = 14$ es $Z_{\text{corr.}} = [(56 + 3) \cdot (0,5)]^{1/2} - [(160 - 56 - 1) p]^{1/2} = -1,74496$

Este valor da una probabilidad de 0,405 que es mucho más cercana al valor exacto que la efectuada con la corrección de Yates. Mientras que si el valor de p es muy cercano a cero o uno, se debería usar una fórmula corregida diferente dada por:

$$Z_c = [0,5 \cdot (8r + 5) \cdot (1-p)]^{1/2} - [0,5 \cdot (8n - 8r - 3) p]^{1/2}$$

Por ejemplo, sea una probabilidad de éxito de 0,05 en un proceso binomial de 50 pruebas y hay que calcular la probabilidad de tener por lo menos 2 éxitos. Entonces, la probabilidad exacta da un resultado de $P_{Bi} = 0,54053$, mientras que la aproximación normal da:

Yates: $Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma = [|2 - 2,5| - 1/2] / 2,121 = 0$; lo que da $P \approx 0,5$

Molenaar: $Z_c = [0,5 \cdot (8r + 5) \cdot (1-p)]^{1/2} - [0,5 \cdot (8n - 8r - 3) p]^{1/2}$

$Z_c = [0,5 \cdot (16 + 5) \cdot (1-0,05)]^{1/2} - [0,5 \cdot (400 - 16 - 3) 0,05]^{1/2} = 0,06397$; lo que da $P \approx 0,5255$

Y con la otra es $Z_{\text{corr.}} = [(4r + 3) \cdot (1-p)]^{1/2} - [(4n - 4r - 1) p]^{1/2} = 0,1423$ O sea $P \approx 0,5566$

Este ejemplo muestra que la corrección de Molenaar es mejor que la de Yates en ambos casos. Y que la segunda ajusta mejor, cuando las probabilidades son muy pequeñas o muy grandes.

Lo mismo puede hacerse con la aproximación normal a la probabilidad de Poisson. Suponiendo que la variable X sea de tipo Poisson, el valor tipificado para usar la aproximación normal es $z = (X - \mu) / \sigma = (X - \mu) / \mu^{1/2}$. Y su valor corregido será:

1) *Corrección de Yates*: $z = [|X - \mu| - 1/2] / \mu^{1/2}$

Por ejemplo, si el valor medio de un proceso de Poisson es $\mu = 5$ y se desea calcular la probabilidad de a) no obtener ningún éxito, b) obtener a lo sumo 2 y c) por lo menos 8 éxitos. Entonces, las probabilidades exactas son: $P_{PO}(x = 0) = 0,0067$; $P_{PO}(x \leq 2) = 0,1246$; $P_{PO}(x \geq 8) = 0,1334$ y efectuando las aproximaciones con la función normal, los valores resultan ser:

a) $z = [|0 - 5| - 1/2] / 5^{1/2} = 2,013$ lo que equivale a una probabilidad de 0,0221

b) $z = [|2 - 5| - 1/2] / 5^{1/2} = 1,118$ y corresponde a una probabilidad de 0,1318

c) $z = [|8 - 5| - 1/2] / 5^{1/2} = 1,118$ igual que la anterior

2) *Correcciones de Molenaar*: $z = 2 [X + 0,75]^{1/2} - 2 [\mu]^{1/2}$

a) $z = 2 [0 + 0,75]^{1/2} - 2 [0]^{1/2} = - 2,740$ o sea $P \approx 0,0031$

b) $z = 2 [2 + 0,75]^{1/2} - 2 [2]^{1/2} = - 1,156$ o sea $P \approx 0,1239$

c) La $P_{PO}(x \geq 8) = 1 - P_{PO}(x \leq 7)$ Con la aproximación es $z = 1,0956$ o sea $P(7) \approx 0,8633$

Luego la $P(x \geq 8) \approx (1 - 0,8633) = 0,1367$. Nuevamente esta corrección es la mejor de ambas.

9.9 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | | |
|---|-------|----------|----------|
| 1) Explicar lo que se entiende por "normal" | | | |
| 2) Las estadísticas gaussianas sirven para determinar los límites normales. | | V | F |
| 3) Se estila adoptar los límites de referencia como "límites normales". | | V | F |
| 4) En análisis clínicos conviene más usar los modelos no paramétricos. | | V | F |
| 5) Los criterios de normalidad son tres: | | | |
| 6) Los valores de referencia de una magnitud clínica dependen de la población referencial. | | V | F |
| 7) Para construir los intervalos de referencia se deben explicar los 5 puntos siguientes: | | | |
| 8) La función de Gauss es simétrica y el área total bajo la curva es uno. | | V | F |
| 9) La curva de Gauss tiene 2 puntos de inflexión. | | V | F |
| 10) El efecto de los errores casuales es desplazar la curva de Gauss sin cambiar su forma. | | V | F |
| 11) El efecto de los errores sistemáticos es modificar el ancho de la curva de Gauss. | | V | F |
| 12) La curva de Gauss queda unívocamente identificada con 2 parámetros que son: | | | |
| 13) En la curva Normal la media coincide con la mediana y con la moda. | | V | F |
| 14) La suma de 2 funciones normales es otra función normal. | | V | F |
| 15) La probabilidad de que un valor caiga en un intervalo dado por ± 2 desvíos es: | | | |
| 16) Si las muestras provienen de una población normal se puede aproximar con la empírica. | | V | F |
| 17) Si n es lo suficientemente grande se puede aproximar la binomial con la normal. | | V | F |
| 18) Si $np > 5$ y $nq > 5$ se puede hacer la aproximación normal a la binomial. | | V | F |
| 19) Si $n \gg 1$ y $p \ll 1$ tal que $np = cte$. Se puede aproximar la normal a la Poisson. | | V | F |
| 20) La suma de n variables aleatorias independientes y normales es otra función normal. | | V | F |

2) Calcular las probabilidades o área bajo la curva de Gauss de cada uno de los siguientes casos:

- a) Z entre 0 y 1,6; b) Z entre $-2,1$ y 0; c) Z entre $-1,96$ y 1,96; d) Z entre $-2,04$ y 1,56;
e) a la izquierda de $Z = -1,73$; f) a la derecha de $Z = -0,56$; g) a la derecha de $Z = 1,89$.

3) Determinar los valores de Z conociendo los valores de probabilidad gaussiana de estos casos:

- a) entre 0 y Z es 0,2734; b) a la izquierda de Z es 0,9463; c) A la derecha de Z es 0,7257;
d) entre $-1,5$ y Z es 0,4332; e) entre $1,27$ y Z es 0,2146.

4) Suponiendo que el peso de los estudiantes de la Facultad se distribuya normalmente con media de 63 Kg y desvío de 7 Kg. Hallar la probabilidad de que uno de ellos escogido al azar pese:

- a) más de 75 Kg; b) menos de 55 Kg; c) entre 55 y 74 Kg; d) exactamente 70 Kg

5) Suponiendo que la altura de los mismos estudiantes se distribuya normalmente con una media de 1,68 m y un desvío de 0,17 m y que sean 4.500 en total. Hallar cuántos de ellos tienen una estatura: a) de más de 1,8 m; b) de menos de 1,6 m; c) entre 1,65 y 1,75 m; d) de 1,72 m

6) Haciendo los supuestos de normalidad correspondientes, hallar las frecuencias esperadas en todos los problemas de los capítulos anteriores donde se hallaron con Binomial, Poisson, etc.

7) Aplicar a los casos anteriores las correcciones de Yates y de Molenaar para decidir cual es la mejor de ambas y en que casos conviene usar una u otra.

8) ¿Qué significan los dos puntos de inflexión?

10

Teoría de muestras

En este capítulo se resume la Teoría de Muestras estadística, la cual trata el concepto de estudiar una población desconocida tomándole muestras, y a través del estudio de las mismas poder hacer inferencias acerca de toda la población. Primero se analiza el caso del muestreo aleatorio simple y estratificado, mostrando el manejo de una tabla de números aleatorios. Luego se ven los tipos no-aleatorios en el muestreo y se discute acerca de las ventajas y desventajas de cada uno de los métodos. Se explican los métodos usados en Bioquímica para lograr que las muestras extraídas a los pacientes cumplan los requisitos de aleatoriedad, aunque sea aproximadamente. Lo mismo para el caso de la industria Farmacéutica y para las simulaciones en los muestreos de mercadeo, usadas en el comercio en general. De manera tal de poder aplicar luego los modelos estadísticos que exigen tal requisito. En la Tabla 3, del fascículo de las tablas, se presenta la Tabla Aleatoria, más conocida como: “Random Numbers”.

10.1 Introducción

Las razones para efectuar un muestreo a una población, en lugar de estudiarla directamente, pueden ser varias como se puntualiza a continuación:

- . *El tamaño de la población es infinito.*
- . *El muestreo es de tipo destructivo.*
- . *La población es finita, pero demasiado grande.*
- . *Sería muy caro estudiar a toda la población y basta con deducciones aproximadas.*
- . *Tomaría demasiado tiempo analizar la población total.*

Cuando las muestras son lo suficientemente grandes, se pueden hacer inferencias analíticas bastante extensas, con pocos y simples recursos, en comparación con técnicas más refinadas de la Estadística. Esto es conveniente desde un punto de vista didáctico. La Teoría del muestreo es el estudio de las relaciones entre una población y las muestras que se extraen de ella. Del análisis de las muestras se pueden *estimar o inferir* datos de la población como su media (μ), varianza (σ^2), etc., llamados *parámetros poblacionales*, denotados usualmente con letras griegas, a partir de los valores obtenidos de la muestra, tales como la media muestral \bar{x} , la varianza muestral DS^2 , etc. Por ejemplo, en Bioquímica el verdadero valor de glucosa de un paciente μ es siempre desconocido y además variable con el tiempo, por ello se le extrae una muestra de sangre para poder estimarlo a través de mediciones. Así, el valor \bar{x} obtenido (ya sea con una o más mediciones) es el que se pone en el informe de la determinación clínica. El problema consiste en extraer muestras lo más representativas posibles, de la población desconocida, para que tengan sen-

tido las estimaciones realizadas a través de ellas. Cuando en Farmacia las mediciones se hacen a través de encuestas; se miden los porcentajes de las respuestas obtenidas. Esto se emplea usualmente en investigaciones de mercado, técnicas de propaganda, estudios poblacionales, etc. Por ejemplo, en una encuesta sobre el uso de determinado producto cada respuesta favorable se puede considerar un éxito, y la proporción de éxitos p en el total de las encuestas realizadas se puede usar para estimar la verdadera proporción π en la población tomada como marco de referencia del estudio.

Cuando la población sea finita y de un tamaño manejable en tiempo y costo, los valores poblacionales se calculan directamente, sin necesidad del muestreo. Por ejemplo, si se trata de revisar diamantes, a nadie se le ocurriría tomar muestras sino que se controlarían uno por uno.

Ahora bien, cuando se efectúan mediciones de magnitudes clínicas de tipo cuantitativo, la idea teórica es que se pueden efectuar infinitas mediciones y, en tal caso, el tamaño de la población será infinito. Es el caso de mediciones repetidas de un mismo objeto como pesar un cuerpo, o medirle su longitud, etc., se puede efectuar todas las mediciones que se deseen. Esto no ocurre cuando el ensayo es destructivo para la muestra. Por ejemplo, en mediciones clínicas de calcio, colesterol, hierro, etc., porque al suero se le adicionan los reactivos químicos y sirve para una sola vez. O bien, cuando se controla la cantidad de los componentes activos en un remedio, no hay otra forma que destruirlo para poder revisarlo. Entonces, el número total de determinaciones posibles dependerá de la cantidad de material disponible, que es finito. En resumen, en estos casos se acostumbra a considerar al valor verdadero de la magnitud medida como desconocido. El único medio para estimarlo es la toma de muestras.

10.2 Muestras aleatorias

Las muestras extraídas deben ser representativas de la población en estudio para que las conclusiones obtenidas sean válidas. El estudio de los diferentes métodos de muestreo y las distintas variantes ocasionadas se llama *diseño de experimentos*. Para extraer una muestra representativa se usa el método del *muestreo al azar* o *aleatorio*, que *asigna a todo integrante de la población la misma probabilidad de ser elegido*. Cuando por lo menos uno de los elementos de la población tiene una probabilidad mayor (o menor) de ser seleccionado que el resto de los integrantes de la misma, entonces se trata de un método de *muestreo no aleatorio*.

Un sorteo es el único método seguro de hacer equiprobables a todos los integrantes de la población en el proceso de selección. Por ejemplo, se le asigna un número identificador a cada uno, luego se colocan estos números en una urna, se mezclan bien, para luego ir sacando los números de la misma hasta completar el tamaño de muestra deseado. O bien, se procede a sortearlos con un bolillero como en la Lotería Nacional. En el caso de poblaciones humanas se puede usar el número de documento de identidad, o el de la libreta universitaria en el caso de alumnos de una Facultad. Se debe insistir en que el sorteo sea al azar. Muchas veces, equivocadamente se cree que es lo mismo elegir “a dedo” a los integrantes de la muestra porque el investigador no tiene preferencias por alguno de ellos, y por lo tanto elegirá a cualquiera. No es así. Se trata de una creencia subjetiva del mismo, indemostrable para los demás. Por ejemplo, hay que sacar una rata de una jaula que contiene muchas de ellas; la mala costumbre es abrir la tapa y tomar una sin

más. Por ser ratas, el investigador cree no tener preferencia por ninguna y, por lo tanto, todas tendrían la misma probabilidad de salir. Sin embargo, en realidad toma a la que se deja atrapar, ya sea por no ser tan veloz para escapar como sus compañeras, ya sea por estar enferma, o por ser muy gorda como para reaccionar ágilmente. La *única forma de asegurar el azar* en la selección es numerar las muestras de alguna manera y hacer un sorteo para elegirla. Es seguro que así nadie podrá ponerle objeciones a la mecánica de selección empleada.

Hay casos especiales donde no es tan simple numerar a los integrantes de la población para poder efectuar el sorteo, como sacar una muestra de granos de arroz de una bolsa de 50 Kg, o fideos de su contenedor, o yerba mate de una ponchada, etc. En tales casos, la solución es la homogeneización de la población previo a la extracción de la muestra. Por ejemplo, si se colocan los granos de arroz en una mezcladora y se agitan el tiempo suficiente, luego se toma la muestra en cualquier sector cuando esté bien homogeneizada. Lo mismo al tomar muestras de agua de río. Se supone que la correntada mezcla bien e impide la formación de grupos de coloides como los formados por el sulfato de aluminio cuando se trata de hacerla potable. La misma idea se aplica en el caso de extracción de sangre a pacientes para efectuar análisis clínicos. El torrente sanguíneo la homogeneiza tan bien que al puncionar en la vena se tiene una muestra representativa del paciente. Ese motivo hace desaconsejable la extracción en otra parte diferente de cuerpo del paciente, como pinchar un dedo o una oreja, excepto por fuerza mayor. Hay casos más difíciles o imposibles de homogeneizar como tomar una muestra representativa de tierra de un campo, o de arena en una playa. Aquí hay que subdividir la superficie en sectores, y efectuar un sorteo para seleccionar los lugares de extracción.

Para realizar un sorteo al azar hay diversos procedimientos. La manera clásica es colocar papeletas numeradas o con los nombres de los participantes en una bolsa, revolver y sacar una por una las mismas sin mirar. Una mejor manera es numerar a cada integrante de la población y usar un bolillero. Nadie, salvo la Lotería Oficial, puede manejar un bolillero con miles de números. Es más práctico usar un bolillero pequeño con diez bolillas numeradas de 0 a 9. Se procede a efectuar un sorteo para el número que va en la primera unidad, se repone la bolilla extraída y se hace un segundo sorteo para el número correspondiente a la decena, un tercero para la centena y así sucesivamente. De esta manera, no importa el tamaño de la población. Por ejemplo, si se tienen 87.000 integrantes se usan las diez bolillas para los tres últimos números; en cambio para los dos primeros se desechan los que no corresponden; si sale un 9 en el caso del primero, se realiza otra extracción para que salga alguno de los posibles. En Estadística se usa la Tabla de Números al Azar (Tabla 3) en lugar de bolilleros.

Dicha tabla consiste de una serie de números obtenidos con sorteos, o con un *software* de computadora generador de los *random numbers*. En la Tabla 3 se muestra una de estas tablas. El procedimiento es como sigue.

- . Se parte de cualquier lugar de la tabla elegido a dedo. O lo que es mejor, se pone el dedo y se eligen dos números para identificar la fila de inicio, y los dos siguientes para identificar la columna de inicio. Como solo hay 45 filas y 50 columnas, si las dos cifras resultan mayores se sigue tomando de dos en dos hasta encontrar un valor adecuado.
- . Con el número de fila y columna se localiza el primer número, desde donde se comienza.

. Se toman los números secuencialmente, empezando con el primero a la derecha del punto de partida y se lee la tabla de derecha a izquierda. Al terminar el renglón se sigue con el renglón de abajo de derecha a izquierda y así sucesivamente.

. También se puede usar otro sentido, como de izquierda a derecha, de arriba hacia abajo o viceversa. Pero *no se puede avanzar en diagonal*.

. Los números leídos se toman en grupos correspondientes al tamaño de la población. De a uno si hay menos de 10 integrantes, de a dos si hay menos de 100, de a tres si son menos de 1000, etc. Cuando la secuencia elegida no corresponde a ninguno de los integrantes, se desecha. Por ejemplo, si hay 85 integrantes y el número que sale es mayor, no se lo toma en cuenta y se sigue con el siguiente grupo de dos números.

La división de a cinco números en la tabla para formar columnas no tiene significado alguno, como el dejar un renglón en blanco cada cinco. Se ponen así para facilitar la lectura de la misma. Algunos autores objetan la elección a dedo del primer número. Su argumento es que al usar mucho la tabla de estas se memorizan ciertos grupos y el inconsciente puede jugar una mala pasada. Para evitar este inconveniente, recomiendan lo siguiente: a dedo se elige el lugar de inicio, pero como se explicó antes los dos primeros números se usan para elegir la columna, y los dos siguientes para elegir la fila, del número inicial como punto de partida. Así, si los cuatro primeros números son 2305, entonces el número de comienzo no es donde se colocó el dedo sino el dígito ubicado en la columna 23 y en la fila 5. En síntesis, con una simple hoja de papel se evita cualquier bolillero.

El proceso para fabricar estas tablas debe cumplir con las condiciones de equiprobabilidad e independencia, esto es:

- Todo número de la tabla tiene la misma probabilidad de salir que los demás. Esto es, en cualquier lugar de la tabla la cifra escogida tiene una probabilidad $p = 0,1$ de aparecer.
- La aparición de cierta cifra en la tabla no condiciona para nada la cifra siguiente ni es condicionada por la anterior.

Además de usarse para escoger *muestras aleatorias simples*, estas tablas tienen otros usos. Por ejemplo, cuando se va a testear a dos grupos de pacientes, uno para control y el otro que recibirá el tratamiento. Del total de pacientes posibles se usa la tabla para seleccionar a los integrantes de cada grupo, pero si hay que tomar en cuenta el sexo entonces se debe hacer un sorteo dentro de los pacientes varones, y otro dentro de las mujeres. Así, se garantiza igual probabilidad para cada *estrato*. Esto es, se divide primero a la población en estratos, como sexo, edad, raza, religión, nivel de educación, nivel de ingresos, etc., y luego se procede a seleccionar con un muestreo al azar dentro de cada estrato. A este tipo de procedimiento se lo denomina *muestreo aleatorio estratificado*. La teoría demuestra que la información obtenida en un muestreo estratificado es mucho más rica que en uno simple.

Casi siempre los pacientes entran a las pruebas clínicas, o a las farmacias, de un modo temporal, o sea, a medida que van llegando. Otra vez se puede usar la tabla para garantizar la aleatoriedad. Para el ejemplo anterior se puede hacer un sorteo para determinar cuál paciente irá a control y cuál recibirá el tratamiento, usando una regla de reparto. Por ejemplo: los que obtengan

de 0 a 4 serán de control y de 5 a 9 para el tratamiento. Si de la tabla se extraen los números: 3, 8, 5, 0, 4 ... significa que al primer paciente que llegue le toca el número 3, o sea, irá para control el segundo un 8 y recibe el tratamiento; lo mismo el tercero al que le tocó un 5 los dos siguientes a control y así sucesivamente. Cuando se completa uno de los grupos, se sigue el orden del sorteo sólo con los del otro hasta terminar. Como se ve, siempre se puede encontrar la manera de aleatorizar una muestra de una manera práctica. En el caso de arribo de clientes a una farmacia se puede seguir un procedimiento similar para, por ejemplo, otorgar un premio a modo de propaganda. En una industria, generalmente se seleccionan en forma aleatoria los productos finales o materias primas que irán al laboratorio de control de calidad.

10.2.1 Aplicaciones en Medicina

Los métodos más comunes se conocen con el nombre de *ciegos* o *doble-ciegos*. Esto se refiere a que hay casos donde el director del proyecto de investigación es la única persona que conoce la identidad de las muestras individuales. Por ejemplo, cuales son las muestras de control y cuales son las de los casos a investigar, o bien cuando en un control de calidad se introducen las muestras de control sin que los operarios del sistema lo sepan. De esta forma, el personal que trabaja efectuando mediciones no sabe (*está ciego*) con cuales muestras su jefe los está controlando. Así se puede verificar el comportamiento del personal de un laboratorio, sin que estos lo sepan. Cuando se hace un trabajo de comprobación de la efectividad de un medicamento, los enfermos son separados en dos grupos, uno al cual se le administra el medicamento (casos) y el otro al cual se le suministra un placebo (control). El paciente nunca sabe si se trata de un caso o de un control. Si el director del proyecto decide hacer un trabajo del tipo *ciego*, entonces evita que su personal sepa que pacientes son los controles y quienes los casos. De esta forma se asegura que ambos grupos sean tratados por igual: mismas condiciones ambientales y alimenticias, misma atención personalizada, etc. Se asegura que las dosis y los placebos tengan las mismas características físicas, para evitar que alguien pueda distinguirlos, y mediante un muestreo aleatorio designa quienes van a formar parte de cada grupo. Por sorteo, también asigna al personal y a los pacientes que van a conformar su grupo de trabajo.

En ciertas ocasiones, tampoco el director conoce quienes son los controles y quienes los casos. Es decir, una tercera persona ajena al proyecto de investigación, es el único que los puede identificar, y solo al terminar el trabajo procede a hacerlo. Generalmente, los nombres de los pacientes y su condición se guardan en un sobre cerrado, para ser abierto al final. A este caso, se lo conoce como un *doble-ciego*. Es otra forma de evitar influencias externas que puedan alterar los resultados como consecuencia de la tendencia humana a tener intereses en uno u otro sentido.

Cuando se trata de evaluar la calidad de un test o un diagnóstico clínico, el mejor procedimiento para evitar influencias ajenas al proyecto es la siguiente:

- 1) El investigador decide el tamaño muestral total N que va a usar en su experimento, de entre todos los casos disponibles que tenga.
- 2) Por casos disponibles se entiende aquellos sujetos que pueden ser clasificados con certeza en sanos o enfermos (los casos dudosos se descartan). Así escoge con un muestreo aleatorio a los casos que va a usar con una proporción del 50%. O sea, $TE = TS = N/2$.

A este método se lo conoce como muestreo basado en la enfermedad. De esta forma los cuatro casos mutuamente excluyentes de la tabla de verdad son naturales, así como los totales de casos diagnosticados como positivos y negativos. Si en el estudio, se usan estos valores, entonces se puede tener la tranquilidad de que los resultados no tengan ningún tipo de influencia. En cambio, si se usan los índices que contengan algunos los valores elegidos por el investigador, como Sensibilidad, Especificidad, etc., ya no se puede estar tan seguro. Notar que si el investigador se decide por un muestreo basado en los diagnósticos, la prevalencia de la enfermedad en la población de referencia podría afectar los datos. Es decir, si el investigador decide adoptar N con un 50% de positivos y el otro de negativos ($TP = TN = N/2$), el total de enfermos que obtenga no sería el mismo en un hospital general, que en un servicio especializado adonde se derivan los pacientes con esa enfermedad. Por ejemplo, la cantidad de enfermos que concurren a un centro médico con un diagnóstico presunto de infarto de miocardio, normalmente es mucho menor porcentualmente, que la cantidad que llega a una unidad coronaria.

En Epidemiología se usan ambas maneras de efectuar los muestreos. En estos casos se trata de armar una tabla de 2×2 para analizar la relación entre un factor de riesgo (o uno de protección como la inmunización de una vacuna) con la enfermedad estudiada. Un *muestreo basado en la exposición* es el que se usa cuando el investigador decide adoptar el número de sujetos expuestos al factor de riesgo y el de no-expuestos. Aquí las costumbres no se basan en un 50% para cada caso sino en criterios médicos derivados del conocimiento que se tiene de la situación, o en casos similares ya estudiados y publicados, o sencillamente en el número de casos disponibles cuando la enfermedad es rara como una meningitis, etc. En cambio, un *muestreo basado en la enfermedad* consiste en que el investigador elige el número total de enfermos y de sanos que va a emplear en su experimento, y generalmente la relación no es del 50% sino en alguna otra derivada de experiencias anteriores. Por ejemplo, en el caso de la meningitis no es raro que la proporción usada sea un caso de enfermedad y cuatro casos como control. En este punto, conviene hacer una aclaración importante:

- 1) El primer paso consiste en seleccionar todos aquellos sujetos que se adapten a los requerimientos específicos de la investigación y descartar a los demás.
- 2) Una vez que se tiene el número de total de sujetos disponibles, el investigador debe decidir el tamaño muestral N generalmente menor que el total.
- 3) De acuerdo al tipo de muestreo que vaya a efectuar (por enfermedad o por exposición) el siguiente paso es adoptar el tamaño muestral de cada total marginal de la tabla.
- 4) Recién entonces, empleando un muestreo aleatorio selecciona a los sujetos.

Esta recomendación se hace, porque a veces se hace la selección al azar primero, y luego se eligen quienes formaran parte de la muestra y quienes no de acuerdo a ciertos criterios médicos. Pero esto invalida la representatividad de la muestra en la población, porque la eliminación de ciertos sujetos debe hacerse antes del sorteo y no después.

10.3 Muestras no aleatorias

Cuando el método de extracción de las muestras no asegure a cada individuo de la población o del estrato, igual probabilidad de ser elegido, entonces la muestra obtenida no es aleatoria. A veces, esto se hace por razones de practicidad en el sentido del costo o del tiempo. Si se desea tomar una muestra probabilística de la población argentina no parece razonable usar a cada

individuo como unidad de muestreo. Lo mismo cuando se desea hacer un muestreo a los escolares de una provincia, es muy difícil empadronar a todos primero para luego sortear, y se tardaría demasiado para ubicarlos uno por uno hasta terminar el trabajo.

En el *muestreo de etapas múltiples* se utiliza para el caso de grandes poblaciones humanas. Acá, la unidad de muestreo en la primera etapa son los departamentos de cada provincia. Se los lista y se hace un primer sorteo para la selección. En una segunda etapa se distingue la población rural de la urbana, subdividiendo en fracciones (diferentes superficies con densidad de población semejante). Otra vez se sortea para elegir, y se continúa con otra división en radios dentro de las fracciones, segmentos dentro de radios, y así sucesivamente. La razón es repartir equitativamente el trabajo del encuestador.

En el *muestreo por conglomerados* se eligen conjuntos donde naturalmente se agrupan los individuos. Es, por ejemplo, el caso de las escuelas para hacer un muestreo alumnos en el sistema educativo, o las facultades para los universitarios. Si se trata de estudiar las condiciones laborales de los empleados de comercio que trabajan en supermercados, primero se empadronan a los lugares naturales de trabajo (supermercados), y luego se sortea entre estos conglomerados para elegir a uno. Luego se entrevista a todos los empleados del supermercado elegido, y se acepta esto como una muestra representativa del sector.

El *muestreo sistemático* se usa para el caso de sucesiones de elementos. Por ejemplo, el caso de las historias clínicas de pacientes, certificados de nacimiento, tarjetas de catálogo en una biblioteca, etc. Son los casos donde la información está en archivos y hay que trabajar con estos para obtenerlas. Se elige una cifra entera, razonable, tomando en cuenta el tamaño de la muestra y el de la población. Por ejemplo, hay que tomar una muestra de tamaño 25 de un archivo que contiene 488 fichas; luego, el cociente entre población y muestra es $488 / 25$, aproximadamente 19. Notar que si se elige 20 el tamaño muestral no llega a 25. Entonces, se cuentan las fichas y a llegar a la décimo novena se la extrae, se sigue hasta la número 38 que será la segunda escogida, y así sucesivamente hasta tener las 25 fichas necesarias. Es también el caso de los soldados que se numeran de 1 en adelante y cada 5 (u otro número cualquiera) dan un paso al frente. Es un método sencillo y rápido de selección.

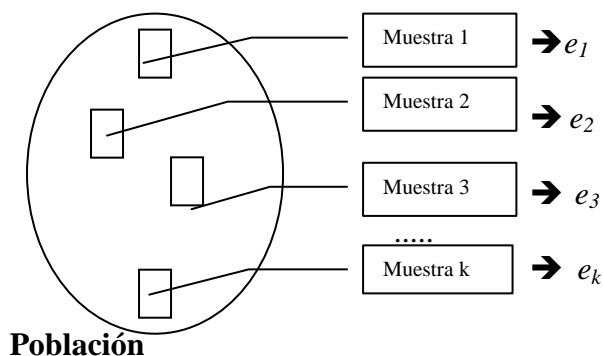
Hay otros casos de muestreo no aleatorios de uso común. Como el de tomar una lista de nombres, cerrar los ojos y con la punta de un lápiz marcar a uno de ellos, para escogerlo. Son los casos de los programas de TV donde se toma la guía telefónica, se la abre en una página cualquiera y se escoge a uno de los números que figuran, para luego hacer un llamado con premio. En los juegos infantiles se hace una ronda con los participantes, se vendan los ojos del que va a elegir, se le hace dar varias vueltas con los ojos vendados para que marque a alguno. Todos los casos vistos tienen una característica común: *los individuos de la población no son equiprobables*.

10.4 Distribuciones de probabilidad en el muestreo

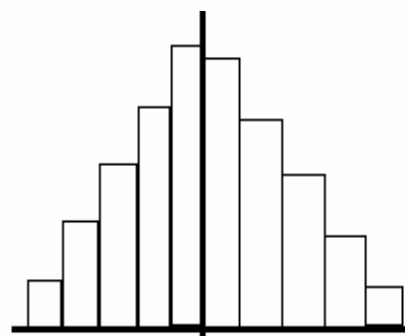
También llamadas *distribuciones muestrales* de un estadígrafo cualquiera obtenido a través de la muestra. La idea es la siguiente: si se toman k muestras, *todas las posibles* de tamaño n (con o sin reemplazamiento) de una población de tamaño N_p , y a cada muestra se le calcula un estadígrafo e (media, mediana, varianza, etc.), se obtienen una serie de k valores: $e_1, e_2, e_3, \dots, e_k$

Estos valores pueden agruparse mediante un histograma de frecuencias para poder apreciar la forma de la distribución de los mismos. En la Figura 10.1 se esquematiza esta situación:

Figura 10.1 Distribuciones muestrales.



Histograma con los valores e_i



De una población cualquiera se extraen k muestras; cada una permite calcular k estadígrafos con los cuales se puede hacer un histograma como el de la derecha de la Figura 10.1. Se aprecia que este histograma adquiere forma de campana si se suavizan los escalones, al achicar los intervalos. Esta curva obtenida a partir de datos muestrales, observados a través del muestreo, tiende asintóticamente a otra curva teórica a medida que k aumenta, y los intervalos se hacen infinitesimales. Dicha curva teórica es la función de Gauss de acuerdo con el Teorema Central del Límite, el principal de la Estadística.

El Teorema Central del Límite permite establecer que, en condiciones muy generales, si la muestra es lo *suficientemente grande*, la distribución teórica de los k valores obtenidos es aproximadamente la función de Gauss. Esta es la base de la *Teoría de las Grandes Muestras*. Las principales *Distribuciones Muestrales* son funciones de Gauss identificadas en forma unívoca con sus dos parámetros μ y SE. En la Tabla 10.1 se presentan estos dos valores para cada uno de los estadígrafos más usuales. En la primera columna de la tabla se muestra cada estadígrafo, en la segunda columna se da la fórmula para el cálculo del error típico de estimación SE. Finalmente en la tercera columna se muestra la estimación puntual para obtener el valor esperado del estadígrafo μ_e , con las aclaraciones respecto al tamaño muestral requerido para que tal estimación sea considerada aceptable.

10.4.1 Distribución muestral de la media

Si el estadígrafo elegido es la media, se tendrán $\bar{X}_1, \bar{X}_2, \bar{X}_3, \dots, \bar{X}_k$ medias muestrales; estas se distribuyen normalmente si k es muy grande. En la práctica, 30 o más valores son suficientes. En la teoría, cuando $k \rightarrow \infty$ entonces la distribución muestral de la media es *asintóticamente normal* y coincidirá con la función de Gauss. Esta distribución tendrá un valor esperado y una varianza que permitirán estimar los respectivos valores poblacionales. O sea,

$$\mu_{\bar{X}} = \mu \qquad \sigma^2_{\bar{X}} = \sigma^2 / n = SE^2(\bar{X}) = VAR(\bar{X})$$

Tabla 10.1. Errores típicos para algunas distribuciones muestrales.

Distribución muestral	Error típico	Notas especiales
Medias	$SE(\bar{x}) = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	Se cumple para muestras grandes o pequeñas. La distribución muestral de medias se ajusta mucho a una normal para $n \geq 30$ incluso para poblaciones no normales. $\mu_{\bar{x}} = \mu$ en todos los casos.
Proporciones	$SE(\pi) = \sigma_{\pi} = \sqrt{\frac{\pi(1-\pi)}{n}}$	Las notas anteriores para medias son igualmente aplicables aquí con $n > 25$ $\mu_p = \pi$ en todos los casos.
Desviaciones típicas	(1) $SE(DS) = \sigma_{DS} = \frac{\sigma}{\sqrt{2n}}$ (2) $SE(DS) = \sigma_{DS} = \sqrt{\frac{\mu_4 - \mu_2^2}{4n\mu_2}}$	Para $n \geq 100$, la distribución muestral de s es muy próxima a una normal. σ_s esta dada por (1) solamente cuando la población es normal (o aproximadamente normal). Si la población no es normal, puede utilizarse (2). Nótese que (2) pasa a ser (1) cuando $\mu_2 = \sigma^2$ y $\mu_4 = 3\sigma^4$, lo que se cumple para poblaciones normales. Para $n \geq 100$, $\mu_s = \sigma$ con gran aproximación.
Medianas	$SE(\text{Med.}) = \sigma_{\text{med}} = \frac{1,2533 \sigma}{\sqrt{n}}$	Para $n \geq 30$, la distribución muestral de la mediana es muy próxima a una normal. Los resultados dados son válidos solamente si la población es normal (o aproximadamente normal) $\mu_{\text{med}} = \mu$
Cuartiles primero y tercero	$SE(Q_2) = \sigma_{Q_2} = \sigma_{Q_3} = \frac{1,3626 \sigma}{\sqrt{n}}$	Las notas anteriores para medianas son igualmente aplicables aquí. μ_{Q_1} y μ_{Q_3} son casi iguales al primero y tercer cuartil de la población. Nótese que $\sigma_{Q_2} = \sigma_{\text{med}}$
Varianzas	(1) $SE(\text{Var}) = \sigma_{\text{var}} = \sigma^2 \sqrt{\frac{2}{N}}$ (2) $SE(\text{Var}) = \sigma_{\text{var}} = \sqrt{\frac{\mu_4 - \mu_2^2}{n}}$	Las notas para desviaciones típicas son igualmente aplicables aquí. Nótese que (2) pasa a ser (1) en caso de que la población sea normal. $\mu_s^2 = \sigma^2 (n-1) / n$ que es casi igual a σ^2 para valores grandes de n .
Coefficiente de variación	$SE(\text{CV}) = \sigma_{\text{CV}} = \frac{\text{CV}}{\sqrt{2n}} \sqrt{1 + 2(\text{CV})^2}$	Aquí $\text{CV} = \sigma / \mu$ es el coeficiente de variación de la población. Los resultados dados son válidos para poblaciones normales (o aproximadamente normales) y $n \geq 100$.

Esto es: la media aritmética de las k medias muestrales obtenidas es aproximadamente igual a la media poblacional (o valor verdadero). Sin embargo, esta aproximación tiene un error de estimación denominado *error típico* o *error estándar* de estimación que en el caso de la media es: $\sigma_{\bar{x}}$. En la bibliografía clínica la nomenclatura más empleada es $SE(\bar{x})$. En la Tabla 10.1 se muestran los valores anteriores para el caso de la media aritmética.

Las relaciones anteriores son válidas solo si la población es infinita, o si es finita, pero el muestreo es con reemplazamiento. Caso contrario, cuando la población es finita y se realizan muestreos sin reposición, entonces se deben ajustar dichas relaciones con:

$$\mu_{\bar{x}} = \mu \qquad \sigma^2_{\bar{x}} = (\sigma^2 / n) [(N_P - n) / (N_P - 1)] = SE^2(\bar{x}) = \text{VAR}(\bar{x})$$

En el Cuadro 10.1 siguiente se presenta un problema de aplicación para un caso donde se conoce a toda la población, los parámetros poblacionales se calculan directamente aplicando las fórmulas vistas en el Tema 4 resultando: $\mu = 4,5$ y $\sigma^2 = 1,25$. Se pueden verificar las relaciones anteriores de dos maneras. En la primera se toman las seis muestras posibles de tamaño 2, para un muestreo sin reemplazamiento. A cada muestra se le calcula su media respectiva, luego con estos 6 promedios se pueden calcular: el promedio y la varianza de esas seis muestras. Ahora, el promedio de todas las medias muestrales da exactamente igual al valor medio poblacional y la varianza de las medias muestrales verifica la relación anterior, si se aplica el factor de corrección para muestras de tamaño finito. La segunda manera (*Bootstrap procedure*) es tomando muestras con reposición, primero se toman todas las 16 muestras posibles con reemplazamiento de tamaño 2. Luego, con las dieciséis muestras se calculan las 16 medias respectivas. Por último, se calcula el promedio y la varianza de estos 16 valores, verificando de nuevo las relaciones vistas más arriba, para el caso de muestras con reposición. Para el otro problema, se suponen conocidos los valores poblacionales, y tomando 50 muestras de tamaño 3 hay que determinar la cantidad de casos donde el resultado esté comprendido en un intervalo (6 ; 7,796). La forma de proceder es calculando primero las probabilidades de obtener esos resultados límites, luego por diferencia calcular la probabilidad gaussiana asociada al intervalo y entonces, multiplicando dicha probabilidad por el tamaño muestral, se puede contestar la pregunta efectuada.

CUADRO 10.1: Distribución Muestral de Medias.

Caso 1) Un jefe de laboratorio quiere determinar cuántos Calcios realiza su personal en una hora de trabajo. Tiene a cuatro bioquímicos trabajando y los resultados obtenidos son:

<i>Bioquímico</i>	<i>Cantidad</i>	
A	4	Si considera a estos datos como su población esta se representa con el conjunto (3 , 4 , 5 , 6) Donde los parámetros poblacionales serán $\mu = 4,5$ y $\sigma^2 = 1,25$ ($\sigma = 1,118$)
B	6	
C	5	
D	3	

Si se toman todas la muestras posibles, sin reemplazamiento, de tamaño 2, se tendrán 6 pares en total, que son las combinaciones de 4 elementos tomados de a 2 ($C = 4! / 2! \cdot 2!$) O sea,

<i>Combinaciones</i>	<i>Resultados</i>	<i>Media muestral</i>	
(A , B)	(4 , 6)	5	El promedio de estas medias es $\frac{5 + 4,5 + 3,5 + 5,5 + 4,5 + 4}{6} = 4,5$ valor igual al poblacional
(A , C)	(4 , 5)	4,5	
(A , D)	(4 , 3)	3,5	
(B , C)	(6 , 5)	5,5	
(B , D)	(6 , 3)	4,5	
(C , D)	(5 , 3)	4	

Esto es $\mu_{\bar{x}} = 4,5 = \mu$

Y la varianza de estos valores es $\sigma^2_{\bar{x}} = 0,417 = (1,25 / 2) [(4-2) / (4-1)]$ verificando la relación vista más arriba. Si se toman todas las muestras con reemplazamiento, hay 16 casos:

(AA) (AB) (AC) (AD) (BA) (BB) (BC) (BD) (CA) (CB) (CC) (CD) (DA) (DB) (DC) (DD)

O sea, las muestras son:

(4;4) (4;6) (4;5) (4;3) (6;4) (6;6) (6;5) (6;3) (5;4) (5;6) (5;5) (5;3) (3;4) (3;6) (3;5) (3;3)

Y las medias de las muestras respectivas son:

4 ; 5 ; 4,5 ; 3,5 ; 5 ; 6 ; 6,5 ; 4,5 ; 4,5 ; 5,5 ; 5 ; 4 ; 3,5 ; 4,5 ; 4 ; 3

Luego el promedio de este conjunto de 16 medias es $\mu_{\bar{x}} = 4,5 = \mu$. Y la varianza de estos valores es $\sigma^2_{\bar{x}} = 0,625 = 1,25 / 2 = \sigma^2 / n$. Las relaciones se cumplen para ambos casos de muestreo, tanto si es con o sin reposición.

El método del *jackknife* consiste en sacar secuencialmente cada dato, calcular la media del grupo y luego reinsertarlo a los restantes para formar cada muestra. Para la población (4, 6, 5, 3) sería:

Muestras	Medias	Promedio general
(6, 5, 3)	4,6666	
(4, 5, 3)	4	4,5
(4, 6, 3)	4,3333	
(4, 6, 5)	5	

Notar que se puede “mirar” la estabilidad de la media que acá varía entre 4 y 5.

Ejemplo: Suponiendo que las proteínas totales se distribuyen normalmente entre los 2.500 varones sanos de una cierta población, con edades de 14 a 50 años, con un valor promedio de 7 g/dl y un desvío de 0,9 g/dl. Un investigador decide tomar 50 muestras aleatorias de tamaño 3, con reemplazamiento. Desea saber en cuántas muestras cabe esperar una media entre 6 y 7,796 g/dl.

El primer paso para resolver este problema es encontrar los parámetros de la distribución muestral de medias, usando los valores poblacionales conocidos. Entonces:

$$\mu_{\bar{x}} \approx 7 \text{ g/dl} = \mu \quad ; \quad \sigma^2_{\bar{x}} \cong \sigma^2 / n = (0,9)^2 / 3 = 0,27 \quad ; \quad \sigma_{\bar{x}} = \text{SE}(\bar{x}) = 0,52$$

Cabe esperar que la distribución muestral de medias sea una curva de Gauss con tales parámetros. Para calcular la probabilidad de que una muestra tenga una media entre 6 y 7,796 g/dl, se tipifican estos valores con:

$$z_6 = (x - \mu_{\bar{x}}) / \sigma_{\bar{x}} = (6 - 7) / 0,52 = -1,92$$

$$z_{7,6} = (x - \mu_{\bar{x}}) / \sigma_{\bar{x}} = (7,796 - 7) / 0,52 = 1,53$$

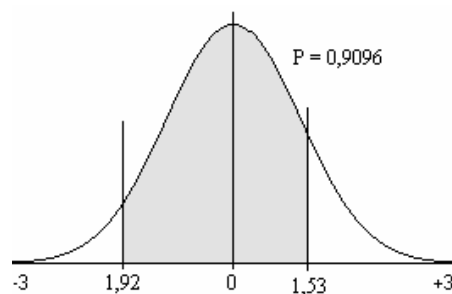
La probabilidad bajo la curva de Gauss tipificada dentro del intervalo - 1,92 ; + 1,53 se obtiene calculando las probabilidades:

$$P(-1,92 ; 0) = 0,4726$$

$$P(0 ; 1,53) = 0,4370$$

$$P(-1,92 ; 1,53) = 0,9096$$

Entonces, el número esperado de muestras buscado será la proporción anterior, multiplicada por el número total de muestras extraídas. Esto es, $(0,9096) 50 = 45,48$ muestras. Es de esperar encontrar 45 muestras con valores de proteínas totales entre 6 y 7,6 g/dl.



10.4.2 Distribución muestral de proporciones

Cuando el estadígrafo obtenido de cada muestra de una población sea la proporción de ocurrencia de un suceso dado, un “éxito”, se trata de una *distribución muestral de proporciones*. Por ejemplo, el caso del lanzamiento de una moneda es uno de población infinita porque se puede realizar la prueba cuantas veces se desee. La probabilidad de sacar un éxito, una cara es $p = 0,5$, la misma para todos los casos posibles, mientras que la probabilidad de “fracaso” al lanzar la moneda y salir seca es $q = 1 - p = 0,5$. Si el experimento consiste en realizar n lanzamientos de la moneda, una cantidad lo suficientemente grande, luego la frecuencia relativa del suceso cara, tendrá una distribución aproximadamente normal de parámetros:

$$\mu_p = \pi \quad \text{y} \quad \sigma^2 \pi = [\pi \cdot (1 - \pi)] / n = SE^2(\pi)$$

$$\text{O bien} \quad \mu_r = r = n \cdot \pi \quad \text{y} \quad \sigma^2_r = n \cdot [\pi \cdot (1 - \pi)] = SE^2(r)$$

CUADRO 10.2: Distribución muestral de proporciones.

Ejemplo) En un laboratorio farmacéutico, históricamente se ha encontrado que el 3% de las ampollitas fabricadas de cierto medicamento salen falladas. Se desea saber cuánto vale la probabilidad de encontrar $r = 20$ defectuosas o menos, en una partida de $n = 500$ producidas.

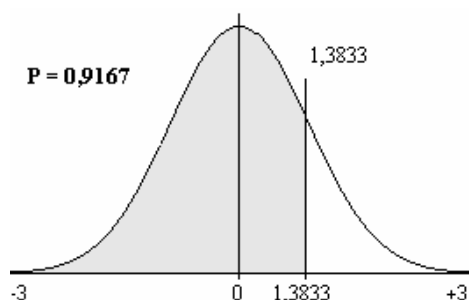
(a) La probabilidad exacta se puede calcular con la fórmula de la Binomial acumulada:

$$P_{Bi}(r \leq 20) = P_{Bi}(r=0) + P_{Bi}(r=1) + \dots + P_{Bi}(r=20) = 0,9202$$

(b) El segundo método es usando la aproximación normal a la binomial con la relaciones siguientes: $\mu_r = n \cdot \pi = 15$; $\sigma^2_r = n [\pi \cdot (1 - \pi)] = (500 \cdot 0,03 \cdot 0,97) = 14,55$; $SE(r) = \sigma_r = 3,8144$

Como se trata de una variable discreta y la función de Gauss es continua, hay que efectuar una corrección por continuidad de Yates o de Molenaar:

$$Z_{20} = [(4 \cdot 20 + 3)(1 - 0,03)]^{1/2} - [(4 \cdot 500 - 4 \cdot 20 - 1)(0,03)]^{1/2} = 1,3833 \text{ O sea } P = 0,9167$$



La probabilidad pedida es el área a la izquierda de $Z_{20} = 1,3883$ en la curva Gauss tipificada.

Con la corrección de Yates sería: $z = [(20 - 15) - 1/2] / 3,8144 = 1,1797$ y $P = 0,881$ O sea una aproximación peor al valor exacto dado en (a). Notar que si no se hubiese hecho la corrección por continuidad, la estimación de la probabilidad pedida sería peor aún: $P = 0,8168$.

Las que se obtienen de la relación vista más arriba con reemplazamiento de $\pi = r / n$. Por muestra lo suficientemente grande, se entiende en la práctica $n > 25$. Por otra parte, cuando la población es finita y el muestreo se realiza sin reposición, entonces las ecuaciones anteriores se modifican con:

$$\mu_p = \pi \quad \text{y} \quad \sigma^2 \pi = [\pi \cdot (1 - \pi) / n] [(N_p - n) / (N_p - 1)] = SE^2(\pi)$$

$$\text{O bien} \quad \mu_r = r = n \cdot \pi \quad \text{y} \quad \sigma^2_r = n [\pi \cdot (1 - \pi)] [(N_p - n) / (N_p - 1)] = SE^2(r)$$

Usando la distribución de Gauss, se pueden calcular las probabilidades asociadas en ejemplos como los presentados en el Cuadro 10.2.

Tanto la distribución de medias como la de proporciones se usan muy frecuentemente en la práctica. El problema básico es conocer los valores poblacionales porque es muy raro saberlos de antemano, salvo el caso de la moneda y otros juegos de azar; lo común es ignorar estos parámetros. Entonces, usando los valores obtenidos del muestreo aleatorio se pueden aproximar los valores desconocidos poblacionales, como se verá en el tema *inferencia estadística*.

10.4.3 Diferencias de medias y de proporciones

Muchas veces es necesario comparar dos poblaciones, como cuando se testea un tratamiento en pacientes contra un control o blanco; o bien, cuando se comparan dos sistemas de medición entre sí (estos sistemas pueden ser: técnicas de laboratorio, marcas comerciales, instrumentos de diferente marca, etc.). El problema es ver si hay diferencias entre ambos casos. El método es encontrar un estadístico para la diferencia (o suma) de ambos estadísticos muestrales. Se tienen dos poblaciones, se efectúa un muestreo aleatorio de tamaño n_1 en la primera de ellas, y se obtienen los valores de un estadígrafo cualquiera: e_1 . La población gaussiana de la distribución muestral tendrá los parámetros μ_{e1} y σ_{e1} . Análogamente, se procede con la segunda población y se obtienen los parámetros μ_{e2} y σ_{e2} . Se pueden combinar ambas poblaciones a través de un estadígrafo que sea la diferencia de ambos: $e = e_1 - e_2$. Este nuevo índice también tendrá una distribución gaussiana, pues la diferencia de dos funciones de Gauss es otra función gaussiana, siempre y cuando las muestras no dependan unas de otras; esto es, sean *independientes*. Sus parámetros son:

$$\mu_{e1-e2} = \mu_{e1} - \mu_{e2} \quad \text{y} \quad \sigma^2_{e1-e2} = \sigma^2_{e1} + \sigma^2_{e2} = SE^2(e_1 - e_2)$$

Si el estadígrafo es la media muestral $e = \bar{x}$ entonces, la *distribución muestral de la diferencia de medias* para poblaciones infinitas, de parámetros $(\mu_1; \sigma_1)$ y $(\mu_2; \sigma_2)$. Entonces:

$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$ es el valor esperado de la diferencia de medias y su error de estimación es:

$$SE^2(\bar{x}_1 - \bar{x}_2) = \sigma^2_{\bar{x}_1 - \bar{x}_2} = (\sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}) = [(\sigma^2_1 / n_1) + (\sigma^2_2 / n_2)]$$

Si el estadígrafo es una proporción $e = \pi$ entonces, la *distribución muestral de la diferencia de proporciones* para poblaciones infinitas, de parámetros $(\mu_1; \sigma_1)$ y $(\mu_2; \sigma_2)$. Luego:
 $\mu_{\pi_1 - \pi_2} = \mu_1 - \mu_2 = \pi_1 - \pi_2$ es el valor esperado de la diferencia de proporciones poblacionales.

$$SE^2(\pi_1 - \pi_2) = \sigma_{(\pi_1 - \pi_2)}^2 = [\pi_1(1-\pi_1)/n_1] + [\pi_2(1-\pi_2)/n_2]$$

Es el error estándar al cuadrado de tal estimación para la diferencia de proporciones.

En el Cuadro 10.3 se presentan dos ejemplos de aplicación de lo visto. En el primer caso se trata de una diferencia de medias usando la vida útil de dos medicamentos, mientras que en el segundo caso se trata de una diferencia de Sensibilidades de dos métodos clínicos (tomando como proporciones a las Sensibilidades) para poder compararlos. En ambos casos las probabilidades encontradas son muy chicas lo que indica diferencias entre las medias muestrales. Para poder decidir si tales diferencias son válidas se emplea la Teoría de la decisión estadística (Tema 12).

CUADRO 10.3: Diferencia de 2 muestras.

Caso 1) La vida útil de un medicamento fabricado por el Laboratorio A es de 1.400 días, con una desviación estándar de 200 días. Por su parte, el mismo medicamento fabricado por el Laboratorio B de la competencia tiene una duración de 1.200 días con un desvío de 100 días. Se eligen 125 medicamentos de cada Laboratorio con un muestreo al azar. Calcular la probabilidad que los del Laboratorio A duren 250 días más que los del B.

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_A - \mu_B = 1400 - 1200 = 200 \text{ días}$$

$$\sigma^2_{\bar{X}_A - \bar{X}_B} = (\sigma^2_{\bar{X}_A} + \sigma^2_{\bar{X}_B}) = [(\sigma^2_A / n_A) + (\sigma^2_B / n_B)] \quad \text{Como } n_A = n_B = 125$$

$$\sigma^2_{\bar{X}_A - \bar{X}_B} = [(200)^2 + (100)^2] / 125 = 400 \quad \text{O sea: } SE(\bar{X}_A - \bar{X}_B) = \sigma_{\bar{X}_A - \bar{X}_B} = 20 \text{ días}$$

La tipificación de esta diferencia se hace con: $Z_{A-B} = (250 - 200) / 20 = 2,5$

Entonces, la probabilidad pedida es igual al área a la derecha de este valor en la curva de Gauss. $P = (0,5 - 0,4938) = 0,0062$. Lo que significa que hay un 0,6% de probabilidad porcentual de que la vida útil del medicamento fabricado por A dure 250 días más que el de su competencia.

Caso 2) La Sensibilidad de una prueba clínica es del 85% si se usa la marca A, y del 80% si es la marca B. Se toman 100 muestras al azar para cada caso, y se desea saber la probabilidad que la diferencia de sensibilidades entre ambas sea del 10% o más. Se estima $\pi_1 \approx 0,85$ y $\pi_2 = 0,8$

$$(a) \quad \mu_{A-B} = \mu_A - \mu_B = 0,85 - 0,80 = 0,05 \quad \text{y} \quad SE_{(A-B)} = \sigma_{A-B} = 0,02681 \text{ pues:}$$

$$\sigma^2_{A-B} = [\pi_1(1-\pi_1)/n_1] + [\pi_2(1-\pi_2)/n_2] = (1/100) [(0,85 \cdot 0,15) + (0,8 \cdot 0,2)] = 0,002875$$

Si se hace la H_0 de que no hay diferencia entre ambas la variable tipificada es:

$$Z_{A-B} = [(0,1 - 0) - (0,85 - 0,80)] / 0,05362 = 0,09325 \quad \text{O sea } P(x < 10) = 0,5371 \text{ y la probabilidad pedida se calcula con } P(X \geq 10) = (1 - 0,5371) = 0,4629$$

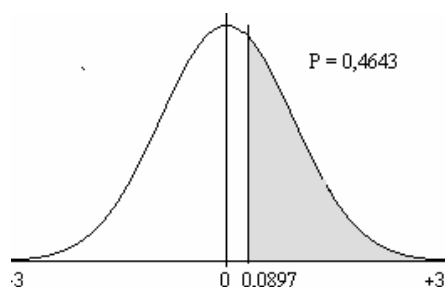
(b) Otra forma de resolver este problema es hacer el supuesto de que las muestras provienen de dos poblaciones diferentes, la primera con una probabilidad de 0,85 y la otra con 0,80.

En ese caso se puede suponer que la proporción $\pi = 0,05$ es la diferencia real entre ambas poblaciones y que tiene un desvío $\sigma = 0,05362$. En $n = 200$ muestras el valor esperado se calcula con $\mu_r = 200 \cdot 0,05 = 10$ y se pide hallar la probabilidad de que $r \geq 10$. O sea $[1 - P_{Bi}(r \leq 9)]$

Calculando el valor tipificado con la corrección de Molenaar resulta:

$$z = [(4 \cdot 9 + 3)(1-0,05)]^{1/2} - [(4 \cdot 200 - 4 \cdot 9 - 1)(0,05)]^{1/2} = 0,0897$$

O sea $P = 0,5357$ y la probabilidad pedida es $(1 - 0,5357) = 0,4643$



El área a la derecha del valor tipificado es $p = 0,4643$ O sea, existe una probabilidad del 46,43% que la marca A tenga una sensibilidad mayor del 10% a la de la marca B.

(c) Utilizando los mismos supuestos que en el punto anterior, pero trabajando con la probabilidad exacta, la respuesta se calcula con:

$$P_{Bi}(x > 10) = 1 - P_{Bi}(x \leq 9) = 1 - [P_{Bi}(x = 0) + P_{Bi}(x = 1) + \dots + P_{Bi}(x = 9)] = 0,4547$$

Notar que ambas aproximaciones normales son bastante buenas. La idea de usar la diferencia de dos poblaciones binomiales como si fuese una sola se explicará más adelante en más detalle.

10.5 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|--|---|---|
| 1) Una muestra es aleatoria si todo componente tiene la misma probabilidad de ser elegido. | V | F |
| 2) El muestreo sistemático es de tipo aleatorio. | V | F |
| 3) Los muestreos no aleatorios son: | | |
| 4) En una tabla de números al azar se puede ir en diagonales a través de ella. | V | F |
| 5) La base de la Teoría de las Grandes Muestras es el Teorema | | |
| 6) Una distribución muestral de un estadígrafo tiende asintóticamente a una normal | V | F |
| 7) En una distribución de medias el valor esperado es y su varianza es | | |
| 8) En el punto anterior: valor esperado y varianza no cambian si la población es finita. | V | F |
| 9) Cuando se realizan muestreos sin reposición hay que aplicar un factor de corrección. | V | F |
| 10) En una distribución muestral de proporciones se puede trabajar igual tamaño muestral | V | F |
| 11) Estas distribuciones valen si se toman todas las muestras posibles de tamaño N | V | F |
| 12) En una distribución de proporciones el valor esperado es y su varianza es | | |

- 13) Solo si las muestras son independientes se puede usar la distribución de diferencias. **V** **F**
14) Los casos vistos en este capítulo suponen la posibilidad de acceder a toda la población **V** **F**
15) Si los parámetros poblacionales son desconocidos no se puede aplicar lo visto más arriba. **V** **F**
16) El valor esperado y varianza de una diferencia de medias se expresa como:.....

2) En una medición de pesos se efectuaron 64 determinaciones de un objeto, resultado una media de 65 g con un desvío estándar de 4 g. Luego se llevó el objeto a una balanza calibrada en el INTI el cual informa un valor de referencia (que puede ser tomado como patrón) de 62,5 g. Determinar cuanto vale la probabilidad del valor tipificado de la media encontrada.

3) Si la Sensibilidad de una prueba clínica A es del 85% para un determinado punto de corte y la de otra prueba B para hacer el mismo análisis es del 90%. Se pide calcular el valor promedio y el desvío de la diferencia entre ambas, junto con su valor tipificado para $TE = TS = 100$.

4) El fabricante de cierto kit para medir glucosa entrega un estándar de 100 mg/dl con un error del 8%. Si un nuevo envase viene con otro estándar de 105 mg/dl y el mismo error, ¿cuánto vale el valor tipificado de la diferencia entre ambos envases si la diferencia de medias fue de 6 mg/dl?

5) Las 64 determinaciones de Glucosa de un paciente muestran una media de 88 g/dl con un desvío estándar de 5 mg/dl. El Bioquímico a cargo decide realizar 36 determinaciones más de la misma sangre pero usando un segundo kit cuya media fue de 90 y un desvío de 4 mg/dl. Tipificar esta diferencia.

6) En una farmacia se realiza una encuesta entre 100 clientes que concurren a la misma. El 75% de los encuestados se mostró favorable al uso de la marca A de jarabe para la tos. Luego de efectuada una propaganda intensa a través de los medios de comunicación, se vuelve a realizar la encuesta a otros 100 clientes y se obtiene una respuesta favorable en el 88% de los casos. Averiguar la probabilidad asociada a la diferencia de proporciones antes y después de la campaña.

7) En una balanza se pesaron 49 veces una cierta cantidad de droga resultando una media de 58 g con un desvío de 2 g. La misma droga se pesó 36 veces en otra balanza y se obtuvo un valor promedio de 55 g con un desvío de 4 g. Calcular la probabilidad asociada a la diferencia de medias entre las dos balanzas empleadas.

8) Suponiendo que las mediciones de glucosa se distribuyen normalmente entre los 5.000 varones sanos de una cierta población, con edades de 4 a 60 años, con un valor promedio de 9,5 y un desvío de 1,8. Un investigador decide tomar 90 muestras aleatorias de tamaño 4, con reemplazamiento. Desea saber en cuántas muestras cabe esperar una media entre 7 y 1,1.

9) Explicar como puede evitarse la influencia del factor humano en un muestreo basado en enfermedad, para la estimación de los índices clínicos.

10) Explicar cuales son los pasos a seguir en un estudio epidemiológico para evitar la influencia del investigador en estudios basados en exposición, en protección y en enfermedad.

11) Cuando se trabaja con proporciones y se puede aproximar con la función normal: ¿Cuál es la mejor corrección por continuidad que se puede usar, para que la aproximación sea lo mejor posible? ¿Cuanto debe valer el tamaño de muestra para poder usar aproximaciones normales?

11

Teoría de la inferencia estadística

En el capítulo anterior se trabajó con muestras tomadas de poblaciones donde sus parámetros se conocían, o bien, se podían tomar todas las muestras posibles de tamaño k . Esto se hizo con fines didácticos porque en la práctica es muy raro encontrarse en tal situación. Lo más común es no conocer la población y el único camino que queda es tomar muestras para poder estudiarla. En este capítulo se presenta la Teoría de Estimación o Inferencia Estadística, para mostrar cómo se pueden obtener estimadores de los parámetros desconocidos de una población cualquiera. Se muestran sus requisitos y también como calcular los llamados *intervalos de confianza*. Estos son intervalos dentro de los cuales, se tiene una cierta probabilidad de encontrar al verdadero valor poblacional de la magnitud biológica en estudio. La idea general es realizar muestreos de poblaciones para luego inferir los parámetros de la misma. No importa si se conoce a priori o cuál es la forma de la función distribución de probabilidades en la población. Si se toman muestras aleatorias e independientes, lo suficientemente grandes, los estadígrafos muestrales que se calculen tendrán una función distribución gaussiana (ver Tabla 4) que permitirá efectuar estimaciones de los respectivos parámetros poblacionales y las probabilidades asociadas a las mismas.

11.1 Introducción

Las razones para efectuar una estimación en una población, en lugar de estudiarla directamente, pueden ser: que el tamaño de la población sea infinito, que el muestreo sea destructivo, que la población sea finita pero demasiado grande, y otras razones como costo o tiempo. Por esto parece más práctico tomar muestras. Hay dos maneras básicas de hacer estimaciones:

Estimaciones puntuales: se estima el parámetro desconocido con un solo valor.

Estimaciones por intervalos: se estima el parámetro desconocido con un intervalo, el cual tiene asociado una cierta probabilidad de ocurrencia.

Siempre es más aconsejable estimar con intervalos en lugar de hacerlo con un solo valor puntual, por las mismas razones que tienen los granjeros para no poner todos los huevos en una misma canasta. En efecto, suponiendo que un paciente concurre a varios bioquímicos a efectuarse el mismo análisis y estos le informan, por ejemplo: Bioq. A = 85 mg/dl; Bioq. B = 75 mg/dl; Bioq. C = 90 mg/dl. El paciente podría llegar a desconfiar de estos valores por la diversidad de resultados obtenidos: a simple vista le parecerían diferentes. Por su parte, el médico que encargó el trabajo podría mandar a hacer de nuevo el análisis. Muy diferente hubiese sido la situación si

en sus informes hubieran usado intervalos como: Bioq. A: (85 ± 15) mg/dl; Bioq. B: (75 ± 4) mg/dl y el Bioq. C: (90 ± 20) mg/dl. En este caso hay coincidencia entre los tres informes, A dice entre 70 y 100, B dice entre 71 y 79 y C dice entre 70 y 110. El intervalo informado por B cae dentro de los otros dos y el de A cae dentro del intervalo informado por C. Esto es, hay una zona donde los tres informes coinciden, cosa imposible de ver si se hubieran usado estimadores puntuales. Por su parte, en la industria farmacéutica lo normal es trabajar con límites superiores e inferiores en las composiciones químicas de los medicamentos, tanto para el proceso productivo como para su posterior control de calidad. Otro tanto ocurre en la industria alimenticia o de cosméticos. La idea es siempre informar o establecer la probabilidad de que el verdadero valor caiga dentro del intervalo informado. La forma más *prudente* de informar es usando intervalos.

11.2 Estimaciones con intervalo

Las estimaciones por intervalo de un parámetro poblacional μ desconocido dan idea de la precisión y exactitud de la inferencia efectuada, junto con la probabilidad de que tal estimación sea cierta. Se calculan a través de los llamados: *intervalos de confianza*. Estos se construyen con la función probabilística del modelo estadístico adoptado para realizar la estimación. Se comienza con el modelo de Gauss empleado generalmente para las *grandes muestras*. En efecto, si la distribución muestral de un estadígrafo es la de Gauss de parámetros $(\mu_e; \sigma_e)$, porque la cantidad de muestras aleatorias e independientes tomadas es lo suficientemente grande, entonces con la Tabla 4 se pueden obtener las distribuciones de varios estadígrafos e y las probabilidades asociadas con la tabla de Gauss tipificada. Usando estas tablas se pueden hacer los dos tipos de estimaciones vistos.

Una *estimación puntual* se logra con: $\mu_e = \mu$ donde, por ejemplo, si e es la media aritmética, el valor μ_e es el promedio de las medias muestrales bajo la condición de tener una cantidad de muestras superior o por lo menos igual a 30. Este tipo de mediciones no es exacto sino que hay una cierta incertidumbre al hacerla, llamada *error típico de estimación*: $\sigma_e = SE(e)$, dado por el error estándar de la distribución muestral multiplicada por un cierto factor. Todo esto se muestra en la segunda columna de la Tabla 10.1. Para el caso de la media es: $SE(e) = \sigma_e = \sigma / \sqrt{n}$. Entonces, se puede construir la *estimación por intervalos* con:

$$\mu \in (\mu_e \pm Z_{\alpha} \cdot \sigma_e) \quad (a)$$

Otra manera de informar lo mismo es calculando los límites superior e inferior del intervalo con: $LS = \mu_e + Z_{\alpha} \cdot \sigma_e$. Y el inferior es $LI = \mu_e - Z_{\alpha} \cdot \sigma_e$. Entonces, la relación (a) se escribe como

$$\mu \in (LI ; LS) \quad (b)$$

Donde Z_{α} es un valor llamado *coeficiente de confianza*. La relación anterior (a) establece que el parámetro desconocido μ está contenido dentro de un intervalo dado por su mejor estimación puntual μ_e más o menos su error típico de estimación σ_e , multiplicado por un coeficiente relacionado con la confianza de que esto sea cierto. La relación (b) establece que el valor verdadero μ está contenido dentro de un intervalo conformado por dos límites, el inferior y el superior. Y la probabilidad de que μ se encuentre dentro de ese intervalo viene dada por el coeficiente de con-

fianza adoptado. Al valor α se lo denomina *nivel de significación* y por lo general es elegido por el investigador. Los usos y costumbres establecidos en Biología establecen tres límites habituales: 95%, 99%, y 99,9%. En la Física o Química, donde los materiales de trabajo son más estables que los biológicos, los requerimientos son menores, llegando a lo sumo al 99%. En casi todas partes el nivel usual de trabajo es del 95%. Estos valores de *Confianza* determinan el *nivel de significación* α , y el coeficiente de confianza Z_α que se obtiene de la curva de Gauss. En la Tabla 11.1 se muestran algunos coeficientes de uso frecuente. A veces, se necesitan intervalos diferentes, de tipo abierto, donde no se necesitan dos valores extremos sino solo uno. En tal caso:

$$\mu \in (-\infty; (\mu_e + Z_\alpha \cdot \sigma_e)] \quad \text{O bien,} \quad \mu \in [(\mu_e - Z_\alpha \cdot \sigma_e); +\infty)$$

Estos intervalos abiertos por izquierda o por derecha se denominan de *una cola*, mientras que los comunes para la estimación de parámetros poblacionales son de *dos colas*. En la Figura 11.1 se muestran las dos formas usuales de una y de dos colas, con los respectivos valores de z_α en las Tablas 11.1 y 11.2.

TABLA 11.1: Niveles de confianza más usados.

Niveles de significación α	Nivel de confianza NC = $(1-\alpha) \cdot 100$	Coficiente de confianza Z_α
0,0005	99,95%	3,29
0,0010	99,90%	3,09
0,0013	99,87%	3,00
0,0050	99,50%	2,58
0,0100	99,00%	2,33
0,0227	97,72%	2,00
0,0250	97,50%	1,96
0,0500	95,00%	1,645
0,1587	84,13%	1,00

Figura 11.1: Probabilidad normal presentada en una y dos colas.

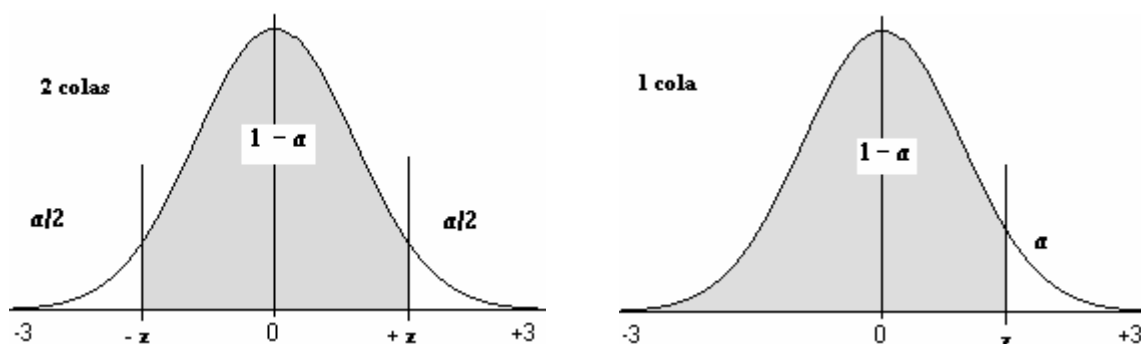


TABLA 11.2: Coeficientes de confianza más usados Z_α

	99,90%	99,00%	95,00%
1 cola	3,09	2,33	1,645
2 colas	3,29	2,58	1,96

En clínica la notación de los intervalos de confianza para un nivel de confianza $NC\%$ es:

$$CI \text{ NC } \% \left(\mu_e \pm Z_{\alpha} \cdot SE(e) \right)$$

Por ejemplo, para un 95% de confianza se expresa con:

$$CI \text{ 95\% } \left([\mu_e - 1,96 \cdot SE(e)] ; [\mu_e + 1,96 \cdot SE(e)] \right) \quad (c)$$

Esta es la notación recomendada para los trabajos en Medicina desde hace unos años para expresar los resultados de investigaciones en publicaciones científicas. Los tres tipos de notaciones expresados con las relaciones (a), (b) y (c) son equivalentes.

11.3 Intervalos para la media y proporciones.

Para el caso de la **media aritmética**, el intervalo se construye buscando en la Tabla 10.1 el error típico de estimación para la media y reemplazando en la ecuación (a) vista más arriba:

$$\mu \in \left(\bar{X} \pm Z_{\alpha} \sigma / \sqrt{n} \right)$$

El nivel de significación adoptado hará variar el ancho del intervalo de confianza. Como puede verse en la Tabla 14.1, a medida que aumenta el Nivel de Confianza aumenta el coeficiente de confianza Z_{α} mientras los demás términos no se alteran. Por su parte, cuanto mayor sea el intervalo, mayor será el error $\Delta x = Z_{\alpha} \sigma / \sqrt{n}$ (recordando la expresión general del error vista antes). Esto significa que se tendrá una menor precisión. Se puede formular ahora una regla general para todos estos casos: a *mayor confianza, menor precisión*. Notar que el supuesto básico para poder usar esta relación es que el valor σ es conocido. Como generalmente esto no es así, se puede usar el valor muestral DS para estimar σ si la muestra es lo suficientemente grande, y para este caso la condición aceptable es $n \geq 30$.

Conviene desarrollar ejemplos para ilustrar esta tesis, como sigue:

Ejemplo 1) Suponiendo que a un paciente se le extrae una muestra de sangre y al suero obtenido se lo fracciona en 50 alícuotas, luego a cada una se le determina la creatinina, y con los valores medidos se obtienen un promedio de 10 mg/dl y un desvío de 2,2 mg/dl. El verdadero valor de la creatinina en el paciente se puede estimar con un nivel de confianza del 95 % ($Z_{\alpha} = 1,96$) con:

$$\mu \in \left(10 \pm 1,96 \cdot 2,2 / \sqrt{50} \right) \text{ mg/dl} = (10,0 \pm 0,6) \text{ mg/dl} \quad \text{O bien: CI 95\% (9,4 ; 10,6)}$$

Eso significa que se tiene una probabilidad del 95 % de encontrar la creatinina real del paciente entre 9,4 y 10,6 mg/dl. Si se quiere aumentar la confianza al 99% el nuevo intervalo tendrá una mayor indeterminación, o sea, el intervalo será más ancho: entre 9,2 y 10,8 mg/dl.

$$\mu \in \left(10 \pm 2,58 \cdot 2,2 / \sqrt{50} \right) \text{ mg/dl} = (10,0 \pm 0,8) \text{ mg/dl} \quad \text{O bien: CI 99\% (9,2 ; 10,8)}$$

Y si todavía se aumenta un poco más al 99,9%:

$$\mu \in (10 \pm 3,29 \cdot 2,2 / \sqrt{50}) \text{ mg/dl} = (10 \pm 1) \text{ mg/dl} \quad \text{O bien: CI 99,9\%}(9 ; 11)$$

Ejemplo 2) Se toman 36 muestras de cierto antibiótico y se les mide su contenido de Sulfato de Neomicina, resultando un promedio de 92 mg con un desvío de 6 mg. Se desea saber el intervalo al 99,9% de la variable medida, tal que el contenido de Sulfato sea menor o igual al valor medio.

Se trata de un caso de intervalo de una sola cola. El 99,9% del área bajo la curva normal tipificada debe quedar a la izquierda de un valor crítico (como el dibujado en el Gráfico 14.1). O sea,

$$\mu \in (-\infty ; (\mu_e + z_{\alpha} \cdot \sigma_e)] = (-\infty ; (92 + 3,09 \cdot 6 / \sqrt{36})) \quad \text{O bien: CI 99,9\%} (-\infty ; 95,09)$$

Esto significa que hay un 99,9% de confianza de que el contenido de Sulfato de Neomicina del medicamento examinado nunca supere 95,09 mg.

Si se hubiese tratado de determinar el intervalo de estimación del verdadero contenido del Sulfato en el medicamento, entonces se debería calcular su intervalo de confianza con:

$$\mu \in (92 \pm 1,96 \cdot 6 / \sqrt{36}) \text{ mg/dl} = (92 \pm 2) \text{ mg/dl} \quad \text{O bien: CI 95\%} (90 ; 94,9)$$

$$\mu \in (92 \pm 2,58 \cdot 6 / \sqrt{36}) \text{ mg/dl} = (92,0 \pm 2,6) \text{ mg/dl} \quad \text{O bien: CI 99\%} (89,4 ; 94,6)$$

$$\mu \in (92 \pm 3,29 \cdot 6 / \sqrt{36}) \text{ mg/dl} = (92,0 \pm 3,3) \text{ mg/dl} \quad \text{O bien: CI 99,9\%} (88,7 ; 95,3)$$

Ejemplo 3) Se realizaron 200 exámenes de Bioestadística. Si se toman 50 de ellos al azar y se obtiene una media de 7,5 puntos con un desvío de 1 punto: ¿cuánto vale el intervalo de estimación para la media de los 200 exámenes con una confianza del 95%?

En este problema, la población no es grande en función de la muestra, entonces conviene efectuar la corrección para estos casos vista en el capítulo anterior:

$$\mu_{\bar{x}} = \mu = 7,5 \quad \text{y} \quad \sigma_{\bar{x}} = 0,123 \quad \text{pues:}$$

$$\sigma_{\bar{x}}^2 = (\sigma^2 / n) [(N_P - n) / (N_P - 1)] = [(1)^2 / 50] [(200 - 50) / (200 - 1)] = 0,0151$$

$$\text{Con esos datos se puede obtener: } \mu \in (7,5 \pm 1,96 \cdot 0,123) = (7,50 \pm 0,24)$$

$$\text{O bien: CI 95\%} (7,26 ; 7,74)$$

Para el caso de las **proporciones** el intervalo queda determinado por:

$$\mu \in [p \pm Z_{\alpha} \cdot \sqrt{\frac{\pi \cdot (1 - \pi)}{n}}]$$

Donde p es la proporción de éxitos obtenida en la muestra y π es la proporción de éxitos esperada o real en la población de referencia. Hay casos donde π es conocida como $\pi = 1/2$ en el lanzamiento de una moneda, $\pi = 1/6$ al lanzar un dado, $\pi \approx 1/2$ en el sexo de un recién nacido, etc. Pero

cuando no se sepa el verdadero valor de p , entonces hay que estimarlo con la proporción de éxitos en la muestra $p \approx \pi$ si y solo si, la muestra es lo suficientemente grande ($n > 40$)

Ejemplo 4) Se lanza una moneda al aire 100 veces y se obtienen 56 caras. El intervalo de confianza del 99 % para la probabilidad de sacar cara con tal moneda se calcula con:

$$p = 56/100 = 0,56 \text{ y } \pi = 0,5 \quad \text{El desvío es: } \sigma_{\pi} = \sqrt{\frac{\pi \cdot (1 - \pi)}{n}} = \sqrt{\frac{0,5 \cdot 0,5}{100}} = 0,05 \quad \text{O sea,}$$

$$\mu \in (0,56 \pm 2,58 \cdot 0,05) = (0,56 \pm 0,13) \quad \text{O bien: CI 99\% (0,43 ; 0,69)}$$

Expresado de otra forma: $\mu \in (56 \pm 13) \%$. Esto es, hay una probabilidad del 99 % de obtener entre 43 y 69 caras, si se arroja 100 veces esa moneda al aire.

Ejemplo 5) En un estudio por cohortes se tuvo 50 casos de enfermos en un total de 200 sujetos. Se esperaba una prevalencia del 20% en la enfermedad; hallar el intervalo de confianza:

$$p = 50/200 = 0,25 \text{ y } \pi = 0,2. \text{ El desvío es: } \sigma_{\pi} = \sqrt{\frac{\pi \cdot (1 - \pi)}{n}} = \sqrt{\frac{0,2 \cdot 0,8}{200}} = 0,03 \quad \text{O sea,}$$

$$\mu \in (0,25 \pm 1,96 \cdot 0,03) \text{ O bien: CI 95\% (0,19 ; 0,31)}$$

11.3.1 Intervalos “a priori” de media y proporciones.

Por intervalos “a priori” se entienden aquellos intervalos de confianza que se pueden obtener *antes* de efectuar el experimento. La idea es saber antes de medir en que intervalo se espera que caiga el resultado obtenido si todo está bien, o sea de acuerdo a las expectativas previas. Simplemente, se trata de rescribir el intervalo definido con las ecuaciones (a), (b) y (c) ya vistas:

$$\boxed{\mu_e \in (\mu \pm Z_{\alpha} \cdot \sigma_e)} \quad (d)$$

Por ejemplo para el caso de la media aritmética sería:

$$\bar{X} \in (\mu \pm Z_{\alpha} \sigma / \sqrt{n})$$

Y para las proporciones:

$$p \in \left[\pi \pm Z_{\alpha} \cdot \sqrt{\frac{\pi \cdot (1 - \pi)}{n}} \right]$$

El significado de estos intervalos es estimar dentro de que valores se espera que caiga el valor muestral, antes de realizar las mediciones del experimento. Notar que esto solo puede aplicarse cuando se conocen “a priori” los valores poblacionales, por ejemplo cuando se está usando un suero control, de referencia o estándar en una medición clínica (donde μ es el valor aceptado). O

bien cuando se conoce la probabilidad π de la población como en el caso de una moneda, el sexo de un recién nacido, las proporciones mendelianas, etc. Como ser:

Ejemplo 6) Suponiendo que se tiene un suero patrón cuyo valor de creatinina es $\mu = 12$ mg/dl y con una especificación del estándar de $\sigma = 1,22$ mg/dl. Si se fracciona dicho suero en 36 alícuotas y se mide a cada una de ellas con el sistema habitual de laboratorio, se desea saber en que rango de valores debe caer la media aritmética de las mediciones a realizar, si el método clínico está bien calibrado, con un nivel de confianza del 95 %.

$$\bar{X} \in (\mu \pm Z_{\alpha} \sigma / \sqrt{n}) = [12 \pm 1,96 \cdot (1,22 / 6)] = (12,0 \pm 0,4)$$

O bien, se espera que \bar{X} caiga dentro del intervalo (11,6 ; 12,4) si el método de medición está funcionando bien, con una confianza del 95%.

Ejemplo 7) En la determinación del SIDA se usa un método cuya especificación establece que tiene una Sensibilidad del 90% y una Especificidad del 80%. Si se van a efectuar 49 determinaciones, averiguar los valores reales que se esperan obtener al 95%.

$$\pi = 0,9 \text{ y el desvío es: } \sigma = \sqrt{\frac{\pi \cdot (1 - \pi)}{n}} = \sqrt{\frac{0,9 \cdot 0,1}{49}} = 0,043 \quad \text{O sea}$$

$$p \in (0,9 \pm 1,96 \cdot 0,043) \quad \text{O bien: Sensibilidad esperada es CI 95\% (0,98 ; 0,82)}$$

Mientras que la Especificidad esperada se puede calcular con:

$$\pi = 0,8 \text{ y el desvío es: } \sigma = \sqrt{\frac{\pi \cdot (1 - \pi)}{n}} = \sqrt{\frac{0,8 \cdot 0,2}{49}} = 0,057 \quad \text{O sea}$$

$$p \in (0,8 \pm 1,96 \cdot 0,057) \quad \text{O bien, la Especificidad estaría en CI 95\% (0,91 ; 0,69)}$$

11.4 Intervalos para el desvío estándar y la varianza.

Para el caso del **desvío estándar** se puede construir el intervalo de confianza recurriendo a la Tabla 4 para obtener el error típico de estimación y así:

$$\mu \in (DS \pm Z_{\alpha} \cdot \frac{\sigma}{\sqrt{2n}})$$

Y para la **varianza** es:

$$\mu \in (DS^2 \pm Z_{\alpha} \cdot \sigma^2 \sqrt{\frac{2}{n}})$$

En ambos casos se usa la aproximación $\sigma \approx DS$ para resolver los problemas prácticos, siempre y cuando la muestra sea lo suficientemente grande ($n \geq 100$). Para ilustrar su aplicación se tomarán los datos de problemas anteriores:

Ejemplo 8) Usando los datos de los ejemplos 1 y 2, pero para $n = 144$ sería:

$$(1) \sigma \in (2,2 \pm 1,96 \cdot 2,2 / \sqrt{2 \cdot 144}) \text{ mg/dl} = (2,20 \pm 0,25) \text{ mg/dl. O bien CI 95\% (2,45; 1,95)}$$

$$\sigma^2 \in (4,84 \pm 1,96 \cdot 4,84 \sqrt{\frac{2}{144}}) \text{ mg/dl} = (4,84 \pm 0,13) \text{ mg/dl. O bien CI 95\% (4,71 ; 4,97)}$$

$$(2) \sigma \in (6,0 \pm 1,96 \cdot 6 / \sqrt{2 \cdot 144}) \text{ mg/dl} = (6,0 \pm 0,7) \text{ mg/dl. O bien CI 95\% (5,3 ; 6,7)}$$

11.5 Propiedades de un estimador

Los conceptos básicos desarrollados en esta sección son para el caso de estimaciones paramétricas, pero se pueden extender también al caso no-paramétrico. Sea una población cualquiera de la cual se extraen n muestras aleatorias e independientes; los valores obtenidos de cada muestra conforman un conjunto que tendrá una cierta distribución muestral que se usará para estimar el parámetro desconocido: μ . Se desea que el estimador e sea bueno. El problema es que se entiende por buen estimador y la respuesta es: *La distribución muestral de e debe concentrarse lo más posible alrededor del valor verdadero del parámetro μ y su dispersión debe ser lo más pequeña posible.*

La técnica más importante para hallar estimadores se debe a R.A. Fisher y es el *Método de Máxima Verosimilitud (Likelihood)*. Consiste en hallar una *función de verosimilitud* para cada caso y así elegir, como estimador de valor poblacional desconocido, al valor que hace máxima a esta función. Su desarrollo matemático excede los límites del presente trabajo quien desee verlo con más detalle puede recurrir a texto, más específicos como los de H. Crámer, etc. Por ahora, basta mencionar que el *estimador máximo verosímil* de un índice de tendencia central es la *media aritmética* y de un índice de dispersión es el *desvío estándar*.

Las propiedades clásicas de los estimadores son:

Estimadores insesgados: un estimador se dice insesgado cuando su valor es igual al correspondiente valor poblacional. Como la media o la mediana para la media poblacional. Caso contrario, se lo llama *estimador sesgado* como la media de una distribución muestral de varianzas.

Cuando el valor esperado del estadígrafo $\Xi(e) = \mu$ el estimador es insesgado.

Por ejemplo:

$$\Xi(\bar{X}) = \mu_{\bar{X}} = \mu$$

$$\Xi(\text{Med}) = \mu_{\text{Med}} = \mu$$

Sin embargo, $\Xi(\sigma^2) = \Xi[\sum(x_i - \mu)^2 / n] = \mu_{\sigma^2} = (n-1) / n \cdot \sigma^2$ (poblacional)

Es un estimador sesgado pues $(n-1) / n$ siempre diferente de la unidad y eso sesga la estimación. Sin embargo, el sesgo se elimina cuando para el cálculo de la varianza se usa la fórmula

$$\Xi(DS^2) = \Xi[\sum(x_i - \mu)^2 / (n-1)] = \mu_{\sigma^2} = \sigma^2 \text{ (muestral)}$$

Estimador eficiente: si las distribuciones muestrales de dos estadígrafos tienen la misma esperanza (o media), aquel que tenga menor error de estimación (varianza) será el más eficiente de ambos. En otro caso, se dice que el *estimador es no eficiente*.

Por ejemplo, las distribuciones muestrales de la media aritmética y de la mediana tienen el mismo valor esperado, que es la media poblacional. Sin embargo, la varianza de la distribución muestral de medias es menor que la de la mediana. Por lo tanto, el estimador eficiente es la media aritmética. En efecto:

$$SE(\bar{x}) = \sigma_{\bar{x}} = \sigma / \sqrt{n} \quad \text{mientras que} \quad SE(\text{Mediana}) = (1,2536) \sigma / \sqrt{n} \quad \text{O sea} \quad \sigma_{\bar{x}} < \sigma_{\text{Med}}$$

Existen otros métodos para la obtención de los estimadores como el de los momentos desarrollados por K. Pearson, el de Ráo-Cramer o el de suficiencia para funciones de densidad conjunta de probabilidad. Sin embargo, para los alcances de esta obra basta saber que los estimadores máximos verosímiles cumplen a la vez con las propiedades de ser insesgados, eficientes y suficientes. En resumen, para efectuar estimaciones apropiadas, cuando se pueda se deben usar las estimaciones por intervalo antes que las puntuales y se deben emplear los estimadores máximos verosímiles para construirlos. Los niveles de confianza para el caso de variables biológicas se adoptan por costumbre en 95 %, 99 % y 99,9 %, y se calculan siempre los tres para poder informar en las publicaciones científicas, como se podrá ver mejor en el próximo capítulo. En Medicina, lo habitual es usar el nivel del 95% para expresar los resultados, lo que se hace extensivo a sus disciplinas derivadas como: Epidemiología, Inmunología, y otras.

11.6 Intervalos de confianza para dos muestras

Sean dos estadígrafos e_1 y e_2 , extraídos de poblaciones normales o con función distribución de tipo gaussiana, el intervalo de confianza para la diferencia de sus respectivos parámetros poblacionales se obtiene con:

$$\mu_{e_1 - e_2} = (e_1 - e_2) \pm Z_{\alpha} \cdot \sigma_{1,2} = (e_1 - e_2) \pm Z_{\alpha} \cdot \sqrt{\sigma_1^2 + \sigma_2^2}$$

Donde el $SE(e_1 - e_2) = \sigma_{1,2}$. Los dos casos más interesantes son la media y las proporciones. Se reemplazan los valores de los estadígrafos y los de sus respectivos errores típicos (ver Tabla 4) en la fórmula anterior y se puede obtener para la media aritmética:

$$\mu_{\bar{x}_1 - \bar{x}_2} = (\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Cuando ambas muestras tienen la misma varianza $\sigma^2 = \sigma_1^2 = \sigma_2^2$ es:

$$\mu_{\bar{x}_1 - \bar{x}_2} = (\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha} \cdot \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Y si además ambas muestras tienen el mismo tamaño muestral $n = n_1 = n_2$ es:

$$\mu_{\bar{x}_1 - \bar{x}_2} = (\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha} \cdot \sigma \sqrt{\frac{2}{n}}$$

Por su parte para la diferencia de proporciones es:

$$\mu_{p_1 - p_2} = (p_1 - p_2) \pm Z_{\alpha} \cdot \sqrt{\frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}}$$

Cuando ambas muestras fueron extraídas de la misma población, o cuando hay razones para creer que los valores poblacionales son iguales, esto es $\pi = \pi_1 = \pi_2$ entonces la ecuación se reduce a:

$$\mu_{p_1 - p_2} = 0 = (p_1 - p_2) \pm Z_{\alpha} \cdot \sqrt{\pi(1 - \pi) \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}$$

Cuando el valor poblacional π es desconocido se lo puede estimar con la aproximación:

$\pi \approx (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$ que es el promedio ponderado de las proporciones muestrales

Para ilustrar estas aplicaciones se presentan los ejemplos siguientes:

Ejemplo 1) Se tomaron 200 muestras aleatorias de presión sistólica a niños cuyos padres son hipertensos, obteniéndose una media de 107 y un desvío de 7. Luego se tomaron 100 muestras de niños cuyos padres tienen la presión sanguínea normal, y se obtuvo una media de 98 con un desvío de 6. Obtener los límites de confianza del 95 % a la diferencia de medias.

En este caso se trata de una diferencia de medias, pero con varianzas diferentes estimadas con las muestras de la manera siguiente:

$$\mu_{\bar{x}_1 - \bar{x}_2} = (\bar{x}_1 - \bar{x}_2) = 107 - 98 = 9 \text{ y } SE(1-2) = \sigma_{1-2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{49}{200} + \frac{36}{100}} = 0,778$$

$$\mu_{1-2} \in (9 \pm 1,96 \cdot 0,778) = (9,0 \pm 1,5). \text{ O bien CI } 95\% (7,5 ; 10,5)$$

$$\mu_{1-2} \in (9 \pm 2,58 \cdot 0,778) = (9,0 \pm 2,0). \text{ O bien CI } 99\% (7 ; 11)$$

$$\mu_{1-2} \in (9 \pm 3,29 \cdot 0,778) = (9,0 \pm 2,6). \text{ O bien CI } 99,9\% (6,4 ; 11,6)$$

Ejemplo 2) Se inoculan dos organismos aislados durante dos epidemias distintas, a dos muestras diferentes obtenidas de la misma población. A las dos semanas se enferman el 68,5 % de las 200 pruebas realizadas con el primer organismo, y el 65,3 % de las 150 pruebas efectuadas para el segundo caso. Hallar un intervalo de confianza de esta diferencia de proporciones encontradas.

En este caso se trata de una diferencia de proporciones y se asume que el valor verdadero de cada población se estima con los valores muestrales hallados: $\mu_{\pi_1 - \pi_2} = p_1 - p_2 = (68,5 - 65,3)\% = 3,2\%$

$$\sigma_{\pi_1 - \pi_2} = \sqrt{\frac{\pi_1(1 - \pi_1)}{n_1} + \frac{\pi_2(1 - \pi_2)}{n_2}} = \sqrt{\frac{0,685 \cdot 0,315}{200} + \frac{0,653 \cdot 0,347}{150}} = 0,051 = 5,1\%$$

Entonces es: $\mu_{1-2} \in (\mu_{\pi_1-\pi_2} \pm Z_{\alpha} \cdot \sigma_{\pi_1-\pi_2}) = (3,2 \pm 1,96 \cdot 5,1) \%$ O bien CI 95% (-6,8 ; 13,2). Y esto significa que hay un 95% de probabilidades de que el verdadero valor de la diferencia de los porcentajes esté entre -6,8% y 13,2%.

Ejemplo 3) Imaginado para el problema anterior que el supuesto básico es que no hay diferencia entre las dos poblaciones, calcular el intervalo de confianza para el valor verdadero poblacional π

Con este supuesto será $\pi = \pi_1 = \pi_2 = (n_1 p_1 + n_2 p_2) / (n_1 + n_2) = 67,13\%$. Y con esta estimación se puede calcular el desvío con:

$$\sigma \approx DS = \sqrt{\pi (1 - \pi) \left[\frac{1}{n_1} + \frac{1}{n_2} \right]} = 4,35\%$$

Entonces la estimación será $\pi \in (67,13 \pm 1,96 \cdot 4,35)$ O bien: CI 95% (58,60 ; 75,66) %

11.7 Intervalo para el cociente de dos proporciones

El intervalo de confianza para el cociente de dos proporciones se logra haciendo una transformación de variables. Sea el cociente $\Phi = \pi_1 / \pi_2$, entonces tomando el logaritmo neperiano de esa variable ($\ln \Phi$), se deduce que tiene una distribución de probabilidades aproximadamente normal si la muestra es lo suficientemente grande. Entonces, tomado el antilogaritmo se pueden obtener los límites del intervalo de confianza, con la ecuación:

$$\Phi e^{[\pm 1,96 SE (\ln \Phi)]}$$

Donde el valor esperado es $\Xi (\Phi) = \mu_{\Phi} \cong \Phi$ se transforma en: $\Xi (\ln \Phi) \cong \ln \Phi$ y el error estándar en: $SE (\ln \Phi)$. De esta forma usando los antilogaritmos es que los límites se pueden calcular tomado el signo positivo para el superior y el negativo para el inferior, como se muestra en los siguientes ejemplos clínicos:

- *Likelihood Ratio de positivos:*

Este índice clínico se calcula con: $LR+ = S / (1 - E) = (vp / TE) / [1 - (vn / TS)]$

O lo que es lo mismo:

$LR+ = (vp / TE) / (fp / TS)$ que es un cociente de dos proporciones.

Donde el error estándar se calcula con:

$$SE [\ln (LR+)] = \sqrt{(1/vp) - (1/TE) + (1/fp) - (1/TS)}$$

Límite superior = $(LR+) e^{[+ 1,96 SE (\ln LR+)]}$

Límite inferior = $(LR+) e^{[- 1,96 SE (\ln LR+)]}$

- *Likelihood Ratio de negativos:*

Este índice clínico se calcula con: $LR^- = (1 - S) / E = [1 - (vp / TE)] / (vn / TS)$

O lo que es lo mismo:

$LR^- = (fn / TE) / (vp / TS)$ que es un cociente de dos proporciones.

Donde el error estándar se calcula con:

$$SE [\ln (LR^-)] = \sqrt{(1/fn) - (1/TE) + (1/vp) - (1/TS)}$$

$$\text{Límite superior} = (LR^-) e^{[+ 1,96 SE (\ln LR^-)]}$$

$$\text{Límite inferior} = (LR^-) e^{[- 1,96 SE (\ln LR^-)]}$$

- *Riesgo Relativo:*

Este índice clínico se calcula con: $RR = EER / CER$

O lo que es lo mismo:

$RR = [a / (a + b)] / [c / (c + d)]$ que es un cociente de dos proporciones.

Notar que en el caso del diagnóstico $a = vp$, $b = fp$, $c = fn$ y $d = vn$. Y el error estándar será:

$$SE [\ln (RR)] = \sqrt{(1/a) - (1/(a + b)) + (1/c) - (1/(c + d))}$$

$$\text{Límite superior} = (RR) e^{[+ 1,96 SE (\ln RR)]}$$

$$\text{Límite inferior} = (RR) e^{[- 1,96 SE (\ln RR)]}$$

- *Odds Ratio:*

$$OR = (a / c) / (b / d) = (a d) / (c b)$$

Es un cociente de dos proporciones pero hay una diferencia, no se trata del cociente de una frecuencia con respecto a un total marginal, sino respecto de otra frecuencia. Por eso el cálculo del error estándar es ligeramente diferente:

$$SE [\ln (OR)] = \sqrt{(1/a) + (1/b) + (1/c) + (1/d)}$$

$$\text{Límite superior} = (OR) e^{[+ 1,96 SE (\ln OR)]}$$

$$\text{Límite inferior} = (OR) e^{[- 1,96 SE (\ln OR)]}$$

Usando el ejemplo numérico del capítulo 4 (Pág. 23), se pueden calcular: $OR = 7,5$ y $OR = 5,6$.

Los intervalos al 95% de confianza resultan:

Para OR = 7,5 será 95% CI (3,7 ; 15,4)

Para RR = 5,6 será 95% CI (2,9 ; 10,7)

Usando otro ejemplo numérico del capítulo 4 (Pág. 9) para el caso de 1000 muestras es:

Para LR+ = 3,6 será 95% CI (3,2 ; 4,1)

Para LR- = 0,13 será 95% CI (0,07 ; 0,24)

Notar que estos intervalos no son simétricos. Los índices no caen en la mitad del intervalo.

11.8 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|---|-------|-------|
| 1) Una estimación de tipo puntual es más recomendable que una por intervalos. | V | F |
| 2) Un intervalo de confianza se arma con el estimador y su error típico. | V | F |
| 3) La fórmula para obtener un intervalo de confianza es: | | |
| 4) El mejor estimador es el máximo verosímil. | V | F |
| 5) La fórmula general para la diferencia de los estadígrafos de dos muestras es: . | | |
| 6) Un estimador es insesgado cuando su valor esperado coincide con el valor poblacional | V | F |
| 7) Los errores típicos de estimación se pueden obtener de | | |
| 8) Un estimador es más eficiente que otro cuando su error típico es mayor. | V | F |
| 9) Un estimador máximo verosímil es insesgado y eficiente a la vez. | V | F |
| 10) La probabilidad asociada a un intervalo de confianza se relaciona con la significación. | V | F |
| 11) El valor coeficiente de confianza se calcula con α y la tabla de Gauss. | V | F |
| 12) El intervalo de confianza para la diferencia de medias es:..... | | |
| 13) El intervalo de confianza para la diferencia de proporciones es: | | |
| 14) A mayor nivel de confianza, mayor ancho del intervalo. | V | F |
| 15) La relación entre el Nivel de significación y el Nivel de confianza es: | | |
| 16) El intervalo de confianza para el cociente de dos proporciones es simétrico. | V | F |
| 17) El error estándar del OR es muy similar al del RR. | V | F |

2) La vida útil de un medicamento A se midió con 200 muestras y dio una media de 1400 días con un desvío de 120 días. El medicamento B se determinó con 100 muestras y se obtuvo una media de 1200 días con un desvío de 80 días. Hallar el intervalo de confianza para la diferencia de vidas útiles de ambos medicamentos, al nivel del 95%, 99% y 99,9%.

3) El colesterol de una paciente medido 30 veces dio una media de 256 mg/dl y un desvío de 32 mg/dl. Encontrar el intervalo de confianza para el 95%.

4) A la muestra de suero de la paciente del problema anterior se la divide en 40 alícuotas y se le mide el colesterol con un nuevo método clínico. Su media es de 280 mg/dl y su desvío de 40 mg/dl. Encontrar el intervalo de confianza para el 99%.

5) Calcular la diferencia de medias entre el método clínico y el nuevo, usando los datos de los problemas 3 y 4 anteriores. Fijarse si el valor cero cae dentro del intervalo. Explicar los resultados obtenidos desde el punto de vista de la exactitud.

6) Calcular la diferencia de desvíos estándar encontrados en los problemas 3 y 4, expresada mediante un intervalo de confianza. Explicar los resultados desde el punto de vista de la precisión en mediciones clínicas.

7) Para hacer un test VDRL se emplea el kit A en 200 pacientes y se encuentran 120 pacientes que dieron un resultado (-). Los aciertos en (+) fueron 70 y 110 en (-). Con estos datos calcular la Sensibilidad y Especificidad de A y B. Si con los mismos pacientes se empleó el Kit B y resultaron 140 pacientes con (-), mientras el número de aciertos en (+) fue de 55 y en (-) 130. Obtener los intervalos de confianza al 95% para cada uno de los dos kits usados.

8) Se midieron 50 alícuotas de un suero con el Kit de Glucosa marca 1, y otros 50 con la marca 2. Los resultados fueron:

Marca 1: entre (80-80,9) 8 veces, (81-81,9) 22 veces, (82-82,9) 15 veces y 5 entre (83-83,9);

Marca 2: entre (80-80,9) 6 veces, (81-81,9) 30 veces, (82-82,9) 10 veces y 4 entre (83-83,9).

- Obtener los intervalos de confianza para las medias de cada método y para los desvío estándar, a un nivel del 95% y del 99%.
- Si el suero era uno de control con un valor de 79 mg/dl, determinar si cae dentro o fuera de los intervalos de confianza calculados en el punto anterior.
- Si la precisión máxima admisible para una glucosa es del 8% de acuerdo con los criterios de Aspen. Comparar este valor límite contra los hallados en cada marca.

9) Se sabe que la concentración de coloides en el agua de un río es del 5,4 /ml. Se tomó una muestra del mismo a la que se le agregó un reactivo. Luego de un rato se extrajo una pequeña cantidad para analizarla en una cámara de recuento. Los resultados fueron:

Nº de coloides: 2 3 4 5 6 7 (nº/ml)
 Frecuencia: 10 20 170 100 90 10

Calcular los intervalos de confianza para el 95% y el 99%.

10) En un recuento de rojos en cámara se espera obtener 4,5 millones por cm³ de cierto paciente. Se le toma una muestra de sangre y los resultados hallados son:

Nº de rojos por millón 2 3 4 5 6
 Frecuencia 10 90 110 150 40

Detectar si el valor esperado del recuento cae dentro o fuera del intervalo de confianza para el 95% de confianza. Discutir los resultados.

11) Calcular los intervalos de confianza al 95% de todos los índices clínicos de calidad

Test	Resultado verificado		Total
	Enfermo	Sano	
(+)	200	40	TP = 240
(-)	10	550	TN = 560
Total	TE = 210	TS = 590	800

Lo mismo para los índices de riesgo diagnósticos (RR y OR).

12

Teoría de la decisión estadística

En este capítulo se presenta la Teoría de la Decisión Estadística como herramienta básica para la toma de decisiones, basadas en evidencia científica. La manera de hacerlo es plantear las hipótesis posibles y luego efectuarle una prueba o *test estadístico*. Llamada en algunas obras: la docimasia estadística. Cuando una conclusión se valida con un test estadístico se la llama de tipo *cuantitativo*, en caso contrario la decisión adoptada es de tipo *cualitativo*, o sea, una decisión tomada en forma subjetiva. El método consiste en definir una probabilidad de aceptación del orden del 95% (o rechazo) de una hipótesis de trabajo planteada, que permite calcular los valores críticos (o límites de aceptación) de un estadígrafo calculado a partir de los valores medidos. La importancia de este tema es muy grande. Basta decir que el objeto final de la Estadística es la toma de decisiones. En este capítulo se presenta el modelo de las grandes muestras usando la función de Gauss, no solo por razones didácticas sino por el hecho de que muchos modelos que se verán más adelante se pueden aproximar con este cuando las muestras sean lo suficientemente grandes. La manera de trabajar y razonar es la misma para todos los otros modelos estadísticos que se irán mostrando a lo largo de este trabajo. Esta herramienta permite la toma de decisiones en muchos casos prácticos, como ver si un método clínico está calibrado, testear precisión y exactitud, plantear métodos de Control de Calidad, comparar entre sí dos técnicas de laboratorio, o entre dos medicamentos, y otros más.

12.1 Hipótesis estadísticas

Las conjeturas que se realizan sobre los parámetros desconocidos de las poblaciones tienen asociadas una cierta probabilidad de ocurrencia. Conviene hacer planteos o supuestos que luego puedan ser verificados en el ámbito probabilístico. La idea es que estas hipótesis pueden ser aceptadas o rechazadas con un *test de validación*. Pero se debe remarcar que nunca se tendrá la certeza acerca de sus resultados. Cuando un test acepta o rechaza una hipótesis estadística, lo hace con el nivel de confianza deseado por el investigador, tan alto como guste, pero nunca del 100%, o sea, de la certeza.

El primer paso es formular hipótesis estadísticas. Esto es, plantear los supuestos realizados en términos matemáticos cubriendo todos los casos posibles. Por ejemplo, se desea testear si la vida útil de un remedio en condiciones normales de almacenamiento es por lo menos de un año. Se pueden plantear dos alternativas: a) la vida útil es de un año o más, lo cual escrito matemáticamente se expresa con: $T \geq 365$ días; y b) todos los demás casos posibles, es decir: la vida útil es menor a un año, descrito con: $T < 365$ días. Otro ejemplo: se quiere ver si no hay diferencias entre un nuevo método de medición clínico y el viejo que se venía usando hasta ahora.. Las

dos hipótesis serían (a) $\mu_{\text{viejo}} = \mu_{\text{nuevo}}$ y (b) $\mu_{\text{viejo}} \neq \mu_{\text{nuevo}}$. O sea, en la primera hipótesis, se suponen equivalentes, mientras que en la segunda se los toma como diferentes. Se denominan:

Hipótesis nula H_0 : es la hipótesis estadística planteada para ser testeada;

Hipótesis Alternativa H_1 : es la hipótesis complementaria de la anterior.

Esto significa que la unión de ambas hipótesis cubre el universo de casos posibles. Como se vio en los ejemplos anteriores, siempre se plantean dos alternativas que son complementarias. Se puede elegir a cualquiera de ellas para ser la H_0 . Conviene usar a la que tiene el signo igual dentro de sus alternativas, pues facilita el cálculo de estadígrafo de prueba. De todos modos es lo mismo, pues muchas veces se plantea una hipótesis para ser rechazada antes que aceptada, y viceversa. Por ejemplo, si se quiere testear un dado para ver si está bien, se plantea: $H_0: p = 1/6$ y $H_1: p \neq 1/6$, donde la hipótesis es que la probabilidad de ocurrencia de cualquiera de sus caras es de $1/6$. Si es una moneda lo que se quiere ensayar, el planteo es similar con una probabilidad de $1/2$. En las comparaciones de dos técnicas clínicas lo que interesa es probar que ambas miden igual el valor de la magnitud en el paciente y por ello se emplea la media aritmética. Sería equivalente a pensar que ambas tienen la misma exactitud. Pero también se puede verificar si ambos métodos tienen similar dispersión usando la varianza como estadígrafo de prueba. En estos casos de comparaciones de dos estadígrafos conviene plantear como H_0 la que contiene la igualdad. El motivo para hacer esto es poder simplificar los valores poblacionales desconocidos, en el cálculo de la tipificada de las diferencias. Así, el valor: $\mu_{\text{viejo}} - \mu_{\text{nuevo}} = 0$ y se evita el tener que usar un suero control (con μ conocido) para hacer las comparaciones.

12.2 Validaciones estadísticas

Las pruebas que se realizan para plantear las hipótesis se conocen con el nombre de *ensayos de validación estadística*. El problema básico es determinar si las diferencias observadas entre el valor obtenido y el valor esperado se deben al azar, o si realmente son diferentes. Por ejemplo, para testear una moneda se arroja al aire 100 veces y se obtienen 55 caras. Entonces, el valor de frecuencia relativa obtenido es de 0,55, mientras que el valor teórico esperado es de 0,5. El problema es poder determinar, en *forma científica*, si esa diferencia observada de 0,05 es porque la moneda está algo cargada o simplemente por casualidad. Notar que si se toma esta decisión en forma subjetiva (*cualitativa*), diferentes personas podrían opinar de manera diferente. Esto es, para algunos la diferencia es tan chica que no vale la pena tenerla en cuenta, pero otros pueden opinar que es demasiado grande para olvidarla, y finalmente un tercer grupo puede creer que los datos son insuficientes para poder tomar una decisión. Este tipo de divergencia de opinión generó un sin número de discusiones estériles entre los científicos del pasado, donde unos apoyaban fervientemente una teoría, mientras los de la escuela contraria se encontraban en franca oposición. Se necesitó un árbitro creíble para ambas partes. Uno que tomase las decisiones en forma *objetiva* (*cuantitativa*), sin ser influenciado por nada ni por nadie y creíble para todos. El método de la prueba estadística surgió entonces como el único aceptable, el método *científico* para decidir. Tanto es así que hoy ningún trabajo de investigación es aceptado sino se realizó una validación estadística de sus conclusiones. Trabajos no validados no se aceptan más en los congresos, reuniones, coloquios o seminarios de discusión. Al poder disponer de una herramienta tan poderosa para probar hipótesis, el trabajo de investigación se aceleró en forma notable. Además, ahora se

dispone de técnicas para el diseño de experimentos, como se podrá ver más adelante en el capítulo del ANOVA (Análisis de la varianza).

Como se vio en el Gráfico 4.1 hay dos maneras de acertar al tomar una decisión, y también hay dos maneras de equivocarse. En estadística se denominan:

Error de tipo I: ocurre cuando se rechaza una hipótesis, que debería haber sido aceptada.

Error de tipo II: ocurre cuando se acepta una hipótesis, que debería haber sido rechazada.

En el comercio, al error de tipo I se lo denomina *riesgo del vendedor*. Ocurre cuando es devuelta una mercadería por no pasar el control de calidad, y en realidad era buena. Por ejemplo, en un lote de 2000 artículos se revisan 5 de ellos para decidir si se lo acepta, y justo se eligen para hacer el control los únicos fallados; entonces se le devuelven al vendedor quien tiene que correr con todos los gastos que demanda una reposición de mercadería, si no es que pierde la venta injustamente. A veces, la mercadería viene de otro continente y los costos son muy altos sin contar con la pérdida de imagen.

Por otro parte está el *riesgo del comprador*, que puede comprar algo de mala calidad por haber revisado una parte pequeña del lote que no puso en evidencia tal cosa. Es el caso de un error estadístico de tipo II. El problema de una compraventa justa es minimizar el riesgo de ambas partes. Para eso, se emplean las Normas IRAM respectivas donde se especifican el riesgo de cada uno, el tamaño del lote de prueba y el número aceptable de piezas falladas. Los contratos de compraventa detallan claramente todas esas condiciones negociadas por ambas partes.

Lo ideal sería minimizar ambos tipos de errores a la vez, lo cual puede hacerse aumentando el número de pruebas. Sin embargo, en la práctica, para un tamaño muestral dado, cuando se achica uno, inevitablemente crece el otro. A veces, un tipo de error puede tener más importancia que el otro, como el caso visto de sensibilidad y especificidad de los índices clínicos donde según la enfermedad conviene minimizar uno u otro. Otras veces, no es posible aumentar el número de pruebas, ya sea por su costo, por falta de material o por el tiempo insumido. Pero la regla de oro de los estadísticos es:

Para minimizar los errores, hay que aumentar el número de pruebas.

12.3 Test de hipótesis

El primer paso para contrastar una hipótesis estadística cualquiera es fijar el *nivel de significación* (α) de la prueba, de acuerdo con el nivel de confianza deseado. En Física, Química y otras ciencias exactas, se suele adoptar un valor crítico $Z_{\alpha} = 1$, considerado suficiente. Sin embargo, en otras disciplinas como Biología, Medicina, etc., que trabajan con materiales menos estables, los niveles de confianza usuales son del 95 %, 99 % y del 99,9 %. En efecto, un trozo de metal tiene menos variabilidad durante el proceso de medición que un material biológico. Cuando se trabaja con seres vivos las magnitudes clínicas como glucosa, colesterol, bilirrubina y otras,

varían casi en forma continua dentro de ciertos límites. La ingesta de medicamentos, el ejercicio, la posición adoptada por el individuo y el tiempo de ayuno, modifican los valores de dichas magnitudes. Esto obliga a los investigadores a tomar límites de tolerancia más amplios.

Cuando se rechaza una hipótesis usando un nivel de significación $\alpha = 0,05$, para un nivel de confianza del 95 %, se dice en el informe que los resultados fueron *significativos*. Esto se denota de varias formas. Algunos autores ponen el estadígrafo hallado y el valor de la significación, otros el valor de la probabilidad correspondiente al estadígrafo medido; otra es usar un asterisco para ello. En Medicina, la recomendación actual es poner todo, en especial el intervalo de confianza (CI), junto con el valor del estadígrafo. Son ejemplos de estos modos:

- ... se obtuvieron resultados significativos ($p < 0,05$)
- ... esta hipótesis no es aceptable $Z_x = 2,07$ ($p = 0,0468$)
- ... se rechazó la hipótesis planteada ($Z = 2,07$) al 95% de confianza.
- ... se rechazó la hipótesis planteada ($Z = 2,07^*$)
- ... se rechaza pues $Z = 2,07$ cae fuera de 95% CI (-1,6 ; 1,6)

Son cinco maneras diferentes de decir lo mismo; sin embargo, la última presenta mayor información al lector. Al poner el estadígrafo Z se le está diciendo que se ha empleado la función de Gauss (o la teoría de las grandes muestras) para realizar la validación, cosa que no es clara en la primera al usar la probabilidad p . Además 95% CI dice que se trata de un intervalo al 95% y da los límites superior e inferior del mismo. En síntesis, la notación más recomendable es la última pues es la que mejor resume todos los datos hallados.

Cuando se rechaza la hipótesis al nivel del 99%, se dice que los resultados fueron *muy significativos* y se denota con dos asteriscos en lugar de usar uno solo o con 99% CI. Por ejemplo, $Z = 2,7^{**}$ o bien $p < 0,01$, también $p = 0,003$, o $Z = 2,7$ cae afuera de 99% CI (-2,58 ; 2,58).

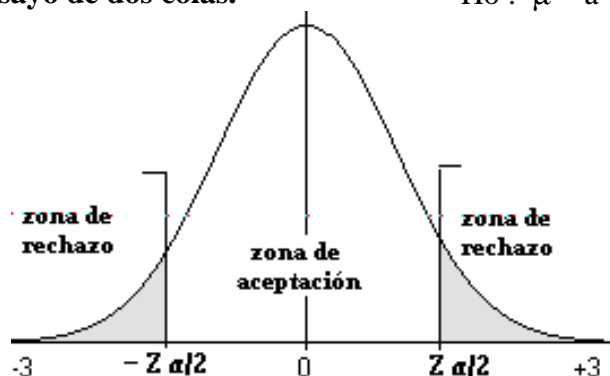
Si se rechaza la hipótesis al nivel del 99,9%, entonces se habla de resultados *altamente significativos* y esto se denota empleando tres asteriscos, o bien con la notación de intervalo 99,9% CI. Como ser, $Z = 3,8^{***}$, o puede ser con $Z = 3,8$ ($p < 0,001$), y también $p = 0,001$, o $Z = 3,8$ cae afuera de 99,9% CI (-3,29 ; 3,29).

En cambio, cuando la hipótesis nula no puede ser rechazada se dice que los resultados *no fueron significativos* y no se coloca ningún asterisco. Por ejemplo, $Z = 1,2$ o $p > 0,05$, o también $Z = 1,2$ cae adentro de 95% CI (-1,96 ; 1,96)

El siguiente paso, una vez definido el nivel de confianza, es determinar las *zonas de aceptación y de rechazo* de la hipótesis planteada. Primero se define el tipo de ensayo a usar. Hay dos tipos: ensayo de *una* o de *dos* colas, de acuerdo a si en la hipótesis nula hay una desigualdad o una igualdad respectivamente. Luego se calcula el valor crítico del ensayo que delimita las zonas buscadas. Cuando la hipótesis nula sea de igualdad, el ensayo será de *dos colas* porque interesa testear los casos de mayor y de menor a la vez. Eso implica calcular dos límites dentro de los cuales estará la zona de aceptación de la (H_0) y a cada lado dos zonas de rechazo - ver Figura 12.1. Cuando en la hipótesis nula figure alguno de los signos de desigualdad, el ensayo tendrá *una cola*, es decir, habrá un solo valor crítico que separa las zonas de aceptación y de rechazo - ver Figura 12.2.

Figura 12.1 : (a) Ensayo de dos colas.

$$H_0 : \mu = a \quad H_1 : \mu \neq a$$

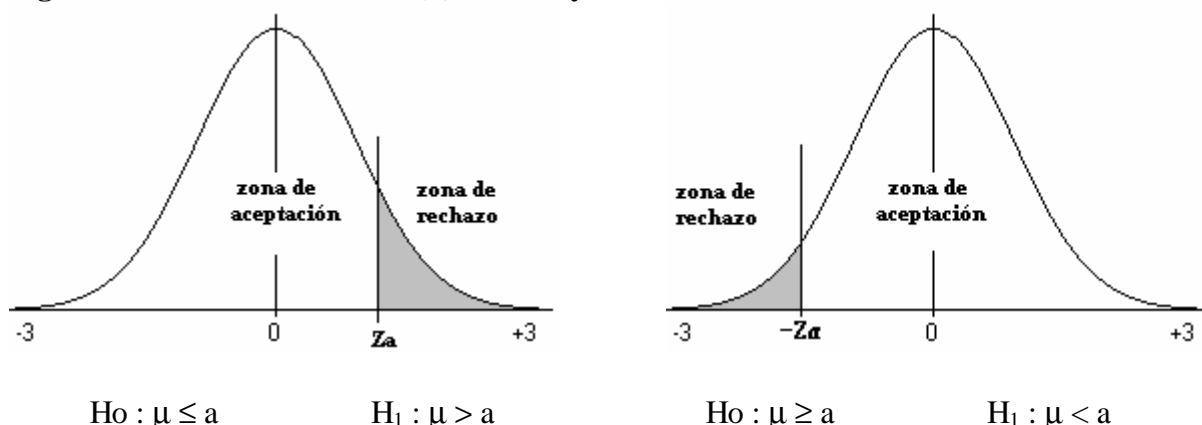


Suponiendo que la hipótesis nula dada $H_0 : \mu = a$, donde a es un valor cualquiera, referente al estadígrafo e que tiene una distribución muestral de tipo gaussiano, definida unívocamente con los parámetros $(\mu_e ; \sigma_e)$. Si se tipifica con $Z = (a - \mu_e) / \sigma_e$, la distribución de la variable tipificada Z será la función normal estandarizada, de parámetros $(\mu = 0 ; \sigma = 1)$ que está tabulada. Si este estadígrafo Z cae dentro de la zona de aceptación, entonces la hipótesis nula será aceptada. Caso contrario, si cae en la zona de rechazo, se considerará válida a la hipótesis alternativa; se rechazará la H_0 con la significación correspondiente. El área total sombreada es el nivel de significación α , por ello se debe repartir en cantidades iguales en ambos extremos de la curva. Eso define el valor crítico $z_{\alpha/2}$ a buscar en tablas. La zona de rechazo es simétrica respecto al origen, cada cola tiene bajo la curva la mitad del nivel de significación adoptado $\alpha/2$. La zona de aceptación de la hipótesis nula tiene bajo la curva una cantidad $1 - \alpha$. Por ejemplo, si se elige un nivel de confianza del 95%, el nivel de significación será $\alpha = 0,05$. Esto significa que se debe dejar un 2,5% en cada cola y el valor crítico será: $Z_{\alpha/2} = 1,96$ que corresponde a la probabilidad buscada.

Si en la hipótesis nula se define una situación de mayor o menor, entonces se trata de un ensayo de una cola. Los dos casos posibles se muestran en la Figura 12.1 (b) siguiente:

Figura 12.1:

(b) Ensayo de una cola.



El área total sombreada es el nivel de significación α y queda a la derecha del valor crítico Z_α , pues en la hipótesis nula se postula que el valor del estadígrafo testado es menor o igual que cierto valor límite dado por a . En cambio, cuando en la hipótesis nula se postula que el parámetro es mayor o igual al límite dado, el área sombreada quedará a la izquierda del valor crítico z_α y

definirá la zona de rechazo. Por ejemplo, si el nivel de significación es del 95 %, el área total será del 5 %, y así queda $Z_{0,05} = 1,645$.

En síntesis, los pasos a seguir en un ensayo de hipótesis son:

- 1) se formulan las hipótesis de trabajo: nula y su alternativa;
- 2) el investigador define el nivel de confianza y calcula el valor crítico de tablas;
- 3) con este valor, quedan establecidas las zonas de rechazo y aceptación de hipótesis;
- 4) se calcula el valor tipificado del estadígrafo usando μ_e ; σ_e y el valor a de la H_0 ;
- 5) se compara este valor con el crítico para aceptar o rechazar la H_0 como se vio antes.
- 6) para informar los resultados se usa principalmente el intervalo de confianza.

Estos pasos son comunes a todos los modelos estadísticos. La diferencia entre ellos reside en la manera de calcular el valor de comparación y en la Tabla estadística que se empleará para el ensayo. Pero la estrategia es similar para todos ellos. Una manera sencilla y práctica de hacer esta comparación es dibujando las zonas de aceptación y rechazo a mano alzada. Luego, se coloca el valor calculado en el paso 4 en el gráfico, y así ver en cuál zona cae, para proceder en consecuencia. Hay otra manera que se está volviendo más popular en la bibliografía médica. Consiste en calcular el intervalo de confianza al nivel fijado por el investigador. El estadígrafo Z que se obtiene a través de los datos experimentales debería caer dentro de tal intervalo si se cumple la H_0 planteada. En cambio, si el mismo cae fuera, eso significa que se puede rechazar la H_0 al nivel de confianza del intervalo. En resumen:

Forma clásica de hacer el test: comparar Z versus $Z\alpha$ para tomar la decisión.

Forma recomendada en clínica: fijarse si μ_e cae dentro o fuera del 95% CI para decidir.

O bien, fijarse si e cae dentro o fuera del 95% CI poblacional

Se presentan tres casos posibles, a saber, que el estadígrafo Z caiga en una de las zonas de aceptación / rechazo, o bien que coincida justo con el límite de ambas zonas. Esto es:

- Si Z cae dentro de la zona de aceptación: los resultados no son significativos y no se rechaza H_0 . Entonces, $Z < Z_{\text{crítico}}$ O bien, μ_e cae dentro del 95% CI.

- Si Z cae dentro de la zona de rechazo: los resultados son significativos y se rechaza la hipótesis H_0 . Se deben verificar los otros niveles de confianza. En este caso será $Z > Z_{\text{crítico}}$ O μ_e cae fuera del 95% CI

- Si Z cae en el límite: los resultados son dudosos. Por lo tanto, se deberán hacer más pruebas para poder llegar a una decisión. Es cuando μ_e coincide con uno de los valores límites del 95% CI. Acá es $Z \equiv Z_{\text{crítico}}$ (no es necesario que sean exactamente iguales puede ocurrir que a juicio del investigador estén “sospechosamente cercanos”).

Conviene remarcar el hecho siguiente: que una hipótesis nula sea aceptada, no quiere decir que se tenga prueba científica de su validez. Puede haber otro modelo científico más sensible que el de Gauss que detecte diferencias cuando este no lo haga. Solo cuando una hipótesis no es aceptada se puede decir: se ha encontrado evidencia científica para rechazar la hipótesis. Es decir, que se valida el rechazo, pero no la aceptación. Por ello muchos investigadores prefieren plantear hipótesis para ser rechazadas. Aprovechando el hecho de que son ellos quienes deciden

cual es la nula y cual es la alternativa. La sensibilidad de un modelo para detectar diferencias en los ensayos es la *robustez* del mismo. *Cuando se pueda se debe elegir el modelo más robusto.* Esto se verá mejor en los capítulos siguientes cuando se estudie la forma de resolver un mismo problema con más un modelo estadístico.

12.4 Modelo de Gauss para una muestra

El modelo de Gauss sirve para validar hipótesis estadísticas cuando se toma una muestra a los efectos de estimar los parámetros de una población. También se lo puede emplear para comparar dos muestras entre sí, bajo el supuesto de que ambas fueron extraídas de la misma población (como se verá en el punto 12.5 siguiente). En este punto se comenzará trabajando con una sola muestra, de tamaño lo suficientemente grande como para que este modelo pueda ser aplicado. De acuerdo con el estadígrafo con el cual se va a trabajar, se puede encontrar en la Tabla 4 el tamaño mínimo para poder emplear a este modelo. Como las muestras son grandes, es más sencillo y barato usar muestras pequeñas (como se verá en el capítulo siguiente). Pero lo que se busca aquí es ilustrar la manera de razonar y proceder en estos casos.

Cabe destacar que en la última década se comienza a hacer una distinción entre resultados *estadísticamente* significativos y *clínicamente* significativos. Por ejemplo, una diferencia de porcentajes $Z = 13\%$ con 95% CI (7% ; 12%) es estadísticamente significativa, pero si solo importan las diferencias de 15% o más, entonces $Z = 13\%$ no es clínicamente significativa.

En teoría, se pueden ensayar todos los estadígrafos muestrales como los presentados en la Tabla 4. Con las correcciones correspondientes a los casos de muestreo sin reposición en poblaciones finitas. Sin embargo, los estadígrafos de uso más frecuentes son los siguientes:

12.4.1 Medias

El modelo de Gauss para las distribuciones de medias muestrales se puede aplicar para muestras de tamaño treinta o mayores ($n \geq 30$) donde $\mu_e = \mu_{\bar{X}} = \bar{X}$ y $\sigma_e = \sigma / \sqrt{n}$; por lo tanto el estadígrafo Z para poder compararlo con el valor crítico de tablas Z_{α} , será:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad \text{donde } \sigma \text{ es desconocido y se lo aproxima con DS.}$$

Ejemplo de calibración: Se realizaron 36 mediciones repetidas de una misma magnitud clínica: glucosa; usando una solución calibrada con $\mu = 90$ mg/dl, los resultados obtenidos arrojan una media de 95 mg/dl y un desvío estándar de 8 mg/dl. Ensayar la hipótesis de que el sistema de medición está calibrado (en exactitud) con un nivel de confianza adecuado.

Aquí conviene plantear la igualdad como una hipótesis nula.

$H_0 : \mu = 90$ mg/dl y el sistema está calibrado en exactitud.

$H_1 : \mu \neq 90$ mg/dl y el sistema no está calibrado, tiene un error de tipo sistemático.

Por lo tanto se trata de un ensayo de dos colas. El estadígrafo de comparación es:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{95 - 90}{8/\sqrt{36}} = 3,75^{***}$$

Los valores críticos de tablas son:

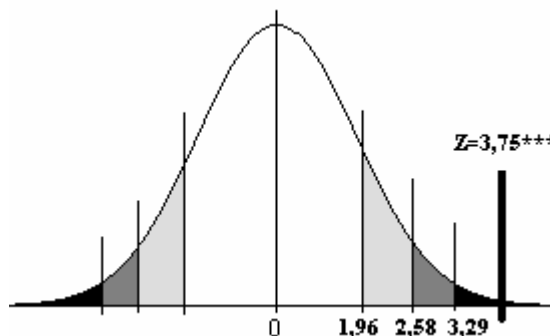
$$Z_{0,95} = 1,96 ; Z_{0,99} = 2,58 \text{ y } Z_{0,999} = 3,29$$

Como $Z \gg Z\alpha$ se rechaza la H_0 . O bien:

$Z = 3,75$ cae fuera de 99,9% CI (-3,29 ; +3,29)

$\mu = 90$ cae fuera de 99,9 % CI (90,6; 99,4)

$\bar{X} = 95$ cae fuera de 99,9% CI (85,6 ; 94,4)



La exactitud del método no es aceptable

Las zonas sombreadas en la figura anterior indican las tres zonas de rechazo de la hipótesis nula. A medida que se oscurece el rechazo es mayor. La primera zona delimitada por $Z_{0,95} = 1,96$ es cuando hay resultados significativos (95%). Como el estadígrafo Z es mayor y cae dentro de esta zona, hasta ahí se dice que se obtuvieron resultados *significativos* como para rechazar H_0 , y se lo denota con $Z = 3,75^*$. En estos casos, se debe continuar comparando con el siguiente nivel del 99%; como $Z_{0,99} = 2,58$ es todavía menor que Z , ahora se puede decir que se obtuvieron resultados *muy significativos* que se denotan con $Z = 3,75^{**}$, para rechazar H_0 . Entonces, se debe continuar con el último nivel acostumbrado del 99,9%. Nuevamente, $Z_{0,999} = 3,29 < Z$; por lo tanto la conclusión final es que se tienen resultados *altamente significativos* $Z = 3,75^{***}$ que prueban la existencia de un error sistemático que ha descalibrado al sistema de medición, en cuanto a su exactitud. Esto significa una *validación estadística* de la conclusión adoptada. La decisión consecuente será calibrar el sistema. La estadística es una herramienta para probar la existencia de tal error, pero no explica la forma de corregirlo. Esto lo debe hacer el investigador con sus conocimientos profesionales y su experiencia técnica. Todavía se puede estimar el módulo de este error sistemático (ES) con: $ES = \text{valor promedio} - \text{valor control} = 5 \text{ mg/dl}$.

También se puede efectuar la validación con los CI. Se pueden plantear tres tipos de CI: El CI de la variable tipificada Z es 99,9% CI (-3,29 ; 3,29) como $Z = 3,75$ cae fuera se rechaza H_0 . Otro tipo de CI es el muestral visto en el capítulo anterior: 99,9 % CI (90,6; 99,4) y como el $\mu = 90$ cae fuera del mismo se rechaza H_0 . El último tipo de CI es el que se forma con los valores poblacionales armando el intervalo con $(\mu \pm Z_{\alpha} \cdot \sigma_e)$ y viendo si $\mu_e = \bar{X}$ cae dentro del mismo. En este caso será: $(90 \pm 4,4)$ para un 99,9% de confianza, como $\bar{X} = 95$ cae fuera se rechaza H_0 . Estos tres tipos de CI son totalmente equivalentes y se pueden usar según convenga, por ejemplo para las diferencias de medias o proporciones conviene el último pues el valor esperado es $\mu = 0$.

Se calculan las probabilidades para el valor $Z = 3,75$ y $Z = -3,75$ con la ecuación gaussiana usando el programa Excel, y se obtiene una probabilidad de rechazo $p = 0,00018$ (esto es la probabilidad dentro de la zona de aceptación es de 0,99982). Como $p < \alpha$ se puede rechazar la hipótesis nula (cuando $Z > 3,5$ ya se puede rechazar H_0 sin ir a tablas). Algunos investigadores prefieren informar de esta manera, diciendo por ejemplo: *se encontró un error sistemático* ($p = 0,00018$). Comparando ambas maneras de informar, se concluye que es más corto y completo informar así: *se encontró un error sistemático de 5 mg/dl; $\mu_e = 95$ con 99,9% CI (90,6 ; 99,4)*. Al ver la Z el lector deduce que se ha empleado el modelo de Gauss y como el valor cae fuera del intervalo de confianza del 99,9%, deduce que se tiene validación estadística de lo afirmado.

También se puede abreviar con: *mediante Gauss se obtuvieron resultados altamente significativos: el sistema está descalibrado en exactitud.*

Para poder calibrar el sistema desde el punto de vista de la precisión habrá que estudiar como se comporta el desvío estándar si se lo compara con uno “patrón”, esto es contra un desvío máximo admisible (σ_{\max}) que se obtiene de los textos de análisis clínicos.

12.4.2 Varianzas

El modelo de Gauss para las distribuciones de varianzas muestrales se puede aplicar para muestras de tamaño cien o mayores ($N \geq 100$), donde $\mu_e = \mu_{\sigma^2} = DS^2$ y $SE(\sigma^2) = \sigma_{\sigma^2} = \sigma^2 \sqrt{2/n}$. Por lo tanto, el estadígrafo Z para poder compararlo con el valor crítico de tablas Z_{α} , será:

$$Z = \frac{DS^2 - \sigma^2}{\sigma^2 \sqrt{2/n}} \quad \text{donde } \sigma \text{ es desconocido y se lo postula con } H_0.$$

Ejemplo de calibración: Continuando con el caso anterior, el investigador completa las 100 muestras obteniendo una nueva media de 94,6 mg/dl con un desvío de 9,65 mg/dl. De los libros de texto surge que la dispersión aceptada para la determinación de Glucosa viene dada por un coeficiente de variación del 10%. Se desea analizar si la dispersión encontrada en el experimento es menor, o por lo menos igual que la aceptable.

El CV% = 10% significa: $0,1 = \sigma_{\max} / \mu = \sigma_{\max} / 90$. De allí surge $\sigma_{\max} = 9$ mg/dl. Se plantean:

H_0 : $\sigma \leq 9$ mg/dl y el sistema tiene una dispersión aceptable.

H_1 : $\sigma > 9$ mg/dl y el error casual es muy grande. El sistema no está calibrado en precisión.

Por lo tanto, se trata de un ensayo de una sola cola y el estadígrafo de comparación se saca con:

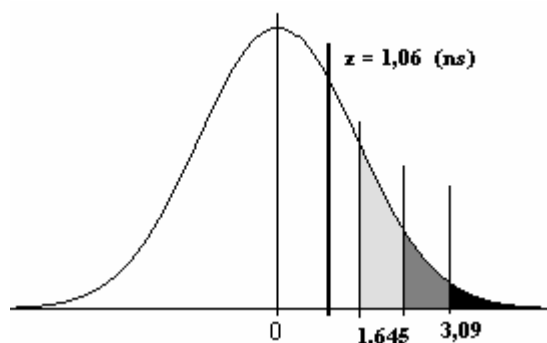
$$Z = \frac{DS^2 - \sigma^2}{\sigma^2 \sqrt{2/n}} = \frac{(9,65)^2 - (9)^2}{(9)^2 \sqrt{2/100}} = 1,06$$

Los valores críticos de tablas son:

$$Z_{0,95} = 1,645 ; Z_{0,99} = 2,33 \text{ y } Z_{0,999} = 3,09$$

Como $Z < Z_{\alpha}$ no se rechaza la H_0 .

$$\sigma^2 \leq 81 \in 95\% \text{ CI } (-\infty ; +87,2)$$



Por lo tanto, los resultados no fueron significativos. Esto es, no hay evidencia suficiente como para pensar que la dispersión no sea aceptable. El error casual no es significativo. Dicho en otros términos, la precisión del sistema de medición es menor que la máxima admisible recomendada en los libros de texto. La forma de expresarlo puede ser:

- no se rechaza la hipótesis nula: $Z = 1,06$ (ns) y también con 95% CI $(-\infty ; +87,2)$;
- no hay evidencia que pruebe una dispersión no aceptable.

12.4.3 Proporciones

El modelo de Gauss para las distribuciones de las proporciones se puede aplicar para muestras de tamaño $n > 25$ o mayores donde $\mu_e = \mu_p = \pi$ y $\sigma_e = SE(\pi) = \sqrt{\pi \cdot (1 - \pi) / n}$. Por lo tanto el estadígrafo Z para poder compararlo con el valor crítico de tablas Z_{α} , será:

$$Z = \frac{p - \pi}{\sqrt{\pi \cdot (1 - \pi) / n}}$$

donde p es el valor experimental de la proporción.

O bien,

$$Z = \frac{r - n \cdot \pi}{\sqrt{n \cdot \pi \cdot (1 - \pi)}}$$

si se trabaja con número de éxitos en vez de proporciones.

Notar que la distribución de las proporciones, o el número de éxitos r, es en realidad una distribución binomial, por lo tanto para que la aproximación sea buena, hay que efectuarle la corrección por continuidad de Yates, o lo que sería mejor la aproximación de Molenaar vista en el punto 9.8.1 anterior.

Ejemplo 1: En un experimento sobre PES (percepción extrasensorial), el sujeto investigado debía adivinar el símbolo de las cartas Zenner que el investigador iba sacando en otro cuarto separado. Cada vez que sacaba una carta tocaba un timbre y el sujeto anotaba el símbolo que creía ver. Se realizó esta prueba 400 veces y acertó 160 veces. Decidir si el sujeto tiene PES.

$H_0 : \mu = \pi = 0,2$ El sujeto está adivinando, $p = 1/5$ (hay 5 símbolos diferentes en las cartas).

$H_1 : \mu > 0,2$ El sujeto tiene capacidad extrasensorial pues acierta más de lo esperado.

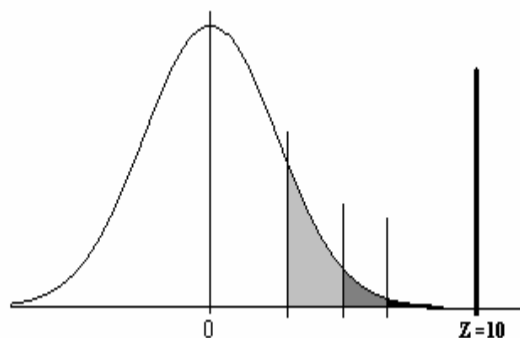
Por lo tanto, se trata de un ensayo de una cola. El porcentaje de aciertos medido en este experimento fue de $p = 160 / 400 = 0,4$. Además $\mu_p = 0,2 = \pi$ y $\sigma_{\pi} = 0,02$. Reemplazando en la primera ecuación

$$Z = (0,4 - 0,2) / 0,02 = 10^{***}$$

Como $Z_{0,95} = 1,645$; $Z_{0,99} = 2,33$ y $Z_{0,999} = 3,09$.

El resultado obtenido es altamente significativo.

Se tiene una fuerte evidencia que este sujeto tiene poder extrasensorial. $p = 0,4$ no cae en 99,9% CI $(-\infty ; +0,26)$



Notar que en realidad se debería hacer la corrección por continuidad, porque se está usando una distribución continua para validar una variable discreta. Pero cuando la evidencia es tan fuerte como en este caso no es necesaria. Para ver esto se puede calcular $Z_{Yates} = 9,94^{***}$ y con la de Molenaar es $Z_M = 8,83^{***}$. Por su parte la probabilidad exacta con la binomial es casi 1.

Ejemplo 2: El fabricante de un medicamento asegura que el mismo tiene un 85% de efectividad para aliviar las neuralgias en medio día. Se tomó una muestra de 500 individuos con neuralgia, a

los cuales se les suministró el medicamento y 400 de ellos se aliviaron en el plazo previsto. Decidir si lo que asegura el fabricante es cierto.

Se trata de un ensayo de una sola cola se pueden postular dos tipos de H_0 : una usando las proporciones de éxitos como se vio en el ejemplo 1, y otra empleando directamente las frecuencias.

$H_0 : \mu \geq 500 (0,85) = 425$ pacientes aliviados y lo que dice el fabricante es cierto.

$H_1 : \mu < 425$ y no es cierta la afirmación realizada.

$$\mu_r = n \cdot \pi = 425 \text{ y } \sigma_r = [(n \cdot \pi (1-\pi))]^{1/2} = 7,98 \approx 8$$

Reemplazando y efectuando las correcciones por continuidad resulta:

$$Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma = [|400 - 425| - 1/2] / 8 = 3,19^{***}$$

$$Z_M = [(4 \cdot 400 + 3)0,15]^{1/2} - [(4 \cdot 500 - 4 \cdot 400 - 1)0,85]^{1/2} = 2,91^{**}$$

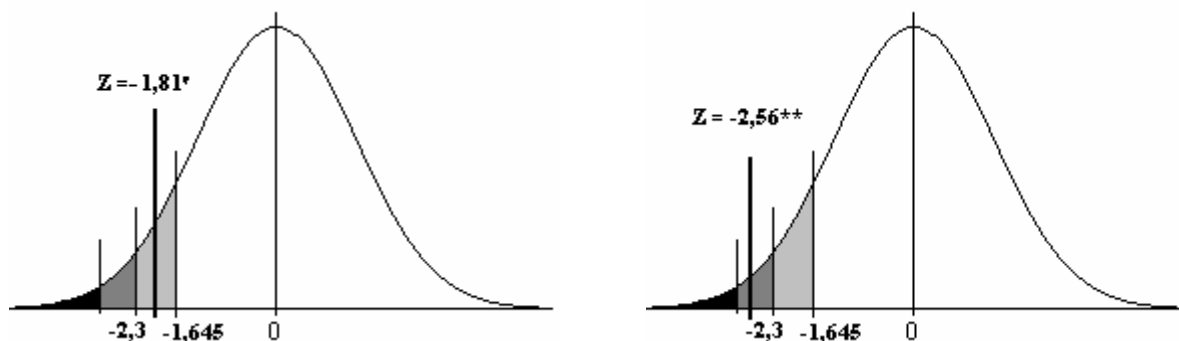
La probabilidad binomial es 0,998407 la cual corresponde a un valor $Z = 2,96$

En conclusión el resultado obtenido es muy significativo. Se encontró evidencia como para rechazar la afirmación del fabricante.

Conviene recordar que la H_0 define el lugar donde se coloca la cola: si en la H_0 se postula mayor o igual, corresponde la cola de la izquierda como en este caso. Si en la H_0 se postula menor o igual, es al revés. Tanto Yates como Molenaar siempre dan valores positivos porque lo que se busca es la probabilidad, y para determinar de cual cola se trata hay que estudiar la H_0 .

Ejemplo 3: El mismo caso anterior, solo que esta vez se curaron 411 individuos. Es todo similar solo que el nuevo valor de Z es:

$Z_{\text{Yates}} = [|r - \mu| - 1/2] / \sigma = [|411 - 425| - 1/2] / 8 = 1,81^*$ Como se trata de la cola izquierda de la distribución el valor de Z debería ser negativo y graficarse como:



Ahora pasan dos cosas: este nuevo valor de Z cae en la zona de rechazo para el 95% de confianza, pero en la zona de aceptación para el 99% de confianza. Estos resultados no son contradictorios como puede parecer a primera vista. Ocurre que se tiene evidencia como para rechazar la H_0 al nivel del 95%, pero no es tan fuerte como para rechazarla al nivel del 99%. Y es por eso que se le colocó un solo asterisco. En cambio, si por ejemplo se hubieran curado 405 individuos, resultaría $Z = -2,56^{**}$. O sea, se rechaza a los niveles del 95% y del 99% pero no alcanza

para rechazar la H_0 al nivel del 99,9% ($Z_{0,999} = -3,09$). En las figuras de más arriba se han graficado estas dos situaciones. Las mismas se deben expresar como sigue:

$$r = 411 \notin 95\% \text{ CI } (412 ; + \infty) \text{ y con eso se deduce que cae dentro del } 99\% \text{ CI}$$

$$r = 405 \notin 99\% \text{ CI } (406 ; + \infty) \text{ y con eso se deduce que cae dentro del } 99,9\% \text{ CI}$$

Notar que si se hubiese usado la aproximación de Molenaar los valores hubieran sido:

$$Z_M = [(4.411 + 3)0,15]^{1/2} - [(4.500 - 4.411 - 1)0,85]^{1/2} = 1,682^*$$

$$Z_M = [(4.405 + 3)0,15]^{1/2} - [(4.500 - 4.405 - 1)0,85]^{1/2} = 2,346^{**}$$

Los valores anteriores corresponden a la cola izquierda de la distribución y por lo tanto deben ser negativos. Las conclusiones son análogas pero los valores dan más cercanos a los límites. Por otra parte se pueden calcular las probabilidades exactas con la probabilidad Binomial que son:

$$P_{Bi} (r \leq 411) = 0,04784 \text{ lo que implica un valor } Z = - 1,666$$

$$P_{Bi} (r \leq 405) = 0,04784 \text{ lo que implica un valor } Z = - 2,378$$

Una vez más se puede apreciar que la aproximación con el método de Molenaar es mejor que el procedimiento clásico de Yates.

12.5 Comparaciones de dos muestras

El modelo de Gauss sirve para comparar dos muestras entre sí y validar la hipótesis estadística de que ambas provienen de dos poblaciones con la misma media. Esto se puede hacer comparando dos medias muestrales para el caso de variables continuas, o dos proporciones en el caso de variable discreta. Si además se pudiese probar que las varianzas son iguales, entonces se podría validar el hecho de que ambas muestras provienen de la misma población normal. Ya que se está suponiendo que tienen igual media y varianza, que son los dos parámetros que definen en forma unívoca a una curva de Gauss.

La hipótesis nula más conveniente es que ambas muestras tienen igual media (o igual proporción), porque de esa forma no se necesitan sus valores en la fórmula del cálculo del estadígrafo Z. Se puede trabajar con tamaños de muestras diferentes, aunque si fueran iguales las fórmulas son más sencillas. Los dos casos principales son:

12.5.1 Comparaciones de dos medias muestrales

Para este caso se pueden emplear las fórmulas vistas en el capítulo anterior para la definición de los intervalos de confianza, y así tipificar la diferencia de medias. Esto es postular:

$$H_0 : \mu_1 = \mu_2 \quad \text{O sea, } \mu(\bar{x}_1 - \bar{x}_2) = 0 : \text{ no hay diferencia entre las medias.}$$

$H_1 : \mu_1 \neq \mu_2$ O sea, $\mu(\bar{x}_1 - \bar{x}_2) \neq 0$: hay diferencia entre las medias

Entonces siempre será un ensayo de dos colas. Y el estadígrafo Z de comparación es:

$$Z = [(\bar{x}_1 - \bar{x}_2) - (\mu_{\bar{x}_1} - \mu_{\bar{x}_2})] / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Aplicando la H_0 a la ecuación anterior queda: $Z = [(\bar{x}_1 - \bar{x}_2)] / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Y si las muestras son de igual tamaño y desvío: $Z = [(\bar{x}_1 - \bar{x}_2)] / [\sigma \sqrt{\frac{2}{n}}]$

Ejemplo 1) Se realizaron 36 mediciones repetidas de glucosa, usando una solución calibrada con $\mu = 90$ mg/dl, con dos marcas comerciales diferentes. Los resultados obtenidos para cada caso son: Marca A (media 95 mg/dl y desvío 8 mg/dl), Marca B (media 93 y desvío 7 mg/dl). Ensayar la hipótesis de que ambas marcas son equivalentes.

$H_0 : \mu_A = \mu_B$ El uso de una marca u otra es indistinto

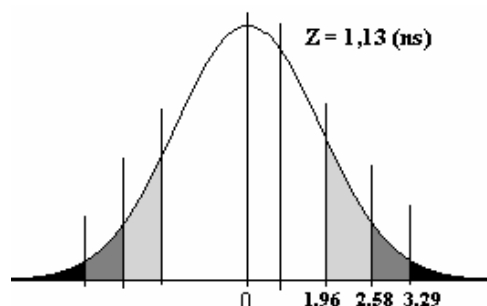
$H_1 : \mu_A \neq \mu_B$ Hay diferencia entre ambas marcas

Donde $(\bar{x}_1 - \bar{x}_2) = 95 - 93 = 2$ y $\sigma_{1-2} = 1,772$

Entonces resulta:

$$Z = [(\bar{x}_1 - \bar{x}_2)] / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 2 / 1,772 = 1,13 \text{ (ns)}$$

El valor 2 cae dentro del CI 95% (- 3,5 ; + 3,5)



Como $Z < 1,96 = z_{\alpha/2}$ no se puede rechazar la hipótesis nula. Este test no arroja evidencia significativa, como para pensar que hay diferencias entre las marcas A y B.

Ejemplo 2) En una industria farmacéutica se cambió un dosificador líquido por uno nuevo. El anterior tenía un promedio histórico de 5,01 ml con un desvío de 0,08 ml en 100 pruebas de calidad. Al actual se le realizaron 50 pruebas arrojando una media de 5,048 ml y un desvío de 0,05 ml. Con esta información se debe decidir si ambos dosificadores son equivalentes.

$H_0 : \mu_N = \mu_V$ Son equivalentes.

$H_1 : \mu_N \neq \mu_V$ Hay diferencia entre ambos.

Usando la fórmula general donde el supuesto principal es que las varianzas son diferentes, pero los valores esperados son equivalentes entonces resulta:

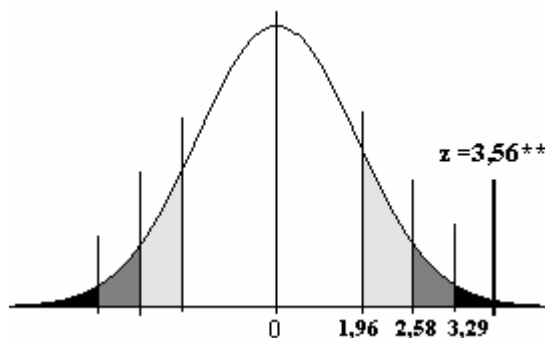
$\sigma_1 \approx DS_1 = 0,08$ y $\sigma_2 \approx DS_2 = 0,05$ con sus tamaños muestrales de 100 y 50 respectivamente:

$$Z = [(5,048 - 5,01)] / \sqrt{\frac{0,05^2}{50} + \frac{0,08^2}{100}}$$

$$Z = 3,56^{***} \notin \text{CI } 99,9\% (-3,29 ; +3,29)$$

$$\mu_N - \mu_V = 0,038 \notin \text{CI } 99,9\% (-0,035 ; +0,035)$$

Hay fuerte evidencia de que no son equivalentes.



Notar que se ha probado que hay una diferencia del 1,2% entre ambos. Lo que significa que cada medicamento fabricado tendrá, en promedio, una cantidad mayor que la histórica lo que implicará un aumento de los costos de producción. Esta evidencia aconseja no usar el nuevo dosificador.

12.6 Comparaciones de dos proporciones

Para este caso se pueden emplear las fórmulas vistas en el capítulo anterior para la definición de los intervalos de confianza de la diferencia de dos proporciones y así tipificarla. O sea:

$$Z = (p_1 - p_2) - (\mu p_1 - p_2) / \sqrt{\frac{\pi_1 \cdot (1 - \pi_1)}{n_1} + \frac{\pi_2 \cdot (1 - \pi_2)}{n_2}}$$

Nuevamente conviene hacer el supuesto de igualdad de proporciones en la hipótesis nula para simplificar la fórmula cuando se pueda. Esto es:

$H_0 : \mu p_1 - p_2 = 0$: No hay diferencia entre ambas proporciones.

$H_1 : \mu p_1 - p_2 \neq 0$: Hay diferencia entre ambas.

También se puede trabajar con las frecuencias en lugar de las respectivas proporciones.

Ejemplo 1) Dos grupos A y B formados por 100 pacientes sufren un cierto tipo de infección. Se administra un nuevo remedio a los del grupo A y un placebo a los del grupo B que es el grupo testigo o control. En todo lo demás, se trata a los 200 pacientes de la misma forma. Transcurrido el tiempo de prueba se verifica que se han curado 75 individuos del grupo A y 65 del B. Decidir si el nuevo remedio ayuda a curar la enfermedad.

$H_0 : \pi_A \leq \pi_B$ Las diferencias detectadas se deben al azar, el remedio no es efectivo.

$H_1 : \pi_A > \pi_B$ Hay diferencia entre ambas y el remedio es efectivo.

Acá conviene hacer un ensayo de una sola cola porque lo que interesa es ver si el remedio sirve, esto es, si el número de curados es mayor en el grupo A que en el grupo B. Los valores muestrales de proporciones se usan para estimar los poblacionales. Entonces:

$$\mu_{1-2} = 0,75 - 0,65 = 0,10 \quad \text{y} \quad \sigma_{1-2} = \sqrt{\frac{0,75 \cdot 0,25}{100} + \frac{0,65 \cdot 0,35}{100}} = 0,0644$$

Con estos valores se puede calcular:

$$Z = 0,10 / 0,0644 = 1,55 \text{ (no significativo)} \quad \mu_{1-2} = 0,10 \text{ cae dentro de CI 95\% } (-\infty ; +0,13)$$

Esto significa que no se tiene prueba científica como para poder afirmar que el suero es efectivo. Si se le realiza la corrección por continuidad el valor de Z es todavía menor ($Z = 1,48$).

Otra forma de resolver este problema es haciendo los supuestos siguientes:

a) Ambas muestras provienen de la misma población, esto es $\pi = \pi_A = \pi_B$

b) La mejor estimación del valor desconocido π es con la media ponderada de las proporciones observadas. Esto es: $\pi \approx [100 (0,75) + 100 (0,65)] / (100 + 100) = 0,7$

Y la varianza se calcula ahora con este valor estimado de 0,7. Esto es:

$$\sigma^2 = \pi (1 - \pi) [(1/n_1) + (1/n_2)] = 0,7 \cdot 0,3 [(1/100) + (1/100)] = 0,0042 \quad \text{O sea, } \sigma = 0,06481$$

c) Se formula la H_0 lo que significa que el valor observado de la diferencia se debe al azar.

$$Z = (p_1 - p_2) / \sigma = 0,10 / 0,06481 = 1,543 \text{ (no significativo)}$$

Notar que los dos procedimientos son muy similares y puede adoptarse cualquiera de ambos en casi todos los casos. La excepción es cuando las diferencias entre los tamaños muestrales es muy grande y cuando las proporciones difieren substancialmente. En todo otro caso, los valores del desvío estándar son muy parecidos como en el ejemplo anterior $\sigma_{1-2} = 0,0644$ y $\sigma = 0,06481$.

Ejemplo 2) Una muestra de 300 clientes de la Farmacia A demostró que el 58% paga de contado. En cambio en otra muestra de 200 clientes de la Farmacia B solo lo hace el 44%. Se desea saber si esa diferencia se debe al azar o si existe alguna causa como su ubicación en la ciudad.

$H_0 : \mu p_1 - p_2 = 0$: No hay diferencia entre ambas.

$H_1 : \mu p_1 - p_2 \neq 0$: Hay diferencia entre ambas.

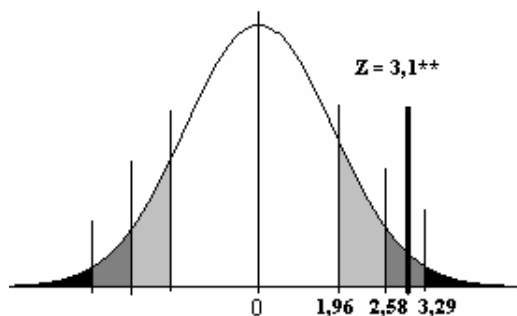
Acá se hace un ensayo de dos colas porque lo que interesa es ver si hay diferencia entre ambas Farmacias. Por lo tanto se calculan:

$$\mu_{1-2} = 0,56 - 0,48 = 0,14 \quad \text{y} \quad \sigma_{1-2} = \sqrt{\frac{0,58 \cdot 0,42}{300} + \frac{0,44 \cdot 0,56}{200}} = 0,0452$$

Con estos valores se puede calcular:

$$Z = 0,14 / 0,0452 = 3,1^{**} \quad \text{pues } Z_{0,95} = 1,96 ; Z_{0,99} = 2,58 \text{ y } Z_{0,999} = 3,29$$

En el caso de desvío ponderado es $p = 0,524$ y $\sigma = 0,0456$ valor muy parecido al anterior σ_{1-2} . Cualquiera de los dos procedimientos dará valores muy parecidos de Z y pueden ser usados indistintamente.



Esto significa que se tienen resultados muy significativos que prueban la diferencia entre la clientela de ambas farmacias, respecto al pago contado.

$$2,58 < Z < 3,29$$

$\mu_{1-2} = 0,14 \notin 99\% \text{ CI } (-0,12 ; +0,12)$ como cae fuera del intervalo se rechaza la H_0

12.7 Intervalos de confianza versus test de hipótesis

En los comienzos del uso de la validación estadística de los resultados experimentales, la costumbre era informar la probabilidad asociada al evento estudiado. Por ejemplo, se informaba que $p < 0,05$ lo que implicaba decir que los resultados fueron significativos, o bien $p = 0,039$ que significa lo mismo, pero usando el valor hallado de probabilidad. Esta costumbre fue criticada porque el lector del trabajo de investigación no podía saber el tipo de test estadístico usado, ni la manera de emplear los datos del estudio. El primer cambio fue introducir el valor del estadígrafo hallado junto con la probabilidad del mismo. Por ejemplo, informar que se encontró un $Z = 2,64$ con $p = 0,0047$, o bien $Z = 2,64^{**}$, de donde el lector podía deducir que se había aplicado la función normal y se habían encontrado valores muy significativos ($p < 0,01$). A esta manera de informar se la denominó: test de hipótesis o valores de probabilidad. En un trabajo de Gardner y Altman publicado en 1986, los autores presentan un estudio comparativo entre el uso de los intervalos de confianza y los test de hipótesis en Medicina.

Según estos autores desde 1950 se produjo un incremento inusitado en el uso de estadística en Medicina, lo que se tradujo en consecuencias desafortunadas porque en vez de analizar los resultados obtenidos clínicamente, lo reducían a una dicotomía del tipo H_0 vs. H_1 . El excesivo uso del test de hipótesis a expensas de otras formas de análisis llegó a tal grado, que en los trabajos de investigación solo se hacía referencia a los valores de probabilidad, o si se acepta o rechaza la H_0 , sin mencionar cosas tales como las concentraciones halladas, las proporciones, sus diferencias etc. La alternativa propuesta fue usar los intervalos de confianza para el enfoque estadístico en los trabajos de investigación, porque entregan más información al lector. Esta es la recomendación para trabajos futuros a publicarse en las revistas médicas y sus asociadas.

El problema básico es que diferencias pequeñas encontradas en una magnitud clínica pueden ser estadísticamente significativas en grandes muestras, mientras que en clínica son irrelevantes. O bien, importantes efectos clínicos pueden aparecer como estadísticamente no significativo solo porque el número de casos estudiados era pequeño. La idea básica es que: *cuando un resultado es estadísticamente significativo, no significa necesariamente que clínicamente lo sea*. Otro problema es que una magnitud clínica no tiene porque poseer una distribución normal en poblaciones humanas, pero si se le toman muestras aleatorias e independientes lo suficientemente

grandes, la distribución muestral va a ser normal independientemente de la población de la que fue extraída. Esto hecho ya fue denotado en los casos de las aproximaciones normales a poblaciones binomiales o de Poisson. Entonces hay que tener cuidado de no suponer a la población como normal, solo porque las muestras extraídas lo sean.

En conclusión la propuesta es incluir toda la información estadística posible en un informe para que tanto el lector como el autor eviten confusiones. Por ejemplo, conviene informar los resultados así: “La diferencia entre las medias muestrales de la presión sistólica entre diabéticos y no diabéticos encontrada fue de 6,0 mm Hg., con un intervalo de confianza del 95% entre 1,1 y 10,9 mm Hg. y el estadístico hallado fue $Z = 2,9$ con una probabilidad asociada de $p = 0,002$.” O en forma sintética: “ $\Delta \bar{x} = 6$ mm Hg, 95% CI (1,1 ; 1,9), $Z = 2,4^{**}$ ($p = 0,002$)”

12.8 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|---|-------|-------|
| 1) La H_0 es la hipótesis que se plantea para ser testeada. | V | F |
| 2) Las H_0 y la H_1 son complementarias. | V | F |
| 3) Explicar que son los riesgos del comprador y del vendedor | | |
| 4) La mejor forma de minimizar los errores es aumentando el número de experimentos. | V | F |
| 5) Explicar los tres niveles de significación usuales en variables biológicas..... | | |
| 6) La zona de aceptación en un ensayo de dos colas es igual a la de una cola. | V | F |
| 7) Explicar los pasos a seguir en un ensayo de hipótesis | | |
| 8) Los resultados son muy significativos si Z cae en la zona de rechazo del nivel 99,9% | V | F |
| 9) Los resultados no son significativos si Z cae en la zona de aceptación. | V | F |
| 10) Si se acepta la H_0 se tiene prueba científica de la misma. | V | F |
| 11) Cuando se rechaza la H_1 se dice que se han validado las conclusiones alcanzadas. | V | F |
| 12) Las pruebas de validación más usuales se hacen con la distribución muestral de | | |

2) La vida útil de un medicamento A se midió con 200 muestras y dio una media de 1400 días, con un desvío de 120 días. El medicamento B se determinó con 100 muestras y se obtuvo una media de 1200 días con un desvío de 80 días. Decidir si hay diferencia entre ambos.

3) Suponiendo que se usa una muestra de sangre control y al suero obtenido se lo fracciona en 30 alícuotas, luego a cada una se le determina la creatinina, y con los valores medidos se obtienen un promedio de 10 mg/dl y un desvío de 2,2 mg/dl. Si el valor real es de 9,5 mg/dl. Decidir si el método clínico está calibrado.

4) Se toman 40 muestras de cierto antibiótico y se les mide su contenido de Sulfato de Neomicina, resultando un promedio de 87 mg con un desvío de 4 mg. Decidir si el contenido del Sulfato está por debajo del límite de control de producción, fijado en 85 mg.

5) Se realizaron 100 exámenes de Bioestadística. Si se toman 30 de ellos al azar entre los varones y se obtiene una media de 6,5 puntos con un desvío de 1,5 puntos, mientras que se eligen otros 30 entre las mujeres obteniendo una media de 7 con un desvío de 1 punto. Decidir si hay diferencias entre las notas de ambos sexos.

6) Se lanza una moneda al aire 100 veces y se obtienen 53 caras. Decidir si la moneda está bien hecha. ¿Qué ocurre si se sacan 58 caras?

7) Se tomaron 200 muestras aleatorias de presión sistólica a niños cuyos padres son hipertensos, obteniéndose una media de 100 y un desvío de 5. Luego se tomaron 100 muestras de niños cuyos padres tienen la presión sanguínea normal, y se obtuvo una media de 95 con un desvío de 4. Decidir si hay diferencia entre ambos grupos.

8) El peso promedio de 50 estudiantes de la Facultad elegidos al azar entre los que toman parte de pruebas atléticas es de 60,5 Kg. con un desvío de 5 Kg. Mientras que otros tantos alumnos, elegidos entre los que no demuestran interés, arrojó una media de 64,5 Kg. con un desvío de 3 Kg. Testear la hipótesis de que los alumnos que participan en pruebas atléticas son más flacos que los otros.

9) Para hacer un test de embarazo se emplea la técnica 1 en 400 pacientes y se encuentran 220 pacientes que dieron un resultado (+). Los aciertos en (+) fueron 180 y 160 en (-). Con estos datos calcular la Sensibilidad y Especificidad de esta técnica. A las mismas pacientes se les aplicó la técnica 2 y resultaron 240 pacientes con (+), mientras el número de aciertos en (+) fue de 190 y en (-) 150. Decidir cuál de ambas técnicas es la mejor para adoptar en nuestro Laboratorio (a un nivel del 95%).

10) Se sabe que la concentración de coloides en el agua de un río es del 5,4 /ml. Se tomó una muestra del mismo a la que se le agregó un reactivo. Luego de un rato se extrajo una pequeña cantidad para analizarla en una cámara de recuento. Los resultados fueron:

Nº de coloides: 2 3 4 5 6 7 (nº/ml)

Frecuencia: 10 20 170 100 90 10

Decidir si hubo algún cambio significativo en la concentración del agua (a un 95% y 99%)

11) Se midieron 50 alícuotas de un suero con el Kit de Glucosa marca X, y otros 50 con la marca Y. Los resultados fueron:

Marca X: entre (80-80,9) 8 veces, (81-81,9) 22 veces, (82-82,9) 15 veces y (83-83,9) 5 veces.

Marca Y: entre (80-80,9) 6 veces, (81-81,9) 30 veces, (83-83,9) 10 veces y (83-83,9) 4 veces

Decidir si hay diferencia en la exactitud y en la precisión de ambos métodos, comparando las medias y desvíos estándar al 95% de confianza. Plantear los intervalos de confianza para cada uno al 95%, 99% y 99,9%.

12) Para hacer un test VDRL se emplea el *Kit A* en 200 pacientes y se encuentran 120 pacientes que dieron un resultado (-). Los aciertos en (+) fueron 70 y 110 en (-). Con estos datos calcular la Sensibilidad y Especificidad de A y B. Si con los mismos pacientes se empleó el *Kit B* y resultaron 140 pacientes con (-), mientras el número de aciertos en (+) fue de 55 y en (-) 130. Decidir cuál de ambas técnicas es la mejor para adoptar en nuestro Laboratorio (a un nivel del 95%).

13) En un recuento de rojos en cámara se espera obtener 4,5 millones por cm³ de cierto paciente. Se le toma una muestra de sangre, y de acuerdo con los resultados hallados decidir (a un nivel de confianza del 95%) si se verifica lo esperado.

Nº de rojos por millón 2 3 4 5 6

Frecuencia 10 90 110 150 40

13

Teoría de pequeñas muestras

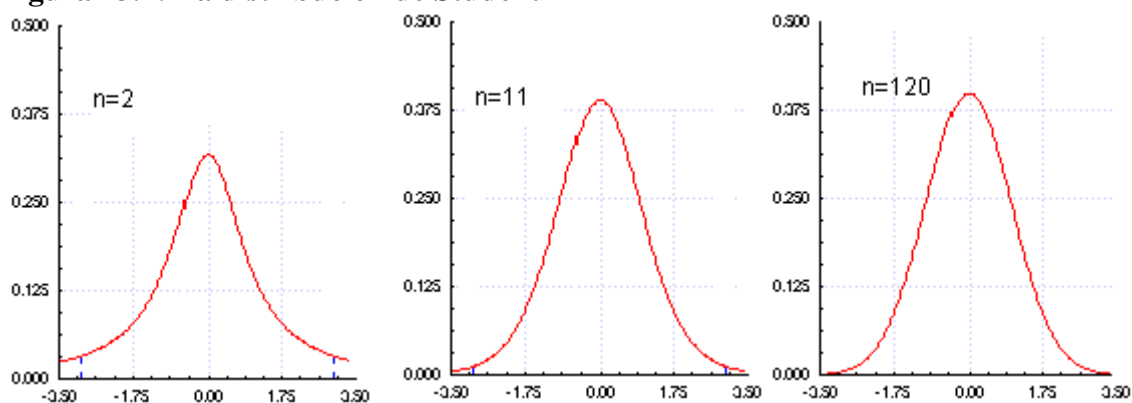
En este capítulo se presentan tres nuevos modelos estadísticos: el llamado t de Student, el modelo de la *Chi-cuadrado* (χ^2) y el modelo F de Fisher. Los tres no requieren ya más del supuesto de un tamaño muestral grande. Ahora con dos o más mediciones se puede trabajar; por eso se usa la expresión *Teoría de pequeñas muestras* para este tema. El empleo de cualquiera de ellos es enteramente similar al visto en el capítulo anterior. Cambia la manera de calcular el estadígrafo de comparación y su respectiva tabla de valores críticos de la distribución muestral. Mientras que el modelo de la t se aplica a medias y proporciones, los dos últimos se usan para el estudio de las desviaciones o dispersiones. También se la llama Teoría Exacta del Muestreo, pues ahora no hay que efectuar la aproximación $DS^2 \cong \sigma^2$ ya que el valor muestral viene en la fórmula de cálculo del estadígrafo de comparación, en lugar del poblacional. Eso hace que no sea necesario efectuar una estimación y se tiene una mayor exactitud que con la gaussiana. Es importante destacar que los tres modelos son válidos tanto para pequeñas como para grandes muestras. Esto amplía el campo de aplicación del modelo de Gauss. Además, al no tener que hacer tantas pruebas disminuye el costo y se gana en tiempo. Todas estas ventajas tienen una contrapartida: se pierde un poco de precisión pues, como se verá, el intervalo de confianza se hace más grande para un mismo caso. Estos modelos se prefieren al de Gauss porque sus ventajas valen la pena, al precio de perder un poco de precisión. Se mostrará su empleo tanto para el caso de una sola muestra de mediciones como para la comparación de dos muestras o grupos de mediciones.

13.1 El modelo de Student

Sea un estadígrafo t calculado para la media con la relación:

$$t = \frac{(\bar{x} - \mu)}{DS/\sqrt{n}}$$

Figura 13.1: La distribución de Student



Si de una población normal, o aproximadamente normal, se extraen muestras aleatorias e independientes y a cada una se le calcula dicho estadígrafo usando los valores muestrales de la media y el desvío estándar, entonces se obtiene una distribución muestral t que viene dada por la fórmula de Student. En realidad, fue obtenida por R. A. Fisher y la bautizó *Student* en honor a W. S. Gosset, quien usaba ese seudónimo para poder publicar sus trabajos en la revista *Biometrika*. Es-

ta función matemática tiene un parámetro que la define en forma unívoca: el número de grados de libertad $\nu = n - 1$ (donde n es el tamaño muestral). El concepto matemático de ν está relacionado con la cantidad de observaciones independientes que se hagan y se calcula con el tamaño muestral n , menos la cantidad k de parámetros poblacionales que deban ser estimados a través de ellas. O sea: $\nu = n - k$. Si se observa la ecuación superior, se ve que el único parámetro poblacional que figura es μ , por lo tanto $k = 1$ y así resulta $\nu = n - 1$. Cuando el tamaño muestral es mayor que 30 la distribución de Student se aproxima mucho a la de Gauss, en el límite ambas son iguales. Es decir que la función Student tiende asintóticamente a la función de Gauss.

Para cada grado de libertad hay una tabla de valores que pueden obtenerse variando el nivel de significación, parecida a la de Gauss. Sería muy engorroso tener una hoja con la tabla para cada grado de libertad. Esto se soluciona de dos formas: una es usando computadoras para resolver los cálculos (programas estadísticos como Mini-Tab, SPSS, Statistica, Excel, etc.). La otra y más común, es preparar una tabla donde en cada fila se coloquen encolumnados los valores críticos más usuales para cada valor de grados de libertad. Como interesan únicamente los valores pequeños, se listan correlativamente de 1 a 30 y luego algunos como 40, 60, 120 e ∞ . Este último tendrá los valores vistos para la normal. Así, en una sola hoja se presentan los valores útiles para el empleo de este modelo, como se muestran en el Tabla 5 del Anexo con las tablas estadísticas.

La distribución de Student, al igual que la de Gauss, es simétrica respecto al origen de coordenadas y se extiende desde $-\infty$ hasta $+\infty$. Pero a diferencia de la normal, puede adoptar diferentes formas dependiendo del número de grados de libertad. Por ejemplo, la que tiene un solo grado de libertad ($n = 2$ y $\nu = 1$), se desvía marcadamente de la normal, como se puede ver en la Figura 13.1 anterior. Luego, a medida que los grados van aumentando, se acerca cada vez más, hasta igualarla en el infinito. Se puede ver esto en las tablas y en la Figura 13.1. Los valores críticos de la Tabla Student, para una confianza del 95 % y dos colas, para 1, 5, 10, 30 y ∞ grados de libertad son 12,71; 2,57; 2,23; 2,04 y 1,96 respectivamente. Estos valores críticos se denotan con sus dos parámetros así: $t_{\alpha; \nu} = t_{0,05; \infty} = 1,96 = z_{\alpha}$.

Los intervalos de confianza para esta distribución se arman en forma análoga a la vista para el caso de Gauss. Con la única diferencia en cómo se calcula el valor crítico $t_{\alpha; \nu}$ en lugar de z_{α} .

$$\mu \in (\mu_e \pm t_{\alpha; \nu} SE(e)) = (\mu_e \pm t_{\alpha; \nu} \sigma_e) \quad \text{Intervalo de confianza con Student}$$

De nuevo, el par de valores (μ_e ; σ_e) se saca de la Tabla 4, con la salvedad que ahora no se usa más la aproximación $DS \cong \sigma$; pues en el cálculo de t se emplea DS directamente. Esto hace que el modelo sea más exacto que el de Gauss. Generalmente, este modelo se aplica al caso de la media, proporciones y sus diferencias o sumas. Para una estimación con 30 o más grados de libertad, se pueden usar tanto el modelo de Gauss, como el de Student. El intervalo es casi igual, salvo que en este último el valor crítico es mayor. En efecto, si se tienen 31 muestras, $t = 2,09$, mientras que $z = 1,96$. Esto hace mayores a los intervalos obtenidos con Student que sus equivalentes gaussianos. Por eso, se dice que el modelo Student tiene menor *precisión* que el de Gauss. La teoría de decisiones se usa en forma análoga, empleando los intervalos de confianza visto más arriba. Pero para poder aplicar este modelo se deben tener en cuenta los requisitos siguientes:

- 1) Las muestras fueron extraídas de una población normal o aproximadamente normal.

- 2) La selección de las muestras se hizo en forma aleatoria.
- 3) Las muestras son independientes entre sí.

Si alguno de ellos no se cumple, las conclusiones que se obtengan no son válidas. Los supuestos se pueden resumir así: *para poder usar Student, se deben tener muestras normales, aleatorias e independientes*. Notar que el error estándar de estimación es $SE(e) = \sigma_e$.

Los casos más frecuentes en la práctica son:

13.1.1 Student para medias muestrales

En este caso $e = \bar{x}$ luego: $\mu_e = \mu$ y $SE(e) = \sigma_e = DS / \sqrt{n}$. Por lo tanto el valor de comparación se calcula con:

$$t = \frac{(\bar{x} - \mu)}{DS/\sqrt{n}}$$

Ejemplo 1) Se desea saber si un instrumento de medición cualquiera está calibrado, desde el punto de vista de la exactitud. Para ello se consigue un valor patrón y se lo mide 10 veces (por ejemplo: una pesa patrón para una balanza, un suero control para un método clínico, etc.). Suponiendo que el resultado de estas mediciones arroja una media de 52,9 y un desvío de 3, usando un patrón de valor 50, se debe determinar si el instrumento está calibrado y la estimación de su error sistemático, si es que se prueba su existencia (no se usan unidades para generalizar este ejemplo).

- $H_0 : \mu = 50$ el instrumento está calibrado en exactitud
 $H_1 : \mu \neq 50$ no está calibrado. Hay un error sistemático

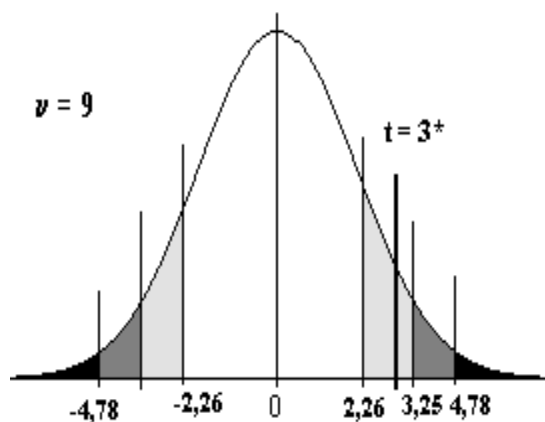
Se trata de un ensayo de dos colas donde hay $v = 10 - 1 = 9$ grados de libertad. De la Tabla 4 se obtienen los valores críticos para el 95% de $t_{0,05;9} = 2,262$, para el 99% de $t_{0,01;9} = 3,25$ y para un nivel del 99,9% es $t_{0,001;9} = 4,781$. Lo que permite establecer las zonas de aceptación y rechazo para testear al estadígrafo t:

$$t = \frac{(52,9 - 50)}{3/\sqrt{10}} = 3^* \quad (p=0,015)$$

$\mu = 50 \notin 95\% \text{ CI}(50,8 ; 55,1)$; o bien
 $\bar{x} = 52,9 \notin 95\% \text{ CI}(47,9 ; 52,1)$

Dibujando las zonas con los valores críticos, el valor de t cae en la de rechazo para el 95% y no alcanza para las otras. La conclusión es que se ha probado la existencia de un error sistemático con una confianza del 95%. Y se estima con:

$$ES \in (\bar{x} - \mu) = 2,90$$



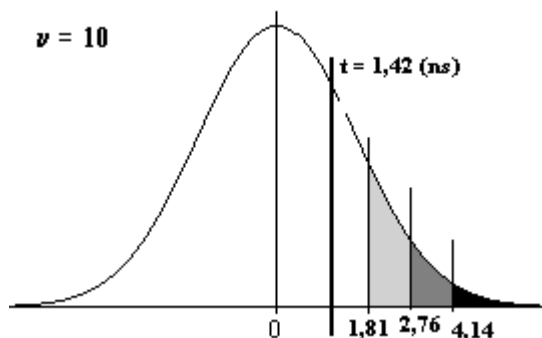
Ejemplo 2) Se midió colesterol total a 11 pacientes varones adultos escogidos al azar los resultados obtenidos arrojan una media de 235 mg/dl y un desvío estándar de 35 mg/dl. Ensayar la hipótesis de que se mantienen por debajo del valor límite de referencia (220 mg/dl).

$H_0 : \mu \leq 220$ mg/dl
 $H_1 : \mu > 220$ mg/dl

El valor de Student para una sola cola es:

$$t = \frac{(235 - 220)}{35 / \sqrt{11}} = 1,42 \text{ (ns)} \quad (p = 0,093)$$

Valor no significativo pues $t_{0,05; 10} = 1,81$
 $\mu = 220$ cae dentro de 95% CI (216 ; 254)



Para el caso de una cola, el valor de tablas para el 95% debe ser el que está en la Tabla 4 para el 90% en dos colas. La idea es que el 10% en dos colas significa el 5% en cada una, por la simetría de la curva de Student. Luego, para $v = 10$, el límite para el 95% será $t = 1,812$ en una cola y $t = 2,228$ para dos colas. En la figura de más arriba se han marcado los límites del 99% y del 99,9% para una sola cola, a los efectos didácticos. La conclusión es que no puede rechazar la hipótesis nula, por lo que debe considerarse un colesterol total admisible desde el punto de vista clínico, por estar por debajo del límite de referencia.

13.1.2 Student para proporciones

En este caso $e = P$ y $\mu_p = \mu = \pi$ luego con $SE(e) = \sigma_\pi = \sqrt{\pi \cdot (1 - \pi) / n}$ se puede obtener el valor del estadígrafo de comparación con la relación:

$$t = \frac{(P - \pi)}{\sqrt{\pi \cdot (1 - \pi) / n}}$$

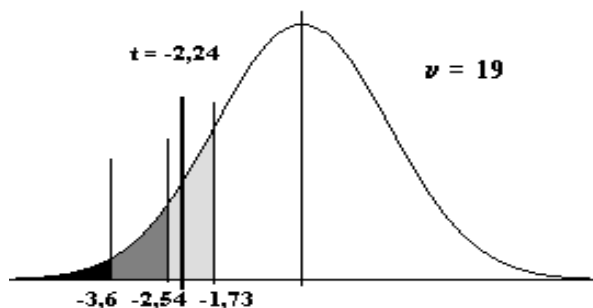
Ejemplo1) Un analgésico de plaza, afirma en su propaganda que alivia el dolor en el 90% de los casos antes de la primer hora luego de su ingesta. Para validar esa información, se hace un experimento en 20 individuos con cefalea. Se observa que fue efectivo en 15 de ellos.

$H_0 : \mu \geq 0,9$ La afirmación es correcta
 $H_1 : \mu < 0,9$ La afirmación es falsa

Es un ensayo de una sola cola (a izquierda)
 El porcentaje de éxitos es: $P = 15 / 20 = 0,75$.
 Con $\mu = 0,9$ y $\sigma_p = \sqrt{0,9 \cdot 0,1 / 20} = 0,067$

$$t = \frac{(0,75 - 0,9)}{0,067} = -2,24^* \quad (p = 0,019)$$

$\mu = 0,9$ cae fuera del CI 95% $(-\infty ; 0,87)$



De tablas: $t_{0,999; 19} = -3,579$ $t_{0,99; 19} = -2,539$
 y $t_{0,95; 19} = -1,729$

El resultado obtenido es significativo ($t = -2,24^*$). Pero la evidencia no alcanza para rechazar la hipótesis a los niveles del 99% y 99,9%. Se la rechaza al nivel de 95% únicamente. Si bien no es tan terminante, se puede afirmar que la aseveración es falsa con un 95% de confianza.

13.2 Student para dos muestras independientes

El modelo de Student también se puede usar cuando se desean comparar dos muestras entre sí, para detectar si hay diferencia significativa entre ellas, debido a algún factor analizado. En primer lugar se analizará el caso de dos muestras independientes como: aplicar dos tipos de remedios a dos grupos de pacientes escogidos al azar, o las mediciones repetidas de una misma magnitud, etc. El otro caso, cuando las muestras no son independientes sino apareadas, se verá en el próximo tema. Una vez más, los supuestos para poder aplicar este modelo se resumen en: *para poder comparar con Student, las dos muestras deben ser normales, aleatorias e independientes.*

Se sacan muestras aleatorias e independientes, de dos poblaciones normales. La idea es averiguar si ambas muestras provienen de la misma población o de poblaciones diferentes. Con eso se puede ver si el efecto de los “tratamientos” aplicados a las muestras es *apreciable*, en cuyo caso las muestras parecerán provenir de diferentes poblaciones. Se usa en los casos donde se compara el efecto de una droga aplicada a un grupo de pacientes, contra otro grupo al cual se le suministra un placebo. También para comparar dos técnicas clínicas y detectar si hay diferencias, por ejemplo: dos marcas comerciales de plaza, dos instrumentos de medición, dos individuos, dos técnicas diferentes (la nueva contra la vieja), dos protocolos, etc. Con estas comparaciones se pueden realizar muchos controles internos en el laboratorio para hacer calibraciones, medir eficacia, etc. Hay una limitación: *solo se pueden comparar dos muestras entre sí a la vez y no más.* Para el caso de tener más de dos muestras, se recurre a los modelos de ANOVA como se verá más adelante.

13.2.1 Comparación de medias

Para estos casos, el valor de Student para validaciones de medias se calcula con:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{DS_1^2}{n_1}\right) + \left(\frac{DS_2^2}{n_2}\right)}}$$

El cual se contrasta con $t_{\alpha; \nu}$ donde $\nu = n_1 + n_2 - 2$ grados de libertad. Hay casos particulares como (a) las muestras son de igual tamaño y (b) son *homocedásticas* (tienen igual varianza). En ambos casos se simplifican las fórmulas de cálculo.

Ejemplo 1) Se aplica un medicamento a 15 pacientes que padecen cierta enfermedad, escogidos al azar, y un placebo a 20 pacientes. En el primer grupo, la desaparición del estado febril se observa a las 19 horas de tratamiento en promedio (con un desvío de 2 hs.). En el grupo control, la mejoría se observa en promedio a las 25 horas con un desvío de 3 horas. Decidir si el medicamento modifica el tiempo de curación.

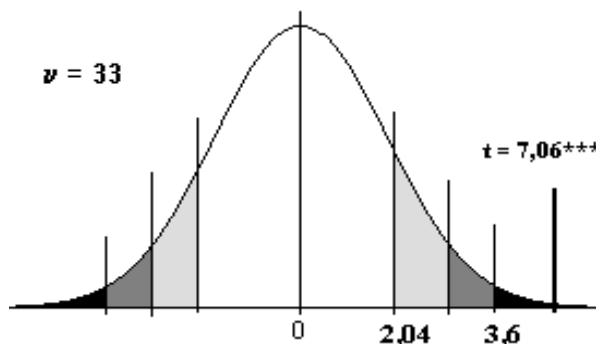
$H_0 : \mu_1 = \mu_2$: el medicamento es inocuo

$H_1 : \mu_1 \neq \mu_2$: el medicamento produce efecto

Es un ensayo de dos colas donde los valores críticos se buscan en la Tabla 5 interpolando entre 30 y 40 grados de libertad.

$$t = \frac{(25 - 19) - (0)}{\sqrt{\left(\frac{9}{20}\right) + \left(\frac{4}{15}\right)}} = 7,06^{***}$$

$\mu_1 - \mu_2 = 0$ cae fuera de CI 99,9% (3,1; 8,9).



Como el valor hallado de t es mucho más grande que el valor crítico de tablas para 33 grados de libertad: $t_{\alpha; v} = t_{0,999; 33} = 3,44$ (ensayo de dos colas y un 99,9% de confianza), la conclusión es: se obtuvieron resultados altamente significativos ($t = 7,06^{***}$) como para rechazar la hipótesis nula. Se tiene una prueba científica del efecto del medicamento.

Ejemplo 2) Se desea verificar si hay diferencia en las mediciones a través de dos métodos clínicos diferentes. Se toma una muestra de suero lo suficientemente grande como para obtener 10 alícuotas. Se distribuyen al azar 5 alícuotas para cada método. Efectuadas las mediciones, con el primero se tuvo una media de 85 mg/dl con un desvío de 8 mg/dl. Mientras que con el segundo se tuvo una media de 83 mg/dl con un desvío de 6 mg/dl.

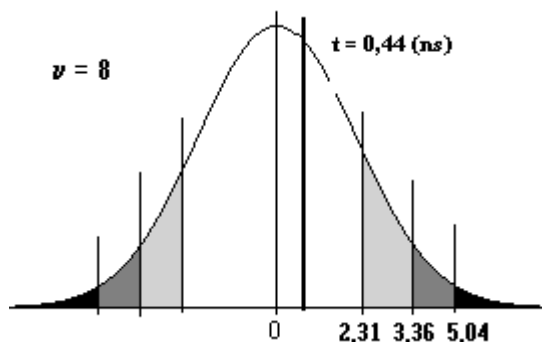
$H_0 : \mu_1 = \mu_2$: no hay diferencia.

$H_1 : \mu_1 \neq \mu_2$: hay diferencia

$$t = \frac{(85 - 83) - (0)}{\sqrt{\left(\frac{64}{5}\right) + \left(\frac{36}{5}\right)}} = 0,44 \text{ (ns)}$$

No se puede rechazar la H_0 . Se concluye que no hay diferencia entre ambos métodos.

$\mu_1 - \mu_2 = 0$ cae dentro de 95% CI(-8,3 ; 12,3)



13.2.2 Comparación de proporciones

Para estos casos, el valor de Student para validaciones de proporciones se calcula la misma fórmula, pero reemplazando los valores esperados con:

$$\mu_{1-2} = (\pi_1 - \pi_2) \quad \text{y} \quad \sigma^2_{1-2} = [\pi_1 (1-\pi_1) / n_1] + [\pi_2 (1-\pi_2) / n_2]$$

Entonces el valor de comparación del modelo Student para este caso es:

$$t = \frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{(\pi_1 (1 - \pi_1) / n_1) + (\pi_2 (1 - \pi_2) / n_2)}}$$

Contrastando con el valor de tablas dado por $t_{\alpha; v}$, con $v = n_1 + n_2 - 2$ grados de libertad.

Ejemplo) Se escogen al azar dos grupos formados por 20 individuos cada uno, entre los que padecen cierta alergia. Se administra una droga curativa al primer grupo y se observa una mejoría en 15 de los casos. Al segundo grupo se le administra un placebo y mejoran 13 de ellos. Ensayar la hipótesis que la droga sirve para curar ese tipo de alergia. Se emplean las hipótesis siguientes:

$H_0 : \mu_{1-2} = 0$ las diferencias observadas se deben al azar.

$H_1 : \mu_{1-2} \neq 0$ la droga produce efecto.

Si se supone que ambas muestras fueron extraídas de la misma población, y por lo tanto no hay diferencias entre las muestras observadas (H_0) $\mu_{1-2} = 0$, eso significa que el porcentaje de curados en dicha población será $\pi = \pi_1 = \pi_2$ y habrá que estimarlo con los datos muestrales, calculando la proporción ponderada con: $p = (\text{total de curados en las muestras} / \text{total muestral}) = (15+13) / 40 = 0,7$. Entonces, sacando factor común en la fórmula de la varianza, esta resulta:

$$SE^2(\pi) = \pi (1-\pi) \cdot [(1/n_1) + (1/n_2)] = \pi (1-\pi) [2/n] = (0,7 \cdot 0,3) (2/20) = 0,021$$

Y es $SE(\pi) = 0,145$; de los datos del problema surgen $P_1 = 15/20 = 0,75$ y $P_2 = 13/20 = 0,65$

$$t = (0,75 - 0,65) / (0,021)^{1/2} = 0,69 < t_{0,95; 38} = 2,02. \mu_{1-2} = 0 \text{ cae dentro de } 95\% \text{ CI}(-0,19 ; +0,39)$$

Un resultado no significativo. Las diferencias observadas no se deben a la droga sino al azar.

13.2.3 Test de equivalencia biológica

Hay ocasiones donde la H_0 no busca establecer si hay o no diferencia entre dos muestras, como las del ejemplo anterior, sino que se trata de establecer si un método clínico o tratamiento nuevo es lo suficientemente bueno como para reemplazar al que se venía usando hasta entonces, el método viejo. Las ventajas de este nuevo método pueden ser: un costo menor, más rápido, menos dañino o peligroso para el paciente, etc. La cuestión básica aquí es ver si, en promedio, la diferencia entre ambos es menor que un cierto valor límite para la magnitud estudiada. Es decir que tal diferencia no implique una inferioridad del nuevo método, desde un punto de vista clínico. Para estos casos la H_0 : La diferencia entre ambos promedios es mayor o igual al valor aceptable y la alternativa es H_1 : Esta diferencia de medias es menor al valor crítico; en cuyo caso ambos métodos pueden ser considerados clínicamente equivalentes. La idea es que, si se rechaza la H_0 se puede usar el método nuevo en lugar del viejo y aprovechar las ventajas que este posee. Pero la decisión se basa más en consideraciones médicas que estadísticas. Entonces, si se trata de magnitudes continuas, se puede usar el test de Student para comparar la diferencia de las dos medias contra el valor crítico δ o máximo aceptable desde el punto de vista clínico. El planteo se hace así: $H_0 : \mu_V - \mu_N = \Delta \geq \delta$. Donde μ_V es el valor poblacional que se obtiene con el método viejo y μ_N con el método nuevo, Δ es la diferencia real entre ambos métodos y δ es la diferencia máxima admisible entre ambos métodos. De esta manera, cuando H_0 pueda rechazarse se tendrá evidencia suficiente como para efectuar el reemplazo, esto es cuando $H_1 : \mu_V - \mu_N = \Delta < \delta$.

Se trata de un ensayo de una sola cola. Pero cuando se trate de ver si en valor absoluto la diferencia entre ambos métodos no supere a un cierto valor δ , porque aquí no interesa tanto que sea menor, sino que también interesa que no sea mayor (dependiendo de la magnitud clínica analizada); entonces la H_0 será : $\mu_V - \mu_N = \Delta = \delta$ y el ensayo será de dos colas. Análogo al visto en el punto anterior. Para ilustrar este procedimiento se usará un ejemplo tomado de la obra de Armitage (pág. 320).

Ejemplo) Sea el índice cardíaco CI (respuesta cardiaca normalizada para la superficie del cuerpo) el cual se mide con un procedimiento invasivo como es el colocar un catéter en el corazón del paciente llamado Termo-dilución (el método viejo) y la unidad de medición son litros por minuto tomado por m^2 de superficie del cuerpo humano. Se ha propuesto una nueva manera de medir esa magnitud con una técnica no invasiva, llamada el método de la Bioimpedancia, en la cual se le adosa un instrumento al cuerpo de paciente en forma externa, y mide en forma eléctrica el valor del CI usando una escala adecuada (el método nuevo). El criterio clínico de aceptación es: el nuevo método se considerará equivalente al viejo cuando, en promedio, el valor obtenido difiera en un 20% respecto al promedio aceptado de 2,75 l / min. / m^2 para el método del catéter. Esto significa que el 20% de tal valor es $\delta = 0,55$. Luego el planteo se hace así:

$$H_0 : |\mu_V - \mu_N| = |\Delta| \geq \delta = 0,55 \text{ o lo que es lo mismo } (\mu_V - \mu_N) = \delta = 0,55$$

$$H_1 : |\mu_V - \mu_N| = |\Delta| < \delta = 0,55 \text{ cuyo equivalente es } (\mu_V - \mu_N) = \delta \neq 0,55$$

Se toma una muestra de $N = 96$ individuos a los cuales se le aplica el método nuevo, los valores encontrados fueron un promedio de 2,68 l / min. / m^2 , y un desvío estándar de 0,26 l / min. / m^2 luego será:

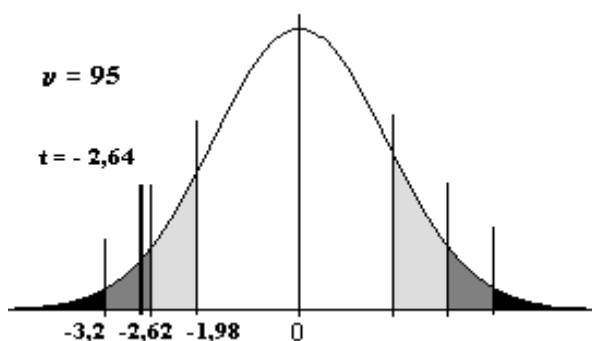
$$t = \frac{(\bar{x} - \mu)}{DS/\sqrt{n}} = (2,68 - 2,75) / (0,26 / \sqrt{96}) = - 0,07 / 0,0265 = - 2,642^{**} > t_{0,99; 95} = - 2,63$$

Lo que indica que hay evidencia significativa como para rechazar la H_0 .

Otra manera de ver lo mismo es cuando se usa el valor límite de confianza, para la diferencia en valor absoluto, la cual es: 0,07. Para un 95% de confianza el error de estimación se calcula con:

$$t_{0,95; 95} \cdot (DS / \sqrt{N}) = 0,053$$

Luego $0,07 + 0,053 = 0,123 < 0,55$ con lo que se puede rechazar la H_0 .



La conclusión final es que se puede usar el método nuevo en lugar del viejo, con una gran ventaja para el paciente, pues ahora ya no tendrá que ser cateterizado para efectuarle su medición del índice cardíaco. A este procedimiento estadístico aparecido en los últimos años en Medicina se lo conoce también con el nombre de *test de equivalencias médicas o biológicas*.

13.3 Student para dos muestras apareadas

El modelo de Student se puede usar para el caso especial de muestras *apareadas*, esto es, cuando se le efectúan dos tratamientos a la misma muestra; por ejemplo, del tipo *antes-después* donde al mismo individuo se lo mide dos veces para ver el efecto del tratamiento realizado, o el caso de método nuevo contra el método viejo, donde al mismo grupo de pacientes se le hacen dos mediciones a cada uno, la del método de rutina habitual y una extra con el nuevo método a probar para decidirse entre ambos. La idea básica es como sigue: se sacan n muestras aleatorias e independientes de una población normal. A cada muestra se le aplican dos “tratamientos” A y B diferentes y lo que interesa detectar es si producen algún efecto apreciable. Este caso es muy diferente al anterior (13.2), si bien las muestras son independientes entre sí, los tratamientos no lo son, porque a un mismo individuo se le aplican ambos tratamientos. Entonces, la misma persona aparecerá dos veces en los resultados: uno en el grupo A y el otro en el grupo B. El truco para resolver este problema de la independencia es trabajar con la diferencia de los resultados de cada par de mediciones efectuadas: $d = x_A - x_B$. Luego se tendrán n diferencias $d_1; d_2; d_3...d_n$, que son independientes entre sí, puesto que cada valor d_i corresponde a un solo individuo. Luego, se le aplica el modelo Student para una sola muestra, ensayando la hipótesis de que no hay diferencias entre ambos grupos. O sea, efectuando la hipótesis: $H_0 : \mu_d = 0$ resultará:

$$t = \frac{\bar{d} - 0}{DS_d / \sqrt{n}}$$

La hipótesis alternativa implica un efecto diferente para cada grupo $H_1 : \mu_d \neq 0$. Si se prueba que el valor esperado del promedio de las diferencias es diferente de cero, entonces el tratamiento aplicado produce un efecto demostrable. Para aclarar estas ideas se presenta el siguiente caso:

Ejemplo) Se escogen 5 pacientes al azar, del grupo que concurre diariamente al Laboratorio de Análisis Clínicos a efectuarse una determinación de Uremia. Las muestras extraídas se miden con el procedimiento habitual y además con una nueva técnica clínica que se desea probar. Ver si hay diferencia entre ambas técnicas. Los resultados expresados en g/l fueron:

Pacientes	Vieja	Nueva	Diferencias
1	0,38	0,33	0,05
2	0,54	0,45	0,09
3	0,22	0,15	0,07
4	0,11	0,09	0,02
5	0,23	0,22	0,01

Promedio = 0,048
 Desvío estándar = 0,033

Con los valores de las diferencias se calculan $\bar{d} = 0,048$ g/l y $DS_d = 0,033$ g/l. Luego:

$$t = \frac{0,048}{0,033 / \sqrt{5}} = 3,25^* > t_{0,95; 4} = 2,776. \text{ O bien, el valor } 0 \text{ cae fuera del CI } 95\% (0,01; 0,09)$$

Se obtuvo evidencia significativa $t = 3,25^*$ que hay diferencia entre ambas técnicas.

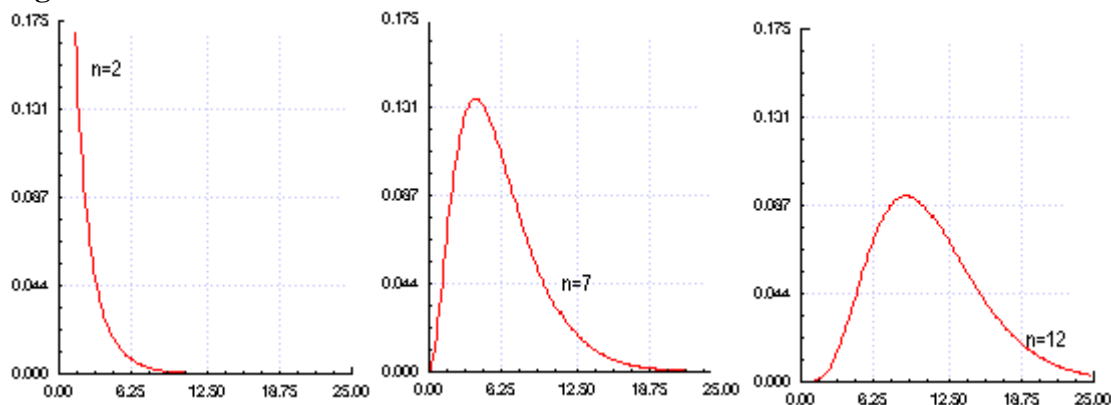
13.4 El modelo de Chi-cuadrado

La función Chi-cuadrado es igual a la función normal elevada al cuadrado. Esto es, el producto de dos distribuciones de Gauss es una distribución de Chi-cuadrado. Si de una población normal, o aproximadamente normal, se extraen muestras aleatorias e independientes, y se le calcula el estadígrafo χ^2 usando el valor muestral de la varianza y el poblacional con:

$$\chi^2 = (n - 1) DS^2 / \sigma^2$$

La distribución muestral de χ^2 viene dada por la fórmula de K.R. Pearson. Esta función matemática está caracterizada por el valor del número de grados de libertad $\nu = n - 1$ (donde n es el tamaño muestral). Al igual que la Student, el valor total del área bajo la curva es igual a la unidad, pero la diferencia principal es que esta no es simétrica respecto al origen, sino que se extiende desde 0 hasta $+\infty$ porque no puede ser negativa. A medida que los grados de libertad aumentan, la curva cambia de forma y sus valores se han tabulado en el anexo de tablas estadísticas (Tabla 6), donde se muestran los valores del área bajo la curva, para los principales valores de χ^2 , a la derecha de éste. O sea, se muestra la zona de rechazo para diferentes niveles de significación y de grados de libertad, lo cuales varían entre 1 y 100. Más allá, conviene usar directamente la función de Gauss.

Figura 13.2: La distribución de Chi-cuadrado



Para cada grado de libertad hay una tabla de valores que pueden obtenerse variando el nivel de significación, parecida a la de Gauss. El problema de calcular los valores críticos, para un nivel de confianza dado, se resuelve de dos maneras: usando computadoras para resolver los cálculos, y la otra más común, usando tablas resumidas en una sola hoja como la que se muestra en la Tabla 6, en forma análoga a la vista para el modelo de Student.

La distribución de χ^2 se usa principalmente para analizar dispersiones. Se compara la dispersión muestral expresada a través de sus cuadrados medios (MS) contra la dispersión poblacional cuantificada a través de la varianza (σ^2). El valor $MS = (n - 1) DS^2 = \nu DS^2$ es otra forma de mostrar la precisión del sistema de medición. El uso más difundido en Bioquímica es para controlar la dispersión de la técnica de Análisis Clínicos empleada en el Laboratorio; se compara

la obtenida en forma experimental, contra un valor considerado como aceptable en los libros de texto denominado la *dispersión máxima admisible* ($\sigma_{\text{máx}}$). Usualmente, se toma el CV%, como se mostró en el ejemplo de varianzas visto en el capítulo anterior. Existen otros criterios, como el de Thonks, que usa un error relativo admisible máximo, y se calcula como un cuarto del rango de los valores normales de referencia, dividido por el valor medio de dicho intervalo (referido a la magnitud clínica en cuestión y expresado en porcentajes). Todo esto se puede ver con mayor detalle en la Tabla 24.1 del Capítulo 24. También se emplea a este modelo para realizar la llamada *prueba de chi-cuadrado* en las comparaciones de frecuencias observadas contra las frecuencias esperadas, con datos de recuento. Más adelante se desarrolla mejor este tema, lo mismo que su uso para testear la independencia de dos o más factores en una Tabla de Contingencia.

En la industria farmacéutica se la usa para analizar la dispersión de los componentes de los productos terminados. Todo remedio fabricado debe cumplir estrictas normas de calidad, generalmente referidas al contenido en peso de sus principales componentes. Se usan dos límites: el superior e inferior, dentro de los cuales se los debe mantener controlados. Este rango de valores define la dispersión máxima admisible y lo ideal es que la dispersión de los productos terminados sea bastante inferior a dicho rango. Ese control de la dispersión es muy similar al explicado más arriba, para los bioquímicos. Para ilustrar estas ideas se presenta un ejemplo en Control de Calidad de equipos de Laboratorio. Pero antes se debe remarcar que para poder aplicar este modelo, se deben tener en cuenta los requisitos siguientes:

1. Las muestras fueron extraídas de una población normal o aproximadamente normal.
2. La selección de las muestras se hizo en forma aleatoria.
3. Las muestras son independientes entre sí.

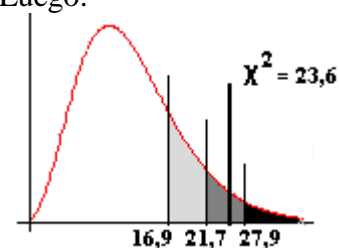
Ejemplo 1) Un bioquímico sospecha que su micro-centrífuga no mantiene constante su velocidad mientras trabaja, lo cual le da una variabilidad indeseada en sus determinaciones. Para controlarla, consigue un tacómetro regulado y mide cada minuto la velocidad durante 10 minutos. Los resultados fueron: una velocidad promedio en las 10 mediciones de 3098 rpm con un desvío de 100,4 rpm. Testear para un error relativo máximo del 2% o menos, si la centrífuga es estable.

El desvío estándar aceptable es: $\sigma_{\text{máx}} = 2\%$ de 3098 rpm = 62 rpm. Luego:

$H_0 : \sigma_{\text{máx}} \leq 62$ rpm: la micro centrífuga es estable
 $H_1 : \sigma_{\text{máx}} > 62$ rpm: la micro centrífuga no es estable

$$\chi^2 = (n - 1) DS^2 / \sigma^2$$

$$\chi^2 = (10 - 1) (100,4)^2 / (62)^2 = 23,6^{**}$$



De la Tabla de valores críticos surge: $\chi^2_{0,99 ; 9} = 21,666$ y $\chi^2_{0,991 ; 9} = 27,877$. Por lo tanto, el bioquímico ha encontrado una muy fuerte evidencia que la velocidad del equipo oscila en forma indeseada, tal como sospechaba. Y deberá ajustarlo si desea disminuir la variabilidad de sus mediciones. Los resultados fueron muy significativos $\chi^2 = 23,6^{**}$.

Ejemplo 2) Un farmacéutico Jefe del Dpto. Control de Calidad en una industria alimenticia, descubre que en su proceso de producción el contenido de ciclamato en su línea de mermeladas die-

téticas varía en forma indeseada. Sospechando que se trata de una falla en el dosificador, decide tomar 10 muestras seguidas del mismo. Encuentra un promedio de 20 gramos con un desvío de 8 gramos. Si en su protocolo de fabricación la variación máxima permitida es del 3%, determinar si el dosificador debe ser corregido.

El desvío estándar aceptable es: $\sigma_{\text{máx}} = 3\%$ de 20 g = 6 g. Luego:

$H_0 : \sigma_{\text{máx}} \leq 6$ g.: el dosificador funciona correctamente

$H_1 : \sigma_{\text{máx}} > 6$ g.: el dosificador debe ser cambiado

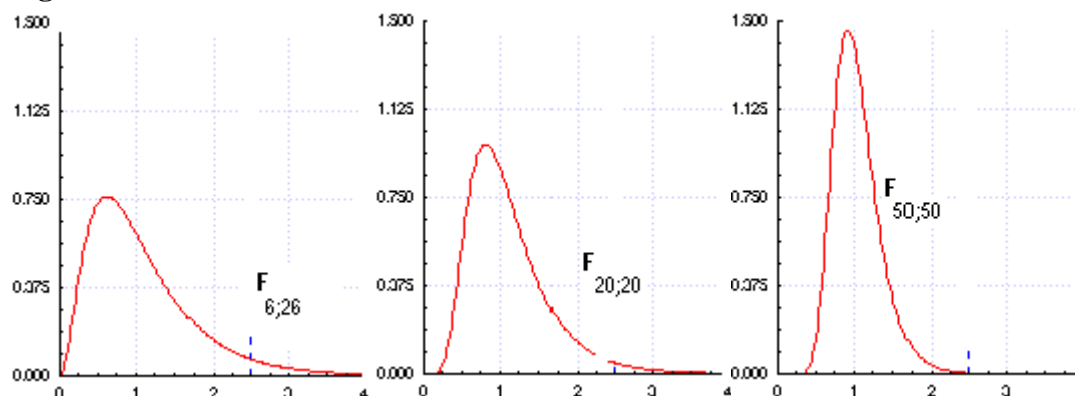
$$\chi^2 = (n - 1) DS^2 / \sigma^2 = (10 - 1) \cdot (8)^2 / (6)^2 = 16 \text{ (no significativo)}$$

De la Tabla de valores críticos surge: $\chi^2_{0,95;9} = 16,9$. Por lo tanto, el farmacéutico no ha encontrado evidencia que respalde sus sospechas. Sin embargo, el valor hallado es muy cercano al crítico, por lo que le convendría hacer más pruebas.

13.5 El modelo de Fisher

Si de dos poblaciones normales, o aproximadamente normales, se extraen dos muestras aleatorias e independientes, y a cada una se le calcula su respectiva varianza, el cociente de ambos valores $F = DS^2_1 / DS^2_2$ (con $F > 1$, esto es, siempre se coloca el más grande como numerador) tendrá una distribución de Fisher, cuyos valores críticos fueron obtenidos por W. Snedecor y se muestran en la Tabla 7 del anexo. Esta tabla se caracteriza por tener dos grados de libertad: el correspondiente al numerador $\nu_1 = n_1 - 1$ y el del denominador $\nu_2 = n_2 - 1$. Programas de computación permiten calcular los valores críticos respectivos. En otra forma se puede usar una tabla de doble entrada como la Tabla 7; su forma se puede ver a continuación:

Figura 13.3: La distribución de Fisher.

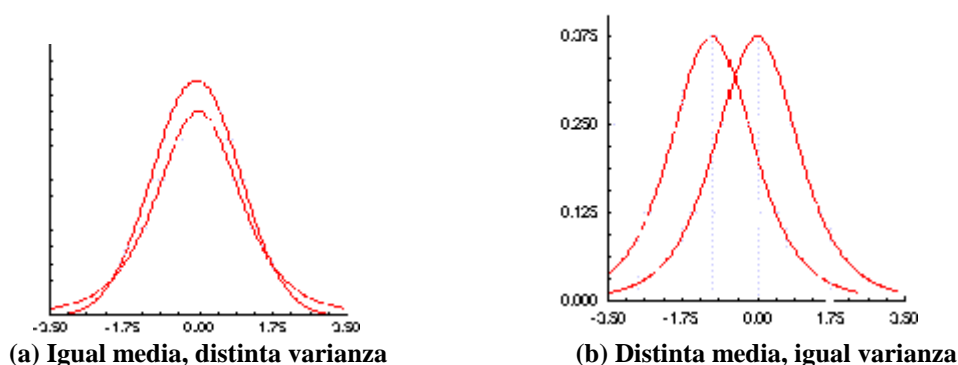


En las Tabla 7 se presenta una hoja para cada nivel de confianza, se eligen los más apropiados como: 95% ; 97,5% ; 99% ; 99,5% y 99,9%. Como siempre, el área total bajo la curva es la unidad y se extiende desde 0 a $+\infty$. La forma es muy parecida a la Chi-cuadrado. En la Figura 13.3 se muestran tres casos, con diferentes grados de libertad, y se marca el valor de $F = 2,5$ con una línea punteada vertical.

El principal uso de esta función es el Análisis de Varianza, que se verá más adelante, y es para cuando se necesita comparar más de dos medias muestrales a la vez. En estos casos la idea es detectar si el efecto de uno o más tratamientos afecta a las muestras testeadas. En cambio, cuando se tiene el caso de dos muestras, la idea es testear si hay homoscedasticidad en las dos poblaciones en estudio. Una vez verificado este supuesto, se puede avanzar más verificando si hay diferencia entre las medias muestrales, y así verificar si ambas muestras tienen igual media y varianza, porque eso significa que en realidad provienen de la misma población normal. Eso probaría que no hay efecto de un tratamiento si se lo compara con un placebo, o que dos técnicas de laboratorio son equivalentes. Si el experimento no verifica esto, entonces se deberá elegir el caso que presente menor varianza, para tener menor variabilidad en las mediciones. En Genética se puede verificar si una generación de crías es más variable en un carácter que la de sus padres. En Sistemática se puede testear si dos poblaciones locales tienen la misma variabilidad. En Bioquímica y Farmacia el uso más frecuente es comparar el error casual de mediciones de laboratorio, al introducir algún efecto o cambiar el método de medición.

En el caso de testear si dos técnicas de laboratorio tienen igual dispersión, o bien, para elegir aquella con mayor precisión, conviene pensar el problema como la incidencia de un factor en estudio en lugar de dos técnicas totalmente diferentes entre sí. Por ejemplo, se trata de una misma práctica, pero se usan dos espectrofotómetros diferentes, y se trata de determinar si la modificación de la varianza se debe al uso de un aparato diferente. El factor acá sería: tipo de espectros. También se puede estudiar la incidencia del factor humano, realizando las mismas mediciones a dos personas diferentes. De esa forma se puede imaginar que las dos muestras provienen de diferentes poblaciones, o que el efecto del factor analizado no es despreciable cuando se rechaza la hipótesis nula. En la Figura 13.4 se muestra el caso de dos poblaciones. En el caso (a) ambas poblaciones tienen la misma media, pero por efecto del error casual sus varianzas son diferentes. Si esta diferencia es significativa, resulta evidenciada por el Modelo de Fisher que permite la comparación de ambas.

Figura 13.4: Dos poblaciones con



En el caso (b) hay un error sistemático que desplaza la media, pero sus varianzas permanecen iguales. Es lo mismo que sumar una constante a todos los valores; ocurre un desplazamiento hacia la derecha. Student se usa para detectar esto cuando se hace el test de comparación de dos medias independientes.

Como se verá más adelante, se puede construir todo un bagaje de métodos para efectuar un Control de Calidad interno en un laboratorio de medición clínica. Por ahora, basta decir que se puede controlar la exactitud con los modelos de Student y la precisión con los de Chi-cuadrado y Fisher. Con esto se pueden comenzar a controlar y calibrar los sistemas de medición. Las limitaciones de todo esto son dos: la primera es que se puede estudiar el efecto del factor analizado en solo dos muestras y no en más de dos. La segunda es que si la calidad se entiende como exactitud y precisión, solo se pueden emplear estos modelos para magnitudes de tipo cuantitativas como las de la Química Clínica, pero no en magnitudes cualitativas como las usuales en Microbiología, Bacteriología, Micología, etc. En magnitudes cuantitativas, por "calidad" se entiende precisión y exactitud, en lugar de la capacidad de una prueba clínica para diagnosticar. Sin embargo, a pesar de estas limitaciones sigue siendo una herramienta sencilla y poderosa de control.

Para poder aplicar este modelo se deben tener en cuenta los requisitos siguientes:

1. Las muestras fueron extraídas de una población normal o aproximadamente normal.
2. La selección de las muestras se hizo en forma aleatoria.
3. Las muestras son independientes entre sí.

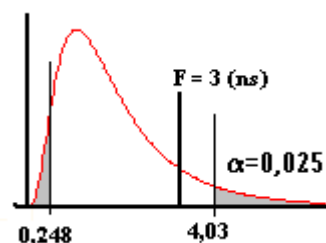
Ejemplo) El jefe de un laboratorio se encuentra con una técnica de medición fuera del control estadístico. Para investigar las causas decide investigar si el factor humano tiene incidencia, y toma una muestra de suero cualquiera la divide en 20 alícuotas. Luego elige 10 de ellas al azar y se las entrega al laboratorista 1 para que haga las determinaciones; las restantes las encomienda al laboratorista 2 para que las mida. Los resultados obtenidos son: $DS^2_1 = 2,4$ es la varianza obtenida por el laborista, 1 y $DS^2_2 = 0,8$ para el otro. Decidir si hay diferencia en dispersión entre ambos.

$H_0 : \sigma^2_1 = \sigma^2_2$: no hay diferencia y el factor humano no incide

$H_1 : \sigma^2_1 \neq \sigma^2_2$: hay diferencia entre ambas personas

Se calcula el estadígrafo de comparación de Fisher

$$F = DS^2_1 / DS^2_2 = 2,4 / 0,8 = 3$$



Como se trata de un ensayo de dos colas, para un nivel del 95% de confianza, se busca en las tablas para: $\nu_1 = \nu_2 = n_1 - 1 = 9$ grados de libertad, mientras que $\alpha = 0,025$ para el límite inferior y $\alpha = 0,975$ para el superior. Estos valores son:

$$F_{0,975; (9,9)} = 4,03.$$

Luego, para calcular el valor no tabulado $\alpha = 0,025$ se aprovecha una propiedad que tiene la función F usando la inversa como sigue:

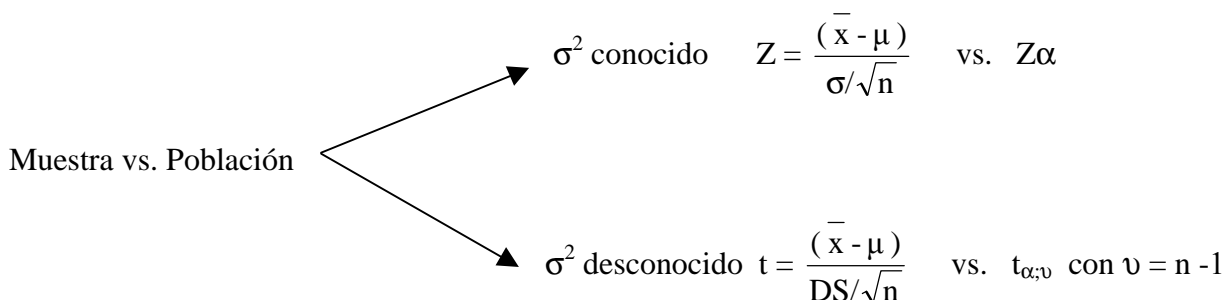
$$F_{0,025; (9,9)} = 1 / F_{0,975; (9,9)} = 1 / 4,03 = 0,248$$

Como el valor hallado $F = 3$ cae dentro de la zona de aceptación, no hay evidencia significativa como para decir que el factor humano tiene incidencia en la dispersión de las mediciones.

13.6 Cuadro resumen

Test de hipótesis para las medias muestrales.

Se puede siempre usar el modelo Student. El de Gauss cuando la muestra sea grande ($n > 30$)



Muestra vs. Muestra:

Caso 1: Muestras independientes con σ_1^2 y σ_2^2 son desconocidos

$$\text{Si } \sigma_1^2 \neq \sigma_2^2 \text{ es } t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{DS_1^2}{n_1}\right) + \left(\frac{DS_2^2}{n_2}\right)}} \text{ vs. } t_{\alpha;v} \text{ con } v = \frac{[(DS_1^2/n_1) + (DS_2^2/n_2)]^2}{\frac{(DS_1^2/n_1)^2}{n_1-1} + \frac{(DS_2^2/n_2)^2}{n_2-1}}$$

$$\text{Si } \sigma_1^2 = \sigma_2^2 \text{ es } t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{(n_1-1)DS_1^2 + (n_2-1)DS_2^2}{n_1 + n_2 - 2}}} \text{ vs. } t_{\alpha;v} \text{ con } v = n_1 + n_2 - 2$$

Caso 2: Muestras independientes con σ_1^2 y σ_2^2 son conocidos

$$\text{Si } \sigma_1^2 \neq \sigma_2^2 \text{ es } Z = [(\bar{x}_1 - \bar{x}_2) - (\mu_{\bar{x}_1} - \mu_{\bar{x}_2})] / \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \text{ vs. } Z\alpha$$

$$\text{Si } \sigma^2 = \sigma_1^2 = \sigma_2^2 \text{ es } Z = [(\bar{x}_1 - \bar{x}_2) - (\mu_{\bar{x}_1} - \mu_{\bar{x}_2})] / \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \text{ vs. } Z\alpha$$

Caso 3: Las muestras son apareadas y se suponen de una misma población con $DS_d \approx \sigma^2$

Se toma $d_i = X_{1i} - X_{2i}$ entonces $t = \frac{\bar{d} - 0}{DS_d / \sqrt{n}}$ vs. $t_{\alpha;v}$ con $v = n - 1$

Test de hipótesis para proporciones

Se puede siempre usar el modelo Student. El de Gauss cuando la muestra sea grande ($n > 25$)

Muestra versus población: π conocido o estimado con $p = r / n$

Gauss: $Z = \frac{p - \pi}{\sqrt{\pi \cdot (1 - \pi) / n}}$ o $Z = \frac{r - n \cdot \pi}{\sqrt{n \cdot \pi \cdot (1 - \pi)}}$ vs. $Z\alpha$

Student: $t = \frac{(P - \pi)}{\sqrt{\pi \cdot (1 - \pi) / n}}$ o $t = \frac{(r - n \cdot \pi)}{\sqrt{n \cdot \pi \cdot (1 - \pi)}}$ vs. $t_{\alpha;v}$ con $v = n - 1$

Muestra vs. Muestra: π_1 y π_2 conocidos

$t = \frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{(\pi_1(1 - \pi_1) / n_1) + (\pi_2(1 - \pi_2) / n_2)}}$ vs. $t_{\alpha;v}$; con $v = n_1 + n_2 - 2$

Ambas muestras provienen de la misma población: $\pi_1 = \pi_2 = \pi$

Se estima $\pi = [(p_1 n_1) + (p_2 n_2)] / (n_1 + n_2)$ y se reemplaza en

$t = \frac{(P - \pi)}{\sqrt{\pi \cdot (1 - \pi) / n}}$ o $t = \frac{(r - n \cdot \pi)}{\sqrt{n \cdot \pi \cdot (1 - \pi)}}$ vs. $t_{\alpha;v}$ con $v = n - 1$

Test de hipótesis para varianzas

Muestra vs. Población: $\chi^2 = (n - 1) DS^2 / \sigma^2$ vs. $\chi^2_{\alpha;v}$ con $v = n - 1$

Muestra vs. Muestra: $F = DS^2_1 / DS^2_2$ vs. $F_{\alpha;v_1;v_2}$

con $DS^2 > DS^2_2$ y $v_1 = n_1 - 1$; $v_2 = n_2 - 1$

13.7 Significación clínica versus estadística

Antes de continuar es necesario realizar una reflexión sobre lo visto hasta ahora, en particular sobre las diferencias conceptuales entre la significación estadística tratada hasta aquí, y lo que se necesita en Medicina que es la significación clínica de los resultados. En 1986/88 Gardner y Altman presentaron un informe en el *BMJ* donde señalaban el peligro del avance inusitado de la estadística en los estudios clínicos. En ese entonces se había llegado a un punto donde, de los resultados obtenidos se informaba el valor (P-value) de probabilidad encontrado en el resumen (por ejemplo: $p = 0,027$), y a veces se olvidaban de poner lo más importante que eran los resultados hallados (por ejemplo, si comparaban dos muestras en lugar de poner en el “abstract” el valor de la diferencia hallada, ponían el valor de p). La recomendación fue indicar con claridad todos los datos posibles como ser:

“La diferencia entre las medias muestrales de presión sistólica de un grupo de 100 pacientes diabéticos y otros 100 no diabéticos, fue de 6 mm Hg, con un intervalo de confianza del 95% entre 1,1 y 10,9 mm Hg. El estadístico $t = 2,4$ resultó significativo con 198 grados de libertad, con un valor asociado de $p = 0,02$.”

O resumido con: “promedio 6 mm Hg, 95% CI (1,1 ; 10,9), $t = 2.4$ con $v = 198$ y $p = 0,02$ ”

A partir de ese entonces, esta es la forma recomendada para presentar la *significación estadística* en los resultados de las investigaciones clínicas. Sin embargo, falta una cuestión vital para el lector de estos temas: ¿ Esa diferencia hallada que implica clínicamente hablando ? Una primera respuesta es la deducción de que los pacientes diabéticos tienen una presión sistólica más alta que los no diabéticos. Pero la principal cuestión clínica es: ¿ Son hipertensos o no ? Y esta pregunta no se puede deducir de la información presentada más arriba. Falta la investigación de la *significación clínica* de los resultados.

En el año 1991, Braitner pone de manifiesto esta situación en un informe publicado en los Anales de Medicina Interna de USA, presentando un ejemplo sencillo que se muestra a continuación:

En una investigación sobre los resultados de un nuevo tratamiento para un cáncer, se trata de comparar la mejoría de los pacientes a los cuales se les aplicó el nuevo tratamiento, con otro grupo a los cuales se les aplicó el tratamiento viejo. En un primer estudio se tomaron 800 pacientes en cada grupo y se observó que 480 individuos respondieron al tratamiento nuevo, mientras que 416 lo hicieron con el viejo. Entonces, el 60% de los pacientes (480/800) respondieron a la nueva droga, mientras que el 52% (416/800) respondieron con la vieja. Se trata de una comparación entre dos proporciones estadísticamente hablando, y su significación se puede calcular con el modelo de comparación de dos muestras independientes de Student o Gauss. Si se efectúan los cálculos se puede ver que hay una diferencia altamente significativa ($p = 0,001$), para la diferencia de proporciones hallada del 8%, que tiene un 95% CI (3 ; 13)%. Esto significa, que hay prueba de que la diferencia entre ambos tratamientos existe y que está entre el 3% y el 13% (su mejor estimación es el 8%). Luego se hizo un estudio semejante en otro lugar, pero solo se utilizaron 50 pacientes en total, y los resultados obtenidos fueron: 15 mejorías con la nueva y 13 con la vieja. Si se aplica ahora el mismo tratamiento estadístico para los datos, hay que emplear

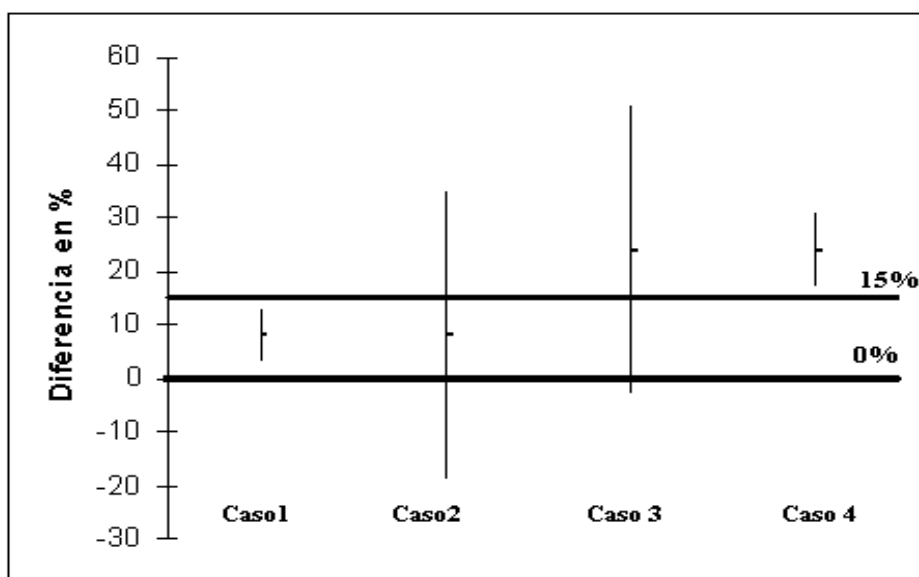
Si se aplica ahora el mismo tratamiento estadístico para los datos, hay que emplear el modelo Student porque las muestras son pequeñas, y se halla un porcentaje de mejoría del 60% para la nueva (15/25) y un 52% (13/25) para la vieja. Nuevamente la diferencia hallada es del 8%, pero esta vez los resultados no son significativos ($p = 0,57$) y no se puede hablar de una diferencia real entre ambos tratamientos, el intervalo de confianza de las diferencias es: 95% CI (-19 ; 35)%. Como el valor 0% (H_0) cae dentro del intervalo, no hay prueba de que el nuevo es mejor que el viejo, es decir no se puede rechazar la hipótesis nula. Se repitió el estudio nuevamente para 50 pacientes y los resultados fueron: Nuevo 60% (15/25) y viejo 36% (9/25), entonces ahora la dife-

N°	Porcentaje que responden al tratamiento		Valor P (p-value)	Significación Estadística	Diferencia entre los métodos en %	
	NUEVO	VIEJO			Puntual	95% CI
1	480/800=60%	416/800=52%	0,001	SI	8%	3% a 13%
2	15/25 = 60%	13/25 = 52%	0,57	NO	8%	- 19% a 35%
3	15/25 = 60%	9/25 = 36%	0,09	NO	24%	- 3% a 51%
4	240/400=60%	144/400=36%	<0,001	SI	24%	17% a 31%

ren-
cia
entre
las
pro-
por-
cio-
nes
es
del
24%
con
un

95% CI (-3 ; 51)% y el valor $p = 0,09$ tampoco es significativo. En un lugar diferente, se repite nuevamente la comparación en 800 pacientes y los resultados se muestran en la última línea del cuadro siguiente:

Si se comparan estos cuatro estudios, se puede ver que dos de ellos son estadísticamente significativos y los otros con menor tamaño muestral no lo son. Se consulta a un panel de expertos que estudian los costos de ambos tratamientos, la mejoría real en los pacientes, el tiempo empleado, la posible remisión de la enfermedad, el riesgo, etc.; quienes concluyen que: Para que valga la pena el cambio de tratamiento, el porcentaje de pacientes mejorados no puede ser menor del 15% (criterio clínico). Ahora, con ese valor crítico mínimo admisible se comparan los cuatro casos:



Se concluye que solo en el cuarto caso hay una diferencia clínica y estadística como para cambiar de tratamiento. La *diferencia significativa clínica* no existe en el primer caso, y es dudosa en los casos 2 y 3. Notar, que *no es lo mismo hablar de significación clínica que estadística*.

13.8 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|---|-------|-------|
| 1) El modelo de Student se usa para testear medias y varianzas. | V | F |
| 2) El modelo de Chi cuadrado se usa para testear proporciones. | V | F |
| 3) Las ventajas de las pequeñas muestras respecto a las grandes muestras son: | | |
| 4) El modelo de Fisher sirve para comparar dos varianzas. | V | F |
| 5) Las desventajas de las pequeñas muestras frente a las grandes son: | | |
| 6) Para emplear modelos de muestras pequeñas se necesitan por lo menos 2 valores. | V | F |
| 7) Con el modelo Student se pueden hacer validaciones de: | | |
| 8) A la teoría de pequeñas muestras se la llama también teoría exacta del muestreo. | V | F |
| 9) Los modelos de muestras pequeñas se pueden aplicar a las grandes, pero no el viceversa. | V | F |
| 10) La Chi cuadrado y la Fisher sirven para estudiar dispersiones. | V | F |
| 11) Student sirve para estudiar exactitud. | V | F |
| 12) Con Student se puede trabajar con 1 y con muestras con las medias y | | |
| 13) No es lo mismo comparar dos muestras independientes que apareadas. | V | F |
| 14) La única diferencia entre Gauss y Student en los intervalos de confianza es el valor Z. | V | F |
| 15) La curva de Student es simétrica respecto al origen, pero Chi cuadrado y Fisher no. | V | F |
| 16) Cuando al mismo individuo se lo mide dos veces, se dice que hay apareamiento. | V | F |
| 17) Las muestras apareadas sirven para estudiar casos del tipo: antes-después. | V | F |
| 18) Estos tres modelos permiten comparar hasta dos muestras. | V | F |
| 19) Las muestras deben provenir de poblaciones normales. | V | F |
| 20) Las muestras pueden escogerse de cualquier manera no tendenciosa. | V | F |
| 21) Si las muestras no son independientes, igual se las puede comprar con la Chi cuadrado. | V | F |
| 22) Las distribuciones de Chi cuadrado y Fisher varían desde cero a infinito. | V | F |
| 23) Los grados de libertad siempre valen N-1. | V | F |
| 24) Para dos muestras, los grados de libertad valen la suma de los tamaños menos dos. | V | F |
| 25) Cuando en la Ho hay un signo de mayor, la zona de rechazo queda a la izquierda. | V | F |
| 26) El signo de desigualdad de la H1 define el lugar del rechazo (izquierda o derecha). | V | F |

- | | | |
|--|----------|----------|
| 27) En las comparaciones de dos muestras conviene usar una igualdad en H_0 . | V | F |
| 28) El error casual se investiga con Chi cuadrado o Fisher. | V | F |
| 29) El error sistemático se investiga con el modelo Student y un patrón. | V | F |
| 30) En Fisher el denominador siempre debe ser mayor que el numerador. | V | F |
| 31) La propiedad inversa permite obtener probabilidades chicas en una Fisher. | V | F |
| 32) Explicar cómo se hace para controlar la exactitud de un método. | | |
| 33) Lo mismo pero respecto a la precisión. | | |
| 34) Explicar el significado del error máximo admisible. | | |
| 35) La tabla de Fisher se caracteriza por tener dos grados de libertad. | V | F |
| 36) Las tablas de Student y Chi cuadrado tienen un solo grado de libertad. | V | F |

2) La vida útil de un medicamento A se midió con 20 muestras y dio una media de 140 días, con un desvío de 10 días. El medicamento B se determinó con 10 muestras y se obtuvo una media de 120 días con un desvío de 80 días. Decidir si hay diferencias entre ellos desde el punto de vista de la exactitud y de la precisión.

3) Una máquina produce grageas con un diámetro de 0,5 cm según su promedio histórico. Para comprobar si se mantiene en buenas condiciones se toma una muestra de 10 grageas, se las mide y se obtiene una media de 0,53 cm con un desvío de 0,03 cm. Ensayar la hipótesis que está bien.

4) Dos tipos de soluciones químicas A y B fueron ensayadas para medirle su pH. Se tomaron 6 muestras de A que dieron una media de 7,48 con un desvío de 0,02. De la B se tomaron 5 muestras y se obtuvo una media de 7,32 con un desvío de 0,03. Con esta información decidir si ambas muestras tienen el mismo pH.

5) La especificación de un medicamento requiere un 12% de cierto principio activo. Se analizaron 20 muestras del medicamento y se detectó un promedio de 12,5% del componente, con un desvío de 0,3%. Decidir si el producto cumple las especificaciones requeridas.

6) Para el problema anterior, se eligen 18 muestras de un segundo lote de producción y resulta una media de 12,1% con un desvío de 0,09. Se pide averiguar:

- a) si este lote cumple las especificaciones requeridas;
- b) si hay diferencias entre las medias de ambos;
- c) si hay diferencias entre las varianzas de los dos lotes.

7) En una fábrica de medicamentos un farmacéutico sugiere cambiar una de las principales máquinas del proceso productivo, para reducir los tiempos de producción y bajar los costos. Un estudio económico financiero encuentra que tal cambio resulta si el tiempo se reduce en por lo menos un 10%. Se hicieron 5 pruebas con el nuevo equipo y resultó una disminución promedio del tiempo de fabricación del 10,5% con un desvío del 0,3%. Averiguar si resulta conveniente la adquisición de la nueva máquina.

8) Se tienen dos poblaciones cuyas varianzas son desiguales: $\sigma_1^2 \neq \sigma_2^2$ y desconocidas. Por ello el investigador tiene que estimarlas con los valores muestrales $DS_1^2 \approx \sigma_1^2$ y $DS_2^2 \approx \sigma_2^2$. Elige al azar 25 individuos de la primera población y obtiene un promedio de 14 con un desvío de 2, mientras que con 26 muestras de la segunda obtiene una media de 18 y un desvío de 3. Con esta información se debe efectuar una comparación de medias con el modelo de Student. Calcular en primer lugar el número de grados de libertad y luego efectuar el test

9) Calcular como resultaría el problema anterior si se supiera que las varianzas poblacionales son iguales o se pudiera hacer semejante supuesto.

10) En un hospital se desea saber si hay diferencia entre dos tratamientos diferentes. Para tratar de saber la efectividad de los mismos para disminuir el dolor causado por la enfermedad. En el primer caso, cuando el paciente es tratado en la sala de emergencias se le aplica una mezcla de un IV hidrocloreto de ranitidina para tratar la dispepsia. Se mide el dolor antes del tratamiento y 45 minutos después de ser aplicado. Al dolor se lo clasifica en una escala de 1 a 10. El segundo tratamiento es darle una mezcla de lidocaina viscosa con un antiácido. Se eligieron 28 pacientes y el primer tratamiento se le aplicó 15 de ellos. La asignación de cada tratamiento fue hecha en forma aleatoria. Los resultados fueron: una media de 3,87 para el primero y de 4,08 para el segundo. Como no hay datos reales del desvío estándar en los ratings de dolor, el investigador decide usar un dato conocido: en ausencia de tratamiento el desvío es $\sigma = 1,73$. Luego decide usar ese dato para estimar el desvío de las diferencias muestrales σ_{Δ} con la relación:

$$\sigma_{\Delta} = 1,73 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{Decidir si hay diferencia entre los dos tratamientos}$$

14

Estadística no paramétrica

En este capítulo se desarrollan modelos estadísticos para realizar validaciones de todo tipo de magnitud biológica, tanto cualitativa como cuantitativa. Se desarrollan modelos equivalentes a los del capítulo anterior, sin las restricciones o supuestos de los mismos. El primer modelo es el de la Binomial, aplicable al caso de una sola muestra, y que puede ser aproximado con el de Gauss cuando las muestras sean grandes. El segundo modelo de rachas en una muestra sirve para validar la aleatoriedad en el orden o secuencia de extracción de muestras de una población. A continuación se desarrollan modelos para comparar dos muestras entre sí. Cuando las muestras sean apareadas se tienen dos modelos: el del Signo y el de Wilcoxon. Si bien este último es más poderoso que el otro porque usa mayor cantidad de información, el del Signo se presenta por su uso difundido en Análisis Clínicos y Farmacia. Cuando las muestras son independientes, se presenta el modelo paramétrico más poderoso, el modelo de la U de Mann-Whitney, casi tanto como su equivalente paramétrico: el modelo Student para dos muestras independientes.

14.1 Conceptos básicos

En los capítulos anteriores se han analizado modelos estadísticos que implican distribuciones continuas con ciertos supuestos básicos para la aplicación de estas técnicas. El principal uso de esos modelos es la estimación de parámetros desconocidos de la población en estudio, para poder hacer pruebas de validación o ensayos de significación y testear así las hipótesis planteadas. Estos supuestos se plantean fundamentalmente sobre el valor que toman los parámetros poblacionales o sobre comparaciones de dos de ellos. Hasta ahora se ha trabajado con magnitudes biológicas de tipo cuantitativas y continuas. A las magnitudes discretas se las ha tratado como proporciones para poder usar los modelos vistos, y cuando se usó el modelo de Gauss se tuvo que hacer una corrección por continuidad. A esta metodología de trabajo se la denomina Estadística Paramétrica, por contraposición a otra donde lo que interesa es comparar distribuciones en lugar de parámetros. Mientras los supuestos usados en la paramétrica especifican la distribución original (generalmente la gaussiana), hay otros casos en la práctica donde no se puede hacer esto, donde no se puede especificar la forma de distribución original. Se requiere entonces otra metodología de trabajo, una estadística de *distribuciones libres*, donde no se necesitan hacer supuestos acerca de la distribución poblacional, donde se puede comparar distribuciones entre sí o verificar supuestos a cerca de la forma de la población. Por ejemplo, verificar el supuesto de normalidad necesario para usar el modelo Student. La solución para estos casos es el empleo de la *Estadística no paramétrica*. Hay ciertas ventajas en su uso, tales como:

- trabajar con magnitudes *cualitativas*, además de las cuantitativas;
- estudiar casos donde no es posible precisar la naturaleza de la distribución;
- ídem para los casos donde los supuestos de la forma poblacional son débiles;

- aplicar el mismo modelo a casi todas las distribuciones en lugar a una sola;
- es más fácil de entender para quienes no poseen base matemática adecuada.

Y también tiene algunas desventajas como:

- cálculos usualmente más engorrosos;
- no extraen tanta información como los paramétricos si se aplican al mismo caso;
- son menos eficientes si las muestras son grandes.

Los modelos paramétricos tienen mayor capacidad para detectar diferencias muestrales que los no paramétricos. Es decir, son capaces de ver una diferencia significativa en casos donde los otros no pueden. Como su poder discriminador es mejor, *siempre que se pueda, conviene usar modelos paramétricos antes que los no paramétricos*, por su mayor sensibilidad para detectar diferencias significativas. A menos que las diferencias sean tan grandes que con cualquier modelo pueden detectarse. Pero como los no paramétricos se aplican casi todos los casos, son más fáciles de entender y no tienen tanta “complicación matemática”; se están poniendo de moda en Bioquímica y Farmacia cada vez más.

El independizarse de la forma de la población llevó a estos modelos a otras aplicaciones no clásicas, como en las ciencias de la conducta, marketing, ciencias sociales, etc. En algunas técnicas, como las pruebas de *rango* o de *orden*, se trabaja con puntajes, que no son verdaderamente numéricos, lo cual ocasiona deformaciones en los datos si se empleasen técnicas paramétricas y el valor de las conclusiones de la validación estadística quedaría menoscabado. Por ejemplo, se pueden asignar rangos por textura, coloración, sabor, olor (magnitudes organolépticas), clasificar por infección con cierto tipo de virus, y otros casos donde no se cumpla el supuesto de homogeneidad de varianzas; acá el modelo de rangos puede ser la salida. Cuando se comparan dos muestras, los modelos paramétricos hacen hincapié en la comparación de las medias, mientras que los no paramétricos fijan su atención en comparar medianas. A continuación se presenta el Modelo de la Binomial para trabajar en una sola muestra.

14.2 La prueba de la binomial

Esta prueba se emplea cuando los resultados del experimento se expresan en forma dicotómica. Los dos tipos de resultados se pueden clasificar en “éxito” o “fracaso” y aplicar la fórmula de la probabilidad binomial para poder calcular la probabilidad de ocurrencia de la hipótesis nula formulada. Comparando esta probabilidad B_x con el nivel de significación adoptado α por el investigador para el experimento, se procede a tomar la decisión de acuerdo con uno de los dos casos posibles siguientes:

- Si $B_x < \alpha$ se rechaza la H_0 , porque significa que la probabilidad de que ocurra esta hipótesis es muy baja, comparada con el nivel de significación adoptado para la validación.
- Si ocurre lo contrario y resulta $B_x > \alpha$, entonces no hay evidencia significativa como para rechazarla. Esto es especialmente útil para muestras pequeñas porque da la probabilidad exacta de la ocurrencia de los sucesos.

Cuando las muestras sean grandes es muy práctico usar la aproximación de la binomial con la normal. Una muestra se considera grande si $n > 25$. Hay tres maneras de resolver un problema de acuerdo con su tamaño muestral. El procedimiento del modelo Binomial se puede resumir en los pasos siguientes:

Paso 1) definir la cantidad de muestras o experimentos a realizar: n ;

Paso 2) medir las frecuencias de ocurrencia en ambos casos posibles (éxito para la menor);

Paso 3) plantear la hipótesis nula y su alternativa;

Paso 4) de acuerdo con tamaño muestral, seguir uno de estos tres caminos:

(I) Si $n \leq 25$ y si $p \cong q$ se usa la Tabla 8 del Anexo de Tablas Estadísticas para determinar el valor crítico en un ensayo de una sola cola $B_x = B_\alpha$. Para un ensayo de dos colas se duplica el valor de tablas $B_x = 2 B_\alpha$. El estadígrafo comparativo es la probabilidad para la frecuencia menor obtenida; o sea, con el valor x (frecuencia menor) se saca de tablas el valor B_α y se saca B_x según si es de una o dos colas. Luego si $B_x < \alpha$ se rechaza H_0 .

(II) Si $n \leq 25$ y si $p \neq q$ habrá que calcular la probabilidad con la fórmula de la binomial de acuerdo con la hipótesis nula. Esto es, se debe obtener la probabilidad acumulativa desde 0 hasta el valor predicho x (o su complementaria si es más corta). De esa forma se calcula B_x . Luego es:

$$P_B = \sum_0^x C(n, x) \cdot p^x (1-p)^{n-x}$$

Entonces se compara P_B con $(1 - P_B)$ y al menor de ellos se lo llama B_x . Este valor se compara con el nivel de significación α adoptado por el investigador, y se toma la decisión de aceptar o rechazar H_0 como se vio más arriba.

(III) Si $n > 25$ se trata de una muestra "grande" y se usa la aproximación normal a la binomial. Es decir, se transforma el valor x de frecuencia observada a unidades tipificadas (Z). Cuidando de efectuar la corrección por continuidad y de tablas se obtiene su probabilidad B_x para compararlo con el nivel de significación α adoptado. O sea:

$$Z = \frac{(X \pm 0,5) - \mu_r}{\sigma_r} = \frac{(X \pm 0,5) - n \cdot p}{\sqrt{n \cdot p \cdot q}}$$

El signo \pm se usa para mostrar las dos alternativas de corrección por continuidad. Si el ensayo es de una sola cola y la H_0 es $x \leq n p$ se usa $(X + 0,5)$. Por su parte, si la H_0 es: $x \geq n p$ se usa $(X - 0,5)$. Una vez calculado Z , se busca en la tabla de la normal su probabilidad B_x la que se compara con el nivel α para poder decidir. En el cuadro 14.1 se presentan ejemplos ilustrativos de cada caso.

CUADRO 14.1: La prueba de la Binomial para una sola muestra

Caso 1 - Muestras pequeñas con equiprobabilidad:

Ejemplo 1) En una maternidad ocurrieron 18 nacimientos en una semana, 11 de los cuales fueron varones. La hipótesis de trabajo es que la proporción de sexos es la habitual.

Si bien este problema se puede resolver usando la fórmula de la binomial, como n es menor que 25 y las probabilidades son aproximadamente iguales, entonces es más práctico usar la Tabla 8 del Anexo. Para un tamaño muestral $n = 18$ y para la frecuencia menor $x = 7$, resulta un valor de probabilidad $B_x = 0,24 > \alpha$. Por lo tanto, no se puede rechazar la hipótesis nula. No hay evidencia significativa que pruebe que la proporción de sexos no es la habitual.

Ejemplo 2) Un profesor le enseñó a un grupo de 20 alumnos dos modelos estadísticos: el A y el B para resolver un mismo problema en la clase de coloquios de la siguiente manera: Seleccionó 10 al azar y les enseñó primero el A y más tarde el B; con los demás procedió al revés: primero el B y luego el A. Finalizado el examen parcial, les tomó a los mismos 20 alumnos el problema explicado en coloquios. Se supuso que por el estrés del examen habría una tendencia a usar el método aprendido primero. Los resultados mostraron 17 con este y 3 con el método aprendido segundo.

H_0 : No hay preferencias por usar el método aprendido antes o después ($p = q = 1/2$).

H_1 : Prefieren el aprendido primero ($p > q$).

Los datos a usar son $n = 20$; $p = q = 0,5$; la menor frecuencia es $x = 3$ entonces usando la Tabla de la Binomial se obtiene una probabilidad $B_x = 0,001$ que es mucho menor que $\alpha = 0,05$. Por lo tanto se rechaza la H_0 y se concluye que hay una especie de regresión hacia lo aprendido primero.

Caso 2 - Muestras pequeñas sin equiprobabilidad: en un laboratorio que fabrica medicamento en ampollas, históricamente 10% de la producción presenta fallas de cierre hermético. Para probar la calidad de un lote fabricado se extraen 5 muestras al azar y se verifican. Si encuentran por lo menos una fallada se devuelve todo el lote, en caso contrario sale a la venta. Se desea saber si con esta política se tiene un nivel aceptable (1%) de rechazos.

Este problema se resuelve con la aplicación de la fórmula binomial, con $p = 0,1$ y $q = 0,9$

$$B_x (x < 1) = P(x = 0) = C(n, x) \cdot p^x (1 - p)^{n-x} = C(5, 0) (0,1)^0 (0,9)^5 = 0,59$$

Hay una probabilidad del 59% de no encontrar ninguna fallada. O sea, en el 41% de los casos habría devolución, lo cual no es conveniente. Si se cambia la política y se rechazan lotes con dos o más falladas. Eso significa que se aceptarían lotes con una o ninguna fallada. Esto es:

$$B_x (x < 2) = P(x = 0) + P(x = 1) = 0,59 + C(5, 1) (0,1)^1 (0,9)^4 = 0,59 + 0,328 = 0,918$$

Ahora, con esta nueva política se rechazarían lotes en el 8% de los casos lo cual es más aceptable. Si se la hace más amplia y se rechazan lotes con tres o más falladas, se tiene:

$$B_x (x < 3) = P(x = 0) + P(x = 1) + P(x = 2) = 0,59 + 0,328 + C(5, 2) (0,1)^2 (0,9)^3 = 0,9909$$

Entonces esta última política es la buscada.

Caso 3 - Muestras grandes: en 90 pacientes se probó un nuevo medicamento y 10 de ellos no se curaron en el plazo previsto. La idea es aceptar la droga si logra mejoría en el 75% de los casos. ¿Qué decisión se debe tomar a la luz de estos resultados experimentales?

Como la muestra es grande, conviene usar la aproximación normal a la binomial. Entonces,
 $H_0 : x \leq n \cdot p = 67,5$ casos curados y la droga no sirve.
 $H_1 : x > 67,5$ Y La droga sirve.

$$\mu_r = n \cdot p = 90 \cdot 0,75 = 67,5 ; SE(r) = \sigma_r = \sqrt{n \cdot p \cdot q} = \sqrt{90(0,75)0,25} = 4,11 ; X = (90-10) = 80$$

$$Z = \frac{(X \pm 0,5) - \mu_r}{\sigma_r} = \frac{(X + 0,5) - n \cdot p}{\sqrt{n \cdot p \cdot q}} = \frac{(80,5) - 67,5}{4,11} = 3,163^{***}$$

Se tiene evidencia altamente significativa de la bondad del medicamento probado. Por otro lado $x = 67,55$ no cae dentro del 99,9% CI (70,4 ; 89,6).

NOTA: si se hubiese calculado con el programa Excel resultarían:

- 1) la aproximación normal con $Z = 3,163$, da una probabilidad de 0,9997028;
- 2) la probabilidad binomial acumulada, que es la exacta da un valor de 0,9992192.

Esto es, $Bx = 0,00078 \lll \alpha = 0,001$ para el 99,9% de confianza y se rechaza H_0 .

14.3 La prueba de rachas de una muestra

Esta prueba es para aleatoriedad. En los casos anteriores vistos se trabajó bajo el supuesto de extracción de muestras aleatorias de la población. De acuerdo con el diseño del experimento se puede lograrlo, sin embargo, hay casos donde no es tan sencillo hacerlo. Otras veces puede que sea necesario probar la aleatoriedad. Para ello se han venido desarrollando una serie de modelos estadísticos que estudian el *orden o secuencia* en que las muestras individuales fueron obtenidas, para probar que la muestra es aleatoria. Este modelo se basa en la cantidad de *rachas* que una muestra exhibe. Por rachas se entiende a una sucesión de símbolos idénticos que pueden estar separados o no por otro tipo de símbolos. Por ejemplo, sea una serie de mediciones de magnitudes dicotómicas identificadas con los símbolos de resultado positivo (+) o negativo (-) a juicio del investigador, de acuerdo con cierto criterio profesional empleado:

Resultados: $\underline{++} \underline{---} \underline{+} \underline{---} \underline{++} \underline{+}$
 N° de rachas: 1 2 3 4 5 6 7

La primera racha empieza con una serie de 2 símbolos positivos, la segunda racha con 3 negativos, la tercera con un positivo, y así sucesivamente hasta la séptima racha con un positivo. El número de rachas es $r = 7$. El número total de rachas indica si una muestra es o no aleatoria. Si se da un número pequeño de rachas puede deberse a una falta de independencia o a una tendencia

temporal. Mientras que si por el contrario hay un número muy grande de rachas, las fluctuaciones cíclicas sistemáticas, en un período corto de tiempo, pueden causar influencia en los valores asignados por el investigador. Por ejemplo, si se lanza al aire una moneda 30 veces y se obtienen 30 rachas es razonable dudar de esa moneda, son demasiadas. Lo mismo ocurriría si se obtienen únicamente dos rachas, como primero todas caras y luego secas.

El número de rachas no depende de la frecuencia de los sucesos. En efecto, dos situaciones bien diferentes pueden tener las mismas frecuencias. Como las del ejemplo anterior de las monedas donde se tiene una frecuencia relativa de $\frac{1}{2}$, en ambos casos. Sin embargo, en el primer caso sale una cara, luego una seca, y así sucesivamente una y una, mientras que en el segundo primero se dan 15 caras seguidas y luego las 15 secas finales. Generalmente se conoce la distribución muestral en muestreos repetidos, lo que permite obtener la probabilidad asociada para poder probar hipótesis. El *procedimiento* para aplicar este modelo es como sigue:

Paso 1) Se calcula el número n_1 de elementos de una clase identificadas por un símbolo y n_2 la cantidad de elementos de la otra.

Paso 2) Se ordenan los $n = n_1 + n_2$ sucesos en el orden en que ocurrieron.

Paso 3) Se cuenta el número r de rachas.

Paso 4) Se postula una hipótesis nula de trabajo que permita calcular la probabilidad asociada.

Paso 5) Se determina la probabilidad que ocurran r rachas, usando H_0 , y se compara con el nivel de significación α adoptado para aceptar o rechazar la H_0 .

Para determinar el valor de la probabilidad se presentan dos casos según el tamaño de la muestra:

(I) *Muestras pequeñas*: si los tamaños muestrales de ambos resultados son menores que 20 se emplea la Tabla 9 del Anexo. En la primera de las tablas, ubicada en la parte superior, se dan valores de r para una (H_0) de menor o igual, tal que su probabilidad asociada es $p = 0,05$ para dos colas (para una cola es la mitad: $p = 0,025$). En la parte inferior se presenta una segunda tabla, es para valores de r con (H_0) de mayor o igual, tal que su probabilidad asociada para dos colas es $p = 0,05$. A la de arriba se la denota con $R_{\text{mín}}$. Y a la de abajo con $R_{\text{máx}}$.

(II) *Muestras grandes*: si alguno de los tamaños muestrales es mayor a 20, se determina el valor de estadígrafo comparación Z con una distribución aproximadamente normal:

$$Z = \frac{r - \mu_r}{\sigma_r} \quad \text{el cual se compara como siempre con } z_\alpha \text{ de Gauss.}$$

Con:

$$\mu_r = 1 + [(2 n_1 n_2) / n]$$

$$\text{VAR}(r) = (\sigma_r)^2 = \frac{(2 n_1 n_2) (2 n_1 n_2 - n)}{n^2 (n - 1)}$$

Para ilustrar mejor estas ideas, se presentan dos problemas resueltos en el Cuadro 14.2 siguiente, uno para muestras pequeñas y el otro para las grandes.

CUADRO 14.2: La Prueba de rachas para una sola prueba

Caso 1 - Muestras pequeñas: en un Laboratorio de investigación se prueba un antiinflamatorio nuevo. Los resultados son aceptables si al segundo día de aplicado al paciente se observa una reducción del 90% en la inflamación; se le asigna (+) a ese caso. Se quiere testear la hipótesis que la sucesión de signos positivos y negativos se produce al azar. La sucesión de los 24 casos analizados fue:

Sucesión: + - + + + + - + + + - - - - + - - + + + - - - -
 Rachas: 1 2 3 4 5 6 7 8 9 10

Se usa un test dos colas porque no se predice la dirección de la desviación que supone. El tamaño de cada muestra es 12. Usando ambas tablas del apéndice se determina la zona de rechazo de la hipótesis nula cuando r es menor o igual a 7 o cuando r es mayor o igual a 19; con esto se define:

Zona de aceptación: $7 \leq r \leq 19$ y como $r = 10$ cae dentro de esta zona y no se rechaza (H_0).

Se concluye que se deben suponer aleatorias a las muestras tomadas.

Caso 2 - Muestras Grandes: en un hospital se forma todas las mañanas temprano, una cola de pacientes esperando su turno para la extracción de sangre. La bioquímica a cargo decide verificar si la colocación de hombres y mujeres es al azar. Anota el sexo de cada uno de los primeros 50 pacientes que entraron al laboratorio. Los resultados fueron:

Sucesos : HH M H M HHH MM H MM H M HH MMM HH MM HH
 Rachas : 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Sucesos : MM H M H M H MM H M HH M HH M H M H M H MM
 Rachas : 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34

Total de rachas: $r = 34$. El total de hombres y mujeres fue: $25 = n_1 = n_2$

$$\mu_r = 1 + [(2 n_1 n_2) / n] = 1 + [2 (25) 25] / 50 = 26$$

$$(\sigma_r)^2 = \frac{(2 n_1 n_2) (2 n_1 n_2 - n)}{n^2 (n - 1)} = \frac{2 (25) 25 \{ [2 (25) 25] - 50 \}}{(50)^2 (50 - 1)} = 12,245 \quad \text{y} \quad \sigma_r = 3,5$$

$$Z = (r - \mu_r) / \sigma_r = (r - 26) / 3.5 = (34 - 26) / 3.5 = 2,29^* \quad (0,01 < P_N(Z_{2,29}) = 0,011 < 0,05)$$

Se rechaza la hipótesis de que los sexos guardan un orden aleatorio al formar la fila. Se encontró evidencia significativa para rechazarla. Si se calcula la probabilidad gaussiana para $Z = 2,29$ resulta $p = 0,011$; esto es, más cerca del rechazo al 99% que al 95% de nivel de confianza.

La prueba de aleatoriedad algunas veces es necesaria para comprobar los supuestos específicos de otros modelos estadísticos. Debería realizarse primero para decidir si el modelo es aplicable porque se verifica su supuesto de aleatoriedad, como en el caso Gauss, Student, Fisher, etc. Para evitar en la prueba de aleatoriedad, lo aconsejable es usar sorteos al azar para la selección de muestras.

14.4 Modelos para dos muestras apareadas

Los modelos no paramétricos para realizar comparaciones entre muestras apareadas son dos: el modelo del Signo se usa cuando la magnitud biológica medida en el experimento es de tipo dicotómica. El modelo de Wilcoxon es más potente porque usa mayor información: el signo y el rango de las diferencias halladas. Ambos modelos usan el *signo* de las diferencias entre los pares de muestras testeados.

Cuando el investigador asigna puntajes a las mediciones que realiza, basado en algún criterio profesional, el problema es traducir eso a una dicotomía. La manera de hacerlo es calcular la mediana de los puntajes y asignar un signo positivo a los resultados mayores y negativo a los menores. Entonces, se ha “dicotomizado” la variable “puntajes” medida y se pueden aplicar estos modelos. Los puntajes se suelen usar mucho en ciencias sociales y humanas, donde el comportamiento es juzgado de acuerdo con una cierta escala. También en las notas de los exámenes en el sistema educativo. Puede hacerse lo mismo con magnitudes cualitativas y cuantitativas, o sea que el campo de aplicación se extiende a cualquier magnitud medida. En el caso dicotómico es directo. En el caso politómico hay que separar en dos grandes grupos, de acuerdo con algún criterio profesional adecuado. En el caso ordinal se procede en forma análoga. Por ejemplo, en los antibiogramas donde lo que se mide es el diámetro de acción del antibiótico en una caja de Petri, los resultados se expresan con tres valores: resistente, poco sensible y muy sensible. Como lo que interesa son los antibióticos potentes, la clasificación es (+) para los calificados como muy sensibles y (-) para los otros dos, donde aplicar tal antibiótico produce poco o ningún resultado en el paciente. Las magnitudes organolépticas (turbidez de orina, sangre en heces, olor color, etc.) son fácilmente adaptables a este modelo. Finalmente, en el caso de magnitudes continuas se debe definir un punto de corte en la variable con la que se está trabajando. Por ejemplo, si se está midiendo el peso de los pacientes, se puede definir: (+) hasta 60 kg (Flacos), y (-) más de 60 kg (Gordos); con la temperatura corporal puede ser (+) hasta 37 °C (normal), y (-) más de 37 °C (afiebrado). Lo cual muestra que *toda magnitud biológica puede ser dicotomizada*.

Las muestras apareadas se diseñan para el caso clásico en que se usa un mismo individuo como su propio control, para hacer comparaciones entre dos casos de un cierto factor en estudio. Por ejemplo, las del tipo “antes-después” donde un factor cualquiera es el tratamiento que se le aplica al individuo, el caso de medirle el peso antes del régimen de adelgazar y después del mismo, o su velocidad para correr antes del entrenamiento y un régimen de comida especial y después, o las medicaciones en Medicina, etc. Otra forma muy común en las investigaciones farmacológicas es la del tipo “droga *versus* placebo”. En Bioquímica se solía recomendar para comparar dos técnicas de laboratorio; ver si la nueva técnica es equiparable a la vieja. Usando a los mismos pacientes que vienen a hacerse determinaciones diariamente, se los mide con ambas técnicas para comparar. El costo de esta manera es más bajo porque una de las dos determinaciones que se le hacen a cada paciente es la de rutina. En cambio, el modelo más potente de Student pa-

ra muestras independientes requiere usar más reactivos, porque el mismo paciente no puede aparecer dos veces en los datos para no perder independencia. Como regla general, se debe recordar que la mejor forma de comparar dos métodos o rutinas de laboratorio es usando mediciones repetidas del mismo valor patrón. Por ejemplo, fraccionando en alícuotas a un mismo suero control para realizar el experimento de comparación. Si la homogeneización se hace correctamente, no hay que preocuparse por los supuestos gaussianos, recordando el Teorema Central del Límite.

También sirve para testear alguna parte de un método con dos variantes; como diferentes marcas comerciales, personal, equipos, instrumentos, etc. Por ejemplo, probar el resultado de dos métodos de fabricación de medicamentos, dos vendedores de una farmacia, dos marcas comerciales de analgésicos, etc. De lo antedicho, se puede inferir el amplio campo de aplicación en la práctica de estos modelos. Su gran limitación es que no se pueden usar para comparar más de dos muestras a la vez.

14.4.1 Modelo del Signo

Esta prueba se basa en los signos de las diferencias observadas entre valores apareados. La independencia se resguarda si la unidad de muestreo es el resultado de comparar a las parejas entre sí, en lugar de usar las observaciones individuales. Cada pareja de datos debe ser independiente de las otras, y la forma de lograrlo fácilmente es seleccionado al azar al elegir los componentes de las muestras. Una ventaja es que puede usarse cuando las observaciones pareadas están simplemente ordenada por rangos. No es necesaria la homogeneidad de las varianzas, ni que las muestras sean extraídas de la misma población. Su desventaja es que elimina mucha información pues la reduce a una dicotomía. Otras ventajas son: la facilidad de aplicación, la reducción de tiempo y costo para el investigador. Los supuestos básicos para aplicar este modelo son:

- La variable subyacente en las determinaciones es continua;
- las pares de muestras son independientes;
- la hipótesis nula es: los signos se distribuyen al azar alrededor de la mediana.

Este último implica que las diferencias deben ser tales que aproximadamente la mitad de ellas sean positivas y la otra mitad negativas. Por supuesto que puede haber *ligas*, esto es, parejas de muestras cuya diferencia sea nula. Para el modelo solo interesan las otras. Se toma en cuenta el número n de diferencias halladas (n_+ : cantidad de signos positivos y n_- : número de negativos). Donde $n_+ + n_- = n$. Este valor puede ser a lo sumo igual al número total N de parejas testeadas. O sea, $N = n + e$ (donde e es la cantidad de empates o ligas encontrados). En resumen, el procedimiento para aplicar este modelo sigue los pasos:

Paso 1) Se efectúan las parejas de mediciones y se determina el signo de diferencia.

Paso 2) Se cuenta el número de (+) y el de (-) hallados y así se calculan: $n_+ + n_- = n$.

Paso 3) Se elige al menor de ellos como $x = n_{\text{mínimo}}$ para efectuar el test.

Paso 4) Dependiendo del tamaño de la muestra se siguen dos caminos:

(I) *Muestras pequeñas* ($n \leq 25$): se usa la Tabla 8 del Anexo de Tablas Estadísticas para determinar el valor crítico en un ensayo de una sola cola $B_x = B_\alpha$. Para un ensayo de dos colas se duplica el valor de tablas $B_x = 2 B_\alpha$. El estadígrafo comparativo es la probabilidad para la frecuencia menor obtenida, o sea con el valor x (frecuencia menor) se saca de tablas el valor B_α y se calcula B_x según si es de una o dos colas. Luego si $B_x < \alpha$ se rechaza H_0 .

(I) *Muestras grandes* ($n > 25$): se usa la aproximación normal a la binomial, para determinar Z en forma análoga a las vistas antes, con la debida corrección por continuidad. Como la (H_0) supone una igualdad entre la cantidad de positivos y negativos, se toma $p = q = 0,5$ y así:

$$Z = \frac{(r \pm 0,5) - \mu_r}{\sigma_r} = \frac{(r \pm 0,5) - n p}{\sqrt{n \cdot p \cdot q}} = [(2r \pm 1) - n] / \sqrt{n} \text{ versus } Z_\alpha$$

Ejemplo 1) Muestras pequeñas: en un laboratorio se desea testear una nueva técnica. Para ello se toman las determinaciones hechas en el día con la manera usual (método viejo) y se las replica con el nuevo procedimiento, obteniéndose los valores:

Paciente	Nuevo	Viejo	Signo
1	2,3	2,2	+
2	2,6	2,5	+
3	3,5	3,4	+
4	2,9	2,9	0
5	2,5	2,4	+
6	3,3	3,1	+
7	1,7	1,8	-
8	2,8	2,8	0
9	2,9	2,9	0
10	3,1	3,2	-
11	2,4	2,3	+
12	3,2	3,3	-
13	2,5	2,4	+
14	2,6	2,5	+
15	3,3	3,2	+
16	2,9	2,8	+
17	2,2	2,1	+

$N = 17$
 $(+) = 11 = n_+$
 $(-) = 3 = n_-$
 $(0) = 3$
 Entonces
 $n = 14$
 $x = 3$

Del total de pruebas $N = 17$ solo interesan los $n = 14$ que no fueron empates ($e = 3$). Con los 14 casos restantes se tienen : $n_+ = 11$ signos positivos y $n_- = 3$ negativos; por lo tanto, se toma el valor mínimo como $x \leq 3$. Entrando a la Tabla 8 con $n = 14$ y $x = 3$, resulta un valor de probabilidad binomial: $B_x = 0,029$ que es menor que el nivel de significación $\alpha = 0,05$, por lo que se debe rechazar la H_0 . Esto es, se tiene evidencia científica que ambos métodos no son equivalentes para las mediciones realizadas. Por la cantidad de signos + parece que el método nuevo arroja valores más grandes que el viejo. Para continuar, es conveniente obtener un suero control y revisar la calibración de los métodos.

Ejemplo 2) Muestras grandes: el centro de estudiantes de cierta Facultad piensa que una de sus autoridades debe ser removida de su cargo. Para ello planifica una campaña de pegatinas y una asamblea estudiantil ad-hoc, donde sus seguidores harán acotaciones significativas para convencer a los concurrentes. Uno del grupo, que cursó Bioestadística, duda de la efectividad del plan y decide buscar evidencia científica para decidir si tuvo efecto la asamblea. Prepara un cuestionario a los asistentes para detectar los cambios "antes y después" de la reunión, y selecciona al azar a 100 de ellos para encuestarlos.

Los casos que interesan son los $30 + 42 = 72$ casos de cambios de opinión. Si la asamblea no afectara a sus concurrentes, cerca de la mitad de los que cambiaron de opinión hubieran pasado de SÍ a NO y la otra mitad de NO a SÍ. Esto es: H_0 : Se esperan 36 cambios porque la asamblea no tuvo efecto. Los casos que interesan son los que cambiaron de opinión $n = 30 + 42 = 72$.

		Antes		
		SI	NO	
	SI	9	42	$\mu_r = n p = 72 (0,5) = 36$
	NO	30	19	$\sigma_r = \sqrt{n.p.q} = \sqrt{72 (0,5)0,5} = 4,243$
Después				$r = 42$ cambiaron de opinión de Si a NO
				$Z = [(42 - 0,5) - 36] / \sqrt{72} = 1,3$

Notar que si se hubiera tomado $r = 30$, el resultado sería el mismo: $Z = 1,3$ (no significativo). Como cae en la zona de aceptación con un nivel del 95% ($Z_\alpha = 1,96$), no se puede rechazar la hipótesis nula. La conclusión es que a pesar de los 42 casos que pasaron de NO a SÍ no hay evidencia significativa como para pensar que la asamblea fue útil a los fines estudiantiles.

Ejemplo 3) Muestras grandes: un farmacéutico que ha realizado transformaciones en el aspecto de su negocio, realiza una encuesta entre sus clientes habituales para conocer su opinión respecto a los cambios. Descarta a los que no emitieron opinión y a los que no tenían un conocimiento previo de su empresa. Elige al azar 60 entre los remanentes. Encontró que 48 de ellos emitieron una opinión que considera positiva y 12 negativas. Ver si obtuvo pruebas de que la imagen de su empresa ha mejorado.

(H_0) $p = q = 0,5$ las opiniones emitidas se deben al azar y no hay mejora de imagen.

De los datos surgen: $\mu_r = n p = 60 (0,5) = 30$; $\sigma_r = \sqrt{n.p.q} = 3,873$ y $r = 48$

$$Z = [(r \pm 0,5) - \mu_r] / \sigma_r$$

Reemplazando con los de opinión positiva es: $Z = [(48 - 0,5) - 30] / 3,873 = 4,52***$

Reemplazando con los de opinión negativa es: $Z = [(12 + 0,5) - 30] / 3,873 = -4,52***$

Y también se puede calcular con la otra fórmula: $Z = [2 (48 - 0,5) - 60] / \sqrt{60} = 4,52***$

Por lo tanto, se rechaza la H_0 de que el cambio de opinión se debe al azar, pues tiene evidencia altamente significativa que justifica su esfuerzo por mejorar la imagen de su farmacia.

14.4.2 La prueba de rangos de Wilcoxon

En la prueba del Signo se utiliza la información acerca de la dirección de las diferencias encontradas en la pareja de muestras. Pero no se considera la *magnitud* relativa de tales diferencias. En cambio, en el modelo de Wilcoxon se toma en cuenta ambas cosas y por eso es más poderoso que el del Signo. Acá se le da un peso a cada signo, relativo a la magnitud de la diferencia encontrada. Se la denomina también: *Prueba de rangos señalados y pares encontrados*. El investigador debe hacer dos cosas básicas al examinar una pareja de datos:

- 1) *determinar en la pareja cual es el "mayor" de ambos;*
- 2) *ordenar por rango las diferencias halladas* (rango acá significa "ranking" u orden)

Entonces puede usar este test para validar la hipótesis nula de que no hay diferencias entre los pares debido al tratamiento aplicado. En casos donde la valoración se hace en forma subjetiva, usando puntajes, también se puede aplicar el modelo de Wilcoxon (Psicología, Sociología, etc.). En la Bioquímica se la puede emplear cuando se valora el tamaño de una reacción ante una droga, vacuna, coloración, etc., con algo similar a un puntaje. En Farmacia ocurre otro tanto cuando se valoran con puntos la eficacia de medicamentos, o con escalas las encuestas de opinión en técnicas de mercadeo, propaganda, etc. *Esta prueba también puede usarse en una muestra única donde deseen hacer inferencias acerca de la mediana:* aquí al valor supuesto en la hipótesis nula se le resta a cada observación realizada y se tienen las diferencias buscadas para hacer el test.

El supuesto básico para poder usarlo es:

las magnitudes con las que se trabaja provienen de una distribución simétrica.

No importa si cada muestra proviene de una población distinta, lo importante es que ambas deben provenir de poblaciones con distribuciones simétricas. El método puede ser resumido en:

Paso 1) Se determina las diferencias D_i entre los pares de observaciones realizadas.

Paso 2) Sin tomar en cuenta el signo, se ordenan en forma creciente. Las ligas se descartan pues en este modelo no se toman en cuenta los empates únicamente se consideran los rangos de las diferencias encontradas (notar que ahora *rango* no significa lo mismo que lo visto en el capítulo 4).

Paso 3) Se coloca el signo a cada uno de los rangos R_i hallados. Se suman entre sí los rangos de las diferencias positivas calculando su total T_+ , y el de las negativas obteniendo T_-

Paso 4) Se elige la menor de ambas sumas y se la define como el estadígrafo T .

Paso 5) Se compara el valor T obteniendo con el valor crítico T_{α} para tomar decisiones.

La hipótesis nula es que los tratamientos aplicados son equivalentes. Esto es, la suma de los rangos positivos y negativos son aproximadamente iguales. Habrá diferencias de ambos signos pero con valoraciones que se equiparan. En cambio, si las sumas de rangos son muy diferentes, se puede deducir que el efecto del factor analizado no es despreciable ni producto del azar.

En este modelo puede haber dos tipos de ligas. La primera ocurre cuando no se aprecia diferencia entre la pareja analizada ($D_i = 0$) y se deja de lado en los cálculos, al igual que en el caso del modelo anterior. La segunda ocurre cuando se produce un empate en el valor de las diferencias ($D_i = D_j = \dots = D_k$). O sea, la diferencia de rangos no es nula sino que tiene el mismo módulo que otra diferencia. Es posible hallar 1, 2, 3, ..., k empates. En este caso, la solución es promediar el valor de los rangos empatados hallados E_i y asignárselo a cada uno de los empates. Esto es, a cada empate se le da un valor $E_i = (R_1 + R_2 + \dots + R_k) / k$.

De acuerdo con el tamaño muestral se pueden presentar dos casos:

(I) *Muestras pequeñas* ($6 \leq n \leq 25$): en estos casos se usa la Tabla 10 del Anexo de rangos señalados de Wilcoxon. Allí se entra con el tamaño muestral y el nivel α , según sea de una o dos colas, para encontrar el valor crítico T_α , el cual se compara con el valor T calculado como la menor, de la suma de los dos rangos positivo y negativo. Si $T \leq T_\alpha$ se puede rechazar la H_0 al nivel de significación α adoptado, caso contrario se acepta. Por ejemplo, si $n = 15$ y $T = 21$ entonces $T_\alpha = 25$ es mayor y se rechaza H_0 al nivel de 0,05 en dos colas. En este caso se rechaza la hipótesis de que ambos grupos no difieren entre sí y se concluye que hay evidencia significativa que prueba que hay diferencia entre ambos. No se puede rechazar al nivel 0,01.

(II) *Muestras grandes* ($n > 25$): cuando la muestra es grande se usa la aproximación normal pues se ha demostrado que para $n > 25$ el estadígrafo T se aproxima a la función de Gauss en forma aceptable. Para tipificar el estadígrafo se usa:

$$Z_T = \frac{T - \mu_T}{\sigma_T} \quad \text{y se lo compara con } Z_\alpha \text{ como es habitual, con:}$$

$$\mu_T = (n/4)(n+1) \quad \text{y} \quad \sigma_T^2 = (n/24)(n+1)(2n+1) = \text{VAR}(T)$$

Donde n es el total de parejas de datos medidas útiles, esto es aquellas parejas cuyas diferencias no son nulas. Entonces, $n = N - D(0)$. Los siguientes ejemplos ilustran el uso de este modelo:

Caso 1) Muestras pequeñas: una determinación clínica se realizó en 8 pacientes que concurren al laboratorio usando el método A; se repitió la medición pero usando otra marca de espectrofotómetro (método B). Los resultados obtenidos fueron los del cuadro siguiente. Averiguar si es lo mismo usar uno u otro espectro.

Paciente	Método A	Método B	Diferencias D_i	Rangos R_i	T
1	87	76	11	8	T = 4
2	66	56	10	7	
3	75	76	-1	-1	
4	95	86	9	6	
5	73	76	-3	-3	
6	88	86	2	2	
7	69	64	5	4	
8	77	70	7	5	

La menor de las sumas de rangos señalados es $T = 4$. Como no hay diferencias nulas encontradas, resulta $N = n = 8$, de la Tabla 10 del Anexo se obtiene el valor $T\alpha = 4$. Como este estadígrafo no es menor que el valor crítico, se rechaza la hipótesis nula de igualdad entre ambos espectros con un nivel significación de 0,05 para un ensayo de dos colas.

Este mismo problema se puede resolver con el modelo del signo. Para ello, se calcula un valor de $x = 2$ y de la tabla respectiva se obtiene una posibilidad asociada $Bx = 2p = 2(0,145) = 0,029$ para una prueba de dos colas. Acá no se puede rechazar la hipótesis nula (H_0) $x = 2$ para un valor de $\alpha = 0,05$. A primera vista se tienen entonces dos resultados diferentes usando estos dos modelos. Es importante destacar que no son contradictorios entre sí. El modelo del Signo tiene menor sensibilidad y no puede detectar diferencia alguna entre ambos métodos clínicos. No puede discriminar la pequeña diferencia que implica cambiar el espectrofotómetro en la técnica realizada. En cambio, el modelo de Wilcoxon, más sensible, ya detecta diferencias entre ambos. Si bien en el límite, pero suficiente para tener una prueba científica. Esto no es extraño pues este modelo, además de emplear la información del signo, agrega más información con los rangos. Es una especie de moderación o prorrateo de los signos: no todos pesan igual a la hora de contarlos.

Si se pudiese aplicar un modelo paramétrico porque se cumplen los supuestos básicos, el apropiado sería el modelo de Student para comparar dos muestras apareadas, visto en el capítulo anterior. En tal caso resultaría:

$$t = \frac{\bar{d} - 0}{DS_d / \sqrt{n}} = \frac{5 - 0}{5,21 / \sqrt{8}} = 2,7^* \text{ que está claramente en la zona de rechazo del nivel } 0,05.$$

Entonces, si este caso cumpliera los requisitos adecuados se podría resolver con los tres modelos: el de Student, que es el más poderoso de los tres, que detecta claramente las diferencias entre ambos espectrofotómetros; el no paramétrico de Wilcoxon que las detecta, pero en el borde de la zona de rechazo; y el más débil del Signo que no llega a detectar las diferencias. Con lo cual se puede plantear una regla general:

Se debe usar el modelo más sensible disponible para cada caso.

Caso 2) Muestras grandes: se desea analizar si el efecto de dos antibióticos de amplio espectro, el A y el B, es equivalente en el tratamiento de infección bacteriana. Para ello se eligen 30 muestras al azar de otros tantos pacientes de bacteriología. Se siembran y desarrollan en cajas de Petri durante 2 días con el mismo caldo de cultivo. Luego se colocan las dos drogas en cada uno, y a la 24 hs se mide el desarrollo de las aureolas, sus diámetros aproximados, como índice del poder de ambas drogas. Interesa destacar las diferencias entre ellas como se muestra en el cuadro siguiente, con $D_i = A_i - B_i$:

Paciente	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Di	-2	0	0	1	0	0	4	4	1	1	5	3	5	3	-1
Rangos	-11,5			4,5			20	20	4,5	4,5	23	16,5	23	16,5	-4,5
Paciente	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Di	1	-1	5	8	2	2	2	-3	-2	1	4	8	2	3	-1
Rangos	4,5	-4,5	23	25,5	11,5	11,5	11,5	-16,5	-11,5	4,5	20	25,5	11,5	16,5	-4,5

El rango menos frecuente es el de signo negativo: $T = 11,5 + 4,5 + 4,5 + 16,5 + 11,5 + 4,5 = 53$
Por su parte, hay 4 casos de diferencias nulas que no se toman en cuenta, entonces resulta el valor $n = 30 - 4 = 26$ y reemplazando en las fórmulas para el cálculo del estadígrafo T queda:

$$\mu_T = (n / 4) (n + 1) = (26/4) (26+1) = 175,5$$

$$\sigma_T^2 = (n / 24) (n + 1) (2n + 1) = (26/24) (26+1) (52+1) = 1550,25 \text{ o sea, } \sigma_T = 39,37 = SE(T)$$

$$Z_T = \frac{T - \mu_T}{\sigma_T} = (53 - 175,5) / 39,37 = -3,11^{**} \quad (-3,29 < -3,11 < -2,58)$$

Se ha encontrado evidencia muy significativa que prueba la diferencia entre ambas drogas testeadas en el experimento. Por lo que se rechaza la (H_0) de igualdad entre ambas (dos colas).

14.5 Modelos para dos muestras independientes

Existe una variedad de modelos no paramétricos para comparar dos muestras entre sí. El grupo de modelos que usan magnitudes discretas (datos de recuentos) se deja para el capítulo dedicado al Análisis de Frecuencias que se verá más adelante. Entre las demás cabe mencionar:

La *prueba de aleatoriedad* para dos muestras independientes es un modelo que emplea valores numéricos como puntajes y se requiere que la magnitud considerada tenga una medida de intervalo. Sirve para comparar a las dos muestras en medidas de posición cuando ambos tamaños muestrales son pequeños. Cuando los tamaños son grandes, los cálculos se tornan muy engorrosos y es preferible usar la aproximación con el Modelo Student. Sin embargo, el modelo de la U puede considerarse una prueba de aleatoriedad aplicada a los rangos de las observaciones realizadas. Whitney, en 1948, mostró que hay ocasiones en que el modelo de la U es más poderoso que el de Student y es la mejor alternativa.

La *prueba de Moses para reacciones extremas* es un modelo con muchas aplicaciones en las ciencias de la conducta. Está diseñado para estudiar las reacciones de los individuos ante una condición experimental que les provoque diferencias en la dirección a conductas extremas. Por ejemplo, la inestabilidad política o una depresión económica, provoque una radicalización en las opiniones políticas o sociales de algunas personas. Esta prueba fue concebida específicamente para ser usada con datos (por lo menos ordinales) organizados para estudiar tales hipótesis. Sin embargo, también es aplicable cuando el investigador espera calificaciones altas para un grupo y bajas para el otro, dentro del área educativa. Pero el mismo autor señala que la prueba de la U es más eficiente en esos casos y por lo tanto preferible a este modelo.

La *prueba de rachas de Wald-Wolfowitz* es aplicable a casos donde la (H_0) supone que las dos muestras fueron extraídas de la misma población, frente a la alternativa que difieren totalmente entre sí. Si las muestras son lo suficientemente grandes esta prueba detecta diferencias en la tendencia central de las poblaciones, en variabilidad, en oblicuidad, casi en cualquier cosa. Sin embargo, exige el supuesto que la magnitud considerada provenga de una distribución continua y que las medidas efectuadas sean por lo menos ordinales. Ambos autores consideran que el

modelo de la U es más potente que el de las rachas, especialmente en el caso de dos muestras extraídas de poblaciones con la misma ubicación. La razón es que la de rachas está diseñada para detectar diferencias de cualquier tipo; en cambio, el modelo de la U es específico para descubrir diferencias de este tipo y no pierde potencia en buscar cualquier desvío.

Repasando los tres modelos mencionados antes, se ve que todos son superados en potencia por el modelo de la U, el cual se constituye en la mejor alternativa frente al modelo paramétrico equivalente para comparar dos muestras independientes: el de Student. Por esa razón no se desarrollan en la presente obra y se deja a la inquietud del lector que puede encontrarlos en los trabajos de Siegel, Hollander, Mc Collough, Moses y otros mencionados en la bibliografía. Otro motivo para desarrollar el modelo de la U es su uso cada vez más extendido en Medicina, Bioquímica y Farmacia.

14.5.1 Modelo U de Mann-Whitney

Este modelo U sirve para testear si dos muestras independientes han sido tomadas de la misma población. Se tiene, por lo menos, una magnitud ordinal de la misma. Este es el modelo no-paramétrico más poderoso para comparar dos muestras cuando no son apareadas. Es para el caso donde se tiene dudas acerca de la verificación de los supuestos que piden el modelo Student, o cuando las medidas son ordinales. La hipótesis de trabajo (H_0), siempre es que ambas muestras provienen de la misma población.

El procedimiento a seguir para usar este método es como sigue:

Paso 1) Se ordenan todos los datos, de menor a mayor, de ambas muestras y en un solo conjunto, cuidando de identificar a cada uno con su muestra respectiva.

Paso 2) Se determina la muestra de referencia. Conviene que sea la de menor tamaño, caso contrario la muestra de control, placebo o blanco.

Paso 3) Comenzando con el menor valor, se cuenta el número de muestras que preceden al primero de la muestra control (U_1); luego se busca el segundo de la muestra control y se cuenta el número de muestras precedente del otro grupo (U_2); después se ubica al tercer valor y se procede en forma análoga para determinar (U_3), y así sucesivamente hasta recorrer toda la muestra de control o referencia. En caso de empate, se le asigna medio punto a cada uno.

Paso 4) Se obtiene el estadígrafo $U = U_1 + U_2 + U_3 + \dots$

Paso 5) Se procede a comparar este valor contra el valor crítico de tablas.

Hay dos tipos de tablas para el modelo U. La Tabla 11 del Anexo corresponde a muestras pequeñas (ninguna de las muestras es mayor que 9), y arroja el valor de la probabilidad del estadígrafo U calculado en el Paso 4. La Tabla 12 corresponde a un tamaño mediano de las muestras (entre 9 y 20 cada una), y da un valor crítico U_α que se debe comparar con el obtenido experimentalmente. Para $n > 20$ se usa la aproximación con la función de Gauss. En la práctica se pueden presentar tres casos posibles:

(I) *Muestras muy pequeñas* ($n_1 < 9$ y $n_2 < 9$): cuando ninguna de las muestras es mayor que 8, se emplea la Tabla 11 del Anexo. El procedimiento para usar esta tabla es como sigue.

Al mayor tamaño muestral se lo denomina n_2 y sirve para ubicar la correspondiente celda dentro de la Tabla 11 (hay 6 celdas una para cada tamaño entre 3 y 8). Al menor tamaño muestral se lo denomina n_1 y se usa para ubicar la columna correspondiente dentro de la celda. El valor de probabilidad (p) se saca cruzando esta columna con la fila que corresponde al estadígrafo U calculado en el Paso 4. La tabla está construida para ensayos de una sola cola, en cuyo caso se hace $U\alpha = p$. Cuando el ensayo sea de dos colas será $U\alpha = 2 p$. Finalmente este valor $U\alpha$ debe compararse con el valor U calculado en el Paso 4. Si el valor de U es tan grande que no aparece en la tabla, se usa la transformación $U' = n_1 n_2 - U$ como se muestra en el ejemplo siguiente:

Ejemplo 1) A 3 alumnos elegidos al azar se les enseñó una nueva técnica de preparar un examen final, un día antes de la fecha, otros 4 alumnos; sin preparación especial, se eligieron para control. Los resultados del examen en el grupo entrenado fueron: (10, 8 y 6), los del grupo control obtuvieron (9, 7, 5 y 4). Ordenado los datos en forma ascendente es:

Pareja:	4	5	6	7	8	9	10
Grupo:	C	C	E	C	E	C	E

La nota más chica es del grupo control y no da puntos al igual que la siguiente. La tercer nota solo (7) tiene un valor del otro grupo, otorga un punto, la cuarta nota del control (9) tiene dos notas menores del otro grupo y entrega dos puntos; luego $U = 0 + 0 + 1 + 2 = 3$.

De tablas con los tamaños muestrales de $n_1 = 3$ y $n_2 = 4$, se entra por la fila de $U = 3$ y se encuentra una probabilidad $p = 0,2$ para una cola. Como este ensayo es para ver si hay diferencias, es de dos colas, por lo que debe duplicarse este valor $U\alpha = 0,4$. Comparando con el nivel habitual de significación es $U\alpha = 0,4 > \alpha = 0,05$. Luego, no se puede rechazar la hipótesis nula pues entra en la zona de aceptación. La preparación especial no tuvo efecto en el rendimiento de los alumnos en el examen final.

Cabe destacar que si en vez del grupo control se hubiese tomado el otro como referencia, el puntaje hubiese sido $U = 2 + 3 + 4 = 9$: la tabla no llega al valor 9 en la celda de $n_2 = 4$. Por lo tanto, se debe usar el valor $U' = n_1 n_2 - U = 12 - 9 = 3$, que resulta ser igual al valor hallado anteriormente.

(II) *Muestras medianas* (n entre 9 y 20): cuando el tamaño muestral de la muestra más grande de las dos es mayor que 9 ($n_2 \geq 9$), la tabla anterior no puede usarse. Se debe emplear en cambio la Tabla 12 del Anexo. No interesa si el tamaño de la otra muestra es menor que 9. Esta tabla arroja valores críticos de U , llamados U_c , que se deberán comparar con el calculado en el Paso 4. No da probabilidades como la Tabla 11 por lo que se hace el ensayo comparando los valores de U para decidir si se acepta o rechaza la H_0 .

Como el trabajo de calcular U puede ser engorroso cuando la muestra es algo grande, se muestra una alternativa de cálculo más sencilla, como sigue:

Pasos 1 y 2) Son iguales a los detallados más arriba.

Paso 3) Se coloca el rango correspondiente debajo de la sucesión ordenada. Así, corresponde el rango 1 al menor de los valores, el 2 al siguiente y así sucesivamente. En casos de empates se coloca el rango promedio entre los empatados.

Paso 4) Se calcula $R1 = r_i + r_j + \dots$ como la sumatoria de todos los rangos hallados correspondientes a la muestra de tamaño n_1 . Y $R2$ a su correspondiente de la muestra n_2 .

Paso 5) Se calculan dos estadígrafos con:

$$U1 = (n_1 \cdot n_2) + (n_1/2) (n_1 + 1) - R1$$

$$U2 = (n_1 \cdot n_2) + (n_2/2) (n_2 + 1) - R2$$

Y se elige al menor de ellos para adoptarlo como el estadígrafo muestral U . Luego si este valor es menor o igual que el valor crítico de tablas, se puede rechazar la H_0 . Si $U > U_c$ es lo contrario.

Ejemplo) En un ensayo de valoración se escogieron 16 jueces al azar. A 10 de ellos se les pidió que calificaran de 0 a 100 el sabor de un nuevo jarabe que se está por lanzar al mercado. A los 6 restantes se les hizo probar el sabor tradicional para usarlos como control. La prueba se hizo a ciegas; ni los jueces ni el organizador sabían cómo serían repartidos. Se obtuvieron los resultados siguientes ordenados de menor a mayor:

Valores	48	52	53	67	73	79	81	85	86	88	90	92	93	95	97	99
Grupo	E	C	C	E	C	E	E	E	C	E	C	E	E	E	C	E
Rango	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Por lo tanto se calculan:

$$R1 = 2 + 3 + 5 + 9 + 11 + 15 = 45$$

$$R2 = 1 + 4 + 6 + 7 + 8 + 10 + 12 + 13 + 14 + 16 = 91$$

$$U1 = (10)(6) + (6/2)(6 + 1) - 45 = 36$$

$$U2 = (10)(6) + (10/2)(10 + 1) - 91 = 24$$

Se adopta $U = 24$ - Notar que $U' = (10)(6) - 24 = 36$ (llega al mismo menor valor) -

De tablas surgen para $n_1 = 6$ y $n_2 = 10$ (una sola cola): $U_{0.05} = 14$; $U_{0.01} = 8$ y $U_{0.001} = 3$

Como el valor muestral $U = 24$ es mayor que los valores críticos, no se puede rechazar la H_0 .

(III) *Muestras grandes* ($n > 20$): cuando el tamaño muestral de una de las muestras es grande, o sea $n > 20$, la distribución de U se acerca a una gaussiana $N(0,1)$, con los parámetros siguientes:

$$Z_U = \frac{U - \mu_u}{\sigma_u} \quad \text{y se lo compara con } Z_\alpha \text{ como es habitual, donde:}$$

$$\mu_u = (n_1 \cdot n_2) / 2$$

$$\sigma_u^2 = (n_1 \cdot n_2) (1/12) (n_1 + n_2 + 1) = \text{VAR}(u)$$

Para el cálculo del estadígrafo U se procede como se explicó para el caso de muestras de tamaño mediano, en el punto anterior.

Ejemplo) Para comparar dos grupos de alumnos luego de un examen parcial, la cátedra eligió al azar a 22 de la carrera de Farmacia y 16 de Bioquímica, a los cuales se les tomaron los mismos temas, el mismo día y a la misma hora. Los resultados ordenados fueron:

P	30	35	40	40	40	40	48	50	55	58	60	60	60	65	65	65	70	70	75
C	B	B	B	B	F	F	B	B	F	F	B	B	F	B	F	F	F	B	F
R	1	2	4,5	4,5	4,5	4,5	7	8	9	10	12	12	12	15	15	15	17,5	17,5	19

P	78	78	78	79	80	80	80	83	84	85	85	85	90	90	90	92	95	95	95
C	B	B	F	F	B	F	F	B	F	B	F	F	B	F	F	F	F	F	F
R	21	21	21	23	25	25	25	27	28	30	30	30	33	33	33	35	37	37	37

$$R_B = 1 + 2 + 4,5 + 4,5 + 7 + 8 + 12 + 12 + 15 + 17,5 + 21 + 21 + 25 + 27 + 30 + 33 = 240,5$$

$$R_F = 4,5 + 4,5 + 9 + 10 + 12 + 15 + 15 + 17,5 + 19 + 21 + 23 + 25 + 25 + 28 + 30 + 30 + 33 + \dots + 33 + 35 + 37 + 37 + 37 = 500,5$$

$$\text{Luego es: } U_B = (n_B \cdot n_F) + (n_B/2) (n_B + 1) - R_B = (16 \cdot 22) + (16/2)(16+1) - 240,5 = 247,5$$

$$U_F = (n_B \cdot n_F) + (n_F/2) (n_F + 1) - R_F = (16 \cdot 22) + (22/2)(22+1) - 500,5 = 104,5$$

Se elige al menor como $U = 104,5$ (Farmacia) y $U' = 22(16) - 104,5 = 247,5$

$$\mu_u = (n_1 \cdot n_2) / 2 = (22 \cdot 16) / 2 = 176$$

$$\sigma_u^2 = (n_1 \cdot n_2) (1/12) (n_1 + n_2 + 1) = 1144. \text{ Luego es } \sigma_u = 33,8 \text{ y se calcula Z con:}$$

$$Z_U = (U - \mu_u) / \sigma_u = (104,5 - 176) / 33,8 = -2,12^*: \text{ Hay una diferencia significativa en el rendimiento de ambos grupos siendo los de Farmacia mejor que los de Bioquímica.}$$

14.6 Modelo de Cochran para n muestras apareadas

Cuando se tiene el caso de N individuos, a cada uno de los cuales se le efectuó una medición de una magnitud clínica dicotómica (o binaria) mediante n técnicas diferentes ($n > 1$), se tiene el caso de mediciones repetidas al mismo paciente. Esto es un total de n valores apareados para cada paciente. Los resultados de cada test se pueden expresar como sano o enfermo, (+) o (-) etc. Lo mismo ocurre cuando n jueces opinan sobre N casos diferentes. Suponiendo que la magnitud clínica es X, entonces el valor X_{ij} corresponde al paciente i medido con el método j . Los resultados obtenidos se pueden expresar como se hace en la Tabla 14.1 siguiente:

Tabla 14.1. N pacientes medidos con n métodos clínicos

Pacientes	Test 1	Test 2	...	Test j	...	Test n	Total
1	X ₁₁	X ₁₂	...	X _{1j}	...	X _{1n}	T _{1.}
2	X ₂₁	X ₂₂	...	X _{2j}	...	X _{2n}	T _{2.}
3	X ₃₁	X ₃₂	...	X _{3j}	...	X _{3n}	T _{3.}
...
i	X _{ij}	X _{ij2}	...	X _{ijj}	...	X _{ijn}	T _{i.}
...
N	X _{N1}	X _{N2}	...	X _{Nj}	...	X _{Nn}	T _{N.}
Total	T _{.1}	T _{.2}	...	T _{.j}	...	T _{.n}	T

Donde como convención se adopta X = 0 cuando es (+) y X = 1 cuando es (-) entonces los valores observados son una serie de ceros y unos. Definiendo las sumatorias:

$$T_{.j} = \sum_{i=1}^N X_{ij} : \text{El total de la columna } j. \quad T = \sum_{i=1}^N \sum_{j=1}^n X_{ij} : \text{El total de todas las observaciones}$$

$$T_{i.} = \sum_{j=1}^n X_{ij} : \text{El total de la fila } i.$$

$$RS = \sum_{i=1}^N (T_{i.})^2 : \text{La suma de los cuadrados de las filas.}$$

$$CS = \sum_{j=1}^n (T_{.j})^2 : \text{La suma de los cuadrados de las columnas.}$$

Entonces se puede efectuar un test estadístico para ver si existe una diferencia significativa entre los tests clínicos llamado: *La prueba Q de Cochran*. El estadístico Q se calcula con:

$$Q = \{(n-1) [(n \cdot CS) - T^2]\} / [n T - RS]$$

El cual tiene aproximadamente una distribución Chi-cuadrado con n – 1 grados de libertad.

Ejemplo) Se han examinado 12 pacientes mediante tres diferentes técnicas para detectar toxoplasmosis. Si los resultados se informan como (+) y (-) detectar si hay diferencias significativas entre los tres métodos. Usando la convención de 1 para los (+) y 0 para los (-) resultó:

Pacientes	Test 1	Test 2	Test 3	Total
1	1	1	0	2
2	0	0	0	0
3	1	1	1	3
4	1	0	1	2
5	1	1	1	3
6	0	0	1	1
7	1	1	1	3
8	0	0	0	0
9	1	1	0	2
10	0	0	1	1
11	0	1	0	1
12	1	0	1	2
Total	7	6	7	20

$$T = \sum_{i=1}^N \sum_{j=1}^n X_{ij} = 20$$

$$RS = \sum_{i=1}^N (T_{i.})^2 = 49 + 36 + 49 = 134$$

$$CS = \sum_{j=1}^n (T_{.j})^2 = 4 + 0 + \dots + 4 = 46$$

$$Q = \{(2) [(138) - 400]\} / [60 - 134]$$

$$Q = 7,08^* > \chi^2_{0,95; 2} = 5,991$$

La conclusión es que se han encontrado diferencias significativas entre los tres métodos.

Para el caso particular de tener que comparar dos tests clínicos es $n = 2$ y el problema se puede analizar como se muestra a continuación:

Tabla 14.2. N pacientes medidos con 2 métodos clínicos

Pacientes	Test 1	Test 2	Total
1	X_{11}	X_{12}	$T_{1.}$
2	X_{21}	X_{22}	$T_{2.}$
3	X_{31}	X_{32}	$T_{3.}$
...
i	X_{i1}	X_{i2}	$T_{i.}$
...
N	X_{N1}	X_{N2}	$T_{N.}$
Total	$T_{.1}$	$T_{.2}$	T

Esto resulta equivalente a presentar los datos en una tabla como la siguiente:

Tabla 14.3. Tabla de Concordancia entre dos tests clínicos

		Test 1		Total
		(+)	(-)	
Test 2	(+)	r_{11}	r_{12}	$r_{11} + r_{12}$
	(-)	r_{21}	r_{22}	$r_{21} + r_{22}$
Total		$r_{11} + r_{21}$	$r_{12} + r_{22}$	N

Donde el significado de las frecuencias es

- r_{11} : los pacientes que muestran ambos resultados (+) : concordancia positiva.
- r_{22} : los pacientes que muestran ambos resultados (-) : concordancia negativa.
- r_{21} : los pacientes que son (+) con el primer test y (-) con el segundo (discordancia).
- r_{12} : los pacientes que son (-) con el primer test y (+) con el segundo (discordancia).

Como los valores X_{ij} solo pueden ser 0 o 1, el total de cada fila de la Tabla 14.2 puede ser 0 cuando ambos resultados son (-), o 1 cuando un resultado es (+) y el otro (-) y 2 cuando ambos resultados son (+). Por su parte, el total de cada columna será igual al total de casos (+) en cada test. Por lo tanto el total de cada columna será:

$$T_{.1} = \sum_{i=1}^N X_{ij} = r_{11} + r_{21} \quad \text{y} \quad T_{.2} = \sum_{i=1}^N X_{ij} = r_{12} + r_{22} \quad \text{Luego el total general es:}$$

$$T = \sum_{i=1}^N \sum_{j=1}^n X_{ij} = T_{.1} + T_{.2} = 2 r_{11} + r_{21} + r_{12} \quad \text{Pero el total ambas filas también es T:}$$

$$T = \sum_{i=1}^N \sum_{j=1}^n X_{ij} = \sum_{i=1}^N T_{i.} = 0 (r_{22}) + 1 (r_{12}) + 1 (r_{21}) + 2 (r_{11}) \quad \text{Por lo tanto}$$

$$RS = 0^2 (r_{22}) + 1^2 (r_{12}) + 1^2 (r_{21}) + 2^2 (r_{11}) = (r_{12} + r_{21}) + 4 r_{11} \quad \text{Y así se puede calcular}$$

$$CS = (r_{11} + r_{21})^2 + (r_{11} + r_{12})^2 \quad \text{Con lo cual la fórmula del estadígrafo Q se reduce a:}$$

$$Q = [2.CS - T^2] / [2 T - RS]$$

Notar que si se define como 1 a los casos de (-) y como 0 a los (+) el resultado es el mismo. Lo más importante para destacar es que si r_{11} y/o r_{22} cambia sus valores Q no varía.

Ejemplo 1) Se han efectuado dos mediciones de VDRL a cada uno de 16 pacientes con dos marcas diferentes de kits. Se desea saber si las marcas son equivalentes. Los resultados fueron

Pacientes	Test 1	Test 2	Total	
1	1	1	2	$T = \sum_{i=1}^N \sum_{j=1}^n X_{ij} = 15$ $RS = \sum_{i=1}^N (T_i)^2 = 64 + 49 = 113$ $CS = \sum_{j=1}^n (T_j)^2 = 4 + 0 + \dots + 1 = 25$ $Q = \{(1) [(2 \cdot 25) - 225]\} / [2 \cdot 15 - 113] = 2,11$ $Q = 2,11 < \chi^2_{0,95; 1} = 3,871$ Los resultados no fueron significativos. Esto indica que no hay diferencias estadísticamente significativas entre ambos tests.
2	0	0	0	
3	1	1	2	
4	1	0	1	
5	1	1	2	
6	0	0	0	
7	1	1	2	
8	0	0	0	
9	1	1	2	
10	0	0	0	
11	0	1	1	
12	1	0	1	
13	0	0	0	
14	0	1	1	
15	0	0	0	
16	1	0	1	
Total	8	7	15	

También se puede resolver este problema con la Tabla de Concordancia:

		Test 1		Total
		(+)	(-)	
Test 2	(+)	5	3	8
	(-)	2	5	7
Total		7	8	15

$$RS = (r_{12} + r_{21}) + 4 r_{11} = (3+2) + 4 \cdot 5 = 25$$

$$CS = (r_{11} + r_{21})^2 + (r_{11} + r_{12})^2 = (5+2)^2 + (5+3)^2 = 49 + 64 = 113$$

$$Q = [2.CS - T^2] / [2 T - RS] = [2 \cdot 25 - 225] / [2 \cdot 15 - 113] = 2,11 < \chi^2_{0,95; 1} = 3,871$$

14.7 Modelo de Cohen-Kappa

Este modelo se usa cuando se tienen dos métodos de diagnóstico diferentes A y B cuyos resultados se expresan en forma de dos o más categorías (dicotómica o polótómica). Cada muestra individual es juzgada por cada uno de los métodos en alguna de las clases posibles. El estadígrafo Kappa (K) da una medida del acuerdo que existe entre ambos métodos clínicos. Para ilustrar este procedimiento se presenta el siguiente ejemplo:

- Se desea ver si se puede reemplazar el antibiótico A que se está usando en el laboratorio, por uno nuevo B. El resultado de cada medición de acuerdo al diámetro de la aureola, se agrupa en tres categorías: R (resistente), S (sensible) y M (muy sensible). Se efectúan 100 pruebas con los resultados siguientes:

Frecuencias Observadas

		B			
		M	S	R	
A	M	44	5	1	50
	S	7	20	3	30
	R	9	5	6	20
		60	30	10	100

Paso 1) Se calcula el valor q como la sumatoria de los productos de los totales marginales de filas y columnas de cada clase, dividido el tamaño muestral:

$$q = q_M + q_S + q_R = [(50 \cdot 60) + (30 \cdot 30) + (20 \cdot 10)] / 100 = 41$$

Paso 2) Se calcula la sumatoria de los valores encontrados en la diagonal principal, o sea el número de las coincidencias encontradas:

$$c = 44 + 20 + 6 = 70$$

Paso 3) Se obtiene el estadígrafo Kappa (K) con:

$$K = (c - q) / (N - q) = (70 - 41) / (100 - 41) = 0,492$$

Esto significa que de todas las muestras que se esperaba que no sean concordantes, solo el 49,2% de ellas son de hecho concordantes. En general, cuanto más cerca este Kappa de 1, mejor será la concordancia encontrada. Pero para el ejemplo presente el máximo valor posible de Kappa será $K = 0,835$. Lo que se deduce del cuadro siguiente:

		B			
		M	S	R	
A	M	50	0	0	50
	S	0	30	0	30
	R	10	0	10	20
		60	30	10	100

Respetando los totales marginales, y haciendo máximas las coincidencias en la diagonal principal, se puede obtener el caso ideal de concordancia máxima posible dados los totales marginales encontrados. Para esa tabla ideal el valor $K = 0,835$. Luego el porcentaje entre el valor de K observado y su máximo posible es del 59,2%

Paso 4) Se calcula el error estándar de Kappa con la formula aproximada siguiente:

$$SE (K) = \sqrt{\frac{q}{N \cdot (N - q)}} = \sqrt{\frac{41}{100 \cdot (59)}} = 0,08$$

Paso 5) El intervalo de confianza al 95% se calcula mediante la aproximación normal

$$\text{Limite inferior} = K - 1,96 SE (K) = 0,492 - 1,96 (0,08) = 0,335$$

$$\text{Límite superior} = K + 1,96 SE (K) = 0,492 + 1,96 (0,08) = 0,649$$

Luego es: K 95% CI (0,335 ; 0,649)

Paso 6) El test estadístico se puede hacer usando la aproximación normal con:

$$z = \frac{c - q}{\sqrt{\frac{q \cdot (N - q)}{N}}} = \frac{70 - 41}{\sqrt{\frac{41 \cdot (59)}{100}}} = 5,9^{***} \quad \text{Resulta ser altamente significativo.}$$

Este resultado indica que se puede rechazar la H_0 y aceptar que existe una asociación estadística muy fuerte entre ambos antibióticos. Este tipo de asociación matemática es aceptado en Medicina como si fuese una asociación clínica. La concordancia entre ambas pruebas se toma como aceptable, clínicamente hablando. Sin embargo, no se debe olvidar que no es una asociación del tipo causa-efecto, sino que se trata del concepto de independencia estadística visto en el Tema 6.

Notar que en el caso poco frecuente de que los métodos A y B terminan con el mismo número de muestras en la categoría M, el mismo número en la S y así hasta la última categoría, en este caso particular R, entonces el valor máximo de Kappa sería uno. Pero cuando se termina eligiendo números diferentes en cada categoría, como en el ejemplo anterior, el valor máximo de Kappa será siempre menor a uno ($K = 0,835$ en el ejemplo).

También se puede analizar el grado de acuerdo dentro de cada categoría por separado. En el ejemplo visto para la categoría M hay 44 acuerdos en total entre A y B, pero el antibiótico A define 6 casos más como si fuesen M, mientras que el antibiótico B define 16 casos más como M. Entonces se tienen en total 66 ($44 + 6 + 16$) muestras definidas como M por ambos antibióticos, de los cuales en solo 44 hay concordancia. Por lo tanto, el porcentaje de acuerdo para la categoría M será de $44/66 = 0,6667$, o sea un 66,7% de concordancia en dicha categoría. Análogamente, la concordancia para S es del 50% y en R es del 25%. Por otro lado, observando la tabla con los máximos acuerdos posibles se ve que para la categoría M hay en total 50 acuerdos en ambos antibióticos, con un total de $50 + 0 + 10 = 60$ entre A y B. Por eso, la máxima proporción de acuerdo posible en esa categoría es $50/60 = 0,833$ (83,3%) para comparar con el 66,7% encontrado, y así sucesivamente.

Este método fue muy popular durante años, pero desde entonces arrecian cada vez más las críticas contra el mismo. Finalmente las conclusiones actuales permiten decir que:

1) Kappa no debería ser aceptada como el estándar de concordancia o el método de referencia para obtenerla, cuando se comparan dos métodos clínicos en forma apareada. Es decir, cuando se mide dos veces a la misma muestra.

- 2) No debería usarse un estadígrafo como K que genera mucha controversia.
- 3) Se debería considerar a K como una de las alternativas disponibles para la decisión final.

Se pueden identificar dos usos posibles de Kappa: a) como una manera de testear la independencia estadística de los métodos y b) como una manera de cuantificar el grado de concordancia entre ambos métodos. Para el primer caso (a) se trata de testear una H_0 sobre que no hay más concordancia de la que podría haber dadas las circunstancias, esto es dado los totales marginales hallados. Pero de esta forma la decisión a tomar sobre si existe una asociación clínica, es puramente *cualitativa*, en vez de cuantitativa como debería ser. Se puede concluir que para este exclusivo propósito el uso de Kappa en clínica es adecuado. Pero para el segundo caso (b) no se puede decir lo mismo, porque no puede cuantificar adecuadamente el grado de concordancia, lo que significa que no se puede tomar como un índice clínico del tipo Sensibilidad, etc. Kappa debe ser interpretada como una proporción de las veces que los métodos podrían ser concordantes debido únicamente al azar, pero como los métodos no se aplican a muestras independientes, sino a muestras apareadas, los métodos son intrínsecamente dependientes y por lo tanto su aplicación o la relevancia de este término, en la concordancia es muy discutible.

El estadígrafo Kappa es apropiado para testear cuando la concordancia excede el grado de acuerdo debido al azar. Y puede ser calculado fácilmente. Estas son las dos ventajas del mismo. La lista de desventajas es muy larga y puede ser encontrada en varios libros mencionados en la Bibliografía.

14.8 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- 1) Los modelos no paramétricos (MNP) sirven para cualquier magnitud biológica. **V F**
- 2) En los MNP no se necesita hacer supuestos acerca de la forma de la distribución original. **V F**
- 3) Mencionar las ventajas y desventajas de trabajar con estadística no paramétrica: **.....**
- 4) Siempre que se pueda conviene usar modelos no paramétricos. **V F**
- 5) Los pasos a seguir para aplicar el modelo de la Binomial son: **.....**
- 6) En el modelo Binomial se pueden dar 3 casos posibles de acuerdo con el tamaño muestral. **V F**
- 7) La prueba de rachas de una muestra sirve para: **.....**
- 8) Los pasos a seguir en una prueba de rachas son: **.....**
- 9) En el modelo Rachas se pueden dar 3 casos posibles de acuerdo con el tamaño muestral. **V F**
- 10) Si las muestras son grandes, el estadígrafo de Rachas tiene una distribución $N(0,1)$. **V F**
- 11) Es lo mismo trabajar con dos muestras apareadas o dos muestras independientes. **V F**
- 12) Los supuestos básicos para poder usar el Modelo del Signo son: **.....**
- 13) Toda magnitud biológica puede ser dicotomizada. **V F**
- 14) Para el modelo del signo una muestra es grande si supera los 30 valores. **V F**
- 15) Los casos antes-después y droga-placebo se pueden estudiar con el modelo del Signo. **V F**
- 16) Si se aplican dos tratamientos al mismo individuo en n casos, las muestras son: **.....**
- 17) La prueba de Wilcoxon es más poderosa que su similar del Signo. **V F**
- 18) La prueba de Wilcoxon usa un peso relativo para cada signo empleado. **V F**
- 19) El supuesto para Wilcoxon es que las muestras provienen de una distribución simétrica. **V F**
- 20) Los pasos a seguir en el modelo de Wilcoxon son: **.....**
- 21) El tamaño muestral para Wilcoxon es grande si: $n > 20$ **V F**

- 22) Se debe usar el modelo más sensible (robusto) en cada caso investigado. **V F**
 23) Para dos muestras independientes los modelos que se pueden usar son:
 24) De todos los modelos para comparar dos muestras independientes, la U es el mejor. **V F**
 25) El modelo de la U es el mejor no paramétrico para poder comparar a muestras
 26) Los pasos a seguir en el modelo de la U son:
 27) En el modelo de la U se pueden presentar 3 casos de acuerdo con el tamaño muestral. **V F**
 28) Para la U las muestras son muy pequeñas si ninguna es mayor de 9. **V F**
 29) Si las muestras son medianas o grandes, los pasos a seguir para la U se modifican en.....
 30) El modelo de Cochran se puede emplear para comparar 3 o más métodos clínicos. **V F**
 31) El tamaño de muestra total N tiene gran influencia en el cálculo de Q. **V F**
 32) El número de casos concordantes influencia el valor de Q. **V F**
 33) Kappa es la manera de cuantificar el concepto de concordancia entre 2 métodos clínicos **V F**
 34) Enumerar las ventajas del estadígrafo Kappa:.....
 35) Realizar un esquema sinóptico sobre cual modelo no paramétrico elegir, en los casos de la realidad.

2) Resolver (cuando se pueda) los problemas desarrollados en los capítulos anteriores y los problemas propuestos, con los modelos no paramétricos vistos en este capítulo.

3) Comparar el poder entre los modelos de punto anterior.

4) Se eligieron 17 días al azar de un semestre y se determinaron las ventas en las sucursales A y B de una misma farmacia. Determinar si hay diferencias significativas en las recaudaciones de cada una.

A	300	350	240	450	140	340	248	250	155	258	360	260	610	165	170	90	120
B	240	320	270	370	200	190	210	180	140	200	290	340	400	200	120	80	150

5) Resolver el mismo problema anterior si el procedimiento es diferente: se eligen 34 días al azar y se asignan a cada farmacia con un sorteo posterior. Explicar cómo haría esta asignación usando una tabla de números al azar.

6) Resolver si hay asociación entre ambos tests clínicos 1 y 2, en los dos casos siguientes:

	Test 1	
Test 2	(+)	(-)
(+)	5	3
(-)	2	5

	Test 1		
Test 2	A	B	C
A	55	2	1
B	2	25	2
C	1	5	7

7) Decidir si hay diferencias significativas entre los diámetros de aureola de 2 antibióticos A y B en cajas de Petri, con los datos siguientes expresados de mm:

Diá.	128	140	178	179	180	185	189	193	194	200	205	210	215	220	220	224	230
Antib.	A	B	B	A	B	A	A	A	B	B	A	B	B	A	B	A	B

15

Análisis de frecuencias

En los capítulos anteriores se han analizado los modelos teóricos sobre muestras, inferencia y decisión estadísticas aplicadas a magnitudes biológicas de tipo continuo en los modelos paramétricos. Luego, se mostraron los modelos no paramétricos para analizar los casos en donde es dudoso el cumplimiento de los supuestos, o bien cuando se trabaja con otro tipo de magnitudes como las cualitativas. Ahora se presentan nuevos modelos estadísticos aptos para magnitudes de tipo discreto, como las de recuento, de uso extendido en Medicina, Bioquímica y Farmacia. En este capítulo se muestran modelos para ensayar hipótesis cuando se trabaja con datos frecuenciales. Por ende, también se aplican a casos en donde, si bien la variable es continua, los datos se han agrupado en clases con sus frecuencias respectivas, o sea, en los hechos se ha “discretizado” a una magnitud de tipo continuo. Lo mismo para el caso de magnitudes cualitativas donde se miden las frecuencias de una clasificación de tipo ordinal o dicotómica. Cualquiera sea el caso, se pueden hacer los supuestos de que las muestras fueron extraídas de cierta población, donde las frecuencias obtenidas representan las proporciones verdaderas de la frecuencia total en la población. Con los modelos teóricos, como la Binomial, Poisson, Gauss, etc. se pueden obtener las frecuencias esperadas. Entonces el problema es comparar las *frecuencias observadas* contra las *esperadas* predichas por la teoría usada. Así se puede comparar una teoría científica contra la realidad en un experimento. Como la coincidencia entre ambas nunca es total, el problema se reduce a poder determinar si las diferencias observadas entre ellas se deben al azar y la teoría es correcta, o de lo contrario se debe rechazar el supuesto realizado. Un supuesto histórico es el de Student (1907) cuando demostró que el error en el conteo celular realizado con un hemocitómetro sigue mejor el modelo Poisson en vez del Normal usado tradicionalmente. Surge así un nuevo concepto, el de *bondad de ajuste*, que se verá mejor en el próximo capítulo junto con los tests de validación respectivos para los casos de ajuste de distribuciones. Aquí se comienza por mostrar la prueba de la Chi-cuadrado aplicado a casos donde los datos se ordenan según un único criterio de clasificación; luego con el G-test se moderniza el uso clásico de la Chi-cuadrado. A continuación se presentan los casos bidimensionales con los llamados *tests de independencia*, en las denominadas *Tablas de Contingencia*. Se profundiza el caso más usual de las tablas de 2 x 2 donde hay dos criterios de clasificación divididos en dos clases cada uno. Los criterios de decisión se basan en la prueba clásica de la Chi-cuadrado, el G-test y el Test exacto de Fisher. En particular, se presenta el concepto del “*odd-ratio*” para efectuar investigaciones clínicas, por su cada vez más difundido uso en Bioquímica y Farmacia, en pruebas de medicamentos y de infecciones por exposición a factores de riesgo, tanto clínicas como epidemiológicas.

15.1 La prueba de la Chi-cuadrado

Por lo general, los resultados de un muestreo casi nunca coinciden con los valores predichos por la teoría o modelo con el cual se está trabajando. Esto es así por las fluctuaciones aleatorias en las mediciones o error de medición. El problema es poder determinar si esas diferencias se deben al azar, o bien no se ajustan al modelo teórico estudiado, en cuyo caso este deberá ser modificado y vuelto a investigar. La *validación estadística* es la herramienta para poder decidir en estos casos. En este punto se verá este problema cuando los datos son de recuentos o de cantidad de casos observados y están agrupados en clases. En los dos casos se está trabajando con *frecuencias*. En la Tabla 15.1 siguiente se muestra el caso general donde ocurren n acontecimientos $A_1, A_2, A_3, \dots, A_n$. Las muestras extraídas de la población en estudio permiten determinar el número de veces en que ocurre cada acontecimiento y sus frecuencias observadas serán: $O_1, O_2, O_3, \dots, O_n$. Por ejemplo, si se está lanzando un dado, habrá 6 acontecimientos posibles, el número de ases observados será O_1 y así sucesivamente. El modelo teórico postulado permite obtener la frecuencia esperada: $E_1, E_2, E_3, \dots, E_n$. En el ejemplo del dado, si se efectuaron 600 lanzamientos, el número esperado de ases será $E_1 = 100$ bajo el postulado de que el dado no está cargado.

Tabla 15.1: Caso general.

Acontecimiento	A_1	A_2	A_3	...	A_n
Frecuencia esperada	E_1	E_2	E_3	...	E_n
Frecuencia observada	O_1	O_2	O_3	...	O_n

La manera clásica de estudiar las diferencias entre frecuencias esperadas y observadas es usar el estadígrafo χ^2 para cuantificar la discrepancia entre ambas. Este se calcula con la ecuación:

$$\chi^2 = [(O_1 - E_1)^2 / E_1] + [(O_2 - E_2)^2 / E_2] + [(O_3 - E_3)^2 / E_3] + \dots + [(O_n - E_n)^2 / E_n]$$

O sea,

$$\chi^2 = \sum_1^n \frac{(O_i - E_i)^2}{E_i}$$

De la fórmula surge que si $\chi^2 = 0$, la frecuencia esperada coincide exactamente con la observada y entonces la teoría predice perfectamente los acontecimientos. Si existe una diferencia cualquiera entre ambas frecuencias, será $\chi^2 > 0$, valor que irá aumentando si no es producto del azar, hasta alcanzar valores significativos que permitan rechazar la hipótesis nula de la igualdad entre la teoría y la realidad.

La distribución muestral del valor χ^2 se aproxima mucho a una del tipo Chi-cuadrado, modelo que se usa para validar hipótesis. Los grados de libertad se calculan para dos casos posibles:

1) *Hipótesis Extrínseca*: es una hipótesis externa a los datos. No se necesita de estos para obtener los parámetros poblacionales en el cálculo de las frecuencias observadas. Por ejemplo, el lanzamiento de una moneda o un dado en teoría de juegos, las leyes de Mendel, etc.

$$v = n - 1$$

2) *Hipótesis Intrínseca*: es una hipótesis interna a los datos. Se necesitan los datos para sacar los parámetros poblacionales. Por ejemplo, si se trabaja bajo el supuesto de normalidad: la media y la varianza de la población se estiman con los datos muestrales -ver el caso 1 del cuadro 4.2- o sea, habrá $r = 2$ grados de libertad perdidos.

$$v = n - r - 1$$

En el caso de las extrínsecas es $v = n - 1$ porque con las primeras $n - 1$ muestras, se puede determinar la restante, entonces como una de ellas se obtiene a partir de las demás, se pierde solo un grado de libertad. Para ilustrar las ideas anteriores se desarrolla el siguiente caso:

Ejemplo: se lanza una moneda al aire 100 veces y se obtienen 58 caras. Ensayar la hipótesis que la moneda está bien.

(H₀) $p = 0,5$ la moneda esta bien.

(H₁) $p \neq 0,5$ la moneda esta "cargada".

De acuerdo con la hipótesis nula, si la moneda está bien se espera un número de 50 caras:

$$\chi^2 = [(O_1 - E_1)^2 / E_1] + [(O_2 - E_2)^2 / E_2] = [(58 - 50)^2 / 50] + [(42 - 50)^2 / 50] = 2,56 \text{ (ns)}$$

Para hallar el valor crítico se toman: $v = n - 1 = 2 - 1 = 1$ grado de libertad y un nivel de confianza del 95%. Entonces de tablas es: $\chi^2_{0,95; 1} = 3,84$ que resulta ser mayor que el valor de comparación $\chi^2 = 2,56$ (valor no significativo), por lo tanto no se puede rechazar la hipótesis nula. Se dice: no hay prueba suficiente como para pensar que la moneda está mal hecha.

15.2 La prueba de G (G-test)

Esta es otra técnica para comprobar la concordancia entre las frecuencias observadas y las esperadas. El llamado *Test de la Razón de Verosimilitudes* (G - test) se ha empezado a usar recientemente y posee propiedades teóricas que lo hacen preferible a la clásica prueba de la Chi-cuadrado. La fórmula usada para el cálculo de la prueba Chi-cuadrado puede escribirse de la manera siguiente:

$$\chi^2 = \sum (O_i - E_i)^2 / E_i = [\sum (O_i)^2 / E_i] - n$$

Donde el *cociente de frecuencias* es el indicador de las diferencias. Cuando la coincidencia sea perfecta, el cociente será E_i y la suma de todos los casos será n , lo cual anulará el valor de comparación: $\chi^2 = n - n = 0$. A medida que aparezcan diferencias entre las frecuencias, cada uno de estos irá aumentando el valor total alejándolo del cero hasta llegar a valores significativos desde

el punto de vista estadístico. Ahora bien, este cociente de las frecuencias $L_i = O_i / E_i$ (o de las probabilidades, si se dividen ambos miembros por n) llamado cociente de *verosimilitudes*, puede emplearse como un indicador de concordancia entre frecuencias, y así se puede construir un test para validar hipótesis. La distribución teórica de este indicador se simplifica con la transformación del valor G , que se calcula con:

$$G = 2 \sum_1^k O_i \ln(O_i / E_i)$$

Donde k es la cantidad de clases o muestras. Este estadígrafo para muestras grandes ($n > 25$) se aproxima bastante a una Chi-cuadrado. Para muestras pequeñas, el valor más exacto se puede lograr a través de la Probabilidad Binomial. El número de grados de libertad, lo mismo que para la prueba de la Chi-cuadrado visto antes, se calcula para dos casos posibles con:

- 1) *Hipótesis Extrínsecas*: $v = k - 1$
- 2) *Hipótesis Intrínsecas*: $v = k - r - 1$

Ejemplo) Se efectúa un cruce genético y el supuesto mendeliano es esperar una proporción de 3:1 de fenotipos en la primera generación filial. Los resultados obtenidos muestran en la prole 178 casos del tipo normal y 22 casos de mutaciones. Para poder testear el grado de concordancia del experimento entre las frecuencias observadas y las esperadas (75% de normales y 25% de mutantes) se puede usar el G-test como sigue:

Fenotipo	Frec. Observada O_i	Frec. Esperada E_i	$O_i \ln(O_i/E_i)$	G
Normales	178	150	30,46	24,8***
Mutantes	22	50	-18,06	
Total	200	200	12,4	

El resultado final $G = 24,8$ debe ser comparado con el valor crítico de tablas, para $v = 1$ grado de libertad. Se obtiene $\chi^2_{0,999; 1} = 10,828$ mucho menor que G . Por lo tanto, se tienen evidencia altamente significativa como para rechazar la hipótesis nula. *No significa que las Leyes de Mendel dejen de cumplirse*, sino que deben buscarse las causas genéticas o experimentales que aclaren lo ocurrido en este caso. Como por ejemplo el concepto de gen ligado, o una falla de método en los cruces, etc. También se puede usar para este mismo caso la prueba de Chi-cuadrado en efecto:

Fenotipo	Frec. Observ. O_i	Frec. Esper. E_i	Diferencias $(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
Normales	178	150	+ 28	784	5,23
Mutantes	22	50	- 28	784	15,68
Total	200	200	0		$\chi^2 = 20,91***$

$\chi^2 = 20,91 \gg \chi^2_{0,999;1} = 10,828$ y se rechaza la hipótesis.

Notar que en ambos casos el valor crítico de tablas es el mismo. Sin embargo, el valor del estadígrafo $G = 24.8 > \chi^2 = 20,91$. Esto significa que G pone más en evidencia la diferencia entre las frecuencias esperadas y observadas, es más eficiente para ello. Se dice que la prueba del G-test es más *sensible* que la prueba de la Chi-cuadrado (χ^2 -test) para detectar diferencias significativas. Viceversa, cuando la hipótesis nula no pueda ser rechazada, es de esperar que el valor G sea de menor valor que el χ^2 para el mismo caso. Se dice que el G-test es más *conservador* (le cuesta más rechazar una hipótesis) que a la prueba de la Chi-cuadrado (ver ejemplo 2).

Se puede generalizar a más casos como el siguiente:

Ejemplo 2) En los experimentos de Mendel con porotos se cosecharon 315 porotos lisos y amarillos, 108 lisos y verdes, 101 rugosos y amarillos y 32 rugosos y verdes. La proporción esperada para este caso es 9 : 3 : 3 : 1. Decidir si los datos observados se ajustan a la teoría mendeliana.

En este caso, el investigador puede considerar que tiene cuatro tipos de porotos diferentes y usar la relación mendeliana para obtener las frecuencias esperadas de cada uno. Las proporciones esperadas son 9/16 para amarillos y lisos, 3/16 para amarillos y rugosos como para verdes y lisos, y 1/16 para rugosos y verdes. Con tales proporciones puede calcular las frecuencias esperadas de las 4 casillas. Por ejemplo, el número esperado de porotos lisos y amarillos será $(9 / 16) N = 312,75$.

Porotos	O _i	E _i	O _i ln (O _i /E _i)	G
Lisos y amarillos	315	312,75	2,258	0,476 (ns)
Rugosos y amar.	101	104,25	-3,199	
Lisos y verdes	108	104,25	3,817	
Rugosos y verdes	32	34,75	-2,638	
Total	556	556	0,238	

Los grados de libertad son: $\nu = (4 - 1) = 3$ y entonces $\chi^2_{0,95;3} = 7,815 \gg \gg G$ por lo que no se puede rechazar la H₀.

Luego si se usa el modelo menos potente de la Chi-cuadrado resulta:

$$\chi^2 = \frac{(315 - 312,75)^2}{312,75} + \frac{(108 - 104,25)^2}{104,25} + \frac{(101 - 104,25)^2}{104,25} + \frac{(32 - 34,75)^2}{34,75} = 0,47$$

Notar que ahora el modelo de la G es más conservador que el de la Chi pues:

$$G = 0,476 > \chi^2 = 0,47$$

Esto muestra que al modelo de la G le cuesta más rechazar una hipótesis nula que al modelo de la Chi-cuadrado, para un mismo caso. Por eso G es más conservador que Chi, y si a esto se le agrega que G es más sensible que Chi para detectar diferencias, se justifica el abandono por parte de los investigadores del clásico modelo de la Chi cuadrado para analizar estos casos.

15.3 Tablas de contingencia

Se tratará ahora el caso de los llamados *tests de independencia*. La noción de independencia estadística o probabilística se vio en el Capítulo 6.2. Allí se mostró que si dos sucesos son independientes, la probabilidad de que ocurran juntos se puede calcular con el producto de sus probabilidades individuales. Por ejemplo, en teoría de juegos la probabilidad de sacar dos secas al lanzar dos monedas al aire es de $1/4$. En Genética la probabilidad de que un grano de maíz sea de color rojo es de $1/2$ y de que sea rugoso $1/3$, por lo tanto la probabilidad de que sea a la vez rojo y rugoso es de $1/6$. Se puede plantear un experimento para verificar la hipótesis anterior. Con el modelo de la Chi-cuadrado se pueden comparar las frecuencias observadas y las esperadas. La proporción mendeliana es de $2 : 2 : 1 : 1$ y la prueba testea dos cosas a la vez: a) las probabilidades de cada una y b) si son independientes. La primera hipótesis es para verificar el modelo de Mendel, pero la segunda prueba si los genes son independientes o si están ligados, es decir, ubicados en el mismo cromosoma.

La Tabla 15.1 muestra un caso de una *tabla de contingencia* del tipo $1 \times n$ identificada como una *tabla de clasificación simple*, y en el mismo punto se muestra cómo se puede plantear la *prueba de la Chi-cuadrado* para resolver un test de hipótesis. Generalizando estas ideas, se llega a un caso más amplio: *Tabla de Contingencia $h \times k$* donde las frecuencias observadas O_{ij} ocupan una tabla de h filas y k columnas. En correspondencia con las mismas se pueden calcular, bajo algún modelo teórico planteado, sus respectivas frecuencias esperadas E_{ij} . Entonces se puede plantear una vez más la prueba de la Chi-cuadrado para este caso como:

$$\chi^2 = \sum \sum (O_{ij} - E_{ij})^2 / E_{ij}$$

Donde O_{ij} : es la frecuencia observada en la columna i ; fila j

E_{ij} : es la frecuencia esperada en la columna i ; fila j

Este estadígrafo se contrasta con el valor crítico $\chi^2_{\alpha; n}$. Los grados de libertad se calculan para dos casos posibles:

1) *Hipótesis Extrínseca*: es una hipótesis externa a los datos, no se necesita de estos para obtener los parámetros poblacionales en el cálculo de las frecuencias esperadas:

$$v = (h - 1) (k - 1)$$

2) *Hipótesis Intrínseca*: es una hipótesis interna, se necesitan los datos para sacar los m parámetros poblacionales:

$$v = (h - 1) (k - 1) - m$$

Ejemplo 1) En cierta Universidad, los ingresantes fueron clasificados en tres grupos principales, de acuerdo con la orientación de sus estudios secundarios en: escuelas de comercio, normal e industrial. Se desea saber si la elección de sus carreras universitarias guarda alguna relación con su origen. Para ello, se tomaron datos de tres grandes facultades con orientaciones similares, tales como: económicas, humanista y ciencias exactas. Se formula la hipótesis nula que la elección de

carrera que hace el ingresante es independiente del tipo de su escuela secundaria. O sea, la proporción de estudiantes inscriptos en las carreras es la misma.

Frecuencias Observadas (Oij)

	Comercio	Normal	Indus.	TOTAL
Economía	524	325	200	1049
Sociales	435	256	254	945
Exactas	256	320	430	1006
TOTAL	1215	901	884	3000

Frecuencias Esperadas (Eij)

	Comercio	Normal	Indus.	TOTAL
Economía	424,9	315,1	309,1	1049
Sociales	382,7	283,8	278,5	945
Exactas	407,4	302,1	296,4	1006
TOTAL	1215	901	884	3000

Los datos obtenidos se vuelcan en una tabla de frecuencias observadas, como se muestra en la tabla de la izquierda arriba. Para obtener las frecuencias esperadas se procede como sigue:

Si los dos factores investigados (Tipo de escuela secundaria y Tipo de Universidad elegida) son independientes, entonces el valor esperado en la primera casilla E11 se puede imaginar como la probabilidad esperada P11 multiplicada por el número total de datos N. La probabilidad teórica P11, de que un alumno verifique las condiciones pedidas en tal casilla (estudió en una escuela de comercio en su secundaria y eligió la facultad de Económicas) será:

$P_{11} = E_{11} / N$ Por otra parte, puede imaginarse la tabla anterior como una partición del universo de alumnos. En particular, la celda $i = 1$ y $j = 1$ es la intersección de dos sucesos: **E**: facultad de económicas y **C**: escuela de comercio. Luego:

$P(E \cap C) = P(E) \cdot P(C)$ Si **E** y **C** son independientes, la probabilidad del suceso $E \cap C$ es igual al producto de sus probabilidades. Y a su vez, cada una es:

$$P(E) = N_E / N = 1049 / 3000 = 0,35 \quad \text{y} \quad P(C) = N_C / N = 1215 / 3000 = 0,405$$

Reemplazando en la primera ecuación es: $P_{11} = P(E \cap C) = E_{11} / N$ entonces,

$$E_{11} = N \cdot P_{11} = N P(E \cap C) = N \cdot P(E) \cdot P(C) = N \cdot (N_E / N) (N_C / N) \quad \text{o sea,}$$

$$E_{11} = (N_E \cdot N_C) / N = (1049 \cdot 1215) / 3000 = 424,9$$

La frecuencia esperada de la primera casilla es igual al producto de los totales marginales correspondientes, dividido el tamaño total de la muestra tomada. Análogamente para las demás casillas se puede concluir que: la frecuencia esperada de una casilla cualquiera Eij se puede obtener con la relación:

$$E_{ij} = N_i \cdot N_j / N$$

Aplicando esta relación a todas las casillas de las frecuencias observadas, se obtienen las frecuencias esperadas Eij que se muestran en la tabla de arriba, en su parte derecha. Una vez que se tienen todos los valores se puede calcular el estadígrafo:

$$\chi^2 = \sum \sum (O_{ij} - E_{ij})^2 / E_{ij} = [(524 - 424,9)/424,9] + [(325 - 315,1)/315,1] + \dots = 191,47 \quad ***$$

valor mucho mayor que el de tablas $\chi^2_{\alpha;v}$ con $v = (3 - 1) (3 - 1) = 4$. O sea, $\chi^2_{0,999;4} = 18,467$

Por lo tanto, se rechaza la hipótesis nula y se concluye que el tipo de escuela secundaria realizada por el ingresante influye en su elección de carrera universitaria, como era de esperar. Se deja al lector el cálculo del estadígrafo G para este caso, donde la evidencia es tan fuerte que no se necesita mayor sensibilidad para rechazar H_0 .

Esta prueba requiere que los datos observados en cada casilla no sean pequeños. Ninguna celda deberá tener frecuencia esperada nula y el 20% de las mismas no debe ser inferior a 20. Si los datos originales no verifican estas condiciones, la solución es agrupar categorías similares para lograr que se verifiquen estas condiciones.

15.4 Tablas de contingencias 2 x 2

Es un caso particular del visto en el punto anterior, pero es el más popular entre los investigadores y por ese motivo se lo tratará con más detalle. Los modelos estadísticos propuestos para analizar este tipo de tablas son tres: el test de la Chi-cuadrado, el G-test y el test exacto de Fisher. Es conveniente, antes de seguir adelante, identificar a tres tipos de casos que se pueden encontrar en la práctica, de acuerdo a como el investigador desea planificar su experimento:

Modelo I: se fija de antemano únicamente el tamaño muestral.

Modelo II: se fija de antemano el tamaño de los totales marginales en uno de los factores.

Modelo III: se fijan de antemano todos los totales marginales de la tabla.

El modelo de la Chi-cuadrado no fue diseñado específicamente para ninguno de los tres modelos y los investigadores lo suelen usar sin hacer distinción entre ellos. Sin embargo, estudios recientes aconsejan lo siguiente:

- Usar el G-test de independencia para los modelos *I* y *II*.
- Usar el test exacto de Fisher para el modelo *III*.

Si un bioquímico elige al azar una cantidad de 100 pacientes para estudiar la incidencia del sexo en la preferencia del paciente por optar por uno de los dos turnos de extracción de la muestra. Tiene dos factores a investigar: Sexo y Turno. No controla la cantidad de hombres y mujeres que concurren a su laboratorio, ni tampoco si se reparten en el turno de la mañana o de la tarde. Por lo tanto este caso será tratado como Modelo *I*. En cambio, si fija de antemano que el 50% de su muestra serán hombres, entonces cambia la situación y ahora tiene un Modelo *II*. Si además elige que el 70% de los seleccionados sea del turno mañana y el resto de la tarde, ya ha definido los cuatro totales marginales y por tanto debe tratar a este caso como Modelo *III*.

Ejemplo 1) En los experimentos de Mendel con porotos detallado más arriba, se puede realizar otro tipo de análisis. Ahora se desea verificar los dos factores: Color y Textura, son independientes (esto es, no pueden estar ligados). Entonces, se plantea este mismo problema como si fuera una tabla de 2x2 (en lugar de 1x4 como antes). En este caso, el investigador no puede fijar ninguno de los totales marginales, a lo sumo selecciona la cantidad N de porotos para la muestra. Las proporciones esperadas en la Tabla de 1 x 4 eran: 9/16 para amarillos y lisos, 3/16 para ama-

rillos y rugosos como para verdes y lisos, y 1/16 para rugosos y verdes. Pero ahora para testear independencia serán las frecuencias esperadas de las 4 casillas. Por ejemplo, el número esperado de porotos lisos y amarillos será $(416 \times 423 / 556) = 316,5$ en lugar de 312,75 cuando se usa la teoría mendeliana. Planteado de esta manera el problema queda:

Frecuencias observadas: Oij

Frecuencias esperadas: Eij

FACTOR B: tersura

FACTOR		Lisos	Rugosos	Total			Lisos	Rugosos	Total
A: color	Amarillos	315	101	416	Amarillos	316,5	99,5	416	
	Verdes	108	32	140	Verdes	106,5	33,5	140	
	Total	423	133	556	Total	423	133	556	

Se trata de un **Modelo I**. Luego se calcula el estadígrafo con:

$$\chi^2 = \frac{(315 - 316,5)^2}{316,5} + \frac{(108 - 106,5)^2}{106,5} + \frac{(101 - 99,5)^2}{99,5} + \frac{(32 - 33,5)^2}{33,5} = 0,118$$

Los grados de libertad son: $\nu = (2 - 1)(2 - 1) = 1$. De tablas se encuentran valores críticos mayores que el estadígrafo calculado y por lo tanto no se puede rechazar la hipótesis nula.

Aplicando el G-test al mismo caso se obtiene:

$$G = 2 \sum O_i \cdot \ln(O_i / E_i)$$

$$G = 2 [315 \ln(315/316,5) + 108 \ln(108/106,5) + 101 \ln(101/99,5) + 32 \ln(32/33,5)]$$

$$G = 2 (-1,497 + 1,511 + 1,511 - 1,466) = 0,059 \text{ (no significativo)}. \text{ Acá } G < \chi^2 \text{ (robustez).}$$

Ejemplo 2) Se eligen 400 pacientes al azar, de entre aquellos que sufren cierta enfermedad. A la mitad se le aplica una nueva droga curativa, y a la otra mitad un placebo. Decidir si la nueva droga es efectiva con los datos encontrados que se muestran en la tabla siguiente.

Frecuencias observadas: Oij

Frecuencias esperadas: Eij

FACTOR B: Curación

FACTOR		SI	NO	Total			SI	NO	Total
A: droga	SI	185	15	200	SI	140	60	200	
	NO(placebo)	95	105	200	NO	140	60	200	
	Total	280	120	400	Total	280	120	400	

Se trata de un Modelo II. Por razones didácticas, se lo resuelve primero con el modelo de la Chi-cuadrado, sin y con la corrección de Yates por continuidad -se trata de datos discretos y la distribución teórica es continua-; esta corrección consiste en restarle la mitad de la menor unidad empleada en valor absoluto, en este caso 0,5. Luego, se aplica el modelo aconsejado el G-test y

por ultimo el G-test con la corrección de Williams por continuidad. En total se resuelve el problema de cuatro formas diferentes, para poder ver las ventajas y desventajas de cada uno:

Caso 1) Empleando el modelo de la Chi-cuadrado es:

$$\chi^2 = \frac{(185 - 140)^2}{140} + \frac{(95 - 140)^2}{140} + \frac{(15 - 60)^2}{60} + \frac{(105 - 60)^2}{60} = 96,43 \text{ *** } \gg \chi^2_{0,999; 1} = 10,83$$

Caso 2) Empleando el modelo de la Chi-cuadrado con la corrección de Yates es:

$$\chi^2 = \frac{(|185 - 140| - 0,5)^2}{140} + \frac{(|95 - 140| - 0,5)^2}{140} + \frac{(|15 - 60| - 0,5)^2}{60} + \frac{(|105 - 60| - 0,5)^2}{60}$$

$$\chi^2 = \frac{(44,5)^2}{140} + \frac{(44,5)^2}{140} + \frac{(44,5)^2}{60} + \frac{(44,5)^2}{60} = 94,3 \text{ ***}$$

Hay evidencia altamente significativa como para rechazar la hipótesis nula. Se tiene prueba científica de que la droga es buena. Tanto con el modelo clásico de la Chi-cuadrado como con este mismo modelo aplicándole la corrección por continuidad (que la hace menos sensible).

Caso 3) Empleando la prueba de G resulta:

$$G = 2 \sum O_i \ln (O_i / E_i) = 2 [185 \ln (185/140) + 95 \ln (95/140) + 15 \ln (15/60) + 105 \ln (105/60)]$$

$$G = 2 (51,56 - 36,84 - 20,79 - 58,76) = 105,38 \text{ ***}$$

Una vez más, se ve la mayor sensibilidad de este modelo frente al Chi-cuadrado $G > \chi^2$

Caso 4) Empleando la prueba de G con la corrección de Williams por continuidad, la cual es preferible a la de Yates, resulta:

$$G_{\text{corr.}} = G / Q \quad \text{con} \quad Q = 1 + (A \cdot B / 6n)$$

Para ello se calculan los siguientes factores

$$A = [n (\text{total marginal de fila 1})^{-1} + n (\text{total marginal de fila 2})^{-1} - 1]$$

$$B = [n (\text{total marginal de columna 1})^{-1} + n (\text{total marginal de columna 2})^{-1} - 1]$$

O sea,

$$A = [(400/200) + (400/200) - 1] = 3$$

$$B = [(400/280) + (400/120) - 1] = 3,76$$

$$Q = 1 + 3 (3,76) / 2400 = 1 + 0,0047 = 1,0047$$

$$G_{\text{corr.}} = G / Q = 105,38 / 1,0047 = 104,89 \text{ ***}$$

El valor ajustado es ligeramente menor que G, pero igualmente significativo. Conviene destacar que esta corrección se impone cuando el valor de G esta cerca del valor crítico de tablas, para despejar cualquier duda. Pero, en el presente ejemplo la diferencia es tan grande que no se hace necesario realizar la corrección. Aquí se hizo por razones didácticas.

Otra vez se puede notar que el G-test pone más en evidencia las diferencias significativas que el test del la Chi-cuadrado. Esto se debe a la mayor sensibilidad del G-test, que se convierte en la opción recomendable, para tratar las tablas de contingencia 2 x 2 diseñada como un Modelo II experimental. Con la aclaración de que se debe usar la corrección de Williams cuando el valor del estadígrafo esté cercano al límite de la zona de rechazo al 95% de confianza.

Los valores críticos usados en las tablas de 2 x 2 son:

$$\chi^2_{1,095} = 3,841 \quad ; \quad \chi^2_{1,099} = 6,635 \quad \text{y} \quad \chi^2_{1,0999} = 10,83$$

En resumen: en una Tabla 2 x 2, diseñada como Modelo I o Modelo II, conviene usar la prueba de la G (G-test). Cuando el valor obtenido experimentalmente esté cercano a 3,841 se debe usar la corrección de Williams para salir de dudas. Cuando supere 6,635 no hace falta tal corrección, salvo para la sutileza de determinar la probabilidad asociada. Y cuando sea mayor de 11 ya no hay dudas de la fuerte evidencia obtenida.

Ejemplo 3) Un cardiólogo decide investigar si los problemas de hipertensión de sus pacientes son independientes de los antecedentes familiares de los mismos. Para ello, escoge 12 historias clínicas que cuentan con esos datos; del total son: 5 con antecedentes y 7 sin, mientras que del total solo 5 no son hipertensos. Clasifica sus datos en una tabla y aplica el test exacto de Fisher para determinar si ambos factores son independientes.

Casos observados

En términos generales

Factor B: Hipertensos

Factor A :		SI	NO	Total
Con	SI	2	5	7
antecedentes	NO	3	2	5
	Total	5	7	12

	SI	NO	Total
SI	a	b	a+b
NO	c	d	c+d
Total	a+c	b+d	N

Es un **Modelo III**. La probabilidad exacta se calcula con la probabilidad multinomial. Escribiendo la tabla de datos observados, en términos generales (la tabla de la derecha) se puede expresar:

$$P = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{n! a! b! c! d!}$$

Tomando logaritmos para facilitar el calculo queda:

$$\ln P = \ln (a+b)! + \ln (c+d)! + \ln (a+c)! + \ln (b+d)! - \ln n! - \ln a! - \ln b! - \ln c! - \ln d!$$

Se puede usar cualquiera de ambas relaciones para resolver el cálculo. Para este caso no se usan los logaritmos porque los valores son chicos. Para usar el Modelo de Fisher se procede así:

Paso 1) Se calcula la probabilidad multinomial de que a lo sumo haya **a** casos en la primera celda.

Esto es, calcular $P(r \leq a) = P(r = 0) + P(r = 1) + \dots + P(r = a)$. Cada uno de estos términos es una Multinomial que se calcula con la fórmula de más arriba.

Paso 2) Se compara esta probabilidad acumulada $P(r \leq a)$ con el nivel de significación $\alpha = 0,05$, para aceptar o rechazar la hipótesis nula como es usual.

Para facilitar la visualización del problema se presentan los casos para obtener la acumulada $r \leq 2$

$r = 2$

2	5
3	2

$r = 1$

1	6
4	1

$r = 0$

0	7
5	0

$$P = \frac{7! 5! 7! 5!}{12! 2! 5! 3! 2!} = 0,26515 = P(r = 2) \quad : \text{ celda de la izquierda.}$$

$$P = \frac{7! 5! 7! 5!}{12! 1! 6! 4! 1!} = 0,04399 = P(r = 1) \quad : \text{ celda del medio.}$$

$$P = \frac{7! 5! 7! 5!}{12! 0! 7! 5! 0!} = 0,00126 = P(r = 0) \quad : \text{ celda de la derecha.}$$

$$P(r \leq 2) = P(r = 0) + P(r = 1) + P(r = 2) = 0,3104 \gg \alpha = 0,05$$

Por lo tanto, no se puede rechazar la hipótesis nula. Los antecedentes familiares son independientes de la hipertensión. Como los datos son muy pocos, el investigador puede aumentar el número de casos para tener un panorama más acorde a sus sospechas.

15.5 Aplicaciones en Farmacia y Bioquímica

En la últimas décadas la aplicación de estadística en Medicina se ha incrementado vertiginosamente. En un comienzo, sus métodos para el diseño y análisis de experimentos comenzaron a usarse en el área de los estudios sobre enfermedades del corazón y cáncer. La incorporación de nuevos índices clínicos -como los vistos en el capítulo 4- empezó a mediados de la década del 70. Hay muchas razones para que haya ocurrido esto; entre ellas se pueden citar a dos: cada vez es más aceptado que la mayoría de las enfermedades son causadas por múltiples factores y el desarrollo de cada enfermedad involucra muchos cofactores diferentes. El análisis multivariado, especial para estos casos, que requiere un nivel matemático no acorde con el general en carreras relacionadas con Medicina, se ha vuelto accesible gracias a la aparición de software de computadora que soluciona los problemas a los que no son especialistas. Además, el hardware necesario para poder usarlo ya está disponible en el mercado, a precios de ensueño en el pasado.

Más recientemente, los estudios epidemiológicos se han incorporado a esta corriente innovadora, con la aplicación de técnicas más sofisticadas. En la problemática del contagio cabe mencionar a dos áreas en particular: los estudios sobre infecciones de tipo hospitalaria y sobre la inmunodeficiencia humana con el virus tipo 1 (HIV-1). Muchas investigaciones se están realizando para estudiar la historia natural de la HIV-1 y los factores asociados con el desarrollo de la

adquisición del síndrome de esta (SIDA). Por su parte, ya es aceptado universalmente que las infecciones hospitalarias no son causadas por un único factor de riesgo, sino que se originan en una compleja interacción entre factores relacionados con el infectado, el agente infeccioso y el medio ambiente que lo rodea. En este capítulo se presentan las medidas de asociación más usuales en este tipo de estudios: el “*riesgo relativo*” (*relative risk*) y la “*eficacia*” (odds ratio), en su forma más elemental, aplicadas a diseños con tablas de 2x2, dejando la regresión logística para estudios más avanzados.

Una enfermedad puede ser estudiada desde dos perspectivas. Una de tipo “observacional” donde el investigador no interviene ni manipula los factores involucrados y deja que la enfermedad siga su curso natural. La otra de tipo “experimental” donde el investigador puede controlar de alguna manera factores que pueden ser de importancia en el proceso de la enfermedad, y se usan para examinar relaciones “causa-efecto” entre los factores involucrados. Los estudios poblacionales por cohortes y los estudios caso-control son del tipo observacional. Los estudios en pruebas clínicas son de tipo experimental, donde un grupo está expuesto a un factor de riesgo o a un tratamiento, y el otro grupo no lo está. El investigador, usando una tabla de números al azar, reparte a los pacientes seleccionados para el experimento, en forma aleatoria, dentro de cada grupo. Un grupo actuará como placebo o control mientras que el otro recibirá el tratamiento o una protección. Si se homogenizan los grupos antes de comenzar el tratamiento, cuidando factores tales como edad, sexo, etc., se supone que se evita el sesgo que produciría alguno de esos factores, si no es tomado en cuenta. Si los grupos son lo suficientemente grandes, se espera que dentro de los resultados el grupo protegido sea más grande que el otro, y aparecerá una significación estadística que le permitirá al investigador tener una prueba científica de la relación causa-efecto entre los factores estudiados. Conviene destacar que esta prueba es de tipo estadístico, lo cual no siempre tiene significado en lo clínico, *que exista significación estadística no siempre implica que haya una significación clínica*.

Hay que ser cuidadoso con la homogeneización de la muestra. Si es “demasiado” homogénea, se tornará difícil extender las conclusiones del estudio a la población en general. Por otra parte, si no lo es, pueden aparecer factores importantes en el desarrollo de la enfermedad, que introduzcan variabilidades engañosas. Por ejemplo, si se hace una investigación sobre infecciones post-operatorias en enfermedades coronarias, estudiando el efecto de determinado antibiótico para prevenirlas, se hace poco creíble extender las conclusiones de este estudio a todas las infecciones de tipo hospitalario. Por otra parte, si la muestra se toma eligiendo a pacientes provenientes de cualquier cirugía, sin importar edad o sexo, ocurre otro tanto. Por ejemplo, incluir muchas cesáreas en el estudio, sin hacer discriminación de las mismas en el total, implicaría un sesgo al sexo femenino y a cierta escala de edades. Esto invalidaría extender las conclusiones del estudio a ambos sexos. El otro punto a tener muy en cuenta, es realizar este tipo de experimentos al estilo “doble-ciego”. Esto es, ni los que intervienen en la investigación, ni los pacientes, deben saber quiénes reciben las drogas y quiénes los placebos. De esta forma se evita cualquier tipo de subjetividad o favoritismo. La diferencia entre este tipo experimental de trabajo (“clinical trials”) y los estudios poblacionales del tipo caso-control, es que la enfermedad ya se ha presentado en este último, a diferencia del primero. Y el investigador busca en la historia pasada del paciente las causas que originaron la enfermedad, los factores que pueden haber influenciado. Su grupo control será otro grupo de individuos que no la padezcan y que tengan características similares a los enfermos, y buscando en la historia pasada la existencia de los mismos factores en estos últimos, podrá sacar conclusiones.

15.6 Odds-ratio y riesgo relativo

En Medicina el *riesgo* puede ser definido como la probabilidad de que ocurra un determinado suceso que implique un peligro para la salud de un paciente. Tal como el riesgo de contraer una septicemia luego de ser operado.

El *factor de riesgo* es una magnitud cualitativa, de tipo dicotómica, donde sus dos resultados posibles indican si el factor está presente o no. Usualmente esto se denota con: SÍ-NO. Son ejemplos de factor de riesgo características tales como: edad, sexo, tratamiento profiláctico antes de una cirugía, inmunización, medicación preventiva, etc. Cuando un grupo de pacientes recibe una droga que se está probando, será el caso SÍ del factor de riesgo, y el otro grupo usado como control o placebo será el caso NO. En cambio, cuando el primer grupo es cateterizado en su cirugía se piensa que está expuesto a una infección (SÍ) y otro grupo donde no fue necesaria (NO) será el de control.

La comparación del riesgo es uno de los objetivos principales de los estudios epidemiológicos, y su cuantificación implica la medición experimental de sus dos índices asociados: *Riesgo Relativo* (RR) y “*Odds-Ratio*” (OR).

El diseño experimental para medir ambos índices es el planteo de una Tabla de Continencia del tipo 2x2. Como el investigador es quien decide el tamaño muestral y cuántos serán expuestos al factor de riesgo, es un caso de *Modelo I*. Suponiendo que toda la población estudiada es homogénea en todas sus características excepto en una: una parte de esta se expone a un factor de riesgo, que se cree es de importancia en causar una determinada enfermedad, y que ninguno de los cuatro casos posibles tiene una ocurrencia nula; el diseño experimental usado generalmente se muestra esquemáticamente en la Tabla 15.2 siguiente:

Tabla 15.2: Frecuencias observadas entre la exposición a un factor de riesgo y la contracción de una enfermedad cualquiera.

Factor de Riesgo	Enfermedad		Total
	SÍ	NO	
O	a	b	$n_1 = a+b$
Característica	c	d	$n_2 = c+d$
Grupal	$n_3 = a+c$	$n_4 = b+d$	n

$$RR = (a \cdot n_2) / (c \cdot n_1)$$

$$OR = (a \cdot d) / (b \cdot c)$$

Riesgo relativo (RR) se define como el cociente entre la probabilidad de contraer la enfermedad de la población expuesta y la probabilidad de contraerla de los no expuestos.

$$RR = P(E) / P(\text{no}E) = (a/n_1) / (c/n_2) = (a \cdot n_2) / (c \cdot n_1) \text{ que es la expresión de más arriba.}$$

Cuando no hay pacientes de la población expuesta que contraigan la enfermedad ($a = 0$) el RR se anula. En cambio, si ningún paciente de los no expuestos contrajo la enfermedad ($c = 0$) entonces el RR se vuelve infinito. Si los factores son *independientes* el $RR = 1$. El RR solo puede ser estimado con estudios prospectivos. La interpretación se puede ver mejor con un ejemplo: se sabe que entre los que sufrieron infarto de miocardio el nivel de colesterol medio es de 300

mg/dl. En la población general el nivel es de 200 mg/dl. Se conoce la distribución de colesterol en ambas poblaciones, la de los infartados y la de la población humana usadas como referencia. Desde el punto de vista médico se toma como valor referencial 250 mg/dl, se considera a un paciente con “colesterol alto” cuando supera dicho valor. Luego se cuentan los casos de encontrados con colesterol alto en los infartados y entre los no infartados. Sea por caso, una relación de $120/180 = 2/3$ entre los infartados y de $1000/9000 = 1/9$ entre los no infartados, entonces se calcula el $RR = (2/3) / (1/9) = 6$. Eso se interpreta así: “Si un paciente tiene un colesterol mayor que 250, su chance de infarto es 6 veces mayor que si tuviese un valor de 200 o menos”. Así se construyen las simplificaciones en los diarios y revistas de divulgación.

Si se trata de un estudio del tipo caso-control, generalmente el RR no puede calcularse y se necesita del OR como una medida de asociación entre ambos factores analizados. Entonces, el *Odds Ratio* (OR) se define como el cociente entre dos “odds” posibles. Un “*odd*” es la relación entre la cantidad de “enfermos” y los “no enfermos” de una población dada. Como hay dos poblaciones, la expuesta y la no expuesta al factor de riesgo, hay dos tipos de “*odd*” posibles y la tasa entre ambos es el valor de OR. Sea:

$P_1 = a / n_1$: La probabilidad de contraer la enfermedad de la población expuesta.

$(1 - P_1) = b / n_1$: la probabilidad de no contraer la enfermedad de la población expuesta

El “*odd*” de la población expuesta es $O_1 = P_1 / (1 - P_1) = a / b$. A su vez,

$P_2 = c / n_2$: la probabilidad de contraer la enfermedad de la población no expuesta.

$(1 - P_2) = d / n_2$: la probabilidad de no contraer la enfermedad de la población no expuesta.

El “*odd*” de la población no expuesta es $O_2 = P_2 / (1 - P_2) = c / d$.

Por lo tanto, su cociente es $OR = O_1 / O_2 = (a \cdot d) / (c \cdot b)$ que es la expresión vista en la Tabla 15.2 de más arriba. Cuando el $OR = 1$ significa que ambos factores estudiados son independientes entre sí. Si a o d se anulan entonces el $OR = 0$; en cambio, si se anulan c o b , se hace infinito. Para el caso usado como ejemplo para el RR, se puede calcular:

$$OR = (120 / 1000) / (60 / 8000) = 16$$

Y se interpreta así: “cuanto más alto el OR, peor”. Otra forma de ver este índice es como el cociente de dos pares de probabilidades. Una es el RR entre los expuestos y el otro es el RR entre los no expuestos:

$$RR1 = P_e / (1 - P_e) = \frac{a / (a+b)}{b / (a+b)} = a/b \quad \text{y} \quad RR2 = P_{no\ e} / (1 - P_{no\ e}) = \frac{c / (c+d)}{d / (c+d)} = c/d$$

Luego es: $OR = RR1 / RR2$

La diferencia en el uso de uno u otro índice reside en el tipo de investigación que se está realizando. *RR* se emplea en un estudio de *incidencia*, donde la frecuencia de algún resultado se calcula entre dos grupos determinados por la presencia o ausencia de alguna característica. En

cambio, *OR* se usa en un *estudio de control* donde los resultados se obtienen en dos grupos, uno expuesto a un factor de riesgo y el otro usado como placebo o control.

Tanto el *RR* como el *OR* son estadísticos cuantitativos, no se asocian a pruebas de inferencia que terminan en un nivel de significación; en ese sentido no importa la distribución, lo que se estima es el intervalo de confianza de ambos valores al 95% o 99%, obtenidos con las relaciones que se muestran a continuación. Luego se trata de ver si el valor de *RR* = 1, o *OR* = 1 cae dentro de dichos intervalos, en cuyo caso se piensa que hay independencia entre los factores:

Test para OR:

Valor esperado: $E(OR) = \mu_{OR} = OR$

Varianza: $E\{(OR^2)\} - \mu_{OR}^2 = \sigma_{OR}^2$. Para simplificar el cálculo se toman logaritmos:
 $\sigma^2(\ln OR) = a^{-1} + b^{-1} + c^{-1} + d^{-1}$

O sea, $\sigma^* = (a^{-1} + b^{-1} + c^{-1} + d^{-1})^{1/2} = SE(\ln OR)$

Luego, los límites del intervalo de confianza de *OR* (95%) son:

Límite superior: $OR_s = OR e^{1,96 \sigma^*}$ 95% CI (*OR*_i ; *OR*_s)
Límite inferior: $OR_i = OR e^{-1,96 \sigma^*}$

Cuando el valor *OR* = 1 cae fuera del intervalo, hay evidencia como para rechazar la *H*₀. Para el 99% de confianza el exponente es $\pm 3,09$ (en lugar de $\pm 1,96$ para el 95%).

Test para RR:

Valor esperado: $E(RR) = \mu_{RR} = RR$

Varianza: $E\{(RR^2)\} - \mu_{RR}^2 = \sigma_{RR}^2$. Para simplificar el cálculo se toman logaritmos:
 $\sigma^2(\ln RR) = a^{-1} - (a + b)^{-1} + c^{-1} - (c + d)^{-1}$

O sea, $\sigma'' = [a^{-1} - (a + b)^{-1} + c^{-1} - (c + d)^{-1}]^{1/2} = SE(\ln RR)$

Luego, los límites del intervalo de confianza de *OR* (95%) son:

Límite superior: $RR_s = RR e^{1,96 \sigma''}$ 95% CI (*RR*_i ; *RR*_s)
Límite inferior: $RR_i = RR e^{-1,96 \sigma''}$

Cuando el valor *RR* = 1 cae fuera del intervalo, hay evidencia como para rechazar la *H*₀. Para el 99% de confianza el exponente es $\pm 3,09$ (en lugar de $\pm 1,96$ para el 95%).

Para ver mejor estas ideas se desarrollan los ejemplos siguientes:

Ejemplo 1) En un hospital pediátrico se tomaron datos de contagio bacteriano post-operatorio, con relación al uso de una cateterización de la arteria umbilical en 340 cesáreas realizadas. Los datos tomados de la obra de Muñoz y Townsend (pág. 963) son ficticios.

Cateterización:	Bacteriemia		Total	OR = $\frac{130 \cdot 80}{30 \cdot 100}$
	SI	NO		
SÍ	130	100	230	OR = 3,47
NO	30	80	110	
Total	160	180	340	

$$\sigma^2 (\ln OR) = a^{-1} + b^{-1} + c^{-1} + d^{-1} = (130)^{-1} + (100)^{-1} + (30)^{-1} + (80)^{-1} = 0,06353$$

$$\sigma^* = (a^{-1} + b^{-1} + c^{-1} + d^{-1})^{1/2} = (0,06353)^{1/2} = 0,252$$

Luego, los límites del intervalo de confianza de OR (95%) son:

$$\text{Límite superior: } OR_s = OR \cdot e^{1,96 \sigma^*} = 3,47 \cdot e^{1,96 (0,252)} = 5,686$$

$$\text{Límite inferior: } OR_i = OR \cdot e^{-1,96 \sigma^*} = 3,47 \cdot e^{-1,96 (0,252)} = 2,116$$

Como el valor: OR = 1 cae fuera de este intervalo 95% CI (2,116 ; 5,686), se tiene prueba estadística de una relación entre la infección y el haber tenido una cateterización operatoria.

Sin embargo, este es un caso de una Tabla de contingencia 2x2, Modelo II. Por lo tanto, conviene más aplicar la prueba de G a los mismos datos. Resulta: $G = 26,33^{***}$ y corrigiendo con Williams es un poco menor $G_{corr} = 26,192^{***}$. Pero ahora se tiene una prueba científica de la falta de independencia entre ambos factores analizados: hay asociación estadística entre el factor analizado y la enfermedad.

Ejemplo 2) En una investigación poblacional se obtuvieron 38.152 datos que fueron clasificados según si sufrieron un infarto o no, y dentro de estos fueron desagrupados de acuerdo a si tenían un valor de glucosa alto (mayor a 1,1). Se trata de establecer si un valor alto de glucosa tiene incidencia en sufrir infartos. Los datos mostrados son ficticios. Los resultados se vuelcan en la tabla siguiente:

Glucosa Alta	Infartados		Total	RR = $\frac{161 \cdot 34768}{2636 \cdot 748}$
	SÍ	NO		
SÍ	161	2475	2636	RR = 2,839
NO	748	34020	34768	
Total	909	36495	37404	

$$\sigma^2 (\ln RR) = a^{-1} - (a + b)^{-1} + c^{-1} - (c + d)^{-1} = (161)^{-1} - (748)^{-1} + (748)^{-1} - (36495)^{-1} = 0,00714$$

$$\text{O sea, } \sigma'' = [a^{-1} - (a + b)^{-1} + c^{-1} - (c + d)^{-1}]^{1/2} = 0,0845 = SE(\ln RR)$$

Luego, los límites del intervalo de confianza de OR (95%) son:

$$\text{Límite superior: } RR_s = RR \cdot e^{1,96 \sigma''} = 2,839 \cdot e^{1,96 (0,0845)} = 3,350$$

$$\text{Límite inferior: } RR_i = RR \cdot e^{-1,96 \sigma''} = 2,839 \cdot e^{-1,96 (0,0845)} = 2,406$$

Como RR = 1 cae fuera del intervalo (2,406 ; 3,350), se rechaza la hipótesis nula de independencia entre ambos factores. Se concluye que las personas con glucosa alta tienen un riesgo

de infarto 284% mayor que los que tienen glucosa normal. Notar que si se hubiese hecho un G-test, el resultado sería $G_{corr} = 118,1^{***}$ y se tiene una evidencia muy fuerte para rechazar la hipótesis nula. Con el test de la Chi cuadrado ocurriría otro tanto.

15.7 Modelos en Epidemiología

Existen dos maneras de tomar las muestras en Epidemiología una es la de hacer un muestreo basándose en la enfermedad y la otra en la exposición. En la primera el investigador selecciona al azar una cierta cantidad de enfermos y otra de no enfermos, mientras que en la segunda selecciona una cierta cantidad de individuos expuestos al factor de riesgo (o de protección) y otra de no expuestos (o no protegidos). Conviene definir dos modelos genéricos para los estudios epidemiológicos:

Modelo D: Este diseño del estudio es específico para estudiar la enfermedad porque se hace un muestreo sobre la base del factor de riesgo. El investigador decide los valores de TE (total de individuos expuestos) y TnE (total de individuos no expuestos).

Modelo E: Este diseño del estudio es específico para estudiar el factor de riesgo (exposición o protección) porque se hace un muestreo sobre la base de la enfermedad. El investigador decide los valores de TD (total de individuos enfermos) y TnD (total de individuos sanos).

Tabla 15.3 Tabla de contingencia 2 x 2 para estudios epidemiológicos

Factor de Riesgo o Protección	Enfermedad		Total
	SÍ	NO	
SÍ	a	b	TE
NO	c	d	TnE
Total	TD	TnD	N

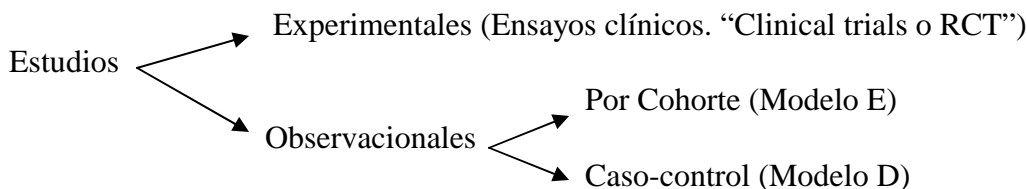
Modelo D: El investigador decide el valor de los totales marginales TE y TnE

Modelo E: El investigador decide el valor de los totales marginales TD y TnD

Por lo tanto, los valores de las cuatro celdas de la Tabla (a, b, c y d) junto con los otros dos totales marginales que no son elegidos por el investigador, serán “naturales” es decir libres de cualquier manipulación. Y lo mismo ocurrirá con los índices calculados a partir de ellos. Esto no se tiene en cuenta siempre, debido a los usos y costumbres habituales. Una mala costumbre es por ejemplo, primero hacer una selección al azar de los individuos enfermos y no enfermos, y luego eliminar a aquellos que no se ajustan a ciertos requerimientos del investigador. Por ejemplo, en un artículo publicado en *The Lancet*, 340: 1074-78, 1992, sobre el análisis de la eficacia de una vacuna contra la meningitis, los investigadores procedieron de esa manera y no hay forma de demostrar que la muestra tomada sea representativa de la población estudiada. Es evidente que no hubo mala intención por parte de los investigadores, porque explican este procedimiento en detalle, pero no pueden demostrar que la selección efectuada final sea representativa ni aleatoria, como tampoco que su criterio no haya influenciado los resultados. Esto es lo mismo que cuando se elige con la mano a una rata de una jaula que contiene muchas para hacer un experimento, es evidente que no hay preferencia por una u otra, pero no hay manera de demostrar que la extracción sea aleatoria. En realidad se atrapó, a la que se dejó agarrar, ya sea porque no es muy veloz, porque esta gorda, enferma, etc. Este problema es como el de la “casta Susana”, no es suficiente

serlo sino que además, hay que parecerlo. Y para ello la única manera es: 1º) De entre todos los individuos disponibles hay que descartar a los que no verifican los requisitos del investigador. 2º) Y luego hacer la selección al azar de entre los remanentes.

Los estudios epidemiológicos pueden ser clasificados como sigue:



En los estudios experimentales, el investigador controla ciertos factores escogidos que podrían tener alguna importancia en el desarrollo de la enfermedad. Lo más común es que manipule el factor (por ejemplo la aplicación de una vacuna en estudios inmunológicos, etc.). En los estudios observacionales el investigador no manipula los factores, estos ocurren en forma la natural. Los estudios por Cohorte enrolan a todos los individuos que estuvieron expuestos a un agente de exposición (por ejemplo HIV-1), tanto como el seguimiento de la historia natural que tuvo la enfermedad en los individuos (por ejemplo casos de SIDA). Se trata del uso de un diseño del estudio como un Modelo E. En cambio, en los estudios del tipo Caso-control, los individuos son seleccionados en base a la exposición (Modelo D) para ver si resultaron enfermos o no. El objetivo central de todos estos estudios, es la *comparación del riesgo* y las cuantificaciones de este concepto se hacen a través de dos índices el Riesgo Relativo y el Odds Ratio. En los estudios por Cohorte y Ensayos clínicos, el RR puede calcularse directamente. En cambio, el OR solo puede estimarse con estudios prospectivos como en lo Caso-control, donde RR no puede ser calculado.

Ambos índices se usan para ver si hay asociación estadística entre el factor y la enfermedad, cuando $OR = 1$ o bien $RR = 1$, se interpreta como que hay independencia estadística. El problema es que no son índices epidemiológicos propiamente dichos. Por ejemplo:

Caso-control (Modelo E)

Situación A

Factor	Enfermo		
	Si	No	
Si	10	10	20
No	90	90	180
	100	100	200

OR = 1 y OR 95% CI (0,4 ; 2,5)

Situación B

Factor	Enfermo		
	Si	No	
Si	90	90	180
No	10	10	20
	100	100	200

OR = 1 y OR 95% CI (0,4 ; 2,5)

Mirando los valores de OR en ambas situaciones, se deduce que no hay asociación estadística entre el factor y la enfermedad. No hay diferencia entre ambas desde el punto de vista estadístico. Sin embargo el porcentaje de expuestos en uno es claramente diferente del otro, porque no es lo mismo tener 20 que 180 individuos expuestos, sobre 200 estudiados. Si se efectúa una comparación de ambos OR siempre sería no significativa, en cambio si se comparasen la proporción de expuestos, o bien los Odds de Exposición, sí sería significativa. Esto muestra que OR no tiene un significado epidemiológico estricto, sino que solo estudia si el factor analizado tuvo algo que ver

con el desarrollo de la enfermedad. En un Caso-control, la enfermedad ya se ha producido en los Casos y no ha ocurrido en los controles. Por lo tanto la conclusión epidemiológica en estos ejemplos es: “Hay muchos factores que pueden ocasionar la enfermedad (dependencia multifactorial) y de acuerdo a los resultados obtenidos, el factor analizado no esta relacionado con la enfermedad, se trata de un “confounding factor”. Ahora bien, si se analizan las tablas anteriores con el test de la Chi-cuadrado resulta $\chi^2 \approx 0$ en ambos casos. La pregunta es ¿ Por que motivo se usa el OR-test en lugar de Chi para hacer la misma cosa ? La respuesta de los epidemiólogos es que OR tiene un significado clínico y el otro no. Pero lo mismo ocurre si hay asociación, como sigue

Situación C

Factor	Enfermo		
	Si	No	
Si	20	8	28
No	80	92	172
	100	100	200

OR = 2,9 y OR 95% CI (1,2 ; 6,9)

Situación D

Factor	Enfermo		
	Si	No	
Si	92	80	172
No	8	20	28
	100	100	200

OR = 2,9 y OR 95% CI (1,2 ; 6,9)

Mirando los valores de OR en ambos casos, la conclusión es que hay asociación entre el factor y la enfermedad. Lo mismo si se usa la Chi cuadrado, porque $\chi^2 \approx 6 > 3,841$. Ahora no es lo mismo tener 28 que 172 expuestos en 200 muestras, desde el punto de vista epidemiológico. Una rápida mirada a los resultados permite ver que la proporción de expuestos cambia de 14% a 86%. Imaginando que la Situación C ocurre después de cierta vacuna aplicada a la población, y la otra situación era antes de la vacuna. Entonces, esta vacuna debe ser relevante porque produjo el cambio en los expuestos (vacunados / no vacunados). Sin embargo, si se comparan ambas situaciones con el índice OR, no se podrá demostrar la eficacia de la vacuna Lo que sí hace OR es mostrar que la vacuna y la enfermedad están asociados (lo mismo que Chi). La pregunta es ahora: ¿ Cual es el objeto de usar OR en lugar de Chi en estos estudios ? La respuesta se deja a cargo del lector.

Ensayos clínicos y estudios por Cohorte (Modelo D)

Situación E

Factor	Enfermo		
	Si	No	
Si	10	90	100
No	10	90	100
	20	180	200

RR = 1 y OR 95% CI (0,4 ; 2,3)

Situación F

Factor	Enfermo		
	Si	No	
Si	10	90	100
No	10	90	100
	20	180	200

RR = 1 y OR 95% CI (0,9 ; 1,1)

El principal objetivo de estos estudios es detectar al factor de riesgo para la enfermedad. La diferencia entre ambos es que en caso por Cohortes el factor de riesgo ocurre naturalmente, en cambio en los Ensayos clínicos es controlado por el investigador. Cuando un grupo es expuesto (o protegido) a cierto factor, se lo compara con otro no expuesto (placebo) y se mide el RR. En los dos ejemplos anteriores el valor de RR = 1 y esto muestra que el factor no está asociado a la enfermedad. Lo mismo si se hace el test de la Chi-cuadrado ($\chi^2 \approx 0$). Sin embargo, no es lo mismo haber encontrado 20 que 180 enfermos en 200 casos. Si la situación E ocurre antes de cierto

hecho y la F después, este hecho deber ser relevante, aunque estadísticamente no lo sea. El índice RR dice únicamente que no hay asociación estadística. La pregunta es ¿ Si RR no está relacionado con la prevalencia encontrada de la enfermedad, y el objeto de estos estudios es ver que pasa con la enfermedad, para que se los usa en lugar de la Chi ? La respuesta de los epidemiólogos es que la prevalencia se puede ver directamente, en cambio el RR muestra que el factor analizado no es responsable de la enfermedad, sino que debe ser otro factor el causante.

Situación G

Factor	Enfermo		
	Si	No	
Si	20	80	100
No	9	91	100
	29	171	200

Situación H

Factor	Enfermo		
	Si	No	
Si	99	1	100
No	45	55	100
	144	56	200

RR = 2,2 y OR 95% CI (1,1 ; 4,6)

RR = 2,2 y OR 95% CI (1,8 ; 2,7)

En los dos casos anteriores se deduce que hay asociación estadística entre el factor y la enfermedad porque RR = 2,2 y el valor RR = 1 no cae dentro del intervalo de confianza (lo mismo si se usa Chi). Pero no es lo mismo obtener 29 que 144 enfermos en 200 casos. Si la Situación G se presenta antes de cierto hecho, y la H después, entonces este hecho debe ser importante porque la prevalencia subió del 14,5% al 72%. Sin embargo, si se comparan los RR obtenidos en ambos casos, estadísticamente no se podrá detectar diferencia alguna y la influencia del hecho no podrá ser validada. Todo lo que hace RR es decirnos que hay asociación entre el factor y la enfermedad, pero no puede detectar un cambio en la prevalencia, a pesar de que se supone que estos estudios se diseñan para analizar la enfermedad. Nuevamente la pregunta es: ¿ Por cual motivo se usa el índice RR si se puede hacer lo mismo con el Chi-test ? La respuesta se la deja al lector.

La conclusión final con estos ocho ejemplos es que: *ni OR, ni RR tienen un verdadero significado epidemiológico, sino que solo sirven para ver la asociación estadística.*

15.8 Análisis de factores ocultos

Existen muchos factores que pueden desencadenar una enfermedad, por eso se habla de fenómenos multifactoriales. Cuando el problema se lo reduce a solo uno, como se hace en una Tabla de 2 x 2, puede ocurrir que ese único factor analizado no explique adecuadamente el fenómeno analizado, por ejemplo

Ejemplo 3) Otro investigador, al revisar el Ejemplo 1 del punto anterior, sospecha que en realidad hay un factor oculto que no se tuvo en cuenta en el caso. Piensa que el tiempo de gestación debió haber sido considerado porque no es lo mismo un nacimiento prematuro que uno normal. Para verificar sus sospechas reorganiza los datos como se muestra a continuación, donde arma dos tablas de contingencia, modificando los factores. Su conclusión ante estos datos es que la asociación entre bacteriemia y cateterización es *aparente*. En realidad, hay una muy fuerte relación entre bacteriemia y tiempo de gestación (OR = 75), y a su vez el tiempo de gestación está muy relacionado (OR = 5) con las cateterizaciones. Su planteo es algo como:

Tiempo de gestación → Cateterización → Bacteriemia
 En lugar del primer ejemplo donde: Cateterización → Bacteriemia

Prematuro:	Bacteriemia		Total	OR = $\frac{150 \cdot 150}{10 \cdot 30}$
	SÍ	NO		
SÍ	150	30	180	
NO	10	150	160	
Total	160	180	340	OR = 75,0

Prematuro:	Cateterización		Total	OR = $\frac{150 \cdot 160}{180 \cdot 80}$
	SÍ	NO		
SÍ	150	30	180	
NO	80	80	160	
Total	230	110	340	OR = 5,0

Aplicando el G-test a los dos casos anteriores resulta para el primero, entre Bacteriemia y Tiempo de gestación, un valor $G = 232,11^{***}$. Mientras que para el segundo, entre Cateterización y Tiempo de gestación, es $G = 43,821^{***}$. En correspondencia con los valores de OR respectivos, pero con demasiada evidencia para mostrar la relación.

Ejemplo 4) Un tercer investigador sospecha que no hay relación entre Cateterización y Bacteriemia, sino que es aparente. Hay una especie de enmascaramiento entre ambas por la influencia del factor Tiempo de gestación. Entonces, para remover tal influencia reorganiza los datos de la manera siguiente:

Cateterización en Prematuros:	Bacteriemia		Total	OR = $\frac{127 \cdot 7}{23 \cdot 23}$
	SÍ	NO		
SÍ	127	23	150	
NO	23	7	30	
Total	150	30	180	OR = 1,68

Cateterización en no premat.:	Bacteriemia		Total	OR = $\frac{5 \cdot 75}{5 \cdot 75}$
	SÍ	NO		
SÍ	5	75	80	
NO	5	75	80	
Total	10	150	160	OR = 1

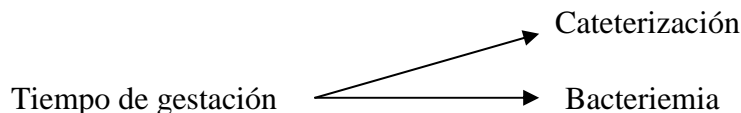
Al examinar los nuevos valores de OR cuando se remueve el factor Tiempo de gestación, se ve una caída brusca de los mismos hasta un valor $OR = 1$ que muestra la independencia de los factores en el caso de los nacimientos en tiempo normal. En efecto, si se piensa en una tabla de independencia, los valores de las frecuencias esperadas de OR se calculan con:

Expuesta:	Enfermedad y/o Contagio		Total
	SÍ	NO	
SÍ	$(a+b)(a+c)/N$	$(a+c)(c+d)/N$	$a + c$
NO	$(a+b)(b+d)/N$	$(b+d)(c+d)/N$	$b + d$
Total	$a + b$	$c + d$	N

Luego, cuando los factores son independientes resulta:

$$\mu (OR) = [(a+b)(a+c) \cdot (b+d)(c+d)] / [(a+b)(b+d) \cdot (a+c)(c+d)] = 1$$

Si se examinan las dos tablas de más arriba con el G-test, resulta para la de arriba $G_{corr.} = 1,035$ y para la de abajo $G = 0$, ambos valores no significativos. Por lo que no se puede demostrar una asociación entre el factor Cateterización y Bacteriemia. Su planteo es de esta forma:



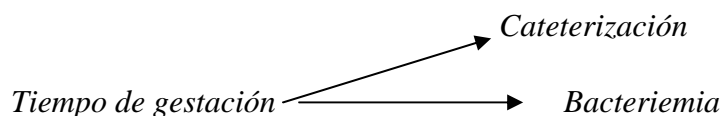
Hay varias cosas para destacar en estos ejemplos (1), (3) y (4). Lo más notable es que el desarrollo del último ejemplo demuestra el análisis incorrectamente efectuado en el primero. La relación real es entre los factores Tiempo de gestación y Bacteriemia. La Cateterización es independiente de la Bacteriemia cuando se remueve el Tiempo de gestación. Lo segundo es que cuando $OR = 1$, entonces el G se anula. Lo tercero es que a medida que OR crece, mucho más rápido crece G. Por último, si se tienen en cuenta los valores críticos de la Chi-cuadrado con un grado de libertad, se deduce que cuando $G \geq 3,841$ hay diferencia significativa y se puede probar asociación entre los dos factores analizados. El problema final es para cuál valor de OR el estadígrafo G comienza a ser significativo. Si se tuviese tal valor, se dispondría de un valor crítico para OR. La conclusión final es:

Se debe tener mucho cuidado en la elección de los factores para evitar enmascaramientos.

Los estadígrafos OR y RR se emplean como números índices, más que para armar un test de hipótesis del tipo G-test o χ^2 -test. Por eso, se estila armar intervalos de confianza de los mismos. De esta manera se puede cuantificar la relación entre los factores de riesgo. Luego, con ver si el valor unitario cae dentro del intervalo, se tiene el test respectivo.

15.9 Factores encajados o jerárquicos

A veces puede ocurrir que un factor de riesgo tenga mucha más importancia para una enfermedad que otro factor incidente. Como una especie de “categorización” o jerarquía de los factores. Imaginando, es como que uno está dentro de otro. Cuando esto ocurra se puede tener la tentación de quitar al menos importante a fin de ver mejor las cosas. En el ejemplo 3, de más arriba, el investigador separa el factor cateterización en dos partes para compararlo con la Bacteriemia, y así no puede comprobar la asociación entre estos dos factores. Su conclusión es que ambos factores son independientes y los separa. Pero como en el ejemplo 2, del mismo punto, otro investigador probó una asociación entre Tiempo de gestación y Cateterización, hace un planteo como:



El problema es que al desglosar los datos para forzarlos a entrar en una tabla de 2x2, el tamaño muestral se achica y eso va contra la primer regla de oro estadística. En un caso como el visto en los ejemplos, con datos hipotéticos donde las significaciones son tan extremas, no parece ser de importancia. Pero en casos reales donde no se tenga la suerte de encontrar valores muy grandes de G, para rechazar hipótesis, el disminuir N puede ser de importancia.

Si se miran los tres ejemplos mencionados con el OR, se puede notar que el máximo valor de $OR = 75$ se da para la combinación de Bacteriemia (B) con Tiempo de gestación (T). El factor Cateterización (C) con (B), tiene un $OR = 3,47$ y con (T) es de $OR = 5$. Mientras que al subdividir la muestra para separar el factor (T) y poder hacer dos comparaciones de (C) y (T), en ambos casos el OR es prácticamente uno. Es claro en el primer ejemplo que la *causa* es (C) y el *efecto* es (B). Cuando se hace la segunda comparación la *causa* es (T) y el *efecto* es (B). En la tercera se plantea como causa (T) y como efecto (C) lo que ya no es tan claro. Otro tanto ocurre en los dos últimos ejemplos. Si hubiera más factores todo sería más confuso. Por lo tanto, conviene ver las cosas de una manera diferente, antes que con la perspectiva de una tabla 2x2. En principio, la causa original, el factor de riesgo real es la cesárea. Esto se subdivide en dos casos: parto prematuro y parto normal (factor T). Por otra parte, también se puede subdividir en otros dos casos, teniendo en cuenta si fue necesario o no efectuar una cateterización. Se conforma así una partición del universo en cuatro casos mutuamente excluyentes entre sí. Si se efectúa una Tabla 2x2 con ellos, se encuentra significación -segunda tabla del ejemplo 2: (T) con (C). Pero parecería mejor imaginar a estos dos factores como uno encajado en el otro. Uno de primer nivel, en este ejemplo sería (T) porque produce mayor significación, y el otro (C) en un segundo nivel (subfactor). Entonces, se podría presentar una Tabla 2x4 para representar mejor las cosas como se muestra a continuación en el ejemplo siguiente:

Ejemplo) Usando los datos de los ejemplos anteriores de Bacteriemia en neo-natos, otro investigador decide presentar los datos con un modelo encajado de tablas y analizar los datos:

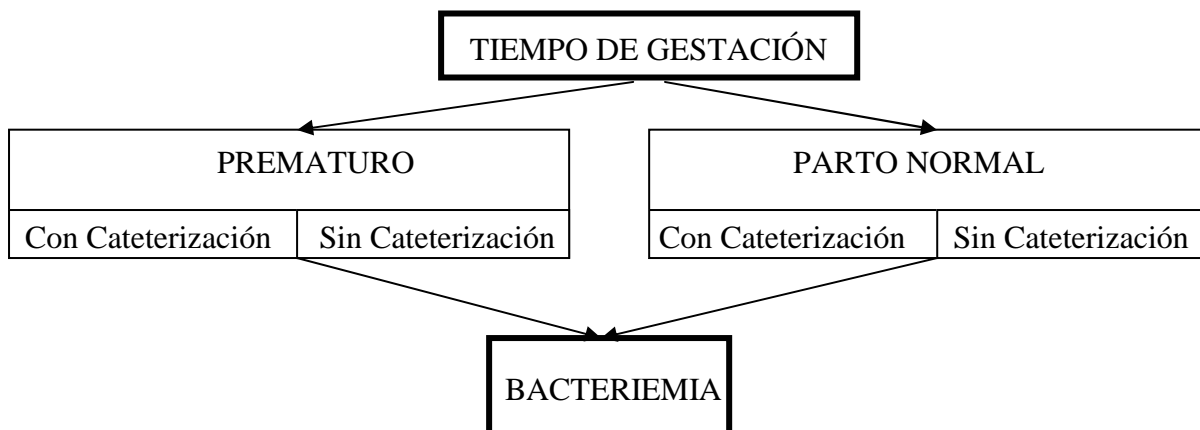
Frecuencias Observadas			Oij		BACTERIEMIA		Total
			SÍ	NO	SÍ	NO	
TIEMPO de GESTACION	Prematuro	Cateteriz.	SÍ	127	23	150	150
			NO	23	7	30	30
	Normal	Cateteriz.	SÍ	5	75	80	80
			NO	5	75	80	80
Total			160	180	340	340	

Frecuencias Esperadas			Eij		SÍ		NO	Total
			SÍ	NO	SÍ	NO		
TIEMPO de GESTACION	Prematuro	Cateteriz.	SÍ	70,59	79,41	150	150	
			NO	14,12	15,88	30	30	
	Normal	Cateteriz.	SÍ	37,65	42,35	80	80	
			NO	37,65	42,35	80	80	
Total			160	180	340	340		

Cálculo del estadígrafo G		
$G_{ij} = 2 O_{ij} \ln(O_{ij}/E_{ij})$		
j = 1	j = 2	i
149,18	-57,001	1
22,4512	-11,4702	2
-20,1882	85,7175	3
-20,1882	85,7175	4
G_j		
131,255	102,964	
$G = G_1 + G_2 = 234,219$		

Al efectuar una tabla de 2x4 con todos los datos obtenidos, usando una clasificación jerárquica, el resultado final es $G = 234,22^{***}$ que es el máximo encontrado. En la tabla 2x2 hecha en el ejemplo 2, con los mismos factores (T) y (B) el valor era ligeramente menor $G = 232$. Notar que la tabla de 4x2 es una yuxtaposición de las dos tablas no significativas, vistas en el ejemplo 3. Con todos los datos juntos se tiene una mejor visualización del problema.

El esquema final de la interacción de los factores se puede presentar con:



Y mirando todos los caminos posibles para ir de T a B, el de mayor riesgo es aquel que combina: Prematuro- Con Cateterización, para contraer la enfermedad. Su contribución particular al valor total del estadígrafo G es netamente superior a los demás.

15.10 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|---|-------|-------|
| 1) El análisis de frecuencias se puede aplicar a cualquier magnitud, dividida en clases. | V | F |
| 2) La validación estadística se puede hacer con: | | |
| 3) El G-test es mejor que la prueba con Chi-cuadrado porque es más sensible. | V | F |
| 4) El G-test es mejor que la prueba con Chi-cuadrado porque es menos robusto. | V | F |
| 5) La expresión matemática para calcular χ^2 es: | | |
| 6) Las hipótesis extrínsecas toman en cuenta el número de parámetros poblacionales r. | V | F |
| 7) Las hipótesis intrínsecas hacen que los grados de libertad sean: n-1. | V | F |
| 8) La corrección de Yates se debe hacer para corregir la continuidad. | V | F |
| 9) Cuando se usa una distribución continua, para testear una variable discreta hay que hacer: | | |
| 10) El G-test se llama test de la razón de verosimilitudes (<i>likelihood</i>). | V | F |
| 11) El cociente de frecuencias es el cociente de verosimilitud. | V | F |
| 12) La distribución teórica de G es la probabilidad multinomial. | V | F |
| 13) No hay que hacer distinciones en G entre hipótesis intrínsecas y extrínsecas. | V | F |
| 14) La función distribución para efectuar el G-test es la Chi-cuadrado. | V | F |
| 15) Las tablas de contingencia se usan para validar la independencia de los factores. | V | F |
| 16) Para calcular el número de grados de libertad en una tabla se usa: (h-1)(k-1)-m | V | F |
| 17) Para el cálculo de frecuencias esperadas en una tabla de contingencia se usa el concepto | | |
| 18) Una frecuencia esperada E_{ij} en una tabla se calcula con: | | |
| 19) Una tabla 2x2 puede ser clasificada en tres modelos diferentes. | V | F |
| 20) La clasificación se basa en cómo fija el investigador el tamaño muestral. | V | F |
| 21) Para un Modelo I y II conviene usar el test exacto de Fisher. | V | F |
| 22) Para corregir el G-test por continuidad se usa la y se calcula con | | |
| 23) Los índices: RR y OR se obtienen de una tabla cualquiera de 2x2. | V | F |
| 24) El RR es un cociente de probabilidades entre los expuestos y los no expuestos. | V | F |
| 25) Si $OR = 1$ los factores analizados son independientes. | V | F |
| 26) El mejor diseño experimental para determinar OR es el "doble-ciego". | V | F |

- | | | |
|--|---|---|
| 27) La eficacia se calcula con el OR. | V | F |
| 28) Si el OR crece, aumenta el valor del estadígrafo G. | V | F |
| 29) Cuando $OR = 1$ resulta $G = 0$. | V | F |
| 30) El OR permite comparar dos factores con validación para ver si hay relación entre ellos. | V | F |
| 31) Se debe tener cuidado de evitar enmascaramientos. | V | F |
| 32) Para estudiar la concordancia entre dos métodos clínicos las muestras son independientes | V | F |
| 33) Cuando se realizan dos mediciones a cada individuo se puede usar el test de Student | V | F |
| 34) Las magnitudes para usar el test de McNemar deben ser dicotómicas. | V | F |
| 35) Si no lo son se puede definir un punto de corte para transformarlas. | V | F |
| 36) El G-test aplicado al caso de McNemar tiene mayor poder que este. | V | F |
| 37) El test de McNemar es igual al de Cochran aplicado a dos muestras. | V | F |
| 38) Lo mismo que antes ocurre si se usa la corrección de Yates. | V | F |
| 39) El test Q de Cochran requiere que las magnitudes sean dicotómicas. | V | F |
| 40) El problema de los test de McNemar, G-test y Cochran es que no toman en cuenta N | V | F |
| 41) El test de Logs Odds Ratio no tiene problemas para estudiar la concordancia. | V | F |
| 42) El Test de Cohen-Kappa es la mejor opción disponible para la concordancia | V | F |
| 43) El Odds de discordancia puede cuantificar el concepto de concordancia. | V | F |
| 44) Explicar el diseño encajado para estudiar los factores de riesgo: | | |

2) En tres turnos de producción en una fábrica: Mañana, Tarde y Noche, se producen accidentes laborales. De acuerdo con los datos siguientes, probar la H_0 que la proporción de accidentes es la misma para los tres turnos laborales.

	Mañana	Tarde	Noche	Total
Empleados	150	180	120	450
Accidentes	5	6	5	16
Total	155	186	125	466

3) Un investigador midió la forma en que se repartían las visitas de ratones a trampas tratadas previamente con orina de ratones, con otras limpias de olor. Probar que el tratamiento efectuado a las trampas no afecta la frecuencia de visita.

	Limpias	Con orina	Total
Visitadas	3	17	20
No visitadas	9	23	32
Total	12	40	52

Datos de Perrota, C.A. Un. Wisconsin, 1954.

4) Un total de 130 ratones fue seleccionado para probar la penicilina, luego de haber sido inyectados con un *staphylococcus aureus* cultivado en un caldo enriquecido. La tabla de supervivencia obtenida fue la siguiente. Verifica si la penicilina tuvo incidencia en la cura.

	Vivos	Muertos	Total
Control	8	12	20
Penicilina	48	62	110
Total	56	74	130

5) Luego de dos años de investigaciones sobre una nueva medicación y tratamiento para un tipo de cáncer, el grupo obtuvo los datos de 80 pacientes elegidos:

	Sin Tratar	Tratados	Total
Sin tratar	11	29	40
Tratados	20	20	40
Total	32	48	80

Hace la prueba de Chi- cuadrado y no le da significativa si hace la corrección de Yates. Se le ocurren tres opciones:
 1) Presentar el informe sin la corrección.
 2) Mentir modificando los datos, bajando de 11 a 10 para hacerla significativa.
 3) Usar el G-test y ver qué pasa.

Decidir cuál de las opciones es la que debería elegir y dar las razones para la misma.

6) Para verificar la concordancia de dos métodos A y B se realizaron 200 mediciones y los resultados fueron los siguientes:

Método B	Método A		Total
	(+)	(-)	
(+)	80	15	95
(-)	10	95	105
Total	90	110	200

- Aplicar los test estadísticos de McNemar con y sin la corrección de Yates, el G-test con y sin la corrección de Williams, el Q-test para dos muestras, el Log-Odds Ratio test y explicar los resultados obtenidos y las conclusiones estadísticas.
- Aplicar a los mismos datos el método del Odds de Discordancia y explicar las conclusiones clínicas.
- Comparar las conclusiones clínicas con las estadísticas y discutir los resultados.

7) Lo mismo que el problema anterior pero con los datos siguientes:

Método B	Método A		Total
	(+)	(-)	
(+)	9978	22	10000
(-)	10	9990	10000
Total	9988	10012	20000

8) Lo mismo que antes pero con los datos

Método B	Método A		Total
	(+)	(-)	
(+)	10	95	105
(-)	80	15	95
Total	90	110	200

Comparar estos resultados con los del problema (6) anterior.

Apéndice 1: Reproducibilidad en casos binarios con el método de Guttman.

El estudio de la reproducibilidad en casos binarios se realiza con un diseño experimental llamado test-retest. Consiste en medir dos veces al mismo individuo, con el mismo método clínico en magnitudes binarias. Los resultados se pueden presentar en una tabla como sigue:

Retest	Test		Total
	(+)	(-)	
(+)	a	b	a + b
(-)	c	d	c + d
Total	a + c	b + d	N

N individuos al azar fueron sometidos al mismo test dos veces. Los resultados se agrupan según si hay concordancias (a y d) o si hay discordancias (c y b) en el estudio. Una perfecta reproducibilidad implica que c y b son nulos.

Para analizar este tipo de estudio en forma estadística se desarrolló el método de Guttman (1947) para ciencias humanas. Allí los tests son cuestionarios y las respuestas son generalmente del tipo verdadero/falso. No se espera que los cuestionarios varíen entre una y otra aplicación, salvo alguna pregunta redactada en forma engañosa. Por lo tanto, la variabilidad encontrada en el estudio es atribuida al sujeto analizado. Sin embargo, en análisis clínicos se puede suponer a la muestra constante (adoptando las precauciones necesarias para que el analito no varíe entre el test y el retest); y por eso la variabilidad se puede atribuir a la inestabilidad del método clínico. Las causas de variabilidad pueden ser: fluctuación de los instrumentos, diferentes reacciones químicas, cambios en los observadores, etc. Este método consiste en calcular dos límites, dentro de los cuales se espera encontrar el porcentaje de reproducibilidad del método.

$$\text{Límite inferior} = (k/k-1) (F-1/k) = 2F - 1$$

$$\text{Límite superior} = (k/k-1) (h - 1/k) = 2h - 1$$

Donde: F es la suma de los valores máximos de cada columna (suma de frecuencias relativas).

G es la suma de las frecuencias relativas de la diagonal principal: [(a + d)/N]

$$h = (1/k) \{ 1 + \sqrt{(k-1)(kG-1)} \} = (1/2) \{ 1 + \sqrt{(2G-1)} \}$$

$$k = 2 \text{ (porque hay dos clases en cada resultado posible (positivo y negativo))}$$

Para ilustrar el procedimiento se presenta el siguiente ejemplo:

Ejemplo: Se aplicó la determinación de VDRL a 872 pacientes dos veces seguidas. A cada muestra del paciente se la dividió en dos alícuotas, usando la primera para el test y la segunda para el retest. Los resultados se muestran a continuación:

Retest	Test		Total
	(+)	(-)	
(+)	400	12	412
(-)	50	410	460
Total	450	422	872

Los valores máximos de cada columna sumados son:
 $F = (400 + 410) / 872 = 0,929$
 $G = (400 + 410) / 872 = 0,929$ luego será
 $h = (1/2) \{ 1 + \sqrt{(2 \times 0,929 - 1)} \} = 0,963$
 Límite inferior = $[2 (0,929)] - 1 = 0,858$
 Límite superior = $[2 (0,963)] - 1 = 0,926$

Donde la reproducibilidad se halla entre el 85,8% y el 92,6%, valor considerado como muy bueno. Por lo tanto se concluye que la reproducibilidad de la VDRL empleada es aceptable desde un punto de vista clínico.

16

Bondad de ajuste

En el capítulo anterior se mostró la primera parte del análisis de frecuencias y se mencionó el empleo de la prueba de Chi-cuadrado y la prueba de G, para determinar si los valores observados se ajustaban a los esperados. En este tipo de ensayos se busca establecer si las diferencias observadas entre los datos reales y los teóricos se deben al azar, o si por el contrario la teoría no es buena para explicar la realidad. A eso se lo denomina ensayos de *bondad de ajuste*. Los problemas vistos cuando las frecuencias esperadas se calculaban con la teoría mendeliana, son en esencia casos de bondad de ajuste. En cambio, en los problemas presentados como tablas de contingencia, se buscaba probar la *independencia* de los factores involucrados entre sí. En este capítulo se tratarán con más detalle los casos de ajuste. El caso clásico y más difundido es la prueba de Chi-cuadrado. Desde 1900 y a través de su revista *Biometrika*, Pearson difundió su prueba ampliamente. El ejemplo más famoso fue el informe de Student sobre el error de recuento en hemocitómetro, donde proponía ajustar con una Poisson los valores observados, en lugar de usar la acostumbrada curva de Gauss. En las últimas décadas la prueba de la G se está imponiendo por su mayor sensibilidad y robustez. Se constituye en la opción más recomendable para el caso paramétrico. Cuando los supuestos no se cumplen, o sencillamente por su facilidad de ejecución, la prueba más recomendable en el campo no paramétrico es la de Kolgomorov-Smirnov, para una y para comparar dos muestras entre sí. Todos estos modelos se presentan a continuación.

16.1 El método clásico de Pearson

Este método aplica el test de la distribución de Pearson: la Chi cuadrado. Consiste en comparar las frecuencias observadas en un experimento, con sus respectivas frecuencias observadas, obtenidas con la distribución teórica que explica el fenómeno estudiado. Hasta la primera década del siglo XX, la distribución Gaussiana era el enfoque aceptado por todos para explicar la Teoría de errores en mediciones. El punto de inflexión histórico a esa tendencia se produce en 1907, con el trabajo de un ingeniero químico W. Gosset, quien en su tarea diaria efectuaba conteos de células de levadura de cerveza en una cámara de recuento de Neubauer denominada hemocitómetro. Gosset sospechó que sus datos se ajustaban mejor a una distribución Poisson que a la de Gauss y para probarlo usó el método de Pearson. Publicó sus trabajos en la revista *Biometrika* dirigida por el mismo Pearson, su mentor en cuestiones matemáticas. Estas publicaciones llegaron a manos de un estudiante avanzado escocés: R.A. Fisher, quien usó un comentario de

Gosset respecto a la posible existencia de una distribución teórica, mejor que la de Gauss en muestras pequeñas, para desarrollar un nuevo modelo: el de Student. Allí comenzó la moderna estadística, con aplicaciones innovadoras en el trabajo experimental. Parece adecuado comenzar este tema con ese caso. En la Tabla 16.1 se presentan los datos de Student (1907) para mostrar el uso del ajuste a una función de Poisson:

Tabla 16.1 El método de la Chi cuadrado.

N	O _i	E _i	O* _i	E* _i	$\chi^2=(O^*i-E^*i)^2/E^*i$	χ^2_{corr}
0	75	66,12	75	66,12	1,193	1,062
1	103	119,02	103	119,02	2,156	2,024
2	121	107,11	121	107,11	1,801	1,674
3	54	64,27	54	64,27	1,641	1,485
4	30	28,92	30	28,92	0,040	0,012
5	13	10,41	17	14,56	0,409	0,258
6	2	3,12				
7	1	0,80				
8	0	0,18				
9 y más	1	0,05				

N : número de células dentro de cada cuadrícula.
O_i frecuencia observada (cantidad de cuadrículas con N células adentro).
E_i : frecuencia esperada obtenida con la fórmula de Poisson.
E*_i : frecuencia agrupada.

Si la frecuencia observada es muy chica ($O_i < 5$), los valores de frecuencia esperada se distorsionan mucho y los valores de χ^2 respectivos contribuyen demasiado al total, pudiendo mostrar significación donde en realidad no la hay. Una regla práctica establecida hace muchos años es agrupar las frecuencias anexas (cuando $O_i < 5$) en un valor manejable. En la Tabla 16.1 como las cuatro últimas frecuencias observadas son chicas, se agrupan con la más cercana, lo mismo que con las esperadas. Entonces, los datos agrupados se muestran en las columnas cuarta y quinta de la tabla ($O^*6 = 13+2+1+0+1 = 17$ y la esperada es: $E^*6 = 10,41+3,12+0,8+0,18+0,05= 14,56$) Luego de efectuar este arreglo, se calculan los valores de χ^2 respectivos que se colocan en la sexta columna y se obtiene un total de $\chi^2 = 7,24$. Si se aplica la corrección de Yates el valor disminuye mucho $\chi^2_{corr} = 6,515$.

Los grados de libertad se calculan de la manera siguiente: Hay 10 clases de frecuencias observadas, pero con el agrupamiento quedan $n^* = n-4 = 6$. En una distribución de Poisson se usa un solo dato poblacional μ , luego resulta $r = 1$. Entonces, los grados de libertad son:

$$v = n^* - r - 1 = (10 - 4) - 1 - 1 = 4$$

De tablas es $\chi^2_{0,95; 4} = 9,488 > \chi^2_{corr} = 6,515$ por lo tanto no se puede rechazar la hipótesis nula. En cambio, si se hubieran estimado las probabilidades con la función de Gauss en vez de la Poisson, no habría sido lo mismo: con Poisson se ajusta mejor que la de Gauss

Como los grados de libertad dependen del número de parámetros poblacionales que se usan con cada modelo teórico, y con el número de clases luego de ser agrupadas, las fórmulas para calcularlos más habituales son:

Tabla 16.2 Grados de libertad

Distribución	Parámetros pob.	r	$\nu = n^* - r - 1$	Donde n^* se calcula con el número de clases, del problema analizado, luego de hacerse la agrupación para que no ocurra que $n < 5$
Poisson	μ	1	$n^* - 2$	
Gauss	$\mu ; \sigma$	2	$n^* - 3$	
Binomial	p	1	$n^* - 2$	

16.2 El método moderno con G-test

Como se vio en el capítulo anterior, el G-test es más sensible y robusto que la prueba de la Chi cuadrado. Por lo tanto, se torna el método más recomendable para analizar el problema de la bondad de ajuste. Su uso es totalmente análogo al visto más arriba, con la salvedad de emplear otro estadígrafo G, para compararlo con los valores críticos de la Chi cuadrado y poder decidir acerca de la hipótesis nula. Aplicando este modelo a los datos de la Tabla 16.1 anterior se tiene:

Tabla 16.3 La prueba de G con Poisson.

N	O _i	E _i	O* _i	E* _i	G _i	Los grados de libertad en este caso son: $\nu = n^* - r - 1 = (10-3) - 1 - 1 = 5$ Y la corrección de Williams es: $Q = 1 + [(k^2 - 1)] / 6N\nu$ Con $k = 10$: número de clases $N = 400$: número de casillas $Q = 1 + [(100-1) / (6 \cdot 400 \cdot 5)] = 1,00825$
0	75	66,12	75	66,12	18,903	
1	103	119,02	103	119,02	-29,780	
2	121	107,11	121	107,11	29,508	
3	54	64,27	54	64,27	-18,804	
4	30	28,92	30	28,92	2,200	
5	13	10,41	13	10,41	5,777	
6	2	3,12	4	4,15	-0,295	
7	1	0,80				
8	0	0,18				
9 y más	1	0,05				
Total	400	400,00	400	400	7,509	

Si la frecuencia esperada es muy chica ($E_i < 3$), los valores de G no se pueden calcular. Una regla práctica es agrupar las frecuencias anexas tal que E^*i sea mayor o igual a 3. En la Tabla 16.2 como las tres últimas frecuencias esperadas son chicas, se agrupan con la más cercana (la sexta) y lo mismo se hace con las observadas: O^*i . Entonces, los datos agrupados se muestran en las columnas cuarta y quinta de la tabla ($O^*7 = 3,12+0,8+0,18+0,05 = 4,15$ y la observada es: $O^*5 = 2+1+0+1 = 4$). Luego de efectuar este arreglo, se calculan los valores de G respectivos, que se colocan en la sexta columna y se obtiene un total de $G = 7,509$. Si se aplica la corrección de Williams resulta $G_{corr} = G / Q = 7,50825 / 1,00825 = 7,45$. Como los grados de libertad son $\nu = 4$ de es $\chi^2_{0,95; 4} = 9,488 > G_{corr} = 7,45$ por lo tanto no se puede rechazar la hipótesis nula.

En el tema 7 se desarrolló un ejemplo de aplicación usando la probabilidad binomial, para estudiar el caso de 6.115 familias con 12 hijos. En el cuadro 7.1 se volcaron los datos de las frecuencias observadas y calculadas. Tomando esa información se prepara la Tabla 16.4 para ilustrar el uso de la prueba G con una población supuesta como binomial:

Tabla 16.4 La prueba de G con la Binomial.

Hijas mujeres	Observadas Oi	Esperadas Ei	O*i	E*i	Gi
0	7	2,346			
1	45	26,082	52	28,428	62,802
2	181	132,836	181	132,836	111,996
3	478	410,016	478	410,016	146,664
4	829	854,247	829	854,247	-49,740
5	1112	1265,628	1112	1265,628	-287,804
6	1343	1367,279	1343	1367,279	-48,124
7	1033	1085,211	1033	1085,211	-101,869
8	670	628,055	670	628,055	86,631
9	286	258,475	286	258,475	57,882
10	104	71,803	104	71,803	77,057
11	24	12,089	27	13,022	39,377
12	3	0,933			

Los grados de libertad son:
 $\nu = n^* - r - 1 = 9$

La corrección de Williams es:
 $Q = 1 + [(k^2 - 1)] / 6N\nu$

Con $k = 11$
 $N = 6.115$ resulta

$Q = 1 + [(121-1)/(6 \cdot 6.115 \cdot 9)]$
 $Q = 1,0003634$

$G_{corr} = 94,83^{***}$

$\chi^2_{0,999; 9} = 27,877 \lll G_{corr}$

Se rechaza la hipótesis nula que suponía una distribución binomial en la población. Se encontró evidencia altamente significativa que demuestra que el ajuste no es bueno.

En el tema 9 se presentó un ejemplo sobre las distribuciones en peso de 195 varones. Los datos mostrados en el Cuadro 9.4 se vuelcan a continuación en las dos primeras columnas de la Tabla 16.5, para poder usar el G-test en el caso de una distribución poblacional gaussiana:

Tabla 16.5 La prueba de G con la Normal.

Intervalos del peso	Observadas Oi	Esperadas Ei	O*i	E*i	Gi
hasta 63,5	3	3,3			
63,5-65,5	14	12,2	17	15,5	3,140693
65,5-67,5	30	31,7	30	31,7	-3,307158
67,5-69,5	48	51,1	48	51,1	-6,008015
69,5-71,5	48	50,7	48	50,7	-5,25359
71,5-73,5	39	31,1	39	31,1	17,6556
73,5-75,5	11	11,7	13	14,9	-3,546708
75,5 y más	2	3,2			
Total	195	195	6115	195	2,68082

Los grados de libertad son:
 $\nu = n^* - r - 1 = 6 - 2 - 1 = 3$

La corrección de Williams es:
 $Q = 1 + [(k^2 - 1)] / 6N\nu$

Con $k = 6$
 $N = 195$ resulta

$Q = 1 + [(36-1)/(6 \cdot 195 \cdot 3)]$
 $Q = 1,00598$

$G_{corr} = 2.664 < \chi^2_{0,95;3} = 11,07$

Para este caso, al número total de clases útiles $k=6$ hay que restarle $r = 2$ grados de libertad pues se usan los datos muestrales para estimar a los dos parámetros poblacionales (μ ; σ) necesarios para obtener las probabilidades gaussianas. Efectuada la corrección de Williams, el valor G_{corr} es menor que el crítico de tablas para el 95% de confianza. Por lo tanto, no se puede rechazar la hipótesis de que los pesos tienen una distribución normal en la población.

16.3 La prueba de Kolgomorov-Smirnov

Se trata de un modelo *no-paramétrico* para comprobar bondad de ajuste. La clasificación de este modelo se debe al hecho que no depende de una distribución original específica; no se sabe a priori la forma de la distribución de donde se sacan los datos. Por ejemplo el modelo de Student requiere una población original de tipo gaussiana, como la de Chi-cuadrado y la de Fisher. Aquí se trata de *distribuciones libres*, sin parámetros de tipo prefijado como μ o σ . Es cuando se está más interesado en comparar distribuciones antes que parámetros, ni tampoco cuando se trata de estimar parámetros (*modelos paramétricos*). El modelo de Kolgomorov-Smirnov se aplica a distribuciones de tipo continuas y es considerada *conservadora*. También se la usa para probar hipótesis acerca de distribuciones discretas. Es un modelo fácil de realizar. Se basa en calcular las diferencias, en valor absoluto, entre las frecuencias acumuladas relativas observadas y las esperadas, en cada clase. Luego se busca la mayor de las diferencias en valor absoluto, y el estadígrafo $D_{\text{máx}}$ así obtenido se compara con el valor crítico de tablas. En este punto de trata del modelo aplicado al caso de una sola muestra y la Tabla 13 del anexo se presentan los valores críticos para validar hipótesis. Este modelo presenta especiales ventajas cuando se lo aplica a muestras pequeñas, y en tales casos no se necesita agrupar en clases. Otra ventaja es que permite construir límites de confianza para la distribución acumulada completa.

$$D_{\text{máx}} = | O_{Ai} - E_{Ai} |_{\text{máx}}$$

Donde O_{Ai} : frecuencia acumulada relativa de la clase i observada
 E_{Ai} : frecuencia acumulada relativa de la clase i esperada

Para comparar este estadígrafo con un valor crítico de tablas, se busca ese valor en tablas entrando con el tamaño muestral N . Hay dos tablas para los valores críticos de esta distribución. La Tabla 13 que se emplea para los casos más comunes donde se usan los datos muestrales para estimar los valores poblacionales, o sea, para cuando se usan *hipótesis intrínsecas*. A su vez, la Tabla 14 es también para Kolgomorov-Smirnov aplicado a una sola muestra, pero para los raros casos donde se usan *hipótesis extrínsecas*.

La Tabla 13 se extiende correlativamente hasta $N = 30$, y luego a partir de allí se usa una aproximación dada por las relaciones siguientes:

Cuadro 16.1 Valores críticos para la prueba de Kolgomorov-Smirnov en una sola muestra.

Hipótesis intrínsecas			
$N > 30$	α	$K\alpha$	$D\alpha = K\alpha / \sqrt{N}$
	0,10	0,805	$0,805 / \sqrt{N}$
	0,05	0,886	$0,886 / \sqrt{N}$
	0,01	1,031	$1,031 / \sqrt{N}$

Significativo
Muy significativo

La Tabla 14 del Anexo es hasta una muestra de tamaño 100. Por lo tanto se debe usar la aproximación asintótica dada por:

$$D\alpha = \sqrt{\frac{-\ln(\alpha/2)}{2N}} = K\alpha / \sqrt{N}$$

Y para los valores críticos usuales se puede preparar otro cuadro resumen:

Cuadro 16.2 Valores críticos para la prueba de Kolgomorov-Smirnov en una sola muestra. Hipótesis extrínsecas

N > 100	α	$K\alpha$	$D\alpha = K\alpha / \sqrt{N}$	
	0,05	1,358	$1,358 / \sqrt{N}$	Significativo
	0,01	1,628	$1,628 / \sqrt{N}$	Muy significativo
	0,001	1,949	$1,949 / \sqrt{N}$	Altamente significativo

Este modelo en muchos casos posee mayor potencia que la prueba de la Chi cuadrado para bondad de ajuste. Su uso es más sencillo y es la recomendada para realizar este tipo de estudios en el área de los análisis clínicos y farmacológicos. Solo es aventajada en sensibilidad por el G-test, pero se puede aplicar a casi todos los casos. Para mostrar el uso de esta prueba se desarrollan los ejemplos siguientes, usando los datos de los casos vistos en el punto anterior:

Ejemplo 1) Aplicar el test al caso visto en la Tabla 16.3 de informe de Student en 1907.

n	Oi	Ei	Oacum	Eacum	OAi	Eai	OAi - Eai	
0	75	66,12	75	66,12	0,1875	0,1653	0,0222	→ D _{máx}
1	103	119,02	178	185,14	0,4450	0,4629	0,0179	
2	121	107,11	299	292,25	0,7475	0,7306	0,0169	
3	54	64,27	353	356,52	0,8825	0,8913	0,0088	
4	30	28,92	383	385,44	0,9575	0,9636	0,0061	
5	13	10,41	396	395,85	0,9900	0,9896	0,0004	
6	2	3,12	398	398,97	0,9950	0,9974	0,0024	
7	1	0,80	399	399,77	0,9975	0,9994	0,0019	
8	0	0,18	399	399,95	0,9975	0,9999	0,0024	
9 y más	1	0,05	400	400	1,0000	1,0000	0,0000	
Total	400	400,00						

Para este problema como se usan los datos muestrales para estimar μ con la media y aplicar Poisson se trata de un caso de hipótesis intrínsecas. Luego será:

$$D\alpha = K\alpha / \sqrt{N} = 0,886 / 20 = 0,0443 > D_{máx} = 0,0222 \quad (95\%)$$

Y no se puede rechazar la hipótesis nula. La misma conclusión que la obtenida con el G-test. Aunque el resultado está muy cerca del límite para una confianza del 95%, alcanza para ver la aplicación de este modelo menos poderoso que el G-test.

Ejemplo 2) Aplicar el test al caso visto en la Tabla 16.4, del estudio de 6115 familias de Sajonia con 12 hijos, y ver la distribución de sexos entre ellos; n es la cantidad de hijas mujeres. Se postula una distribución binomial en esta población

n	Oi	Ei	Oacum	Eacum	OAi	EAI	OAI- EAI
0	7	2,346	7	2,346	0,0011	0,0004	0,0007
1	45	26,082	52	28,428	0,0085	0,0046	0,0039
2	181	132,84	233	161,26	0,0381	0,0264	0,0117
3	478	410,02	711	571,28	0,1163	0,0934	0,0228
4	829	854,25	1540	1425,5	0,2518	0,2331	0,0187
5	1112	1265,6	2652	2691,2	0,4337	0,4401	0,0064
6	1343	1367,3	3995	4058,4	0,6533	0,6637	0,0104
7	1033	1085,2	5028	5143,6	0,8222	0,8412	0,0189
8	670	628,06	5698	5771,7	0,9318	0,9439	0,0121
9	286	258,48	5984	6030,2	0,9786	0,9861	0,0076
10	104	71,803	6088	6102	0,9956	0,9979	0,0023
11	24	12,089	6112	6114,1	0,9995	0,9998	0,0003
12	3	0,933	6115	6115	1	1,0000	0,0000
Total	6115	6115					

→ Dmáx

El valor crítico para hipótesis intrínsecas es:

$$D\alpha = K\alpha / \sqrt{N} = 0,886 / 78,2 = 0,0113 < Dmáx = 0,0228 \text{ (95\%)}$$

Y

$$D\alpha = K\alpha / \sqrt{N} = 1,031 / 78,2 = 0,0132 < Dmáx = 0,0228 \text{ (99\%)}$$

Por lo tanto se tiene evidencia muy significativa para rechazar la hipótesis de una distribución binomial en la población.

Ejemplo 2) Aplicar el test al caso visto en la Tabla 16.5, del estudio del peso de 195 varones suponiendo una distribución Normal del peso en la población.

Peso	Oi	Ei	Oacum	Eacum	OAI	EAI	OAI-EAI
hasta 63,5	3	3,3	3	3,3	0,0154	0,0169	0,0015
63,5-65,5	14	12,2	17	15,5	0,0872	0,0795	0,0077
65,5-67,5	30	31,7	47	47,2	0,2410	0,2421	0,0010
67,5-69,5	48	51,1	95	98,3	0,4872	0,5041	0,0169
69,5-71,5	48	50,7	143	149,0	0,7333	0,7641	0,0308
71,5-73,5	39	31,1	182	180,1	0,9333	0,9236	0,0097
73,5-75,5	11	11,7	193	191,8	0,9897	0,9836	0,0062
75,5 y más	2	3,2	195	195,0	1,0000	1,0000	0,0000
Total	195	195					

→ Dmáx

El valor crítico para hipótesis intrínsecas es:

$$D\alpha = K\alpha / \sqrt{N} = 0,886 / 13,96 = 0,0635 > D_{\text{máx}} = 0,0308 \text{ (95\%)}$$

Y por lo tanto no se puede rechazar la hipótesis nula, de una distribución gaussiana de los pesos.

16.4 Test de Kolgomorov-Smirnov para 2 muestras

Es un modelo estadístico para verificar si *dos muestras independientes* han sido extraídas de la misma población o de poblacionales con igual función distribución. La prueba de dos colas es sensible a cualquier clase de diferencia entre las distribuciones de donde fueron extraídas ambas muestras. La prueba de una cola se usa para decidir si los valores de la población, de donde se extrajo a una de las muestras, son estocásticamente mayores (o menores) que las de la otra población. Por ejemplo, para probar la predicción de que los valores de un grupo experimental serán mejores que los del grupo control.

Como en el caso de una sola muestra, el modelo compara las dos frecuencias acumulativas relativas de ambas muestras y busca la mayor diferencia encontrada. Este valor multiplicado por ambos tamaños muestrales, se contrasta contra un valor de tablas. Sean las muestras A y B, con un tamaño muestral de N_A y N_B respectivamente, y con $D_{\text{máx}}$ como la mayor diferencia encontrada en valor absoluto, entonces el valor de K se calcula con:

$$K = N_A \cdot N_B \cdot D_{\text{máx}}$$

Si este valor K es mayor que el de tablas $K\alpha$ para un nivel de significación dado, se rechaza la hipótesis nula de que ambas muestras provienen de la misma población. Cuando el tamaño de muestras es menor a 25 se pueden usar las Tablas 14 del Anexo. Cuando las muestras son grandes se puede emplear una aproximación, como se muestra al final de las Tablas 14, donde se da el valor crítico de $D\alpha$ para comparar con $D_{\text{máx}}$. Finalmente, en la Tabla 15 y 16 del Anexo se presenta una versión simplificada de las tablas, para el caso de igual tamaño muestral, o sea cuando $N_A = N_B = n$; en esta última también se expresan los valores de $D_{\text{máx}}$ a comparar con $K\alpha$. Para ilustrar estas ideas se presenta el caso siguiente:

Ejemplo 1) De un estudio publicado por la *J. Amer. Med. Ass.* De Friedman, M. Et al., 217, pág. 929-932 (1971) se sacaron los datos de la tabla siguiente:

N	1	2	3	4	5	6	7	8	9	10	11
Grupo A	3,6	2,6	4,7	8,0	3,1	8,8	4,6	5,8	4,0	4,6	
Grupo B	16,2	17,4	8,5	15,6	5,9	9,8	14,9	16,6	15,9	5,5	7,5

A dos grupos de personas, descansados y luego de ingerir una infusión de Hipoclorito de argenina, se le midió los niveles pico de la hormona de crecimiento en plasma. Los sujetos se clasificaron en dos grupos. Grupo A: se trata de individuos relativamente predispuestos a enfermedades coronarias, cuyo temperamento se caracteriza por un excesivo sentido de competitividad, liderazgo y urgencia de tiempo. Grupo B: estos son relativamente resistentes a tener problemas co-

ronarios y cuyo temperamento es lo inverso del grupo anterior. Estudios anteriores del mismo equipo (1950) indicaban que el Grupo A tenía mayor tendencia a sufrir enfermedades coronarias que los del Grupo B. En este trabajo, se investiga la incidencia del temperamento en la cantidad de hormona de crecimiento estimulada con la argenina.

La hipótesis de trabajo es que ambos grupos deben diferir. O sea, ambos grupos no provienen de la misma población. Se plantea entonces como

$H_0 : P (X \leq a) \geq P (Y \leq a)$ No hay diferencia entre ambas muestras.

$H_1 : P (X \leq a) < P (Y \leq a)$ El nivel hormonal es menor en el Grupo A que en el B.

Para desarrollar este modelo se deben seguir los pasos siguientes:

Paso 1) Se ordena cada grupo en orden ascendente. Con el rango se obtiene el número de clases, como en los histogramas, pero tratando de tener el mayor número de clases posible, para ambas muestras. Los resultados se muestran en la primer columna del cuadro siguiente.

Paso 2) Se calculan las frecuencias de cada grupo en las mismas clases, luego sus frecuencias acumuladas y finalmente, dividiendo por el respectivo tamaño muestral, se calculan las frecuencias acumuladas relativas, que se muestran en las columnas 2 y 3.

Paso 3) Para cada clase, se calculan las diferencias en valor absoluto de las frecuencias acumuladas relativas de cada grupo. Y se busca el valor máximo de esas diferencias $D_{m\acute{a}x} = 0,7$

Paso 4) Se determina cual tabla de valores críticos usar. En este caso, se trata de muestras pequeñas con diferente tamaño muestral (Tabla 15) Con $N_A = 10$ y $N_B = 11$ se obtiene de tablas los valores críticos: $K_{0,05} = 60$; $K_{0,01} = 77$ y $K_{0,001} = 89$

Paso 5) Se comparan estos valores críticos con el estadígrafo K de Kolgomorov-Smirnov para dos muestras: $N_A \cdot N_B \cdot D_{m\acute{a}x} = K = 10 \cdot 11 \cdot 0,7 = 77$

	Grupo A	Grupo B	Diferencias
2,5 - 3,4	0,2	0	0,200
3,5 - 4,4	0,4	0	0,400
4,5 - 5,4	0,7	0	0,700 (D_{máx})
5,5 - 6,4	0,8	0,182	0,618
6,5 - 7,4	0,8	0,182	0,618
7,5 - 8,4	0,9	0,273	0,627
8,5 - 9,4	1	0,364	0,636
9,5 - 10,4	1	0,455	0,545
10,5 - 11,4	1	0,455	0,545
11,5 - 12,4	1	0,455	0,545
12,5 - 13,4	1	0,455	0,545
13,5 - 14,4	1	0,455	0,545
14,5 - 15,4	1	0,545	0,455
15,5 - 16,4	1	0,818	0,182
16,5 - 17,4	1	1	0,000

Se concluye que se debe rechazar la hipótesis nula y aceptar la alternativa. O sea, se tiene evidencia muy significativa $K = 77^{**}$ de que ambas muestras no provienen de la misma población, los valores hormonales del Grupo A son menores que los del Grupo B.

16.5 Test de Bondad de ajuste con repetición

Es un modelo estadístico para los casos en que los datos para un test se hacen de manera repetida. La manera más simple es agrupar los datos repetidos en una especie de “pool” de datos y usar los totales para hacer la prueba. Si bien se respeta la regla de oro de maximizar los datos, el problema es que se puede perder información a causa del agrupamiento y así no tener un panorama mejor. Para mostrar un caso de este tipo se presenta el ejemplo siguiente:

Ejemplo 1) Una industria farmacéutica encara un estudio sobre el efecto de un nuevo analgésico. Para ello, efectúa relevamiento en 8 ciudades diferentes de un país elegidas al azar, donde se efectúa la misma prueba. Todos los casos analizados con respuesta positiva, se dividen según el sexo. La hipótesis de trabajo es que este factor no tiene influencia en los resultados obtenidos. Los datos obtenidos se muestran en la tabla siguiente. Con los cuales se realiza un G-test como se vio en el punto 15.2 del capítulo anterior:

Sexo	Observadas	Esperadas	G	Como $\chi^2_{0,999; 1} = 10,828$ los resultados fueron altamente significativos y se tiene una muy fuerte evidencia para rechazar H_0
Femenino	616	530	185,26	
Masculino	444	530	-157,22	
Total	1060	1060	28,04***	

El farmacéutico a cargo del estudio estadístico deber rechazar la hipótesis del 50% para cada sexo, parecería que este factor tiene algo que ver. Sin embargo, antes de tomar una decisión al respecto, decide aprovechar al máximo los datos recogidos, con el desglose en las 8 localidades y enfoca el estudio desde otro ángulo:

Caso 1) Test de independencia (tabla de contingencia)

Frecuencias Observadas

Loc.	f	m	Total
1	83	47	130
2	77	43	120
3	110	96	206
4	92	58	150
5	51	31	82
6	48	61	109
7	70	42	112
8	85	66	151
Total	616	444	1060

Frecuencias Esperadas

Loc.	f	m	Total
1	75,547	54,453	130
2	69,736	50,264	120
3	119,71	86,287	206
4	87,17	62,83	150
5	47,653	34,347	82
6	63,343	45,657	109
7	65,087	46,913	112
8	87,751	63,249	151
Total	616	444	1060

Coloca los datos observado en una tabla como la de arriba y para el cálculo de las frecuencias observadas, usa el concepto de independencia. Es decir, multiplicando los totales marginales de cada casilla y dividiendo por $N = 1060$. En la primer casilla será $E_{11} = (130 \cdot 616) / 1060 = 75,547$ y así sucesivamente completa la tabla de frecuencias esperadas de más arriba. Luego hace el test:

Loc.	Gf	Gm	Gf+Gm
1	15,61784	-13,8356	1,782242
2	15,26	-13,4239	1,836087
3	-18,6161	20,48096	1,864869
4	9,923213	-9,27915	0,644065
5	6,924128	-6,35701	0,567118
6	-26,6275	35,34652	8,719038
7	10,1883	-9,29289	0,895414
8	-5,41473	5,619838	0,205107
Tot	7,25518	9,25876	16,51394

G = 16,51394

Para cada casilla se calcula el valor:

$$G_{ij} = 2 O_{ij} \cdot \ln(O_{ij} / E_{ij})$$

La suma total para filas y columnas es:

$$G = 16,514 *$$

Los grados de libertad son: $\nu = (8-1)(2-1) = 7$

De tablas $\chi^2_{0,95; 7} = 14,067$

Se tiene prueba (95%) de *heterogeneidad* en los datos. Esto significa, que el factor Sexo, no es independiente del factor Localidad. En la tercer columna se coloca el valor $G_f + G_m$ para poder realizar otro tipo de análisis. Comparando el peso que tiene cada localidad en el total, se nota que más del 50% del valor total de G se debe a la Localidad 6, mientras que las demás no parecen ser significativas, porque tienen una contribución parecida en el total. El investigador ahora sospecha que los resultados inesperados del primer análisis que efectuó, se originan en la gran desproporción encontrada en esta localidad. Cosa que deberá ser investigada con mucho cuidado.

En realidad, con cada localidad se puede armar una tabla 2x2, usando los datos observados en ella, y para obtener los esperados se usa una estimación del porcentaje poblacional. Esto es, la probabilidad de que sea de sexo femenino es $P(f) = 616 / 1060 = 0,581132$ y para el masculino es: $P(m) = 444 / 1060 = 0,418868$. Entonces, en una localidad cualquiera, por ejemplo la 6, se tendrá una frecuencia esperada femenina $E_f = 109 \cdot 0,581132 = 63,34$ y masculina $E_m = 109 \cdot 0,418868 = 45,66$ lo que se obtiene multiplicando la probabilidad por el tamaño muestral de la localidad. Se puede entonces armar una tabla para esa localidad como la siguiente :

Sexo	Observadas	Esperadas	G
Femenino	48	63,34	-26,628
Masculino	61	45,66	35,347
Total	109	109	8,719

Como $\chi^2_{0,95; 1} = 3,841$ los resultados fueron a significativos y se tiene evidencia para rechazar H_0

Realizado el test de G, resultó significativo pues $G = 8,719 * > \chi^2_{0,95; 1} = 3,841$. Análogamente se pueden plantear los G-test por localidad y comparar siempre contra el mismo valor crítico. Eso se muestra en la última columna de la tabla al principio de esta página. Se concluye que la única localidad con diferencia significativa es la Localidad 6. Como esto no es la tendencia general, algo debe haber ocurrido en tal localidad para tener esos resultados. Algo que merece ser investigado con cuidado en un próximo experimento, que deberá ser planeado cuidadosamente.

Se puede hacer un resumen de estos resultados calculando G agrupado = 1,5004 que se calcula usando $O_i = 616$; $O_j = 444$ y $E_i = 1060 (9/16)$; $E_j = 1060 (7/16)$

Tests realizados	Grados de libertad	Estadígrafo G
Total Datos Agrupados	1	1,500
Total por localidad	7	16,514*
TOTAL	8	18,014*

16.6 Análisis de concordancia

En el análisis de frecuencias el problema general es ver si los datos experimentales se ajustan a una teoría dada, la cual trata de interpretar la realidad en términos matemáticos. Esto se vio en varios capítulos anteriores y en el presente, con varios de los modelos presentados. Cuando el concepto de independencia matemática, visto en el capítulo 6, es la base de cálculo para obtener las frecuencias esperadas teóricamente se habla de tablas de contingencia. En cambio, cuando se usa otra teoría como la Mendeliana, la de Gauss, etc. para el cálculo de las frecuencias esperadas se habla de bondad de ajuste o de tablas de contingencia para el ajuste. En el fondo ambos tipos de tablas son dos formas de abordar el mismo problema: los datos muestrales versus la realidad. Lo que ocurre es que a veces la realidad se la toma de una teoría cualquiera y otras de la teoría de la independencia estadística. El problema básico es que no siempre la *asociación estadística significa asociación clínica*. Para ilustrar este concepto se desarrollará a continuación el caso del Análisis de Concordancia (*Agreement*) donde se mostrará que el significado estadístico de la asociación, a veces no tiene el sentido clínico del problema. Un problema habitual en Medicina es ver cuando un método “nuevo”, puede ser usado para reemplazar al habitual que se usa en el laboratorio (el “viejo”). El método más común de hacer esto es medir a cada individuo con los dos métodos a la vez, como se explicara en el Modelo de Student para muestras apareadas y en los modelos no paramétricos equivalentes.

Cuando a cada individuo se lo mide con más de un método a la vez, a las muestras obtenidas se las considera apareadas. En el caso general habrá k métodos, aplicado a n individuos y así el total de datos será $N = n \cdot k$. Se trata de casos especiales de Tablas de contingencia donde se puede aplicar el modelo de Cochran para analizar estos casos. Para el caso particular de tener $k = 2$ métodos clínicos, hay varios modelos que se pueden usar, se trata del caso de individuos medidos con ambas técnicas a la vez, y entonces resulta un caso particular de las tablas de contingencia de 2×2 . En estos casos la idea es armar la tabla de una manera tal que evite el apareo de los datos. El requisito básico para aplicar los modelos de tablas de contingencia es exigir la independencia de los datos, por lo tanto se debe evitar que un mismo paciente aparezca en dos lugares de la misma tabla, con un diseño como se mostró en la Tabla 14.3 anterior:

Cuadro 16.3 Tabla de Concordancia

Test 2	Test 1		Total
	(+)	(-)	
(+)	r_{11}	r_{12}	$r_{11} + r_{12}$
(-)	r_{21}	r_{22}	$r_{21} + r_{22}$
Total	$r_{11} + r_{21}$	$r_{12} + r_{22}$	N

El modelo más usado para este análisis es el Test de McNemar, corregido con Yates. Pero además se pueden usar el G-test de McNemar corregido con Williams, el test no paramétrico de Cochran y otros dos menos conocidos como el del Log-Odds Ratio y el test Cohen-Kappa. Como se mostrará a continuación ninguno de estos cinco modelos puede ser recomendado para el análisis clínico del problema, por lo que la Cátedra ha desarrollado otro enfoque usando los conceptos clínicos primero y la Estadística después.

16.6.1 Modelo de McNemar

El test de McNemar para analizar la significación entre dos métodos clínicos, se basa en el supuesto siguiente: Asumiendo que hay una población en la cual se usan dos métodos en cada individuo de la misma, para medir una magnitud dicotómica cuyo resultado se expresa como positivo (+) o negativo (-) con cada una de ellas, entonces habrá cuatro conjuntos mutuamente excluyentes y colectivamente exhaustivos, si se ubica a cada uno de los individuos en una de las cuatro celdas de la tabla de concordancia, y así se tiene una partición de la población. Luego si se eligen al azar un total de N individuos los datos obtenidos se pueden presentar como en la Tabla 16.3 anterior. El interés principal se centra en los valores discordantes, los cuales seguirán una distribución Binomial con un parámetro igual a 0,5. El test de *dos proporciones correlacionadas* usa el cuadrado del estadígrafo normal tipificado y puede ser analizado con la distribución de Chi-cuadrado. Aplicando la corrección de Yates por continuidad resulta:

$$z^2 = (|r_{12} - r_{21}| - 1)^2 / (r_{12} + r_{21})$$

La Ho es que no hay asociación entre ambos métodos y se la testea con el valor de tablas para 1 grado de libertad y un nivel de significación α : $\chi^2_{(\alpha;1)}$ Entonces cuando $z^2 > 3.841$ se puede rechazar la Ho con un 95% de confianza. Esto es considerado por algunos clínicos como una prueba de la existencia de concordancia entre ambos métodos, cosa que es bastante discutible como se verá más adelante.

Ejemplo 1) Un Bioquímico decide testear dos técnicas de laboratorio para diagnosticar cierta enfermedad, para elegir entre ambas. Los resultados se clasifican como positivo cuando se la detecta y negativo cuando el paciente aparenta estar sano. Se quiere descubrir si hay asociación entre el carácter positivo-negativo y el procedimiento de diagnóstico. Para ello escoge 100 pacientes al azar y les aplica ambas técnicas. Los resultados obtenidos se pueden presentar como:

Resultado	Método de medición		Total
	Técnica A	Técnica B	
(+)	70	52	122
(-)	30	48	78
Total	100	100	200

Lo primero a notar es que hay 100 pacientes en el experimento, pero el total de la tabla es 200. Lo que indica que habrá pérdida de independencia, pues los pacientes se ubicaron en más de una celda de la tabla y no se pueden aplicar los modelos vistos. Por eso McNemar propuso cambiar la unidad muestral.

Técnica B	Técnica A		Total
	(+)	(-)	
(+)	38	14	52
(-)	32	16	48
Total	70	30	100

Ahora los datos son los mismos pero ordenados de otra forma: ++ ; +- ; -+ y --. De esa manera, cada paciente aparece una sola vez en la tabla y se mantiene la independencia. McNemar propuso una fórmula con distribución Chi-cuadrado.

Se toman en consideración el par de valores donde los métodos difieren ($r_{12} = 32$ y $r_{21} = 14$) pero no se toman en cuenta los casos donde coinciden. El estadígrafo de McNemar sin la corrección de Yates es:

$$\chi^2 = (r_{12} - r_{21})^2 / (r_{12} + r_{21})$$

O sea, $\chi^2 = (14 - 32)^2 / (14 + 32) = 7.04^{**}$

Y con la corrección de Yates es $\chi^2_{\text{corr}} = (|r_{12} - r_{21}| - 1)^2 / (r_{12} + r_{21}) = 6,28^* > \chi^2_{0,95;1} = 3,841$

Donde se puede apreciar que la corrección vuelve más conservador a este test, es decir le va a costar más rechazar la H_0 . La conclusión estadística es que: hay evidencia significativa de asociación entre ambas técnicas, estas no son independientes. Desde el punto de vista clínico, a esto se lo interpreta como que existe *concordancia* entre ambas técnicas y entonces, la Técnica 1 puede reemplazar a la Técnica 2 y viceversa. Esta es la manera habitual de hacer el análisis de concordancia. Sin embargo, hay dos objeciones básicas que se le pueden hacer:

Objeción 1: No se toma en cuenta, ni el tamaño muestral total N, ni el número de concordancias halladas.

Objeción 2: Cuando la cantidad de discordancias de un tipo sea muy similar a la del otro tipo, el resultado siempre será no significativo, sin importar ni el tamaño de las mismas.

Por ejemplo, suponiendo ahora que los resultados obtenidos en 400 muestras sean:

Método nuevo	Método viejo		Total
	(+)	(-)	
(+)	180	22	202
(-)	10	188	198
Total	190	210	400

$\chi^2 = (22 - 10)^2 / (10 + 22) = 4,5^* > \chi^2_{0,95;1} = 3,841$
 Sin la corrección de Yates hay concordancia

$\chi^2 = (|22 - 10| - 1)^2 / (10 + 22) = 3,78 < 3,841$
 Con la corrección de Yates no hay concordancia

Luego, la conclusión es que no se puede reemplazar al método viejo con el nuevo si se toma en cuenta la corrección de Yates. Esto, desde el punto de vista estadístico. Sin embargo se puede ver que si se mantiene el número de discordancias en 32 y se aumenta el número de muestras a 400 millones (en vez de 400) las conclusiones serían las mismas. Pero desde el punto de vista clínico, no es lo mismo tener 32 discordancias en 400 muestras que en 400 millones. El sentido común indica que obtener 32 discordancias en 400 millones de casos, es prueba más que suficiente para poder reemplazar a un método con el otro. Y esta es la base de la primera objeción.

Por otra parte, con respecto a la segunda objeción, suponiendo que en el ejemplo de más arriba se hubiesen obtenido $r_{12} = r_{21} = 1$, el valor del estadígrafo sería prácticamente nulo y no se podría rechazar H_0 . El problema es que lo mismo ocurre si hubiese sido $r_{12} = r_{21} = 199$. Pero, no es lo mismo tener 2 discordancias que 398 discordancias en 400 casos desde el punto de vista clínico. El sentido común nos dice que en el primer caso la concordancia es casi perfecta, mientras que en el segundo casi no existe.

Ambas objeciones muestran que el Test de McNemar puede ser útil desde el punto de vista estadístico, pero *no puede ser recomendado* para ser usado en Clínica. Lo que ocurre es que el concepto de independencia estadística es en realidad lo que estudia McNemar. Y no tiene nada que ver con el concepto clínico de concordancia, porque que exista asociación estadística, no implica necesariamente que haya asociación clínica.

16.6.2 Modelo de Cochran

Este test puede ser aplicado para el caso de que haya $k = 2$ muestras. Notar que si se calcula el estadígrafo de Cochran $Q = [2 \cdot CS - T^2] / [2 T - RS]$ para el último ejemplo es:

$$T = (r_{12} + r_{21}) + 2 r_{11} = (22 + 10) + 2 \cdot 180 = 392$$

$$RS = (r_{12} + r_{21}) + 4 r_{11} = (22 + 10) + 4 \cdot 180 = 752$$

$$CS = (r_{11} + r_{21})^2 + (r_{11} + r_{12})^2 = (180 + 10)^2 + (180 + 22)^2 = 76.904$$

$$Q = [2 \cdot 76.904 - 153.664] / [2 \cdot 392 - 752] = 144 / 32 = 4,5^* > \chi^2_{0,95;1} = 3,841$$

Notar que el valor del estadígrafo Q es exactamente igual al valor de χ^2 de McNemar sin la corrección de Yates. Este resultado muestra que hay evidencia muy significativa para rechazar la hipótesis nula. La población *no tiene* proporciones iguales de positivos respecto a los dos procedimientos de diagnóstico. El bioquímico concluye que los métodos analizados, el nuevo y el viejo son concordantes. Es decir, que decide usar ambos métodos indistintamente.

Sin embargo las dos objeciones planteadas en el punto anterior, se pueden plantear nuevamente para este caso porque el valor de Q es igual al de McNemar, como puede demostrarse fácilmente haciendo pasaje de términos. Es decir:

$$Q = [2 \cdot CS - T^2] / [2 T - RS] = \chi^2 = (r_{12} - r_{21})^2 / (r_{12} + r_{21})$$

Nuevamente entonces se puede concluir que este test no puede ser recomendado para Clínica.

16.6.3 G-test de McNemar

El G-test para el caso de McNemar consiste en calcular el Likelihood para el estadígrafo basándose en la distribución Multinomial, este método también se llama *individuos doblemente testeados*. Hay dos probabilidades básicas: la observada y la esperada, entonces el logaritmo natural del cociente entre ambas es $G/2$, donde el estadígrafo G se distribuye aproximadamente como una Chi-cuadrado con un grado de libertad. Se calcula con:

$$G = 2 \{r_{12} \ln[2r_{12}/(r_{12} + r_{21})] + r_{21} \ln[2r_{21}/(r_{12} + r_{21})]\}$$

Y la corrección de Williams por continuidad es $q = 1 + (1 / 2 N)$, o sea $G' = G / q$. Para ilustrar este procedimiento se puede usar el ejemplo anterior:

Método B	Método A		Total
	(+)	(-)	
(+)	180	22	202
(-)	10	188	198
Total	190	210	400

$$G = 2 \{22 \cdot \ln[2 \cdot 22 / (22 + 10)] + 2 \cdot 10 \ln[2 \cdot 10 / (10 + 22)]\}$$

$$G = 4,551 \text{ y } q = 1 + [1 / (180 + 188)] = 1,00125 \text{ luego}$$

$$G' = 4,54^* > \chi^2_{0,95;1} = 3,841$$

En síntesis, analizando los datos con los tres tests presentados hasta ahora, las conclusiones estadísticas serían:

- Si se aplica el test de McNemar corregido con Yates: No hay asociación entre ambos métodos.
- Si se aplica el test de Cochran o el McNemar sin corregir: Hay asociación significativa.
- Si se aplica el G-test: Hay asociación significativa.

Se puede ver entonces que el G-test tiene mayor poder de discriminación que los restantes y es el más recomendable desde el punto de vista estadístico. El de Cochran es bastante similar y la diferencia entre ambos es pequeña y como es un test no paramétrico su campo de aplicación es más grande. El método clásico de McNemar pierde su poder al ser corregido con Yates. La conclusión estadística sería la siguiente: Se tiene evidencia significativa de la concordancia entre ambos métodos. Pero también al G-test se le pueden efectuar las dos objeciones básicas, pues el tamaño muestral N solo aparece en la corrección de Williams, cuyo peso relativo es muy pequeño en el resultado final. Notar que si N tiende a infinito, q tiene a uno y la corrección no existe.

Por lo tanto *la utilidad de estos tres tests desde el punto de vista clínico no es aceptable*. A ninguno de estos tres estadígrafos se lo puede considerar como un índice de asociación clínico, sino solamente de asociación estadística.

16.6.4 El test del log-Odds Ratio

Este test se aplica en realidad para las Tablas de la verdad vistas en el Gráfico 4.1 cuando se explicaron los índices clínicos. Pero si uno de los dos métodos puede ser considerado como el de referencia (por ejemplo la inmuno fluorescencia para la toxoplasmosis, etc.), entonces la Tabla de la verdad se convierte en una de contingencia para muestras apareadas. Así la Tabla 16.3 puede ser considerada como una Tabla de la Verdad y se puede aplicar este modelo estadístico. El logaritmo natural del OR llamado L, es simétrico alrededor del cero, cosa que no ocurre con OR. El cuadrado de L dividido por su desvío estándar SD (L) se distribuye como una Chi-cuadrado con un grado de libertad. La Ho : El valor esperado: $\Xi (L)$ es nulo, es lo mismo que decir: no hay significación estadística entre los métodos. Y el estadígrafo se calcula con:

$$L = \ln \left\{ \frac{(r_{11} + 0.5)(r_{22} + 0.5)}{(r_{12} + 0.5)(r_{21} + 0.5)} \right\} \quad y$$

$$SE^2(L) = \left[\frac{1}{(r_{11} + 0.5)} + \frac{1}{(r_{22} + 0.5)} + \frac{1}{(r_{12} + 0.5)} + \frac{1}{(r_{21} + 0.5)} \right] \quad \text{luego será}$$

$$\chi^2 = \frac{[L - \Xi(L)]^2}{SE^2(L)} = \frac{L^2}{SE^2(L)} \quad \text{para ser contrastado con } \chi^2_{(0.05; 1)} = 3.841$$

Aplicando este método al ejemplo del punto anterior resulta $\chi^2 = 164,1 \gg \chi^2_{(0.05; 1)}$ y hay una fuerte evidencia como para rechazar la Ho. En este modelo ninguna de las dos objeciones anteriores puede ser aplicada, pero desgraciadamente hay una nueva:

Objeción 3: Si los valores de ambas diagonales en la tabla son intercambiados, χ^2 no varía.

Esto significa que si ahora fuese: $r_{11} = 10$, $r_{22} = 22$, $r_{12} = 188$ y $r_{21} = 180$, el resultado sería de nuevo $\chi^2 = 164,1$. Estadísticamente significa lo mismo, pero clínicamente no es lo mismo encontrar 32 que 368 discordancia en 400 muestras. *Tampoco se puede recomendar este test.*

16.6.5 El test de Cohen-Kappa

En este test visto en el capítulo 14 no se puede hacer ninguna de las tres objeciones presentadas más arriba. Sería el único que resiste al sentido común. Sin embargo existen muchas objeciones respecto a su uso como:

- Kappa no es realmente una medida de concordancia corregida por el azar.
- Kappa es un índice de concordancia del tipo “ómnibus”, no distingue entre varios tipos de fuentes de desacuerdo. Para la tala de 2 x 2, no distingue entre un caso (+ -) de otro que sea (- +).
- Kappa resulta influenciado por la prevalencia y las categorías analizadas. Como resultado este índice es raramente comparable con procedimientos, poblaciones y estudios de corte. (Thompson y Walter, 1988, Feinstein y Cicchetti, 1990)
- Kappa puede resultar bajo aunque haya un alto nivel de concordancia, incluso cuando los valores individuales sean exactos. (Uebersax, 1988)
- Kappa requiere que dos métodos usen el mismo tipo de categorías, y a veces el interés está basado en analizar la consistencia entre diferentes categorías (por ejemplo, un método usa las categorías A, B y C, mientras que el otro usa A, B, C, D y E)
- Hay tablas que pretenden categorizar los rangos de Kappa como: “buenos”, “justos”, “pobres”, “malos” etc., las que no deben ser usadas por su arbitrariedad

El estadístico kappa para el análisis de concordancia se calcula con:

$$\text{Kappa} = \frac{\text{concordancia observada} - \text{concordancia esperada}}{1 - \text{concordancia esperada}}$$

Donde la concordancia observada es el nivel de concordancia = $(r_{11} + r_{22}) / N$

Y la concordancia esperada con: $[(r_{11} + r_{12})(r_{11} + r_{21}) / N] + [(r_{22} + r_{12})(r_{22} + r_{21}) / N]$

Aplicando este método al ejemplo anterior resulta kappa = 0,84, resultado que puede ser calificado como una concordancia casi perfecta, de acuerdo a la tabla de calificaciones especificada por Guyatt, G. en el *JAMA* (2002):

Calificación de la concordancia de acuerdo al valor de kappa

Valor de kappa	Calificación de la concordancia
0	Muy pobre
Entre 0 y 0,02	Pobre
Entre 0,2 y 0,4	Ligera
Entre 0,4 y 0,6	Moderada
Entre 0,6 y 0,8	Substancial
Entre 0,8 y 1,0	Casi perfecta

Feinstein y Cicchetti en 1990, presentaron dos paradojas básicas de kappa: 1) Alta concordancia pero bajo kappa. 2) Totales marginales desbalanceados producen valores más altos de kappa, que si estuvieran balanceados (como era de esperar). Por ello, la recomendación oficial a los clínicos es que utilicen el procedimiento de phi en lugar de kappa.

16.6.6 El procedimiento phi

Este método se conoce también como: concordancia chance-independiente, porque es independiente del nivel de concordancia, solo depende del Odds Ratio. Lo cual supone una ventaja al no depender tanto de la prevalencia en los totales marginales como kappa. Su valor se calcula con la relación siguiente:

$$\text{Phi} = \frac{(ad) - (bc)}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

Si se aplica al ejemplo anterior resulta $\text{phi} = 0,84$, el mismo valor que el obtenido por kappa. Sin embargo a pesar de las recomendaciones para los médicos, phi presenta las mismas desventajas de kappa y si se aplica a las paradojas de Feinstein tampoco las resuelve. Hay dos problemas:

Problema 1: Puede ocurrir que kappa y phi sean nulos, (o sea una gran discordancia), no importa en nivel de concordancia hallado en el estudio. Cuando $(a + b) = (c + d)$ entonces siempre resultará $\text{kappa} = \text{phi} = 0$. Por ejemplo, sean los dos casos siguientes:

Caso 1

Método B	Método A		Total
	(+)	(-)	
(+)	3600	60	3660
(-)	60	1	61
Total	3660	61	3721

Caso 2

Método B	Método A		Total
	(+)	(-)	
(+)	20	400	420
(-)	1	20	21
Total	21	420	441

En el Caso 1 se puede ver a simple vista la gran concordancia observada en el estudio, pues en 3720 casos se encontraron 120 discordancias. Esto implica un nivel de concordancia del 97%, y sin embargo el valor de kappa es nulo, lo mismo que el valor de phi.

En el Caso 2 se puede ver de un vistazo que no hay concordancia, pues se encontraron solo 40 concordancias en 441 casos, lo que implica un nivel de concordancia del 9%. Acá también los valores de kappa y phi son nulos, lo que estaría de acuerdo con los valores observados del nivel de concordancia. Pero el punto es que cuando se cumple la condición vista más arriba, entonces no importa la concordancia observada, kappa y phi se anularán y serán ciegos ante estos hechos clínicos. Esto muestra, que ninguno de ambos procedimientos se basa en concepto clínico de la concordancia sino en el estadístico.

Problema 2: Kappa y phi pueden ser negativos, no importa el tipo de concordancia hallada en el estudio. Cuando $a = 0$, o bien $d = 0$, los valores de kappa y phi serán negativos, no importa como son los demás valores hallados.

En las dos tablas siguientes se muestran dos situaciones bien opuestas, y sin embargo los valores de kappa y phi se parecen mucho. En el Caso 3 la concordancia es muy clara, y sin embargo el valor $\text{kappa} = \text{phi} = -0,01$. Mientras que en el Caso 4 la discordancia es muy clara y los valores calculados son: $\text{kappa} = -0,04$ y $\text{phi} = -0,13$, indicando la pobre concordancia hallada.

Caso 3

Método B	Método A		Total
	(+)	(-)	
(+)	390	5	395
(-)	5	0	5
Total	395	5	400

Caso 4

Método B	Método A		Total
	(+)	(-)	
(+)	0	260	260
(-)	10	250	260
Total	10	510	520

16.6.7 El método de visión dual

La concordancia sirve para ver si un método clínico puede ser reemplazado por otro. La idea no es ver cual de los dos es el mejor, sino simplemente si se pueden intercambiar. Para analizar el problema de la concordancia se propone una nueva idea: “Visión dual de la concordancia”. Se basa en un supuesto clínico principal:

Dos métodos clínicos concuerdan cuando tienen la misma sensibilidad y especificidad.

Esto se verifica, si y solo si $r_{11} = r_{22}$. En efecto, si se supone que uno de los dos métodos es el ideal, entonces la Tabla de concordancia se transforma en una Tabla diagnóstica y se pueden calcular la sensibilidad y especificidad del otro método. Ahora, haciendo la suposición inversa se pueden calcular el otro par de sensibilidad y especificidad. La única forma de que se hagan iguales es cuando $r_{11} = r_{22}$. Los tests estadísticos para verificar ese supuesto son el McNemar G-test con la corrección de Williams, el Q-test de Cochran, o el clásico Chi-cuadrado con la corrección de Yates. Pero esto solo tendría las objeciones vistas más arriba, por lo tanto se necesita estudiar además el punto de vista clínico usando el nivel de concordancia, o los Odds de discordancia.

El procedimiento de visión dual consiste de dos etapas:

1) *Visión estadística*: Lo primero es verificar si hay concordancia estadísticamente hablando. Para eso se puede usar cualquiera de los tests estadísticos mencionados. Sin embargo, la sugerencia es usar el G-test porque es el más potente de los tres como se vio antes. O bien, el Q-test cuando alguna de las frecuencias observadas de discordancia sea nula. La idea es obtener el valor de G corregido y compararlo con el valor crítico de la Chi-cuadrado para 95% de confianza. Cuando sea $G > 3,841$ entonces hay prueba estadística que muestra que los métodos no tienen la misma sensibilidad y especificidad. Si $G < 3,841$ no hay evidencia como para pensar que no se puedan intercambiar los métodos.

Pero esto como estos no es suficiente porque estos tests no tienen en cuenta ni el tamaño muestral, ni el nivel de concordancia, pueden aparecer paradojas clínicas como las vistas, y por lo eso se necesita de una segunda etapa.

2) *Visión clínica*: Lo segundo es ver si la concordancia encontrada es suficiente desde el punto de vista clínico. Para ello se necesita definir un criterio clínico de aceptación. Por ejemplo:

La concordancia es aceptable cuando la observada sea mayor o igual que un nivel de concordancia crítico ($\lambda_{crítico}$) definido por los médicos para cada enfermedad

La idea es encontrar el intervalo de confianza del 95% para el nivel de concordancia observado en el estudio, y ver si el valor crítico cumple con la condición de ser mayor o igual que el límite inferior del intervalo. Por ejemplo, si el criterio médico es: *La concordancia es aceptable cuando no hay más de 10 discordancias en 100 casos*. Esto significa que el nivel crítico de concordancia se establece en un 90%. Pero también se puede usar el otro índice: Odds de Discordancias, donde $DO = 10/90$, o sea DO es 1 : 9 Clínicamente, la concordancia será aceptable cuando no haya más de 1 discordancia, y 9 concordancias en 10 casos.

Casos posibles:

a) Si hay concordancia estadística en el primer paso se decide directamente con el segundo paso. Esto significa que todo el peso de la responsabilidad por el cambio recae en el médico, y no en estadística como es ahora.

b) Si no hay concordancia estadística en el primer paso significa que algo pasará con la sensibilidad y especificidad si se hace el intercambio de métodos. Y eso hay que estudiarlo con más cuidado (análisis más incisivo).

b-1) Ahora se hace el segundo paso: Si no se acepta por el criterio clínico adoptado, entonces significa que: estadísticamente y clínicamente hay prueba como para rechazar el intercambio.

b-2) Si en cambio, hay aceptación clínica uno se siente inclinado a aceptar el cambio, porque este criterio es el decisivo. El problema es que no hay en la literatura médica criterios de concordancia que sirvan de guía para todos los casos y para todas las enfermedades. Entonces el médico antes de aceptar tiene que hacer un estudio más cuidadoso:

Análisis más incisivo: Suponiendo que el Método 1 es el de referencia (o la verdad) entonces se pueden obtener la sensibilidad, especificidad e Índice de Youden del Método 2 (digamos S_2 , E_2 y Y_2). Viceversa, si suponemos al Método 2 como el de referencia se puede obtener lo mismo para el Método 1 (digamos S_1 , E_1 y Y_1). Potencialmente hablando, hacer el intercambio de métodos puede significar un cambio en estos índices, que se puede estimar con: $\Delta S = S_1 - S_2$ y análogamente: ΔE y ΔY .

Entonces la idea es ver si un cambio de digamos $\Delta S = 20\%$ es aceptable para el tipo de enfermedad que se está estudiando: Tipo I (el error más grave es un falso negativo, y hay que estudiar la variabilidad potencial de la sensibilidad). Tipo II (lo más grave es un falso positivo, y hay que estudiar la variabilidad potencial de la especificidad) y el Tipo III (los dos errores son igualmente graves, y hay que estudiar la variabilidad potencial del índice de Youden).

Solo el criterio clínico puede decidir en estos casos. Por ejemplo, hay enfermedades donde no sería aceptable un DO de 1 : 9 (90% de nivel de concordancia) como HIV, ciertos cánceres, etc.

Conclusión: Son necesarias las dos etapas y además el análisis más incisivo cuando el G-test te avisa de que hay un problema con los índices. Pero, siempre el criterio decisivo es el del médico. Estadística es solo para ayudar, pero no para decidir.

Todas las cuentas necesarias para hacer este nuevo procedimiento, se pueden hacer con un algoritmo médico desarrollado por la cátedra, y que se puede bajar libremente de su página web.

El factor común entre todos los modelos anteriores presentados es que, en alguna parte de su deducción aparece el concepto de independencia matemática, y luego del estadígrafo obtenido se deriva la conclusión clínica. La idea básica es que: *asociación estadística, no tiene por que ser asociación clínica*. Por todo esto esta cátedra propone un camino alternativo para resolver la cuestión: Como se muestra a continuación:

Problema: Analizar la concordancia entre tres nuevos métodos clínicos y uno viejo

Caso 1			Caso 2			Caso 3		
	Método 1			Método 1			Método 1	
Método 2	+	-	Método 3	+	-	Método 4	+	-
+	180	22	+	180	14	+	110	84
-	10	188	-	10	196	-	80	126
	190	210	190	210	400	190	210	400
	202	198		194	206		194	206
G = 4,54 *	λ 95% (89 ; 95)%		G = 0,6	λ 95% (92 ; 96)%		G = 0,1	λ 95% (54 ; 64)%	
λ = 92%	Kappa = 0,84		λ = 94%	Kappa = 0,88		λ = 59 %	Kappa = 0,18	
Phi = 0,84	ΔS = 5,6 %		Phi = 0,88	ΔS = 2 %		Phi = 0,18	ΔS = 1,2 %	
ΔE = 5,4 %	ΔY = 0,1 %		ΔE = 1,8 %	ΔY = 0,1 %		ΔE = 1,2 %	ΔY = 0 %	

En el Caso 1 hay diferencia significativa entre la sensibilidad y especificidad de acuerdo al G-test que es significativo. Cosa que no ocurre en los otros dos casos. La diferencia potencial de sensibilidad es del 5,6%, de especificidad del 5,4% y de Youden 0,1%. La concordancia observada es muy alta (92%). Por lo tanto, de acuerdo a la visión estadística se rechaza la concordancia y con el análisis más incisivo, se ve que la variabilidad potencial de los índices no es tan alta, por lo que se podría aceptar cambiar el Método 2 por el viejo Método 1 para ciertas enfermedades.

En el Caso 2 como G-test no es significativo, se usa directamente la visión clínica y se nota que el intervalo de confianza permite aceptar un valor crítico de hasta el 96%, porque el valor observado fue del 94%. Esto es suficiente para la mayoría de las enfermedades y no hay dudas que se puede reemplazar el Método 1 con el Método 3.

En el Caso 3, G no es significativo, pero la concordancia observada es muy baja (59%). Con la visión clínica se deduce que se puede aceptar un valor crítico de concordancia de hasta el 64%; lo cual no es aceptable para muchas enfermedades. Por lo tanto el Método 4 no es aceptable desde un punto de vista clínico.

La conclusión final de tipo cualitativo, es que la mejor opción para efectuar el reemplazo del método viejo es el Método 3. Para efectuar una conclusión de tipo cuantitativo, habría que comparar estadísticamente los cuatro casos, usando un modelo que todavía no se descubre.

16.7 Resolución de las paradojas de Feinstein

En 1990 fueron publicadas dos paradojas básicas para mostrar las falencias del procedimiento de kappa, y proponer una solución para las mismas, la cual no fue aceptada con el paso del tiempo. En este punto se usará el nuevo procedimiento de visión dual para hacerlo, y de paso mostrar que el procedimiento phi tampoco las resuelve. Por lo que no se puede aceptar a este procedimiento como la solución, por el contrario hay más fallas para el mismo:

Paradoja 1: Nivel de concordancia alto, con bajo kappa y phi

Caso A				$\lambda = 85\%$	Caso B				$\lambda = 85\%$
				kappa = 0,7					kappa = 0,32
				phi = 0,7					phi = 0,33
Método 2	Método 1				Método 2	Método 1			
	+	-				+	-		
+	40	9	49	G = 0,58	+	80	10	90	G = 1,64
-	6	45	51		-	5	5	10	
	46	54	100	λ 95% CI (78 ; 92)		85	15	100	λ 95% CI (78 ; 92)

En los dos casos anteriores A y B, el nivel de concordancia es el mismo: 85%. Sin embargo, comparando ambos casos resultan ser los valores de kappa y phi son muy diferentes entre sí. Esta disparidad inesperada ilustra las fallas de ambos procedimientos. Feinstein atribuye esta disparidad a la diferencia entre los totales marginales, porque mientras en el caso A están balanceados, no ocurre lo mismo en el Caso B. Ahora bien, con este argumento se muestra que no pueden ser aceptables desde un punto de vista clínico, porque no puede ser que la decisión del reemplazo de un método por otro, dependa de la prevalencia de la enfermedad que se encuentre.

Ambas paradojas se pueden resolver usando el procedimiento de visión dual. En ambos casos el G-test no es significativo, lo que indica que no habrá grandes diferencias de sensibilidades ni especificidades. O sea, desde el punto de vista estadístico no hay razones para rechazar la concordancia. Usando la visión clínica, se deduce que en ambos casos un nivel crítico del 92% sería aceptable. Y como para la mayoría de las enfermedades esto es suficiente, no hay dudas que se pueden intercambiar los métodos en ambos casos.

Paradoja 2: Totales marginales muy desbalanceados, producen un valor de kappa y phi más alto, que si estuvieran balanceados (como era de esperar de acuerdo a lo visto más arriba).

Los Casos C y D presentan el mismo nivel de concordancia observado: 60%. Los totales marginales del Método 2 son los mismos, pero en el Método 1 están invertidos. De acuerdo a la visión de kappa, uno esperaría encontrar mejor valor en el Caso C que en el D, pero esto no es así, sino todo lo contrario, lo que constituye una paradoja clínica para Feinstein.

Aplicando el procedimiento dual, la visión estadística muestra una diferencia entre ambos casos, mientras en el Caso C el G-test no es significativo, en el otro sí lo es. Por lo tanto, se puede aplicar la visión clínica directamente en el Caso C para observar que el nivel crítico máximo admisible es del 70%, lo cual no es suficiente para muchas enfermedades. La conclusión para este caso sería: no hacer el intercambio de métodos de acuerdo al criterio clínico.

Caso C				$\lambda = 60\%$	Caso D				$\lambda = 60\%$
				kappa = 0,13					kappa = 0,26
				phi = 0,13					phi = 0,31
Método 2	Método 1				Método 2	Método 1			
	+	-				+	-		
+	45	15	60	G = 2,5	+	25	35	60	G = 25***
-	25	15	40		-	5	35	40	
	70	30	100	λ 95% CI (50 ; 70)		30	70	100	λ 95% CI (50 ; 70)

En el caso D hay que hacer un análisis más incisivo para estudiar las variabilidades potenciales de los tres índices principales de acuerdo a cada tipo de enfermedad. Se encuentra que $\Delta S = 42\%$, $\Delta E = 38\%$, $\Delta Y = 4\%$ lo que sería inaceptable para las enfermedades del Tipo I y II, para las del

Tipo III podría ser aceptable en algunos casos, sin embargo usando la visión clínica en tal caso, se concluye lo mismo que anteriormente y el intercambio de métodos no puede ser aceptado. O sea, la paradoja deja de ser tal para este procedimiento, en cambio comparando kappa con phi, sigue existiendo. Esta segunda paradoja se refleja de nuevo en los dos casos siguientes:

Caso E				$\lambda = 90\%$	Caso F				$\lambda = 90\%$
				$\text{kappa} = 0,44$					$\text{kappa} = 0,74$
				$\text{phi} = 0,44$					$\text{phi} = 1$
	Método 2	Método 1				Método 1			
		+	-			+	-		
+		85	5	90	$G = 0$	70	10	80	$Q = 10^{***}$
-		5	5	10		0	20	20	
		90	10	100	λ 95% CI (84 ; 96)	70	30	100	λ 95% CI (84 ; 96)

Aquí la paradoja es que a pesar de que los totales marginales están “simétricamente” desbalanceados en la misma dirección (similar prevalencia), los valores de kappa y phi son diferentes y en el Caso E no reflejan la alta concordancia observada (90%). Esto se resuelve usando el procedimiento de visión dual.

En el Caso E, se puede ver la concordancia a simple vista. En efecto, al ser iguales las frecuencias de discordancia se deduce que $G = 0$ y por lo tanto desde un punto de vista estadístico, no hay diferencias. Usando la visión clínica, se detecta un valor muy alto de nivel de concordancia máximo admisible (96%). Por lo tanto, no hay dudas que el intercambio de métodos puede realizarse. Pero esto no ocurre en el Caso F, a pesar de la alta concordancia observada, porque ahora la visión estadística muestra una variabilidad significativa entre la sensibilidad y especificidad de ambos métodos. Acá no se puede aplicar el G-test porque una frecuencia es nula, y hay que usar el Q-test que resulta altamente significativo. Entonces, realizando el análisis más incisivo se ve que las variabilidades potenciales son: $\Delta S = 13\%$, $\Delta E = 33\%$, $\Delta Y = 21\%$ lo que sería inaceptable para mayoría de las enfermedades.

16.8 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|--|----------|----------|
| 1) Bondad de ajuste significa que tan bien se parecen los datos reales a los teóricos. | V | F |
| 2) El método clásico para efectuar una prueba de bondad de ajuste es | | |
| 3) Se acostumbra a agrupar las clases con sus anexas cuando la frecuencia observada es < 3 . | V | F |
| 4) Para el cálculo de los grados de libertad se usa el número de clases agrupadas. | V | F |
| 5) El G-test comparado con el clásico es mejor porque es | | |
| 6) Cuando la frecuencia esperada en un G-test es menor que 5 hay que agrupar las clases. | V | F |
| 7) El método de Kolgomorov-Smirnov (KS-test) busca la diferencia máxima de frecuencias. | V | F |
| 8) El KS-test se puede usar para una sola muestra o para comparar dos entre sí. | V | F |
| 9) Los pasos a seguir en un KS-test para una sola muestra son | | |
| 10) Da lo mismo usar hipótesis extrínsecas que intrínseca en un KS-test. | V | F |
| 11) El KS-test para dos muestras se puede usar si estas son independientes. | V | F |
| 12) Las tablas estadísticas para el KS-test de una y dos muestras son iguales. | V | F |
| 13) Los pasos a seguir para efectuar un KS-test en dos muestras son..... | | |
| 14) El estadígrafo de comparación en el KS-test es el mismo. | V | F |
| 15) Conviene ordenar los datos en forma creciente antes de usar el KS-test para 2 muestras. | V | F |
| 16) Un test de bondad de ajuste con repetición solo permite comparar los datos agrupados. | V | F |
| 17) Las pruebas que se pueden plantear en un test con repetición son | | |

- 18) La variabilidad total en un test de bondad de ajuste se cuantifica con.....
 19) Heterogeneidad de los datos repetidos implica encontrar significación en el G-test . **V F**
 20) El G de datos agrupados, más el G de datos desagrupados es igual al G total. **V F**
 21) Explicar las fallas que presentan los tests de McNemar, Log(OR), kappa y phi:
 22) Explicar el método de visión dual.
 23) La concordancia es un concepto estadístico. **V F**
 24) Cuando una de las frecuencias de discordancia es nula, hay que usar el Q-test . **V F**
 25) Ídem anterior, pero ocurre que OR es infinito y phi tiende a uno. **V F**
 26) En la concordancia, la visión clínica es más importante que la visión estadística. **V F**
 27) Explicar y resolver las paradojas de Feinstein.
 28) Las mismas fallas de kappa, ocurren en phi pero más ligeramente. **V F**

2) Aplicar el KS-test a los tres problemas vistos en el punto 16.1 y comparar las conclusiones obtenidas con las del texto usando el G-test.

3) Ídem anterior para los problemas vistos en capítulos anteriores.

4) Se efectuaron 1000 pruebas de efectividad de un nuevo medicamento, en 5 grupos de 200 individuos enfermos cada uno, elegidos al azar con el método de doble ciego. Los resultados se clasificaron en SI cuando fueron efectivos, y NO en caso contrario. Con esta información decidir si el medicamento es efectivo.

Caso	SI	NO	Total
1	120	80	200
2	110	90	200
3	98	102	200
4	130	70	200
5	125	75	200
Total	583	417	1000

5) Para los problemas siguientes se pide analizar la concordancia entre los métodos clínicos para ver si el Método 1 (el viejo) puede ser reemplazado por alguno de los nuevos:

Caso 1

		Método 1		
Método 2		+	-	
+	285	25	310	
-	15	275	290	
	300	300	600	

Caso 2

		Método 1		
Método 3		+	-	
+	166	14	180	
-	10	410	420	
	176	424	600	

Caso 3

		Método 1		
Método 4		+	-	
+	390	10	400	
-	30	170	200	
	420	180	600	

Realizar el análisis con los cinco tests estadísticos tradicionales y discutir los resultados desde un punto de vista estadístico y desde un punto de vista clínico.

6) Aplicar el método de visión dual en el problema anterior.

Apéndice 1: Método de Quintin McNemar para analizar la concordancia

Cuando los clínicos deben evaluar una magnitud usando un criterio propio, a menudo no concuerdan en sus observaciones. Esto es palpable cuando se trata de medir magnitudes organolépticas tales como olor, turbidez, etc. Entre 1947 y 1955 McNemar introdujo una manera estocástica de analizar la concordancia entre dos observadores en estudios psicológicos, y a partir de allí se fue incorporando a Medicina y otras ciencias. Siguiendo su nomenclatura original, el razonamiento seguido fue:

Tabla de valores observados

		Observador 1		
		+	-	
Observador 2	+	a	b	a + b
	-	c	d	c + d
		a + c	b + d	N

a y d son las concordancias entre ambos
 c y b son las discordancias entre los dos
 N es el número total de casos analizados

La hipótesis nula planteada es que el número de los dos tipos de discordancias debe ser el mismo, para que haya una concordancia aceptable entre ambos observadores. Brevemente:

$$H_0) E(b) = E(c)$$

Tomando como mejor estimación del total de discordancias: $E(b + c) = b + c$, entonces los valores esperados en las celdas respectivas serán:

$$E(b) = E(c) = (b + c) / 2$$

Para examinar estadísticamente los resultados obtenidos, se puede usar la distribución Chi-cuadrado entre valores observados y esperados de la manera siguiente:

$$\chi^2 = \frac{\{b - [(b + c)/2]\}^2 + \{c - [(b + c)/2]\}^2}{[(b + c) / 2]}$$

Resolviendo esta ecuación resulta:

$$\chi^2 = (b - c)^2 / (b + c)$$

La cual puede ser interpretada como una distribución Chi-cuadrado con un grado de libertad. A su vez si se efectúa la corrección de Yates por continuidad resulta:

$$\chi^2 = (|b - c| - 1)^2 / (b + c) \text{ la cual se puede expresar como se vio más arriba con}$$

$$z^2 = (|r_{12} - r_{21}| - 1)^2 / (r_{12} + r_{21})$$

La recomendación es usar la corrección por continuidad cuando $(b + c) < 10$. En los demás casos puede ser obviada porque no contribuye grandemente a la aproximación.

Apéndice 2: Propuesta de Feinstein y Cicchetti para resolver las paradojas de kappa

En 1990 aparecieron dos trabajos de estos autores, en el primero presentan dos paradojas que muestran fallas en el método de Kappa, y en el segundo una forma de solucionarlas. Todo esto para los casos binarios.

La idea es obtener un índice general de concordancia para los casos positivos y otro para los casos de negativos. Por ejemplo, si el total de acuerdo en positivos para el primer observador resulta ser $r_{11} + r_{12} = n_1$ y para el segundo observador es $r_{11} + r_{21} = f_1$, entonces los valores esperados para la concordancia en positivos será:

$$\Xi (+) = (n_1 + f_1) / 2 \quad \text{Análogamente para el caso de los negativos.}$$

$$\Xi (-) = (n_2 + f_2) / 2 \quad \text{Donde } r_{22} + r_{12} = f_2 \quad r_{22} + r_{21} = n_2$$

Tabla de Concordancia

		Test 1		
Test 2		(+)	(-)	Total
(+)		r_{11}	r_{12}	$r_{11} + r_{12} = n_1$
(-)		r_{21}	r_{22}	$r_{21} + r_{22} = n_2$
Total		$r_{11} + r_{21} = f_1$	$r_{12} + r_{22} = f_2$	N

La proporción de positivos se puede calcular teniendo en cuenta que r_{11} es el valor observado:

$$P_{\text{pos}} = r_{11} / \Xi (+) = 2 r_{11} / [(n_1 + f_1)]$$

$$P_{\text{neg}} = r_{22} / \Xi (-) = 2 r_{22} / [(n_2 + f_2)]$$

El índice kappa debería ser acompañado siempre por los valores de estas dos proporciones en los informes, para tener una indicación más clara de lo que está pasando. Notar que los valores esperados se calculan dentro de las fórmulas de las proporciones.

17

Análisis de Varianza

En este primer capítulo dedicado al tema del Análisis de Varianza, se verán los conceptos fundamentales, supuestos básicos y forma de cálculo del modelo más elemental, llamado *unifactorial*. En los capítulos subsiguientes se verán modelos *bifactoriales* y los temas relacionados con los mismos. Paralelamente se mostraran modelos no paramétricos, para los casos donde uno o más de los supuestos básicos no se verifiquen. En adelante, se emplearán el nombre que le puso J.W. Tukey al tema: ANOVA, una manera resumida y practica de referirse a la materia en cuestión y de uso muy difundido en la literatura actual.

Los modelos de ANOVA son el capítulo fundamental de la Estadística porque son algo más que unas técnicas aplicadas. Son la herramienta fundamental para adentrarse en la naturaleza de la variación de los acontecimientos; permiten discernir mejor las causas de los fenómenos y los efectos de los factores involucrados. Es una herramienta indispensable para el bioquímico y farmacéutico moderno. No solo lo introduce en la misma Naturaleza de las cosas, sino que es la herramienta básica para el diseño de experimentos. Toda vez que necesite buscar las causas que hayan descontrolado sus técnicas de laboratorio o de producción, podrá usar la filosofía de estos modelos para realizar su investigación.

El ANOVA puede ser considerado como una manera de verificar si dos o más medias muestrales fueron extraídas de una misma población o de poblaciones con el mismo valor esperado, para una magnitud clínica dada. En consecuencia, cuando estas medias muestrales no sean coincidentes habrá que suponer que provienen de poblaciones diferentes por el efecto causado por un factor en estudio. Como por ejemplo, comparar las medias muestrales de un placebo *versus* las medias de muestras con diferentes dosis de un medicamento. O bien, la comparación entre sí de varias marcas comerciales, proveedoras de drogas o kits de medición, como además comparar varios operadores, o equipos, o pipetas entre sí, etc.

Cuando se trabaja con dos muestras se usa el modelo de Student para muestras independientes como se vio en los temas anteriores. El ANOVA es un método más general, que se extiende a más de dos muestras y se puede demostrar que coincide con Student si se aplica a solo dos muestras. Es decir, el modelo de Student es un caso particular del ANOVA.

Estos modelos desarrollados por R. A. Fisher a principios de este siglo, tienen una distribución teórica esperada: la función F, tabulada por G. W. Snedecor, sita en la Tabla 7 del Anexo.

Es decir, que los tests estadísticos a realizar se basan en comparar el valor muestral calculado con los datos medidos: F contra un valor crítico de tablas F_{α} . La idea básica del método es que si las muestras son *normales, independientes y aleatorias*; y se supone que todas tienen la misma varianza (*homocedásticas*), entonces, para que provengan de una misma población se necesita únicamente que las medias muestrales sean todas iguales. Esta será la hipótesis nula H_0 que se usará en todos los modelos de ANOVA, junto con los cuatro supuestos mencionados.

17.1 Introducción

Para comenzar, conviene unificar la terminología que se usará en todos los modelos de ANOVA que se irán desarrollando. En el Cuadro 17.1 siguiente, se presenta la manera de volcar los datos recopilados. Para simplificar los cálculos se supone que se ha seguido la regla de oro: *Todas las muestras tienen el mismo tamaño*. El lector interesado en desarrollar las ecuaciones para muestras de distinto tamaño, puede recurrir a los textos, como el de Sokal, para analizar los cambios. Pero en el presente trabajo, se intenta reducir al mínimo la terminología matemática para facilitar la comprensión del lector no avezado.

Cuadro 17.1 Presentación de datos en una ANOVA simple

N°	Factor A						
	Grupo 1	Grupo 2	Grupo 3	...	Grupo j	...	Grupo a
1	X11	X12	X13	...	X1j	...	X1a
2	X21	X22	X23	...	X2j	...	X2a
3	X31	X32	X33	...	X3j	...	X3a
.
i	Xi1	Xi2	Xi3	...	Xij	...	Xia
.
n	Xn1	Xn2	Xn3	...	Xnj	...	Xna
Total	T.1	T.2	T.3	...	T.j	...	T.a

Este modelo es de tipo unifactorial, es decir se emplea para estudiar el efecto de un cierto Factor A sobre las muestras tomadas. A cada muestra se la denomina grupo. Hay un total de **a** grupos. En cada grupo se han realizado **n** determinaciones de la magnitud clínica **X**. Además,

X_{ij} : es la determinación número *i* del grupo *j*

T_j : es la suma de las *n* observaciones correspondientes al grupo *j*

$\bar{X}_j = T_j / n$: es el valor promedio del grupo *j*.

$T = T.1 + T.2 + T.3 + \dots + T.j + \dots + T.a$: es el total de todas las observaciones X_{ij} hechas.

$N = n \cdot a$: es el número total de datos medidos.

$$\bar{X} = T/N = \sum_1^a \bar{X}_j / a = \sum_1^n \sum_1^a X_{ij} / N : \text{ es el promedio general de todos los datos.}$$

Para los totales, se usa un punto para denotar que se ha sumado a lo largo del subíndice i. Por ejemplo, T.2 es el total de los datos de la muestra número 2, o sea: T.2 = X12 + X22 + ... + Xn2. En estas condiciones se puede calcular la varianza total con :

$$DS^2 = SS_T / v_T = \sum_1^n \sum_1^a (X_{ij} - \bar{X})^2 / (N - 1)$$

La varianza total de los datos es el cociente entre la suma de los cuadrados totales SS_T y los grados de libertad totales $v_T = N - 1$. Mientras que la suma de cuadrados totales SS_T , es la sumatoria para N datos, del cuadrado de las diferencias, entre cada valor y su promedio general.

El motivo para escribir la varianza como un cociente entre la suma de los cuadrados y los grados de libertad es que ambos términos pueden ser divididos en dos partes. Una de esas partes se explica por el efecto del factor analizado y la otra parte es la inexplicada, debida al error aleatorio de las mediciones. En efecto :

$$v_T = N - 1 = N - a + a - 1 = (N - a) + (a - 1) = v_D + v_E$$

$v_E = a - 1$: Son los grados de libertad *entre* las muestras.

$v_D = N - a$: Son los grados de libertad *dentro* de las muestras.

Por su parte se puede describir la suma de cuadrados como:

$$SS_T = \sum \sum (X_{ij} - \bar{X})^2 = \sum \sum [(X_{ij} - \bar{X}_j) + (\bar{X}_j - \bar{X})]^2$$

Efectuando el desarrollo del cuadrado y simplificando queda :

$$SS_T = \sum \sum (X_{ij} - \bar{X}_j)^2 + \sum \sum (\bar{X}_j - \bar{X})^2$$

$$SS_T = SS_D + SS_E$$

La suma de cuadrados totales, se ha particionado en dos términos, la suma de cuadrados dentro de las muestras SS_D , calculada como la sumatoria del cuadrado de las diferencias entre cada observación y el promedio del grupo al cual pertenece y la suma de cuadrados entre las muestras SS_E , calculada como la sumatoria del cuadrado de las diferencias entre los promedios de cada grupo y el promedio general o media muestral.

Como se puede ver la SS_D , tiene un total de N términos cuadráticos libres, pero hay una cantidad de relaciones en el calculo de las medias grupales que le restan a grados de libertad, por lo tanto los grados de libertad de esa suma de cuadrados será $v_D = (N - a)$. Por su parte, como hay a medias grupales, la SS_E tendrá a términos cuadráticos libres, a los que se le debe restar un término, por la relación de vinculo entre ellos con la media general. O sea, $v_E = (a - 1)$.

Si se divide las sumas de los cuadrados por sus grados de libertad respectivos, se obtiene una nueva cantidad con todas las características de una varianza, denominada cuadrados medios, que se denota: **MS** (del término inglés: **Mean of Squares**).

Luego, los nuevos estadígrafos hallados son :

$MS_E = SS_E / v_E$: Son los cuadrados medios *entre* las muestras.

$MS_D = SS_D / v_D$: Son los cuadrados medios *dentro* de las muestras.

Notar que, la suma de ambos términos, **no es igual** al cuadrado medio total. Los que se particionan en dos partes son las **SS** y los grados de libertad, pero no los **MS**. Y allí reside el principal hecho que permite el estudio con los modelos de ANOVA; una medida cuantitativa de la variabilidad de las mediciones se divide en dos partes: Una atribuible al efecto del factor que se desea investigar y la otra, es el remanente, inexplicada, atribuida al error de medición.

R. A. Fisher demostró que el cociente entre estos dos estadígrafos tiene una distribución F y por lo tanto se puede plantear un test de hipótesis de la manera siguiente:

Se calcula: $F = MS_E / MS_D$ y se lo compara con un valor crítico de tablas $F_{\alpha; (a-1); (N-a)}$

Cuando $F > F_{\alpha}$ se rechaza la H_0 planteada, de que todas las muestras provienen de la misma población. En la forma habitual vista en los capítulos anteriores para testear la hipótesis nula, de que todas las medias son iguales. Si se rechaza la hipótesis nula, entonces se debe continuar para averiguar en cuál, o cuales, muestras hay diferencias significativas. Esto se verá en el Tema 18.

17.2 Formas cortas de cálculo

Las ecuaciones anteriores parecen complicadas, sin embargo hay maneras de simplificar los cálculos para obtener el estadígrafo principal F, como se muestra a continuación:

$$SS_T = \sum \sum (X_{ij} - \bar{X})^2 = \sum \sum (X_{ij} - (T/N))^2 = \sum \sum (X_{ij})^2 - (T^2 / N)$$

Así, los cuadrados medios totales, se pueden calcular como la sumatoria de los cuadrados de todas las observaciones realizadas, menos el *término de corrección*: T^2 / N , que se calcula como el cuadrado del total de las observaciones, dividido el tamaño muestral del experimento. Por su parte, se puede simplificar el:

$$SS_E = \sum \sum (\bar{X}_{.j} - \bar{X})^2 = \sum \sum [a \cdot T_{.j} - (T/N)]^2 = (1/n) \sum_1^a (T_{.j})^2 - (T^2 / N)$$

Así, los cuadrados medios entre las muestras, se puede calcular como la sumatoria de los cuadrados de los totales muestrales dividido su tamaño muestral, menos el término de corrección. Finalmente, la suma de cuadrados entre las muestras se obtiene como la diferencia de los dos términos anteriores.

Entonces, ya se está en condiciones de armar el denominado: “Cuadro de ANOVA”, volcando en el mismo los datos principales:

Cuadro 17.3 : Cuadro de ANOVA.

Variabilidad	Suma de cuadrados	Grados de Libertad	Cuadrados medios	$F = MS_E / MS_D$
Entre los Grupos	SS_E	v_E	MS_E	
Dentro de los grupos	SS_D	v_D	MS_D	
Total	SS_T	v_T		

Los diez pasos a seguir para la simplificación de los cálculos usando las relaciones anteriores son las siguientes:

Paso 1) Se calcula el gran total T como la suma de todas las observaciones.

Paso 2) Se obtiene el llamado término de corrección T^2 / N

Paso 3) Se calcula la suma de los cuadrados de todas las observaciones realizadas:

$$T_x^2 = \sum_1^n \sum_1^a (X_{ij})^2$$

Paso 4) Se calcula la suma de los cuadrados de los totales grupales T.j y se divide por n:

$$T_G^2 = (1/n) \sum_1^a (T.j)^2$$

Paso 5) Se obtiene el valor de SS_E restando el término de corrección al valor calculado en el caso anterior. O sea: $SS_E = Paso 4 - Paso 2$.

Paso 6) Se obtiene el valor de SS_T restando a la cantidad obtenida en el Paso 3 el término de corrección. O sea: $SS_T = Paso 3 - Paso 2$.

Paso 7) Se calcula el valor de $SS_D = SS_T - SS_E$.
 O sea: $SS_D = Paso 6 - Paso 5$.

Paso 8) Se calculan los grados de libertad respectivos y se completa el cuadro de ANOVA.

Paso 9) Se calcula el estadígrafo $F = MS_E / MS_D$

Paso 10) Se obtienen de las tablas los valores críticos para los niveles de significación buscados (95%, 99% y 99,9%) y se decide si se acepta o rechaza la hipótesis nula, como es usual.

En estos diez pasos se resume el método de cálculo de F, a cuentas sencillas y rápidas de efectuar. Sin embargo, existen programas de computadoras específicos para resolver un ANOVA, tales como el Statistics, Stat-Graphics, Mat-Acad, etc. Por su parte, una planilla de cálculo cualquiera como Lotus o Excel, permiten programar los pasos anteriores a fin de simplificar el trabajo de cálculo.

17.3 Aplicación en Control de Calidad

Para ilustrar los conceptos anteriores se presentan ejemplos de aplicación, relacionados con el tema del Control de Calidad. La idea es practicar las fórmulas anteriores y comparar con las técnicas relacionadas presentadas en los capítulos precedentes.

Ejemplo 1) En un sistema Regional de Control de Calidad, con 15 laboratorios afiliados se quiere investigar las fluctuaciones entre 3 diferentes maneras de medir RGR (Recuento de Glóbulos Rojos). La primera forma es usando equipos automatizados de recuento, o contadores hematológicos, tales como el Technicon H301 y similares. De entre todos los afiliados que usen ese método se eligen cinco de ellos al azar para conformar el Grupo 1. La segunda manera es usando el método microhematocrito y para ello, se eligen al azar, otros cinco afiliados que usan tal método para conformar el Grupo 2. Finalmente, el Grupo 3 se conforma con otros cinco laboratorios seleccionados al azar, de entre los que usan otros métodos, como por ejemplo el macrohematocrito, recuento en cámara, etc. Los 15 laboratorios siguen un programa de Control de Calidad interno y se suponen calibrados. Se envía a cada laboratorio una muestra ciega, con una sangre calibrada en el laboratorio de referencia de: $\mu \in (2,9467 \pm 0,0004) \cdot 10^6$ gl/ml.

Los valores observados se muestran en el cuadro 19.3. Se necesita sacar algunas conclusiones de los datos obtenidos.

Cuadro 17.4 : Valores de RGR expresados en 10^6 gl/ml

N°	Grupo1	Grupo2	Grupo3
1	2,94	3,01	3,30
2	2,96	3,10	3,20
3	2,95	3,04	3,20
4	2,96	3,06	3,10
5	2,94	3,04	3,30
Total	14,75	15,25	16,10
Media	2,95	3,05	3,22
$\sum x^2$	43,5129	46,5169	51,87
DS	0,01	0,03317	0,08367

Para usar este modelo, se comienza con los diez pasos descriptos más arriba hasta conseguir el resultado buscado:

Paso 1) Se calcula el gran total $T = 14,75 + 15,25 + 16,10 = 46,1$

Paso 2) Se obtiene el llamado término de corrección $T^2 / N = (46,1)^2 / 15 = 141,68066$

Notar que *no se ha redondeado* el resultado. Como se verá a continuación, si se realiza una cosa así, se pierden las sutiles diferencias con la que trabaja el modelo.

Paso 3) Se calcula la suma de los cuadrados de todas las observaciones realizadas. O sea:

$$T_x^2 = \sum_1^n \sum_1^a (X_{ij})^2 = (2,94)^2 + (2,96)^2 + \dots + (3,30)^2 = 141,8998$$

Paso 4) Se calcula la suma de los cuadrados de los totales grupales T_j y se divide por n :

$$T_G^2 = \sum_1^a (T_j)^2 / n = (1/5) [(14,75)^2 + (15,25)^2 + (16,1)^2] = 141,867$$

Paso 5) Se obtiene el valor de SS_E restando a la cantidad obtenida en el *Paso 4* el término de corrección.

$$SS_E = 141,867 - 141,68066 = 0,18634$$

Paso 6) Se obtiene el valor de SS_T restando a la cantidad obtenida en el *Paso 3* el término de corrección.

$$SS_T = 141,8998 - 141,68066 = 0,21914$$

Paso 7) Se calcula el valor de $SS_D = SS_T - SS_E$.

$$SS_D = 0,21914 - 0,18634 = 0,0328$$

Paso 8) Se calculan los grados de libertad respectivos y se completa el cuadro de ANOVA

$$v_T = N - 1 = 14 ; v_D = (N - a) = 12 ; v_E = (a - 1) = 2$$

Variabilidad	SS	v	MS
Entre grupos	0,18634	2	0,09317
Dentro grupos	0,0328	12	0,0027333
Total	0,21914	14	

$$F = 34,1***$$

Paso 9) Se calcula el estadígrafo de comparación con $F = MS_E / MS_D = 34,1***$

Paso 10) Se obtiene de tablas los valores críticos $F_{0,95;2;12} = 3,89$ $F_{0,999;2;12} = 12,967$

La conclusión es que debe rechazarse la hipótesis nula. La muestra no proviene de la misma población. Para los enunciados se concluye que las tres técnicas de laboratorios arrojan valores diferentes. Pero para poder comparar entre sí, los grupos testeados se necesitan más conceptos del ANOVA, que permiten avanzar un poco más en la investigación. Con este análisis, solo se sabe que hay una diferencia entre los tres grupos testeados, pero no se puede determinar entre quienes.

Como a estas alturas no se dispone de los modelos adecuados para poder efectuar las comparaciones entre grupos, que se verán en el próximo capítulo, todo lo que se puede aplicar es lo visto en pequeñas muestras. Esto es, se puede proceder a realizar comparaciones de dos muestras entre sí, con el modelo de Student para muestras independientes. Entonces, habrá tres comparaciones posibles: Grupo1 vs. Grupo 2; Grupo 2 vs. Grupo 3 y Grupo 1 vs. Grupo3.

Cuadro 17.5 : Comparaciones de muestras entre sí.

Caso 1) Comparación del Grupo 1 versus el Grupo 2:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{DS_1^2}{n_1}\right) + \left(\frac{DS_2^2}{n_2}\right)}} = \frac{(2,95 - 3,05) - 0}{\sqrt{\left(\frac{(0,01)^2}{5}\right) + \left(\frac{(0,0332)^2}{5}\right)}} = -6,45^{***}$$

Pues $t_{\alpha;v} = t_{0,999;8} = 5,041$: La conclusión es que ambas técnicas arrojan diferentes resultados y sus diferencias son altamente significativas.

Caso 2) Comparación del Grupo 3 versus el Grupo 2:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{DS_1^2}{n_1}\right) + \left(\frac{DS_2^2}{n_2}\right)}} = \frac{(3,22 - 3,05) - 0}{\sqrt{\left(\frac{(0,08367)^2}{5}\right) + \left(\frac{(0,0332)^2}{5}\right)}} = +4,25^{**}$$

Pues $t_{\alpha;v} = t_{0,99;8} = 3,36$: La conclusión es que ambas técnicas arrojan diferentes resultados y sus diferencias son muy significativas.

Caso 3) Comparación del Grupo 1 versus el Grupo 3:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{DS_1^2}{n_1}\right) + \left(\frac{DS_2^2}{n_2}\right)}} = \frac{(2,95 - 3,22) - 0}{\sqrt{\left(\frac{(0,01)^2}{5}\right) + \left(\frac{(0,08367)^2}{5}\right)}} = -7,16^{***}$$

Pues $t_{\alpha;v} = t_{0,999;8} = 5,041$: La conclusión es que ambas técnicas arrojan diferentes resultados y sus diferencias son altamente significativas.

Conclusión: Los 3 grupos difieren entre sí.

17.3.1 Control de exactitud

Para controlar la exactitud de una técnica clínica o industrial, se necesita de un patrón. Con este valor se pueden comparar los valores promedios de cada grupo analizado, para ver si cada uno está calibrado. El procedimiento es sencillo y se puede realizar de dos maneras. La primera es realizar un test de hipótesis con el modelo Student. La segunda es determinar el intervalo de confianza de cada valor promedio encontrado, para cada grupo, y ver si el valor patrón cae dentro o fuera del mismo. Cuando no caiga dentro hay evidencia como para pensar que el sistema está descalibrado, hay un error de tipo sistemático que también se puede calcular.

Ejemplo 2) Con los datos del ejemplo anterior determinar en forma analítica el control de exactitud que se le puede realizar a los tres grupos investigados, con un test y con un intervalo.

Método 1) Calibración con el modelo Student

H_0 : El método está calibrado : $\mu = \bar{X}$

H_1 : El método no está calibrado: $\mu \neq \bar{X}$

$$t = \frac{(\bar{x} - \mu)}{DS/\sqrt{n}} \text{ versus } t_{\alpha; v}$$

Grupo 1: Contador Hematológico.

$$t = \frac{(\bar{x} - \mu)}{DS/\sqrt{n}} = \frac{(2,95 - 2,9467)}{0,01/\sqrt{5}} = 0,738 \text{ versus } t_{\alpha; v} = t_{0,95; 4} = 2,776$$

No hay evidencia como para rechazar la H_0 . Como el resultado no fue significativo, se acepta la hipótesis nula que el Grupo 1 está calibrado, con respecto al laboratorio de referencia. A su vez, se puede proceder con cada dato individual. El programa regional, le envía esta información a cada uno de ellos, y le suele agregar un ranking. Así, el laboratorio que envió el dato número 3 resultara el primero de los 15, pues es el más cercano al valor control; mientras los que enviaron los datos número 1 y 5 salen en el segundo puesto y así sucesivamente. Para el investigador significa que su técnica de rutina debería ser la que utiliza el contador hematológico, por ser la más exacta de todas y estar calibrada. Esto se deduce, comparando las medias muestrales halladas.

Grupo 2: Microhematocrito.

$$t = \frac{(\bar{x} - \mu)}{DS/\sqrt{n}} = \frac{(3,05 - 2,9467)}{0,0332/\sqrt{5}} = 6,96^{**} \text{ versus } t_{\alpha; v} = t_{0,99; 4} = 4,604 \text{ y } t_{0,999; 4} = 8,610$$

Hay evidencia como para rechazar la H_0 al (99% de confianza). Como el resultado fue muy significativo, se rechaza la hipótesis de que el grupo 2 está calibrado. Se concluye que hay un error sistemático con respecto al laboratorio de referencia. Esto significa que el programa regional debe informar a los 5 laboratorios del grupo número 2 de esta situación, además de darles el ranking obtenido a cada uno de ellos, deberá arbitrar las soluciones del caso. Si no se usa una sangre calibrada, la única conclusión a sacar es que no es lo mismo una técnica que la otra, pero cuando se tiene el valor patrón, para el responsable de cada laboratorio de ese grupo, significa que se ha descubierto y probado un error sistemático que afecta a su técnica usual, evaluado en $ES = 3,05 - 2,9467 = 0,1033 \cdot 10^6$ gl/ml. Luego le quedan dos caminos: El más rápido y sencillo es corregir los resultados que logre, restando el valor del ES a cada medición efectuada con esa técnica. El segundo camino, más científico, es ponerse a investigar las causas que originan ese tipo de error, y así poder corregir su técnica.

Grupo 3: Otras técnicas.

Para los laboratorios que no emplean alguna de las dos técnicas anteriores resulta:

$$t = \frac{(\bar{x} - \mu)}{DS/\sqrt{n}} = \frac{(3,22 - 2,9467)}{0,08367/\sqrt{5}} = 7,303^{**} \text{ versus } t_{\alpha;v} = t_{0,99;4} = 4,604 \text{ y } t_{0,999;4} = 8,610$$

Hay evidencia como para rechazar la Ho al (99% de confianza). Como el resultado fue muy significativo, se rechaza la hipótesis nula de que el grupo 3 está calibrado. Se concluye que los 5 laboratorios que usan otras técnicas, como el recuento en cámara o el Macrohematocrito, deben ser informados de la gran diferencia que tienen con el laboratorio de referencia, se ha encontrado evidencia muy fuerte de un ES = 0,2733 gl/ml. Y para responsable de cada uno de estos 5 laboratorios del Grupo 3, significa que la técnica usual, no solo debe ser desechada como confiable, sino que es la peor de todas. Como era de esperar, de acuerdo a la bibliografía y por ello se desaconseja su uso en el laboratorio moderno.

Método 2) Calibración con intervalos de confianza (95%)

Ho : El método está calibrado: $\mu \in (\bar{X} \pm \Delta X)$ Con $\Delta X = t_{0,95;4} \cdot DS/\sqrt{n}$
 H1 : El método no está calibrado: $\mu \notin (\bar{X} \pm \Delta X)$ y $\mu = 2,9467$

Grupo	\bar{X}	ΔX	Límite Superior	Límite inferior	μ cae
1	2,95	0,012	2,962	2,938	Adentro
2	3,05	0,041	3,091	3,009	Fuera
3	3,22	0,104	3,324	3,116	Fuera

Se rechaza la Ho para los dos últimos casos.

17.3.2 Control de precisión

Para controlar la precisión de una técnica clínica o industrial, se necesita de un valor referencial de la *desviación máxima admisible* ($\sigma_{m\acute{a}x}$). En Bioquímica, tales valores se toman de algunos criterios existentes, tales como: usar el error relativo máximo, el criterio de Thonks, el de Aspen, etc. (que se verán más adelante). O bien, en los libros de texto de Análisis Clínicos. En Farmacia, se usan los mismos en los Laboratorios de Control de Calidad industriales. Pero, cuando se trata de sistemas de producción, los límites vienen especificados en el protocolo de fabricación del medicamento. En resumen se necesita definir primero un $\sigma_{m\acute{a}x}$ y luego efectuar un test de hipótesis con la distribución de Chi Cuadrado. El planteo es:

Ho : El sistema tiene una precisión aceptable: $DS \leq \sigma_{m\acute{a}x}$
 H1 : El sistema no tiene una precisión aceptable: $DS > \sigma_{m\acute{a}x}$

Lo que se testea con: $\chi^2 = (n - 1) DS^2 / \sigma_{m\acute{a}x}^2$ versus $\chi^2_{\alpha;v}$

Para ilustrar estas ideas se desarrolla el siguiente ejemplo, donde se usan los datos del ejemplo de calibración anterior:

Ejemplo) Usar los datos del Cuadro 17.4, con un CV%_{máx} del 5%, para decidir si los Grupos tienen una precisión aceptable.

Grupo	DS	$\sigma_{m\acute{a}x} = 0,05 \cdot \mu$	χ^2	$\chi^2_{0,95;4}$	Resultado
1	0,01	0,1473	0,018	9,488	No se rechaza Ho
2	0,03317		0,203		No se rechaza Ho
3	0,08367		1,299		No se rechaza Ho

Conclusión: Todos los grupos tienen una precisión aceptable.

17.3.3 Control del factor humano

Para controlar la influencia del factor humano en las mediciones clínicas se debe hacer medir lo mismo a varios operadores diferentes. Así, se pueden comparar los valores medidos por cada uno, entre sí, con un modelo de Anova. En el ejemplo siguiente se usa un modelo de un factor para ilustrar el método, sin embargo, cuanto más factores se tomen en cuenta, mejor será la sensibilidad del modelo estadístico para detectar las diferencias.

Cuadro 17.5.1 : Comparaciones de 3 operadores distintos

1	27	27	20	18	Σ
	32	30	22	20	
	30	28	22	21	
	29	30	20	18	
	30	32	22	20	
	35	30	23	20	
					606
2	30	24	22	22	Σ
	30	30	22	24	
	30	30	23	23	
	30	30	23	20	
	30	30	24	24	
	30	30	25	22	
					628
3	30	26	19	22	Σ
	31	32	22	21	
	32	30	22	21	
	36	36	22	19	
	34	31	25	24	
	30	29	24	23	
					641
Σ Σ					1875

Fuente: Dras. Cech, N. ; Coschiza, M. y Lodeiro, N.

Ejemplo 1) Se desea saber si hay diferencias entre los tres operadores de un laboratorio en la medición de Antibiógramas. Lo que se mide es el diámetro del halo que deja cada antibiótico en la caja Petri donde se efectuó el cultivo. Los datos que se muestran en el Cuadro 17.5.1 corresponden a 6 pacientes diferentes, empleando 4 antibióticos para cada uno de ellos. Acá se resuelve el problema usando un modelo de 1 factor, pues no se toman en cuenta los diferentes antibióticos. Es como si se tuviesen 26 diámetros a medir, por los 3 operadores, con lo que se obtienen un total de 72 datos.

Paso 1) Se calcula la suma total de datos $T = \sum_1^a T_c = 1.875$

Paso 2) Se calcula la suma del cuadrado de todos los datos: $T_x^2 = \sum_1^a \sum_1^b X_{ijk}^2 = 50.459$

Paso 3) Se calcula el término de corrección $T^2 / N = (1.875)^2 / 72 = 48.828$

Paso 4) Se calcula la suma de los cuadrados entre los operadores:

$$T_E^2 = \sum_1^a \sum_1^n T_{ik}^2 / a.n = 48.854$$

Paso 6) Se calcula la suma de los cuadrados respectivos:

$$SS_{total} = T^2 - T^2 / N = Paso 2 - Paso 3 = 1.631$$

$$SS_{entre} = T_E^2 - T^2 / N = Paso 4 - Paso 3 = 26 \text{ (operadores)}$$

$$SS_{dentro} = SS_{total} - SS_{entre} = 1.631 - 26 = 1.605$$

Paso 7) Se calcula el cuadro de ANOVA como sigue:

Tabla 20.2: Cuadro de ANOVA para los datos de Tabla 20.1

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre Filas (operadores)	26	2	13	0,56 (ns)
Dentro grupos (error)	1.605	69	23,26	
Total	1.631	71		

Los resultados no fueron significativos, por lo que se concluye que no hay diferencias entre la forma de medir de los tres operadores. Es decir, el factor humano no influye en los datos.

17.4 Modelos I y II del ANOVA

De acuerdo a la forma en que fue diseñado el experimento, es diferente la manera de proseguir usando el modelo de ANOVA. Hay dos modelos posibles, los cuales fueron descriptos por primera vez por Eisenhart en 1947.

Modelo I : se supone que el efecto de factor analizado es *constante* en cada grupo.

Modelo II : se supone que el efecto del factor analizado es *aleatorio* en cada grupo.

Hasta acá, tanto el Modelo I como el Modelo II, usan los mismos cálculos y se llega al Cuadro de ANOVA. Si no se encuentran diferencias significativas entre las muestras, no tiene sentido seguir adelante. Pero, si se rechaza la H_0 , entonces el problema siguiente es descubrir donde las diferencias se hacen significativas.

Cuando el investigador selecciona los diferentes grupos a investigar siguiendo un criterio determinado, quiere decir que cada grupo es elegido a propósito. En cuyo caso, se supone que el efecto del factor en cada grupo es una constante, que lo afecta por igual en todas sus mediciones. Aunque, esa constante puede variar de grupo a grupo, y ese efecto justamente es lo que se está tratando de descubrir. Es una especie de error sistemático actuando sobre el grupo y modificando su media de manera tal que la diferencia con el promedio general se hace significativo. Por esto, es el investigador quien opta por el Modelo I, al elegir cada grupo a su modo.

En cambio, cuando el investigador selecciona al azar, entre todos los casos posibles, a los grupos de trabajo, entonces se trata de un Modelo II y se supone que el efecto del factor en cada grupo es de tipo aleatorio. Por ejemplo, si un farmacéutico desea investigar el efecto de los antibióticos sobre cierta enfermedad, y mediante un sorteo, selecciona a 5 de ellos, entre los vigentes en el mercado, entonces se trata de un Modelo II. Si en cambio, tal selección la hace buscando a los que contengan sulfato de Neomicina, de 5 marcas diferentes, entonces se trata de un Modelo I porque tal selección no la hizo al azar.

Cuando se realiza un diseño en Modelo II, generalmente se trata de estudios poblacionales, donde se trata de determinar si el factor de análisis, produce variabilidad en las muestras. El problema allí, es calcular el porcentaje de incremento en la varianza, que es la medida cuantitativa de la dispersión en general. La varianza total, se descompone en dos términos, uno debida al factor y otro remanente, resto o error. El porcentaje de varianza debida al factor se denomina: *Componente Añadida de Varianza* y en el próximo capítulo se verá como calcularla.

Cuando se efectúa un experimento diseñado como Modelo I, es porque se está interesado en determinar y cuantificar, el efecto constante del factor en la muestra. De lo que se trata en tal caso, es de poder comparar entre sí las medias muestrales para detectar si hay diferencia significativa entre ellas. En los ejemplos vistos en el punto de calibración anterior, la elección de los tres grupos no se hizo al azar, sino que se eligieron a propósito y por lo tanto se trataba de un Modelo I. Las comparaciones entre medias se hicieron con el modelo Student, el que solo puede comparar de a dos muestras cada vez. Entonces, el camino fue hacer todas las comparaciones posibles de a pares. Sin embargo, cuando la cantidad de muestras es grande, este proceder puede ser muy engorroso y difícil de entender. Para estos casos, hay una forma más elegante, basada en el ANOVA, denominada: *Comparaciones Múltiples*. Objeto central del próximo capítulo.

Para entender mejor la utilidad de estos conceptos, conviene hacerse una pregunta básica: ¿ Por que el segundo valor del primer grupo del ejemplo anterior fue 2,96 en lugar de cualquier otra cantidad ? Una primera aproximación a la respuesta es pensar que esto se debe a que fue extraído de una población con una media $\mu \in (2,9467 \pm 0,0004) \cdot 10^6$ gl/ml, pero si se desconoce tal valor, se puede usar su mejor estimación:

$\bar{X} = T / N = 3,07 \cdot 10^6$ g/ml. Luego como primera aproximación se puede imaginar la relación:

$$X_{21} \cong \mu$$

Pero en este caso el valor de X_{21} es 2,96 y $\mu = 3,07$ (o con sangre calibrada $\mu = 2,9467$). Hay una diferencia de 0,11 unidades que sigue existiendo. Para el Modelo I, se puede efectuar una segunda aproximación y pensar que ese valor forma parte del grupo 1 cuya media es de 2,95. Luego el efecto de medir con un contador automático provoca en las mediciones una diferencia constante α_1 , para el grupo 1 de $(3,07 - 2,95)$. O sea, un valor negativo $\alpha_1 = -0,12$ a lo largo de todo el grupo. Entonces, como una segunda aproximación se puede expresar :

$$X_{21} \cong \mu + \alpha_1 = 3,07 - 0,12 = 2,95$$

Todavía hay una diferencia de 0,01. Pues $X_{21} = 2,96$ y con la segunda aproximación se lo estima en 2,95. La pregunta siguiente es ¿ Por qué este valor individual no puede ser totalmente explicado ? La respuesta es: porque hay una variabilidad inherente de tipo genético o ambiental que afecta el valor medido. Pero, si se trabaja con una misma sangre se ha acotado la variabilidad genética, que aparece en las mediciones biológicas entre individuos. Queda todavía la variabilidad ambiental, para acotarla se puede realizar todas las mediciones en un mismo ámbito, con temperaturas y humedad controladas. En el ejemplo visto: en caso de trabajar con varios laboratorios, ubicados en diferentes lugares físicos; no hay manera de reducir la variabilidad ambiental. De todas maneras, se podría haber establecido a los laboratorios intervinientes en el experimento, las condiciones ambientales, tales como humedad ambiente, temperatura, alimentación eléctrica regulada, etc. De manera tal, de dejar constantes esos valores y acotar la variabilidad ambiental. Pero, siempre hay algo incontrolable, por más que trabaje en las condiciones asépticas de un quirófano, todavía falta fijar variaciones propias de mediciones repetidas. Y en el hipotético caso, de fijar ese tipo de variabilidad, todavía queda el irremediable error de medición. Entonces, la tercera aproximación y final que se puede plantear es:

$$X_{21} = \mu + \alpha_1 + \sigma_{ij}$$

Donde se supone que el tercer término agregado, es un valor particular que adopta una función error de medición, de tipo gaussiana con parámetros $(0 ; \sigma_{ij})$. Entonces, se puede expresar el Modelo I en términos generales con:

Modelo I de ANOVA :

$$X_{ij} = \mu + \alpha_1 + \sigma_{ij}$$

Esto ya es una mejor explicación, para saber porque se produce una medida en particular y arroja cierta luz sobre la naturaleza de las mediciones. El ANOVA permite cuantificar estas ideas para poder trabajar en optimizaciones.

En un Modelo II, se supone que el efecto del factor en el grupo es de tipo aleatorio, en lugar de constante. Entonces, en lugar de poner un valor constante α_i dentro de un grupo cualquier

ra, se plantea un efecto de tipo aleatorio dado por A_i , que tiene una distribución gaussiana de parámetros $(0 ; \sigma_A)$

Modelo II de ANOVA :

$$X_{ij} = \mu + A_i + \sigma_{ij}$$

17.5 Supuestos básicos del ANOVA

De acuerdo a lo visto, para un modelo unifactorial de ANOVA se necesitan hacer los cuatro supuestos siguientes:

- 1) *Las muestras deben ser aleatorias.*
- 2) *Las muestras deben ser independientes.*
- 3) *La población de donde son extraídas debe ser gaussiana.*
- 4) *Sus varianzas deben ser iguales.*

Si se verifican estos supuestos, entonces con hipótesis nula de igualdad de medias muestrales se prueba si todas las muestras fueron extraídas de la misma población, contra la hipótesis alternativa de que el efecto del factor analizado es significativo.

17.6 Modelo no paramétrico de Kruskal-Wallis

Cuando no se verifique alguno de los cuatro supuestos básicos del Anova (normalidad, aleatoriedad, independencia y homoscedasticidad), o bien, la magnitud estudiada no sea continua, se tiene el recurso de usar un modelo no paramétrico equivalente: *Modelo de Kruskal-Wallis*.

Este tipo de modelo se hizo muy popular debido a que los cálculos son muy sencillos y no hay que preocuparse por las hipótesis del Anova. Aunque claro está, el Anova tiene mayor efectividad para detectar las diferencias entre las medias muestrales. La manera de proceder con el modelo de Kuskal-Wallis es muy similar al de la U de Mann y Whitney, se trata de ordenar primero los datos en forma creciente, ignorando la división en grupos. Se le asigna un rango a cada dato y en el caso de empates o "ligas" se usa el rango promedio entre ellos. Luego se reemplaza cada dato de la tabla original, por su rango equivalente y con los nuevos datos se calcula el estadígrafo H de Kruskal-Wallis, cuya función distribución se muestra en la Tabla 17 del Anexo.

El estadígrafo H depende de dos cosas: el rango de cada uno de los valores obtenidos y del tamaño muestral respectivo. Su cálculo es algo engorroso pero muy sencillo de efectuar como se muestra en el ejemplo siguiente.

$$H = \left[\frac{12}{\sum_1^a n_i (\sum_1^a n_i + 1)} \sum_1^a \frac{(\sum_1^{n_i} R)_i^2}{n_i} \right] - 3 \left(\sum_1^a n_i + 1 \right)$$

Si se tiene en cuenta que para el caso de muestras iguales $N = \sum_1^a n_i = a \cdot n$ resulta:

$$H = \left[\frac{12}{N(N+1)} \sum_1^a \frac{(\sum_1^{n_i} R)_i^2}{n} \right] - 3(N+1)$$

Para ilustrar este modelo se toman los datos del Cuadro 17.6 de un estudio de diferentes drogas para el tratamiento de una enfermedad, tanto puras como mezcladas y un control o placebo. Se mide el tiempo de curación en días, desde que los pacientes comienzan el tratamiento.

Cuadro 17.6: Estudio comparativo de las drogas A, B y C.

Nº	Placebo	Droga A	Droga B	Mezcla A+B	Droga C
1	125	107	108	108	112
2	117	108	111	109	116
3	120	110	106	108	115
4	125	109	108	111	113
5	115	112	107	107	114
6	121	110	106	106	112
7	117	110	111	108	115
8	117	107	110	107	115
9	126	109	107	107	112
10	118	111	108	109	117

Fuente: Ejemplo derivado del libro de Sokal y Rohlf

Los pasos a seguir para resolver este modelo son los siguientes:

Paso 1) Se calculan los rangos de cada valor obtenido:

valor	rango	R	valor	rango	R	valor	rango	R	valor	rango	R	valor	rango	R
106	1	2	108	11	14	109	21	19,5	112	31	31,5	117	41	42,5
106	2	2	108	12	14	110	22	23,5	112	32	31,5	117	42	42,5
106	3	2	108	13	14	110	23	23,5	112	33	31,5	117	43	42,5
107	4	7	108	14	14	110	24	23,5	113	34	34	117	44	42,5
107	5	7	108	15	14	110	25	23,5	114	35	35	118	45	45
107	6	7	108	16	14	111	26	27,5	115	36	37,5	120	46	46
107	7	7	108	17	14	111	27	27,5	115	37	37,5	121	47	47
107	8	7	109	18	19,5	111	28	27,5	115	38	37,5	125	48	48,5
107	9	7	109	19	19,5	111	29	27,5	115	39	37,5	125	49	48,5
107	10	7	109	20	19,5	112	30	31,5	116	40	40,0	126	50	50

El tiempo de cura medido en días se ordena de menor a mayor, sin tomar en cuenta a que grupo pertenece cada dato. El valor mínimo obtenido fue de 106 días y el mayor de 126. Estos valores se vuelcan a un cuadro como el de más abajo, en las primeras columnas de cada bloque de 10 datos. Hay 5 bloques de 3 columnas cada uno. En la segunda columna de cada bloque se coloca el rango correspondiente: 1 para el menor y así sucesivamente hasta llegar al mayor de rango 50. En caso de empates, se asigna a los datos empatados, el rango promedio. Por ejemplo, los tres primeros datos, tienen el mismo valor 106, le tocan los rangos 1, 2 y 3, luego su promedio 2 es asignado a cada uno de ellos.

Paso 2) Luego se reemplazan los datos del cuadro original por sus respectivos rangos:

Cuadro 17.7 : Reemplazo de los datos del Cuadro 17.6 por sus rangos.

Nº	Placebo	Droga A	Droga B	Mezcla A+B	Droga C	Totales
1	48,5	7	14	14	31,5	
2	42,5	14	27,5	19,5	40	
3	46	23,5	2	14	37,5	
4	48,5	19,5	14	27,5	34	
5	37,5	31,5	7	7	35	
6	47	23,5	2	2	31,5	
7	42,5	23,5	27,5	14	37,5	
8	42,5	7	23,5	7	37,5	
9	50	19,5	7	7	31,5	
10	45	27,5	14	19,5	42,5	
Total	450	196,5	138,5	131,5	358,5	1275

Paso 3) Luego se reemplazan los datos obtenidos en la fórmula de H :

$$H = \left[\frac{12}{50(50+1)} \cdot \frac{(450)^2 + (196,5)^2 + (138,5)^2 + (131,5)^2 + (358,5)^2}{10} \right] - 3(50+1) = 38,109$$

Paso 4) Dado que se encontraron empates, el valor de H debe ser corregido dividiendo por D :

$$H_{\text{corregido}} = H / D \quad \text{y para ello D se calcula con:}$$

$$D = 1 - \frac{\sum_1^m T_j}{(N-1)N(N+1)} \quad \text{donde } T_j = (E_j^3 - E_j) \text{ con } m = 10 \text{ empates } E_j$$

En el primer empate hay 3 valores empatados en 106 días, luego es:

$$T_1 = (E_1^3 - E_1) = (3^3 - 3) = 24$$

En el segundo y tercer empate hay 7 valores empatados en 107 días, luego es:

$$T_2 = (E_2^3 - E_2) = (7^3 - 7) = 336 = T_3$$

Luego hay cuatro empates seguidos de 4 valores de 109 a 112 y otros dos en 115 y 117

$$T_4 = (E_4^3 - E_4) = (4^3 - 4) = 60 = T_5 = T_6 = T_7 = T_8 = T_9$$

Finalmente hay un solo empate de 2 valores en 125

$$T_{10} = (E_{10}^3 - E_{10}) = (2^3 - 2) = 6$$

Luego el total de los puntajes por empate es:

$$\sum_1^m T_j = 24 + 336 + 336 + 6 \cdot (60) + 6 = 1062$$

Reemplazando estos valores, el factor de corrección es:

$$D = 1 - \frac{\sum_1^m T_j}{(N-1)N(N+1)} = 1 - \frac{1062}{49(50)51} = 1 - 0,0085 = 0,99151$$

Entonces $H_{\text{corregido}} = H / D = 38,109 / 0,99151 = 38,44$

Para muestras grandes el estadígrafo H se distribuye aproximadamente según una distribución Chi cuadrado, mientras que si los tamaños muestrales son pequeños ($N < 6$) se puede emplear la distribución exacta Tabla 17 del Anexo. En este caso, como $N=50$ se formula la hipótesis nula H_0 de que los 5 grupos testeados no difieren en “localización”, es decir no hay diferencia entre ellos, y se contrasta con $\chi^2_{\alpha(v)} = \chi^2_{0,01(5-1)} = 13,277$. Como este resulta menor que el valor del estadígrafo $H_{\text{corregido}} = 38,44^{**}$ se rechaza la hipótesis nula. Se puede afirmar que hay evidencia científica de que hay diferencias entre los grupos investigados.

17.7 Comparación de varias proporciones

En los capítulos anteriores se analizó la comparación de dos proporciones en dos casos básicos. La primera era usar los modelos de Gauss y Student para hacerlo usando el error de muestreo. La segunda era usar el modelo de la Chi-cuadrado para analizar la independencia en tablas de contingencia de 2×2 . En estos casos la magnitud era dicotómica y se analizaban dos factores (Tabla de la Verdad, Tabla de contingencia, Epidemiología, etc.). Cuando más de dos proporciones deben ser comparadas, el cálculo del error estándar entre un par de proporciones, se parece mucho a las comparaciones múltiples que se verán en el próximo capítulo. Sin embargo, el test general para determinar si hay diferencias significativas entre todas las proporciones que se quieren analizar es análogo al F-test del Anova. En realidad es una extensión del análisis de la Chi-cuadrado para varias proporciones.

Suponiendo que hay k grupos de observaciones, y que en el grupo número i de tamaño n_i hay una cantidad de individuos r_i que muestran cierta característica (presente / ausente, sano / enfermo, etc.); entonces la proporción de individuos con la característica será $p_i = r_i / n_i$. Los datos se pueden presentar como en el Cuadro 17.8

Cuadro 17.8 Comparación de k proporciones.

Grupos	1	2	3	...	i	...	k	Todos los grupos
(+)	r_1	r_2	r_3	...	r_i	...	r_k	R
(-)	$n_1 - r_1$	$n_2 - r_2$	$n_3 - r_3$...	$n_i - r_i$...	$n_k - r_k$	N - R
Total	n_1	n_2	n_3	...	n_i	...	n_k	N
Proporción	p_1	p_2	p_3	...	p_i	...	p_k	$P = R / N$

Las frecuencias de esta tabla de contingencia de 2 x k son dicotómicas (+) y (-) y se muestran en las dos filas de datos. En la tercer fila están los tamaños grupales y en la cuarta las proporciones de positivos respectivas. El test de la Chi-cuadrado sigue el mismo esquema presentado antes, donde para cada una de las frecuencias observadas, se puede calcular su frecuencia esperada con el producto de los respectivos totales marginales dividido el tamaño muestral, o sea:

$E_i = (\text{Total de la fila } i \times \text{Total de columna } i) / N$ y el valor del estadígrafo con:

$$\chi^2 = \sum_1^n \frac{(O_i - E_i)^2}{E_i}$$

La H_0 es que todas las k proporciones de positivos son iguales, porque provienen de la misma población de donde fueron extraídas aleatoriamente. Se hace el test comparando el valor muestral contra el de tablas de la Chi-cuadrado, con k - 1 grados de libertad. No se requiere corrección por continuidad a menos que las frecuencias observadas sean muy pequeñas, ya que se usan los mismos totales marginales para el número de tablas individuales de 2 x 2, y el número total es muy grande. Por tanto, estos se ajustan mucho mejor a la distribución Chi-cuadrado que en cada una de las tablas individuales de 2 x 2 que la componen. La proporción promedio se puede obtener usando la media ponderada con:

$$P = R / N = \frac{\sum_1^k n_i \cdot p_i}{\sum_1^k n_i} \quad \text{y} \quad Q = 1 - P$$

La contribución de cada grupo al total de χ^2 es, por ejemplo para el grupo i, igual a:

$(r_i - P n_i)^2 [(1 / P n_i) + (1 / Q n_i)]$ Donde reemplazando y manipulando la fórmula queda:

$$\chi^2 = \sum_1^k \frac{n_i \cdot (p_i - P)^2}{P \cdot Q}$$

Esta fórmula es equivalente a la anterior vista en el Capítulo 15 y 16, pero los supuestos básicos han sido reemplazados en dos aspectos: a) la variación de las proporciones siguen una distribución binomial en lugar de la normal y b) la varianza verdadera ha sido estimada con $(P Q) / n_i$. A Medida que los tamaños muestrales aumentan, estas aproximaciones pierden importancia. Para ilustrar este procedimiento se presenta el ejemplo siguiente, tomado del libro de Armitage y Berry (página 207): Se analizaron muestras de materia fecal de 288 individuos, agrupados por edades, para ver la presencia de huevos de la *Schistosoma mansoni*.

Frecuencias observadas

Edad (años)	0-9	10-19	20-29	30-39	40-49	Total
Positivo	14	16	14	7	6	57
Negativo	87	33	66	34	11	231
Total	101	49	80	41	17	288

Frecuencias esperadas

Edad (años)	0-9	10-19	20-29	30-39	40-49	Total
Positivo	20	9,7	15,8	8,1	3,4	57
Negativo	81	39,3	64,2	32,9	13,6	231
Total	101	49	80	41	17	288

Contribución de χ^2 para cada celda ($\chi^2 = 10,36$)

Edad (años)	0-9	10-19	20-29	30-39	40-49	Total
Positivo	1,79	4,09	0,21	0,15	2,07	8,31
Negativo	0,44	1,01	0,05	0,04	0,51	2,05
Total	2,23	5,10	0,26	0,19	2,58	10,36

Hay 4 grados de libertad y el valor crítico de tablas es $\chi^2_{0,95; 4} = 9,49$; por lo tanto la H_0 debe ser rechazada. Hay diferencias significativas entre las proporciones halladas. Por otra parte se pueden calcular:

$$P = R / N = \frac{\sum_1^k n_i \cdot p_i}{\sum_1^k n_i} = 0,1979 \quad \text{y} \quad Q = 1 - P = 0,8021 \quad \text{Entonces,}$$

$$\chi^2 = \sum_1^k \frac{n_i \cdot (p_i - P)^2}{P \cdot Q} = \{101 [(14/101) - 0,1979]^2 + \dots + 17 [(6/17) - 0,1979]^2\} / [(0,1979) (0,8021)] = 10,37^*$$

La pequeña discrepancia entre los valores de χ^2 calculados por ambos métodos se deben a errores de redondeo, pero en ambos casos la H_0 debe ser rechazada. Si en lugar de la proporción de positivos, se hubiese usado la de negativos el resultado final es el mismo.

17.8 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|--|----------|----------|
| 1) Con los modelos de ANOVA se pueden comparar más de dos muestras a la vez. | V | F |
| 2) Y también más de un factor, con varias muestras a la vez. | V | F |
| 3) Para poder aplicar los modelos de ANOVA las muestras deben ser: | | |
| 4) En ANOVA se pueden usar muestras de diferente tamaño. | V | F |
| 5) La SS total se particiona en dos partes que son: | | |
| 4) Los grados de libertad se dividen en dos partes, lo mismo que la Suma de Cuadrados. | V | F |
| 5) Los tres MS se pueden calcular con:..... | | |
| 6) Los MS entre y dentro, sumados, son igual a los MS totales. | V | F |
| 7) Con el Cuadro de ANOVA se pueden presentar los resultados en forma resumida. | V | F |
| 8) El estadígrafo de comparación F se calcula como el cociente de los MS entre y dentro. | V | F |
| 9) La distribución teórica del F es una distribución Chi Cuadrado. | V | F |
| 10) Los diez pasos a seguir para calcular el Cuadro de Anova en forma rápida son: | | |
| 11) La ventaja del ANOVA respecto a Student es que puede comparar más de 2 muestras. | V | F |
| 12) En el Modelo I de ANOVA se supone que el factor actúa en forma constante. | V | F |
| 13) En el Modelo II de ANOVA se supone que el factor actúa en forma aleatoria. | V | F |
| 14) El factor actúa constante del Modelo I, actúa por igual en todas las muestras. | V | F |
| 15) El modelo de Kruskal-Wallis es el equivalente no paramétrico del ANOVA. | V | F |

- 16) En caso de empates se debe calcular el rango promedio entre los empates o ligas. **V F**
 17) La idea de K-W es reemplazar los valores originales por sus rangos y calcular un H. **V F**
 18) El factor de corrección se usa siempre y vale H/D. **V F**
 19) El estadígrafo H se distribuye según una función F de Fisher. **V F**
 20) Para tamaños muestrales muy pequeños $N < 40$ se puede usar la Tabla 17 del Anexo. **V F**
 2) Se desea investigar el efecto de los alimentos balanceados en la cría de pollos para un productor de la zona. El experimento consiste en pesar los pollos antes de comenzar y al final de un mes de pruebas. Las diferencias en peso encontradas en cada uno se muestran en la tabla siguiente. Como control se alimenta a un grupo de la forma tradicional. Se escogen al azar 10 pollos por grupo.

Nº	Control	Marca 1	Marca 2	Marca 3
1	150	207	230	221
2	160	208	235	225
3	140	210	228	219
4	135	209	240	217
5	155	212	238	225
6	151	210	226	222
7	147	220	234	223
8	137	207	225	217
9	146	209	239	215
10	138	211	237	224

Averiguar si hay diferencias significativas entre las marcas testeadas usando:

- El modelo de ANOVA.
- El modelo de Kruskal Wallis
- Comparar las medias usando el Modelo de Student.
- ¿Cuál es la diferencia si las marcas hubiesen sido escogidas al azar entre las del mercado?

3) Resolver el problema de ejemplo 1, del punto 17.3, usando el modelo no paramétrico de Kruskal Wallis, y comparar con los resultados allí obtenidos.

4) Un bioquímico desea averiguar si hay diferencia entre tres marcas comerciales de kits, para efectuar la técnica de glucosa. Para ello, prepara un “pool” de sueros, lo homogeneiza bien y los fracciona en 15 alícuotas. Asigna cinco viales para cada marca y mide la glucosa para el mismo suero en forma repetida. Si la homogeneización está bien realizada, se dan por cumplidos los supuestos de ANOVA. Los resultados se muestran en la tabla siguiente. Decidir si existen diferencias significativas entre las marcas probadas en el experimento.

Medición Nº	Marca 1	Marca 2	Marca 3
1	1,04	1,10	1,21
2	1,05	1,11	1,19
3	1,06	1,09	1,18
4	1,07	1,07	1,17
5	1,05	1,08	1,21

5) Un farmacéutico desea saber si una promoción produce algún efecto en sus ventas. Para ello toma como control el promedio histórico de ventas de los días miércoles, escogidos al azar, de

sus datos. Tal día se tomó como referencia luego de un sorteo. Prueba un mes con el primer tipo de promoción, deja pasar dos meses, y prueba un mes con un segundo tipo de promoción que incluye premios a los clientes. Sus resultados se muestran en la tabla siguiente, sobre la base de ellos decidir si las promociones dan resultado.

Medición Nº	Control	Promo 1	Promo 2
1	245	230	276
2	176	252	289
3	198	261	293
4	217	245	282

6) En una industria farmacéutica hay cuatro líneas de producción para la fabricación de analgésicos, con distinta tecnología. Sus rendimientos horarios son similares, pero el encargado desea averiguar si la cantidad de productos rechazados es la misma para los cuatro. Para ello toma de los registros históricos de producción 7 días elegidos al azar, del último semestre. Sus resultados fueron :

Dato Nº	1	2	3	4
1	452	322	298	340
2	379	345	312	358
3	412	367	280	345
4	320	341	310	362
5	350	372	235	370
6	390	317	304	326
7	378	324	320	368

Decidir si hay diferencias entre las cuatro tecnologías.

7) Se ha determinado una magnitud clínica mediante cuatro operadores diferentes usando una misma técnica y con la misma muestra sanguínea. El objeto es detectar si hay diferencias debido al factor humano en los resultados del laboratorio. Los datos obtenidos son:

Dato Nº	A	B	C	D
1	45	44	40	44
2	43	44	44	45
3	41	45	45	46
4	42	43	46	45

a) Se pide determinar usando el modelo de ANOVA unifactorial si hay diferencia entre los cuatro operadores. En caso de encontrarla, investigar usando el modelo Student todos los pares posibles para determinar entre cuales de ellos hay diferencias.

b) Si se sabe que el valor real de la magnitud es 45 estudiar a cada operador en exactitud y precisión para determinar un orden de mérito entre los mismos.

18

Comparaciones múltiples

Los diferentes métodos de hacer comparaciones múltiples se emplean *sólo* cuando el resultado del ANOVA resulta significativo. En tal caso, se sabe que existen diferencias entre las muestras, pero sin poder especificar entre cuales de ellas. Se necesita, entonces, alguna forma de poder compararlas entre sí, y alcanzar así el objetivo final del ANOVA. Por ejemplo, decidir si un placebo difiere realmente de los demás grupos, y todavía, si entre éstos algunos son diferentes de otros. Los métodos de *comparaciones múltiples* son la herramienta adecuada para este fin, pues permiten un análisis muy sutil de la significación encontrada. Para un *Modelo II* el trabajo es mucho más simple, solo se trata de cuantificar a la componente añadida de varianza. En cambio, para un *Modelo I* se debe analizar primero si las comparaciones fueron *planeadas* antes de realizar el experimento o después. Cuando se analizan los datos y se encuentran resultados sorprendidos, no previstos, a veces conviene un análisis posterior que se diseña en función de estos resultados, en cuyo caso se deben usar los métodos para comparaciones *no planeadas*. Existen numerosos modelos estadísticos para todas estas posibilidades, tanto para tamaños muestrales iguales como distintos en esta parte se presentan aquellos más recomendables por su potencia, sensibilidad y sencillez.

18.1 Componente añadida de varianza

Las variaciones detectadas con un Modelo II en el ANOVA se presentan en el cuadro respectivo, como proviniendo de dos fuentes, una debida a la variabilidad entre los grupos y la otra debida al error de medición. Cuando no se detecta significación en el ensayo, ocurre que los valores de los cuadrados medios involucrados son aproximadamente iguales; entonces su cociente será cercano a la unidad, esto es:

$$F = MS_E / MS_D \approx 1 \quad (\text{resultados no significativos})$$

Por otra parte, el valor esperado del estadígrafo F es:

$$E(F) = E(MS_E) / E(MS_D) = (\sigma^2 + n \sigma_A^2) / (\sigma^2)$$

Luego para que este cociente sea aproximadamente la unidad, la componente añadida de varianza σ_A^2 debe ser nula. Por lo tanto, se dice que si el Cuadro de ANOVA resulta no significativo en un Modelo II, es porque la componente añadida no influye. En cambio, si el Cuadro resulta ser significativo, entonces la componente añadida no es despreciable, y se está interesado en el valor de su contribución a la variabilidad total, más que en su valor en sí. No interesa su magnitud, sino el porcentaje que tiene en la variabilidad total. El valor de interés se calcula con el cociente porcentual entre el valor de la componente añadida y el total. De la relación anterior resulta entonces:

$$E(MS_E) - E(MS_D) = (\sigma^2 + n \sigma_A^2) - (\sigma^2) = n \sigma_A^2$$

Donde n es el número de mediciones de cada grupo. O sea :

$$\sigma_A^2 = [E(MS_E) - E(MS_D)] / n$$

Entonces la mejor estimación de la componente añadida de varianza es su valor muestral :

$$\sigma_A^2 \approx DS_A^2 = [(MS_E - MS_D)] / n$$

Pero en los hechos, más que la magnitud absoluta de la componente añadida DS_A^2 interesa obtener el valor porcentual de su contribución al total (P%) y este valor se calcula con:

$$P\% = 100 (\sigma_A^2 / \sigma^2) \approx 100 [(DS_A^2) / (DS^2)]$$

$$\text{Donde } DS^2 = MS_E + MS_D$$

Por ejemplo, para el Cuadro de ANOVA visto en el capítulo anterior (Paso 8 del ejemplo en el punto 17.3), suponiendo que los grupos se hubiesen escogido al azar, en lugar de cómo se hizo, se trataría de un Modelo II y así el valor de la contribución porcentual debida a la componente añadida resulta ser :

$$MS_E = 0,09317 \quad ; \quad MS_D = 0,00273 \quad ; \quad n = 5 \quad \text{O sea, } F = 34,1^* \text{ (resultado significativo)}$$

Para el caso del Modelo II se puede obtener :

$$DS_A^2 = (1/n) (MS_E - MS_D) = (1/5) (0,09317 - 0,00273) = 0,01809$$

$$DS^2 = MS_E + MS_D = 0,09317 + 0,00273 = 0,0959$$

Entonces resulta:

$$P\% = 100 (0,01809 / 0,0959) = 18,87 \%$$

Así el 18,87% de la variabilidad total del experimento, se debe a la componente añadida de varianza entre los grupos. El factor aleatorio incidente es muy importante. Para este problema, significa que el tipo de técnica empleada para medir el RGR debe ser tenido muy en cuenta para lograr estudiar la concordancia entre los diferentes laboratorios.

18.2 Comparaciones múltiples “ a priori ”

Cuando se trata de un modelo I de Anova, lo que más interesa es poder comparar las medias muestrales entre sí, una vez que se sabe que hay diferencia significativa entre ellas. Hay dos formas básicas de poder efectuar estas comparaciones :

Comparaciones “a priori”: Son aquellas comparaciones planificadas previamente, durante la etapa del diseño experimental. Es decir, las que el experimentador cree que va a encontrar diferencias significativas, *antes* de hacer el experimento.

Comparaciones “a posteriori”: Son aquellas comparaciones no planificadas de antemano. Surgen a partir de los datos experimentales, cuando el investigador descubre diferencias inesperadas y quiere testearlas.

Para explicar mejor el funcionamiento de este tipo de comparaciones se desarrolla un ejemplo de aplicación :

Un laboratorio investiga la composición de un medicamento nuevo, para tratar cierta enfermedad infecciosa. Decide testear 3 drogas A, B y C con la misma composición porcentual en la fórmula y un cóctel hecho con dos de ellas A y B con la mitad del porcentaje para cada una. Para ello, busca 50 pacientes escogidos al azar entre los que padecen la enfermedad con grados similares de avance y pertenecientes a un mismo estrato social y educacional. Lo que se mide es la cantidad de días que tardan en curarse completamente. Para los 5 casos a estudiar, escoge al azar 10 pacientes. Al primer grupo les suministra un placebo, al segundo la droga A, al tercero la B, al cuarto el cóctel (50% de A y 50% de B) y al quinto la droga C. Los resultados obtenidos se muestran en el cuadro siguiente:

Cuadro 18.1 : Ejemplo de Anova de 1 factor. (Datos del Cuadro 17.6)

Nº	Placebo	Droga A	Droga B	Mezcla A+B	Droga C	Totales
1	125	107	108	108	112	
2	117	108	111	109	116	
3	120	110	106	108	115	
4	125	109	108	111	113	
5	115	112	107	107	114	
6	121	110	106	106	112	
7	117	110	111	108	115	
8	117	107	110	107	115	
9	126	109	107	107	112	
10	118	111	108	109	117	
Total	1201	1093	1082	1080	1141	5597
Media	120,1	109,3	108,2	108	114,1	
Σx^2	144383	119489	117104	116658	130217	627851
$(1/n)(\Sigma x)^2$	144240,1	119464,9	117072,4	116640	130188,1	627605,5

$$(1/N)(\Sigma \Sigma x)^2 = (5597)^2/50 = \boxed{626528,18}$$

Se trata de un diseño experimental de un Modelo I de Anova de un factor. Se calculan los totales por grupo, y la media de cada uno. El gran total es $T = 5.597$ días. La suma de los cuadrados de cada grupo y el total de la suma de cuadrados $\sum \sum x^2 = 627.851$. Además, para cada grupo se calcula el cuadrado de su total dividido su tamaño $(1/n)(\sum x)^2$, su suma y, finalmente, el término de corrección $T^2 / N = (1/n)(\sum \sum x)^2 = (5.597)^2 / 50 = 626.528,18$. (Ver Cuadro anterior). Con todos estos datos se pueden calcular las sumas de cuadrados así:

$$SS_t = \sum \sum x^2 - T^2 / N = 627.851 - 626.528,18 = 1.322,82$$

$$SS_e = \sum (1/n)(\sum x)^2 - T^2 / N = 627.605,5 - 626.528,18 = 1.077,32$$

$$SS_d = SS_t - SS_e = 1.322,82 - 1.077,32 = 245,5$$

Con estos datos se arma la Tabla de Anova, para detectar si hay diferencias entre las medias muestrales, se compara el estadígrafo F contra los valores críticos de tabla.

Cuadro 18.2 : Tabla de Anova del Cuadro 18.1.

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre grupos	1.077,32	4	269,33	49,33***
Dentro grupos	245,5	45	5,46	

Total 1.322,82 49

NOTA: A las sumas de cuadrados se las divide por sus grados de libertad respectivos y se obtienen los cuadrados medios. El estadígrafo F se obtiene haciendo el cociente de los dos cuadrados medios.

Para los tres niveles usuales: $F_{0,05(4,45)} = 2,58$; $F_{0,01(4,45)} = 3,77$; $F_{0,001(4,45)} = 5,57$. Como $F = 49,33***$ se rechaza la hipótesis nula porque se tiene evidencia altamente significativa de diferencia entre las medias muestrales.

A esta altura de los cálculos se han descubierto diferencias altamente significativas entre los 5 grupos testeados. Pero no se puede establecer entre cuales grupos en una forma individual, que es justamente lo que más interesa. Por lo tanto, se necesitan modelos estadísticos que permitan comparar a los grupos entre sí. Esto es, los *modelos de comparaciones múltiples*. Para el problema que se está estudiando, el diseñador del experimento busca comparar varias cosas: (1) El grupo control contra los grupos a los cuales se le suministró el medicamento. (2) Luego quiere comparar las drogas puras contra el cóctel de dos de ellas y (3) Comparar las drogas entre sí. Esto lo diseñó antes de hacer las mediciones, por lo tanto corresponde usar un modelo “a priori”.

Hay dos formas de hacer estas comparaciones. La primera es hacer todas las comparaciones posibles entre dos grupos, usando el modelo t-Student aplicado al caso de dos muestras independientes, como se vio en el capítulo anterior. La segunda forma es descomponer la suma de cuadrados del tratamiento en comparaciones separadas, usando el modelo F-Fisher como continuación del Anova. Esta última, es la manera adecuada porque es la más sencilla y elegante de ambas. Matemáticamente ambos casos son equivalentes.

La regla general para realizar las comparaciones múltiples “a priori” es sencilla:

Paso 1) Para comparar k grupos de tamaño n_i se toma la suma del grupo elegido para comenzar las comparaciones T_1 , se lo eleva al cuadrado y se lo divide por su tamaño muestral, resulta entonces:

$$S_1 = T_1^2 / n_1 = (1201)^2 / 10 = 144.240,1 \quad (\text{grupo placebo})$$

Paso 2) Se suman los totales de los grupos restantes $T_r = \sum_{i=2}^k T_i = T_2 + T_3 + \dots + T_k$

$$T_r = (1093 + 1082 + 1080 + 1141) = 4396$$

Paso 3) Se eleva al cuadrado este último total y se lo divide por el tamaño muestral de todos esos grupos, calculando así el segundo término buscado:

$$S_2 = (T_r)^2 / (n_2 + n_3 + \dots + n_k) = (4396)^2 / 40 = 483.120,4$$

Paso 4) Se calcula el término de corrección, sumando todos los grupos componentes de la comparación, elevándola al cuadrado y dividiendo por el tamaño total de las muestras integrantes

$$T = \sum_{i=1}^k T_i = T_1 + T_2 + \dots + T_k = 5597 \quad S_3 = T^2 / (n_1 + n_2 + \dots + n_k) = T^2 / N$$

$$S_3 = (5597)^2 / 50 = 626.528,18$$

Paso 5) Se calcula la suma de cuadrados para la comparación buscada como:

$$SS (\text{placebo vs. medicamentos}) = S_1 + S_2 - S_3 = 832,32$$

Para este caso hay un grado de libertad, por lo tanto $MS = SS / v = 832,32$

Paso 6) Se calcula el estadígrafo $F = MS (\text{placebo vs. medicamentos}) / MS \text{ dentro}$

$$F = 832,32 / 5,46 = 152,44$$

Paso 7) Se compara con los valores de tabla para niveles del 95%, 99% y 99,9% para rechazar la hipótesis nula en la forma acostumbrada.

$$F_{0,05; (1,45)} = 4,05 \quad F_{0,01; (1,45)} = 7,23 \quad F_{0,001; (1,45)} = 12,4$$

En conclusión, se rechaza la hipótesis nula con resultados altamente significativos $F = 152,44^{***}$

Esto significa, que se tiene evidencia científica muy fuerte de las diferencias entre el grupo de control y los demás grupos, se puede concluir que el efecto del medicamento sirve para curar la enfermedad infecciosa (hay validación estadística). Este era el objetivo principal de la investigación, pero ahora, se puede continuar más allá, gracias a las bondades de este modelo. Por ejemplo, se puede investigar si hay diferencias entre el grupo al cual se le suministró el cóctel de droga y los demás que fueron tratados con las drogas puras. Para ello, se repite la misma técnica anterior, pero usándola con los datos de los grupos remanentes.

Paso 1) Se calcula la primera suma de cuadrados con los datos del grupo que recibió el cóctel de drogas:

$$S_1 = T_2^2 / n_2 = (1080)^2 / 10 = 116.640$$

Paso 2) Se suman los totales de los grupos restantes $T_r = \sum_3^k T_i = T_3 + \dots + T_k$

$$T_r = (1093 + 1082 + 1141) = 3316$$

Paso 3) Se eleva al cuadrado este último total y se lo divide por el tamaño muestral de todos esos grupos, calculando así el segundo término buscado:

$$S_2 = (T_r)^2 / (n_3 + \dots + n_k) = (3316)^2 / 30 = 366.528,53$$

Paso 4) Se calcula el término de corrección, sumando todos los grupos remanentes de la comparación, elevándola al cuadrado y dividiendo por el tamaño total de las muestras integrantes

$$T' = \sum_2^k T_i = T_2 + T_3 + \dots + T_k = 4396 \quad S_3 = T'^2 / (n_2 + \dots + n_k) = T'^2 / N$$

$$S_3 = (4396)^2 / 40 = 483120,4$$

Paso 5) Se calcula la suma de cuadrados para la comparación buscada como:

$$SS (\text{cóctel vs. drogas puras}) = S_1 + S_2 - S_3 = 48,13$$

Para este caso hay un grado de libertad, por lo tanto $MS = SS / v = SS = 48,13$

Paso 6) Se calcula el estadígrafo $F = MS (\text{cóctel vs. drogas puras}) / MS \text{ dentro}$

$$F = 48,13 / 5,46 = 8,82$$

Paso 7) Se compara con los valores de tabla para niveles del 95%, 99% y 99,9% para rechazar la hipótesis nula en la forma acostumbrada.

$$F_{0,05; (1,45)} = 4,059 \quad F_{0,01; (1,45)} = 7,2525 \quad F_{0,001; (1,45)} = 12,45$$

En conclusión, se rechaza la hipótesis nula con resultados muy significativos $F = 8,82^{**}$

Se tiene evidencia científica muy fuerte de las diferencias entre el grupo al que se le suministró el cóctel y los restantes grupos, se puede concluir que el efecto del medicamento mezclado difiere muy significativamente de los puros. Por último, se puede hacer una comparación más entre los diferentes tipos de drogas. Lo que se busca ahora, es detectar si hay diferencia entre ellas y para esto se repite la técnica anterior, usando los tres grupos remanentes.

Paso 1) Para comparar los 3 grupos remanentes:

$$S_1 = (1082)^2 / 10 = 117.072,4 \quad S_2 = (1093)^2 / 10 = 119.464,9 \quad S_3 = (1141)^2 / 10 = 130.188,1$$

Paso 2) Se calcula el término de corrección, sumando todos los grupos componentes de la comparación, elevándola al cuadrado y dividiendo por el número total de muestras integrantes:

$$T'' = \sum_3^k T_i = T_3 + T_4 + T_5 = 3316 \qquad S_4 = T''^2 / (n_3 + n_4 + n_5)$$

$$S_4 = (3316)^2 / 30 = 366.528,53$$

Paso 3) Se calcula la suma de cuadrados para la comparación buscada como:

$$SS \text{ (entre drogas puras)} = S_1 + S_2 + S_3 - S_4 = 196,87$$

Para este caso hay dos grados de libertad, por lo tanto $MS = SS / v = SS / 2 = 98,435$

Paso 4) Se calcula el estadígrafo $F = MS \text{ (entre drogas puras)} / MS \text{ dentro}$

$$F = 98,435 / 5,46 = 18,03$$

Paso 5) Se compara con los valores de tabla para niveles del 95%, 99% y 99,9% para rechazar la hipótesis nula en la forma acostumbrada.

$$F_{0,05; (2,45)} = 3,204 \qquad F_{0,01; (2,45)} = 5,110 \qquad F_{0,001; (2,45)} = 8,085$$

En conclusión, se rechaza la hipótesis nula con resultados altamente significativos $F = 18,03^{***}$

Se tiene evidencia científica muy fuerte de las diferencias entre los diferentes tipos de drogas investigados. Todo esto se puede resumir en un Cuadro de Anova similar.

Cuadro 18.3: Comparaciones múltiples (Datos del ejemplo en Cuadro 18.1).

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre grupos	1077,32	4	269,33	49,33***
Control vs. Drogas	832,32	1	832,32	152,44***
Cóctel vs. Drogas	48,13	1	48,13	8,82**
Drogas entre sí	196,87	2	98,44	18,03***
Dentro grupos	245,5	45	5,46	
Total	1322,82	49		

Observando la tabla anterior puede notarse, que la suma de cuadrados entre grupos (1077,32) se descompuso en tres términos, al igual que los grados de libertad respectivos. Lo que permite efectuar tres ensayos de hipótesis extras y sacar las conclusiones correspondientes. No se puede descomponer en más de cuatro comparaciones porque no alcanzarían los grados de libertad, para poder hacer una comparación múltiple de tipo *ortogonal*. Cuando se quieran hacer más comparaciones, hay que emplear modelos *no ortogonales* como el de Dunn-Sidak que el lector interesado puede encontrar en la bibliografía como Biometry de Sokal-Rohlf (Edición 1981).

18.3 Comparaciones múltiples “ a posteriori ”

Cuando se desean hacer comparaciones sobre la base de los resultados obtenidos, es decir, cuando las comparaciones no fueron planificadas de antemano por el experimentador, hay que usar los modelos “a posteriori”. En estos modelos se deben distinguir los dos casos posibles:

Modelos “a posteriori” con tamaños muestrales iguales:

- Modelo de Tukey (*T-method*).
- Modelo de Tukey corregido (*T'-method*) cuando las muestras son aproximadamente iguales.
- Modelo de Welsch (*Welsch-method*).
- Modelo de Dunn-Sidák.

Modelos “a posteriori” con tamaños muestrales distintos:

- Modelo de Hochberg (*GT2-method*).
- Modelo de Tukey – Kramer (*TK-method*).
- Modelo de Student – Neumann – Keuls (*SNK-method*).
- Modelo de Scheffé.
- Modelo de Gabriel (*SS-STP method*) para hacer todas las comparaciones posibles.

Todos estos métodos son llamados también modelos *no ortogonales* y pueden encontrarse en la bibliografía. Las diferencias entre uno y otro se basan en la potencia para discriminar diferencias significativas entre dos medias muestrales. Tienen ventajas y desventajas que los investigadores en Estadística siguen discutiendo, en un campo abierto hasta hoy. Para ser consecuente con el objetivo de esta obra, que busca la sencillez a fin de facilitar la comprensión del lector y la simplicidad de cálculos, se ha decidido desarrollar a dos de ellos. Uno para el caso de tamaños iguales de muestras (Tukey) y el otro para tamaños distintos. Pero en ambos, se prefiere mostrar las comparaciones gráficas sugeridas por Gabriel (1978 y 1980) en lugar de los tediosos cálculos para comparar en una tabla todos los casos posibles.

La idea básica comienza en el modelo de Student para comparar dos medias muestrales independientes:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{DS_1^2 / n_1 + DS_2^2 / n_2}} \quad \text{versus} \quad t_{\alpha; n-1}$$

Como se hace el supuesto de homoscedasticidad, resulta $DS_1^2 = DS_2^2 = DS^2$

Si los tamaños de muestras son iguales entonces:

$$\sqrt{DS_1^2 / n_1 + DS_2^2 / n_2} = \sqrt{DS^2 / n + DS^2 / n} = DS \cdot \sqrt{2/n}$$

Pero si los tamaños muestrales son diferentes, como el cuadrado medio intra-grupo MS_d , resulta ser una especie de promedio entre todas las varianzas muestrales, se puede obtener la expresión:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{2 \cdot MS_d / n}} \quad \text{versus} \quad t_{\alpha; n-a}$$

Si se supone que los valores esperados son iguales ($\mu_1 = \mu_2$) y en el caso límite, cuando el valor del estadígrafo t es igual al de tablas la comparación anterior queda:

$$(\bar{x}_1 - \bar{x}_2)_{\text{límite}} = t_{\alpha; n-a} \sqrt{2 \cdot MS_d / n} = \text{LSD} \quad (\text{para un ensayo de dos colas})$$

Donde LSD es una especie de “*mínima diferencia significativa*”. Es un valor único, que una vez calculado, le permite al investigador revisar cualquier pareja de medias, para decidir si difieren significativamente entre sí. Por ejemplo, para el problema presentado en el Cuadro 18.1:

$$\text{LSD} = t_{0,05;45} \cdot \sqrt{2 \cdot (5,46) / 10} = 2,017 \cdot 1,045 = 2,107$$

Por lo tanto, cualquier pareja de medias que difieran en 2,107 son significativamente diferentes una de otra, con probabilidad $p < 0,05$. Aplicando estos conceptos al problema, se vuelven a comprobar las diferencias testeadas (usando el promedio entre las drogas puras).

Sin embargo, debe recalarse que estas comparaciones *son válidas, sólo si fueron planificadas de antemano*. Para poder incluir las comparaciones no planificadas, se puede buscar una fórmula equivalente a la anterior, tal como:

$$\text{MDS} = (\text{valor crítico de tablas}) \cdot (\text{error típico})$$

Donde MDS es la *mínima diferencia significativa*, usando una distribución estadística apropiada para el caso a decidir con un test y de donde obtener un valor crítico. Con un error típico de referencia para el estadígrafo elegido. Los diferentes modelos mencionados más arriba, buscan optimizar la distribución siguiendo varias líneas de pensamiento. Pero, el criterio más difundido es emplear el “*test del rango studentizado*” cuya tabla de valores críticos puede verse en la Tabla 18 del Anexo. Entonces, para esa distribución estadística queda:

$$\text{MDS} = Q_{\alpha(a; n-a)} \sqrt{MS_d / n}$$

Donde la tabla de rangos studentizados posee dos argumentos: a = número de grupos analizados y el número de grados de libertad ($n-a$). Para un nivel de significación como $\alpha = 0,05$ elegido, se puede calcular el valor del MDS para el ejemplo de la investigación de la Tabla 18.1:

$$\text{MDS} = Q_{0,05(5; 45)} \sqrt{5,46 / 10} = 4,02 \sqrt{0,546} = 2,969$$

Para hacer las comparaciones “*a posteriori*” se mostrará el método de Tukey con la sugerencia gráfica desarrollada por Gabriel en el Cuadro 18.4 siguiente. Para no tener que hacer todos los pares de combinaciones posibles en una tabla engorrosa, Gabriel sugirió calcular los límites superior e inferior de cada media muestral estudiada, determinando así un intervalo de confianza dentro del cual, el verdadero valor puede estar en cualquier parte, con una confianza obtenida a

partir del nivel de significación α elegido por el investigador. Luego se grafican estos intervalos y cuando no se superpongan entre sí, la diferencia entre las medias será significativa. De esta forma, con un simple vistazo, se pueden detectar las diferencias buscadas.

Cuadro 18.4 : Comparaciones no planificadas para tamaños muestrales iguales.

Usando los valores del problema de la Tabla 18.1 se procede como sigue:

Paso 1 : Se calcula el valor de la mínima diferencia significativa MSD con :

$$MDS = Q_{\alpha(a; n-a)} \sqrt{MS_d / n} \quad \text{En este caso es:}$$

$$MDS = Q_{0,05(5; 45)} \sqrt{0,546} \quad \text{Interpolando entre 40 y 60 para los grados de libertad resulta:}$$

$$MDS = 2,9691$$

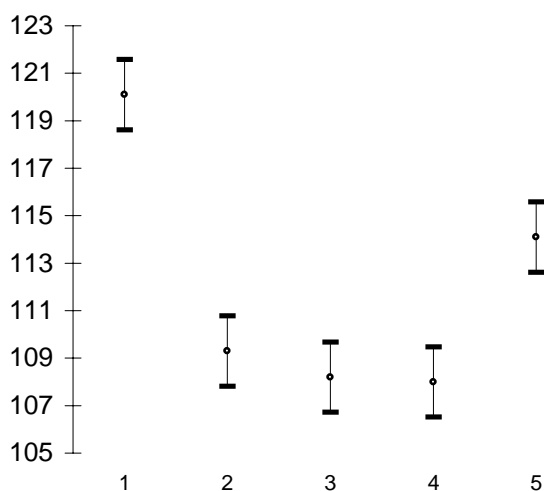
Paso 2 : Se calculan los límites superior e inferior para cada media muestral, sumando y restando la mitad de MSD a cada media muestral respectivamente, con las relaciones:

$$LSi = \bar{x}_i + MSD/2 \quad \text{y} \quad Lli = \bar{x}_i - MSD/2 \quad \text{Notar que: 95\%CI (Lli ; LSi)}$$

	Placebo	Droga A	Droga B	Mezcla A+B	Droga C
LSi	121,58	110,78	109,68	109,48	115,58
Medias	120,1	109,3	108,2	108	114,1
Lli	118,62	107,82	106,72	106,52	112,62

Paso 3 : Se representan estos valores como se muestra en el Gráfico 18.1 siguiente. Ahí se puede ver claramente que el intervalo del Grupo1 (placebo) no se superpone con los demás y se concluye que difiere significativamente de los otros. Queda demostrado que el medicamento reduce el tiempo de cura. Además, el grupo 5 (droga C) difiere de los demás, o sea las drogas A y B. Se concluye que el medicamento con droga C tiene efectividad, pero no tanta como las otras drogas. Como los intervalos entre las drogas A, B y su cóctel se superponen, se puede deducir que no hay diferencia entre ellos. Por lo tanto, la sospecha de que el cóctel era mejor que cada una por separado no se pudo confirmar.

Gráfico 18.1 : Comparaciones entre medias muestrales de la tabla anterior.



18.4 Tamaños muestrales distintos

Cuando se desean hacer comparaciones con tamaños muestrales distintos, el método sigue siendo el mismo en el caso de las comparaciones planificadas, porque al descomponer el Anova en partes, se tiene en cuenta el tamaño muestral individual, como se mostró antes. Pero, para el caso de comparaciones “a posteriori” el método desarrollado en el punto anterior deja de ser válido. De todos los modelos propuestos para trabajar en este caso, de tener diferentes tamaños muestrales se elige el método GT2 simplificado. Gabriel en 1978 propuso hacer más sencillos los engorrosos cálculos del método de Hochberg (GT2), presentado en 1974, con una manera gráfica aproximada. Al igual que en el punto anterior, se pueden visualizar fácilmente las diferencias significativas, observando los intervalos que no se superponen. De hecho, los cálculos son muy similares, excepto que se usa otra distribución estadística, denominada *Máximo módulo studentizado de dos colas*, que se muestra en la Tabla 19 del Anexo.

En este caso, la Mínima Diferencia Significativa es del tipo:

$$MDS_{ij} = (\text{valor crítico de tablas}) \cdot (\text{error típico}) = M_{\alpha(k; v)} \sqrt{DS_i^2 + DS_j^2}$$

Donde $M_{\alpha(k; v)}$ se obtiene de la Tabla 19 con $k = (a/2)(a-1)$ y $v = N - a$
 Por su parte el error típico se puede relacionar con el MS_d resultando:

$$MDS_{ij} = M_{\alpha(k; v)} \sqrt{(MS_d / n_i + MS_d / n_j)}$$

Con estos datos se arma una tabla de todas las comparaciones posibles para poder sacar las conclusiones necesarias. Sin embargo, como esto sería muy engorroso, se prefiere desarrollar el método simplificado por Gabriel, en un ejemplo a continuación:

En una investigación sobre gastos farmacéuticos semanales, relacionado con los enfermos internados, en 8 hospitales de una provincia; se eligieron 128 casos, tratando de mantener la proporción de sexos y con similares características de internación y con edades similares. Los datos recogidos expresados en cientos de pesos, se muestran en el Cuadro 18.5 :

Cuadro 18.5: Gastos farmacéuticos en internados (por \$100).

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	Hospital 6	Hospital 7	Hospital 8
Medias	352.4	282.2	312.4	295.4	276.2	252.4	234.6	263.8
n_i	20	18	18	12	16	14	18	12
DS_i^2	45.8	40.4	20.5	30.4	28.8	30.4	20,8	20.4

Paso 1) Se calcula la varianza ponderada usando las muestrales con :

$$MS_d = \frac{\sum_1^8 (n_i - 1) \cdot DS_i^2}{\sum_1^8 (n_i - 1)} = \frac{(20-1)45,8 + (18-1)40,4 + \dots + (12-1)20,4}{(19+17+17+11+15+13+17+11)} = 3.645,1 / 120 = 30,4$$

Los grados de libertad asociados son $v = N - a = (128 - 8) = 120$

Paso 2) Buscando en tablas con $k = (a/2)(a-1) = (8/2)(8-1) = 28$ y $v = 120$ resulta:

$$M_{\alpha(k;v)} = M_{0,05(28;120)} = 3,934$$

Paso 3) Se calculan los límites superior e inferior para cada media muestral, para armar los intervalos de comparación, en forma similar a la vista. Cuando dos intervalos no se superpongan significará que hay diferencia significativa entre las dos muestras.

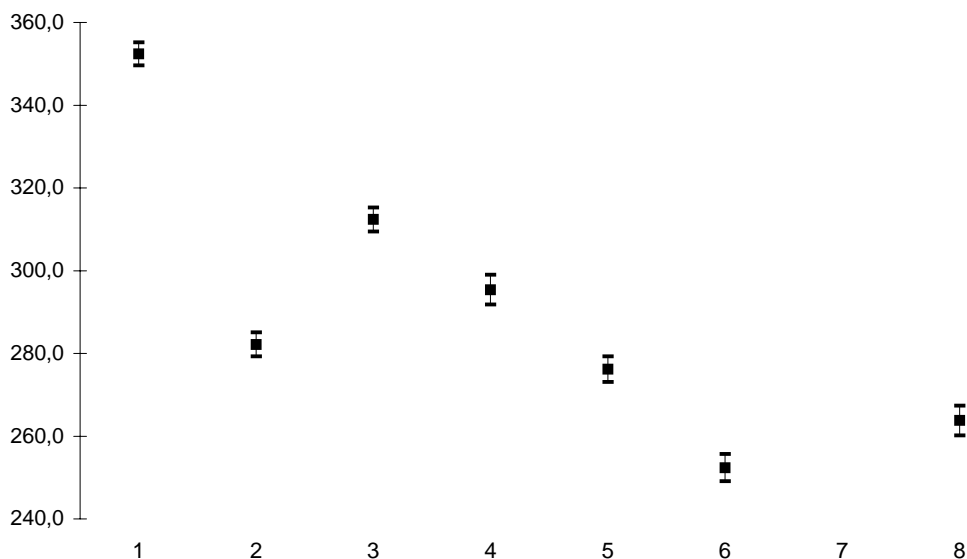
$$LSi = \bar{x}_i + E_i \quad y \quad Lli = \bar{x}_i - E_i \quad \text{Con}$$

$$E_i = (1/\sqrt{2}) \cdot M_{\alpha(k;v)} \cdot \sqrt{MS_d / ni} = 0,7071 \cdot 3,934 \cdot \sqrt{30,4} \cdot \sqrt{1/ni} = 15,34 \sqrt{1/ni}$$

Con esta relación se pueden ahora calcular, los intervalos de confianza para una de las medias muestrales, como se muestra a continuación con los 95% CI (Lli ; LSi):

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	Hospital 6	Hospital 7	Hospital 8
Ei	3,43	3,62	3,62	4,43	3,84	4,10	3,62	4,43
LSi	355,83	285,82	316,02	299,83	280,04	256,50	238,22	268,23
Medias	352,40	282,20	312,40	295,40	276,20	252,40	234,60	263,80
Lli	348,97	278,58	308,78	290,97	272,36	248,30	230,98	259,37

Paso 4) Se grafican los límites de los intervalos, para detectar superposiciones:



Si bien la mayoría son diferentes, se pueden destacar dos cosas, mientras en el Hospital 1 los gastos son máximos, en el Hospital 6 son mínimos. Con esa información, el investigador encontró uno de los motivos principales de las diferencias: En el primer caso los médicos no recibían según el Vademécum, mientras que en el Hospital 6 sí. Lo cual le permitió reducir los gastos en forma sencilla y rápida.

18.5 Diseño básico de experimentos

Cuando se desea hacer un experimento se debe identificar claramente la magnitud clínica en los objetivos del mismo. Si es más de una magnitud se deben emplear los análisis de regresión y correlación que se verán más adelante o el análisis multivariado que está más allá de los objetivos de la presente obra. Pero ahora, ya se está en condiciones de discutir la parte básica.

Las “reglas de oro” a tener en cuenta son:

- *Cuanto mayor sea el número de muestras, más confiables serán las conclusiones.*
- *Siempre que se pueda, se deben tomar muestras de igual tamaño.*
- *Se debe tratar de usar modelos paramétricos en lugar de no paramétricos (el más potente).*
- *Cuando haya que investigar a más de dos grupos, diseñar con modelos de Anova y planificar de antemano las comparaciones entre muestras a realizar.*

La primera regla apunta a mejorar el nivel de confianza y a potenciar el modelo elegido usado para la toma de decisiones. La segunda busca simplificar los cálculos porque las fórmulas empleadas en los casos de diferente tamaño se complican demasiado. La tercera se debe al hecho que muchas veces, en un mismo experimento, se pueden emplear varios modelos estadísticos y siempre los paramétricos son mucho más potentes que sus equivalentes no paramétricos.

La cuarta regla indica que hay que tratar de usar los modelos de Anova cada vez que se pueda. En efecto, la capacidad que tienen de descomponer la varianza total del experimento en términos aditivos es decisiva. Toda medición estará afectada de una cierta indeterminación llamada error y otro tanto ocurre con los grupos de mediciones del experimento. Nunca puede saberse a ciencia cierta las causas que originan la variabilidad en los resultados experimentales. Todo lo que puede hacerse es acotarla a términos manejables en la práctica; por ejemplo para medir la longitud de una mesa basta una regla común, no es necesario hacerlo en décimas de milímetro con otra regla especial. Sin embargo, a veces, el investigador necesita saber cuanta variabilidad genera el efecto de un cierto factor. La manera tradicional de hacerlo, era dejar fijos o constantes todos los demás factores que él sospecha como causas de variabilidad, y en forma controlada hace variar a uno solo de ellos para estudiarlo.

Por ejemplo, si se desea conocer la influencia del factor humano en las mediciones, se usa siempre el mismo equipo, los mismos reactivos, el mismo método, una única muestra (fraccionada en alícuotas de ser necesario), a la misma temperatura y humedad ambiental, etc. Y con mediciones repetidas de la misma muestra, se hace operar a un individuo, luego a otro y así hasta el último. De esta forma, si todo queda casi constante, la única fuente de variabilidad será la de usar diferentes personas. Antes, se les hacía hacer una medición a cada uno y la varianza total era una forma de cuantificar la variabilidad del factor humano más la inherente al sistema de medición que es inevitable. Luego, si una sola persona repetía la misma cantidad de mediciones, la varianza total obtenida se la podía suponer como la del error casual. Si se comparaban ambas se-

ries de mediciones se podía tener una idea de las diferencias en exactitud con el modelo Student, y las diferencias en precisión comparando las varianzas con el modelo de Fisher.

Pero desde la aparición del Anova toda esta operatoria fue dejada un lado. Por ejemplo, en el mismo experimento ahora se hace realizar a cada uno de los individuos la misma cantidad de mediciones junto a la persona de control, y con el modelo de Anova de un factor se puede descomponer la varianza total en dos partes: una debida al factor humano (MSe) y la otra, remanente (MSd) debida al error. Con un Modelo II se puede estimar la relación porcentual entre ambas, pero con un Modelo I se puede todavía comparar al control contra las otras personas y además a éstas entre sí. Todas estas ventajas, se logran con el mismo gasto y con la misma duración del experimento que a la manera tradicional. El secreto está en un buen diseño del experimento y en la planificación de las comparaciones, antes de realizarlo.

Aún más, se puede comparar más de un factor a la vez usando otros modelos de Anova que se verán más adelante. En el ejemplo anterior, además de estudiar la influencia del factor humano, se puede analizar la influencia de otros factores como: tipo de instrumento, marca de reactivos, técnicas diferentes, etc. Y para ello se pueden usar Modelos de Anova Multifactoriales con o sin replicación.

18.6 Comparaciones no paramétricas

Cuando se encuentren diferencias significativas, con el modelo de Kruskal-Wallis, como en el ejemplo del capítulo anterior, se debe continuar el estudio con las comparaciones múltiples para el caso de Estadística No Paramétrica. En este punto se mostrará el modelo STP para el caso de muestras de igual tamaño, basado en el estadístico U de Mann-Whitney. Si las muestras son de diferente tamaño, se puede emplear el test de la U para dos muestras y trabajar sobre todos los pares posibles. Otra vez se tiene una ventaja con muestras de igual tamaño, porque se tiene un modelo que simplifica los cálculos, de por sí muy engorrosos. Se explicará este modelo con el ejemplo anteriormente mencionado.

Paso 1) Se ordenan los datos por grupo en orden creciente, como se muestra abajo:

Cuadro 18.6: Los datos del Cuadro 17.6 ordenados en forma creciente.

Nº	Placebo	Droga A	Droga B	Mezcla A+B	Droga C
1	115	107	106	106	112
2	117	107	106	107	112
3	117	108	107	107	112
4	117	109	107	107	113
5	118	109	108	108	114
6	120	110	108	108	115
7	121	110	108	108	115
8	125	110	110	109	115
9	125	111	111	109	116
10	126	112	111	111	117
Total	1201	1093	1082	1080	1141

Paso 2) Para cada par de muestras hay que calcular el valor Up de la manera siguiente: hay que contar para cada dato de una muestra, el número de datos de la otra muestra que tienen un valor más bajo, se le asigna un punto a cada uno, y el número de datos empatados con él a los que se le asigna ½ punto. Por ejemplo, si se comparan las muestras placebo y droga A, se puede notar que el valor más chico del grupo placebo, es mayor cualquiera de los de A. Por lo tanto si se toma el primer valor de Placebo: 115, hay 10 valores de A menores que él y le corresponden 10 puntos. Luego se toma el segundo valor 117 y ocurre lo mismo, por lo que son 10 puntos más y así sucesivamente para los restantes, se tendrá un total de 10. 10 = 100 puntos. Otro tanto ocurre si se compara el placebo contra la droga B y contra el cóctel de drogas A+B. Esto es,

$$C(\text{placebo-droga A}) = 10 + 10 + \dots + 10 = 100 \text{ puntos} = C(\text{placebo-droga B}) = C(\text{placebo-cóctel})$$

Pero con la droga C, la cosa cambia, ahora si se toma el valor más chico del placebo: 115 se puede ver que hay 5 valores más chicos y 3 valores iguales, o sea, es un total de 5+1,5 puntos para ese primer caso. Para el segundo dato :117 se tienen 9 valores más chicos y un empate, o sea un total de 9,5 puntos. Para el tercer y cuarto dato ocurre lo mismo. Para los restantes datos todos los valores son menores y le tocan 10 puntos. Esto es,

$$C(\text{placebo-droga C}) = 6,5 + 9,5 + 9,5 + 9,5 + 10 + 10 + 10 + 10 + 10 = 95 \text{ puntos}$$

Análogamente, tomando ahora el grupo de la droga A, se lo va comparando contra los otros tres grupos como sigue:

$$\begin{aligned} C(\text{droga A-droga B}) &= 3 + 3 + 5,5 + 7 + 7 + 7,5 + 7,5 + 7,5 + 9 + 10 = 67 \text{ puntos} \\ C(\text{droga A-droga A+B}) &= 2,5 + 2,5 + 5,5 + 8 + 8 + 9 + 9 + 9 + 9,5 + 10 = 73 \text{ puntos} \\ C(\text{droga A-droga C}) &= 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 1,5 = 1,5 \text{ puntos} \\ C(\text{droga B-droga A+B}) &= 0,5 + 0,5 + 2,5 + 2,5 + 5,5 + 5,5 + 5,5 + 9 + 9,5 + 9,5 = 50,5 \text{ puntos} \\ C(\text{droga B-droga C}) &= 0 \text{ (todos los datos de droga C son mayores)} \\ C(\text{droga A+B-droga C}) &= 0 \end{aligned}$$

Paso 3) Ahora se determina el valor Up correspondiente para cada comparación C realizada en el paso anterior, eligiendo el mayor de los dos casos posibles: $Up = C$ ó $Up = n^2 - C$ y se vuelca en una tabla como la siguiente:

Cuadro 18.7 : Puntajes del estadígrafo Up .

Up	Placebo	Droga A	Droga B	Cóctel A+B	Droga C
Placebo	/				
Droga A	100**	/			
Droga B	100**	67	/		
Cóctel A+B	100**	73	50,5	/	
Droga C	95**	98,5**	100**	100**	/

El valor crítico para Up es :

$$U_{\alpha(a;n)} = (n^2 / 2) + n \cdot Q_{\alpha(k;\infty)} \sqrt{\frac{(2n+1)}{24}}$$

Donde se usa la tabla del rango studentizado para $k = 5$ y un número infinito de grados de libertad, donde es $Q_{0,05}(5; \infty) = 3,858$ y $Q_{0,01}(5; \infty) = 4,603$; $n = 10$. Luego:

$$U_{0,05}(5; 10) = (100 / 2) + 10 \cdot 3,858 \sqrt{\frac{(2(10)+1)}{24}} = 50 + 36,089 = 86,089 \approx 86,1$$

$$U_{0,01}(5; 10) = (100 / 2) + 10 \cdot 4,603 \sqrt{\frac{(2(10)+1)}{24}} = 50 + 43,057 = 93,057 \approx 93,1$$

Se comparan estos valores con los de la tabla anterior y las diferencias muy significativas se las indica con un doble asterisco. Mirando la tabla en sentido vertical se observa que en la primera columna, el placebo difiere significativamente de los otros cuatro grupos. La droga A solo difiere de la droga C. La droga B difiere del cóctel y de la droga C. El cóctel difiere de la droga C. Mirando en sentido horizontal se puede ver que la droga C difiere del resto, el cóctel del placebo y B, la B del placebo lo mismo que la A.

Otra forma de ver las diferencias en su conjunto es con:

$$\mu_{\text{placebo}} \gg \mu_{\text{droga C}} \gg \mu_{\text{droga A}} \approx \mu_{\text{droga B}} \approx \mu_{\text{cóctel A+B}}$$

Donde el símbolo $>$ representa diferencias significativas y el símbolo \gg se usa para las diferencias muy significativas. Como se puede ver, hay mucha coincidencia con los resultados del modelo paramétrico visto más arriba. Sin embargo, esto no siempre es así, pues por lo general este modelo no discrimina tan bien, las diferencias entre medias muestrales como lo hace su equivalente modelo paramétrico.

18.7 Bioequivalencia

Los llamados tests de *Bioequivalencia* (o pruebas de equivalencia) han tenido un desarrollo reciente en Bioestadística. La diferencia de este tipo de test con los vistos anteriormente, es que aquí interesa ver si el efecto de un tratamiento clínico cualquiera, es superior en promedio al del placebo. O sea, en lugar de plantear una hipótesis nula habitual de que no hay diferencia entre las medias de los grupos, para poder así rechazarla cuando haya evidencia suficiente, la estrategia en los test de equivalencia consiste en suponer que la diferencia entre las medias es igual o mayor un valor de la magnitud clínica que se considera inferior clínicamente hablando.

De esta forma, la hipótesis alternativa queda: *La diferencia de medias es menor que una cantidad clínicamente relevante, y por lo tanto los tratamientos pueden ser considerados equivalentes*. Notar que, en caso de rechazar la hipótesis nula, habrá validación estadística de Bioequivalencia entre los métodos. Por ejemplo, se tiene un tratamiento viejo y aparece uno nuevo con ciertas ventajas como de menor costo, más rápido, menos invasivo, etc. Entonces el problema reside en ver si ambos métodos son equivalentes desde un punto de vista clínico, para poder efectuar el reemplazo del viejo con el nuevo.

Ejemplo: Se desea medir el índice cardíaco (CI), que es la relación entre el flujo sanguíneo expresado en litros por minuto, y la superficie de cuerpo del paciente. Esta magnitud continua se mide con un método invasivo llamado: termo dilución (TD), que consiste en colocar un catéter en el corazón para realizar la medición. Se propone un nuevo método llamado la bioimpedancia (BI), que consiste en adherir un instrumento externo en la piel del cuerpo del paciente, el cual mide una diferencia eléctrica. La gran ventaja es que este método es no invasivo. Se busca decidir si hay Bioequivalencia entre ambos métodos para poder efectuar el reemplazo. El valor referencial es de 2,75 litros/minuto/m². Cuando el paciente tiene un valor muy por encima o por debajo, significa que hay un problema con el flujo del torrente sanguíneo. Esta y otras pruebas ayudan al médico a decidir el tratamiento adecuado para cada caso. El método TD es más confiable que el método BI, por lo tanto se adopta el siguiente criterio:

Si el nuevo método BI, cae dentro del 20% del valor del viejo método (TD), entonces ambos métodos pueden ser considerados equivalentes.

Se sabe que el valor de corte es 2,75 litros/minuto/m². El 20% de este valor es $\delta = 0,55$, entonces se puede plantear la siguiente hipótesis nula:

$H_0 : | \mu_{TD} - \mu_{BI} | \geq 0,55$ - El método BI cae fuera del intervalo admisible

$H_1 : | \mu_{TD} - \mu_{BI} | < 0,55$ - El método BI cae dentro de la dispersión admisible del 20%

Se miden un total de $N = 68$ individuos con el BI y resulta un promedio de 2,68 litros/minuto/m², con un desvío de 0,26 litros/minuto/m², entonces el error estándar de estimación será:

$$DS (BI) = 0,26 / (68)^{1/2} = 0,0265 \text{ litros/minuto/m}^2,$$

Y el test para probar la H_0 es de una sola cola para $v = 68 - 1 = 67$ grados de libertad. De tablas se obtiene el valor Student con $t_{0,95; 67} = 1,67$, entonces es:

$$t = (2,75 - 2,68) / 0,0265 = 2,64 > 1,67 \text{ por lo tanto se rechaza la hipótesis nula.}$$

Esto significa que se tiene prueba estadística para validar la H_1 , y se pueden considerar equivalentes a ambos métodos. O bien, hay Bioequivalencia entre el método TD y el BI. Notar que otra manera de hacer lo mismo es calcular:

$$| \mu_{TD} - \mu_{BI} | + t_{0,95; 67} DS / \sqrt{N} = 0,07 + 1,67 \cdot 0,0265 = 0,114 < 0,55$$

Lo que muestra que la diferencia encontrada es menor que la máxima admisible de 0,55 y por lo tanto hay Bioequivalencia.

Cuando la pregunta sea si $\mu_{TD} > \mu_{BI}$ entonces usando las hipótesis anteriores se trata de un test de una sola cola, en cambio si la pregunta es $\mu_{TD} ? \mu_{BI}$ se tratará de un test de dos colas y dependiendo de ello se usarán distintos valores de tablas. Sin embargo, la forma más sencilla es usando el intervalo de confianza, abierto para una cola y cerrado para dos. En el caso de intervalo abierto solo interesa el valor límite como se mostró más arriba: si es menor hay equivalencia y si es mayor no la hay.

18.8 Problemas propuestos

- | | | |
|---|---|---|
| 1) Tiene sentido hacer comparaciones múltiples, cuando el Anova es significativo. | V | F |
| 2) Las comparaciones múltiples se pueden usar en el Modelo I y II de Anova. | V | F |
| 3) Los modelos de comparaciones múltiples solo sirven para comparaciones planeadas. | V | F |
| 4) En comparaciones no planificadas solo se pueden usar modelos no Paramétricos. | V | F |
| 5) La componente añadida de varianza se usa en los Modelos II de Anova. | V | F |
| 6) La componente añadida es un porcentaje de la variabilidad total del experimento. | V | F |
| 7) Conviene planificar las comparaciones antes de hacer el experimento. | V | F |
| 8) El modelo de Tukey-Kramer se usa cuando hay tamaños muestrales diferentes. | V | F |
| 9) Los modelos de Gabriel se usan en un gráfico para simplificar las comparaciones. | V | F |
| 10) Con el modelo Student se pueden comparar hasta dos medias muestrales. | V | F |
| 11) La MDS es el producto entre un valor de tablas y el error típico del modelo. | V | F |
| 12) El MDS para muestras de diferente tamaño no es una constante. | V | F |
| 13) Para aumentar la confiabilidad hay que aumentar el número de muestras. | V | F |
| 14) Conviene usar los modelos Paramétricos siempre que se pueda. | V | F |
| 15) Para simplificar los cálculos conviene tomar muestras de igual tamaño. | V | F |
| 16) Conviene diseñar el experimento usando la estrategia del Anova. | V | F |
| 17) No se puede medir la incidencia del factor humano en un experimento. | V | F |
| 18) El modelo de Kruskal-Wallis es el Anova de 1 factor de tipo no paramétrico. | V | F |
| 19) No se pueden hacer comparaciones múltiples no paramétricas. | V | F |
| 20) Explicar el concepto de Bioequivalencia | | |

21) En un experimento sobre la acción de un medicamento contra el Colesterol, se midió la reducción del mismo al cabo de una semana de iniciado el tratamiento. Determinar si hay diferencia significativa entre los cuatro grupos testeados:

Nº	40-44 años	45-49 años	50-54 años	55 años y más
1	189	175	186	100
2	238	205	190	128
3	250	198	170	110
4	190	200	160	139
5	210	180	195	154

- a) En caso de que corresponda hacer comparaciones múltiples si se planificó comparar los mayores de 54 años contra todos los demás, y luego los más jóvenes contra los dos intermedios.
 b) Buscar todas las comparaciones posibles con un modelo paramétrico y otro no paramétrico.

22) En un experimento de Bacharat (1940) sobre el factor de difusión de una sustancia, la cual si está presente en el momento de la inoculación incrementa la mancha que esta produce en la piel de un conejo. Se eligieron al azar 6 localidades geográficas diferentes y a 6 conejos de cada uno. La idea era estudiar si la ubicación geográfica tenía influencia. Se midió el tamaño de la marca en cm^2 . Los datos obtenidos fueron los siguientes:

Fuente de	Suma de	Grados de	Cuadrados	
-----------	---------	-----------	-----------	--

Variación	cuadrados	Libertad	Medios	F
Entre conejos	12.8333	5	2.56673	4,39
Dentro de conejos	17.5266	30	0.5832	
Total	30.2599	35		

Hallar la componente añadida de varianza.

23) Averiguar si las medias son diferentes en las tasas de citocromo-oxidasa:

-24 hs. después de una inyección de metaxyclo n = 5 ; media = 24,8 ; varianza = 0,81

-Control n = 3 ; media = 19,7 ; varianza = 1,96

(aplicada al macho de la especie *Periplaneta*).

24) Aplicar los modelos de comparaciones múltiples a los problemas planteados al final del capítulo anterior, cuando los resultados sean significativos. Discutir acerca del significado de los resultados encontrados para cada caso.

25) Se ha medido un suero control de cierta magnitud clínica con cinco técnicas diferentes, provenientes cada una de un laboratorio diferente. Se trata de establecer si los laboratorios miden lo mismo, o si hay diferencia entre ellos. El suero control tiene $\mu = 135$ mg/dl con un desvío estándar $DS = 2$ mg/dl y ha sido enviado por el laboratorio de referencia. Se lo fracciona en 50 alícuotas y se lo reparte equitativamente entre los 5 laboratorios que participan de la experiencia. Ninguno de ellos sabe que se trata del mismo suero, ni tampoco el valor del mismo. Solo se les informa que tienen que medir 1 muestra ciega de suero cada vez que se les envía. Se eligen al azar 10 días diferentes del mes, para efectuar el envío. No interesa ver la diferencia entre los días del mes, sino las diferencias que se observan de las mediciones repetidas. Luego de realizada la experiencia los valores informados son:

Nº	Labor. 1	Labor. 2	Labor. 3	Labor. 4	Labor. 5
1	140	133	128	136	145
2	142	132	127	135	142
3	138	134	129	137	143
4	139	136	132	134	141
5	141	137	130	135	142
6	137	135	133	136	140
7	136	138	131	134	141
8	139	134	129	136	143
9	140	136	132	135	144
10	141	135	133	134	140

Se pide:

- Detectar si hay diferencia entre los laboratorios con un Cuadro de Anova
- Mediante el método de Tukey compara los laboratorios entre sí.
- Usando el método gráfico de Gabriel presentar todas las comparaciones.
- Con el valor patrón decidir cuales laboratorios están calibrados en exactitud.
- Con un desvío máximo aceptable del 5%, controlar la precisión de cada laboratorio.

- f) Introduciendo el valor patrón en el gráfico de Gabriel ver la calibración en exactitud y comparar estas conclusiones con las del punto (d).
- g) Discutir sobre el sentido de las diferencias halladas y como deberían ser comunicadas a cada laboratorio involucrado.

26) Decidir si hay diferencia entre las drogas testeadas de acuerdo a los resultados siguientes

Nº	Placebo	Droga 1	Droga 2	Droga 3
1	450	350	250	280
2	460	360	240	290
3	455	370	245	295
4	470	360	250	285
5	465	380	255	280

Usando cualquiera de las técnicas vistas en este capítulo para efectuar comparaciones múltiples:

- a) Las comparaciones planeadas fueron:
 - Placebo versus drogas
 - Droga 1 versus las otras dos
 - Droga 2 versus droga 3

- a) Usar los métodos de Gabriel para comparaciones no planeadas.

28) En un estudio de control de parásitos, cada rata fue inyectada con 500 larvas del parásito llamado *Nippostrongylus muris*. Diez días después fueron sacrificadas y se contó el número de gusanos adultos. Se quiere dilucidar la siguiente cuestión: ¿Hay diferencia en la resistencia a la invasión parasitaria por grupos de ratas suministradas por diferentes proveedores? Se analizaron 4 grupos diferentes, formados por 5 ratas cada uno. Los resultados fueron:

Nº	Grupo 1	Grupo 2	Grupo 3	Grupo 4
1	279	378	172	381
2	338	275	335	346
3	334	412	335	340
4	198	265	282	471
5	303	286	250	318

19

Análisis de Anova encajado

En este capítulo se tratan los modelos de ANOVA de un factor que tienen más de un nivel de análisis. No se trata de dos factores por separado, sino que dentro del factor principal de análisis, se consideran otros “factores” subordinados. Para evitar la confusión con los términos, se prefiere decir que hay dentro del factor un *nivel* principal y dentro de éste, uno o más subniveles de análisis. Es como si estos últimos estuviesen encajados en el primero. Esto permite un poco más de sofisticación en el estudio de las variabilidades que afectan a los experimentos. En la primera parte se explican los diferentes modelos de ANOVA encajado, de acuerdo a como son los modelos empleados en cada nivel. Luego se presentan las formas cortas de cálculo y ejemplos de aplicación.

19.1 Introducción

El modelo de Anova de un factor visto en los capítulos precedentes, es con mucha frecuencia insuficiente para abarcar toda la complejidad de un experimento dado y para extraer del mismo toda la información importante. En el presente capítulo se verá el caso en que cada grupo de muestras está subdividida en clases escogidas aleatoriamente. Por ejemplo, en un experimento donde se estudiarán 5 caldos de cultivo, se pueden preparar 3 cajas de Petri para cada caldo, y en cada una de estas, incubar la muestra extraída a un paciente. Así, los caldos de cultivo serán los grupos principales y las cajas los subgrupos dentro de estos. La ventaja respecto al Anova de un factor, es que ahora se pueden observar las diferencias entre los 3 tipos de siembra para estar seguros que las diferencias observadas entre las cajas se deben a la diferente composición de los caldos. Podría darse que diferencias morfológicas o ambientales y otras accidentales, entre los diferentes microbios cultivados en las diferentes cajas, ocasionen diferencias entre el resultado de los cultivos al final de la experiencia. La única manera de separar estos dos efectos es tener dos o más cajas de Petri para cada caldo de cultivo, y con el Anova, separar las variabilidades de estos efectos. Si no se encuentran diferencias significativas entre las cajas, se pueden atribuir las diferencias observadas a los distintos tipos de caldos de cultivo. Todavía, si se encuentran diferencias significativas entre las cajas, con la componente añadida de varianza entre las cajas, se puede comparar la variación entre los caldos para ver si son significativas, sobre la base de la variabilidad entre cajas tratadas de la misma manera.

A este tipo de análisis, se lo conoce como *Anova encajado o jerárquico*, pues una clasificación subordinada está encajada dentro del nivel más alto de clasificación. En el nivel subordinado los grupos siempre deben ser escogidos en forma aleatoria. Por lo tanto, siempre se tratará

de un Modelo tipo II de Anova en los subniveles. En cambio, en el nivel principal los grupos se pueden escoger en forma aleatoria (*Anova encajado puro*), o bien, en forma especial como en un Modelo tipo I de Anova (*Anova encajado mixto*). Además, se pueden seguir subdividiendo las clases en otros subgrupos y tener así más de dos niveles. Por ejemplo, para el estudio de una droga aplicada a seres humanos se pueden preparar 5 dosis y un placebo (clásico Modelo I) para el primer nivel. Dentro de cada grupo se pueden escoger 10 pacientes en forma aleatoria para conformar el segundo nivel de análisis. De cada paciente se pueden tomar tres muestras para evaluar la variabilidad de lecturas repetidas y conformar así el tercer nivel subordinado, para cuantificar la parte de la varianza debida al “error” de medición. Pero en el segundo nivel se puede determinar la variabilidad entre pacientes y en el primero la debida a las dosis. Otros usos importantes del modelo puro se ven frecuentemente en Genética cuantitativa, Sistemática, etc.

19.2 Modelos teóricos y formas de cálculo

Los datos medidos en un Anova encajado, pueden explicarse en forma similar a la vista como Modelos I y II del Anova de un factor. Con las diferencias siguientes:

$$X_{ijk} = \mu + A_i + B_{ij} + \varepsilon_{ijk} \quad \text{Modelo de Anova encajado puro}$$

Donde: X_{ijk} : es el k-ésimo dato medido, de la subclase j, del grupo i

μ : es la media Paramétrica de la población

A_i : es la contribución aleatoria del efecto del grupo i

B_{ij} : es la contribución aleatoria del efecto del subgrupo j, del grupo i

ε_{ijk} : es el error de la medición k, en subgrupo j, del grupo i

Se supone que A_i se distribuye en forma gaussiana con media 0 y varianza σ_i^2 ; mientras que B_{ij} también se distribuye normalmente con media 0 y varianza σ_{BCA}^2 , donde se indica que la varianza es del subgrupo B dentro del grupo A. Por su parte ε_{ijk} tiene media 0 y varianza σ^2

Mientras que para el modelo mixto la relación anterior se transforma en :

$$X_{ijk} = \mu + \alpha_i + B_{ij} + \varepsilon_{ijk} \quad \text{Modelo de Anova encajado mixto}$$

Donde: X_{ijk} : es el k-ésimo dato medido, de la subclase j, del grupo i

μ : es la media Paramétrica de la población

α_i : es la contribución fija del efecto del grupo i

B_{ij} : es la contribución aleatoria del efecto del subgrupo j, del grupo i

ε_{ijk} : es el error de la medición k, en subgrupo j, del grupo i

Esta fórmula es la misma que la vista más arriba, solo que el efecto del factor principal es fijo en lugar de aleatorio, esto es que α_i es una constante para todo los datos del grupo i.

Para presentar las formas de cálculo en este modelo de Anova, se muestra el siguiente:

Ejemplo 1) Una empresa farmacéutica tiene sucursales en todo el país. Luego de lanzar una nueva línea de productos de cosmética decide analizar la variabilidad de sus ventas. Para ello escoge

3 sucursales al azar y en cada una de ellas se sortearon 4 días cualquiera del mes. Como se trabaja 24 horas en 3 turnos, para separar la varianza error se sortearon 2 turnos en cada caso. Su propósito al tomar 4 días al azar, diferentes en cada sucursal, es ver si las diferencias entre sucursales, añaden varianza a la variabilidad total de las ventas debidas a razones estacionales y de mercado. La razón es mejorar la planificación general de las compras para no quedarse con un stock excesivo de mercadería. Como las sucursales (nivel 1) se escogieron al azar, lo mismo que los días del mes (subnivel 2) y los turnos (mediciones repetidas dentro del subnivel 2). Se trata claramente de un *Modelo de Anova encajado puro*. Muy diferente hubiera sido si las sucursales las elige a propósito para ver si hay diferencia significativa entre ellas; en tal caso sería un modelo mixto. Los datos obtenidos son el monto total de venta diario, expresados en pesos, de la nueva línea de cosmética. Notar que se podría tomar las ventas mensuales de cada sucursal durante un semestre y diseñar un modelo simple de Anova. Pero el diseño escogido no necesita esperar varios meses para trabajar los datos, sino que tomando una sola quincena ya es suficiente para comenzar a seguir el proceso de venta de la nueva línea lanzada.

Cuadro 19.1: Ventas diarias en \$ de una nueva línea cosmética en 3 sucursales

	Suc. 1				Suc. 2				Suc. 3			
	1	2	3	4	1	2	3	4	1	2	3	4
1	175,5	233,4	252	210,3	209,4	168	152,1	191,4	169,8	233,1	209,7	187,2
2	178,5	242,7	250,8	204,9	209,4	163,5	147,9	197,4	172,5	237,0	207,6	193,5
T _{jk}	354	476,1	502,8	415,2	418,8	331,5	300	388,8	342,3	470,1	417,3	380,7
T _k	1748,1				1439,1				1610,4			

Paso 1) Se calculan los totales diarios T_{jk} de las mediciones repetidas (X_{ijk}) y el total por sucursal. Entonces hay N = a . b . n = 3 . 4 . 2 = 24 datos y es:

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n \sum_{l=1}^2 X_{ijkl} = \sum_{i=1}^a T_{ik} = 1748,1 + 1439,1 + 1610,4 = 4797,6 = T$$

Paso 2) Se calcula la suma del cuadrado de todos los datos:

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}^2 = (175,5)^2 + (178,5)^2 + \dots + (193,5)^2 = 980.569,44 = T_x^2$$

Paso 3) Se calcula la suma de los cuadrados de los totales de los subgrupos, dividido por el tamaño muestral respectivo (en este caso todos los tamaños son iguales a 2):

$$\sum_{i=1}^a \sum_{j=1}^b T_{jk}^2 / n_{jk} = (1/n) \sum_{i=1}^a \sum_{j=1}^b T_{jk}^2 = (1/2)[(354)^2 + (476,1)^2 + \dots + (380,7)^2] = 980.436,15 = T_b^2$$

Paso 4) Se calcula la suma de los cuadrados de los totales de los grupos, dividido por su tamaño muestral respectivo (en este caso todos los tamaños son iguales a n.b = 8)

$$\sum_{i=1}^a T_k^2 / n_k = (1/n.b) \sum_{i=1}^a T_k^2 = (1/8)[(1748,1)^2 + (1439,1)^2 + (1610,4)^2] = 965.031,32 = T_a^2$$

Paso 5) Se calcula el término de corrección T² / N = TC = (4797,6)² / 24 = 959.040,24

Paso 6) Se calculan las sumas de cuadrados necesarias

$$SS_{\text{total}} = T_x^2 - TC = (\text{Paso 2} - \text{Paso 5}) = (980.569,44 - 959.040,24) = 21529,2$$

$$SS_{\text{grupos}} = T_a^2 - TC = (\text{Paso 4} - \text{Paso 5}) = (965.031,32 - 959.040,24) = 5991,08$$

$$SS_{\text{subgrupos}} = T_b^2 - T_a^2 = (\text{Paso 3} - \text{Paso 4}) = (980.436,15 - 965.031,32) = 15404,83$$

$$SS_{\text{dentro}} = T_x^2 - T_b^2 = (\text{Paso 2} - \text{Paso 3}) = (980.569,44 - 980.436,15) = 133,29$$

Paso 7) Se calculan el cuadro de Anova encajado como sigue:

TABLA DE ANOVA

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre grupos	5991,08	(a-1) = 2	MSe /a=(5991,08/2)=2995,54	MSe / MSeg = 1,75 (ns)
Entre subgrupos	15404,83	a(b-1) = 9	MSeg /a(b-1)=(15404,83/9)=1711,7	MSeg / MSd =154,2***
Dentro subgrupos	133,29	ab(n-1) = 12	MSd/ab(n-1)=(133,29/12) =11,1	
Total	21529,2	23		

Si en lugar del modelo encajado, se hubiera hecho el modelo unifactorial simple, la suma de cuadrados entre las sucursales (entre grupos) sería la misma $SS_{\text{entre suc.}} = 5991,08$. Y la suma de cuadrados error, dentro de los grupos, sería la diferencia entre la total y la que hay entre las sucursales, $SS_{\text{error}} = (21529,2 - 5991,08) = 15538,12$. Esto significa que, el estadígrafo $F = (5991,08 / 2) / (15538,12 / 21) = 4,05^*$ sería significativo.

Cabe preguntarse la razón por la cual, en este modelo hay diferencias significativas entre las sucursales y en el encajado no. La respuesta es que en el ANOVA simple la variabilidad remanente o error contiene la debida a la aleatoriedad de ventas diarias entre los turnos. Y al no hacer esa discriminación el análisis se torna confuso, muestra una significación disfrazada porque el cuadrado medio remanente o error ha cambiado. Ahora se puede discriminar mejor y se puede ver que la mayor parte de la variabilidad se debe a la que hay entre los turnos. A medida que el investigador subdivida en más niveles, mejor podrá discriminar las cosas. *La idea básica en diseño de ANOVA es tratar de reducir lo más posible el MS remanente debido al error.*

La riqueza del modelo encajado reside, en este caso, en que la suma de cuadrados remanente o error (dentro de las sucursales) puede descomponerse en dos partes: una al considerar los diferentes turnos (entre turnos, dentro de las sucursales) y la nueva remanente más chica. Hay menor error (indeterminación) y se pueden encontrar mejores explicaciones a las cosas. Se debe continuar con un análisis de la componente añadida de varianza:

Paso 8) Se calculan las componentes añadidas con:

La variabilidad error (dentro de subgrupos) es: $\sigma^2 = 11,1$

La variabilidad entre subgrupos, es decir entre turnos dentro de las sucursales es:

$$\sigma_{BcA}^2 = (MS_{\text{subgrupos}} - MS_{\text{error}}) / n = (1711,7 - 11,1) / 2 = 850,3$$

La variabilidad entre grupos (sucursales) es:

$$\sigma_A^2 = (MS_{\text{grupos}} - MS_{\text{subgrupos}}) = (2995,54 - 1711,7) / 8 = 160,48. \quad \text{Luego,}$$

$$\text{Variabilidad total} = \sigma^2 + \sigma_{BcA}^2 + \sigma_A^2 = 11,1 + 850,3 + 160,48 = 1021,88. \quad \text{Entonces:}$$

σ^2 % representa el $(11,1 / 1021,88) \cdot 100 = 1,08\%$ de la variabilidad total (el error).

σ_{BcA}^2 % representa el $(850,3 / 1021,88) \cdot 100 = 83,21\%$ del total (turnos dentro de sucursales).

σ_A^2 % representa el $(160,48 / 1021,88) \cdot 100 = 15,71\%$ del total (entre las sucursales).

19.3 ANOVA encajado de tres niveles

Para este modelo aparece un nuevo subnivel y entonces, otra vez, quedan dos tipos de planteos o diseños experimentales posibles: el Modelo puro y el mixto.

$$X_{ijkl} = \mu + A_i + B_{ij} + C_{ijk} + \varepsilon_{ijkl} \quad \text{Modelo de Anova encajado puro}$$

Donde: X_{ijk} : es el k-ésimo dato medido, de la subclase j, del grupo i

μ : es la media Paramétrica de la población

A_i : es la contribución aleatoria del efecto del grupo i

B_{ij} : es la contribución aleatoria del efecto del subgrupo j, del grupo i

C_{ijk} : es la contribución aleatoria del efecto del sub-subgrupo k, del subgrupo j en grupo i

ε_{ijkl} : es el error de la medición l, en sub-subgrupo k, en subgrupo j, del grupo i

Se supone que A_i se distribuye en forma gaussiana con media 0 y varianza σ_i^2 ; mientras que B_{ij} también se distribuye normalmente con media 0 y varianza σ_{BcA}^2 , donde se indica que la varianza es del subgrupo B dentro del grupo A. Además, C_{ijk} se distribuye normalmente con media 0 y varianza σ_{CcB}^2 (se indica el sub-subgrupo C dentro del subgrupo B). Y por su parte ε_{ijkl} tiene media 0 y varianza σ^2 y se trata de la varianza remanente o error.

Mientras que para el modelo mixto la relación anterior se transforma en :

$$X_{ijkl} = \mu + \alpha_i + B_{ij} + C_{ijk} + \varepsilon_{ijkl} \quad \text{Modelo de Anova encajado mixto}$$

Donde: X_{ijk} : es el k-ésimo dato medido, de la subclase j, del grupo i

μ : es la media Paramétrica de la población

α_i : es la contribución fija del efecto del grupo i

Y los demás términos son similares al visto modelo puro.

19.3.1 Formas cortas de cálculo

Para ilustrar este modelo se irá desarrollando la forma de cálculo con el modo paso a paso, para facilitar su comprensión. Se ha tomado un experimento donde el primer nivel se escoge en forma deliberada, esto es, se trata de un modelo mixto.

Ejemplo 2) Se ha medido el contenido glucosa en sangre humana. Se tiene un grupo control y dos preparados con dos drogas diferentes A y B. Para cada uno de los tres grupos (a = 3), se eligieron dos pacientes al azar (b = 2). A cada paciente se le tomó una muestra en tres horarios diferentes del día escogidos al azar (c = 3). Luego se hicieron tres repeticiones de la determinación para disminuir el error remanente (n = 3). Los datos medidos se muestran en el Cuadro 19.2 siguiente, junto con los cálculos preliminares:

Cuadro 19.2: Contenido de glucosa en sangre

Tratamientos:	Control			Droga A			Droga B											
Pacientes	1			2			1			2								
Horario	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3			
Datos	99	97	95	96	98	94	88	88	86	80	83	80	78	80	78	79	80	78
	98	96	96	94	99	94	87	89	84	81	82	81	79	81	76	80	81	76
	99	98	94	95	98	92	86	86	85	80	80	82	78	80	74	78	82	74
Suma:	296	291	285	285	295	280	261	263	255	241	245	243	235	241	228	237	243	228
Total preparados:	872			860			779			729			704			708		
Total Grupos :	1732			1508			1412											

Paso 1) Se calcula el total de las observaciones realizadas:

$$T = \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=3} X_{ijkl} = 99 + 98 + 99 + 97 + \dots + 76 + 74 = 4.652$$

Paso 2) Se calcula el total del cuadrado de las observaciones realizadas:

$$T_x^2 = \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=3} X_{ijkl}^2 = (99)^2 + (98)^2 + (99)^2 + \dots + (74)^2 = 404.092$$

Paso 3) Se calcula la suma de los cuadrados de los totales de los sub-subgrupos (horarios) dividido por su tamaño muestral (n = 3) respectivo:

$$T_c = (1/n) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \left(\sum_{l=1}^{n=3} X \right)^2 = (1/3) [296^2 + 291^2 + 285^2 + \dots + 228^2] = 404.044,667$$

Paso 4) Se calcula la suma de los cuadrados de los totales de los subgrupos (pacientes) dividido por su tamaño muestral respectivo. (n . c = 9)

$$T_b = (1/n.c) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \left(\sum_{k=1}^{c=3} \sum_{l=1}^{n=3} X \right)^2 = (1/9) [872^2 + 860^2 + 779^2 + 729^2 + 704^2 + 708^2] = 403.905,111$$

Paso 5) Se calcula la suma de los cuadrados de los totales de los grupos (tratamientos) dividido por su tamaño muestral respectivo. (b. n . c = 18)

$$T_a = (1/b.n.c) \sum_{i=1}^{a=3} \left(\sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=3} X \right)^2 = (1/18) [1732^2 + 1508^2 + 1412^2] = 403.757,333$$

Paso 6) Se calcula el término de corrección:

$$T^2 / N = (1/a.b.c.n) \left(\sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=3} X \right)^2 = (1/54) [4.652^2] = 400.761,185$$

Paso 7) Se calcula la suma de cuadrados total:

$$SS_T = \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X^2_{ijkl} - T^2 / N = T X^2 - T^2 / N = \text{Paso 2} - \text{Paso 6}$$

$$SS_T = 404.092 - 400.761,185 = 3.330,815$$

Paso 8) Se calcula la suma de cuadrados entre los grupos (entre tratamientos):

$$SS_E = (1/b.n.c) \sum_{i=1}^{a=3} \left(\sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X \right)^2 - T^2 / N = T_a - T^2 / N = \text{Paso 5} - \text{Paso 6}$$

$$SS_E = 403.757,333 - 400.761,185 = 2.996,148$$

Paso 9) Se calcula la suma de cuadrados dentro de grupos (pacientes dentro de tratamientos):

$$SS_{\text{subgrupos}} = (1/n.c) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \left(\sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X \right)^2 - (1/b.n.c) \sum_{i=1}^{a=3} \left(\sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X \right)^2 = T_b - T_a$$

$$SS_{\text{subgrupos}} = \text{Paso 4} - \text{Paso 5} = 403.905,111 - 403.757,333 = 147,778$$

Paso 10) Se calcula la suma de cuadrados dentro de subgrupos (horarios dentro de pacientes):

$$SS_{\text{sub-subgrupos}} = (1/n) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \left(\sum_{l=1}^{n=2} X \right)^2 - (1/n.c) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \left(\sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X \right)^2 = T_c - T_b$$

$$SS_{\text{sub-subgrupos}} = \text{Paso 3} - \text{Paso 4} = 404.044,667 - 403.905,111 = 139,556$$

Paso 11) Se calcula la suma de cuadrados remanente o error (datos dentro de tratamientos):

$$SS_T = \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \sum_{l=1}^{n=2} X^2_{ijkl} - (1/n) \sum_{i=1}^{a=3} \sum_{j=1}^{b=2} \sum_{k=1}^{c=3} \left(\sum_{l=1}^{n=2} X \right)^2 = T X^2 - T_c = \text{Paso 2} - \text{Paso 3}$$

$$SS_T = 404.092 - 404.044,667 = 47,333$$

Paso 12) Se prepara la Tabla de ANOVA:

Fuente de Variación	v	SS	MS	F
Entre Tratamientos	(a-1) = 2	2996,148	1498.074	30,41*
Pacientes dentro de Tratamientos	a(b-1)=3	147,778	49,259	4,24*
Horarios dentro de pacientes	ab(c-1)=12	139,556	11,629	8,85***
Datos dentro de Tratamientos (error)	abc(n-1)=36	47,333	1,315	

Total 53 3330,815

En la primer columna se indican las diferentes fuentes de variación analizadas. En la segunda columna los grados de libertad respectivos. Las sumas de cuadrados y los cuadrados medios se colocan en la tercer y cuarta columna. Finalmente se calculan los estadígrafos F como sigue:

$$F_a = MS_a / MS_b = 1.498,074 / 49,259 = 30,41^* > F_{0,95; 2; 3} = 9,55$$

$$\text{Pero } 30,41 < F_{0,99; 2; 3} = 30,81$$

Por lo tanto, se encontraron diferencias significativas entre los 3 grupos analizados.

El segundo test se hace para comprobar si hay diferencias significativas entre los pacientes dentro de los tratamientos que recibieron. El estadígrafo es:

$$F_b = MS_b / MS_c = 49,259 / 11,629 = 4,24^* > F_{0,95; 3; 12} = 3,49$$

$$\text{Pero } 4,24 < F_{0,99; 3; 12} = 5,95$$

Se encontraron diferencias significativas entre los pacientes, dentro de los tratamientos hechos.

El tercer test, se efectúa para comprobar si hay diferencias significativas entre los horarios dentro de los pacientes. El estadígrafo es:

$$F_b = MS_c / MS_{\text{error}} = 11,629 / 1,315 = 8,85^{***} > F_{0,999; 12; 36} = 3,76$$

Se encontraron diferencias altamente significativas entre los horarios dentro de pacientes.

Como el primer test resultó significativo es necesario realizar las comparaciones múltiples entre los grupos investigados. Para los demás subniveles, como hay significación, se debe realizar la cuantificación de las componentes añadidas de varianza. En cuyo caso es:

Paso 13) Estimación de las componentes de la varianza:

$$DS^2_{BCA} = (49,259 - 11,629) / 6 = 6,27 \text{ (pacientes dentro de tratamientos: } n = 6)$$

$$DS^2_{CCB} = (11,629 - 1,315) / 3 = 3,438 \text{ (horarios dentro de pacientes: } n = 2)$$

$$DS^2 = 1,315 \text{ (remanente o error)}$$

La variación total de los niveles analizados, se calcula como la suma de las tres anteriores:

$$VAR = DS^2_{BCA} + DS^2_{CCB} + DS^2 = 11,02$$

Paso 14) Se calculan las componentes añadidas de varianza como un porcentaje:

$$(DS^2_{BCA} / VAR) 100 = (6,27 / 11,02) 100 = 56,9\% \text{ (pacientes dentro de tratamientos)}$$

$$(DS^2_{CCB} / VAR) 100 = (3,438 / 11,02) 100 = 31,2\% \text{ (horarios dentro de pacientes)}$$

$$(DS^2 / VAR) 100 = (1,315 / 11,02) 100 = 11,9\% \text{ (error)}$$

La notación con los paréntesis en los primeros pasos anteriores se hizo para denotar la manera de generalizar este tipo de ANOVA a más de tres niveles. Por cada subnivel que se agregue habrá un nuevo subíndice, o sea, una sumatoria más de los valores medidos. Luego el cálculo se hace así: Los dos primeros pasos son análogos, esto es, la suma de todos los valores encontrados para obtener el gran total y la suma de los cuadrados de los datos.

En el tercer paso habrá tantas sumatorias como subíndices (subniveles) haya, luego se coloca el paréntesis en la última de ellas para elevar al cuadrado los subtotales comprendidos. En el cuarto paso se corre el paréntesis una posición a la izquierda para que abarque las dos últimas sumatorias y así elevar al cuadrado los subtotales respectivos. Y así sucesivamente hasta llegar al término de corrección, que es cuando el paréntesis abarca a todas las sumatorias.

Para obtener las Sumas de Cuadrados se comienza restando las cantidades obtenidas en el Paso 2 y el término de corrección, para obtener el SS entre los grupos. A continuación, para el último subnivel, se resta el término de corrección al Paso inmediato anterior a éste. Y así sucesivamente. Por ejemplo, si el Paso 8 es el cálculo del término de corrección, la primera diferencia se hace entre: Paso 8 – Paso 7, la siguiente Paso 7 – Paso 6, luego Paso 6 – Paso 5 y así hasta llegar al Paso 2 – Paso 3 que da la remanente o error.

El Cuadro de ANOVA es enteramente similar al visto, y debe verificarse el total de los grados de libertad volcados, lo mismo que el total de la suma de cuadrados. Los MS se obtienen, como siempre, dividiendo los SS por sus respectivos grados de libertad. Y los estadígrafos F se obtienen dividiendo el MS del primer renglón por el MS del segundo, luego el MS del segundo por el MS del tercero y así sucesivamente. Los demás pasos a seguir dependen de sí se encuentra significación, en la misma forma que la del ejemplo visto.

19.4 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|--|-------|---|
| 1) El ANOVA encajado tienen ventajas respecto al ANOVA simple. | V | F |
| 2) El encajado puede separar la variabilidad debida a los distintos subniveles estudiados. | V | F |
| 3) Mencionar las ventajas de este modelo:..... | | |
| 4) El modelo puro del encajado es de Modelo I en todos los niveles. | V | F |
| 5) Explicar como se diseña un modelo mixto de ANOVA encajado: | | |
| 6) El encajado se llama también jerárquico porque toma en cuenta los distintos niveles. | V | F |
| 7) Explicar teóricamente los modelos puros y mixtos. | | |
| 8) Al volcar los datos conviene realizar los totales preliminares. | V | F |
| 9) El primer paso siempre es encontrar el total de las mediciones efectuadas. | V | F |
| 10) En el segundo paso hay que calcular el total de los cuadrados de las observaciones. | V | F |
| 11) En el tercer paso se calcula la SS entre los grupos analizados (segundo nivel). | V | F |
| 12) Explicar como se generalizan los demás pasos:..... | | |
| 13) Siempre hay que calcular la componente añadida de varianza. | V | F |
| 14) Los estadígrafos F para cada nivel se calculan dividiendo su MS por el MSError. | V | F |
| 15) Hay que comparar F con un valor de tablas, con los grados de libertad respectivos. | V | F |
| 16) Si hay significación en el primer nivel, hay que hacer comparaciones múltiples. | V | F |
| 17) Cuando en los subniveles haya significación se debe calcular la componente añadida. | V | F |
| 18) Explicar como se generaliza este modelo a más de tres niveles:..... | | |

2) Resolver numéricamente el siguiente problema:

Tratamientos:	Control			Droga 1			Droga 2											
Pacientes	1			2			1			2								
Dieta	1	2	3	1	2	3	1	2	3	1	2	3						
Datos	450	431	446	445	440	449	478	470	468	480	477	483	510	518	501	499	512	502
	440	425	442	448	443	450	476	465	463	475	479	485	520	519	509	505	515	505

3) Ídem anterior

Pacientes	1			2			1			2			1			2		
Horario	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Datos	131	131	136	150	140	160	157	154	147	151	147	162	134	138	135	138	139	134
	130	125	142	148	143	150	145	142	153	155	147	152	125	138	136	140	138	127

20

ANOVA para más de un factor

El modelo de ANOVA encajado estudia una sola fuente de variación como factor de análisis, el cuál se subdivide en subgrupos de tipo jerárquico para su análisis como fuentes de variación separadas. Es decir, fuentes o factores encajados dentro de una clasificación superior. En el presente capítulo se verán fuentes de variación de igual rango, desarrollando modelos para analizar los casos de más de un factor. Para comenzar se tratará el caso más común: ANOVA de dos factores, y el caso de ANOVA multifactorial (tres o más factores) se deja la inquietud para el lector interesado en avanzar un poco más que los límites planteados para la presente obra. Por ejemplo, se pueden usar estos modelos en Control de Calidad, así se puede investigar el efecto que tiene en la técnica bioquímica empleada el factor humano y las diferentes marcas de kits comerciales en el mercado, para el caso de dos factores. O bien, incluir diferentes espectrofotómetros como un tercer factor de análisis. El cálculo del modelo de ANOVA de dos factores se puede clasificar en dos casos básicos: el primero, *sin repetición*, es cuando hay un solo dato por cada combinación posible de los dos factores y el segundo, *con repetición*, cuando hay más de un dato para cada caso o combinación de factores posibles.

En el modelo de ANOVA de dos factores con repetición aparece un nuevo concepto estadístico de gran utilidad para el investigador: la *interacción* de ambos factores analizados. Cuando el efecto de ambos factores actuando en conjunto es mayor que el que tienen por separado, se dice que hay *sinergia*, una especie de interacción positiva. En cambio, si el efecto conjunto es menor, se habla de *interferencia*. Esto agrega una nueva hipótesis de trabajo a las cuatro vistas para los modelos de ANOVA: Normalidad, Aleatoriedad, Independencia y Homoscedasticidad. Este quinto supuesto básico es: *No hay interacción*. Por su parte, si no se realiza más de una medición por cada combinación de los dos factores, no se puede detectar la interacción y no se requiere del quinto supuesto.

Al final se muestran algunos casos especiales de uso difundido en la práctica y se explica la manera de generalizar los modelos de ANOVA cuando hay más de dos factores en estudio. En forma paralela se van muestra el modelo no paramétrico de Friedman, para poder trabajar en caso que no se pueda (o no se quiera) verificar los supuestos básicos.

20.1 ANOVA de dos factores con repetición

Antes de comenzar el estudio de este modelo conviene repasar la diferencia conceptual entre un modelo de ANOVA encajado de dos niveles y un ANOVA de dos factores, porque a menudo se los confunde. La manera de diferenciarlos es ver si existe correspondencia entre los grupos o clases. Cuando existe tal correspondencia, entonces el modelo apropiado es el de dos factores. En cambio, cuando no hay correspondencia entre las clases y tal factor representa solamente subdivisiones aleatorias de las clases dentro de otro factor, se trata de un modelo encajado. Para aclarar mejor estas ideas se presenta el Gráfico 20.1 siguiente:

Gráfico 20.1: ANOVA de dos factores versus ANOVA encajado de dos niveles.

ANOVA encajado de dos niveles

Edad	Niños		Jóvenes		Adultos	
Sexo	1	2	1	2	1	2
	*	*	*	*	*	*
	*	*	*	*	*	*
	*	*	*	*	*	*
	*	*	*	*	*	*
	*	*	*	*	*	*
	*	*	*	*	*	*

ANOVA de dos factores

	Edad	Niños		Jóvenes		Adultos	
Sexo	1	*	*	*	*	*	*
		*	*	*	*	*	*
		*	*	*	*	*	*
	2	*	*	*	*	*	*
		*	*	*	*	*	*
		*	*	*	*	*	*

En este caso se hicieron 6 mediciones para cada combinación de edad y sexo. En el caso de un modelo de ANOVA de 2 factores, uno sería la edad, clasificada en 3 grupos: niños, jóvenes y adultos y el segundo factor el sexo. Esto es claramente un modelo I para ambos casos, porque no tendría sentido pensar que se hizo un sorteo para el sexo, entre todos los sexos posibles. Por lo tanto, no puede ser un modelo encajado porque el segundo nivel debe ser un Modelo II. Se podría dibujar la tabla de manera tal de tener al sexo en el primer nivel, y encajado dentro del mismo las edades como segundo nivel. Entonces habría que pensar que se hizo un sorteo entre todas las edades posibles, lo que no tiene sentido. Se puede ver así que en este caso solo se puede tratar de un modelo de dos factores. Otra forma de verlo es considerando que el subgrupo 1 de niños, es por sorteo uno de los posibles, y que no tiene nada que ver con el subgrupo 1 de jóvenes y el subgrupo 1 de adultos, para que sea un modelo encajado. En cambio, para el caso de dos factores, el subgrupo 1 es lo mismo para las tres clases de edades planteadas. Se pueden generalizar estas ideas, pensando que en un ANOVA de dos factores donde cada grupo que se tome debe ser el mismo tipo en todos los grupos del otro. Por ejemplo, el grupo Sexo 2, es igual para los tres grupos de edades. En cambio, en el encajado, puede ser cualquiera que salga por sorteo.

Antes de plantear el modelo teórico conviene ilustrar el procedimiento de cálculo en el modelo de ANOVA de dos factores. Para ello, se eligió un ejemplo con repetición. Se trata de un estudio sobre el tiempo de cura de un dolor de cabeza de acuerdo a dos factores: El factor A es la ingesta de aspirinas para aliviar el dolor y el otro factor B es la ingesta de café con el mismo objetivo. Se tomaron 3 pacientes para cada una de las 4 combinaciones posibles, elegidos al azar, de los factores a estudiar y se le midió el tiempo que tomó el alivio del dolor medido en horas. Los datos se muestran en la Tabla 20.1 siguiente:

Tabla 20.1: ANOVA de dos factores con repetición

		Factor A: Aspirina			
		SI	No	Σ	
Factor B:	SI	1,6	4,3		
		1,9	4,1		
		1,8	3,6		
Café	Σ	5,3	12	17,3	
		3,6	4,3		
		3,2	5,1		
		3,8	4,9		
		Σ	10,6	14,3	24,9
		$\Sigma\Sigma$	15,9	26,3	42,2

Paso 1) Se calcula la suma total de datos $T = \sum_1^a \sum_1^b \sum_1^n X_{ijk} = \sum_1^a \sum_1^b T_{jk} = \sum_1^a T_c = 42,2$

Paso 2) Se calcula la suma del cuadrado de todos los datos:

$$T_x^2 = \sum_1^a \sum_1^b \sum_1^n X_{ijk}^2 = (1,6)^2 + (1,9)^2 + \dots + (4,9)^2 = 163,82$$

Paso 3) Se calcula la suma de los cuadrados de los totales de las combinaciones de factores, dividido por el tamaño muestral respectivo. En este caso todos los tamaños son iguales a 3, hay tres mediciones repetidas por cada celda, combinación de los dos factores. Y hay cuatro celdas en total, cada una con una suma de los datos que contiene, entonces:

$$T_c^2 = \sum_1^a \sum_1^b T_{jk}^2 / n = (1/n) \sum_1^a \sum_1^b T_{jk}^2 = (1/3)[(5,3)^2 + (12)^2 + (10,6)^2 + (14,3)^2] = 162,98$$

Paso 4) Se calcula el término de corrección $T^2 / N = (42,2)^2 / 12 = 148,4$

Paso 5) Se calcula la suma de los cuadrados de los totales del grupo B (café) y se divide por su tamaño muestral respectivo. En este caso hay 6 datos para cada caso y dos subtotaes, luego:

$$T_b^2 = \sum_1^a \sum_1^n T_{ik}^2 / a.n = (1/6)[(17,3)^2 + (24,9)^2] = 153,22$$

Paso 6) Se calcula la suma de los cuadrados de los totales del grupo A (aspirinas) y se divide por su tamaño muestral respectivo. En este caso hay 6 datos por cada caso y dos subtotaes, luego:

$$T_a^2 = \sum_1^b \sum_1^n T_{ik}^2 / b.n = (1/6)[(15,9)^2 + (26,3)^2] = 157,42$$

Paso 7) Se calculan las sumas de cuadrados necesarias

$$SS_{\text{total}} = T_x^2 - T^2 / N = \text{Paso 2} - \text{Paso 4} = 15,42$$

$$SS_{\text{dentro}} = T_x^2 - T_c^2 = \text{Paso 2} - \text{Paso 3} = 0,84$$

$$SS_{\text{filas}} = T_b^2 - T^2 / N = \text{Paso 5} - \text{Paso 4} = 4,82 \quad (\text{café})$$

$$SS_{\text{columnas}} = T_a^2 - T^2 / N = \text{Paso 6} - \text{Paso 4} = 9,02 \quad (\text{aspirinas})$$

$$SS_{\text{interacción}} = T_c^2 - (T^2 / N) - T_b^2 - T_a^2 = (\text{Paso 3} - \text{Paso 4}) - SS_{\text{filas}} - SS_{\text{columnas}} = 0,74$$

La suma de cuadrados total sigue siendo la suma de las sumas de cuadrado entre y dentro de los grupos. Aquí, la suma de cuadrados entre se descompone en tres términos: una debida al grupo A, otra al grupo B y la última debida a la interacción. Entonces:

$$SS_{\text{total}} = (SS_{\text{filas}} + SS_{\text{columnas}} + SS_{\text{interacción}}) + SS_{\text{dentro}}$$

Paso 8) Se calcula el cuadro de ANOVA como sigue:

Tabla 20.2: Cuadro de ANOVA para los datos de Tabla 20.1

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre Filas (café)	4,82	(b-1)=1	4,82	45,9***
Entre columnas(aspirinas)	9,02	(a-1)=1	9,02	85,9***
Interacción	0,74	(a-1)(b-1)=1	0,74	7,05*
Dentro grupos (error)	0,84	a.b(n-1)=8	0,105	
Total	15,42	11		

Los tres estadígrafos de comparación F se obtienen dividiendo cada MS encontrado por el MSerror, a diferencia del modelo encajado donde los cocientes se hacían en cascada. Los valores de tablas son : $F_{0,95; 1; 8} = 5,32$; $F_{0,99; 1; 8} = 11,26$ y $F_{0,999; 1; 8} = 25,42$. Comparando con los tres estadígrafos se deduce que los tres son significativos. La ingesta tanto de aspirinas como de café alivian el dolor de cabeza con resultados altamente significativos. Sin embargo se detectó un efecto de interacción entre ambos factores $F = 7,05 *$, lo que significa que el efecto de ambos factores actuando en forma simultánea, tiene un efecto mayor que si actuaran por separado. Esto invalida el quinto supuesto de Anova, y lo que debe hacerse es usar métodos no paramétricos para completar el estudio.

En este problema la interacción resultó significativa y se debe recordar que representa la influencia de un factor sobre el otro. Los efectos de ambos factores actuando en conjunto no son

simplemente aditivos, sino que cualquier combinación de ellos puede arrojar una contribución conjunta tanto positiva (*sinergia*) como negativa (*interferencia*). La conclusión del ejemplo anterior, es que hay sinergia al tomar un café y una aspirina en forma simultánea para aliviar el dolor de cabeza. El alivio aparece más rápido que si se tomasen por separado.

20.2 Supuestos básicos y modelos teóricos

Los supuestos básicos para los modelos de ANOVA con repetición de dos o más factores son:

- Las muestras deben ser *aleatorias*.
- Las muestras deben provenir de poblaciones *normales* o gaussianas.
- Las muestras deben ser *independientes*.
- Las varianzas de las muestras son iguales (*homoscedasticidad*).
- No debe existir *interacción*.

Pueden darse tres casos:

- Ambos factores fueron diseñados como Modelo I (*Modelo I puro*).
- Ambos factores fueron diseñados como Modelo II (*Modelo II puro*).
- Uno de ellos como Modelo I y el otro como Modelo II (*Modelo Mixto*).

Esta clasificación permite plantear los modelos teóricos respectivos con:

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad \text{Modelo I de ANOVA de dos factores puro}$$

Donde: X_{ijk} : es el k-ésimo dato medido, del grupo j de un factor, del grupo i del otro.

μ : es la media Paramétrica de la población.

α_i : es la contribución fija del efecto del grupo i.

β_j : es la contribución fija del efecto del grupo j.

$(\alpha\beta)_{ij}$: es el efecto de la interacción de los factores A y B.

ε_{ijk} : es el error de la medición k, en grupo i y en grupo j.

En este caso, el ε_{ijk} tiene una distribución normal de parámetros $(0 ; \sigma^2)$. Para el ejemplo anterior resulta que ambos términos fueron significativos. En el caso del factor A (aspirinas) el tiempo de alivio del dolor es menor entre los que la toman respecto de los que no lo hacen. Luego, la contribución fija (α_i) del efecto del factor A es significativa y constante para cada grupo. Por su parte los que toman café (factor B) se alivian más rápido que aquellos que no lo hacen. El hecho que el tiempo de alivio es mayor para los que no toman no es por azar, pues hay diferencias significativas entre los grupos sobre la base de los datos medidos en este experimento. Por lo tanto, la contribución fija (β_j) del factor B es importante. Como hay interacción., o sea $(\alpha\beta)_{ij}$ es significativa para el análisis de los hechos, denota que la combinación de ambos factores produce un alivio más rápido, y es un hecho que debe tomarse muy en cuenta. Como esta interacción es significativa no se cumple el quinto supuesto de los modelos de ANOVA, y para efectuar el análisis no queda más remedio que recurrir a modelos no paramétricos equivalentes.

Si se trata de un modelo II en ambos grupos la expresión anterior resulta ahora:

$$X_{ijk} = \mu + A_i + B_j + \varepsilon_{ijk} \quad \text{Modelo II de ANOVA de dos factores puro}$$

Donde: X_{ijk} : es el k-ésimo dato medido, de la subclase j, del grupo i.

μ : es la media Paramétrica de la población.

A_i : es la contribución aleatoria del efecto del grupo i

B_j : es la contribución aleatoria del efecto del grupo j.

$(AB)_{ij}$: es el efecto de la interacción de los factores i y j.

ε_{ijk} : es el error de la medición k, en grupo j y en grupo i.

El término ε_{ijk} tiene una distribución normal de parámetros $(0 ; \sigma^2)$; el término A_i tiene una distribución normal de parámetros $(0 ; \sigma_a^2)$; el término B_j tiene una distribución normal de parámetros $(0 ; \sigma_b^2)$ y el término de la interacción $(AB)_{ij}$ tiene una distribución normal de parámetros $(0 ; \sigma_{axb}^2)$. Para el tercer caso, de modelo mixto, se trata de un modelo análogo a los dos de más arriba, donde uno de los efectos será una constante para todo el grupo y el otro aleatorio. Por su parte, si no hay repetición sólo se necesitan los cuatro primeros supuestos, ya vistos para el caso de ANOVA de un factor, pues no es posible detectar la interacción cuando hay un solo valor, para cada caso de combinación de factores como se trata en el siguiente punto.

20.3 ANOVA de dos factores sin repetición

A menudo resulta demasiado caro o difícil medir más de un dato por lugar, o bien, las mediciones arrojan valores tan parecidos, que resulta innecesario repetirlas. Ahora la suma de cuadrados entre los grupos, se particiona en dos términos, uno para cada grupo, pues no hay interacción. La diferencia entre estas dos sumas y la total corresponde la suma de cuadrados remanente o error (dentro de los grupos). En los modelos teóricos no aparece el término debido a la interacción y lo demás es similar. Se necesitan los primeros cuatro supuestos básicos y los cálculos matemáticos para hallar la tabla de ANOVA son similares. Entonces, parece más práctico tomar directamente un ejemplo para ilustrar estos modelos, como se muestra a continuación:

Tabla 20.3: Modelo Mixto de ANOVA bifactorial sin repetición.

FACTOR B	Dosis	FACTOR A				Σ
		Marca 1	Marca 2	Marca 3	Marca 4	
	1	71,4	72,0	73,8	74,4	291,6
	2	67,8	67,2	68,7	69,6	273,3
	3	66,6	66,3	66,3	66,6	265,8
	4	63,6	65,4	63,0	63,6	255,6
	5	55,2	57,9	57,0	56,4	226,5
	6	40,5	43,2	42,6	41,4	167,7
	7	29,4	29,7	31,2	28,8	119,1
	8	18,0	18,0	18,9	18,9	73,8
	9	17,4	17,7	18,0	17,4	70,5
	10	16,8	16,8	16,5	16,8	66,9
	Σ	446,7	454,2	456,0	453,9	1810,8

En la Tabla 20.3 se muestran los datos de un experimento donde se prueban 4 tipos de marcas comerciales existentes en el mercado, elegidas al azar entre los 15 antibióticos del mismo estilo disponibles para el público. Se planifican 10 dosis diferentes a ser administradas, de menor a mayor concentración, a pacientes de cierta infección. Y se mide el tiempo que tardan en curar. Como las marcas (factor A) fueron elegidas aleatoriamente, pero no las dosis (factor B), se tiene un Modelo Mixto de Anova de 2 factores sin repetición. El cual se resuelve como sigue:

Paso 1) Se calcula la suma total de datos $T = \sum_1^a \sum_1^b X_{ij} = 1.810,8$

Paso 2) Se calcula la suma del cuadrado de todos los datos:

$$T_x^2 = \sum_1^a \sum_1^b X_{ij}^2 = (71,4)^2 + (67,8)^2 + \dots + (16,8)^2 = 101.077,02$$

Paso 3) Se calcula la suma de los cuadrados de las columnas, dividido por el tamaño muestral respectivo (en este caso $b = 10$):

$$T_a^2 = \sum_1^a T_i^2 / b = (1/b) \sum_1^a T_i^2 = (1/10)[(446,7)^2 + (454,2)^2 + (456)^2 + (453,9)^2] = 81.979,97$$

Paso 4) Se calcula la suma de los cuadrados de las filas, dividido por el tamaño muestral respectivo (en este caso $a = 4$):

$$T_b^2 = \sum_1^b T_j^2 / a = (1/a) \sum_1^b T_j^2 = (1/4)[(291,6)^2 + (273,3)^2 + \dots + (66,9)^2] = 101.051,78$$

Paso 5) Se calcula el término de corrección $T^2 / N = (1.810,8)^2 / 40 = 81.974,92$

Paso 6) Se calcula la suma de cuadrados total con:

$$SS_t = T_x^2 - (T^2 / N) = \text{Paso 2} - \text{Paso 5} = 101.077,02 - 81.974,92 = 19.102,1$$

Paso 7) Se calcula la suma de cuadrados entre columnas (factor A) con:

$$SS_A = T_a^2 - (T^2 / N) = \text{Paso 3} - \text{Paso 5} = 81.979,97 - 81.974,92 = 5,05$$

Paso 8) Se calcula la suma de cuadrados entre filas (factor B) con:

$$SS_B = T_b^2 - (T^2 / N) = \text{Paso 4} - \text{Paso 5} = 101.051,78 - 81.974,92 = 19.076,86$$

Paso 9) Se calcula la suma de cuadrados residual (error) con:

$$SS_{\text{error}} = SS_t - SS_A - SS_b = \text{Paso 6} - \text{Paso 7} - \text{Paso 8}$$

$$SS_{\text{error}} = 19.102,1 - 5,05 - 19.076,86 = 20,19$$

Paso 10) Se arma la Tabla de ANOVA como calculando el estadígrafo F como cociente entre los dos MS respecto del MSerror o remanente. Por su parte los grados de libertad son $a - 1 = 3$ para los días, $b-1 = 9$ para las profundidades y $a \cdot b = 27$ para el error. Los valores de SS se sacan de los pasos anteriores y así:

TABLA DE ANOVA

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre columnas (A)	5,05	$(a-1)=3$	1,68	2,25 (ns)
Entre filas (B)	19.076,86	$(b-1)=9$	2.119,65	2.833,8***
Dentro grupos (error)	20,19	$(N+1-a-b) =27$	0,748	
Total	19.102,1	39		

El cálculo de los grados de libertad es sencillo. Se resta 1 de la cantidad de grupos a y b para poder obtener los grados de libertad respectivos, en este caso resultan ser 3 y 9 respectivamente. El número total de grados de libertad es $N-1 = 39$. Y si se le restan los dos anteriores se obtienen los grados de libertad dentro de las muestras o error: $39-9-3 = 27$ o $40+1-4-10 = 27$.

Hay diferencias altamente significativas entre filas, es decir, el tiempo de cura disminuye a medida que aumenta la dosis. No se detectan diferencias entre las diferentes marcas comerciales porque usan similares componentes activos, por lo que no hace falta calcular la componente añadida de varianza.

Una aplicación frecuente de este modelo sin repetición es el *test repetido de los mismos individuos*, en el cual analiza en forma repetida al mismo grupo de individuos a lo largo de un período de tiempo. Estas personas se escogen en forma aleatoria y sirven como repetición, por eso a este factor se lo diseña como Modelo II. En cambio, el factor tiempo se lo considera como efecto fijo del tratamiento y se lo diseña como Modelo I. Este tipo de test considera que las respuestas de los diferentes individuos son paralelas en el tiempo, se supone entonces que no hay interacción entre el tiempo y los individuos, cuando se trata de hallar la componente añadida de varianza debida al factor aleatorio. Por ejemplo, si se trata de medir el crecimiento de los individuos en períodos fijos de tiempo, o cuando se estudia la aparición de alguna respuesta posterior a un tratamiento aplicado al mismo grupo de personas en experimentos fisiológicos o psicológicos.

Otros usos muy difundidos son: el estudio del crecimiento en términos de inmunidad, después de la aplicación de antígenos, la alteración de respuestas a consecuencia de un tratamiento y las medidas de aprendizaje luego de un cierto número de experimentos. La idea básica en los modelos de más de un factor es poder separar la variabilidad entre grupos, en tantos términos como factores de análisis se usen. De esta forma, se puede tener un panorama más completo y simplificar los experimentos, reduciendo el costo y el tiempo de manera conveniente. Un avance grande, si se piensa en el clásico modelo Student de dos muestras.

20.4 Modelo de bloques aleatorizados

Este tipo de modelos de dos factores se aplica frecuentemente en la investigación agrícola, pero también en otros campos biológicos y en experimentos de laboratorio. Por ejemplo, cuando se aplica un fertilizante a una parcela, se obtiene una medida expresada como el producto de la cosecha, o bien, el sembrar en diferentes tipos de suelo. Análogamente, se puede medir el aumento de peso si se suministra a los animales de cría diferentes tipos de alimento balanceado, o bien, si se consideran diferentes razas para determinar su rendimiento. Para ilustrar el uso de este modelo se eligió su aplicación más frecuente ilustrada en la Figura 20.2 siguiente:

Figura 20.2: Bloques aleatorizados en un campo.

B	A	E	D	C	Bloque 1
B	C	E	D	A	Bloque 2
D	B	E	A	C	Bloque 3
C	B	A	E	D	Bloque 4

Un terreno ha sido subdividido en cuatro parcelas (bloques) de similares características dentro de cada una de ellas como por ejemplo, el tipo de suelos, la inclinación del terreno, etc. A cada parcela se le aplican cinco tratamientos diferentes como ser, diferentes fertilizantes, tratamientos con insecticidas, etc. La idea es comprobar si se detectan diferencias significativas por el efecto de los fertilizantes (tratamiento) y para ello se necesitan por lo menos cinco terrenos útiles. Sin embargo, las cinco lecturas de las cosechas levantadas no dan una idea de las diferencias entre los fertilizantes. Por lo tanto, se necesitan repeticiones para poder obtener un término de error. En el ejemplo graficado se escogieron para ello cuatro parcelas. Entonces a cada parcela se la divide en cinco partes, una para cada fertilizante, y se sorteá el lugar donde será colocado cada uno de ellos. Los resultados se muestran arriba. Como regla general los terrenos dentro de cada parcela deben ser muy semejantes. Entonces, se puede investigar también las diferencias entre bloques para detectar cual tipo de terreno es el más conveniente. Así, el efecto del fertilizante se considera como fijo (Modelo I) y el de los terrenos como aleatorio (Modelo II). Luego, el modelo teórico para este tipo de diseño será:

$$X_{ij} = \mu + \alpha_i + B_j + \epsilon_{ij} \quad \text{Modelo de ANOVA para los bloques aleatorizados}$$

Donde: X_{ij} : es el i -ésimo dato medido, del grupo j .

μ : es la media Paramétrica de la población.

α_i : es la contribución fija del efecto del grupo i (fertilizantes)

B_j : es la contribución aleatoria del efecto del grupo j . (terrenos)

ϵ_{ij} : es el error de la medición i -ésima del grupo j .

Los cálculos y Tabla de ANOVA son semejantes a los vistos más arriba, cuando no hay interacción. En el ejemplo siguiente se ilustra el uso de este modelo. En la Figura 20.3 se mues-

tran los datos de una cosecha de soja, medida en kilos, donde se han empleado 3 fertilizantes en 4 tipos de parcelas de terreno diferentes, caracterizadas por tener distintos tipos de suelos, con diferentes pendientes. Cada parcela se escogió de la forma más homogénea posible y se realizó un sorteo para repartir los fertilizantes dentro de cada una.

Figura 20.3: Bloques aleatorizados en un campo.

925(C) | 958(A) | 986(B) Bloque 1

971(A) | 1051(B) | 952(C) Bloque 2

891(B) | 829(C) | 927(A) Bloque 3

955(C) | 1010(B) | 971(A) Bloque 4

FACTOR B bloques (terrenos)	FACTOR A fertilizantes			Σ
	A	B	C	
1	958	986	925	2869
2	971	1051	952	2974
3	927	891	829	2647
4	971	1010	955	2936
Σ	3827	3938	3661	11426

1) Se calcula la suma total de datos $T = \sum_{i=1}^a \sum_{j=1}^b X_{ij} = 11.426$

Paso 2) Se calcula la suma del cuadrado de todos los datos:

$$T_x^2 = \sum_{i=1}^a \sum_{j=1}^b X_{ij}^2 = (958)^2 + (971)^2 + \dots + (955)^2 = 10.914.748$$

Paso 3) Se calcula la suma de los cuadrados de las columnas, dividido por el tamaño muestral respectivo (en este caso $a = 4$):

$$T_a^2 = \sum_{i=1}^a T_i^2 / a = (1/a) \sum_{i=1}^a T_i^2 = 10.889.173$$

Paso 4) Se calcula la suma de los cuadrados de las filas, dividido por el tamaño muestral respectivo (en este caso $b = 3$):

$$T_b^2 = \sum_{j=1}^b T_j^2 / b = (1/b) \sum_{j=1}^b T_j^2 = 10.900.847$$

Paso 5) Se calcula el término de corrección $TC = T^2 / N = (11.426)^2 / 12 = 10.879.456$

Paso 6) Se calcula la suma de cuadrados total con:

$$SS_t = T_x^2 - TC = \text{Paso 2} - \text{Paso 5} = 35.292$$

Paso 7) Se calcula la suma de cuadrados entre columnas (factor A) con:

$$SS_A = T_a^2 - (T^2 / N) = \text{Paso 3} - \text{Paso 5} = 9.717$$

Paso 8) Se calcula la suma de cuadrados entre filas (factor B) con:

$$SS_B = T_b^2 - (T^2 / N) = \text{Paso 4} - \text{Paso 5} = 21.391$$

Paso 9) Se calcula la suma de cuadrados residual (error) con:

$$SS_{\text{error}} = SS_t - SS_A - SS_b = \text{Paso 6} - \text{Paso 7} - \text{Paso 8} = 4.184$$

Paso 10) Se arma la Tabla de ANOVA como sigue:

TABLA DE ANOVA

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre columnas (A)	9.717	(A-1)=2	4.858	10,23**
Entre filas (B)	21.391	(B-1)=3	7.130	6,97*
Dentro grupos (error)	4.184	(N-A-B)=6	697	
Total	35.292	11		

Hay diferencias significativas entre columnas, esto es entre los diferentes tipos de fertilizantes probados. Para continuar este análisis hay que efectuar las comparaciones múltiples con el método de Tukey. Mientras que entre las filas existen diferencias muy significativas, por lo que hay que calcular la componente añadida de varianza σ_B^2 .

$$\sigma_B^2 = \frac{b(a-1)MS_{\text{error}} + (b-1)MS_B}{ab-1} = (1/11) (8 \times 697 + 3 \times 7.130) = 2.451,45$$

La eficiencia relativa de la varianza error de los bloques aleatorizados completos σ_B^2 , con respecto a la variabilidad remanente o error MS_{error} , puede obtenerse con la relación siguiente:

$$\sigma_B^2\% = \frac{\sigma_B^2}{MS_{\text{error}}} \cdot 100 = (2.451 / 697) \cdot 100 = 351,6\%$$

Esto significa tener una variabilidad debida a las diferencias entre los tipos de terreno, donde se hizo el experimento, tres veces y media mayor que la inexplicada debida al error.

20.5 Modelo de Friedman

Este modelo es de tipo no paramétrico para el caso de los bloques aleatorizados en un ANOVA de dos factores sin repetición. Estos *tests* son para diseños que si fuesen realizados mediante un ANOVA serían Modelo I para un factor, mientras que el otro representa los bloques en el experimento. Para un test que se basa en magnitudes ordinales no sería apropiado un diseño del tipo Modelo II que calcule las componentes añadidas de varianza, pues solo interesa la “localización” de las diversas muestras, dentro de la ordenación total.

Para ver el procedimiento del Modelo de Friedman, se usaran los mismos datos de la medición de cuatro marcas de antibióticos elegidas al azar de entre las disponibles en el mercado, vistos en la Tabla 20.3 anterior:

Paso 1) Se ordenan los datos dentro de cada bloque (marcas) y se calculan los rangos:

FACTOR B	Dosis	FACTOR A				Σ Rij
		Marca A	Marca B	Marca C	Marca D	
	1	10	10	10	10	40
	2	9	9	9	9	36
	3	8	8	8	8	32
	4	7	7	7	7	28
	5	6	6	6	6	24
	6	5	5	5	5	20
	7	4	4	4	4	16
	8	3	3	3	3	12
	9	2	2	2	2	8
	10	1	1	1	1	4

Como se puede ver, la regularidad de los rangos hallados se debe a la variación uniforme del tiempo de cura, a medida que la dosis aumenta.

Paso 2) Para cada fila se suman los rangos respectivos, y los totales Rij se colocan en la última columna.

Paso 3) Se calcula el estadígrafo χ^2 con:

$$\chi^2 = \left\{ \left[\frac{12}{ab(a+1)} \right] \sum_i^a \left(\sum_j^b R_{ij} \right)^2 \right\} - 3b(a+1) = \frac{12}{10(4)(10+1)} (40^2 + 36^2 + \dots + 4^2) - 3(4)(10+1)$$

$$\chi^2 = (12/440) \cdot 6.160 - 132 = 36 ***$$

Este valor hay que compararlo con el de tablas $\chi^2_{0,001;9} = 27,877$ resultando ser altamente significativo, como era de esperar de acuerdo al análisis anteriormente realizado.

20.6 Sensibilidad de los modelos de ANOVA

Otra de las reglas de oro en el diseño de los experimentos dice: *Cuanto más factores se tomen en cuenta en el análisis de los datos, mayor será la sensibilidad alcanzada para detectar las diferencias buscadas.* La estrategia de los modelos de Anova es tratar de descomponer la variabilidad dentro de las muestras (error) lo más posible. Así, cuando se toma un solo factor se tiene la variabilidad total descompuesta en dos términos: entre las muestras y dentro de las mismas. Pero si se toma en cuenta un segundo factor con los mismos datos, ahora la variabilidad dentro de las muestras se descompone en dos términos: una debida al nuevo factor y un nuevo remanente al que se lo denomina error. Esto hace que el error se haga más chico y por lo tanto la sensibilidad del modelo aumenta. Aún más, si el segundo factor se toma con repetición el error se achica todavía más porque hay que descomponerlo en otros dos términos: uno para la interacción y el nuevo remanente (error) que será menor que el anterior. Análogamente, si se toman más de dos factores y con repetición el error seguirá disminuyendo.

Para ilustrar lo anterior se pueden usar los datos de mediciones de antibiogramas vistos en el punto 17.3.3, donde se deseaba averiguar si el factor humano tenía influencia en los resultados obtenidos. Ahora, se van a emplear dos factores, el factor A: los 4 tipos de antibióticos usados en los 6 pacientes (columnas) y el factor B: los 3 operadores del laboratorio (filas). Lo que se hace ahora es tomar los tipos de antibióticos como el segundo factor. No es necesario hacer ningún estudio estadístico para saber que habrá diferencias significativas entre los diámetros del halo de inhibición de los antibióticos. Cabría preguntarse entonces: ¿Cuál es el motivo para incluir el segundo factor si ya se sabe el resultado? La respuesta es seguir la regla de oro y tomar en cuenta todos los factores posibles para disminuir el error remanente y tener así una mayor sensibilidad para estudiar lo que realmente interesa que es el factor humano. En la primer resolución de este problema no se pudieron detectar diferencias entre los tres operadores, pero ahora, las cosas cambian como se verá a continuación: El estudio se hace sobre cultivos de *E.Coli* usando 6 cepas distintas y se mide es el diámetro del halo de inhibición de cuatro antibióticos:

CTX : cefotaxima
AKN : ampicacina

CAZ: ceftazidima
GEN : gentamicina

Los 24 diámetros usados fueron medidos por tres operadores diferentes quienes usaron la manera usual que tienen de hacer la tarea. La idea es que si se llegan a detectar diferencias significativas entre ellos (como ocurrió) entonces, se deberá buscar la manera de estandarizar el procedimiento de lectura escribiendo un protocolo específico para ello. De esa forma, se evitará la incidencia del factor humano en los resultados informados por el laboratorio para los antibiogramas. Los valores que siguen son reales y fueron obtenidos por tres bioquímicos del curso de postgrado:

$$\text{Paso 1) Se calcula la suma total de datos } T = \sum_1^a \sum_1^b \sum_1^n X_{ijk} = \sum_1^a \sum_1^b T_{jk} = \sum_1^a T_c = 1.875$$

Paso 2) Se calcula la suma del cuadrado de todos los datos:

$$T_x^2 = \sum_1^a \sum_1^b \sum_1^n X^2_{ijk} = 50.459$$

Tabla 20.3: 6 Antibiogramas medidos por 3 operadores con 4 antibióticos distintos

	CTX	CAZ	AKN	GEN	Σ
1	27	27	20	18	
	32	30	22	20	
	30	28	22	21	
	29	30	20	18	
	30	32	22	20	
	35	30	23	20	
Σ	183	177	129	117	606
2	30	24	22	22	
	30	30	22	24	
	30	30	23	23	
	30	30	23	20	
	30	30	24	24	
	30	30	25	22	
Σ	180	174	139	135	628
3	30	26	19	22	
	31	32	22	21	
	32	30	22	21	
	36	36	22	19	
	34	31	25	24	
	30	29	24	23	
Σ	193	184	134	130	641
$\Sigma\Sigma$	556	535	402	382	1875

Fuente: Dras. Cech N.; Coschiza M. y Lodeiro N.

Paso 3) Se calcula la suma de los cuadrados de los totales de las combinaciones de factores, dividido por el tamaño muestral respectivo ($n = 6$).

$$T_c^2 = \sum_1^a \sum_1^b T_{jk}^2 / n = (1/n) \sum_1^a \sum_1^b T_{jk}^2 = 50.222$$

Paso 4) Se calcula el término de corrección $T^2 / N = (1.875)^2 / 72 = 48.828$

Paso 5) Se calcula la suma de los cuadrados de los totales del grupo B (operadores) y se divide por su tamaño muestral respectivo, con $a = 4$ y $n = 6$.

$$T_b^2 = \sum_1^a \sum_1^n T_{ik}^2 / a.n = 48.854$$

Paso 6) Se calcula la suma de los cuadrados de los totales del grupo A (antibióticos) y se divide por su tamaño muestral respectivo, con $b = 3$ y $n = 6$.

$$T_a^2 = \sum_1^b \sum_1^n T_{ik}^2 / b.n = 50.161$$

Paso 7) Se calculan las sumas de cuadrados necesarias

$$SS_{total} = T_x^2 - T^2 / N = Paso 2 - Paso 4 = 1.631$$

$$SS_{dentro} = T_x^2 - T_c^2 = Paso 2 - Paso 3 = 237$$

$$SS_{filas} = T_b^2 - T^2 / N = Paso 5 - Paso 4 = 26 \text{ (operadores)}$$

$$SS_{columnas} = T_a^2 - T^2 / N = Paso 6 - Paso 4 = 1.332 \text{ (antibióticos)}$$

$$SS_{interacción} = T_c^2 - (T^2 / N) - T_b^2 - T_a^2 = (Paso 3 - Paso 4) - SS_{filas} - SS_{columnas} = 36$$

La suma de cuadrados total, sigue siendo la suma de las sumas de cuadrado entre y dentro de los grupos. Aquí, la suma de cuadrados entre se descompone en tres términos: una debida al grupo A, otra al grupo B y la última debida a la interacción. Entonces:

$$SS_{total} = (SS_{filas} + SS_{columnas} + SS_{interacción}) + SS_{dentro}$$

Paso 8) Se calcula el cuadro de ANOVA como sigue:

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre Filas (operadores)	26	(b-1)=2	13	3,3*
Entre columnas(antibióticos)	1.332	(a-1)=3	444	112,4****
Interacción	36	(a-1)(b-1)=6	6	1,5 (ns)
Dentro grupos (error)	237	a.b(n-1)=60	3,95	

Total

1.631

71

Notar que ahora se descubren diferencias significativas entre los 3 operadores, cosa que no se podía detectar con el modelo de un factor usado antes. El valor del $SS_T = 1.631$ es el mismo para los dos modelos empleados. En el de un factor se descompone en dos términos SS_{entre} los operadores que vale 26 y el remanente por el error es $SS_{dentro} = 1.605$. En cambio, en el de dos factores el remanente se reduce a 237, lo que le da mayor sensibilidad al modelo. Lo que ocurre es que al tomar un segundo factor en cuenta, ahora se tiene una mejor explicación del fenómeno. Lo que antes era el remanente de 1.605, ahora se descompone en tres términos: uno debido al efecto de los antibióticos empleados $SS_a = 1.332$, otro debido a la interacción entre ambos factores $SS_i = 36$ y el nuevo remanente de 237 ($1.605 = 1.332 + 36 + 237$). Por su parte, también se descomponen los grados de libertad: $69 = 3 + 6 + 60$; pero el valor relevante para el error no baja tanto como el valor SS y por eso el MSerror baja de 23,3 a 3,95. Entonces, al efec-

tuar el cociente para determinar el valor del estadígrafo F como el denominador sigue igual, pero el denominador se achica en casi seis veces ahora resultó ser significativo. Esto es, ahora puede detectar las pequeñas diferencias existentes entre los tres operadores.

Si se continúa el análisis usando comparaciones múltiples con el modelo de Tukey se obtienen los resultados siguientes:

	Operador 1	Operador 2	Operador 3
LI i	24,536	25,5	25,994
Promedio	25,25	26,2	26,71
LS i	25,964	29,9	27,423

Se descubre que no hay diferencias entre el operador 1 y el 2, ni entre el operador 2 y el 3. Las diferencias, se encuentran entre los operadores 1 y 3. Investigado el asunto, se descubrió que mientras el primero medía radios usando un punto imaginario en el centro del disco con el antibiótico, los otros dos medían siempre los diámetros directamente. Como conclusión, se modificó el protocolo de la técnica de la manera siguiente: “Se deben medir los diámetros siempre que se pueda. Sólo cuando la superposición entre los halos sea muy grande y los valores estén cerca de los críticos, se usará la medición de los radios, pero repitiendo esto por lo menos 3 veces y en diferentes posiciones promediando los resultados”.

Como conclusión final se puede ver la practicidad de tomar todos los factores posibles en cuenta, aún cuando alguno pueda parecer inútil. La reducción de la variabilidad remanente (error) puede hacer detectar diferencias que de otra manera permanecerían ocultas como en el ejemplo visto. No siempre ocurrirá esto, pero el agrandar un poco los cálculos con la misma inversión efectuada para realizar las mediciones, siempre justifica el aumento de sensibilidad de los modelos de ANOVA.

20.7 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|---|----------|----------|
| 1) El ANOVA de dos factores puede ser con o sin repetición | V | F |
| 2) Los modelos con repetición permiten analizar mejor los datos. | V | F |
| 3) Los 5 supuestos básicos para estos modelos son: | | |
| 4) Los modelos con repetición permiten detectar el fenómeno de interacción. | V | F |
| 5) Explicar la diferencia entre modelos encajados de dos niveles y estos:..... | | |
| 6) Para reconocer un encajado se debe mirar si hay o no repetición. | V | F |
| 7) Hay tres modelos posibles que son:..... | | |
| 8) El modelo teórico puro o mixto se puede plantear con:..... | | |
| 9) El primer paso para calcular este ANOVA es calcular la suma de cuadrados. | V | F |
| 10) Se requiere del valor promedio grupal para poder compararlos entre sí. | V | F |
| 11) El término cuadrático para cada factor, se calcula sumando el cuadrado del total. | V | F |
| 12) Explicar como se obtienen las SS necesarias para la Tabla de ANOVA:..... | | |
| 13) La SS entre las muestras se descompone en dos términos, uno por factor. | V | F |
| 14) La SS total es la suma de las SS dentro y entre las muestras. | V | F |
| 15) Cada MS se calcula dividiendo los SS por sus respectivos grados de libertad. | V | F |

- 16) Presentar una Tabla de ANOVA en forma genérica:.....
- 17) Los estadígrafos F para cada factor son el cociente de sus MS respecto de MSerror **V F**
- 18) Cuando se detecta interacción significativa, se deje de cumplir un supuesto básico. **V F**
- 19) Una sinergia es una interacción de tipo negativa. **V F**
- 20) Explicar el concepto de interferencia y sinergia:.....
- 21) Un modelo I puro tiene en ambos factores un diseño del tipo Modelo I. **V F**
- 22) Un modelo mixto debe tener el segundo factor como modelo II. **V F**
- 23) Explicar como se diseña un experimento con los modelos teóricos posibles:.....
- 24) La idea básica de estos modelos es descomponer la variación total en partes. **V F**
- 25) Explicar un diseño experimental con bloques aleatorizados:.....
- 26) En los bloques aleatorizados se puede detectar la interacción. **V F**
- 27) El modelo no paramétrico equivalente a los bloques es el de Kruskal-Wallis. **V F**
- 28) El modelo de Friedman se puede usar en casos con repetición. **V F**
- 29) El diseño con un Modelo II es el más apropiado para el modelo de Friedman. **V F**
- 30) Explicar resumidamente el modelo no paramétrico de Friedman:.....

2) Hacer un esquema que muestre un resumen de todos los modelos vistos hasta ahora, para una , dos y más muestras; con uno o más factores en el caso paramétrico.

3) Agregar al esquema anterior los modelos no paramétricos equivalentes.

4) Discutir acerca de las ventajas y desventajas de cada uno de los modelos vistos, sus supuestos básicos y su campo de aplicación.

5) Armar un cuadro sinóptico que muestre en forma simplificada como se deben usar todos los modelos estadísticos vistos hasta ahora, para el caso de tener que trabajar con una sola magnitud clínica, y agregando los comentarios pertinentes a un costado.

6) En un laboratorio de análisis se desean comparar entre sí a tres técnicas para la determinación de Glucosa, elegidas expresamente para ello. Así se obtiene suero de tres pacientes escogidos al azar entre los concurrentes diarios. El suero de cada paciente se fracciona en 9 alícuotas iguales. Para probar cada técnica se eligen al azar a tres muestras de suero de cada uno. Los veintisiete valores medidos se vuelcan en la tabla siguiente.

Paciente	Técnica		
	1	2	3
A	0,98	0,99	1,13
	1,01	0,95	1,07
	0,98	1,01	1,07
B	0,92	0,91	1,03
	0,86	0,9	1,07
	0,92	1,01	1,02
C	0,84	0,86	0,98
	0,79	0,79	1,01
	0,8	0,81	1,01

Se desea averiguar:

- a) Si hay diferencias entre las 3 técnicas analizadas.
- b) En caso de encontrarlas usar un modelo apropiado para las comparaciones múltiples.
- c) Interesa determinar la componente añadida de varianza entre los pacientes.
- 7) Un farmacéutico a cargo de una cadena comercial con seis sucursales, escoge al azar tres días de un bimestre, y busca el total de ventas del día para cada caso seleccionado. Los datos los vuelca en una tabla como la siguiente.

FACTOR B (Sucursal)	FACTOR A (días)		
	05/06/00	15/07/00	23/06/00
1	250	280	267
2	340	335	320
3	290	280	240
4	500	568	520
5	180	190	182
6	380	400	390

Se pide calcular:

- a) Si hay diferencia significativa entre las sucursales.
- b) Si hay diferencia significativa entre los días.
- c) Como no sabe si se cumplen además de el modelo de ANOVA de dos factores adecuados, volver a resolver el problema con un modelo no paramétrico.

21

Modelos para más de una variable

Hasta ahora se han visto diferentes modelos estadísticos para el caso de una sola magnitud biológica medida. Pero en los experimentos es frecuente tratar el caso donde hay más de una variable involucrada. En este capítulo se tratará el caso de dos variables con los modelos de *Regresión* y *Correlación estadística*. El caso de más de dos variables excede los límites del presente trabajo, destinado a los alumnos de las carreras de Farmacia y Bioquímica, porque requiere del manejo de un nivel matemático (espacios vectoriales n-dimensionales) que estos no poseen. Se comienza en este capítulo con el caso más sencillo de relación lineal simple entre dos magnitudes biológicas cualesquiera. Es decir, una relación del tipo $Y = X = f(X)$. A continuación se trata el modelo más general con relaciones del tipo $Y = a + bX$, la ecuación de una recta o polinomio de primer grado. Modelo conocido con el nombre de *Análisis de Regresión lineal*. Se presentan los métodos cortos de cálculo y el planteo de ensayos de hipótesis respectivos. A continuación se generaliza la regresión para el caso de polinomios de más de un grado y se muestran las transformaciones de variables, convenientes para linealizar los cálculos.

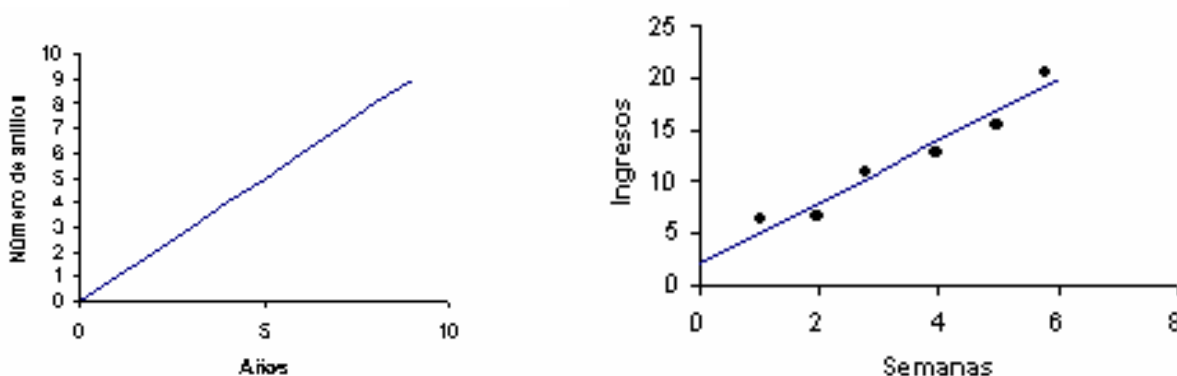
21.1 Introducción

Antes de comenzar, es conveniente aclarar una confusión frecuente en la bibliografía. Hay textos que usan los cálculos de regresión y correlación para los mismos casos por lo similares que son. A veces el lector se confunde y piensa que puede emplear ambos modelos en un mismo problema. Esto no es así, hay una diferencia fundamental entre ellos. El *Análisis de Regresión* se usa cuando el investigador sabe que existe una relación entre las variables porque hay una teoría o investigaciones previas que la han descubierto. Por ejemplo, la relación entre espacio y tiempo ya se sabe que es la velocidad, o como la relación entre voltaje e intensidad de corriente eléctrica. En estos casos, el investigador suele estar interesado en verificar experimentalmente tal relación y el objeto de la regresión es encontrar la curva que mejor ajuste a sus datos experimentales. Cuando no conoce la relación exacta, sino que trata de encontrar una curva de tipo práctica en los llamados modelos empíricos, la idea es tener una manera rápida de relacionar dos magnitudes; como por ejemplo, en la industria cuando se relaciona la velocidad de un motor con el rendimiento del generador eléctrico que mueve. Por su parte, el *Análisis de Correlación* se emplea cuando el investigador sospecha que ambas magnitudes están relacionadas, pero no tiene idea de una ecuación que las combine. Por ejemplo el caso de peso y talla, donde todo lo que se sospecha es que a mayor talla, mayor peso, pero nadie ha descubierto una fórmula que las relacione. Con esto en mente se comienza con:

21.2 Análisis de regresión

La forma más común de concebir las relaciones entre pares de magnitudes es del tipo *causa-efecto*. Lo que trata el análisis estadístico es establecer la forma y la significación de las relaciones funcionales entre las dos variables. La demostración de la relación causa-efecto es tema del procedimiento del método científico, y queda a cargo del investigador. Estadística trata de verificar la función matemática que permite predecir que valores de una variable Y corresponden a valores dados de una variable X . Se suele escribir como $Y = f(X)$, donde X es la variable independiente.

Figura 21.1: Rectas de regresión.



(a) Número de anillos de un árbol (b) Ingresos semanales en una Farmacia (miles \$)

El caso más simple de una recta de regresión es del tipo $Y = X$ donde la recta pasa por el origen de coordenadas y su inclinación es de 45° . Este es el caso de la relación entre el número (Y) de anillos de un árbol y su edad en años (X). Ver caso (a) de la figura anterior. El caso más general es cuando la recta no pasa por el origen y su inclinación es cualquiera. La relación matemática es del tipo $Y = a + b X$ donde a es el punto por donde la recta corta al eje Y cuando $X = 0$ y b es la tangente del ángulo de inclinación. Ver caso (b) de la figura anterior donde se muestra la relación de los ingresos de una Farmacia medidos en miles de pesos con el tiempo expresado en semanas, con una ecuación expresada con: $Y = 2 + 3 X$. Aquí, se supone que se han medido los ingresos reales en una Farmacia y se encontró la recta que mejor ajusta a esa serie de puntos con el Análisis de Regresión.

En todo ejemplo real, las observaciones no coinciden exactamente con la recta de regresión debido a los errores casuales que afectan las mediciones. En Biología se suponen causas de tipo genético y ambientales para explicar la aleatoriedad, además de los errores de medición. Esto significa que para un dado valor de X , el valor de Y que le corresponde no será exactamente: $a + b X$, sino que esta ecuación usando el valor de X arroja el *valor esperado* de Y denominado Y^* . Entonces, para cada valor medido de X se tendrán dos valores: el valor medido en el experimento Y y su valor esperado calculado por la recta de regresión Y^* . La diferencia entre ambos ($Y - Y^*$) debe ser lo más chica posible, para tener una buena aproximación. La idea básica del Análisis de Regresión es minimizar matemáticamente el cuadrado de estas diferencias con el método de los *mínimos cuadrados*.

21.3 Diseños experimentales

Los diseños experimentales en regresión son dos: el Modelo I y el II. Ambos se basan en cuatro hipótesis básicas.

1. *La variable independiente se mide sin error.* Esto significa que está bajo el control del investigador y se consideran “fijos” a los valores de X que eligió. Por ejemplo, al manipular las dosis de un cierto medicamento hipotensor, se fijan estos valores de X y por lo tanto no se la puede considerar como una variable aleatoria. En cambio, el valor de la presión sanguínea del paciente no puede ser fijada por el investigador; entonces Y varía en forma libre. En forma análoga a la vista en los modelos de Anova, se considera *Modelo I* de regresión, al caso donde los valores de X pueden ser manipulados a voluntad. Cuando esto no es así, entonces se tiene el *Modelo II* de regresión donde ambas variables se consideran aleatorias porque no están bajo el control de investigador. Por ejemplo, se toma una muestra aleatoria de una población y a cada individuo seleccionado se le mide su presión sanguínea y su nivel hormonal. Ambas variables no quedan bajo el control del investigador y deben ser consideradas aleatorias.

2. *El valor esperado de la variable Y^* para un dado valor de X , se determina con la relación:*

$$E(Y) = \mu_y = Y^* = \alpha + \beta X$$

Se usan las letras griegas para describir una relación Paramétrica entre las variables.

3. *Para cualquier valor de X , los valores de Y se distribuyen independiente y normalmente.*

$$Y^* = \alpha + \beta X + \varepsilon$$

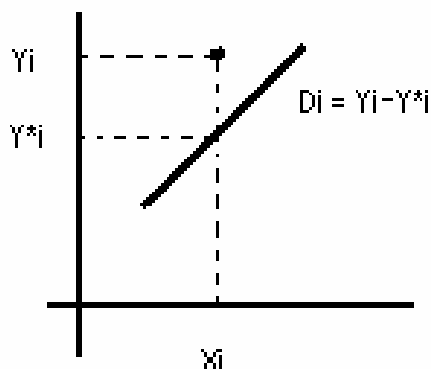
Donde ε es un término de error con una distribución normal de media igual a cero y desvío σ . Esto supone que cada valor X tiene un gran número de valores posibles de Y a hacer la medición, con una distribución normal, cuyo eje de simetría es una vertical (eje z : dentro de un espacio tridimensional imaginario) que pasa por el punto Y^* de la recta de regresión, orientada sobre la línea que une el punto X con él Y^* correspondiente. Esto para los casos donde hay más de un valor de Y para cada valor de X .

4. *Todas las muestras a lo largo de la línea de regresión son homocedásticas.* Se supone que todas las distribuciones normales mencionadas en el punto anterior tienen la misma varianza.

21.4 Cálculos básicos en regresión

La manera más sencilla de ilustrar estos cálculos es partiendo del concepto básico de la recta de regresión: minimizar el cuadrado de las diferencias entre el valor medido Y_i y el valor correspondiente esperado por la recta Y^*_i . El caso más sencillo es cuando hay un solo valor medido de Y_i , para cada valor X_i . En la Figura 21.2 siguiente se muestra el planteo:

Figura 21.2: Diferencias a minimizar.



Cada punto medido tiene un par de coordenadas ($X_i ; Y_i$). Con la recta se estima el valor Y^*_i . La diferencia entre el valor medido y el estimado es $D_i = Y_i - Y^*_i$. Cuanto mejor sea la estimación, menor será esta diferencia. El método consiste en *minimizar* la suma de sus cuadrados: derivando respecto de las dos incógnitas a y b , igualando a cero y despejando. Queda un sistema de dos ecuaciones con dos incógnitas, que al resolverlo permiten hallar las denominadas *ecuaciones paramétricas de regresión*.

Para minimizar se usan las relaciones (ver Apéndice 1):

$$\frac{\partial}{\partial a} \sum D_i^2 = 0 \qquad \frac{\partial}{\partial b} \sum D_i^2 = 0$$

Resolviendo estas relaciones se obtienen:

$$\left. \begin{aligned} \sum Y_i &= a \cdot N + b \sum X_i \\ \sum X_i Y_i &= a \cdot \sum x_i + b \sum X_i^2 \end{aligned} \right\} \text{Ecuaciones normales o paramétricas de regresión.}$$

Ejemplo) Se ha medido la altura de 15 padres y de sus hijos primogénitos en metros. Hallar la recta de regresión. Los datos son:

	X (padres)	Y (hijos)	X ²	X Y	Y ²
1	1,71	1,75	2,9241	2,9925	3,0625
2	1,67	1,76	2,7889	2,9392	3,0976
3	1,62	1,64	2,6244	2,6568	2,6896
4	1,75	1,76	3,0625	3,08	3,0976
5	1,59	1,64	2,5281	2,6076	2,6896
6	1,81	1,77	3,2761	3,2037	3,1329
7	1,69	1,72	2,8561	2,9068	2,9584
8	1,68	1,73	2,8224	2,9064	2,9929
9	1,76	1,75	3,0976	3,08	3,0625
10	1,72	1,72	2,9584	2,9584	2,9584
11	1,79	1,81	3,2041	3,2399	3,2761
12	1,68	1,69	2,8224	2,8392	2,8561
13	1,64	1,66	2,6896	2,7224	2,7556
14	1,68	1,71	2,8224	2,8728	2,9241
15	1,58	1,62	2,4964	2,5596	2,6244
SUMA	25,37	25,73	42,9735	43,5653	44,1783

Reemplazando en las ecuaciones paramétricas resulta

$$\left. \begin{aligned} 25,73 &= a(15) + b(25,37) \\ 43,5653 &= a(25,37) + b(42,9735) \end{aligned} \right\} \text{ De donde se calculan } a \text{ y } b.$$

Sin embargo hay una forma más corta. La recta de regresión debe pasar necesariamente por el centro de gravedad de los datos, es decir por el valor medio de X e Y. Esto es, si se divide la primer ecuación por N en ambos miembros, queda igual a:

$$(1/N) \sum Y_i = a + (b/N) \sum X_i \quad \text{O sea:} \quad \bar{Y} = a + b \bar{X} \quad \text{O bien:} \quad a = \bar{Y} - b \bar{X}$$

Se puede calcular los valores medios de la tabla anterior resultando:

$$\bar{X} = 25,37 / 15 = 1,6913 \text{ m} \quad \bar{Y} = 25,73 / 15 = 1,7153 \text{ m}$$

Reemplazando por los valores hallados se despeja: $a = (1,7153 - b \cdot 1,6913)$

Reemplazando el valor de a en la segunda ecuación, queda una sola ecuación con una incógnita:

$$43,565 = (1,72 - b \cdot 1,69) 25,37 + b 42,97 = 1,72(25,37) + b [(42,97) - (1,69)(25,37)]$$

Despejando se calcula: $b = -0,7$. Con este valor, se puede calcular a, reemplazando en la fórmula de valores medios:

$$\bar{Y} = a + b \bar{X}$$

$$1,72 = a + (-0,7)(1,6913)$$

Resultando: $a = 2,9$

Y la ecuación de la recta es: $Y = 2,9 - 0,7 X$

Si en lugar de calcular la regresión de Y sobre X (X es la variable independiente), se desea calcular la regresión de X sobre Y, porque Y es la variable independiente, entonces las ecuaciones quedan muy similares, pero con otras incógnitas:

$$\left. \begin{aligned} \sum X_i &= c \cdot N + d \sum Y_i \\ \sum X_i Y_i &= c \cdot \sum Y_i + d \sum Y_i^2 \end{aligned} \right\} \text{ Ecuaciones normales para regresión de X sobre Y.}$$

Aplicando estas nuevas relaciones al ejemplo anterior resulta:

$$\left. \begin{aligned} 25,37 &= c(15) + d(25,73) \\ 43,5653 &= c(25,73) + d(44,1783) \end{aligned} \right\} \text{ De donde se calculan } c \text{ y } d.$$

Para simplificar se usa:

$$\bar{X} = c + d \bar{Y} \quad \text{O bien:} \quad c = \bar{X} - d \bar{Y}$$

Reemplazando por los valores hallados se despeja: $c = (1,6913 - d 1,7153)$

Reemplazando el valor de a en la segunda ecuación, queda una sola ecuación con una incógnita:

$$43,5653 = (1,6913 - d 1,7153)(25,73) + d 44,1783$$

$$0,04815 = d (0,04363) \quad \text{y despejando resulta:} \quad d = 1,1 \quad \text{y así se obtiene} \quad c = - 0,2$$

Y la ecuación de esta otra recta es: $X = - 0,2 + 1,1 Y$

Como se puede apreciar, la regresión de Y sobre X difiere de la regresión de X sobre Y. Solo cuando la coincidencia entre los puntos reales y la recta de regresión sea perfecta, entonces ambas rectas de regresión serán iguales.

Existen otras maneras de simplificar los cálculos, usando un cambio de coordenadas. Esto se hace, colocando el origen en el centro de gravedad de los datos. Sencillamente, a cada valor de X e Y se le resta su media respectiva. Las ecuaciones quedan más cortas. El lector interesado podrá encontrar estas técnicas en la Bibliografía específica.

21.5 Cálculos cortos en regresión

La manera más rápida de efectuar los cálculos, incluso los necesarios para los tests de hipótesis, se ilustra en el ejemplo siguiente:

Ejemplo) Nelson (1964) midió la pérdida en peso (en mg) de 9 tandas de veinticinco coleópteros *Tribolium*, luego de seis días de no ingerir comida a nueve humedades diferentes. Los datos son:

Hum.	Peso	x =	y =	x ²	y ²	x.y	Y*	D =	D ²	y* =	y ^{*2}
X	Y	X-50,39	Y-6,02					Y-Y*		Y*-6,02	
0	8,98	-50,389	2,9578	2539,04	8,748449	-149,04	8,704	0,276	0,07616	2,68181	7,1921
12	8,14	-38,389	2,1178	1473,707	4,484983	-81,299	8,0654	0,0746	0,00557	2,04314	4,1744
29,5	6,67	-20,889	0,6478	436,3457	0,419616	-13,531	7,134	-0,464	0,21527	1,11175	1,236
43	6,08	-7,3889	0,0578	54,59568	0,003338	-0,4269	6,4155	-0,3355	0,11254	0,39325	0,1546
53	5,9	2,6111	-0,122	6,817901	0,014938	-0,3191	5,8833	0,0167	0,00028	-0,13897	0,0193
62,5	5,83	12,111	-0,192	146,679	0,036949	-2,328	5,3776	0,4524	0,20463	-0,64458	0,4155
75,5	4,68	25,111	-1,342	630,5679	1,80156	-33,705	4,6858	-0,0058	3,3E-05	-1,33647	1,7861
85	4,2	34,611	-1,822	1197,929	3,320494	-63,069	4,1801	0,0199	0,00039	-1,84208	3,3933
93	3,72	42,611	-2,302	1815,707	5,300227	-98,1	3,7544	-0,0344	0,00118	-2,26786	5,1432
Σ	453,5	54,2	0	8301,389	24,13056	-441,82	54,2	0	0,61606	0	23,514
	50,3889	6,022	:Medias				6,0222				
b										-0,053222	
a										8,704027	

Paso 1) En las primeras dos columnas se vuelcan los datos, se suman y se saca el promedio.

Paso 2) En la tercer y cuarta columna, se colocan los valores de las dos primeras menos sus respectivos promedios y se determinan: $y = f(x)$ dos variables que pasan por el centro de gravedad de los datos, de acuerdo a la transformación efectuada:

$$x = X - \bar{X} = X - 50,3889 \quad \text{e} \quad y = Y - \bar{Y} = Y - 6,022$$

Paso 3) En la quinta y sexta columnas se calculan los cuadrados de los nuevos valores x e y . Mientras que en la séptima se calcula su producto.

Paso 4) Ahora se pueden calcular los coeficientes de la recta de regresión a y b con la ecuación de la recta de regresión de x sobre y , escrita en el nuevo sistema de coordenadas:

$$y = b x = \left(\frac{\sum xy}{\sum x^2} \right) \cdot x \quad \text{Por lo tanto: } b = (-441,82) / (8301,389) = -0,0532$$

Es decir se calcula la pendiente de la recta b como el cociente entre los totales de la quinta y séptima columna. Conocido este valor, se puede calcular el otro considerando que la recta pasa por el centro de gravedad de los puntos, o sea por sus valores promedio:

$$\bar{Y} = a + b \bar{X} \quad \text{O sea: } a = \bar{Y} - b \bar{X} = 6,022 - (-0,0532) 50,389 = 8,704$$

$$\bar{Y} = 8,7 - 0,05 \bar{X}$$

21.6 Ensayos de hipótesis en regresión

La manera de ensayar la hipótesis de que la regresión existe es con un Cuadro de Regresión, similar al visto en el caso del Anova. Para ello se divide a la suma de cuadrados que representa la variabilidad total en dos términos, lo mismo que con los grados de libertad, y se encuentra un estadígrafo F con distribución de Fisher. Para ello se empieza con:

Variación Total: De una variable cualquiera Y , se calcula como la suma de cuadrados de las diferencias entre cada valor medido y su promedio (es la suma de cuadrados en Y). Esto es:

$$VT = \sum (Y - \bar{Y})^2$$

Si al término entre paréntesis se le suma y le resta una misma cantidad Y^* este no se altera. Realizando el reemplazo y las cuentas, resulta:

$$VT = \sum (Y - \bar{Y})^2 = \sum (Y - Y^* + Y^* - \bar{Y})^2 = \sum (Y - Y^*)^2 + \sum (Y^* - \bar{Y})^2 = VNE + VE$$

Donde VNE es la *Variación No Explicada* porque los valores de Y se comportan en forma aleatoria o no previsible. Mientras que el segundo término, cada valor de la diferencia tiene un patrón

El cuadrado medio explicado se debe a la regresión lineal y mide la cantidad de variación de Y, tomada en cuenta por la variación de X. Si a este valor se le resta de la variación total, la variación residual o remanente es la no explicada y se la usa como el cuadrado medio error. El valor del estadígrafo F se lo calcula como el cociente entre estos cuadrados medios. En este caso su valor tan grande hace altamente significativo el rechazo de la hipótesis nula, que suponía que no había regresión. O sea, se tiene evidencia científica validada por el modelo, de que gran parte de la varianza encontrada puede ser explicada por la regresión de Y sobre X.

Se pueden hacer otro ensayo de hipótesis acerca del “coeficiente de regresión”: b como se muestra a continuación:

(H₀) $\beta = 0$ No hay regresión en la población de donde se tomaron las muestras.

(H₁) $\beta \neq 0$ Hay regresión.

El ensayo usa el modelo de Student con $t = (b - \beta) / DS_b$ versus $t_{\alpha;v} = t_{\alpha;n-2}$
Donde el error típico de estimación de b está dado por:

$$DS_b = \sqrt{\frac{MS_{\text{error}}}{\sum x^2}} = \sqrt{\frac{0,088}{8301,39}} = 0,003256$$

Luego $t = (-0,05322 - 0) / 0,003256 = -16,35^{***}$ ($t_{0,001;7} = 5,408$)

Y se rechaza la hipótesis nula con valores altamente significativos.

También se pueden hallar los Límites de confianza para el coeficiente de regresión con:

$$\beta \in (b \pm t_{\alpha;v} DS_b)$$

Donde para un 95% de confianza resulta $t_{\alpha;v} = t_{0,05;7} = 2,356$. O sea:

$$\beta \in (-0,053 \pm 0,008) \rightarrow 95\% \text{ CI } (-0,061 ; -0,045)$$

Se concluye que el valor verdadero del coeficiente de regresión β está en dicho intervalo, con un 95% de probabilidades a favor.

21.7 Regresión por el origen: Recta de Calibración

En muchas oportunidades, la teoría empleada para la regresión exige que la recta pase a través del origen de coordenadas. Entonces, ya se tiene un punto para el cual no se encontrará variaciones en el muestreo. Tal punto debe tratarse en una forma diferente a otro cualquiera observado. Un ejemplo de esto es el caso visto del número de anillos de un árbol y su edad. Otro, más frecuente en Bioquímica y Farmacia, es el caso de la *Recta de Calibración* de un instrumento de laboratorio cualquiera, como una balanza, un espectrofotómetro, etc. Generalizando, para todo instrumento que requiera hacer el “ajuste del cero” antes de comenzar a usarlo. En una balanza, esto es la primera pesada en vacío, cuando se ajusta la escala al cero luego de ser nivelada.

En un espectrofotómetro, es la primera lectura colocando agua destilada en la cubeta. Para ilustrar este caso se presentan los siguientes datos, con dos enunciados diferentes:

Ejemplo 1) Para determinar si un espectrofotómetro está calibrado se han medido 13 valores de referencia o patrones (**Y**) (soluciones calibradas) y los valores observados de transmitancia (**X**) se muestran en la siguiente tabla – ver caso práctico en Apéndice 2:

N	X	Y
1	13,6	52
2	13,9	48
3	21,1	72
4	25,6	89
5	26,4	80
6	39,8	130
7	40,1	139
8	43,9	173
9	51,9	208
10	53,2	225
11	65,2	259
12	66,4	199
13	67,7	255
Total	528,8	1929

$$\sum X^2 = 26.062,1 \qquad \sum Y^2 = 356.259$$

$$\sum XY = 95.755,7 \qquad b = \frac{\sum XY}{\sum X^2} = 3,67$$

Teniendo en cuenta que la recta pasa por el origen, la ecuación de la misma es:

$$Y^* = 3,67 X$$

Para efectuar la validación estadística se usan las relaciones siguientes:

$$\sum x^2 = \sum (X - \bar{X})^2 = \sum X^2 - (\sum X)^2 / N = 26.062,1 - [(528,8)^2 / 13] = 4.552,14$$

$$\sum y^2 = \sum (Y - \bar{Y})^2 = \sum Y^2 - (\sum Y)^2 / N = 356.259 - [(1.929)^2 / 13] = 70.025,08$$

$$\sum xy = \sum XY - (\sum X)(\sum Y) / N = 95.755,7 - [(528,8)(1.929) / 13] = 17.289,92$$

Entonces ahora se pueden calcular las variabilidades siguientes:

$$VT = \sum (Y - \bar{Y})^2 = \sum y^2 = 70.025,08 \text{ con 12 grados de libertad}$$

$$VE = \sum (Y^* - \bar{Y})^2 = \frac{(\sum xy)^2}{\sum x^2} = \frac{(17.289,92)^2}{4.552,14} = 65.670,51 \text{ con 1 grado de libertad}$$

$$VNE = \sum (Y - Y^*)^2 = VT - VE = 70.025,08 - 65.670,51 = 4.354,57 \text{ con 11 grados de libertad}$$

Con estos datos se puede armar la tabla de regresión siguiente

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Explicada (debida a la regresión lineal)	65.670,51	1	65.670,51	165,9***
No Explicada (error)	4.354,57	11	395,87	

Total 70.025,08 12

Se tiene prueba altamente significativa de que existe la regresión de Y sobre X, aunque se ha tomado la ecuación general sin tener en cuenta que pasa por el origen.

Cuando se tengan dudas que la recta de regresión pasa por el origen, o sea $a = 0$, se puede hacer otra validación estadística con:

H_0 : $a = 0$ La recta pasa por el origen.

H_1 : $a \neq 0$ La recta no pasa por el origen.

El valor muestral se calcula considerando que en el centro de gravedad de los datos debe ser:

$$y = Y - \bar{Y} = b x = \left(\frac{\sum xy}{\sum x^2} \right) \cdot (X - \bar{X})$$

$$Y - 148,4 = 3,67 (x - 40,68) = 3,67 x - 149,29$$

$$Y = - 0,89 + 3,67 x$$

Entonces, con esta recta se tendrá la contribución atribuible a la media con:

$$\text{La variación debida a la media es: } (\sum Y)^2 / N = (1929)^2 / 13 = 286.233,9$$

$$\text{La variación debida a b es: } VE = \frac{(\sum xy)^2}{\sum x^2} = 65.670,5$$

Entonces la total es $286.233,9 + 65.670,5 = 351.904,4$ con 11 grados de libertad

$$\text{Como la suma de cuadrados debida a la regresión es: } (\sum XY)^2 / \sum X^2 = 351.819$$

La regresión adicional debida al ajuste de la media es: $SS m = (351.904 - 351.819) = 85$

Como tiene un grado de libertad es: $MS m = SS m / 1 = 85$ entonces,

$$F = MS m / MS error = 85 / 395,87 = 0,215 \text{ No significativo}$$

Se concluye que no hay evidencia que muestre que la recta no pasa por el origen.

Sin embargo el problema principal aquí, no es determinar si hay regresión (porque ya se sabe que así será), esto es no hace falta probar que $b \neq 0$; sino determinar que el valor encontrado no difiera significativamente del *factor* de dilución correcto, que es lo que usa el bioquímico para sus mediciones. Imaginando que el valor esperado del factor de dilución es $\beta = 3,5$:

Ho : $b = \beta = 3,5$ El sistema está bien calibrado para hacer las mediciones.

H1 : $b \neq \beta$ El sistema no está bien calibrado.

Se usa el modelo de Student con:

$$t = (b - \beta) / DS_b \quad \text{versus} \quad t_{\alpha;v} = t_{\alpha;n-2}$$

Donde el error típico de estimación de b está dado por:

$$DS_b = \sqrt{\frac{MS_{\text{error}}}{\sum x^2}} = \sqrt{\frac{395,87}{4.552,14}} = 0,295$$

Luego: $t = (3,67 - 3,5) / 0,295 = 0,58$ (no significativo) ($t_{0,95;11} = 2,201$)

Por lo que no se puede rechazar la hipótesis nula. No hay prueba estadística como para creer que el sistema de medición no está calibrado. Se puede usar otro ejemplo, si se tratase de una balanza, la idea es que cada valor medido Y, de los patrones utilizados X, sea $Y = X$. Esto es, una recta que pasa por el origen a 45° , o bien $\beta = 1$.

Ejemplo 2) En un estudio de Bacteriología, se midieron las reversiones inducidas a la independencia por 10^7 células sobrevivientes (Y), por dosis (ergs / bacterias) 10^{-5} (X) de *Escherichia coli* estreptomyceno-dependiente, sometidas a radiación ultravioleta monocromática de $2,967 \text{ \AA}$ de longitud de onda. (Datos de Zelle, M.R.; Univ. Cornell, del libro de Steel y Torrie).

															Total
X	13,6	13,9	21,1	25,6	26,4	39,8	40,1	43,9	51,9	53,2	65,2	66,4	67,7	528,8	
Y	52	48	72	89	80	130	139	173	208	225	259	199	255	1929	

Para simplificar los cálculos se han tomado los mismos datos del enunciado anterior. Por lo tanto la recta de calibración está dada por la ecuación:

$$Y^* = 3,67 X$$

Habrá 3,67 retornos inducidos por dosis. Entonces la recta de regresión de retornos inducidos Y, por la dosis administrada X se expresa de la manera anterior.

Lo primero es probar que la recta de regresión existe, esto es probar que $b \neq 0$ como se hizo en el caso anterior. Luego si llegase a haber dudas respecto a si la recta pasa o no, por el origen de coordenadas, se puede hacer un test como se vio más arriba. El recurso utilizado aquí es transformar la variable medida para que el resultado se aproxime a una línea recta.

21.8 Más de un valor de Y

Muchas veces en la práctica se encuentran casos donde se obtienen más de un valor de Y por cada valor de X controlado por el investigador. El ordenamiento de los datos se parece mucho al orden en que se presentan las observaciones en un análisis de varianza de un factor. Para ilustrar este punto, se ha desarrollado un ejemplo con tamaños muestrales diferentes que es el caso más completo.

Ejemplo 1) Un investigador mide los porcentajes p de supervivencia del coleóptero *Tribolium castaneum* a cuatro densidades de siembra. Los datos de los porcentajes fueron transformados con $Y = \arcsen \sqrt{p}$ porque cumplen mejor los supuestos de distribución normal y homoscedasticidad. El número de huevos por gramo de arena es la variable X controlada por el investigador. La supervivencia se calcula por el número de insectos que llegan a la edad adulta. Se preparan 4 densidades de siembra diferente y los resultados, transformados se muestran a continuación:

	Densidad de siembra X			
	5 / g	20 / g	50 / g	100 / g
Supervivencia Y	61,68	68,21	58,69	53,13
	58,37	66,72	58,37	49,89
	69,30	63,44	58,37	49,82
	61,68	60,84		
	69,30			
Total	320,33	259,21	175,43	152,84
Ni	5	4	3	3
Medias	64,07	64,80	58,81	50,95

Ref: Ejemplo de Sokal-Rohlf (pág. 480)

Se desea saber si existen diferencias en la supervivencia de los cuatro grupos y además, si se puede establecer una línea de regresión de supervivencia sobre la densidad.

Para resolver este caso se procederá en dos etapas. En la primera se busca mediante modelos de ANOVA decidir si hay diferencia entre los grupos de siembra. En la segunda se procede con el análisis de regresión. Por regla general, si no hay significación en ANOVA es bastante improbable que exista una línea de regresión.

Etapas 1) ANOVA: Se procede con los pasos habituales para este modelo:

Paso 1) Se calcula la suma total de las observaciones $T = 907,81 = \sum \sum Y$

Paso 2) Se calcula la suma de los cuadrados de los datos: $T_x^2 = 55.503,6547$

Paso 3) Se calcula la suma de los totales grupales al cuadrado, divididos por su tamaño muestral respectivo:

$$T_x = [(320,33)^2 / 5] + [(259,21)^2 / 4] + [(175,43)^2 / 3] + [(152,84)^2 / 3] = 55.364,968$$

Paso 4) Se calcula el término de corrección $T^2 / N = 54.941,2664$

Paso 5) Se calculan las sumas de cuadrados con:

$$SS_T = \text{Paso 2} - \text{Paso 4} = 55.503,6547 - 54.941,2664 = 562,3883$$

$$SS_E = \text{Paso 3} - \text{Paso 4} = 55.364,9680 - 54.941,2664 = 423,7016$$

$$SST = SST - SSE = 562,3883 - 423,7016 = 138,6887$$

Paso 6) Los grados de libertad son $v_T = N - 1 = 14$; $v_E = a - 1 = 3$ y $v_D = N - a = 11$

Paso 7) Se formula la Tabla de ANOVA como sigue:

Variación	SS	v	MS	F
Entre grupos	423,7016	3	141,2339	11,2**
Dentro de grupos	138,6777	11	12,6079	
Total	562,389	14		

Los grupos difieren muy significativamente entre sí.

Etapa 2) Regresión: Ahora se debe comprobar si las diferencias entre los valores de supervivencia pueden ser explicados por una regresión lineal sobre la densidad de siembra.

Paso 8) Se calcula la sumatoria de los valores de X multiplicados por su tamaño muestral con:

$$\sum N_i X = 5 (5) + 4 (20) + 3 (50) + 3 (100) = 555$$

Paso 9) Se calcula la sumatoria de los valores de X^2 multiplicados por su tamaño muestral con:

$$\sum N_i X^2 = 5 (5)^2 + 4 (20)^2 + 3 (50)^2 + 3 (100)^2 = 39.225$$

Paso 10) Se calcula la sumatoria de los productos de X e \bar{Y} por su respectivo tamaño muestral con:

$$\sum N_i X \bar{Y} = \sum X \sum Y = 5 (320,33) + 20 (259,21) + 50 (175,43) + 100 (152,84) = 30.841,35$$

Paso 11) Se calcula el término de corrección para X con:

$$TC_x = \frac{(\sum N_i X)^2}{\sum N_i} = (\text{Paso 8})^2 / N = (555)^2 / 15 = 20.535$$

Paso 12) Se calcula la suma de cuadrados de x con:

$$\sum X^2 = \sum Ni X^2 - TC_x = \text{Paso 9} - \text{Paso 11} = 39.225 - 20.535 = 18.690$$

Paso 13) Se calcula la sumatoria de los productos:

$$\sum_{x,y} = \sum X(\sum Y) - \frac{(\sum NX)(\sum \sum Y)}{\sum Ni} = \text{Paso 10} - \frac{\text{Paso 8} \cdot \text{Paso 1}}{N} = -2.747,62$$

Paso 14) Se calcula la suma de los cuadrados explicada con:

$$SS_{ex} = \frac{(\sum XY)^2}{\sum X^2} = (\text{Paso 13})^2 / \text{Paso 12} = (-2.747,62)^2 / (18.690) = 403,9281$$

Paso 15) Se calcula la suma de cuadrados no explicada:

$$SS_{no\ ex.} = SS_E - SS_{ex.} = \text{Paso 5} - \text{Paso 14} = 423,7016 - 403,9281 = 19,7735$$

Paso 16) Se arma la Tabla de Anova completada con el análisis de regresión:

TABLA DE ANOVA + REGRESION

Fuente de Variación	Suma de cuadrados	Grados de Libertad	Cuadrados Medios	F
Entre Grupos**	423,7016	3	141,2339	11,20**

Regresión lineal	403,9281	1	403,9281	40,86*
Desvíos respecto a la Regresión	19,7735	2	9,8868	< 1(ns)

Dentro grupos (error)	138,6867	11	12,6079	
Total	562,3883	14		

Para comprobar si las desviaciones respecto de la regresión lineal son significativas se hace el ensayo: $F = MS_{Y;X} / MS_D < 1$ por lo tanto se acepta la H_0 , que las desviaciones respecto a la regresión lineal son nulas. Esto significa que no hay variación residual, o dispersión, alrededor de la línea de regresión. Por lo tanto se acepta a la recta como una buena explicación.

El siguiente ensayo es para determinar si existe la regresión lineal, es decir si b difiere significativamente de cero. Para eso se hace el ensayo:

$$F = MS_b / MS_{Y;X} = (403,9281) / (9,8868) = 40,86^* \gg F_{0,95; 1; 2} = 18,5$$

Luego se tiene evidencia significativa, como para afirmar que existe una recta de regresión que explica la regresión lineal de la supervivencia, respecto a la densidad de siembra. Resta entonces encontrar dicha recta:

Paso 17) Se calcula el coeficiente de regresión b con:

$$b = \frac{(\sum XY)}{\sum X^2} = \text{Paso 13} / \text{Paso 12} = (-2.747,62) / (18.690) = -0,14701$$

Paso 18) Se calcula la ordenada al origen a con:

$$a = \bar{Y} - b \bar{X} = (T / N) - b [(\sum Ni X) / N] = (907,81 / 15) - (-0,14701 [555 / 15]) = 65,96$$

Por lo tanto la recta de regresión se expresa con:

$$Y^* = 65,96 - 0,14701 \cdot X$$

Se ha probado que a medida que la densidad de siembra aumenta, la supervivencia disminuye. Y que esta relación se puede expresar con la ecuación de arriba.

21.9 Curvas de regresión

El caso anterior era el más simple, cuando la curva de regresión es una recta. Pero para el caso más general la *curva de regresión* toma una forma polinomial con:

$$Y^* = a + b X + c X^2 + d X^3 + \dots$$

La idea es que cualquier curva puede ser aproximada con un desarrollo en serie polinomial. Ahora se tiene un conjunto de potencias crecientes de la variable independiente X, cada una con un coeficiente de regresión diferente: a, b, c, d, etc. Por ejemplo, en el caso de una parábola habrá tres términos polinómicos. A medida que se utilicen potencias más altas, el ajuste de la curva de regresión a los datos reales, será cada vez mejor. Sin embargo, con cada potencia añadida se perderá un grado de libertad y se necesitarán más mediciones. Si n = 5 datos, para el cuadrado medio residual o error, los grados de libertad son n-2 = 3, entonces el polinomio mayor que se podrá usar es el de tercer grado. Por otra parte, es muy raro encontrar polinomios de más de tres grados en las investigaciones biológicas. Los más comunes son:

$$Y^* = a + b X \quad (\text{recta de regresión})$$

$$Y^* = a' + b' X + c' X^2 \quad (\text{parábola de regresión})$$

$$Y^* = a'' + b'' X + c'' X^2 + d'' X^3 \quad (\text{parábola cúbica de regresión})$$

Los coeficientes de cada una de las tres curvas anteriores son diferentes, por lo que deben ser calculados cada vez. Luego de obtenida la recta, se puede aumentar una potencia de X y buscar la parábola de regresión. Pero entonces, hay que recomenzar los cálculos de nuevo y por regla general, los nuevos coeficientes hallados (a', b') son diferentes de los anteriores (a, b). Como estas regresiones polinomiales son ajustes empíricos, si al comprobar la significación esta resulta significativa, significa que ahora se tiene un mejor ajuste que el lineal y conviene intentar la parábola cúbica.

Se recomienza todo de nuevo y los nuevos coeficientes serán diferentes a los anteriores con lo que la significación deberá ser testeada otra vez. Es claro, que si antes de comenzar se hubiese tenido información acerca del tipo de polinomio buscado, se hubiera comenzado por allí y no con la recta. Los cálculos y ensayos relacionados con este tema se pueden encontrar en el libro en Steel y Torrie (1960) mencionados en la bibliografía.

21.10 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|--|---|---|
| 1) Para un mismo problema se puede usar Regresión o Correlación. | V | F |
| 2) El análisis de regresión se usa cuando se conoce la relación teórica $Y = f(X)$. | V | F |
| 3) Para el análisis de peso y talla se puede usar la regresión. | V | F |
| 4) El método de los mínimos cuadrados es el usado en regresión. | V | F |
| 5) Explicar los diseños experimentales posibles en regresión:..... | | |
| 6) En un Modelo I la variable dependiente se mide sin error. | V | F |
| 7) Explicar las 4 hipótesis básicas de estos modelos:..... | | |
| 8) Todas las muestras a lo largo de la curva de regresión son homocedásticas. | V | F |
| 9) Las ecuaciones paramétricas de regresión se deducen minimizando la suma de cuadrados. | V | F |
| 10) Expresar el par de ecuaciones normales de regresión:..... | | |
| 11) La ecuación de regresión de Y sobre X, es igual a la de X sobre Y. | V | F |
| 12) Explicar los pasos básicos para los cálculos cortos de regresión:..... | | |
| 12) La variación Total es la suma del cuadrado de las diferencias entre cada Y y su media. | V | F |
| 13) La Variación No Explicada es la suma del cuadrado de las diferencias de Y con X. | V | F |
| 14) La Variación explicada es la diferencia de VT – VNE. | V | F |
| 15) El ensayo de hipótesis en regresión se hace con una Tabla de ANOVA. | V | F |
| 16) Para testear si hay regresión se puede usar el Modelo de Student. | V | F |
| 17) Para testear si la recta pasa por el origen hay que:..... | | |
| 18) Una curva de calibración instrumental debe pasar por el origen. | V | F |
| 19) Es lo mismo que haya uno o más valores de Y por cada X en los cálculos. | V | F |
| 20) Cuando hay más de un valor de Y primero se testea que haya regresión. | V | F |
| 21) Las curvas de regresión se resuelven con la técnica de polinomios ortonormales. | V | F |
| 22) Las transformaciones más comunes para regresión son:..... | | |
| 23) Para las curvas de dosis-mortalidad se usa la transformación probit. | V | F |

2) Se desea encontrar la recta de regresión para los siguientes datos, y luego validar los resultados obtenidos tanto para a como para b:

X	Y
0	5
10	20
20	40
30	60
40	78
50	98
60	118
70	135
80	152

3) Se desea saber si la balanza investigada está calibrada. Los datos recogidos son:

X	1	2	5	10	20	50	100	150	200	500	1000	1500	2000	3000
Y	1,2	1,9	4,8	10,5	21	49	98	152	202	496	1010	1520	1986	3020

4) Encontrar la recta de regresión para cuando hay 5 valores de Y por cada X, realizando las validaciones correspondientes:

		X			
		15	20	30	50
Y		65	60	53	43
		68	58	52	44
		67	59	51	42
		69	57	54	41
		70	58	50	40

5) Encontrar la recta de regresión para cuando hay 3 valores de Y por cada X, realizando las validaciones correspondientes:

		X			
		15	20	30	50
Y		65	60	53	43
		68	58	52	44
		67	59	51	42

6) Para los datos de la tabla siguiente decidir si existe una recta de regresión entre las dos variables presentadas.

X	95	115	135	155	175	195	221	235	255	275
Y	90	100	120	140	160	170	180	190	200	215

7) Para los datos de la tabla siguiente decidir si existe una recta de regresión entre las dos variables presentadas.

X	Y
40	16
60	28
80	43
100	55
120	79
140	96
160	110
180	125
200	150

Apéndice 1:

La diferencia entre un valor medido (Y_i) y su valor estimado (Y_i^*) es $D_i = Y_i - Y_i^*$
Para minimizar la suma de los cuadrados de estas diferencias se usan las relaciones:

$$\frac{\partial}{\partial a} \sum D_i^2 = 0$$
$$\frac{\partial}{\partial a} \sum [Y_i - (a + bX_i)]^2 = 0$$

Si la sumatoria es convergente, la derivada de la suma es igual a la suma de las derivadas

$$0 = \sum 2.[Y_i - (a + bX_i)][0 - 1 - 0] = (-2) \sum [Y_i - a - bX_i] = \sum Y_i - \sum a - \sum bX_i$$

Operando con esta relación se obtiene la primera ecuación paramétrica de la recta que es:

$$\sum Y_i = a.N + b \sum X_i$$

Para obtener la segunda ecuación paramétrica de la recta de regresión se usa la relación:

$$\frac{\partial}{\partial b} \sum D_i^2 = 0$$
$$\frac{\partial}{\partial b} \sum [Y_i - (a + bX_i)]^2 = 0 \quad \text{Análogamente al anterior si la sumatoria converge es:}$$

$$0 = \sum 2.[Y_i - (a + bX_i)][0 - 0 - X_i] = (-2) \sum [Y_i.X_i - a.X_i - b.X_i^2]$$

Resolviendo esta relación se obtiene la segunda ecuación paramétrica:

$$\sum X_i Y_i = a. \sum x_i + b \sum X_i^2$$

Para mostrar que es un mínimo, basta derivar una vez más y ver que el resultado sea positivo

$$\frac{\partial^2}{\partial^2 a} \sum D_i^2 = (-2) (-N) = 2 N$$

$$\frac{\partial^2}{\partial^2 b} \sum D_i^2 = (-2) (-\sum X_i^2) = 2 \sum X_i^2 \quad (\text{que siempre será un número positivo})$$

Con esto se demuestra que la recta de regresión minimiza la suma de los cuadrados de las diferencias entre cada valor medido y su correspondiente valor estimado con la ecuación de la recta. Por eso también se la llama: *recta de mínimos cuadrados*.

Apéndice 2:

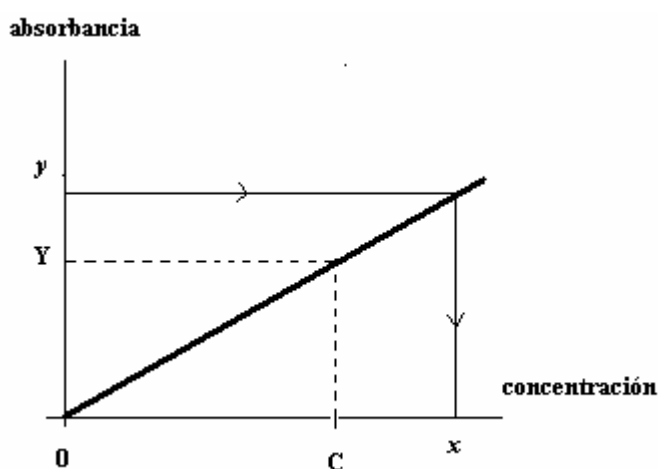
En forma diaria el bioquímico prepara su espectro para realizar las mediciones. Lo primero es el ajuste de cero. Para esto carga agua destilada en la cubeta del espectro (previamente nivelado) y el resultado tiene que dar una absorbancia nula (o 100% de transmitancia). Ajusta el aparato para tener esos valores, es decir para una concentración nula ($X = 0$) debe tener una absorbancia nula ($Y = 0$). Este es el primer punto de la recta de calibración denominado ajuste de cero.

Para sacar el segundo punto usa una sustancia calibrada, control o estándar (que viene provisto con cada kit de mediciones). Suponiendo que el estándar (de glucosa) tiene una concentración de 0,9 (expresada en las unidades habituales de trabajo), entonces usa el kit para proceder a la medición y obtiene una ligera coloración de la sustancia final que coloca en la cubeta del espectro. Esta coloración hace que no toda la luz enviada atraviese la cubeta, sino que un cierto porcentaje será absorbido (o desviado) y mide con el espectro la absorbancia ($Y = A$) que corresponde a una sustancia que tiene una concentración ($X = C$) de glucosa.

La costumbre es calcular el *factor* (k) con esos dos valores como si fuera una regla de tres simple: si el paciente tiene una absorbancia y entonces tendrá una concentración x calculada con:

$$x = (C / A) y = k y$$

Lo que se está haciendo en realidad es usar una recta de calibración:



Donde la tangente del ángulo de la recta es la inversa del factor. A cada paciente se le mide la absorbancia y para obtener la concentración buscada x .

Notar que este procedimiento diario implica que solo dos puntos son suficientes para obtener una recta de calibración del espectro. Cuando en realidad habría que usar más puntos y con ellos el procedimiento de regresión, para obtener una buena recta de calibración.

22

Análisis de Correlación

En este capítulo se continúa con el estudio de estadísticas en dos dimensiones. Cuando se trata de medir el *grado de asociación entre dos magnitudes* biológicas cualquiera se usa el Análisis de Correlación, mientras que cuando se trata de la relación funcional entre ambas se usa el Análisis de Regresión, como se vio en el capítulo anterior. Aquí, se comienza comparando los conceptos de regresión y correlación para ver cuando se aplica uno u otro modelo. Luego se explica la fórmula del producto-momento de Pearson y sus relaciones matemáticas básicas. Los cálculos básicos se ejemplifican para muestras pequeñas y medianas. Se plantean los tests de significación en correlación más comunes. Y por último se muestran algunas aplicaciones básicas de la correlación.

22.1 Introducción

No siempre resulta sencillo saber cual de los dos métodos se debe emplear en el estudio de un problema dado. Ha habido confusión en la literatura y entre los investigadores, al respecto. Por eso, se insistirá una vez más en tratar de distinguir claramente entre ambos casos. Es muy frecuente encontrar casos de correlación tratados como una regresión y viceversa. Hay varias razones para ello:

- Las relaciones matemáticas entre ambos modelos son muy estrechas, se puede pasar de uno a otro con mucha facilidad en los cálculos y eso siempre ha sido una tentación muy fuerte. Básicamente, el cuadrado del coeficiente de correlación es el cociente entre la variación explicadas y la total, que se calculan exactamente igual en ambos modelos:

$$r = \frac{(\sum xy)}{\sqrt{\sum x^2 \sum y^2}} \text{ es el } \textit{coeficiente de correlación} \text{ (fórmula del producto-momento de K.R. Pearson)}$$

$$b = \frac{\sum xy}{\sum x^2} \text{ es el coeficiente de regresión lineal}$$

Y la relación entre ambos es: $r^2 = b \frac{\sum xy}{\sum y^2}$

Esta última relación no tiene ningún significado conceptual, simplemente se trata de una analogía en los cálculos. Desdichadamente, muchos la usan como si fuera lo mismo y mezclan la regresión con la correlación, lo cual no es correcto.

- En los textos antiguos nunca se hizo una distinción lo suficientemente clara entre los dos conceptos. Aún hoy, esto no está totalmente superado, sobretodo por las reimpressiones actuales de los textos clásicos escritos en las primeras décadas del siglo pasado. Hay autores que usan ambos términos como si fuesen sinónimos, lo que aumenta más la confusión.

- Los métodos empíricos desarrollados para ingeniería y otras disciplinas, se usan para simplificar las relaciones reales a términos prácticos. Por ejemplo, las curvas de rendimiento de motores eléctricos, de máquinas térmicas, etc. En esos casos, no interesa la relación entre ambas magnitudes sino obtener una gráfica empírica para el uso diario.

- A veces, aunque el método escogido es el correcto, los datos disponibles no permiten aplicarlo.

Se puede revisar este tema desde el punto de vista del investigador. Tomando por caso un experimento donde se desea establecer el contenido de colesterol en la sangre humana como función del peso corporal. Para independizarse del factor edad y sexo se escoge en forma aleatoria un grupo de personas del mismo sexo y edad, midiendo en cada una ambos valores. Con el grupo de valores así obtenido se puede calcular la regresión entre ambas magnitudes. Sin embargo, el investigador no tiene bajo su control a ninguna de las variables; por lo tanto, las condiciones básicas de un Modelo I de regresión no se cumplen porque ambas variables han sido medidas con error. Casos como este abundan en la bibliografía a pesar de no ser legítimos. Se podría pensar que se trata de un Modelo II de regresión, pero salvo casos muy especiales como el de Berkson, esto tampoco es correcto. En un Modelo II la relación entre ambas variables puede escribirse con la ecuación:

$$\mu_y = \alpha + \beta \mu_x$$

El verdadero valor de Y es igual al valor poblacional de α , más valor poblacional del coeficiente de regresión multiplicado por el verdadero valor de X. Pero como ambas variable se miden con error, será $Y = \mu_y + \epsilon_y$ (donde ϵ_y es el término de error expresado con una distribución normal) y $X = \mu_x + \epsilon_x$ (donde ϵ_x es el término de error de la variable X).

Luego, la covarianza de ambas sería:

$$\sum XY = \sum (\mu_x + \epsilon_x)(\mu_y + \epsilon_y) = \sum [(\mu_x \mu_y) + (\mu_y \epsilon_x) + (\mu_x \epsilon_y) + (\epsilon_x \epsilon_y)] \quad \mu_x$$

Donde se espera que, salvo el primer término, todos los demás se anulen pues las partes debidas al error son independientes entre sí y además de ambos valores esperados. Sin embargo, las desviaciones $D_i = Y_i - Y^*i$ no son ahora independientes de Y^*i , por lo que se invalidan los tests de significación convencionales. El caso especial de Berkson, es cuando las variables independientes se miden con error, aunque son "controladas" por el investigador que posee información acerca de las respectivas varianzas y sabe que los valores X y ϵ_x no están correlacionados.

En conclusión, lo correcto en el caso del experimento de nivel de colesterol y el peso corporal no es hacer una regresión, sino un análisis de correlación entre ambas variables, para el cual esos datos son convenientes, si lo que se busca es alguna ecuación que describa la dependencia de Y (colesterol) sobre el peso (X). El caso contrario es cuando se quiere hallar un coeficiente de correlación, cuando los datos se han calculado apropiadamente como de regresión. Por ejemplo: medir los latidos del corazón en función de la temperatura ambiente. En este caso, el investigador puede decir que los valores de temperatura fueron elegidos al azar. Pero en el fondo se sabe que tales valores son “controlables” por el investigador mediante un adecuado sistema de frío. Se somete al sujeto a una cierta temperatura (elegida al azar o no) regulada por un termostato y se miden los latidos. O sea, se puede calcular el coeficiente de correlación con estos datos, pero solo sería un valor numérico en vez de un valor paramétrico de la correlación. A pesar, de que se puede relacionar el cuadrado de este coeficiente con el cociente entre la variación explicada y la total, no es de ninguna manera una indicación de correlación paramétrica. Se pueden resumir estos conceptos en un cuadro como el siguiente:

Cuadro 22.1: Regresión versus Correlación.

El investigador quiere:	Y aleatoria, X fija	X e Y aleatorias
Establecer y estimar la <i>dependencia</i> de una variable sobre la otra.	Modelo I de Regresión.	Modelo II de Regresión. (caso especial de Berkson)
Establecer y estimar el grado de <i>asociación</i> entre ambas variables	No es significativa.	Análisis de Correlación. (test de significación válido solo si X e Y son variables normales bidimensionales)

22.2 Fórmula del producto-momento

La forma más general de esta fórmula para obtener el *coeficiente de correlación r* es:

$$r = \frac{S_{xy}}{DS_x DS_y}$$

Dónde:

S_{xy} : es la covarianza entre ambas variables.

DS_x : es el desvío estándar de una variable cualquiera X.

DS_y : es el desvío estándar de otra variable cualquiera Y.

$$S_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{(n - 1)} = \frac{\sum x y}{(n - 1)}$$

$$DS_x = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} = \sqrt{\frac{\sum x^2}{n - 1}} \quad \text{Análogamente, } DS_y = \sqrt{\frac{\sum y^2}{n - 1}}$$

Reemplazando y simplificando la fórmula del producto momento queda:

$$r = \frac{\sum x y}{\sqrt{\sum x^2 \sum y^2}} : \text{coeficiente de correlación}$$

Se puede reagrupar esta ecuación de la manera siguiente:

$$r^2 = \frac{(\sum xy)^2}{\sum x^2 \sum y^2} = \frac{(\sum xy)^2}{\sum x^2} \cdot \frac{1}{\sum y^2} = \sum (x^*)^2 \cdot \frac{1}{\sum y^2} = \frac{\sum (X^* - \bar{X})^2}{\sum (Y - \bar{Y})^2}$$

$$r^2 = \frac{\text{Variación Explicada en X}}{\text{Variación Total en Y}} \quad (\text{ver punto 21.3 del capítulo anterior})$$

En una correlación, el cuadrado del coeficiente de correlación es el cociente entre la Variación Explicada de una de las variables y la Variación Total de la otra. Este cociente, se denomina *coeficiente de determinación* y puede variar entre cero y uno. En un caso extremo, cuando la covarianza entre ambas variables no existe, valdrá cero y se piensa que no hay ningún grado de correlación entre ambas. En el otro extremo, cuando el grado de asociación entre ambas variables es perfecto valdrá uno. Por su parte, el coeficiente de correlación variará entre -1 y $+1$.

22.3 Cálculo del coeficiente de correlación

La manera más sencilla es resolviendo el siguiente ejemplo. En una investigación se eligieron al azar nueve individuos de aproximadamente 30 años, de una misma ciudad, considerados sanos. A cada uno de ellos se le midió el peso y el nivel de colesterol en sangre. Los resultados se muestran a continuación. Hallar el coeficiente de correlación entre ambas variables.

Colesterol	Peso	x =	y =	x ²	y ²	x.y	
X	Y	X-214	Y-73				
210	70,2	-4	-2,8	16	7,84	11,2	
122	62,4	-92	-10,6	8464	112,36	975,2	
309	95,4	95	22,4	9025	501,76	2128	
198	68,9	-16	-4,1	256	16,81	65,6	
260	75,2	46	2,2	2116	4,84	101,2	
230	76	16	3	256	9	48	
175	64,5	-39	-8,5	1521	72,25	331,5	
198	64,2	-16	-8,8	256	77,44	140,8	
224	80,2	10	7,2	100	51,84	72	
Total	1926	657	0	0	22010	854,14	3873,5

Media 214 73

$$r^2 = 0,7981013$$

Paso 1) Se vuelcan los datos a una tabla como la anterior y se calculan los totales y promedios respectivos.

Paso 2) Se calculan las diferencias de cada dato respecto a su media y se vuelcan los resultados en las columnas tercera y cuarta.

Paso 3) Los valores obtenidos en el paso anterior se elevan al cuadrado y se colocan en las columnas quinta y sexta. Y en la última columna se coloca el producto de los valores de tercer y cuarta columna.

Paso 4) Los totales de las tres últimas columnas se usan para el cálculo del coeficiente:

$$\sum x^2 = 22010 \quad ; \quad \sum y^2 = 854,14 \quad ; \quad \sum x y = 3873,5$$

$$\text{Luego } r = \frac{3873,5}{\sqrt{(22010)(854,14)}} = 0,8933$$

22.4 Tests de significación en correlación

Hay varias maneras de efectuar los tests de significación en correlación, de acuerdo a la hipótesis nula que se plantee. Básicamente se agrupan en cuatro casos:

- . $H_0 : \rho = 0$ No hay correlación. El verdadero valor de r es nulo.
- . $H_0 : \rho = r$ El coeficiente de correlación vale r .
- . $H_0 : \rho_1 = \rho_2$ No hay diferencias entre dos coeficientes de correlación.
- . $H_0 : \rho_1 = \rho_2 = \rho_3 = \dots = \rho_k$ Todas las muestras provienen de la misma población.

Caso 1) Cuando se postule $H_0 : \rho = 0$, se tiene el caso más sencillo en la correlación de dos variables. El planteo supone que no existe la correlación entre ambas variables y hay tres maneras de proceder. La primera es usando la Tabla 20 del Anexo donde en la primer columna, para un número $k = 1$, y los grados de libertad obtenidos con $v = n - 2$, se obtienen los valores críticos para 95% y 99% de confianza. La segunda forma es usando el test clásico con el modelo Student. Y la tercera, cuando las muestras son grandes ($n > 50$), con la transformación Z de Fisher, estadígrafo con una distribución aproximación normal.

Ejemplo 1) Usando los datos del problema visto en el punto anterior determinar si los resultados permiten validar que hay correlación.

Para esto, se plantea una hipótesis a ser rechazada para tener la validación buscada. Esto es:

$H_0 : \rho = 0$ No hay correlación. El verdadero valor de r es nulo.

$H_0 : \rho \neq 0$ Hay correlación. El verdadero valor de r es diferente de cero.

Método 1) Se busca en la primer columna de la Tabla 20 del Anexo, con $v = n - 2 = 7$ grados de libertad y resulta: $R_{0,95;7} = 0,666$ y $R_{0,99;7} = 0,798$

Entonces $r = 0,8933^{**}$ y se rechaza la hipótesis nula. Por lo tanto, se tiene evidencia muy significativa de correlación.

Método 2) Se usa el modelo Student de dos colas, comparando con tablas al estadígrafo $t_{\alpha;7=n-2}$:

$$t = \frac{(r - 0)}{\sqrt{\frac{(1 - r^2)}{(n - 2)}}} = r \cdot \sqrt{\frac{(n - 2)}{(1 - r^2)}} = 0,8933 \cdot \sqrt{\frac{(9 - 2)}{(1 - 0,8933^2)}} = 5,26^{**} > t_{0,99;7}$$

La conclusión es similar a la vista con el método anterior.

Ejemplo 2) Phillips en 1929 midió las longitudes de la vena transversal en las alas derecha e izquierda de 500 abejas obreras y obtuvo un coeficiente de correlación $r = 0,837$. Determinar si el coeficiente de correlación es significativo.

$H_0 : \rho = 0$ No hay correlación. El verdadero valor de r es nulo.

$H_0 : \rho \neq 0$ Hay correlación. El verdadero valor de r es diferente de cero.

Se emplea el estadígrafo t^* y se lo compara con la t de Student para infinitos grados de libertad. O sea, se trata de una distribución gaussiana y por ese motivo se prefiere llamarlo $z = t^*$. Donde Z se calcula con la transformación de Fisher y el desvío estándar poblacional es la inversa de la raíz cuadrada del número de mediciones, menos 3. Esto es:

$$t^* = z = \frac{(Z - 0)}{\frac{1}{\sqrt{n - 3}}} = Z \sqrt{(n - 3)}$$

donde $Z = 0,5 \ln \left[\frac{1+r}{1-r} \right] = 0,5 \ln [11,27] = 1,2111$

$z = 1,2111 (22,29) = 27^{***} \gg z_{0,001} = z_{\alpha}$ de la función de Gauss

Aplicando la fórmula gaussiana al valor $z = 27$ se obtiene una probabilidad muy pequeña (del orden de 10^{-6}) lo que permite rechazar la hipótesis nula con mucha evidencia significativa.

Caso 2) El otro caso es $H_0 : \rho = a$ la hipótesis supone un valor cualquiera a para el coeficiente de correlación poblacional. Aquí se emplea otra vez el modelo de Fisher, para obtener un estadígrafo z con una distribución gaussiana, para muestras grandes ($n > 50$) con:

$$z = \frac{Z - \mu_z}{\sigma_z} \quad \text{Donde} \quad \mu_z = 0,5 \ln \left[\frac{1+a}{1-a} \right] \quad \text{y} \quad SE(z) = \sigma_z = \frac{1}{\sqrt{n - 3}}$$

Si la muestra es más pequeña $10 < n < 50$ se debe efectuar una corrección.

Ejemplo 3) Con los datos del problema anterior se desea comprobar sí:

$H_0 : \rho = 0,5$ En este caso se emplea la transformación Z de Fisher con:

$H_1 : \rho \neq 0,5$

$$z = \frac{Z - \mu_z}{\sigma_z}$$

Donde:

$$\mu_z = 0,5 \ln \frac{(1 + \rho)}{(1 - \rho)} = 0,5 \ln (1,5/0,5) = 0,5493$$

$$\sigma_z = 1 / \sqrt{(n - 3)} = SE(z) = 0,045$$

$$Z = 0,5 \ln \frac{(1 + r)}{(1 - r)} = 0,5 \ln (11,27) = 1,2111$$

Por lo tanto:

$$z = (1,2111 - 0,5493) / 0,045 = 17,71*** \text{ (versus el valor gaussiano } z_{\alpha})$$

Se rechaza la hipótesis nula con valores altamente significativos. La probabilidad de que $\rho = 0,5$ es muy pequeña.

Ejemplo 4) Se ha tomado una muestra de $n = 12$ pares de datos de una población y se obtuvo un coeficiente de correlación $r = 0,8652$. Se desea saber si el coeficiente poblacional es igual a $0,8$.

$H_0 : \rho = 0,8$ En este caso se emplea la transformación Z de Fisher con la corrección

$H_1 : \rho \neq 0,8$ de Hotelling:

$$z = \frac{Z^* - \mu_z^*}{\sigma_z^*}$$

Donde:

$$Z = 0,5 \ln \frac{(1 + r)}{(1 - r)} = 0,5 \ln (1,8652/0,1348) = 1,3137$$

$$\mu_z = 0,5 \ln \frac{(1 + \rho)}{(1 - \rho)} = 0,5 \ln (1,8/0,2) = 1,0986$$

Pero ahora se debe corregir con:

$$Z^* = Z - [(3Z + r) / (4n - 4)] = 1,3137 - [(4,8062) / 44] = 1,2045$$

$$\mu_z^* = \mu_z - [(3\mu_z + \rho) / (4n)] = 1,0986 - [(4,0958) / 48] = 1,0133$$

$$\sigma_z^* = SE(z) = 1 / \sqrt{n - 1} = 1 / \sqrt{11} = 0,3015$$

Luego es: $z = \frac{Z^* - \mu_z^*}{\sigma_z} = (1,2045 - 1,0133) / (0,3015) = 0,634$ (no significativo)

Con este resultado no se puede rechazar la hipótesis nula.

Caso 3) Ahora es $H_0 : \rho_1 = \rho_2$ Se supone que hay dos muestras tomadas de la misma población y tienen el mismo coeficiente de correlación poblacional. Se usa la transformación de Fisher para la comparación de ambos coeficientes de correlación, en forma similar a la tratada en el modelo gaussiano para diferencia de medias muestrales independientes. Esto es el estadígrafo:

$z = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$ se distribuye normalmente y se puede comparar con el de tablas.

Ejemplo 5) Se desea determinar si hay diferencias significativas entre dos coeficientes de correlación $r_1 = 0,8$ y $r_2 = 0,5$, calculados con muestras de tamaños 50 y 60 respectivamente.

$H_0 : \rho_1 = \rho_2$

$H_1 : \rho_1 \neq \rho_2$

$Z_1 = 0,5 \ln \frac{(1+r_1)}{(1-r_1)} = 0,5 \ln (1,8/0,2) = 1,0986$

$Z_2 = 0,5 \ln \frac{(1+r_2)}{(1-r_2)} = 0,5 \ln (1,5/0,5) = 0,5493$

Luego es:

$$z = (1,0986 - 0,5493) / \sqrt{(1/47) + (1/57)} = 0,5493 / 0,197 = 2,79^{**}$$

Se rechaza la hipótesis nula con resultados muy significativos.

22.5 Comparación entre dos o más coeficientes

Hay oportunidades en que se miden las dos magnitudes clínicas en varias muestras. Se obtienen de esta forma varios coeficientes de correlación. Esto es, como si se tuviesen varias mediciones del mismo índice. Lo que se desea averiguar es si todos ellos provienen de la misma población cuyo verdadero coeficiente de correlación es $\Xi (r) = \rho$. Puede ser aplicado en casos donde se toma en cuenta la raza, la zona geográfica donde habitan, la edad, el sexo, etc. Una manera de decidir si los coeficientes de correlación hallados son homogéneos entre sí, es encontrar el valor poblacional con una estimación a través de todas las muestras y realizar una validación a través del Modelo de Fisher para correlación. Se puede hacer tomando una suma de cuadrados ponderada de los valores de los coeficientes r , transformados con Z , la cual se distribuye con una distribución Chi cuadrado con $N - 1$ grados de libertad.

Para ilustrar el procedimiento se ha elegido el siguiente caso:

Ejemplo) En un estudio del Ministerio de Salud Pública, en diez localidades de la provincia de Misiones, se tomaron muestras al azar de diferentes tamaños de su población y a cada individuo seleccionado se le midió peso y talla. El sexo se repartió mitad y mitad en cada grupo. Las edades corresponden a individuos que se hallan cursando la escuela primaria. En la Tabla 24.1 siguiente:

Tabla 24.1: Correlación entre talla y peso en 10 localidades misioneras.

n	n-3	r	Z	Z ²	(n-3)Z	(n-3)Z ²
200	197	0,42	0,4477	0,20044	88,1969	39,4858
150	147	0,45	0,4847	0,23493	71,2509	34,5353
140	137	0,49	0,5361	0,2874	73,4457	39,3742
130	127	0,48	0,523	0,27353	66,421	34,7382
90	87	0,57	0,6475	0,41926	56,3325	36,4753
120	117	0,51	0,5627	0,31663	65,8359	37,0459
160	157	0,47	0,5101	0,2602	80,0857	40,8517
50	47	0,55	0,6184	0,38242	29,0648	17,9737
160	157	0,49	0,5361	0,2874	84,1677	45,1223
170	167	0,61	0,7089	0,50254	118,3863	83,924

Total: **1370** **1340** **733,1874** **409,5264**

En la primer columna se colocan los respectivos tamaños muestrales de cada localidad. En la tercer columna se vuelcan los respectivos coeficientes de correlación encontrados. Los pasos a seguir son:

Paso 1) Se calculan los tamaños muestrales menos tres y se colocan en la segunda columna.

Paso 2) Se obtienen los valores de r transformados con la relación Z:

Por ejemplo, para la primer localidad será:

$$Z_1 = 0,5 \ln \frac{(1+r_1)}{(1-r_1)} = 0,5 \ln (1,42 / 0,58) = 0,4477$$

Paso 3) Una vez calculados todos los valores de Z se vuelcan en la cuarta columna, y sus cuadrados se colocan en la quinta columna. O sea, en la quinta columna se colocan los resultados de la cuarta multiplicados entre sí.

Paso 4) Luego a cada valor de estas columnas se lo multiplica por su tamaño muestral respectivo y los resultados se colocan en las dos últimas. O sea, (n-3) se multiplica por la cuarta columna y el resultado se coloca en la sexta. Lo mismo con (n-3) por la quinta y se coloca en la séptima.

Paso 5) Se obtiene el tamaño muestral corregido T con:

$$N = \sum Ni = 200 + 150 + \dots + 170 = 1.370$$

$$T = \sum (Ni - 3) = (200-3) + (150 - 3) + \dots + (170 - 3) = 1.340$$

O bien, $T = N - 3 (10) = 1.370 - 30 = 1.340$

Paso 6) Se calcula el promedio de Z con:

$$\Xi(Z) = \frac{\sum (N_i - 3) Z_i}{\sum (N_i - 3)} = \frac{\text{Total de columna sexta}}{\text{Total de columna segunda}} = (733,1874) / 1340 = 0,54715$$

El valor promedio de Z ponderado es: $\Xi(Z) = 0,547155$

Paso 7) Se calcula la suma de cuadrados de Z ponderada con:

$$SSz = \sum (N_i - 3) Z_i^2 = \text{Total de la columna séptima} = 409,5264$$

Paso 8) Se calcula el término de corrección:

$$TC = \Xi(Z) \cdot \sum (N_i - 3) Z_i = \text{Paso 6} \cdot \text{Total de la columna sexta} = (0,547155) \cdot (733,1874)$$

$$TC = 401,167$$

Paso 9) Se efectúa el test de homogeneidad:

H_0 : $r_1 = r_2 = \dots = r_k$ Todos los coeficientes provienen de la misma población.

H_1 : Los r_i son diferentes y no provienen de la misma población.

El estadígrafo de comparación es:

$$X^2 = SSz - TC = \sum (N_i - 3) Z_i^2 - \Xi(Z) \cdot \sum (N_i - 3) Z_i = 409,5264 - 401,1677 = 8,359$$

La comparación es:

$$X^2 = 8,3587 < \chi^2_{0,95; 9} = 16,919$$

No hay suficiente evidencia como para rechazar la hipótesis nula.

Cálculo del coeficiente de correlación poblacional.

Paso 10) Con el valor promedio de Z, se calcula la mejor estimación de ρ :

$$\Xi(r) = \frac{e^{2\bar{z}} - 1}{e^{2\bar{z}} + 1} = \frac{e^{2(0,547155)} - 1}{e^{2(0,547155)} + 1} = 0,4984 \approx \rho \quad (\text{Notar que no es el promedio ponderado de los } r).$$

22.6 Modelo no paramétrico de Kendall

Hay ocasiones en que se sabe que los datos no se distribuyen mediante una distribución normal bi-variante. Otras veces las magnitudes medidas no son de tipo continuo sino cualitativas de tipo ordinal, que permitan comparar a las dos muestras, tales como puntajes, etc. En todos estos casos no se pueden emplear los modelos vistos hasta ahora y se requiere de un modelo no paramétrico equivalente como el de Kendall, para una correlación ordenada. Hay otros modelos para distintas variantes; el lector interesado los puede encontrar en la obra de Siegel indicada en la bibliografía. La condición es que los datos se puedan medir en por lo menos una escala ordinal

El ejemplo típico es como se ordenan a los alumnos de un curso de Bioestadística por sus calificaciones de parciales y de concepto. Por ejemplo, se puede afirmar que A es el mejor alumno, B es el segundo mejor estudiante, que C y D son iguales entre sí pero no son tan buenos como B y así sucesivamente. En otra materia con los mismos alumnos, otro profesor puede ordenarlos según su criterio, respecto a como se comportan en sus clases. Se supone que los dos grupos de valores este correlacionados entre sí porque los alumnos son los mismos, aunque sometidos a diferentes exigencias. Sin embargo, para probar que esto ocurre se necesita un test de validación de correlación ordenada como el de Kendall. En una Farmacia se pueden ordenar los diferentes artículos de acuerdo a la cantidad vendida de cada uno y construir una tabla ordinal con los mejores del ranking. Lo mismo se puede repetir en una sucursal de la misma farmacia. Se espera que haya una correlación entre ambos listados con la ubicación ordenada de los artículos, y otra vez el modelo de Kendall puede aplicarse para verificar si esto es cierto. En un laboratorio de Análisis Clínicos se puede desarrollar la muestra extraída a un paciente en un caldo de cultivo apropiado. Luego de hacer el antibiograma, se pueden ordenar los resultados de acuerdo al diámetro de la aureola medido en la caja de Petri. Con la muestra de otro paciente que tenga la misma enfermedad se procede en forma similar, y se espera que los dos grupos de datos ordenados guarden alguna correlación. Para ilustrar el procedimiento se simulan datos para un problema como el de los antibiogramas, para 15 pacientes, midiendo el diámetro de la halo de inhibición.

N	Y1	R1	Y2	R2
1	8,7	8	5,95	9
2	8,5	6	5,65	4
3	9,4	9	6	10
4	10,0	10	5,7	6,5
5	6,3	1	4,7	2
6	7,8	5	5,5	3
7	11,9	15	6,4	15
8	6,5	2	4,18	1
9	6,6	3	6,15	13
10	10,6	12	5,93	8
11	10,2	11	5,7	6,5
12	7,2	4	5,68	5
13	8,6	7	6,13	12
14	11,1	13	6,3	14
15	11,6	14	6,03	11

A la muestra de cada paciente se la siembra en un caldo de cultivo apropiado. Se prueban dos antibióticos 1 y 2. Los diámetros de las aureolas resultantes se colocan en la segunda y cuarta columna, bajo Y1 e Y2 respectivamente. Luego se calculan los rangos que corresponden a cada muestra y se colocan en las columnas R1 y R2.

Así, el dato más pequeño de la muestra 1 es 6,3 y le corresponde el rango 1, el segundo con rango 2 es 6,5 y así sucesivamente hasta el más grande 11,9 con rango 15.

En la otra serie de datos se procede igual. Como hay un empate entre el rango 6 y 7, se coloca el rango promedio a ambos ($Y = 5,7$).

El coeficiente de correlación ordenada de Kendall es un estadígrafo y no un parámetro, pero generalmente se lo simboliza con la letra griega tau (τ). La fórmula para este estadígrafo es:

Si no hay empates $\tau = O / N(N-1)$

Si hay empates
$$\tau = \frac{O}{\sqrt{[N(N-1) - \sum E_2][N(N-1) - \sum E_1]}}$$

N es el tamaño muestral, en este caso es $N = 15$

O es una cantidad calculada en función del orden, que se puede determinar de varias formas.

$\sum E_1$: Es la suma de las cantidades empatadas en Y1.

$\sum E_2$: Es la suma de las cantidades empatadas en Y2.

La idea es que si la variable Y2 está perfectamente correlacionada con Y1, entonces deberían tener los mismos rangos, es decir, ordenadas de igual manera. Sin embargo, si la correlación es menor el orden de los Y2 no se corresponderá con los de Y1. La cantidad O mide el grado en que la *segunda* variable corresponde al orden de la primera. Su valor máximo será: $N(N-1)$.

A continuación se explican los pasos a seguir para resolver el problema:

Paso 1) Si una de las variables, como en este caso, tiene empates, se ordenan los pares según la variable que no los tenga. Si ambas tienen empates, entonces puede ser cualquiera. A continuación del ordenamiento según la primera variable, se colocan los correspondientes rangos de la otra. Para cada caso hay que calcular el número de rangos de la segunda más grande. Comenzando con $R1 = 1$, le corresponde $R2 = 2$. Salvo el $R2 = 1$, todos los demás son mayores, entonces habrá 13 rangos $R2$ mayores que $R2 = 2$ y así sucesivamente. Como se muestra a continuación:

R1	R2	Rangos que son superiores al R2 en cada caso:
1	2	13; 5; 3; 4; 12; 9; 10; 6,5; 6,5; 8; 14; 11; 15. O sea, el total: $C1 = 13$ rangos.
2	1	Ídem, es $C2 = 13$
3	13	14; 15 O sea, el total es $C3 = 2$
4	5	12; 9; 10; 6,5; 6,5; 8; 14; 11; 15. O sea, el total es $C4 = 9$
5	3	4; 12; 9; 10; 6,5; 6,5; 8; 14; 11; 15. Esto es, $C5 = 10$
6	4	Ídem anterior: $C6 = 9$
7	12	14; 15. O sea, $C7 = 2$
8	9	10; 14; 11; 15. Esto es, $C8 = 4$
9	10	14; 11 ; 15. O sea, $C9 = 3$
10	6,5	(6,5); 8; 14; 11; 15. En empate se suma medio punto: $C10 = 4,5$
11	6,5	8; 14; 11; 15. Ahora será: $C11 = 4$
12	8	14; 11; 15. Esto es, $C12 = 3$
13	14	Solo hay una mayor: 15. Por lo tanto, $C13 = 1$
14	11	Ídem: $C14 = 1$
15	15	$Y C15 = 0$

Paso 2) Se calcula la suma de puntajes C_i como: $C = \sum C_i = 13 + 13 + 2 + \dots + 1 = 78,5$

Paso 3) Se obtiene el valor del ordenamiento con:

$$O = 4 \sum C_i - N(N-1) = 4C - N(N-1) = 4(78,5) - [15(15-1)] = 314 - 210 = 104$$

Paso 4) Se calcula el estadígrafo de Kendall con:

Como hay empates
$$\tau = \frac{O}{\sqrt{[N(N-1) - \sum E_2][N(N-1) - \sum E_1]}}$$

Donde: $\sum E_1 = 0$ porque Y1 no tiene empates.

Y $\sum E_2 = 2$ porque hay 2 empates en Y2 con los rangos 6 y 7

Entonces:

$$\tau = \frac{104}{\sqrt{[15(15-1) - 0][15(15-1) - 2]}} = 104 / 209 = 0,4976$$

Paso 5) para comprobar la hipótesis de que hay correlación se puede usar la Tabla 21 del Anexo, o si las muestras son grandes, se puede usar la aproximación normal.

$$\tau = 0,4976^* \text{ versus } \tau_{\alpha;N} \text{ De tablas es } \tau_{0,95; 15} = 0,390 \text{ y } \tau_{0,99; 15} = 0,505$$

Luego se han encontrado resultados significativos que prueba que hay correlación: Esto es, se rechaza $H_0 : \tau = 0$.

Realmente la tabla 21 es exacta cuando no hay empates. En casos como el presente solo es aproximada y conviene usar una tabla especial presentada por Burr (1960).

22.7 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|---|---|---|
| 1) El coeficiente de correlación se usa para medir el grado de asociación de 2 variables. | V | F |
| 2) Las relaciones matemáticas entre los coeficientes b y r son muy estrechas y sin sentido. | V | F |
| 3) Si X e Y son aleatorias corresponde hacer un Modelo II de regresión. | V | F |
| 4) La correlación se usa cuando el investigador busca establecer el grado de asociación. | V | F |
| 5) Presentar en un cuadro los 4 casos posibles sobre que hacer con Regresión y Correlación... | | |
| 6) El coeficiente r es el cociente entre covarianza y el producto de los DSx y DSy . | V | F |
| 7) Explicar la fórmula del producto-momento de Pearson:..... | | |
| 8) El cociente entre la VE de X y la VT de Y es el coeficiente de correlación. | V | F |
| 9) El cuadrado del coeficiente de regresión es el coeficiente de determinación . | V | F |
| 10) Explicar los pasos a seguir para seguir los cálculos en correlación. | | |
| 11) Explicar los pasos para testear si hay correlación:..... | | |
| 12) Para testear correlación hay dos métodos: el de Student y otro con la Tabla 20. | V | F |
| 13) La transformación de Fisher permite usar la Chi cuadrado para el test de correlación. | V | F |

- 14) El $\ln[(1+r)/(1-r)]$ transforma r en Z. V F
 15) Los grados de libertad de Z son N-3. V F
 16) La tipificación de Z permite testear $H_0: \rho = a$ V F
 17) Si las muestras son grandes $n > 50$ se puede aproximar z con gauss. V F
 18) Si N está entre 10 y 25 se debe efectuar la corrección de Hotelling. V F
 19) Cuando se compara más de dos muestras, se testea que todas provengan de una población V F
 20) Para estimar el valor poblacional de r no se puede usar un promedio directo. V F
 21) Explicar los pasos a seguir en un test de homogeneidad de r en más de 2 muestras:.....
 22) El modelo de Wilcoxon se puede usar para testear si hay correlación. V F
 23) El modelo no paramétrico para la correlación es el de:.....
 24) Las magnitudes deben ser por lo menos ordinales en un test no paramétrico de correlación V F

2) Para los datos de la tabla siguiente decidir si:

- a) Existe correlación entre las dos variables.
 b) Si $\rho = 0,6$

Nº	X	Y
1	80	5
2	100	15
3	120	25
4	140	44
5	160	67
6	180	72
7	200	81
8	220	90
9	240	106
10	260	120

3) Resolver el mismo problema anterior con el modelo de Kendall.

4) Resolver el siguiente problema con el modelo de Kendall.

N	Y1	Y2
1	10	120
2	11	105
3	14	120
4	22	130
5	25	129
6	29	140
7	33	160
8	38	154
9	41	152
10	45	170
11	52	190
12	55	180

23

Control de Calidad instrumental

En este capítulo se desarrollan diferentes métodos para controlar precisión y exactitud en las mediciones de los laboratorios de Análisis Clínicos. El concepto de *calidad* se toma en ese sentido. Un instrumento será de buena calidad cuando tenga una exactitud inferior a la menor unidad de su escala y una precisión acorde. Para *controlar* este hecho se usan modelos estadísticos relacionados con el concepto de *calibración*. Usando un patrón de la magnitud clínica que mide el instrumento, se realizan una serie de mediciones, y entonces el valor promedio de estas deberá coincidir con el valor del patrón para que sea *exacto*, mientras que su dispersión no deberá superar un valor prefijado como el *máximo admisible*. En los instrumentos, por lo general la precisión del equipo viene especificada por el fabricante. En otros casos, la experiencia clínica en mediciones es el parámetro por el cual se suele guiar al profesional, para especificar tal valor. El modelo Student para una sola muestra se emplea para estudiar o probar la coincidencia entre el valor medio medido y el esperado; o sea, controlar la exactitud del instrumento. El modelo de la Chi cuadrado prueba el desvío estándar de las mediciones contra el máximo aceptable, es decir, controla la precisión. Estos conceptos se ejemplifican a lo largo de este capítulo, con los instrumentos más comunes del laboratorio como pipetas, buretas, centrífugas, capilares, espectrofotómetros, estufas de cultivo, etc.

Algunos instrumentos requieren un solo punto de calibración, otros en cambio necesitan más puntos de control. Por razones didácticas, el primer caso se desarrolla en este capítulo pues el concepto básico es el mismo. El segundo, llamado la *recta de calibración*, donde se toman varios puntos, se explicó con más detalle en el Tema 21, correspondiente a la *Regresión Estadística*. A veces, se necesita comparar varios instrumentos entre sí, para decidir por el más *conveniente*. En este capítulo, se muestran las comparaciones de a pares, usando el modelo Student para dos muestras independientes, para contrastarlos en exactitud. El modelo F de Fisher se usa para comparar sus variaciones y testarlos en precisión. Cuando se necesita comparar de dos a la vez, se debe usar los modelos de ANOVA.

23.1 Introducción

El problema básico del control es determinar que se tendrá primero: si un patrón, para calibrar un sistema de medición, o un sistema calibrado para fabricar un patrón. Cada laboratorio en particular deberá contestarse esa pregunta de acuerdo a sus posibilidades. Por caso, si se dispone de una balanza de precisión bien calibrada, se puede preparar un control acuoso para calibrar una técnica de laboratorio. En cambio, si se disponen de pesas patrones, se puede calibrar la balanza primero para luego preparar el control. Una vez resuelto este problema, se procede a calibrar el instrumento sin olvidar al factor humano. En efecto, salvo que el instrumento sea total-

mente automatizado, es la dupla equipo-hombre el sistema de medición elemental y deben tomarse en cuenta ambos factores, como posibles fuentes de error. Por ejemplo, un error de tipo sistemático como atraso al medir tiempos, sobrepeso en una balanza, etc. se debe al instrumento; pero un error de paralaje, cuando el observador no se coloca bien enfrente de la escala de lectura, o no pone a cero una serie de pesadas, ya es cuestión del hombre. En cada uno de los dos factores se pueden producir desviaciones que alteren significativamente los resultados finales de la medición. Por lo tanto, es necesario diseñar controles para poder detectar tales desviaciones por separado, lo mismo que para todo el conjunto. La ventaja principal es que, detectando por separado las fluctuaciones, se facilita la tarea de corregir los defectos de tipo sistemático. Pero, siempre existirá una fuente de variación imposible de anular debido al azar (errores casuales), que deberá ser por lo menos acotada numéricamente.

Una vez calibrados los instrumentos, el paso siguiente es calibrar las técnicas de laboratorio, una por una, cosa que se verá en el capítulo siguiente. El tercer paso es mantener fiable a lo largo del tiempo el sistema de medición para asegurar los valores que se informan con el mismo, esto es, el "producto final" del laboratorio. El método usual para controlar regularmente una técnica de laboratorio es la denominada: *Carta de control de calidad*. Esto se verá en el Tema 25 de control estadístico de calidad.

En resumen, se calibran los instrumentos y luego se calibra el sistema por primera vez, luego, con la carta, se controla que se mantenga calibrado y cuando surge algún problema se vuelve al principio. Esto es, a recalibrar todo de nuevo. La estadística avisa de los problemas y valida las calibraciones, pero no explica como solucionar el problema. Esto se debe buscar en el conocimiento profesional del bioquímico o farmacéutico actuante, de su experiencia en control de calidad y en calibración de instrumentos.

23.2 Propagación de errores

De acuerdo a la *Hipótesis de Haeguen-Bessel*, el error total casual de una medición cualquiera se debe a una multitud de causas aleatorias e independientes, unas de otras, que por sí mismas tienen una influencia infinitésima, pero tal que, la suma de todas ellas tiene una magnitud apreciable. Si a cada causa se la considera una variable independiente y aleatoria, en condiciones muy generales, el *Teorema Central del Límite*, establece que: la sumatoria de todas ellas es otra variable aleatoria; tipificando, esa suma se tendrá una distribución asintóticamente gaussiana. Así, se justifica el empleo del modelo de Gauss para los errores casuales en teoría de mediciones. En los capítulos de muestreo y de estimación estadística, se encuentra la manera de plantear el intervalo de confianza para una serie de mediciones del tipo.

$$\mu \in (\bar{x} \pm \Delta x)$$

Donde se expresa que el verdadero valor μ , de una magnitud x , pertenece a un intervalo formado por su mejor estimación puntual \bar{x} más (o menos) el error de medición Δx . En el modelo de Gauss, Δx es el producto del valor crítico de confianza Z_α por el error típico σ / \sqrt{n} de estimación. En el modelo Student, los cambios son: (a) Usar el valor medido muestral DS en lugar

del valor desconocido poblacional σ y (b) Usar el valor crítico $t_{\alpha; n-1}$ en vez de Z_{α} . Cuando no se tenga más de una medición, entonces Δx será directamente el error de estimación (la menor unidad de la escala de lectura, o su mitad en el caso de escala continua).

Antes de seguir con el tema de las calibraciones, se hace necesario introducir un nuevo concepto a la teoría de errores gaussianas, presentada en el Tema 2. Se trata de cómo se propagan los errores cuando hay más de una variable en juego, en el cómputo del error total.

Sea una magnitud cualquiera $Y = F(x_i)$; donde F es una función de n magnitudes clínicas o variables x_1, x_2, \dots, x_n . Si se mide cada una de las n magnitudes se obtienen:

$$\mu_x \in (\bar{x}_i \pm \Delta x_i)$$

donde:

μ : es el valor verdadero de la magnitud clínica medida x_i

\bar{x}_i : es el promedio de las mediciones efectuadas.

Δx_i : es el error de medición para un cierto nivel de significación α .

Una vez efectuadas todas las mediciones se trata de calcular el valor final:

$$\mu_Y \in (\bar{y} \pm \Delta y)$$

donde:

μ_Y : es el valor verdadero de la magnitud clínica buscada Y .

\bar{y} : es el promedio calculado con la fórmula $Y = F(x_i)$, o sea: $\bar{y} = F(\bar{x}_i)$.

Δy : es el error propagado por las x_i , para un cierto nivel de significación.

Con

$$(\Delta y)^2 = \sum_1^n \left[\left(\frac{\partial y}{\partial x_i} \right) \Delta x_i \right]^2$$

Es decir, el valor del error total al cuadrado, es igual la sumatoria de los cuadrados del producto entre la derivada parcial de la función respecto a cada variable, y el error de la misma.

Ejemplo 1) En la fabricación de un patrón acuoso de glucosa, se pesó varias veces una cantidad de glucosa pura, estimada con una confianza del 95% en $(98,4 \pm 0,4)$ g y una cantidad de agua tridestilada estimada al 95% en $(92,5 \pm 0,3)$ dl. Calcular la concentración de la sustancia cuando ambas se mezclan. La relación entre ambas magnitudes es:

$$Y = \text{Masa de glucosa} / \text{Volumen de agua} = x_1 / x_2 = M / V$$

Cuyo valor promedio se calcula con:

$$\bar{y} = 98,4 \text{ g} / 92,5 \text{ dl} = 1,06378 \text{ g} / \text{dl}$$

El error total que afecta a la concentración es:

$$(\Delta y)^2 = \sum_i^n \left[\left(\frac{\partial Y}{\partial x_i} \right) \Delta x_i \right]^2 = \left(\frac{\Delta M}{V} \right)^2 + \left(\frac{\Delta V}{V^2} \right)^2 = (4,32 \cdot 10^{-3})^2 + (3,45 \cdot 10^{-3})^2 = 30,565 \cdot 10^{-6}$$

Luego es: $\Delta y = 5,53 \cdot 10^{-3} \approx 6 \cdot 10^{-3}$

Finalmente, el resultado se expresa como:

$$\mu_Y \in (\bar{y} \pm \Delta y) = (1,064 \pm 0,006) \text{ g/dl} \rightarrow 95 \% \text{ CI } (1,052 ; 1,076)$$

Esto es, hay un 95% de confianza en tener una solución acuosa de glucosa cuya concentración verdadera de está entre 1,052 y 1,076 g / dl, la cual se puede usar como patrón acuoso.

23.3 Calibraciones de instrumentos

Hay muchos instrumentos diferentes en un laboratorio; aquí solo se mostrarán algunos de ellos en términos generales, con el objeto de ejemplificar el uso de las técnicas estadísticas empleadas en la calibración. Para comenzar conviene hacer una reflexión: Un buen instrumento para las calibraciones es aquel cuyo error de apreciación es muy pequeño, respecto a la de los otros instrumentos que se emplean en la técnica clínica. Por ejemplo, la recomposición de sueros liofilizado y otro tipo de diluciones se hacen usualmente con pipetas, una pipeta calibrada común de 1 a 10 ml, llega a tener un error de apreciación mínimo de una décima de ml, mientras que las comunes llegan a 0,5 ml. Ahora bien, si se usa una balanza eléctrica de precisión que tiene un error de 0,0001 g en su escala de lectura, entonces resulta más conveniente usar la balanza para medir la cantidad de agua necesaria que la pipeta. Parece natural comenzar por ahí.

23.3.1 Calibración de balanzas

Para poder calibrar una balanza se necesita un juego de pesas patrones con certificación del INTI (Instituto Nacional de Tecnología Industrial). La idea general es cubrir todo el rango de pesadas usuales en el laboratorio para testear una serie de puntos de su escala, que llega hasta los 200 g. Así, se traza una recta de calibración, comparando los valores promedios obtenidos, contra los valores patrones indicados en cada pesada como se vio en el Tema 22. Basta por ahora, tomar un solo punto, el del valor más usual en la recomposición de liofilizados. Sea por ejemplo, usar la pesa patrón de 2g para ello. El procedimiento es como sigue:

Paso 1) Nivelar la balanza, girando los tornillos de regulación en los pies de apoyo, hasta centrar la burbuja de aire en el indicador de nivel de vidrio. Conviene aislarla de vibraciones, apoyando sobre telgopor, cartón corrugado etc.

Paso 2) Colgar el platillo en su gancho, cerrar las tapas de vidrio laterales y trabar la balanza

Paso 3) Luego de encender la balanza se controla el ajuste de cero

Paso 4) Se coloca la pesa de 2 g con cuidado sobre el platillo y con las tapas cerradas se hace la primer pesada gruesa con la balanza semi destrabada. Luego se la destraba totalmente y se hace la pesada fina con el avance micrométrico hasta tener el valor final de la primera pesada

Paso 5) Se traba la balanza volviendo a cero los indicadores y se apaga. Se enciende nuevamente repitiendo los pasos anteriores hasta obtener el segundo valor. Y así sucesivamente por lo menos cinco veces. Ahora se tienen 5 pesadas repetidas de la pesa patrón.

Paso 6) Con los $n = 5$ valores obtenidos se calculan el promedio y el desvío estándar de las pesadas efectuadas con la balanza.

Paso 7) Se realiza el test estadístico para validar *exactitud*, usando el modelo Student

$H_0 : \mu = 2 \text{ g}$ Y la balanza tiene buena exactitud, está calibrada.

$H_1 : \mu \neq 2 \text{ g}$ La balanza tiene un error de tipo sistemático.

Se calcula :
$$t = \frac{\bar{x} - \mu}{DS / \sqrt{n}} \quad \text{versus} \quad t_{\alpha, v}$$

Cuando el estadígrafo hallado t , sea menor que el crítico de tablas $t_{\alpha, v}$, se acepta la hipótesis nula; en cambio, si la situación es al revés se tiene evidencia significativa de descalibración. Lo mismo puede hacerse con otras pesas y seguir controlando puntualmente. En este caso de mediciones repetidas de la misma magnitud, se verifican los supuestos básicos de este modelo: por lo tanto, no se necesita recurrir al modelo no paramétrico equivalente, como el de la U de Mann-Whitney.

Una vez controlada la exactitud se procede con la precisión. Se fija un error relativo aceptable para este instrumento, por ejemplo no debe superar el 1%. De acuerdo a eso:

Paso 8) Se calcula el desvío estándar máximo admisible $DS_{adm.} = 0,01 \cdot 2\text{g} = 0,02\text{g}$

Paso 9) Se testea la hipótesis con el modelo de la Chi-cuadrado:

$H_0 : \sigma_{adm} \leq 0,02 \text{ g}$ Y la balanza tiene una precisión aceptable.

$H_1 : \sigma_{adm} > 0,02 \text{ g}$ Hay mucha dispersión.

Se calcula :
$$\chi^2 = (n-1) DS^2 / \sigma^2 \quad \text{versus} \quad \chi_{\alpha; v-1}$$

Si no se rechazan ambas hipótesis nulas, entonces se puede pensar que la balanza tiene una exactitud y precisión aceptables para el control de la calidad propuesto. O sea: *La balanza esta bien calibrada.*

23.3.2 Calibración de pipetas

Para poder calibrar una pipeta, se usa una balanza eléctrica bien calibrada y agua tridestilada. La idea es cargar con agua una pipeta cualquiera y luego pesar el agua recogida en una balanza, para verificar si el volumen de agua coincide con su peso. De esta manera, se puede realizar varios tipos de controles como:

- Control de precisión y exactitud de una pipeta.
- Verificar lo mismo en un lote de pipetas.
- Controlar el factor humano, usando una misma pipeta, pero con diferentes personas.
- Comparar dos tipos de pipetas entre sí para detectar si hay diferencias entre ellas.

Los modelos estadísticos a usar son: el modelo de Student y el Chi-cuadrado para una sola muestra, y los modelos de Student para dos muestras independientes y el de Fisher para comparar dos muestras entre sí.

El método consiste en realizar una serie de n mediciones con la misma pipeta y con el mismo volumen de agua. La variabilidad en las mediciones puede deberse al operador o a la balanza. El operador puede cometer errores de paralaje al medir; por ejemplo, si se ubica mal frente a la misma, o no la pone vertical al sostenerla con la mano en lugar de usar un apoyo fijo, o no verifica que la línea indicadora en la pipeta pase por el centro del menisco, etc. A su vez, la balanza tiene fluctuaciones como: vibraciones indeseadas (se soluciona con una base sólida, horizontal y aislada con material tipo tergopol que absorbe las vibraciones); vibraciones por la fuente eléctrica de alimentación (se soluciona con un estabilizador tipo UPS que mejora este tipo de problemas); mal cierre de las ventanas laterales; etc. Todas estas posibles causas, si bien pueden ser evitadas con un protocolo adecuado, no dejan de incidir en las mediciones afectando la precisión y exactitud. Por otro lado, pueden existir defectos constructivos de la pipeta tales como: caras no paralelas, dilatación por calor, etc., los que van a atentar contra la exactitud. El método de calibración es:

Paso 1) Colocar un frasco de vidrio adecuado en tamaño, limpio y seco, dentro de la balanza y luego pesarlo para tener el valor de la tara.

Paso 2) Se mide un volumen de agua prefijado con la pipeta y luego se vuelca el agua dentro del frasco en la balanza. Se pesa todo el conjunto y por diferencia de pesadas, se calcula el primer valor buscado.

Paso 3) Se repite este procedimiento n veces y con los n valores se calcula el promedio de las pesadas y su desvío estándar.

Paso 4) Se verifica exactitud con el modelo Student y precisión con la Chi-cuadrado. Así se decide si la pipeta esta bien calibrada.

Ejemplo 1) Se desea controlar una pipeta de 5 ml. Para ello, se carga 5 veces al máximo la pipeta con agua tridestilada y se procede a pesar. Decidir si la pipeta esta bien calibrada si los datos obtenidos fueron:

Cuadro 23.1: Control de una pipeta de 5 ml

Nº	Pesada	Diferencia de pesadas	Valor final	Promedio	Desvío
----	--------	-----------------------	-------------	----------	--------

T	72,3465				
1	77,3876	77,3876 - 72,3465	5,0411		
2	82,4251	82,4251 - 77,3876	5,0375		
3	87,5213	87,5213 - 82,4251	5,0962	5,05932	0,04083
4	92,5348	92,5348 - 87,5213	5,0135		
5	97,6431	97,6431 - 92,5348	5,1083		
Total			25,2966		

Control de exactitud: Modelo de Student para una sola muestra.

Ho : $\mu = 5$ ml Está calibrada.

H1 : $\mu \neq 5$ ml Hay un error sistemático

$$t = \frac{\bar{x} - \mu}{DS/\sqrt{n}} = \frac{5,05932 - 5}{0,04083/\sqrt{5}} = 3,248^* > t_{0,95;4} = 2,776$$

Se rechaza la Ho y se puede estimar el error sistemático de esta pipeta en ES = 0,05932 ml. Por lo tanto, se puede corregir cada medición que se haga con esta pipeta restando esa cantidad al valor medido. O bien, desecharla directamente por otra mejor calibrada.

Control de precisión: Modelo de Chi cuadrado.

Ho : $\sigma \leq 0,1$ ml Tiene una precisión aceptable.

H1 : $\sigma > 0,1$ ml Hay un error casual no aceptable.

$$\chi^2 = (n-1) DS^2 / \sigma^2 = (5 - 1) (0,04083)^2 / (0,1)^2 = 0,6668 < \chi_{0,95;4} = 11,143$$

El valor máximo admisible de dispersión se tomó como el valor mínimo de la graduación de la pipeta (0,1 ml) y los resultados no son significativos, esto es, la pipeta aprueba el control de precisión. En este caso, más que la pipeta, lo que está verificando es la interacción entre el observador y la pipeta. Es más probable que de haber un error casual muy grande, la mayor parte de la variabilidad se deba al observador.

En cambio, si se toma un $\sigma \leq 0,01$ ml; el estadígrafo resultante sería $\chi^2 = 66,68^{***}$, un resultado altamente significativo de falta de precisión. O sea, a medida que se exigen valores máximos admisibles más estrictos, es más fácil perder precisión. Parece razonable usar la menor unidad de la escala de lectura del instrumento como límite máximo.

En el ejemplo anterior, como el instrumento es de escala continua, se puede apreciar hasta la mitad de la menor unidad de la escala (error de apreciación). Si se usa tal valor, sería:

Ho : $\sigma \leq 0,05$ ml y entonces $\chi^2 = 2,67$ ml valor no significativo. La pipeta se considera con precisión aceptable, aún cuando su valor máximo admisible se redujo a la mitad.

Esta metodología se puede extender a otros casos de verificación como los siguientes:

Controlar un lote de pipetas del laboratorio:

Para esto se eligen n pipetas al azar (o a todas) cuya capacidad sea por lo menos de 1 ml. Se elige como valor de testigo a un valor como el de la pipeta más pequeña, por caso: 2 ml. Se carga a cada una de las pipetas con 2 ml de agua y se procede a pesar los volúmenes obtenidos, siguiendo el esquema del ejemplo anterior, por diferencias de pesadas. Entonces se pueden verificar varias cosas de acuerdo al diseño del experimento:

(a) Con el modelo Student para una muestra, se puede decidir si la exactitud es aceptable, no importa con cual de las pipetas habituales se esté trabajando.

(b) Con el modelo de la Chi cuadrado, se puede decidir si la dispersión es aceptable con:

- Un solo operador de las pipetas.
- Todos los operadores habituales, cada uno con su pipeta de trabajo usual.

Estudiar el efecto del factor humano en la medición con pipetas:

Esto es, descubrir si con una misma pipeta, pero empleada por diferentes personas, la precisión y exactitud no se ven afectadas. Para ello, se elige alguna de las pipetas calibradas con el procedimiento descrito en el primer ejemplo, que haya pasado los controles de exactitud y precisión. Luego, cada miembro del laboratorio afectado al tema debe medir los 5 ml de agua. Se tiene así un grupo de pesadas a las que se les aplica el mismo método para determinar si aun se mantiene con una calidad aceptable. Caso contrario, significa que el factor humano afecta las mediciones clínicas del laboratorio. Para descubrir quien o quienes son los responsables y así poder enseñarles una mejor manera de medir, se prueba a los sospechosos haciéndoles repetir el método del primer ejemplo, hasta descubrir quien o quienes no pasan la prueba. Aquel que no pase el test estadístico deberá ser mejor entrenado hasta lograr que trabaje con una calidad aceptable.

Comparar dos tipos de pipetas:

La idea es detectar si hay diferencias entre ellas. Pueden ser dos marcas comerciales, dos capacidades diferentes, etc. También se pueden comparar dos personas o dos lotes diferentes. El modelo adecuado para controlar exactitud en estos casos es el de Student para comparar dos muestras independientes. Para verificar la precisión se debe usar el modelo F de Fisher para compararlas entre sí. Sea por ejemplo, usar la pipeta del primer ejemplo como la A, y con otro tipo de pipeta repetir el procedimiento para la pipeta B. Por ejemplo,

Cuadro 23.2: Control de otra pipeta de 5 ml

Nº	Pesada	Diferencia de pesadas	Valor final	Promedio	Desvío
T	72,3465				
1	77,3601	77,3601 - 72,3465	5,0136		
2	82,3986	82,3986 - 77,3601	5,0385		
3	87,3913	87,3913 - 82,3986	4,9927	5,01106	0,01848
4	92,4066	92,4066 - 87,3913	5,0153		
5	97,4018	97,4018 - 92,4066	4,9952		
Total			25,0553		

Control de exactitud de la pipeta B: Modelo de Student para una sola muestra.

Ho : $\mu = 5$ ml Está calibrada.

H1 : $\mu \neq 5$ ml Hay un error sistemático

$$t = \frac{\bar{x} - \mu}{DS/\sqrt{n}} = \frac{5,01106 - 5}{0,01848/\sqrt{5}} = 1,338 < t_{0,95;4} = 2,776$$

No se rechaza la Ho y se puede considerar que tiene buena exactitud.

Control de precisión de la pipeta B: Modelo de Chi cuadrado.

Ho : $\sigma \leq 0,1$ ml Tiene una precisión aceptable.

H1 : $\sigma > 0,1$ ml Hay un error casual no aceptable.

$$\chi^2 = (n-1) DS^2 / \sigma^2 = (5 - 1) (0,01848)^2 / (0,1)^2 = 0,0876 < \chi_{0,95;4} = 11,143$$

No se rechaza la Ho y se puede considerar que tiene buena precisión.

Como conclusión la pipeta B ha pasado los controles de exactitud y precisión, por lo que puede considerarse con calidad aceptable. Suponiendo que el mismo operador ha realizado ambos grupos de mediciones, ahora se puede pensar que el error sistemático de la pipeta A se origina en la misma y no en el operador. Pero también se pueden comparar ambas pipetas entre sí:

Comparación de pipetas A y B en exactitud: Modelo Student para dos muestras independientes.

Ho : $\mu_A = \mu_B$ No hay diferencia entre las pipetas.

H1 : $\mu_A \neq \mu_B$ Hay diferencia entre ambas.

$$t = \frac{(\bar{x}_A - \bar{x}_B) - 0}{\sqrt{(1/n)(DS^2_A + DS^2_B)}} = \frac{5,05932 - 5,01106}{\sqrt{(1/5)\{(0,04083)^2 + (0,01848)^2\}}} = 2,413^* < t_{0,95;8} = 2,306$$

Se encontró una diferencia significativa entre ambas pipetas.

Comparación de pipetas A y B en precisión: Modelo de Fisher.

Ho : $\sigma_A = \sigma_B$ No hay diferencia entre las pipetas.

H1 : $\sigma_A \neq \sigma_B$ Hay diferencia entre ambas.

$$F = DS^2_A / DS^2_B = (0,04083)^2 / (0,01848)^2 = 4,882$$

Los valores críticos son: $F_{0,975;4;4} = 9,6$ y $F_{0,025;4;4} = 1 / F_{0,975;4;4} = 1/9,6 = 0,104$

Como el valor $F = 4,882$ está dentro de la zona de aceptación, no se puede rechazar la hipótesis nula y se considera que ambas pipetas son equivalentes en precisión. La conclusión final es que conviene descartar la primer pipeta y quedarse con la segunda.

Es conveniente resumir el uso de los modelos estadísticos vistos en este tema:

Calibración usando un patrón

Control de exactitud: Modelo Student para una sola muestra

Control de precisión: Modelo Chi-Cuadrado para una sola muestra

Comparación de dos muestras

Control de exactitud: Modelo Student para dos muestras independientes

Control de precisión: Modelo F de Fisher para dos muestras independientes

23.3.3 Calibración de capilares

Para controlar capilares lo más sencillo es dedicarse a la precisión de los mismos pues se pueden usar métodos directos y baratos. Sea el caso de los capilares usados para el método del *microhematocrito* (mH). Se pueden cargar con la misma sangre entera 10 tubos elegidos al azar del envase de 500 tubos. Se colocan estas muestras en una micro centrífuga, y se procesan todos a la vez. Una vez efectuadas las determinaciones con el mismo ábaco y el mismo operador, se tienen 10 valores que deberían ser iguales en el caso ideal. Si se toman los recaudos apropiados al procesar las muestras, la variabilidad mayor se deberá principalmente a variaciones en el diámetro de los capilares usados. Naturalmente, siempre existen fluctuaciones por azar que se suman al total de los errores detectados. El problema es determinar si el total de la variabilidad encontrada se halla por debajo del límite de tolerancia aceptable. De la bibliografía clínica, surge un coeficiente máximo de variación del orden del 2%. Pero cada laboratorio puede auto imponerse un límite razonable, no supere al $CV = 7\%$ del macrohematocrito o el $CV = 5\%$ para el mH.

Con los n valores obtenidos se calculan la media y el desvío estándar muestral. La idea es controlar que este valor no supere el límite máximo admisible, dado por la relación $\sigma_{\text{máx}} = 2\%$. Con estos datos se pueden formular un test de hipótesis con el modelo de la Chi-cuadrado. Para verificar el supuesto $DS \leq \sigma_{\text{máx}} = 2\%$.

De esta manera se controla si la dispersión del diámetro de los capilares es significativo para las mediciones del mH. A su vez, si se repite el procedimiento con otra marca comercial de capilares, se pueden plantear un test de validación entre ambas marcas, usando el modelo F de Fisher para contrastar varianzas. Lo mismo para comparar dos micro centrífugas, dos ábacos de lectura y aún, a dos operadores entre sí.

El control de exactitud es más costoso por la dificultad de hacerse de una "sangre patrón", en el mercado. Esa dificultad se subsana a través de algún método indirecto, obteniendo un control de tipo secundario. Por ejemplo, medir la sangre usada en un contador hematológico bien calibrado, que se acepte como referencia. Así, cualquier sangre puede ser utilizada para el experimento, si se toma el costo de enviarla a calibrar en un laboratorio adecuado. Esto es más barato que comprar una sangre calibrada en el mercado, como otro tipo de solución.

Pero si se tiene una sangre calibrada, entonces se puede hacer lo mismo que en el control de las pipetas. Esto es, se postula que el valor medio encontrado del mH no difiere del valor referencial. Y se valida con el modelo de Student para una sola muestra. Cuando la hipótesis nula no puede ser rechazada, se pueden considerar aceptables los restantes capilares del envase. Por el contrario, si se rechaza la misma, se tiene evidencia científica como para reclamar la devolución de esta mercadería fallada al proveedor.

23.3.4 Calibración de micropipetas

El método de calibración de las micropipetas es enteramente similar al visto para las pipetas en el punto anterior. La diferencia fundamental es que, en lugar de usar agua tridestilada, se necesita una sustancia mucho más pesada como el mercurio bidestilado, para lograr un peso apreciable con poca cantidad. La ventaja es que pequeños volúmenes de mercurio tienen un peso considerable y no se pierde la precisión de la balanza. Se aprovecha mejor el bajo coeficiente de variabilidad de ésta. La desventaja es la alta dilatación del mercurio con la temperatura, por lo que se necesita corregir este factor. La otra, es la necesidad de un lavado exhaustivo con ácido para eliminar todo el resto de mercurio de la micropipetas y, además, su mayor costo en comparación con el costo del agua.

Existe una gran diversidad de micropipetas en el laboratorio clínico. Resulta importante el tener un buen control sobre este instrumental si se tiene en cuenta que es la base de muchas determinaciones habituales. Por ejemplo, las pipetas de Thoma para medir eritrocitos y leucocitos se emplean para diluir la sangre entera de los pacientes y así determinar el conteo de la cantidad de glóbulos por mm^3 . En la Figura 23.1 se han esquematizado dos de estas pipetas: la pipeta para recuento de los rojos consta de un tubo capilar graduado y dividido en diez partes. Marcado con un 0,5 en la quinta señal y con 1 en la décima. Luego viene una ampolla o pera para efectuar la dilución con una bolita de cristal rojo en su interior. En la parte superior tiene una marca con la indicación 101. En la otra pipeta empleada para diluir la sangre en el recuento de blancos, la bolita dentro de la ampolla es blanca, la marca superior dice 11, mientras en el resto son similares. El volumen de la pipeta de eritrocitos esta constituido por una parte del capilar (hasta 0,5) de sangre entera; luego se aspira el líquido diluyendo hasta la marca 101 (es decir el resto del capilar y la ampolla). De esta forma se logra efectuar diluciones 1: 200 con un volumen total de 1 cm^3 . En la pipeta para diluciones en recuento de leucocitos, se procede en forma análoga llegando hasta la marca de 0,5, y luego con diluyente hasta la marca 11, para obtener una dilución de 1: 20. Debe destacarse que el líquido diluyente que queda en el capilar no entra en el cálculo, por lo tanto, debe ser eliminado antes de realizar llenado de la cámara. Aquí es fácil ver que una mala calibración en la micropipeta influirá notablemente en recuento final; en especial el capilar.

Otras de la micropipetas empleadas en los laboratorios clínicos, es la pipeta Sahli. Esta micropipeta tiene una capacidad de 0,02 cm^3 y se carga en ella sangre entera tomada de una punción capilar, o bien, de sangre venosa. Este volumen de sangre debe ser colocado en una cubeta que contenga 5 cm^3 de una solución (llamada de Drabkin). Mezclando adecuadamente y dejando reposar no menos de diez minutos, se forma la *cianometahemoglobina*. Con fotómetro de filtro, o espectrofotómetro, se puede determinar la concentración de la hemoglobina en la sangre. Esta proteína conjugada es la componente principal de los glóbulos rojos, ya que sirve para el trans-

porte de oxígeno y del CO₂ en la sangre; tiene una importancia muy grande en todos los procesos corporales al llevar el O₂. Totalmente saturada, la hemoglobina transporta 1,34 cm³ de O₂ por gramo. Una persona adulta tiene en su organismo unos 600 g de hemoglobina, los que son capaces de transportar unos 800cm³ de oxígeno para el cuerpo humano. Esta micropipeta se esquematiza en la figura en la posición más a la derecha de la Figura 23.1.

Figura 23.1: Micropipetas

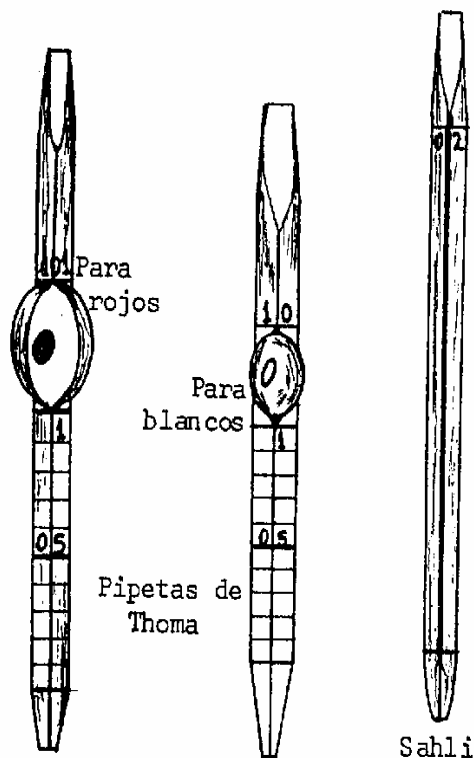
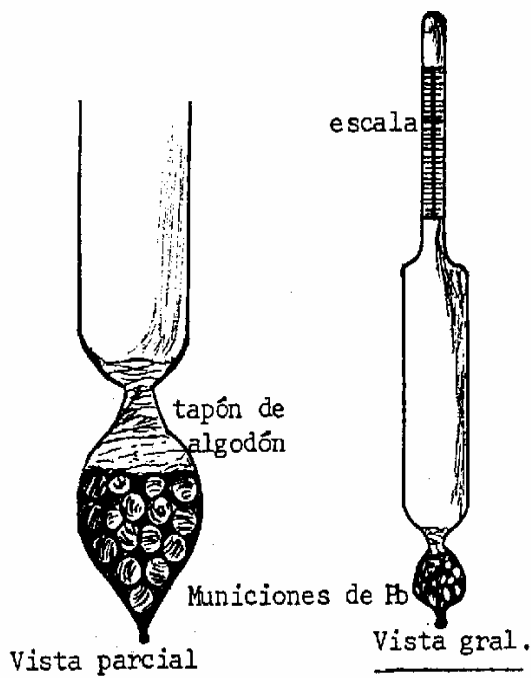


Figura 23.2: Urodensímetros



Vista la importancia de tener muy bien calibradas las micropipetas para el trabajo clínico, en particular la exactitud de los volúmenes muy pequeños que miden, se procederá a mostrar un método sencillo de calibración. Stevenson et al publicaron en 1951 un trabajo en *American Journal of Clinical Pathologist* en el cual se basa este experimento. La idea básica es llenar la micropipeta con un volumen de un líquido de densidad conocida, pesar dicho volumen y verificar si coincide con el indicado en la micropipeta. Como los volúmenes de trabajo son muy pequeños, no se puede usar agua destilada en muchos casos, debido que la masa a pesar es muy pequeña y se requiere mucha precisión en la balanza. Suponiendo que se llena la pipeta con mercurio bidesulfado, por ejemplo una pipeta Sahli cuyo volumen es de 0,02 cm³, entonces se puede pesar el mercurio contenido dentro de la pipeta si se lo coloca en un frasco vacío. Sea M la masa de un frasco vacío, M' la masa del frasco con el mercurio que contenía la micropipeta, entonces la masa del mercurio es la diferencia entre ambas pesadas: $m = M' - M$
 Si se mide la temperatura del mercurio durante la pesada se puede ir a una tabla como la siguiente, para determinar la densidad del mismo a la temperatura del trabajo:

Tabla 23.1: Variación de la densidad del mercurio con la temperatura.

Temperatura °C	Densidad Hg g/cm ³
20,0	13,547
20,5	13,546
21,0	13,545
21,5	13,544
22,0	13,543
22,5	13,542
23,0	13,541
23,5	13,540
24,0	13,539
24,5	13,538
25,0	13,537

Temperatura °C	Densidad Hg g/cm ³
25,5	13,5465
26,0	13,5340
26,5	13,5330
27,0	13,5320
27,5	13,5310
28,0	13,5300
28,5	13,5290
29,0	13,5280
29,5	13,5270
30,0	13,5260
30,5	13,5250

De esta forma el volumen de la micropipeta (V) debe obtenerse con:

$$V = m / d$$

Ejemplo 1) Se ha pesado mercurio bidestilado con una pipeta Sahli. Los valores obtenidos son:

M = 39,8731 g ; M' = 40,1311 g y la temperatura es de 29 °C. Luego,

$$m = 40,1331 - 39,8731 = 0,2580 \text{ g} \quad \text{A } 29 \text{ °C es } d = 13,528 \text{ g / cm}^3 \text{ de tablas}$$

$$V = 0,2580 / 13,528 = 0,01907 \text{ cm}^3$$

Entonces el factor de corrección para tal pipeta será:

$$k = 0,01907 / 0,02 = 0,96$$

Por lo tanto, cuando se mida 0,02 cm³ de mercurio con la pipeta, hará que usar el factor de corrección y queda así 0,01907 cm³. El problema ahora es poder determinar el error que afecta a ese valor, para ello se debe usar el concepto de propagación de errores:

a) Hay que considerar la medición de la temperatura, el termómetro usado tiene un error de apreciación de 0,5 °C. Entonces, $t \in (28,5 ; 29,5) \text{ °C}$ y $d \in (13,527 ; 13,529)$. O sea se puede expresar como $d \in (13,528 \pm 0,001) \text{ g / cm}^3$. O sea, $\Delta d = 0,001 \text{ g / cm}^3$

b) Hay que tomar en cuenta el error de la masa, como se ha calculado el valor por diferencia de pesada, el error $\Delta m = 0,0002$ será el doble que el error de apreciación.

c) Aplicando la fórmula de propagación de errores se puede calcular:

$$(\Delta V)^2 = \sum_1^n \left[\left(\frac{\partial V}{\partial x_i} \right) \Delta x_i \right]^2 = \left(\frac{\Delta m}{d} \right)^2 + \left(\frac{\Delta d}{d^2} \right)^2 = \left(\frac{0,0002}{13,528} \right)^2 + \left(\frac{0,001}{183} \right)^2 = 0,025 \cdot 10^{-8}$$

$\Delta V = 0,16 \cdot 10^{-4} \text{ cm}^3$. Por lo tanto, la estimación de volumen de la micropipeta es:

$$V \in (0,019070 \pm 0,000016) \text{ cm}^3$$

Notar, que el valor marcado en la micropipeta es de $0,02 \text{ cm}^3$ y cae fuera del intervalo anterior. Esto implica la presencia de un error sistemático evaluado $ES = 0,00093 \text{ cm}^3$ por defecto. Luego, para calibrar la micropipeta hay dos caminos: o se remarca con una punta de diamante, o bien, se corrigen los valores con fórmulas. Para finalizar se debe lavar la micropipeta con ácido nítrico concentrado y luego con agua, hasta eliminar todo rastro del mercurio.

23.3.5 Calibración de urodensímetros

El método de urodensímetro se basa en sumergir el aparato en el seno de la masa líquida, donde recibe un empuje de abajo hacia arriba igual al peso del volumen del líquido desalojado. Cuando se equilibran el peso del aparato y el empuje recibido, este flota dejando sumergido un cierto volumen calibrado que permite leer en una escala la densidad del líquido directamente. Como el volumen del aparato siempre es el mismo, los diferentes empuje que reciba dependen de la densidad de los líquidos analizados. Variando el peso del aparato se puede hacerlo flotar de manera de poder leer en la escala. Para ello, se colocan granallas de plomo en el bulbo de su extremo inferior (ver Figura 23.2) y se regula su peso. En la parte superior tiene una varilla con la escala graduada en densidad. Hay densímetros especiales con escalas graduadas en unidades especiales, por ejemplo, el alcoholímetro que mide directamente la graduación alcohólica de bebidas; otras miden en los llamados "grados Baumé" donde el ácido sulfúrico comercial tiene una densidad de 66° ($d = 1,85 \text{ g/cm}^3$).

En clínica, este aparato se usa para medir densidades de orina y de ahí su nombre. Su escala se gradúa entre 1 y $1,06 \text{ g/cm}^3$ para 15°C de temperatura. Los valores de referencia varían de acuerdo a la edad del paciente: en lactantes va de $1,001$ a $1,035 \text{ g/cm}^3$ y en adultos de $1,016$ a $1,035 \text{ g/cm}^3$. Las orinas de baja densidad se denominan *hipostenúricas* ($d < 1,007 \text{ g/cm}^3$). Los solutos disueltos en una orina normal influyen mucho en su densidad (fosfato 25%, cloruros 25% y urea 20%). Este método tiene varias correcciones: una es por temperatura, otra es por el contenido de glucosa y proteínas. Esto escapa a los alcances del presente trabajo, por lo que se dará una idea general de la calibración. Los pasos a seguir son:

Paso 1) Se prepara un patrón acuoso con ácido sulfúrico puro y agua tridestilada, hasta lograr una concentración adecuada. Por ejemplo, $2,029 \text{ g}$ de ácido en 100 cm^3 de agua, da una solución con $d = 1,015 \text{ g/cm}^3$.

Paso 2) Se llena el vaso del urodensímetro con el patrón, hasta $3/4$ de su capacidad y se coloca con cuidado el aparato, dándole una ligera rotación y dejando que flote libremente. Evitando la formación de burbujas. El aparato no debe tocar las paredes del recipiente.

Paso 3) Se mide la temperatura y se lee en la escala usando el fondo del menisco, efectuando si corresponde, las correcciones por temperatura.

Paso 4) Se repite varias veces el procedimiento y con los valores obtenidos se calcula el promedio y el desvío estándar de las mediciones.

Paso 5) Se realizan los tests estadísticos para controlar precisión y exactitud, como se vio en los puntos anteriores tratados.

23.4 Calibración de espectrofotómetro

El método para calibrar este instrumento debe ser la recta de calibración. Sin embargo, cuando se desee controlar un valor en particular se simplifica la recta a un solo caso. La idea básica del control es preparar patrones calibrados de absorbancia y con ello revisar periódicamente el valor leído por el equipo. Si se tienen definidos los límites de aceptación para cada una de estas lecturas, se puede controlar si el espectro se mantiene en los límites de calidad aceptables. Es importante seguir un protocolo de mantenimiento y limpieza preventivo.

Para preparar los patrones hay varias maneras. Lo más sencillo y caro es adquirir los patrones en el INTI o en algún comercio especializado. Lo más aconsejable es recurrir al fabricante del aparato, que puede asesorar gratuitamente al respecto. Además si éste cumple con las Normas Internacionales de Calidad, tal como la ISO 9001 o 9002, seguramente tiene toda esta temática resuelta. La otra manera más barata, pero más complicada, es que el laboratorio se fabrique los patrones por sí mismo y luego los mande calibrar a los centros referencia. En la República Argentina esta función la cumple el Instituto Nacional de Tecnología Industrial (INTI) a cuyo cargo están los patrones del Sistema Internacional de Unidades y sus derivados. Es el único ente legalmente habilitado para cumplir esa función.

Un espectrofotómetro es básicamente un equipo que emite un rayo de luz que atraviesa una cubeta de caras paralelas con la sustancia a analizar. Mide la absorbancia (o transmitancia) de la tal sustancia, y con un cierto factor informa la concentración del analito buscada. Se pueden fabricar valores de referencia en forma económica. Una manera es conseguir acrílicos de diferentes tonalidades, y tallar con él cubetas de caras paralelas que simulen un cierto líquido adentro. Más barato aún es llenar cubetas con una mezcla de agua y tintas de colores, logrando tonalidades específicas para el control. Luego se sellan estas cubetas con láminas de vidrio y pegamento epoxi transparente colocado por fuera. Una vez que se tenga el conjunto de cubetas coloreadas, muy estables en el tiempo, se las manda a calibrar a un laboratorio de referencia. Y a partir de allí, se cuenta con un juego de valores referenciales que permiten controlar al espectro en forma permanente. Los pasos a seguir en un control son:

Paso 1) Con agua tridestilada se llena la cubeta y se calibra el 0 de la escala.

Paso 2) Se elige una de las cubetas patrones con la absorbancia que se desea controlar. Se repite **n** veces esta medición y se calcula el promedio y desvío estándar de los valores medidos.

Paso 3) Con el método Student se controla exactitud y con Chi-cuadrado precisión, en forma totalmente análoga a las descriptas más arriba.

23.5 Calibraciones microbiológicas

Los instrumentos principales en Microbiología son dos: el baño maría y la incubadora o estufa de cultivo. Los límites de tolerancia internacionales para estos dos equipos son:

Baño maría : Entre 36 y 38° C o sea (37° C \pm 1° C)
Estufa de cultivo: Entre 34 y 36° C o sea (35° C + 1° C)

La manera de ver si cumplen estos requisitos es bastante sencilla se debe tomar nota de los valores de temperatura a intervalos regulares de tiempo y verificar si están calibrados. En principio hay dos maneras de hacerlo. La primera vez que se hace un control de este tipo, se pueden tomar los valores de la temperatura cada 15 o 30 minutos, hasta juntar entre 20 y 30 valores. Es decir, basta con un día para completar el experimento. Una vez que se verifica que el equipo está calibrado se puede proceder con la segunda forma de control. En este caso se toma la temperatura en forma diaria y se vuelca en una Carta de Control de calidad, como se explicará más adelante. El empleo de las cartas es para asegurarse que el equipo se mantenga estable en el tiempo. Una vez por mes se debe efectuar una limpieza a fondo a modo de mantenimiento preventivo de acuerdo a lo que aconsejan las normas internacionales.

Ejemplo 1) Para controlar por primera vez un baño maría se tomaron cada media hora los valores de temperatura indicados por un termómetro que fue previamente calibrado. Los resultados de 30 mediciones arrojaron un promedio de 37,5° C con un desvío de 0,42° C. Se desea verificar si el instrumento tiene una exactitud y precisión aceptables. (Datos: Dra. Desimoni M.C.)

Para realizar el control de exactitud se puede emplear el modelo de Gauss o el más exacto de Student teniendo en cuenta que los requisitos son: (37° C \pm 1° C). Entonces:

$$t = \frac{\bar{x} - \mu}{DS/\sqrt{n}} = \frac{37,5 - 37}{0,42/\sqrt{30}} = 6,52 \gggg t_{0,999; 29} = 3,659$$

Se ha encontrado evidencia altamente significativa que muestra un error sistemático de 0,5° C. Para controlar la precisión se usa el modelo de la Chi-cuadrado teniendo en cuenta que el desvío máximo admisible es de 1° C. Entonces:

$$\chi^2 = (n-1) DS^2 / \sigma^2 = (30 - 1) (0,42)^2 / (1)^2 = 5,12 < \chi_{0,95; 29} = 42,557$$

Resulta que no hay evidencia como para pensar que la precisión del equipo no es aceptable.

Ejemplo 2) Para controlar una estufa de cultivo de tomaron 27 determinaciones de la temperatura a intervalos de 15 minutos. Se obtuvo una media de 35,41° C y un desvío de 0,41° C. Se desea saber si el equipo tiene exactitud y precisión aceptables para las normas internacionales. (Datos de la Dra. Piccoli, L.I.)

Para realizar el control de exactitud se puede emplear el modelo Student teniendo en cuenta que los requisitos son: $(35^\circ \text{C} \pm 1^\circ \text{C})$. Entonces:

$$t = \frac{\bar{x} - \mu}{DS/\sqrt{n}} = \frac{35,41 - 35}{0,41/\sqrt{27}} = 5,2 \gg t_{0,999; 26} = 3,707$$

Se ha encontrado evidencia altamente significativa que muestra un error sistemático de $0,41^\circ \text{C}$

Para controlar la precisión se usa el modelo de la Chi-cuadrado teniendo en cuenta que el desvío máximo admisible es de 1°C . Entonces:

$$\chi^2 = (n-1) DS^2 / \sigma^2 = (27 - 1) (0,41)^2 / (1)^2 = 4,37 < \chi_{0,95; 29} = 38,885$$

Resulta que no hay evidencia como para pensar que la precisión del equipo no es aceptable.

El análisis de ambos aparatos en los ejemplos anteriores muestra un error de tipo sistemático en ambos. Mientras que el baño maría calienta más de lo debido, la estufa trabaja calentando menos de lo necesario. La conclusión es que se deben enviar los equipos al fabricante para que regule mejor la termocupla interna de los mismos. Notar que es importante efectuar estos controles con termómetros previamente calibrados para evitar falsos resultados. Lo hallado en ambos ejemplos se puede presentar con intervalos de confianza de la manera siguiente:

El intervalo de confianza para un 95% es: $(\bar{x} \pm \Delta x) = (\bar{x} \pm t_{0,95; n-1} DS / \sqrt{n})$

Luego hay que fijarse si el valor poblacional μ cae dentro o fuera de este intervalo

Aplicando la ecuación anterior a los ejemplos es:

Baño María: 95% CI $(37,34 ; 37,66)^\circ \text{C}$ El valor esperado 37°C cae fuera del mismo

Estufa: 95% CI $(35,25 ; 35,57)^\circ \text{C}$ El valor esperado 35°C cae fuera del mismo

23.6 Otras calibraciones

Hay instrumentos automáticos que incluyen su propio sistema de Control de Calidad. Sin embargo, es conveniente que el profesional pueda realizar controles sobre este tipo de equipamiento, usando el bagaje de los métodos vistos.

Contador hematológico: Estos cuentan las células hasta tres veces antes de informar rojos, blancos y plaquetas. Algunos como el Technicon H 6000 brinda gráficos de la campana de Gauss para recuento de blancos y plaquetas, además del hemograma completo y de otros valores. La manera de controlar un equipo así, es conseguir sangre de referencia o control y con mediciones repetidas de la misma muestra se puede validar precisión y exactitud. Pero, en caso de no tener disponible un patrón, con cualquier sangre se puede controlar por lo menos, la precisión, y luego estudiar la exactitud comparando las mediciones hechas con otras técnicas.

Auto analizador clínico múltipara métrico: Un equipo como el Metrolab 2100, puede efectuar 18 tests por muestra a 36 sueros simultáneos. Su capacidad es de 100 a 180 tests por hora y emplea una cantidad de 350 µl de suero por test. Nuevamente, consiguiendo suero control se lo puede colocar hasta en 36 viales, para tener otras tantas mediciones y controlar así su precisión y exactitud en forma análoga a los anteriores.

23.7 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|---|-------|-------|
| 1) Para controlar un instrumento conviene usar una recta de calibración. | V | F |
| 2) Explicar el concepto de propagación de errores en mediciones:..... | | |
| 3) El modelo Student se usa para controlar exactitud. | V | F |
| 4) Explicar los pasos a seguir para calibrar una balanza:..... | | |
| 5) Se puede controlar un aparato y su operador en conjunto o por separado. | V | F |
| 6) Explicar los pasos a seguir para controlar una pipeta:..... | | |
| 7) Ídem anterior, pero para una micropipeta:..... | | |
| 8) Para controlar precisión se usa el modelo de la Chi-cuadrado. | V | F |
| 9) Para controlar la dispersión de dos aparatos se usa el modelo de la Chi-cuadrado. | V | F |
| 10) Resumir el concepto general de calibraciones usando patrones:..... | | |
| 11) Explicar las diferencias principales en calibrar capilares y micropipetas:..... | | |
| 12) La variación de volumen por temperatura no influye demasiado. | V | F |
| 13) Los urodensímetros son totalmente diferentes a los alcoholímetros. | V | F |
| 14) Para calibrar espectrofotómetros conviene recurrir al consejo del fabricante. | V | F |
| 15) Explicar las causas por las cuales el profesional debe aprender calibraciones:..... | | |

24

Control de Calidad Metodológico

En este capítulo se desarrollan diferentes formas de controlar la *calidad* en las mediciones de los laboratorios de análisis industriales o bioquímicos. Específicamente, de los métodos o técnicas usadas en la práctica diaria de los mismos. El primer paso hacia la calidad es calibrar los instrumentos, el segundo es calibrar los métodos y el tercero asegurar que estas calibraciones sean estables en el tiempo, mediante las cartas de control que se verán en el próximo tema. El concepto de *calidad* se toma aquí en el sentido de precisión y exactitud, como condición necesaria para alcanzar el objetivo principal: *la capacidad del método clínico para diagnosticar*. El campo de aplicación se restringe a las magnitudes clínicas de tipo cuantitativo. Otra limitación es que se trata de controles de tipo puntual, es decir con respecto a un solo valor de la magnitud, en lugar de controles del tipo: *recta de calibración*, donde se toman varios puntos. En la última sección se trata el Control de Calidad en la parte Microbiológica de acuerdo a las recomendaciones de la Confederación Latinoamericana de Bioquímica Clínica.

24.1 Conceptos básicos

Otra vez, el problema básico del control es determinar que se tendrá primero: si un patrón para calibrar un sistema, o un sistema calibrado para fabricar un patrón. Por caso, si se dispone de un suero control bien controlado, se puede calibrar una técnica de laboratorio. En cambio, si se dispone de una técnica calibrada, se pueden fabricar patrones secundarios como un *pool* de sueros, para controlar en el tiempo la estabilidad del método en forma económica y simple.

Resuelto este problema, cuando se dispone de un suero control o similar se procede a calibrar la técnica en su conjunto, o en algunas de sus partes. En efecto, un sistema de medición clínico se puede descomponer en cuatro partes intervinientes:

-*Factor humano*: Todas las personas participantes en la medición.

-*Instrumentos y equipos*: Los que se emplean para realizar la medición como espectrofotómetro, centrífugas, estufas, pipetas, buretas, etc.

-*Drogas y reactivos*: Son los kits comerciales para la determinación, agua tridestilada para las diluciones, ácidos para lavar, etc.

-*Método*: Se trata de los varios pasos que componen toda técnica: *su protocolo*.

Cada uno de estos factores, puede producir desviaciones que alteren significativamente los resultados finales de la medición. Por lo tanto, es necesario diseñar controles para poder detectar tales desviaciones en cada una de las partes, lo mismo que para todo el conjunto.

Cuando se hayan calibrado todas las técnicas de laboratorio, y se mantengan controladas a lo largo del tiempo con las cartas estadísticas, se puede plantear un objetivo más amplio. Tal como comparar el laboratorio con los demás de la zona y asegurar que todos los laboratorios produzcan valores analíticos con una calidad adecuada. Este tipo de controles *inter-laboratorios* se inscriben dentro de un marco de excelencia exigida por la medicina moderna. El requerimiento básico para lograrlo es que, todo el personal involucrado en el proceso esté imbuído de las causas de las imprecisiones analíticas, los modos de detectarlas y corregirlas pero, sobre todo, que sea conciente de la necesidad y utilidad del Control de Calidad Estadístico.

24.2 Sueros controles y patrones

Para poder calibrar técnicas clínicas se necesitan sustancias de control como:

- *Patrones acuosos*: Se preparan con agua tridestilada, a la cual se le adicionan las sustancias puras o proteínas, de manera tal de tener una concentración adecuada a los fines de control.

- *Sueros calibrados*: Se adquieren en el mercado y fueron calibrados por laboratorios de referencia legalmente autorizados. A su vez, estos sueros pueden provenir de sueros *humanos* o de *animales* (es mejor el humano, pero es el más caro).

- *Sueros de referencias o secundarios*: Usando los sueros primarios, se calibra el sistema de medición, y luego, empleando un *pool* de sueros, obtenido con el sobrante del mismo laboratorio, se procede a calibrarlo. Este *pool* ya calibrado se puede emplear para seguir controlando al sistema en forma más económica.

A los fines de conservación, estos patrones se almacenan congelados o liofilizados que es como se venden en el mercado. Si bien el tipo liofilizado es el más estable, tiene el problema de los posibles errores agregados al efectuar la dilución. A su vez el *pool* congelado puede variar su composición, por la degradación de alguno de sus componentes en el tiempo o por la condensación de la humedad ambiente, si no se toman recaudos apropiados.

La técnica mas difundida en los laboratorios es el empleo del *pool* de sueros séricos o de orinas para realizar el control. Hay varias razones para ello: su costo es bajo, su conservación no es costosa ni demasiado sofisticada y principalmente, es la más representativa de todas. En efecto, un *pool* se arma con el adecuado suero sobrante de los pacientes del mismo laboratorio, por eso sus componentes son representativos de la población humana que lo circunda. Mucho más adecuada, que si se emplea un suero de otro país o de tipo animal.

Cuando en la década del cincuenta se comenzaron a efectuar estudios comparativos, entre distintos laboratorios de una misma ciudad (Washington, Tokio, Roma, París, etc.) se repartieron muestras de soluciones acuosas, sueros y orinas provenientes de un mismo *pool*. Como las muestras tenían un mismo origen, se esperaban valores similares obtenidos en los distintos laborato-

rios de la ciudad. Los resultados obtenidos fueron alarmantes. Se encontraron discrepancias por fuera de los límites propuestos como aceptables, en más del 80% de los casos. Como no se podían clausurar a semejante cantidad de laboratorios a la vez, se optó por la política de darles un tiempo para corregir sus técnicas de medición. Y eso fue el surgimiento de las técnicas de control de calidad y de las distintas asociaciones profesionales que se ocupan del tema. Muy pronto surgieron los planes de control entre laboratorios, como una forma de asegurar la calidad de sus resultados. De allí surgieron una serie de necesidades:

- Las muestras de control debían ser equivalentes entre sí.
- Los componentes debían ser estables durante prolongados períodos de almacenamiento.
- Debía disponerse de grandes cantidades para su empleo en diferentes laboratorios.
- Se necesitaba asegurar la estabilidad del producto durante los traslados.

Todo esto creó con el tiempo, industrias dedicadas a la fabricación de sustancias de control. En la actualidad, la mayoría de los productos utilizados provienen de un gran pool de sueros en forma liofilizada. Los fabricantes de muestras de control almacenan ingentes cantidades de plasma sanguíneo obtenido de diversas fuentes: bancos de sangre con productos vencidos, sobrantes de hospitales donantes, etc. Se trasladan congelados a -7° C hasta la planta de producción, donde se los almacena hasta lograr miles de litros de plasma. Luego se procede a desfibrarlos y suplementarlos con diversos agentes para obtener las concentraciones adecuadas, y a continuación se los mezcla, filtra y almacena en viales adecuados. Estos viales son liofilizados y tratados con Nitrógeno. Los cuidados, controles y verificaciones durante todo el proceso de fabricación deben ser en extremo exhaustivos, como para poder asegurar un $CV < 1\%$. La humedad residual en el proceso de liofilización se baja hasta un 2%, y se trabaja con el suero a bajas temperaturas para evitar deterioros. En general, se asegura un $CV = 1\%$ en el producto final. Los sueros liofilizados suelen ser muy estables, salvo casos aislados de pérdida de glucosa por contaminación bacteriana, o cambios producidos por actividad enzimática. En la práctica, conviene emplear los sueros liofilizados para controles de exactitud, luego fabricar patrones secundarios suficientes, como para un semestre por los menos. Entonces, con el material no evaluado sobrante se pueden realizar los controles de precisión, a fin de bajar los costos totales del control de calidad. Los requisitos a cumplir por estos materiales, de acuerdo a las normas internacionales son:

- Estabilidad de los componentes de por lo menos un año.
- Almacenamiento a -20° C o liofilización.
- Resultados negativos para anticuerpos con antígenos HIV I y HIV II, HCV y HBV.
- Concentraciones cercanas a las habituales del laboratorio.
- Valores que faciliten las decisiones médicas.
- Variaciones entre viales que no superen un CV del 1%.
- Baja turbidez.
- Provenientes de sangre humana (o en su defecto de animales).

24.3 Protocolo de las técnicas

La principal cualidad de una técnica clínica es su estabilidad y repetibilidad para hacer las determinaciones. Cualquier cambio, por pequeño que parezca a primera vista, puede alterar el resultado final de la medición, lo cual invalida las comparaciones necesarias para detectar los problemas cuando la técnica se sale de control. Por eso, es necesario el establecimiento en cada técnica de un *protocolo* de medición, para minimizar al máximo las desviaciones. La idea básica es saber las causas que producen efectos indeseables, para evitarlas con el protocolo del método. Por ejemplo, si se va a realizar un estudio microbiológico de orina, se sabe que el recipiente que la contiene no debe estar contaminado, entonces el protocolo debe establecer la obligación de emplear recipientes previamente esterilizados. Lo ideal sería disponer de una especie de manual, que contenga el protocolo a seguir, en términos generales, para cada una de las prácticas de laboratorio que figuran en el Vademécum, o por lo menos, de las más comunes. Con una herramienta de esa naturaleza se podría unificar en el país la metodología de los laboratorios y uniformar mejor los resultados informados. Como la cantidad de información al respecto es demasiado grande, y día a día se incrementa con reportes de investigadores en las revistas especializadas, es casi imposible que un solo laboratorio confeccione un conjunto de protocolos actualizados con las últimas novedades. Lo apropiado sería que las universidades en conjunto con las asociaciones profesionales, mantengan un sistema al efecto, recurriendo a los especialistas de cada área. Mientras tanto, parece conveniente recurrir a los lineamientos generales publicados por ECCLS, IFCC y WHO en Copenhague (1992), titulados “Pacientes y muestra” y “Prácticas correctas en las mediciones clínicas descentralizadas”, puntualizando algunos aspectos a tener en cuenta.

En la *Fase Preanalítica* los puntos a considerar son:

1) *Preparación del paciente*: Los factores referidos al paciente pueden clasificarse en dos grupos principales. Aquellos que no se pueden cambiar y los otros que pueden controlarse por medio del mismo paciente, el laboratorio o el médico. Ejemplos del primer caso son, sexo, edad, embarazo, fase del ciclo menstrual, origen étnico, etc. Mientras que en el segundo caso una intervención activa de los participantes puede manejar factores como: tensión mental (stress), ejercicio, trabajo muscular intenso, dieta, medicación, ingesta de etanol, fumar, postura, tiempo de reposo y/o tiempo de ayuno previo, procedimientos como masaje o palpación, etc.

2) *Tiempo de muestreo*: Los cambios en los sistemas biológicos ocurren usualmente siguiendo *ritmos biológicos* como el ritmo menstrual de 28 días, o el circadiano de 24 hs. Por lo tanto, es necesario la comprensión del efecto de estos ritmos para programar la toma de muestras e interpretar sus resultados. Deben tomarse las muestras en el mismo momento del ciclo si se van a comparar resultados *intra* o *inter* individualmente. Por su parte, las pruebas dinámicas o pruebas de supresión o estimulación, se usan para estudiar la respuesta de los órganos a estímulos negativos o positivos, lo que implica una previa programación en cuanto al tiempo. Por ejemplo, la creatinquinasa es una isoenzima que requiere un tiempo de 6 a 30 hs después del infarto del miocardio para obtener resultados diagnósticos confiables.

3) *Cantidad de prestaciones*: La cantidad o volumen de muestra a extraer necesita el conocimiento del número de prácticas a realizar y sus posibles derivaciones o repeticiones, para que sea adecuada. A su vez, esto requiere conversaciones entre médicos y especialistas para determinar la

combinación, conveniencia, economía, etc., de la batería de prácticas a realizar. Por supuesto, se debe tener conocimiento de cada protocolo para tener información acerca de:

- Cantidad mensurable necesaria en la práctica.
- Interpretación de resultados.
- Indicaciones a darle al paciente antes de la recolección.
- Formas de tomar la muestra, manera de almacenarla y de manejarla.
- Listado de errores y/o equivocaciones que se deben evitar.

4) Identificación correcta: El paciente debe ser correctamente identificado, con todos sus datos personales y sus muestras con una etiqueta que especifique: su nombre, edad y sexo, la fecha de la toma y el tipo de muestra. Esta etiqueta debe ser colocada en el frasco y no en la tapa, también debe soportar la humedad del freezer o la luz solar si afecta la muestra.

5) Toma de muestra: Es necesario que la toma de muestra se haga correctamente, bajo condiciones que eviten la aparición de errores, tanto de medición como de interpretación. El tipo de muestras debe ser escogido con cuidado de acuerdo a la investigación a realizar. Las muestras de sangre pueden ser de tipo: arterial, venosa o capilar; las de orina pueden ser aleatorias, del chorro medio o con horarios; las microbiológicas deben tomarse del lugar adecuado; por ejemplo, en una investigación de gonorrea el hisopo se ubica en la región cervical y no en la vaginal. Un detallado protocolo e instrucciones escritas con explicaciones verbales al paciente, son la mejor forma de prepararse para la toma de muestras. En la bibliografía clínica se encuentra muy detallado el caso de extracción de sangre por venipunción como ejemplo de lo anterior.

6) Interferencia: Algunas sustancias solubles se pueden escurrir en tubos, viales o tapones y pueden afectar notablemente el resultado final. Por ejemplo, el 2 butoxietilo (tris) provoca el desplazamiento de drogas y otros analitos, causando redistribución de eritrocitos y plasma. Los tubos revestidos también son afectados y pueden interferir en la unión antígeno-anticuerpo o inhibir actividad enzimática. Lo recomendado es *evitar el contacto entre tapón y sangre*. Los separadores de suero pueden interferir en la veracidad de las mediciones, o la luz solar afectando las mediciones de bilirrubina, etc.

7) Tipo de muestras: Existen diferentes tipos de muestras y cada uno necesita de recaudos especiales que no se detallaran aquí, baste mencionar algunos de ellos:

- Muestras de sangre.
- Muestras de orina.
- Muestras de saliva.
- Muestra de líquido amniótico.
- Muestras de líquidos cefalorraquídeo.
- Muestras de heces fecales.

Y otros tipos menos comunes como las muestras genitales, vaginales, cervicales, uretrales, faríngeas, esputos, pus, o exudaciones, fluido serosos, superficies cutáneas o mucosas, y otras especiales como las biopsias.

8) Manejo de muestras: Todas las precauciones que se adopten para el manejo de muestras nunca alcanzan para asegurar un 0% de riesgo. No se debe olvidar que toda muestra debe ser manejada como *potencialmente infecciosa* y se debe cuidar la seguridad tanto del operador como la del resto del personal. Evitar contaminación de recipientes o del medio ambiente, aún cuando no sean obvios los riesgos de una infección. Otras precauciones importantes son: el evitar el contacto en-

tre tapón y muestra, lavado y secado cuidadoso, evitar evaporaciones, exposición a luz solar o a temperaturas excesivas, añadir los conservantes o preservativos y aditivos de forma que no altere la muestra mas allá de lo admisible.

9) *Almacenamiento y transporte de muestras*: Además de la esterilización previa, conviene un cerramiento hermético y sellado en caso de transporte fuera del laboratorio. Es importante tomar en cuenta el tiempo entre la extracción y el procesamiento, tratando de que sea lo mas corto posible. En su defecto, habrá que almacenar la muestra en refrigeradores o a temperatura ambiente cuando el frío provoque interferencia en la medición.

Todo lo anterior corresponde a la fase *pre-analítica* del método, o sea, desde la toma de la muestra, hasta el proceso de medición. La fase *analítica* del método se ve a continuación.

24.4 Protocolo de la fase analítica del método

En el apartado anterior se mostró la necesidad de contar con un protocolo detallado para la fase pre-analítica de los métodos clínicos. Se listaron una serie de puntos a tener en cuenta cuando se redacta un protocolo adecuado para cada caso. Eso cubre la parte de controlar la calidad desde el punto de vista de las precauciones a adoptar en la recolección de las muestras. La otra parte que falta, es la redacción del protocolo de la técnica de medición adoptada en cada caso particular, no solo debe ser claro y sencillo, sino que es imprescindible que todo el personal esté imbuído de la necesidad de seguir las normas a rajatabla. La única manera de calibrar una técnica es que el procedimiento sea repetitivo en las mediciones de control e igual al habitual que se usa en la rutina diaria.

Conviene aclarar un concepto: la *repetibilidad* de los resultados de mediciones de diferentes pacientes, esta sujeta a las condiciones siguientes:

- Mismo procedimiento de medición.
- Mismo observador, equipos e instrumentos.
- Misma ubicación.
- Mismas condiciones de uso.
- Mediciones realizadas en forma sucesiva (mínimo período de tiempo)

Por su parte, el concepto de *reproducibilidad* de mediciones, se refiere a la proximidad entre los resultados de las mediciones de la misma magnitud medida, sus condiciones son:

- Mismo procedimiento de medición.
- Diferentes observadores, equipos o instrumentos.
- Distintas ubicaciones.
- Distintos momentos.

La manera de calibrar es similar a las ya vistas. Con un suero control adecuado, se puede controlar la exactitud con el modelo Student y la precisión con la Chi-cuadrado. Análogamente, se puede comparar dos muestras para estudiar el efecto de algún factor sobre las mismas, para

detectar las causas de salida de control del sistema de medición o para investigación. De la comparación entre el promedio de las mediciones repetidas de un mismo suero control y el valor calibrado del mismo, se detecta la presencia de un error sistemático y se lo valida con el test estadístico. De probar la existencia del mismo, se lo cuantifica como la diferencia entre ambos valores y se pueden corregir las mediciones realizadas, o bien revisar el método a fin de eliminarlo. Si la experiencia no alcanza para detectar las causas de un error sistemático, se las puede investigar a través de comparaciones de medias haciendo variar un factor sospechoso de modo controlado. Y así, hasta encontrar los agentes que provocan el error. Si no se los puede eliminar, por lo menos se debe acotarlos.

La manera de proceder con la precisión es algo diferente a la vista en el control instrumental. Ahora, el criterio para adoptar ($\sigma_{\text{máx}}$) *la desviación máxima admisible* se basa en consideraciones clínicas, antes que en la precisión propia del instrumento. Esto se debe, a que en este caso, se calibran métodos cuyo objetivo final y principal es el diagnóstico. El *criterio de Thonks* es el más popular y antiguo adoptado por los bioquímicos.

$$CV\%_{\text{máx}} = 0,75 \frac{\text{Rango del intervalo de referencia}}{\text{Media del intervalo de referencia}} 100$$

De donde, se puede obtener para el caso de una calibración:

$$\sigma_{\text{máx}} = \mu \cdot CV\%_{\text{máx}}$$

Siendo μ el valor calibrado del suero control. Hay dos excepciones:

- 1) Para metabolitos en general se acepta un 10 % en el CV%.
- 2) Para enzimas en general es aceptable un CV% hasta un 20%.

En la Tabla 24.1 se presentan ejemplos de los valores de *Thonks* en las prácticas más usuales, comparándolos con los de otro nuevo criterio: el de *Aspen*. Este último fue presentado en la conferencia internacional del mismo nombre en 1977. Se basa en que: *la variabilidad biológica en una medición clínica es la misma, independientemente de la edad de paciente, del método elegido, del número de pacientes y de la condición de sano o enfermo crónico*. Se calcula como la mitad de la variabilidad biológica intraindividual expresado en términos del CV.

Tabla 24.1: Valores de desviación máxima aceptable.

Sangre o Plasma	Thonks		Monitoreo de Aspen		Prueba de Tamiz	
	EMA	CV% _{máx}	EMA	CV% _{máx}	EMA	CV% _{máx}
Glucosa	10,0	5,0	5,6	2,8	9,6	4,8
Urea	10,0	5,0	11,1	5,6	17,7	8,9
Colesterol	10,0	5,0	8,5	4,3	15,1	7,5
Urato	10,0	5,0	11,9	6,0	18,0	9,0
Proteína	7,1	3,6	3,0	1,5	4,8	2,4
Creatinina	10,0	5,0	6,3	3,2	11,9	6,0
Ion Sodio	1,6	0,8	0,5	0,3	0,8	0,4
Ion Potasio	5,3	2,7	1,6	0,8	2,2	1,1

EL EMA (*Error Máximo Admisible*) se define de acuerdo a varios criterios, algunos se basan en los intervalos de referencia poblacionales, otros totalmente subjetivos de acuerdo a la opinión de expertos, algunos generosos, otros exigentes. Todos suponen la eliminación de los errores sistemáticos y un nivel de significación del 95 %. Luego se define el *Coficiente analítico de variación* como la mitad del EMA: su desventaja es su gran dependencia con los valores poblacionales y el sistema de medición empleado.

Una característica importante del suero como material de control, con un valor asignado en forma analítica, es la aparición de *efectos de tamiz*. En cambio, con sueros acuosos esto no ocurre. A veces, un calibrador simple puede no ser adecuado al caso en cuestión y también se debe validar. La tabla anterior ilustra lo dicho con algunos ejemplos. Es importante que cada laboratorio determine por algunos de estos criterios, el valor máximo de variabilidad aceptable ($\sigma_{\text{máx}}$) para cada técnica en particular. Luego podrá hacer experimentos para ver si su técnica se mantiene dentro de esos valores y si es estable en el tiempo. Cosa que *debe constar* en el protocolo correspondiente.

24.5 Controles del método de rutina

Una vez conseguida la cantidad suficiente de sustancia de control (patrones) como para calibrar una técnica específica, definida su máxima dispersión admisible y con un protocolo detallado a seguir en la rutina, se está en condiciones de comenzar a calibrarla. La idea, una vez más, es que mediciones repetidas de una misma muestra fraccionada en alícuotas tienen una distribución de los errores de tipo gaussiano; en las condiciones explicadas en la Teoría de Pequeñas Muestras. Por eso, se pueden aplicar los modelos de Student, Chi-cuadrado y Fisher.

Cuando se efectúe una calibración por primera vez, los pasos a seguir son:

Paso 1) Tratar la sustancia control de la misma manera que la utilizada en la rutina diaria, respetando estrictamente el protocolo establecido para la técnica de medición.

Paso 2) Realizar n mediciones usando n alícuotas de la misma muestra y empleando al mismo operador, equipo y drogas en todos los casos. Así, se reduce al máximo la variabilidad metodológica (repetibilidad).

Paso 3) Con los n valores obtenidos se calculan la media \bar{x} y el desvío estándar DS.

Paso 4) Testear exactitud con el modelo de Student para una sola muestra; comparando la media muestral \bar{x} con el valor del patrón μ y determinar la existencia de errores de tipo sistemático.

Paso 5) Testear precisión con el modelo de la Chi-cuadrado comparando el desvío estándar muestral DS contra la máxima admisible $\sigma_{\text{máx}}$ y determinar la existencia de errores casuales de una magnitud indeseable.

Paso 6) Si en alguna de las dos pruebas estadísticas hechas, se detectan variaciones que hagan rechazar la hipótesis nula, entonces hay que revisar todo el sistema de mediciones y corregir lo que sea necesario, hasta lograr que la sucesión de pasos 1 a 5 tenga éxito.

Ejemplo 1) Se dispone de un suero control calibrado para Proteínas en 8,24 g/dl. Aplicando la rutina diaria se efectúan 9 mediciones y se obtiene un promedio de 7,9 g/dl con un desvío estándar de 0,3g/dl. Decidir si la técnica empleada por el laboratorio está calibrada.

Para controlar la exactitud se formula la hipótesis de que no existe error sistemático y se aplica el modelo de Student para una muestra. O sea, $H_0 : \mu = 8.24\text{g/dl}$

$$t = \frac{(7,9 - 8,24)}{0,3 / \sqrt{9}} = - 3,4^{**} \quad \text{Se decide rechazar la hipótesis planteada con el 99\% de confianza.}$$

Se encontró evidencia muy significativa que prueba de la existencia de un error sistemático por defecto de ES = - 0.34 g/dl. Entonces hay dos opciones, la primera es corregir las mediciones restando el valor de ES a cada una de ellas, la segunda, es revisar la técnica de medición hasta detectar la causa del error producido y corregirla.

Para controlar la precisión se emplea el criterio de Thonks y se calcula:

$$\sigma_{\text{máx}} = \mu \cdot \text{CV\%}_{\text{máx}} = 8,24 \text{ g/dl} \cdot 3.6\% = 0,297 \text{ g/dl}$$

Luego se formula: $H_0 : \sigma \leq 0.297 \text{ g/dl}$ y el sistema tiene una precisión aceptable.

$$\chi^2 = (9 - 1) \frac{(0,3)^2}{(0,297)^2} = 8,16 \text{ valor no significativo, o sea, no hay prueba de dispersión indeseable.}$$

Ejemplo 2) El bioquímico a cargo del caso anterior, decide investigar el origen del ES encontrado. Sospecha que se debe a una contaminación del kit empleado y decide abrir uno nuevo para investigar. Repite lo actuado con 4 mediciones y obtiene una media de 8,15 g/dl con un desvío de 0,25 g/dl. A la luz de ésta información, decidir si el problema se ha solucionado.

$$t = \frac{(8,15 - 8,24)}{0,25 / \sqrt{4}} = - 0,72 \quad \text{Valor no significativo}$$

Se concluye que el problema estaba en el Kit. Se solucionó el error sistemático. Se puede hacer un nuevo test para ver si hay diferencia entre los kits con:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{DS_1^2}{n_1}\right) + \left(\frac{DS_2^2}{n_2}\right)}} = \frac{(8,15 - 7,9) - (0)}{\sqrt{\left(\frac{0,25^2}{4}\right) + \left(\frac{0,3^2}{9}\right)}} = 1,56 \text{ (no significativo)}$$

El cual se contrasta con $t_{\alpha;v}$ donde $v = n_1 + n_2 - 2$ grados de libertad. Resultando que no se puede rechazar la hipótesis nula. Esto es no hay diferencia entre los kits. Notar que *a pesar saber que uno de ellos está descalibrado y el otro no*, cuando se los compara entre sí, no se detectan diferencias en exactitud.

Con respecto a la comparación en dispersión de ambos kits, se puede usar el modelo de Fisher con $H_0 : \sigma_1^2 = \sigma_2^2$ y se supone que ambos tienen la misma precisión.

$$F = \frac{(n_2 - 1)DS_1^2}{(n_1 - 1)DS_2^2} = \frac{(9 - 1)0,25^2}{(4 - 1)0,3^2} = 1,85 \text{ valor no significativo.}$$

Por lo que tampoco se detecta diferencias entre ambos kits, como era de esperar.-

24.6 Control de Calidad en Microbiología

Los siguientes conceptos han sido extractados de la obra “Mejoría Continua de la Calidad” editada por Confederación Latinoamericana de Bioquímica Clínica. Se recomienda seguir un programa de Control de Calidad Interno, estructurado, exacto y periódico, complementado con un programa externo. El programa de control interno, consiste en realizar en forma periódica una comprobación de las condiciones operativas de los aparatos de los instrumentos, juntos con los antisueros, medios de cultivo reactivo y tinciones empleados en la rutina diaria. El Control de calidad de los equipos, requiere de supervisar y registrar las operaciones que se realizan. En la Tabla 24.2 se resumen los controles a realizar al respecto.

También se debe controlar los medios de cultivo para asegurar su preparación adecuada. Se realizan tres tipos de controles de calidad para los medios:

1) *Control de esterilidad*: De cada lote disponible se eligen al azar de un 2 a un 5 % de los mismos. Luego se los somete a control incubando durante dos días para bacterias a 35-37°C o durante 5 días para hongos a temperatura ambiente.

2) *Control de calidad*: Algunos medios de cultivo se incuban con dos especies, por lo menos, de microorganismos que crecen en el medio. Uno que revela un grupo de especificaciones medias y otro con distintas especificaciones. Por ejemplo, si un medio de cultivo como el Agar SS es selectivo uno de los microorganismos como la *Salmonella spp.* crecerá y el otro en cambio no crecerá, como la *Escherichia Coli*, en una aerobiosis de 24 horas. En la Tabla 24.4 se muestran algunos microorganismos para realizar el control de calidad.

3) *Control de fecha de caducidad*: Se debe revisar que el medio de cultivo no se haya pasado de su fecha límite. Cada caja debe estar etiquetada con la aclaración del nombre del cultivo y la fecha de la preparación del mismo. A partir de ahí, se puede deducir su fecha de caducidad siguiendo los datos mostrados en la Tabla 24.3. Pero en general duran 15 días si se los guarda en heladera (4°C) y sin bolsa de plástico; mientras que los envasados en bolsas duran cuatro veces más.

Esta serie de recomendaciones efectuadas por la Confederación Latinoamericana se basan en la idea de que estos controles debe ser rutinarios en un laboratorio clínico, a diferencia de lo usual, que es recurrir a ellos cuando se presentan los problemas. El mantenimiento preventivo es dentro de un laboratorio, industria, etc., la base para evitar males mayores. Siempre es mejor prevenir que curar, en especial en mediciones clínicas. En la Tabla 24.2 se resumen estos aspectos:

Tabla 24.2: Control Periódico de Equipos.

Equipo	Procedimiento	Periodo	Límites de Tolerancia	Precauciones
Refrigerador	Registrar Temperatura	Diario	$(x \pm 1)^{\circ}\text{C}$	Limpieza mensual Descongelar cada 6 meses
Congelador	Registrar temperatura	Diario	$(x \pm 1)^{\circ}\text{C}$	Limpiar y descongelar cada 6 meses
Incubadora	Registrar temperatura	Diario	35 a 37 °C	Limpieza mensual
Incubadora CO ₂	Registrar temperatura y conc. de CO ₂	Diario	35 a 37 °C 5 % a 10 %	Limpieza mensual
Baño María	Registrar temperatura	Diario	36 a 38 °C	Limpieza mensual
Placa de Calentamiento	Registrar temperatura	Diario	$(x \pm 1)^{\circ}\text{C}$	Limpiar cada vez que se use
pH (medidor)	Solución Calibradora	Cada vez que se use	$x \pm 0.1$	Limpiar cada vez que se use
Autoclave	Esporas de <i>B.Stearot</i>	Semanal	No Crecimiento	Limpieza mensual y agregar agua
Microscopio	Limpieza Correcta	Cada vez que se use		Mantenimiento general cada 6 meses
Jarra de anaeróbicos	Tira de azul de metileno	Cada vez que se use	De azul a blanco	Reactivo catalizador después de usarse y Cambio cada 3 meses
Centrífuga	Comprobar r.p.m.	Mensual	$x \pm x 0,05$	Inspección general y cada 6-12meses
Campana de Seguridad	Flujo de aire y luz UV	Cada seis Meses	45 a 55 pies / min.	Limpieza diaria, inspección general y mantenimiento anual
Medio Ambiente	Registrar temperatura	Diario	22 a 30 °C	Inspección general anual
Humedad	Registro	Diario	30 a 60 %	Evitar aparatos que la modifiquen

Tabla 24.3: Período de uso de caja con medios de cultivos almacenados a 4°C.

Medio	Refrigerador 4°C sin bolsa de plástico	Refrigerador 4°C con bolsa de plástico
Agar Sangre	15 días	50 días
Agar Chocolate	15 días	60 días
Agar Thayer-Martin	15 días	60 días
Agar Mueller-Hinton	15 días	70 días
Agar Mac Conkey	15 días	70 días
Agar EMB	15 días	70 días
Agar CLED	15 días	70 días
Agar Saboureaud	21 días	90 días
Agar de Manitol - Salt	15 días	60 días
Agar SS	6 días	10 días

Tabla 24.4: Microorganismos usados en el control de calidad de medios de cultivos.

Medio	Organismo Control	Incubación	Reacción Esperada
Agar sangre	<i>Streptococcus grupo A</i> <i>Streptococcus pneumoniae</i> <i>Listeria manocyctogen</i>	Aerobiosis 24 h	Crecimiento. Beta hemólisis Crecimiento. Alfa hemólisis Crecimiento. Beta hemólisis
Agar chocolate	<i>Haemophilus influenza</i> <i>Neisseria gonorrhoeae</i>	Aerobiosis 24 h y CO ₂	Crecimiento Crecimiento
Agar Hektoen	<i>Salmonella spp.</i> <i>Escherichia coli</i> <i>Shigelia spp.</i>	Aerobiosis 24 h	Colonias verdes, centro negro Col.Naranjas,crec. ligeramente inhibido Colonias verdes
Agar de DNAsa	<i>Serratia marcescens</i> <i>Enterobacter cloacae</i>	Aerobiosis 24 h	Zona clara (conHCl) Sin zona clara (R ⁻)
Agar de EMB (Levine)	<i>Escherichia coli</i> <i>Proteus mirabilis</i>	Aerobiosis 24 h	Crec. Brillo verde metálico Crecimiento. Sin color
Agar de MacConkey	<i>Escherichia coli</i> <i>Proteus mirabilis</i> <i>Enterococcus faecalis</i>	Aerobiosis 24 h	Crec.Colonias Rojas Crec. Sin color. No crec. No crecimiento
Agar de Manitol-salt	<i>Staphylococcus aureus</i> <i>Escherichia coli</i>	Aerobiosis 24 h	Crec. Colonias Amarillas No crecimiento
Agar SS Salmonella-Shigella	<i>Salmonella spp.</i> <i>Escherichia coli</i>	Aerobiosis 24 h	Crec. No color. Centro Negro. No crecimiento
Agar de Thayer-Martin	<i>Neisseria gonorrhoeae</i> <i>Escherichia coli</i> <i>Staphylococcus aureus</i>	Aerobiosis 24 h y CO ₂ Aerobiosis 24 h Aerobiosis 24 h	Crecimiento. Sin color No crecimiento No crecimiento

24.6.1 Control de Calidad para exámenes bioquímicos

Si bien hay prácticas bioquímicas que se realizan en cajas con medios de cultivos, la mayoría se efectúan en tubos. Y por lo tanto, estos medios deberán ser sometidos a control en forma análoga a como se procede con las cajas. Por ejemplo algunas pruebas de identificación se realizan en cajas Petri como la prueba de Cap, la DNAsa, etc. Pero también se incluyen las supervizaciones de medios de diferenciación, formas de identificación, reveladores de reactivos y reactivos que puedan revelar actividad bacteriana. Los tres tipos de controles para los ensayos en tubos son análogos a los anteriores:

1) *Control de esterilidad*: Se realiza incubando a temperaturas entre 33 a 37 °C, durante 5 días, a un grupo de tubos seleccionados al azar del lote disponible. Se debe verificar que no haya ningún tipo de crecimiento.

2) *Control de fechas de caducidad*: Hay dos tipos de medios que pueden contener el tubo: agar o medio líquido. De eso depende la fecha de caducidad, según la forma de almacenamiento bajo diferentes condiciones. En la Tabla 24.5 se muestran los casos más comunes en agar y en líquido, los tipos de tapón, cerraduras y con sellos de seguridad.

3) *Control de calidad*: Se siembran dos tubos, con dos tipos de microorganismos que contienen el mismo caldo de cultivo o medios con agar. Los microorganismos se eligen para observar una diferenciación entre ambos, de manera tal, que uno genere una reacción típica y el otro no. Puede ser de distintas características fisiológicas o de crecimiento. Por caso, en el medio Lisina Descarboxilasa, una aerobiosis 24 hs con *Escherichia Coli* mostrará un crecimiento de color rojo. Mientras que *A. Calcoaceticus*, *v lwoffii* crecerá sin color rojo. En la Tabla 24.6 se muestra una lista de los microorganismos de control más frecuentes de la rutina diaria, sus características y sus reacciones típicas.

Tabla 24.6 Organismos de Control para algunos medios en tubos –Exámenes Bioquímicos

Medio	Organismo control	Incubación	Reacción Esperada
Arginina descarboxilasa	Enterobacter Cloacae Proteus mirabilis	Aerobiosis 24 hs	Crec. Color azuloso (+) Crec. Color amarillo (-)
Citrato de Simmons	Klebsiella pneumoniac Escherichia coli	Aerobiosis 24 hs	Crec.color azul (+) No crec. Permanece verde (-)
Fenilalanina (Fe Cl3)	Proteus mirabilis Escherichia coli	Aerobiosis 24 hs	Crec.Color verde (+) Crec. Color original (-)
Indol (Kovac' s)	Escherichia coli Klebsiella pneumoniac	Aerobiosis 24 hs	Crec.Color Rojo (+) Crec. Sin Color (-)
Agar de Kligler	Escherichia coli Shigella flexeri Proteus mirabilis	Aerobiosis 24 hs	Acida en la superficie y profundo Alcalina en superficie y acida en profundo. Alcalina en superficie y profundo
Agar de Lisina-hierro	Salmonella typhimurium Shigella flexneri Proteus mirabilis	Aerobiosis 24 hs	Morado en superficie y profundo, H2S+ Morado en superficie y amarillo en profundo Fase sup. Roja, inf. Amarilla
Lisina Descarboxilasa	Salmonella typhimurium Shigella flexneri	Aerobiosis 24 hs	Crec. Color azuloso Crec. Color amarillo
Indol-Ornitina Motilidad Agar semisolido	Escherichia coli Klebsiella pneumoniac	Aerobiosis 24 hs	Mot. (+), Indol (+), Ornitina (+) Mot. (-), Indol (-), Ornitina (-)
Nitrato (Griess)	Escherichia coli A.Calcoaceticus, v. Twoffi	Aerobiosis 24 hs	Crec. Color Rojo Crec. No color rojo.
Dextrosa de O-F	Pseudomonas aeruginosas Aerobiosis Staphylococcus aureus A.Calcoaceticus, v. Twoffi	Aerobiosis 24 hs- 48 hs	Crec. Oxidativo (o) Crec. Fermentativo F Crec.Reacción alcalina
Azúcar triple-hierro	Shigella flexneri Citrobacter Pseudomonas aeruginosas	Aerobiosis 24 hs	Superficie alcalina ,ácido en profundo Superficie ácida y tubo H2S (+) Alcalina en superficie y tubo
Urea de Christensen	Proteus mirabilis Klebsiella pneumoniac Escherichia coli	Aerobiosis 24 hs	Rosa todo el medio Rosa solo en superficie Crec. Sin Color
Agar semi sólido Motilidad	Proteus mirabilis Klebsiella pneumoniac	Aerobiosis 24 hs	Crec.medio nebuloso (+) Crec. Sin borde ondulados (-)
Vogues Proskauer Sin reactivo	Klebsiella pneumoniac Escherichia coli	Aerobiosis 24 hs	Crec. Color rojo Crec. Permanece color

Tabla 24.5: Período de usos de algunos agar y caldos de cultivos.

Medio	Tapón no hermético 4°C Semanas	Tapón no hermético	Cerradura hermética, en bolsa de plástico	Sello de seguridad
		Temp. Ambiente Semanas	Temp. Ambiente Meses	Temp. Ambiente Meses
Caldo de Mueller-Hinton	3 a 4	1 a 2	2 a 4	3 a 6
Infusión cerebro-corazón	3 a 4	1 a 2	2 a 4	3 a 6
Caldo de Soya – Trypticase	3 a 4	1 a 2	2 a 4	3 a 6
Caldo de Triptofano	3 a 4	1 a 2	2 a 4	3 a 6
Caldo de nitrato	3 a 4	1 a 2	2 a 4	3 a 6
Caldo de selenita	3 a 4	1 a 2	2 a 4	3 a 6
Caldo de Tetrionato	3 a 4	1 a 2	2 a 4	3 a 6
Caldo de Azúcar al 1%	3 a 4	1 a 2	2 a 4	3 a 6
Agar inclinado	3 a 4	1 a 2	2 a 4	3 a 6
Agar de Billis Esculina	3 a 4	1 a 2	2 a 4	3 a 6
Agar de Fenilalanina	3 a 4	1 a 2	2 a 4	3 a 6
Agar de Hierro-Lisina	3 a 4	1 a 2	2 a 4	3 a 6
Agar Mot-Indol Ornitina	3 a 4	1	2 a 4	3 a 6
Agar de Mueller-Hinton	3 a 4	1 a 2	2 a 4	3 a 6
Agar nutritivo	3 a 4	1 a 2	2 a 4	3 a 6
Agar O-F	3 a 4	1	2 a 4	3 a 6
Agar sangre	3 a 4	1 a 2	2 a 4	3 a 6
Agar triple azúcar –hierro	3 a 4	1 a 2	2 a 4	3 a 6
Urea de Christensen	3 a 4	1 a 2	2 a 4	3 a 6

24.7 Problemas propuestos

- 1) Las cuatro partes de un sistema de medición son: V F
- 2) Las sustancias de control son Patrones acuosos y sueros calibrados. V F
- 3) El suero control más conveniente se hace con suero de animales. V F
- 4) Lo más barato para hacer el control es usar un pool de sueros. V F
- 5) El protocolo es la parte del método de medición. V F
- 6) Las normas internacionales piden un año de estabilidad a los patrones liofilizados. V F
- 7) Los sueros control no deben estar contaminados con HIV I y HIV II. V F
- 8) En la fase pre-analítica del protocolo hay 9 puntos a tener en cuenta que son:
- 9) La repetibilidad de un método implica igual procedimiento, observador y equipos. V F
- 10) La reproducibilidad se refiere a mismos resultados en distintos momentos. V F
- 11) Para obtener la dispersión máx. admisible se pueden usar los criterios de Aspen. V F
- 12) La variabilidad biológica de una técnica es independiente de la edad del paciente. V F
- 13) Toda sustancia de control debe manipularse como si estuviera contaminada. V F
- 14) Cuando se efectúa una calibración por primera vez los 6 pasos a seguir son:
- 15) La exactitud de un método de medición se testea con el modelo de:
- 16) La precisión de un método de medición se testea con el modelo de:.....
- 17) Las hipótesis nulas para testear exactitud y precisión son:
- 18) El control de los equipos debe ser periódico. V F
- 19) Las esporas usadas para controlar un autoclave son las
- 20) Las incubadoras y baños maría deben limpiarse
- 21) Hay que descongelar cada semestre las:
- 22) Cada vez que se use hay que limpiar los medidores de pH y
- 23) Para los medios de cultivo se eligen al azar de un ... % a un% de los mismos.

25

Control de Calidad Estadístico

En los dos capítulos anteriores de control de calidad se han desarrollado los dos primeros pasos de control, a saber, primero se deben calibrar todos los instrumentos del laboratorio para poder avanzar hacia el siguiente paso, calibrar todos los métodos de medición usados en la rutina diaria. En ese punto, cuando todo está ajustado y controlado, se empieza a trabajar. A medida que pasa el tiempo, es de esperar la aparición de fluctuaciones indeseables que alteren los resultados. Es aquí donde aparece la necesidad de la tercera etapa de control de calidad: *la estadística*, donde cada día se controla que las rutinas se mantengan dentro de un rango de calidad aceptable. Para esto se utilizan las llamadas *cartas de control de calidad* derivadas de la solución del mismo problema en la industria: Las cartas de Shewhart de la década del 30. Para un laboratorio de análisis clínicos, Levey-Jennings presentaron en la década del 50 el método más usado por los laboratorios en la actualidad. En la década del 80 Westgard le agrega una serie de reglas y la técnica de las sumas acumulativas para retroalimentar la información. En nuestro país, la Asociación Bioquímica Argentina (ABA) ha modificado ligeramente el trabajo de Westgard, para adaptar estas reglas a la realidad local. En todos los casos, la idea básica es preparar una gráfica con un valor central y dos bandas de tolerancia por encima y por debajo de la misma. Cada tanto se mide una muestra de control con la rutina diaria, si éste valor cae fuera de las zonas y/o reglas de aceptación se debe revisar el sistema, que así se considera fuera de control. Mientras en USA recomiendan una medición de control cada veinte de rutina, la ABA sugiere una por día. Con el advenimiento de las computadoras, otras técnicas más sofisticadas como las del análisis de tendencia son más accesibles para la mayoría que en el pasado.

Además de éste tipo de control de calidad dentro del laboratorio, existen programas de control *intra-laboratorios*. Para este tipo de programas se reparten muestras ciegas a los participantes, quienes deben medirlas y enviar sus resultados al laboratorio de referencia, el que se encarga de evaluarlas y remitirles las conclusiones, el ranking entre los demás participantes y una serie de informaciones útiles.

25.1 Conceptos básicos

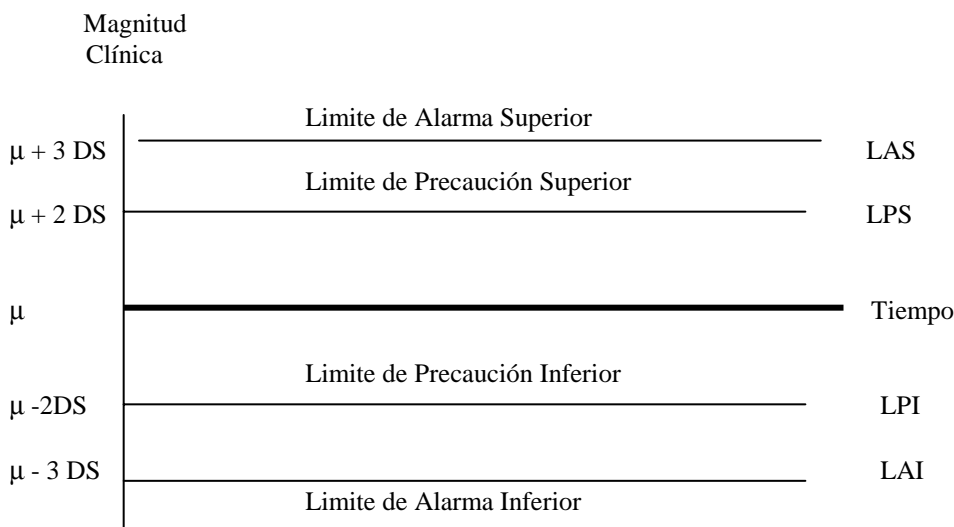
Las técnicas de medias y espectro emplean una gráfica de sentido horizontal en el eje de ordenadas se coloca en el centro el valor esperado μ (valor certificado del suero control), por encima de este se grafican tres líneas a 1, 2 y 3 DS, lo mismo se hace por debajo de μ . En el eje de abscisas se coloca el tiempo (en días), o bien, el número de corridas que se van efectuando. En la Figura 25.1 siguiente se muestra un esquema de este tipo.

Los límites de las zonas de control dentro de las Cartas se definen como sigue:

Límites de precaución: Están dados por la relación ($\mu \pm 2 DS$).

Límites de alarma: Están dados por la relación ($\mu \pm 3 DS$).

Figura 25.1: Diseño básico de las cartas de control.



En términos generales, cuando un valor medido cae por fuera de la franja de alarma es decir fuera del intervalo (LAS, LAI), significa que el sistema está fuera de control. Entonces, debe ser detenido y hasta que no se corrija el problema no se puede seguir con la rutina diaria. En cambio, cuando el valor del suero control medido ese día, cae dentro de la zona comprendida entre los límites de precaución y alarma; entonces se toma un primer aviso de que algo puede llegar a ocurrir y sacar al sistema de control. Finalmente, si el valor cae dentro de (LPS, LPI) se interpreta como que el sistema está controlado y se puede seguir.

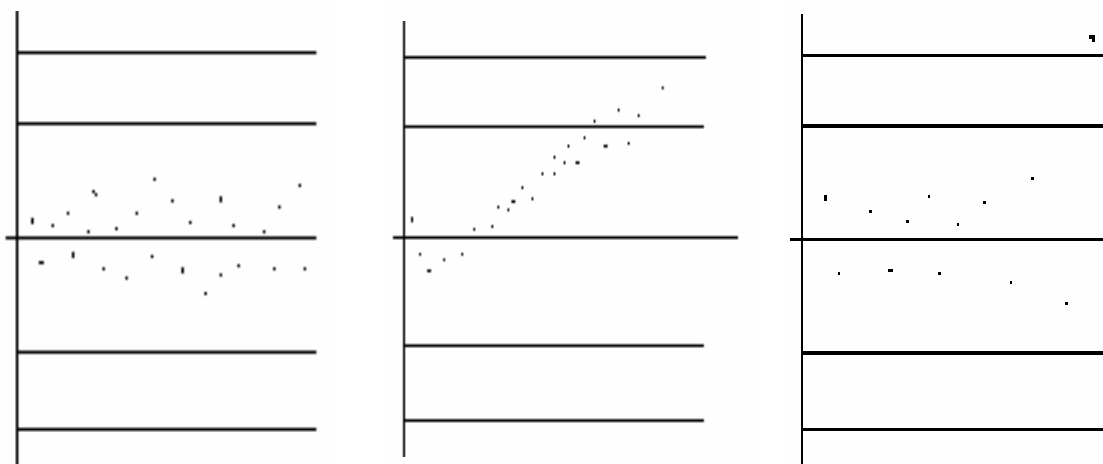
Hay una serie de reglas para interpretar algunas situaciones dadas por la ABA, como se explicará más adelante. Además de todas ellas, también se pueden detectar tendencias mirando la gráfica diaria. Existe el peligro de los falsos rechazos y de las falsas aceptaciones. O sea, valores que hacen detener el proceso, cuando en realidad no debiera ser así y viceversa. Una vez más, se tiene el cuadro de situación ya visto en el tema de índices clínicos, en el Tema 4. Lo ideal es que la carta de control posea unas reglas de trabajo que minimicen éste problema.

La idea general es que el valor del DS es una estimación del verdadero valor de σ en la población. Entonces, se trabaja con los límites poblacionales. Sin embargo, una objeción fue presentada en un trabajo anterior. Si se mira bien, lo que se está haciendo es una estimación de la media de las mediciones y los límites deberían ser DS / \sqrt{n} . Pero, esto angostaría la franja de aceptación de las cartas, cosa que no parece ser del agrado de los usuarios finales.

25.2 Cartas de control de Levey-Jennings

Estas cartas se interpretan de acuerdo a ciertas reglas simples como se explicó más arriba. Se considera al sistema bajo control cuando, el valor medido ese día, de las muestras patrones (suero control en viales medidos con la rutina diaria) caen en la zona de aceptación delimitada por ambos límites de precaución. Es decir, si el valor medido de glucosa por ejemplo es g , y al colocarlo en la gráfica $g \in (LPS, LPI)$ entonces, la glucosa está controlada y se puede comenzar a medir a los pacientes del día. Esta situación se muestra en el caso (a) de la Figura 25.2. Cuando los valores de control de dos días consecutivos, se encuentran entre el límite de precaución y de alarma, y del mismo lado, el sistema se considera fuera de control. Por lo tanto hay que detenerse y revisar el sistema para corregir el defecto, como se muestra en el caso (b) de la Figura 25.2. La otra situación de fuera de control, ocurre cuando un solo valor cae fuera de los límites de alarma, como se muestra en el caso (c) de la Figura 25.2. Para los niveles de precaución, la probabilidad de un falso rechazo, aumenta con el número de determinaciones N . Por ejemplo, si $N = 1$ la probabilidad es 0,05; para $N = 2$ es 0,09; para $N = 3$ es 0,14; etc. En cambio cuando se rechaza fuera de los límites de alarma, los valores de falso rechazo están entre 0,01 y 0,02 para valores de N comprendidos entre 8 y 10.

Fig. 25.2 Ejemplos de cartas de control



(a) Sistema bajo control

(b) Fuera de control

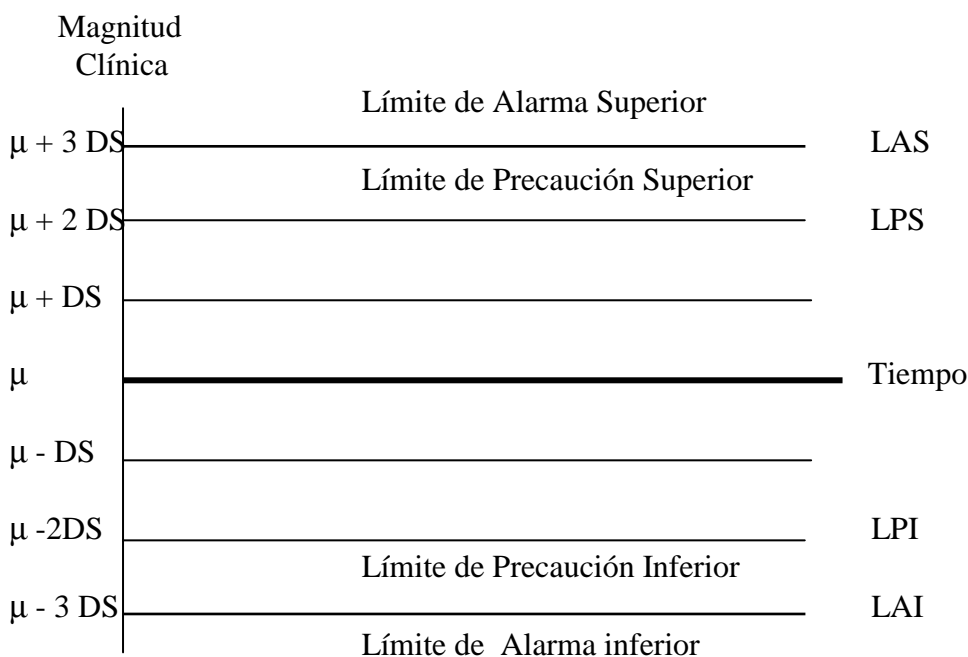
(c) Fuera de control

Además de esas reglas, mirando el gráfico se pueden interpretar una serie de cosas, por ejemplo en el caso (b) anterior: A simple vista se nota una cierta tendencia del sistema a salirse de control mucho antes de llegar al caso de corte. Si se ven varios valores sucesivos, por encima (o por debajo) del promedio, aunque se mantengan dentro del control, habrá una fuerte sospecha de un error de tipo sistemático que por el momento no llega a ser significativo. Un Bioquímico prudente se pondrá a investigar que está pasando antes de salirse de control. El hecho de tener oscilaciones bruscas en ambos sentidos, en forma consecutiva, alerta sobre la posible presencia de un error de tipo casual. Cuando ambos errores empiezan a actuar en conjunto, en forma progresiva, la tendencia de los puntos en el tiempo comienza a cambiar en forma apreciable en la carta de control.

25.3 Cartas de control Westgard

Estas cartas se denominan también “multiregla”. Son más útiles para el bioquímico sin mucha experiencia en la detección de problemas mirando los datos de la carta, pues se analizan con una serie de reglas y una serie de cambios más sutiles que los vistos en el punto anterior. Los datos de control y las cartas se presentan al estilo de Levey-Jennings, pero ahora se agrega otras líneas de control al nivel $\pm 1 DS$.

Figura 25.3: Carta de Control de Westgard



El grupo de reglas recomendado por la Asociación Bioquímica Argentina es:

Regla 1 :3 DS Se debe detener el sistema cuando un punto cae por fuera de la zona de alarma.

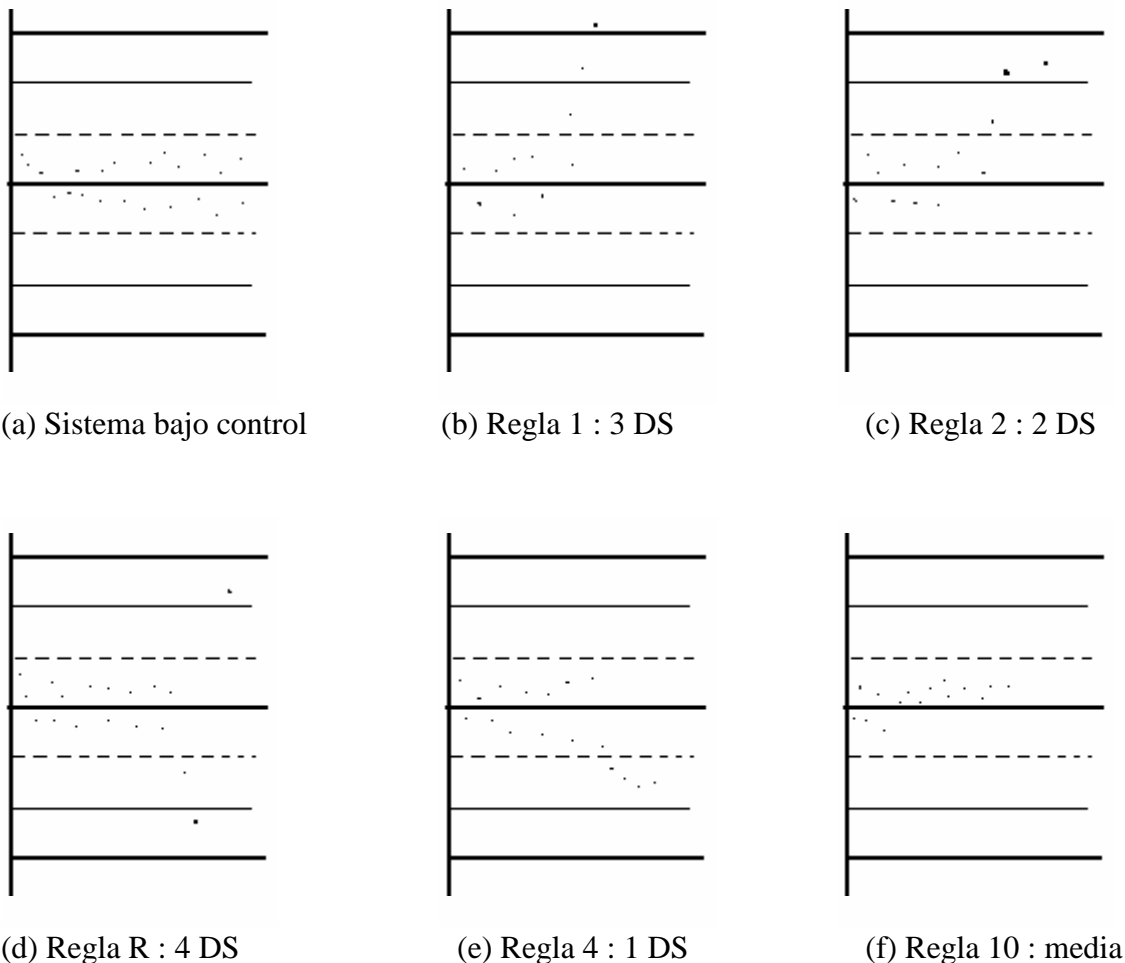
Regla 2 :2 DS Cuando 2 observaciones consecutivas, caen entre los límites de precaución y de alarma, del mismo lado, se considera al sistema fuera de control.

Regla R : 4 DS Se debe detener al sistema cuando dos observaciones sucesivas, están separadas entre sí por 4DS o más.

Regla 4 :1 DS Si cuatro mediciones consecutivas caen en la zona definida entre 1 DS y 2 DS del mismo lado, se debe considerar al sistema fuera de control.

Regla 10 : Media Si 10 mediciones consecutivas, se ubican por encima (o por debajo) del valor promedio, aún cuando estén dentro de los límites aceptables, se debe parar al sistema.

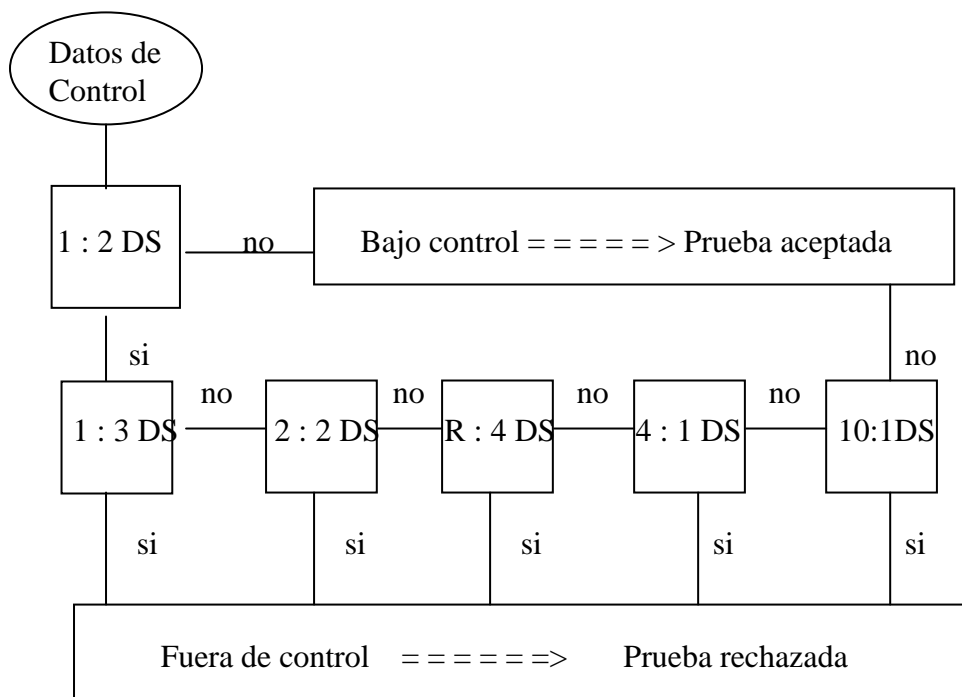
Fig. 25.4 Ejemplos de las reglas de la ABA.



En la Figura 25.4 puede verse una ilustración de las reglas de la ABA. En el caso (a) todos los puntos están dentro de la zona de control y no se ha violado ninguna regla, por que se puede decir que el sistema está *bajo control estadístico* de calidad. En el caso (b) la regla 1 : 3 DS ha sido violada, porque el último punto está encima del LAS, en este caso deben detenerse las mediciones y arreglar el sistema antes de continuar. El consejo de la ABA, es realizar de inmediato una nueva medición con el suero control, para asegurarse que realmente esté fuera de control y que no se trate de un accidente fortuito. Si de nuevo cae por fuera de los LAS, entonces sí hay que detener el sistema. En el caso (c) no se cumple la regla 2 : 2DS, porque las dos últimas mediciones seguidas están en la zona comprendida entre los LPS y LAS. También en este caso, la ABA aconseja repetir la última medición. En el caso (d) el anteúltimo valor cae por debajo del LPI y el siguiente, por encima del LPS, están separados por una distancia mayor a los 4DS y se viola la regla R : 4 DS. En el caso (e) las últimas 4 mediciones están por encima de 1 DS, aunque por debajo del LPS, de todas formas se viola la regla 4 : 1 DS y hay que detener el sistema. La evidencia de la aparición de un error sistemático por exceso es suficiente para parar. Por último, en el caso (f) hay 10 valores por encima del valor de referencia. En este caso en particular, la ABA sugiere detener el sistema con solo 9 casos. Se considera que se ha roto la regla 10 : media y hay que detener al sistema. Todos estos casos, debe ir acompañados por una medida prudente de realizar una nueva medición antes de detener totalmente el sistema. La prudencia, aconseja ir analizando diariamente la tendencia de los puntos en la carta, para prevenir sorpresas.

En la Figura 25.5 se presenta un esquema lógico, a seguir para tomar las decisiones apropiadas, según las reglas anteriores. Es un diagrama de flujo como el de programación en Análisis de Sistemas o programación de computadoras. Ayuda a facilitar la comprensión de las reglas de Westgard en las decisiones a tomar. Fue presentado por este autor en su trabajo referido en la bibliografía de este capítulo.

Figura 25.5 Diagrama lógico de las reglas de Westgard.



Cuando se mide un dato de control se procede siguiendo el diagrama lógico de arriba. Se supone que el sistema está bajo control estadístico. La primera cuestión es ver si cae fuera de los límites de precaución. Si la respuesta es no, el sistema sigue bajo control. Si la respuesta es sí, entonces se deben controlar las reglas de Westgard. Lo primero es ver si no supera los límites de alarma (regla 1 : 3 DS), si la respuesta es si entonces se debe detener al sistema, se lo considera *fuera de control* y se rechaza la prueba. Se sigue verificando si el punto no es el segundo en caer por encima del mismo límite de precaución (regla 2 : 2 DS), o bien si la distancia con la medición de control anterior, no supera los 4 DS (regla R : 4 DS). Cada vez que la respuesta sea si, el sistema de medición se detiene. La otra cuestión es ver si no se ha violado la regla 4 : 1 DS, esto es que no sea el cuarto punto seguido en caer por encima de 1 DS, del mismo lado. Por último se debe verificar que no sea el décimo (o noveno) punto en caer por encima, o por debajo de la media (regla 10 : media). Si la respuesta sigue siendo no, entonces se considera al sistema bajo control hasta la próxima medición de control.

Además de lo visto, este método multiregla se puede usar con repetición. Esto es, en lugar de una sola carta de control por magnitud, trabajar con dos. Una de ellas, con valores de control cercanos al valor referencial superior de la magnitud clínica en la población. Por ejemplo, en el caso de la glucosa, trabajar con un valor de 1 mg/dl, pues el valor de referencia es 1,06 mg/dl. La otra carta de control se hace con un suero control con valor cercano al límite inferior de referen-

cia en la población. De esta manera, se puede estudiar mejor el comportamiento del método de medición, en las proximidades de los valores de *alarma médica*. Se debe recordar que en realidad, esos son los valores más importantes en lugar de uno intermedio. La riqueza de información lograda mejora mucho la prestación de laboratorio de análisis clínicos. Como contrapartida, se aumenta el costo del servicio de control de calidad.

Otra forma de trabajar, es usando una carta para los valores promedio de la magnitud en la población y otra carta con los valores diluidos para la zona inferior de referencia, o aumentados para la parte superior. Esto, de acuerdo a la magnitud clínica en cuestión y a la disponibilidad de fondos del laboratorio.

25.4 Cartas corregidas

Este tipo de cartas, reutilizan la información que se recoge diariamente en los controles de calidad, para mejorar la carta con los nuevos datos. El procedimiento a seguir puede ser descrito como sigue:

Paso 1) Suponiendo que el sistema haya sido calibrado y ajustado con los métodos descritos en los dos capítulos anteriores, el primer paso es conseguir una cantidad de suero control o “pool” de sueros calibrado, suficiente como para seis meses de trabajo.

Paso 2) Para construir la primer carta de control, se deben hacer por lo menos veinte mediciones repetidas seguidas del suero control y con estos datos se calculan la media y el desvío estándar.

Paso 3) Cada día, o cada 20 mediciones de rutina, se mide un valor control que se vuelca en la carta. Se sigue el diagrama lógico explicado en el punto anterior para determinar si el sistema sigue bajo control estadístico de calidad.

Paso 4) A medida que pasa el tiempo se van juntando datos y la idea es aprovechar esta información extra. Por ejemplo, se realizaron veinte mediciones de control rutinario y en ese momento se tienen cuarenta valores del mismo suero control, realizadas con la misma metodología. Con todos los valores reunidos, se recalculan la media y el desvío estándar. Cuanto más mediciones se tengan, más confiable es la estimación realizada. Y con esa información se construye la segunda carta de control, corregida con los datos extra.

Paso 5) Se repite el proceder visto en el *Paso 3* hasta juntar otros veinte datos. Ahora con los sesenta datos disponibles, se repite el *Paso 4*. Se construye la segunda carta de control corregida. Y así sucesivamente, hasta terminar con el stock del suero control.

Como durante todo este proceso, se dispone de una técnica clínica calibrada y controlada, conceptualmente es lo mismo que tener un aparato de medición de referencia, entonces se la puede emplear para calibrar un nuevo “pool” de sueros fabricado en el mismo laboratorio. De esta forma, usando los sobrantes diarios del suero de los pacientes tratados en el día, se puede ir acumulando una gran cantidad de suero representativo de la población que concurre a atenderse en el laboratorio. Solo se necesita un congelador para mantener a -7° C el “pool” que se va juntando. Una vez que se acumularon varios litros de suero se procede a fabricar más suero control.

Al tener un volumen suficiente para un semestre, se descongela todo el stock acumulado de una sola vez. Se mezcla y homogeneiza de la mejor manera posible y luego se llenan varios miles de viales con el suero. Cada vial se cierra herméticamente y se vuelve a guardar en el congelador, excepto veinte de ellos por cada magnitud clínica a controlar, que se separan para calibrarlos con el método usual calibrado disponible, como se vio en el *Paso 2* anterior.

Con los valores de calibración de este nuevo stock de sueros se completa la fase de fabricación. Entonces, se lo almacena hasta que se termine el stock del suero control anterior y se requiera al nuevo recién fabricado.

De esta manera, un laboratorio de análisis clínicos puede autoabastecerse a bajo costo de sueros calibrados para el control de calidad estadístico.

Ejemplo) Los datos acumulados en el control de calidad por un laboratorio se presentan en la tabla siguiente. Se pide calcular los nuevos valores para ir corrigiendo las cartas de control.

Mes	Cantidad	Promedio	Desvío estándar
0		100,5	2,1
1	20	100,5	2,1
2	20	101,1	2,5
3	20	100,6	2,4
4	20	100,4	2,2
5	20	100,7	2,6

Acumulado	Promedio corregido	Desvío est. corregido
0		
20	100,5	2,10
40	100,3	2,31
60	100,4	2,34
80	100,4	2,31
100	100,5	2,37

Las técnicas de control explicadas hasta ahora se conocen como *Métodos de media y Espectro* para diferenciarlas de una nueva propuesta: *Análisis de tendencias*. Cambrowski propuso en 1975 la idea de agregar la información lograda en el control diario de a una por una, en lugar de esperar veinte de ellas para poder corregir la carta de control. Acá la idea es hacer una nueva carta, cada vez que se mide un valor de control. Mediante el empleo de cálculo exponencial se pueden obtener nuevas estimaciones de la media y el desvío estándar, pero con una característica muy interesante: *Se le da más peso a las últimas mediciones que a las primeras*. Esto es, las mediciones más antiguas van perdiendo importancia a medida que pasa el tiempo, en forma gradual. Esto parece aplicar mejor la idea de “actualización” de las cartas de control. A primera vista, esto puede aparentar ser muy laborioso, pero desde la irrupción de las computadora de mesa tipo PC, se hace sencillo el fabricar estas cartas con un programa adecuado.

Finalmente, se debe recordar que un análisis más profundo de las características de una serie de mediciones, se puede realizar con los modelos de ANOVA. Con esos modelos se puede salvar la limitación más grande de estas cartas: *Solo se puede controlar un factor por vez*. En cambio, con el ANOVA se podría analizar el efecto de más de un factor interactuando con las muestras. Por ejemplo, estudiar la incidencia del factor humano controlando varios operadores, ver el efecto de diferentes marcas comerciales de kits, varios tipos de instrumentos, etc.

25.5 Especificaciones del error en mediciones clínicas

En el supuesto de que el sistema de medición clínica tenga un comportamiento estable, la forma de tratar el error más común en esta época es considerar al error total de la medición, como la suma de los errores sistemáticos y casuales. Lo que se puede resumir en la ecuación siguiente:

$$TE \geq ES + 2 DS$$

Donde: TE : Error total; ES : error sistemático y 2 DS error casual con un 95% de confianza. Este supuesto tiene dos componentes: uno para la exactitud y otro para la precisión de acuerdo a los conceptos desarrollados previamente. Sin embargo, en la literatura clínica de hoy los términos usados cambian ligeramente, a saber:

Imprecisión (I%): Se refiere a la dispersión entre mediciones repetidas de la misma muestra y se calcula con el desvío estándar de 20 o más mediciones de un material estable. Se prefiere el término imprecisión antes que el de precisión, porque el desvío estándar o el coeficiente de variación, describe el error o dispersión entre las mediciones. También el término error casual se usa indistintamente, y otros términos como: imprecisión estable, error casual estable o imprecisión inherente. Estos se refieren al comportamiento estable del sistema de medición. Al principio este valor se calcula con las mediciones repetidas, pero después se estiman con las mediciones efectuadas para las cartas de control de calidad del mismo material de control. El valor 2 DS se relaciona con un intervalo de confianza del 95% pero también se pueden usar 3 DS para el 99,9% y ocasionalmente 4 DS para tener el más alto valor de confianza (casi el 100%). Este último valor se recomienda usar para poner un valor más ajustado a la carta de control y tener un margen de seguridad apropiado en las cartas modernas.

Sesgo o Bias (B%): Se refiere a la cercanía del valor medio de las 20 mediciones efectuadas con el valor de referencia o control. En inglés se usa el término "Inaccuracy" o "bias" para describir el mismo concepto visto antes como error sistemático (ES). El término inexactitud estable se usa para aclarar que se trata de un sistema que está trabajando apropiadamente en condiciones estables. Cuando se comparan dos métodos, el que se venía usando y uno nuevo (cambio del pool de sueros etc.), la primera estimación de la inexactitud se calcula con la diferencia de media entre ambos, o a través de la regresión estadística para las concentraciones ciertas magnitudes clínicas. Esta primer estimación se compara con el error máximo admisible para juzgar si es aceptable en la rutina del laboratorio. Más adelante, cuando se tiene trabajando a la carta de control, se la puede calcular como la diferencia entre el promedio de los valores control de la carta y el valor promedio calculado al principio con las 20 mediciones repetidas del material de control.

Error Total (TE%): Describe la combinación de los dos términos anteriores como una suma algebraica de los mismos, cuando ambos se combinan para dar el peor de los casos posibles en una medición clínica individual. Conceptualmente es la suma de los errores casuales y sistemáticos. Suponiendo que no hay error sistemático, el valor $z = 1,65$ incluye el 90% de confianza en un ensayo de dos colas, y para $z = 2, 3$ o 4 se incluyen el 95%, 99,9%. En cambio, cuando la inexactitud sea lo suficientemente grande, es como tener un ensayo de una cola y los valores de z serán 1,65, 2, 3 o 4 para tener un 95%, 97%, 99,9 y casi el 100% respectivamente. Por ejemplo, para que una medición de colesterol sea aceptable debería tener un 3% de error sistemático y 3% de coeficiente de variación, para que el total máximo admisible sea del 9%.

Cuadro 25.1: Imprecisión, Inexactitud y Error total admisible en Europa. Los valores mostrados son los propuestos y entre paréntesis los interinos. Expresados como el CV% a un nivel de 95%

Analito	I%	B%	TE%
Acid phosphatase	4.5	2.1 (9.0)	9.5 (16.4)
Alanine aminotransferase	13.6	13.6	36.0
Albumin	1.4 (1.8)	1.1 (2.8)	3.4 (5.8)
Alkaline phosphatase	3.4	6.4	12.0
Amylase	3.7	6.5 (7.4)	12.6 (13.5)
Aspartate aminotransferase	7.2	6.2	18.0
Bicarbonate	2.3 (4.9)	1.6 (4.6)	5.4 (12.7)
Bilirubins	11.3	9.8	28.4
Calcium	0.9 (1.5)	0.7 (1.8)	2.2 (4.3)
Chloride	0.7 (1.0)	0.5 (1.4)	1.7 (3.0)
Cholesterol	2.7	4.1	8.6
Creatine kinase	20.7	19.8	54.0
Creatinine	2.2	2.8 (4.4)	6.4 (8.0)
Digoxin	3.8 (4.7)	3.9	10.2 (11.7)
Gamma-Glutamyl transferase	7.4	21.8	34.0
Glucose	2.2	1.9 (4.4)	5.5 (8.0)
IgA	2.2 (3.8)	12.5	16.1 (18.8)
IgG	1.9 (3.7)	5.0	8.1 (11.1)
IgM	2.3 (5.4)	12.7	16.5 (21.6)
Iron	15.9	8.9	35.1
Lactate dehydrogenase	3.9	4.1 (7.8)	10.5 (14.2)
Lithium	3.6	4.2	10.1
Magnesium	1.1 (2.6)	1.6 (2.2)	3.4 (6.5)
Phosphate	4.0	3.1 (8.0)	9.7 (14.6)
Potassium	2.4	1.6 (4.8)	5.6 (8.4)
Proteins	1.4	1.5 (2.8)	3.8 (5.1)
Thyroxine (total)	3.4 (4.1)	4.1 (6.8)	9.7 (13.6)
Transferrin	2.4 (4.0)	2.3 (4.8)	6.3 (11.4)
Triacylglycols	11.5	15.6	34.6
Triiodothyronine	4.0 (4.7)	5.5 (8.0)	12.1 (15.8)
Urate	4.2	4.0 (8.4)	10.9 (15.3)
Urea	6.3	5.3	15.7

Ref: Fraser, C.G. et al., Proposed quality specifications for imprecision and inaccuracy of analytical systems for Clinical chemistry, *Eur J Clin Clin Biochem*, 1992: **30**: 311-317

Este modelo presentado solo se usa para una situación estable del sistema de medición y no puede ser usado para planear o seleccionar un sistema de control de calidad porque no incluye un componente derivado del control de calidad. O sea: *este modelo supone que no se necesita hacer control de calidad*. Lo cual no parece razonable dado que la evidencia acumulada en los últimos años, muestran que los sistemas de medición no son estables, incluso esto sería contradictorio con las recomendaciones de las autoridades sanitarias que recomiendan el uso de carta de control. Un modelo más realista necesita incluir la información de las cartas de control para testear los procedimientos habituales y eso se logra con el modelo siguiente:

25.5.1 El error en sistemas no estables.

En el supuesto de que el sistema de medición clínica no tenga un comportamiento estable, se debería incluir la información del control de calidad en la ecuación anterior. Agregando dos términos más a la ecuación, de manera de usar la capacidad de detección del sistema de control para los errores sistemáticos ΔES y casuales ΔRE , como sigue:

$$TE \geq ES + \Delta ES \cdot DS + 1,65 \cdot DS \cdot \Delta RE \quad (95\%)$$

El Control de Calidad puede contribuir a la variabilidad total que tiene un test cuando trabaja en condiciones no estables, porque tiene una cierta capacidad de detección del error., por ejemplo: el tamaño de la muestra que se necesita para poder detectarlos. Este tamaño puede variar de acuerdo a las reglas de control adoptadas en cada caso particular. Los nuevos valores en la ecuación anterior significan:

ΔES : La contribución aleatoria en el error sistemático (ES) que es detectable por un sistema de Control de Calidad

ΔPE : La contribución aleatoria en el error casual que es detectable por el Control de Calidad.

El factor $z = 1,65$ permite obtener un promedio máximo de defectuosos del 95% cuando se declara fuera de control al proceso. Este valor depende de las reglas adoptadas y del número de controles usados en el procedimiento.

Si se toman en cuenta otros factores el modelo es aún más complicado. Cuando se incorpora el concepto de componentes pre-analíticos y analíticos, se deben incluir las variabilidades debidas al tipo de espécimen o su condición, las ocasionadas por el muestreo y las debidas a la propia variación biológica del sujeto analizado. La idea general es sumar algebraicamente los diferentes tipos de errores sistemáticos, y para el error casual se calcula el valor de $z = 1,65$ multiplicado por la raíz cuadrada de la suma de las varianzas componentes del error casual.

$$TE \geq [ES + \Delta ES \cdot DS + ES_m] + 1,65 [(DS \cdot \Delta RE)^2 + (DS^2_{dentro}) + (DS^2_{entre})]^{1/2}$$

Donde ES_m : es el error sistemático debido al muestreo.

DS^2_{dentro} : el la varianza dentro de los sujetos por la variación biológica.

DS^2_{entre} : el la varianza entre los sujetos por la variación muestral.

Por ejemplo, si la variabilidad biológica de un paciente alrededor del valor verdadero homeostático de 200 mg/dL es $DS^2_{entre} = 6,5\%$ para el colesterol, entonces hay un 95% CI(174 ; 226). Lo cual significa que más de la mitad del intervalo de aceptación recomendado, se consume con la variabilidad biológica del paciente. Por lo tanto, el resto de las variabilidades se debe ajustar al intervalo remanente. La variabilidad biológica dentro del paciente $DS^2_{dentro} = 6,5\%$ es más grande que el desvío máximo admisible del 3%. Por lo tanto, la variabilidad biológica predomina sobre las demás y esto complica la interpretación clínica de los resultados. La variabilidad combinada del los componentes biológicos y analíticos se espera que sea del 7,2%, lo cual es un incremento muy pequeño sobre la parte biológica del 6,5% esperada.

Tabla 25.1: Especificaciones deseables europea de la variabilidad clínica

Analito	Variación Biológica		Especificación deseable		
	CVw	CVg	I(%)	B(%)	TE(%)
11-Desoxycortisol	21.3	31.5	10.7	9.5	27.1
17-Hydroxyprogesterone	19.6	52.4	9.8	14.0	30.2
5'Nucleotidase	23.2	19.9	11.6	7.6	26.8
5'-Hidroxiindolacetate, concentration, 24 h	20.3	33.2	10.2	9.7	26.5
a1-Acid Glycoprotein	11.3	24.9	5.7	6.8	16.2
a1-Antichymotrypsin	13.5	18.3	6.8	5.7	16.8
a1-Antitrypsin	5.9	16.3	3.0	4.3	9.2
a1-Globulins	11.4	22.6	5.7	6.3	15.7
a1-Microglobulin, concentration, first morning	33.0	58.0	16.5	16.7	43.9
a2-Antiplasmin	6.2	---	3.1	---	---
a2-Globulins	10.3	12.7	5.2	4.1	12.6
a2-Macroglobulin	3.4	18.7	1.7	4.8	7.6
a2-Microglobulin output, overnight	29.0	32.0	14.5	10.8	34.7
a-Amylase	9.5	29.8	4.8	7.8	15.7
a-Amylase (pancreatic)	11.7	29.9	5.9	8.0	17.7
a-Amylase concentration, random	94.0	46.0	47.0	26.2	103.7
a-Carotene	35.8	---	17.9	---	---
a-Tocopherol	13.8	13.3	6.9	4.8	16.2
Acid phosphatase	8.9	8.0	4.5	3.0	10.3
Acid phosphatase tartrate resistant	10.8	13.3	5.4	4.3	13.2
Acid phosphatase activity, prostatic (PAP)	33.8	---	16.9	---	---
Activated partial thromboplastine time	2.7	8.6	1.4	2.3	4.5
Adenosine deaminase	11.7	25.5	5.9	7.0	16.7
Alanine aminopeptidase	4.1	---	2.1	---	---
Alanine aminotransferase	24.3	41.6	12.2	12.0	32.1
Albumin	3.1	4.2	1.6	1.3	3.9
Albumin, concentration, first morning	36.0	55.0	18.0	16.4	46.1
Aldosterone	29.4	40.1	14.7	12.4	36.7
Aldosterone, concentration	32.6	39.0	16.3	12.7	39.6
Alkaline phosphatase	6.4	24.8	3.2	6.4	11.7
Alkaline phosphatase, bone isoform	6.6	35.6	3.3	9.1	14.5
Alkaline phosphatase, placental	19.1	---	9.6	---	---
Ammonia, output	24.7	27.3	12.4	9.2	29.6
Androstendione	11.5	51.1	5.8	13.1	22.6
Angiotensin converting enzyme	12.5	27.7	6.3	7.6	17.9
Antithrombin III	5.2	15.3	2.6	4.0	8.3
Apolipoprotein A1	6.5	13.4	3.3	3.7	9.1
Apolipoprotein B	6.9	22.8	3.5	6.0	11.6
Ascorbic Acid	25.8	22.9	12.9	8.6	29.9
Aspartate aminotransferase	11.9	17.9	6.0	5.4	15.2
β2-Microglobulin	5.9	15.5	3.0	4.1	9.0
Basophile, count	28.0	54.8	14.0	15.4	38.5
β-Carotene	36.0	39.0	18.0	13.3	43.0
β-Criptoxantin	36.7	---	18.4	---	---
β-Globulins	10.1	9.1	5.1	3.4	11.7
Bilirubin total	25.6	30.5	12.8	10.0	31.1
Bilirubin conjugated	36.8	43.2	18.4	14.2	44.5
CA 125 antigen	13.6	46.5	6.8	12.1	23.3
CA 15.3 antigen	5.7	42.9	2.9	10.8	15.5

Tabla 25.1: Especificaciones deseables europea de la variabilidad clínica

Analito	Variación Biológica		Especificación deseable		
	CVw	CVg	I(%)	B(%)	TE(%)
CA 19.9 antigen	24.5	93.0	12.3	24.0	44.3
CA 549 antigen	9.1	33.4	4.6	8.7	16.2
Calcium	1.9	2.8	1.0	0.8	2.4
Calcium, concentration	27.5	36.6	13.8	11.4	34.1
Calcium, output	26.2	27.0	13.1	9.4	31.0
Carbohydrate deficient transferrin	7.1	38.7	3.6	9.8	15.7
Carcinoembryonic antigen	9.3	55.6	4.7	14.1	21.8
Ceruloplasmin	5.7	11.1	2.9	3.1	7.8
Chloride	1.2	1.5	0.6	0.5	1.5
Cholesterol	6.0	15.2	3.0	4.1	9.0
Cholinesterase	7.0	10.4	3.5	3.1	8.9
Cholinesterase, catalitic activity	5.4	10.3	2.7	2.9	7.4
Cholinesterasa, immunoreactive	6.4	---	3.2	---	---
C3 complement	5.2	15.6	2.6	4.1	8.4
C4 complement	8.9	33.4	4.5	8.6	16.0
Copper	8.0	19.0	4.0	5.2	11.8
Copper	4.9	13.6	2.5	3.6	7.7
Cortisol	20.9	45.6	10.5	12.5	29.8
C Peptide	9.3	13.3	4.7	4.1	11.7
C-Propeptide type I procollagen	8.2	17.6	4.1	4.9	11.6
C-Reactive protein	52.6	84.4	26.3	24.9	68.3
C-Telopeptide type I collagen/creatinine, first urine	35.1	---	17.6	---	---
C-Telopeptide type I procollagen	8.0	28.8	4.0	7.5	14.1
Creatine kinase	22.8	40.0	11.4	11.5	30.3
Creatine kinase MB, %	6.9	42.8	3.5	10.8	16.5
Creatine kinase MB, activity	19.7	24.3	9.9	7.8	24.1
Creatine kinase MB, mass	18.4	61.2	9.2	16.0	31.2
Creatinine	4.3	12.9	2.2	3.4	6.9
Creatinine clearance	13.6	13.5	6.8	4.8	16.0
Creatinine, concentration	24.0	24.5	12.0	8.6	28.4
Creatinine, output	11.0	23.0	5.5	6.4	15.4
Cystein	5.9	12.3	3.0	3.4	8.3
Dehydroepiandrosterone sulfate	3.4	30.0	1.7	7.5	10.4
Deoxypyridinoline/creatinine	14.7	15.1	7.4	5.3	17.4
Dipeptidyl-peptidase IV	8.2	14.5	4.1	4.2	10.9
Elastase-PI	13.6	16.4	6.8	5.3	16.5
Eosinophils, count	21.0	76.4	10.5	19.8	37.1
Epinephrine	25.3	---	12.7	---	---
Epinephrine	48.3	---	24.2	---	---
Erythrocytes, count	3.2	6.1	1.6	1.7	4.4
Estradiol	30.4	---	15.2	---	---
Estradiol	22.6	24.4	11.3	8.3	27.0
Factor V coagulation	3.6	---	1.8	---	---
Factor VII coagulation	6.8	19.4	3.4	5.1	10.7
Factor VIII coagulation	4.8	19.1	2.4	4.9	8.9
Factor X coagulation	5.9	---	3.0	---	---
Ferritine	14.9	13.5	7.5	5.0	17.3
Fibrinogen	10.7	15.8	5.4	4.8	13.6

Tabla 25.1: Especificaciones deseables europea de la variabilidad clínica

Analito	Variación Biológica		Especificación deseable		
	CVw	CVg	I(%)	B(%)	TE(%)
Follicle stimulating hormone	10.1	32.0	5.1	8.4	16.7
Free estradiol	22.8	---	11.4	---	---
Free estradiol	38.6	---	19.3	---	---
Free testosterone	9.3	---	4.7	---	---
Free testosterone	51.7	---	25.9	---	---
Free thyroxine (F14)	7.6	12.2	3.8	3.6	9.9
Free Triiodothyronine (FT3)	7.9	---	4.0	---	---
Fructosamine	3.4	5.9	1.7	1.7	4.5
γ -Globulins	14.6	12.3	7.3	4.8	16.8
γ -glutamyltransferase	13.8	41.0	6.9	10.8	22.2
Globulins, total	5.5	12.9	2.8	3.5	8.0
Glucose	4.9	7.7	2.5	2.3	6.3
Glucose-6-phosphate-1-dehydrogenase	32.8	31.8	16.4	11.4	38.5
Glutathione peroxidase	7.2	21.7	3.6	5.7	11.7
Glycated albumin	5.2	10.3	2.6	2.9	7.2
Glycated total protein	0.9	11.6	0.5	2.9	3.7
Glycohemoglobin	5.6	---	2.8	---	---
Haptoglobin	20.4	36.4	10.2	10.4	27.3
Hematocrit	2.8	6.4	1.4	1.7	4.1
Hemoglobin	2.8	6.6	1.4	1.8	4.1
HDL cholesterol	7.1	19.7	3.6	5.2	11.1
HDL 1 cholesterol	5.5	27.2	2.8	6.9	11.5
HDL 2 cholesterol	15.7	40.7	7.9	10.9	23.9
HDL 3 cholesterol	7.0	14.3	3.5	4.0	9.8
Hydroibutyrate dehydrogenase	8.8	---	4.4	---	---
Hydroxiprolin/minute, night urine	36.1	38.8	18.1	13.2	43.0
Homocysteine	7.7	29.9	3.9	7.7	14.1
Immunoglobulin A	5.0	36.8	2.5	9.3	13.4
Immunoglobulin G	4.5	16.5	2.3	4.3	8.0
Immunoglobulin M	5.9	47.3	3.0	11.9	16.8
Immunoglobulins K chain	4.8	15.3	2.4	4.0	8.0
Immunoglobulins λ chain	4.8	18.0	2.4	4.7	8.6
Insulin	21.1	58.3	10.6	15.5	32.9
Interferon receptor	14.0	20.0	7.0	6.1	17.7
Iron	26.5	23.2	13.3	8.8	30.7
Lactate dehydrogenase 1 isoform	6.3	10.2	3.2	3.0	8.2
Lactate dehydrogenase 2 isoform	4.9	4.3	2.5	1.6	5.7
Lactate dehydrogenase 3 isoform	4.8	5.5	2.4	1.8	5.8
Lactate dehydrogenase 4 isoform	9.4	9.0	4.7	3.3	11.0
Lactate dehydrogenase 5 isoform	12.4	13.4	6.2	4.6	14.8
Lactate	27.2	16.7	13.6	8.0	30.4
Lactate dehydrogenase	6.6	14.7	4.3	4.3	11.4
Lactoferrin	11.8	23.7	5.9	6.6	16.4
Leukocytes, count	10.9	19.6	5.5	5.6	14.6
LDL Cholesterol	8.3	25.7	4.2	6.8	13.6
LDL Cholesterol (direct)	6.5	---	3.3	---	---
LDL receptor mRNA	21.5	13.6	10.8	6.4	24.1
Lipase	23.1	33.1	11.6	10.1	29.1
Lipoprotein (a)	8.5	85.8	4.3	21.6	28.8

Tabla 25.1: Especificaciones deseables europea de la variabilidad clínica

Analito	Variación Biológica		Especificación deseable		
	CVw	CVg	I(%)	B(%)	TE(%)
Lutein	23.7	---	11.9	---	---
Luteinizing hormone	14.5	27.8	7.3	7.8	19.8
Lycopene	43.1	---	21.6	---	---
Magnesium	5.6	11.3	2.8	3.2	7.8
Magnesium	18.3	16.4	9.2	6.1	21.2
Magnesium	3.6	6.4	1.8	1.8	4.8
Magnesium, concentration	45.4	37.4	22.7	14.7	52.2
Magnesium, output	38.3	37.6	19.2	13.4	45.0
Mucinous carcinoma-associated antigen	10.1	39.3	5.1	10.1	18.5
Mean corpuscular hemoglobin	1.6	5.2	0.8	1.4	2.7
Mean corpuscular hemoglobin concentration	1.7	2.8	0.9	0.8	2.2
Mean corpuscular volume	1.3	4.8	0.7	1.2	2.3
Mean platelet volume	4.3	8.1	2.2	2.3	5.8
Myoglobin	13.9	29.6	7.0	8.2	19.6
Monocytes, count	17.8	49.8	8.9	13.2	27.9
N-Acetyl Glucosaminidase, concentration	42.2	18.1	21.1	11.5	46.3
N-Acetyl Glucosaminidase, output	42.4	18.2	21.2	11.5	46.5
Neutrophiles, count	15.1	32.8	8.1	9.1	22.4
Nitrogen, output	13.9	24.2	7.0	7.0	18.4
Norepinephrine	9.5	---	4.8	---	---
Norepinephrine	19.5	---	9.8	---	---
N- Propeptide type I procollagen	7.4	---	3.7	---	---
N- Teloepptide type I collagen /creatin, 1 ^o urine	23.1	---	11.6	---	---
Osmolality	1.3	1.2	0.7	0.4	1.5
Osteocalcin	6.3	23.1	3.2	6.0	11.2
Oxalate, concentration	44.0	18.0	22.0	11.9	48.2
Oxalate, output	42.5	19.9	21.3	11.7	46.8
pCO ₂	4.8	5.3	2.4	1.8	5.7
pH [H ⁺]	3.5	2.0	1.8	1.0	3.9
pH (pH units)	0.2	---	0.1	---	---
Phosphate	8.5	9.4	4.3	3.2	10.2
Phosphate, concentration	26.4	26.5	13.2	9.4	31.1
Phosphate, output	18.0	22.6	9.0	7.2	22.1
Phospholipids	6.5	11.1	3.3	3.2	8.6
Potassium	13.6	13.4	6.8	4.8	16.0
Potassium	4.8	5.6	2.4	1.8	5.8
Potassium, concentration	27.1	23.2	13.6	8.9	31.3
Potassium, output	24.4	22.2	12.2	8.2	28.4
Plasminogen	7.7	---	3.9	---	---
Platelet, count	9.1	21.9	4.6	5.9	13.4
Platelet distribution wide	2.8	---	1.4	---	---
Plateletcrit	11.9	---	6.0	---	---
Prealbumin	10.9	19.1	5.5	5.5	14.5
Prolactin (men)	6.9	61.2	3.5	15.4	21.1
Prolyl endopeptidase	16.8	13.9	8.4	5.5	19.3
Protein C	5.8	55.2	2.9	13.9	18.7
Protein S	5.8	63.4	2.9	15.9	20.7
Protein, concentration	39.6	17.8	19.8	10.9	43.5

Tabla 25.1: Especificaciones deseables europea de la variabilidad clínica

Analito	Variación Biológica		Especificación deseable		
	CVw	CVg	I(%)	B(%)	TE(%)
Protein, output	35.5	23.7	17.8	10.7	40.0
Protein, total	2.7	4.0	1.4	1.2	3.4
Prostatic specific antigen (PSA)	14.0	72.4	7.0	18.4	30.0
Prothrombin time	4.0	6.8	2.0	2.0	5.3
Phosphate tubular reabsorption	2.7	3.3	1.4	1.1	3.3
Pyridinoline/creatinine, morning spot	8.7	17.6	4.4	4.9	12.1
Pyruvate	15.2	13.0	7.6	5.0	17.5
Red cell distribution wide	3.5	5.7	1.8	1.7	4.6
Retinol	14.8	18.3	7.4	5.9	18.1
Rheumatoid factor	8.5	24.5	4.3	6.5	13.5
SCC antigen	39.4	35.7	19.7	13.3	45.8
Selenium	12.0	14.0	6.0	4.6	14.5
Selenium	12.0	12.0	6.0	4.6	14.5
Sex hormone binding globulin (SHBG)	12.1	42.7	6.1	11.1	21.1
Sodium	1.8	12.4	0.9	3.1	4.6
Sodium	51.0	36.4	25.5	15.7	57.7
Sodium	0.7	1.0	0.4	0.3	0.9
Sodium, concentration	24.0	26.8	12.0	9.0	28.8
Sodium, output	28.7	16.7	14.4	8.3	32.0
Superoxide dismutase	17.1	10.5	8.6	5.0	19.1
Superoxide dismutase	12.3	4.9	6.2	3.3	13.5
T3-uptake	4.5	4.5	2.3	1.6	5.3
Testosterone	8.8	21.3	4.4	5.8	13.0
Testosterone	17.3	22.8	8.7	7.2	21.4
Testosterone	25.0	---	12.5	---	---
Thyroglobulin	13.0	25.0	6.5	7.0	17.8
Tissue polypeptide antigen (TPA)	28.7	40.4	14.4	12.4	36.1
Tissue polypeptide specific antigen (TPS)	36.1	108.0	18.1	28.5	58.3
Thyroid stimulating hormone (TSH)	19.7	27.2	9.9	8.4	24.6
Thyroxine binding globulin (TBG)	6.0	6.0	3.0	2.1	7.1
Thyroxine (T4)	6.0	12.1	3.0	3.4	8.3
Transferrin	3.0	4.3	1.5	1.3	3.8
Triglyceride	20.9	37.2	10.5	10.7	27.9
Total catecholamines, concentration	24.0	32.0	12.0	10.0	29.8
Triiodothyronine (T3)	8.7	14.4	4.4	4.2	11.4
Vanilmandelic Acid concentration	22.2	47.0	11.1	13.0	31.3
Urate	8.6	17.2	4.3	4.8	11.9
Urate, concentration	24.7	22.1	12.4	8.3	28.7
Urate, output	18.5	14.4	9.3	5.0	21.1
Urea	12.3	18.3	6.2	5.5	15.7
Urea, concentration	22.7	25.9	11.4	8.6	27.3
Urea, output	17.4	25.4	8.7	7.7	22.1
VLDL Colesterol	27.6	---	13.6	---	---
Von Willebrand factor	0.001	28.3	0.0005	7.1	7.4
Water	3.1	0.1	1.6	0.8	3.3
Zeaxanthine	34.7	---	17.4	---	---
Zinc	11.0	14.0	5.5	4.5	13.5
Zinc	9.3	9.4	4.7	3.3	11.0

Ref: Ricós, C., Alvarez V., Cava F. et al; Desiderable especifications for Total Error, Imprecisión and Bias, derived from biological variation; *Scand J Clin lab Invest*; **59**, 491-500, 1999

25.6 Requerimientos de CLIA para la calidad

Los requerimientos en USA para la calidad analítica de los Análisis Clínicos se expresan como los criterios para testear la capacidad de una prueba clínica (“proficiency testing criteria”) y decidir si tiene un comportamiento aceptable. En el Registro Federal de USA, fueron publicados el 28/02/92; tomo 57 (40): páginas 7002/186. Estas especificaciones pueden ser aplicadas en las cartas de control de Westgard como los *Requerimientos de calidad analítica*. Y se presentan en la tabla siguiente:

Tabla 25.2 Requerimientos de la CLIA

Test o Analito	Perfomance acceptable
Química de rutina	
Alanine aminotransferase	control \pm 20%
Albumin	control \pm 10%
Alkaline phosphatase	control \pm 30%
Amylase	control \pm 30%
Aspartate aminotransferase (AST)	control \pm 20%
Bilirubin, total	control \pm 0.4 mg/dL o \pm 20% (mayor)
Blood gas pO ₂	control \pm 3 SD
Blood gas pCO ₂	control \pm 5 mm Hg o \pm 8% (mayor)
Blood gas pH	control \pm 0.04
Calcium, total	control \pm 1.0 mg/dL
Chloride	control \pm 5%
Cholesterol, total	control \pm 10%
Cholesterol, high dens. lipoprotein	control \pm 30%
Creatine kinase	control \pm 30%
Creatine kinase isoenzymes	MB elevado (presente o ausente) o control \pm 3 SD Creatinine
Creatinine	control \pm 0.3 mg/dL o \pm 15% (mayor)
Glucose	control \pm 6 mg/dL o \pm 10% (mayor)
Iron, total	control \pm 20%
Lactate dehydrogenase (LDH)	control \pm 20%
LDH isoenzymes	LDH1/LDH2 (+ o -) o control \pm 30%
Magnesium	control \pm 25%
Potassium	control \pm 0.5 mmol/L
Sodium	control \pm 4 mmol/L
Total protein	control \pm 10%
Triglycerides	control \pm 25%
Urea Nitrogen	control \pm 2 mg/dL o \pm 9% (mayor)
Uric acid	control \pm 17%
Toxicología	
Theophylline	control \pm 25%
Tobramycin	control \pm 25%
Procainamide (and metabolite)	control \pm 25%
Quinidine	control \pm 25%
Valproic acid	control \pm 25%

Tabla 25.2 Requerimientos de la CLIA (continuación)

Hematología	
Cell identification	90% o mayor consenso en identificación
White cell differentiation	control \pm 3 SD (% de diferentes tipos de células blancas)
Erythrocyte count	control \pm 6%
Hematocrit	control \pm 6%
Hemoglobin	control \pm 7%
Leukocyte count	control \pm 15%
Platelet count	control \pm 25%
Fibrinogen	control \pm 20%
Partial thromboplastin time	control \pm 15%
Prothrombin time	control \pm 15%
Endocrinología	
Cortisol	control \pm 25%
Free thyroxine	control \pm 3 SD
Human chorionic gonadotropin	control \pm 3 SD o (positivo o negativo)
T3 uptake	control \pm 3 SD por método
Triiodothyronine	control \pm 3 SD
Thyroid stimulating hormone	control \pm 3 SD
Thyroxine	control \pm 20% o 1.0 mcg/dL (mayor)
Inmunología General	
Alpha-1 antitrypsin	control \pm 3 SD
Alpha-fetoprotein	control \pm 3 SD
Antinuclear antibody	control \pm 2 dilución o (pos. o neg.)
Antistreptolysin O	control \pm 2 dilución o (pos. o neg.)
Anti-Human Immunodeficiency virus	Reacción o no reactiva
Complement C3	control \pm 3 SD
Complement C4	control \pm 3 SD
Hepatitis (HBsAg, anti-HBc, HBeAg)	Reactiva (positivo) o no reactiva (negativo)
IgA	control \pm 3 SD
IgE	control \pm 3 SD
IgG	control \pm 25%
IgM	control \pm 3 SD
Infectious mononucleosis	control \pm 2 dilución o (pos. o neg.)
Rheumatoid factor	control \pm 2 dilución o (pos. o neg.)
Rubella	control \pm 2 dilución o (pos. o neg.)