



ESTADÍSTICA



ESTADÍSTICA

¿Qué se entiende por estadística?



ESTADÍSTICA

DEFINICIONES



Un sentido moderno de la palabra Estadística está relacionado con

"el desarrollo y aplicación de **métodos y técnicas** para recoger datos para que, mediante su **análisis**, se pueda llegar a la **interpretación** de un **fenómeno** ya sea biológico, social como también un proceso de producción, todos ellos a través de conceptos probabilísticos".

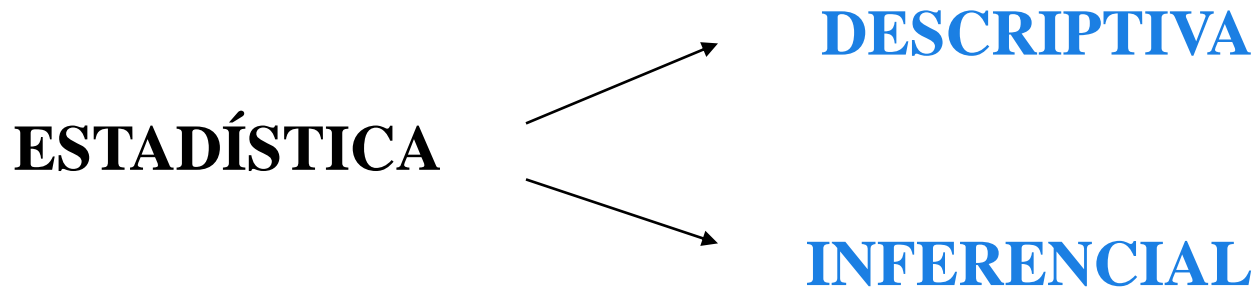


"Como **ciencia**, la estadística utiliza los resultados y la metodología del cálculo de Probabilidades y tiene por objeto la elaboración de métodos y modelos que permitan la **descripción**, el análisis, la interpretación y la **predicción** de fenómenos reales cuando estos son de carácter **aleatorio**".

¿Qué es la estadística?

Definición:

Diremos que la **ESTADÍSTICA** es una ciencia relacionada con el **METODO CIENTIFICO** en la colección y análisis de datos, muchas veces con el objeto de deducir o inferir conclusiones y tomar decisiones ante condiciones de incertidumbre.



Estadística Descriptiva:

Incluye la obtención, organización, resumen y presentación de los datos en forma convenientemente útil y de fácil comunicación además de calcular medidas con esta información.

Estadística Inferencial:

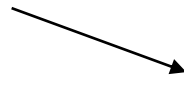
Se orienta a lograr generalizaciones y toma de decisiones, es decir, a partir de los datos de la *muestra* concluir sobre una *población*.

La estadística

ESTADÍSTICA



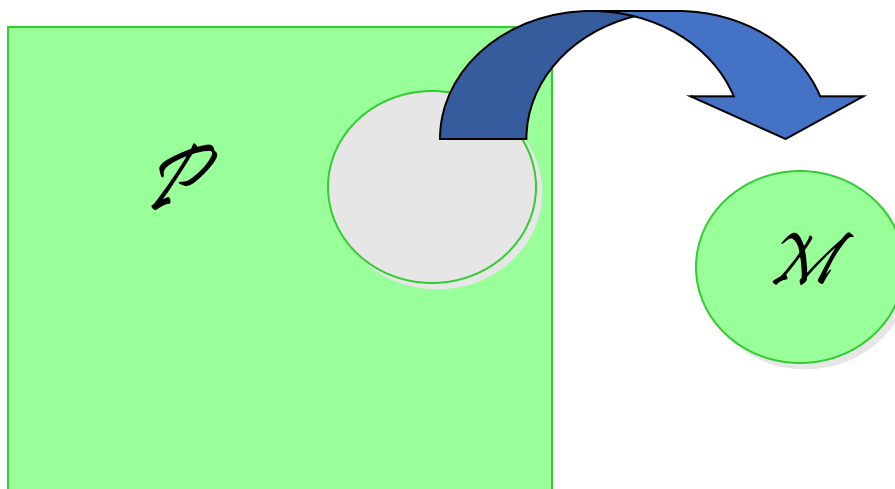
ANÁLISIS EXPLORATORIO



ANÁLISIS CONFIRMATORIO

Conceptos: Población - Muestra

El término **población** se refiere a los elementos del universo respecto a los cuales se quieren obtener conclusiones o tomar decisiones. Es el conjunto de todos los individuos u objetos que poseen alguna **característica común** observable (Población biológica). A cada elemento se le puede asociar una medición que bien puede ser numérica o cualitativa dependiendo de la característica que se quiera estudiar (Población estadística). La población puede ser finita, infinita numerable o infinita no numerable.



El término **muestra** se refiere al **subconjunto** seleccionado de la población de interés. Ésta selección se realiza mediante algún plan.

Población: es un conjunto cuyos elementos tienen una o más características comunes. Este conjunto puede estar formado por elementos animados o no, o por las mediciones que sobre esos elementos se realicen.

Muestra: es una parte o subconjunto de la población previamente especificada. Sus componentes deben tener la o las mismas características comunes que los de la población de donde ella proviene.

• Parámetro: Población

• Estimador: Muestra

MUESTREO

MUESTREO

- **Definición:**

es la técnica empleada para recoger una muestra a partir de una población o un subgrupo de la población.

- **Objetivo:**

estimar parámetros (por ej. medias, prevalencias, etc.), probar hipótesis, es decir inferir sobre la población, basándonos en la información contenida en la muestra.

Por qué el muestreo???

En algunas oportunidades,

- cuando la población es infinita,
- cuando la población es homogénea,
- cuando el proceso es destructivo,

se utiliza por necesidad el muestreo

Requisitos de la muestra

- Cualquier muestra para ser **representativa** del UNIVERSO: **CUALITATIVAMENTE** debe ser :
 - **Homogénea**: estar compuesta sólo por elementos que pertenecen al Universo previamente definido.
 - **Adecuada**: al incluir todas las variables esenciales de los elementos que existen en el universo.
 - **No viciada** : al presentar variaciones de determinadas variables en mas o menos la misma frecuencia que en el universo.

Y CUANTITATIVAMENTE debe ser lo suficientemente grande.

Clasificación de Métodos de Muestreo

1.- Por la Forma de Considerar un Evento

- Sin Reposición
- Con Reposición

2.- Por la Forma de Tomar la Muestra

- Opinático – Sin norma (no aleatorios)
- Aleatoria - Simple
 - Sistemática
 - Estratificada
 - Conglomerados...

3.- Por el número de Muestras

- Simple
- Múltiple

Muestreo Aleatorio

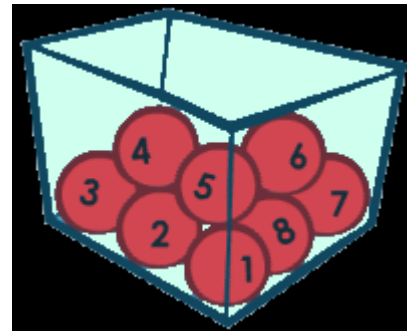
- Se dice que la muestra es aleatoria cuando la manera de selección de cada elemento de la población hace que cada uno tenga **igual oportunidad** de ser seleccionado.
- El método de selección es decisivo en las conclusiones que se pueden obtener de la muestra.

Muestreo Aleatorio Simple: M.A.S.

Es un método de selección de “n” unidades sacadas de “N”, de tal manera que cada una de las muestras del mismo tamaño tiene la misma probabilidad de ser escogida.

En la práctica, una muestra correspondiente a un M. A. S. es sacada unidad por unidad:

- Las unidades de la población son numerados del 1 al N.
- A continuación son seleccionados “n” números aleatorios entre 1 y N, ya sea de tablas o de una urna como en la lotería.
- Por último, se obtienen los “n” del total de “N”.



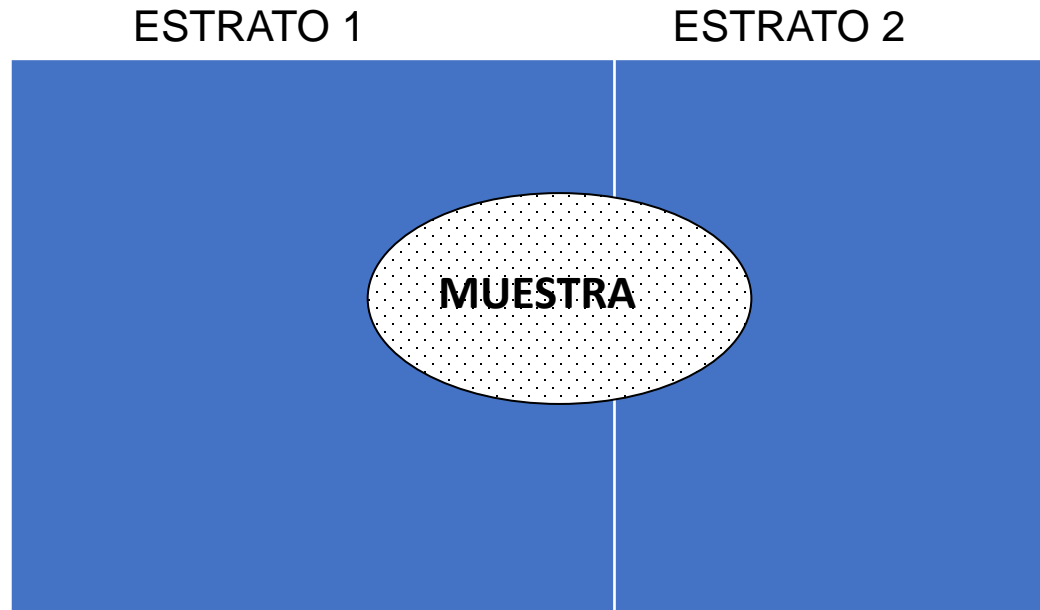
Muestreo Aleatorio Estratificado

Se emplea cuando la **población está agrupada** en pocos estratos, cada uno de ellos con muchos individuos. Consiste en sacar un M. A. S. de cada uno de los estratos.

Los estratos son “homogéneos” dentro y heterogéneos “entre”.

Los Estratos, a veces, son de diferente tamaño; la muestra, por consiguiente, para ser representativa debe contener elementos de cada estrato en forma proporcional a la composición de la población. (Esto se llama afijación proporcional, la que no siempre resulta ser la más conveniente por cuanto los costos de muestreo en cada uno de los estratos pueden ser distintos).

Muestreo Aleatorio Estratificado



Muestreo Aleatorio Sistemático

Se utiliza cuando **las unidades** de la población **están**, de algún modo, **totalmente ordenadas** (o se las puede ordenar) . Para seleccionar una muestra se aprovecha la ordenación de las unidades.

Para seleccionar una muestra de tamaño n

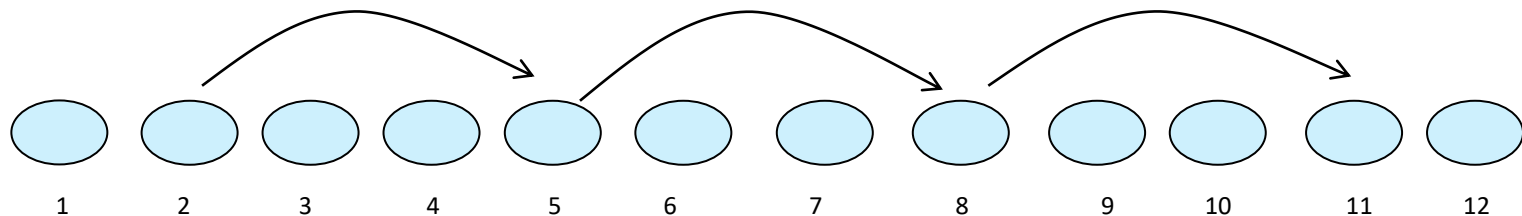
- se decide el tamaño de la muestra n y se calcula $K = N/n$
- se toma una unidad al azar de la primera subpoblación y
- de ahí en adelante cada k -ésima unidad.

Si n_1 es la unidad seleccionada de la primera población, entonces las siguientes observaciones serán $n_2 \leftarrow n_1 + K$, $n_3 \leftarrow n_2 + K \dots$

Muestreo Aleatorio Sistemático

Es preciso un listado de los elementos, que normalmente sigue algún criterio coincidente o no con el interés de investigación: orden alfabético, domicilio, momento o tiempo en que se produce un hecho.

- En primer lugar se selecciona el coeficiente de elevación o intervalo de muestreo ($K=N/n$) (#población/#muestra).(Ej.: $N=12$; $n=4 \Rightarrow K=3$)
- A continuación se elige al azar un número menor o igual al valor del intervalo de muestreo. La primera unidad muestral es el número elegido.(Ej.: $2 < 3$)
- De ahí en adelante los elementos se incluirán sistemáticamente. La segunda unidad muestral se determina sumando a ese primer número el valor del intervalo de muestreo y así sucesivamente hasta completar el tamaño de la muestra.



Muestreo por Conglomerados

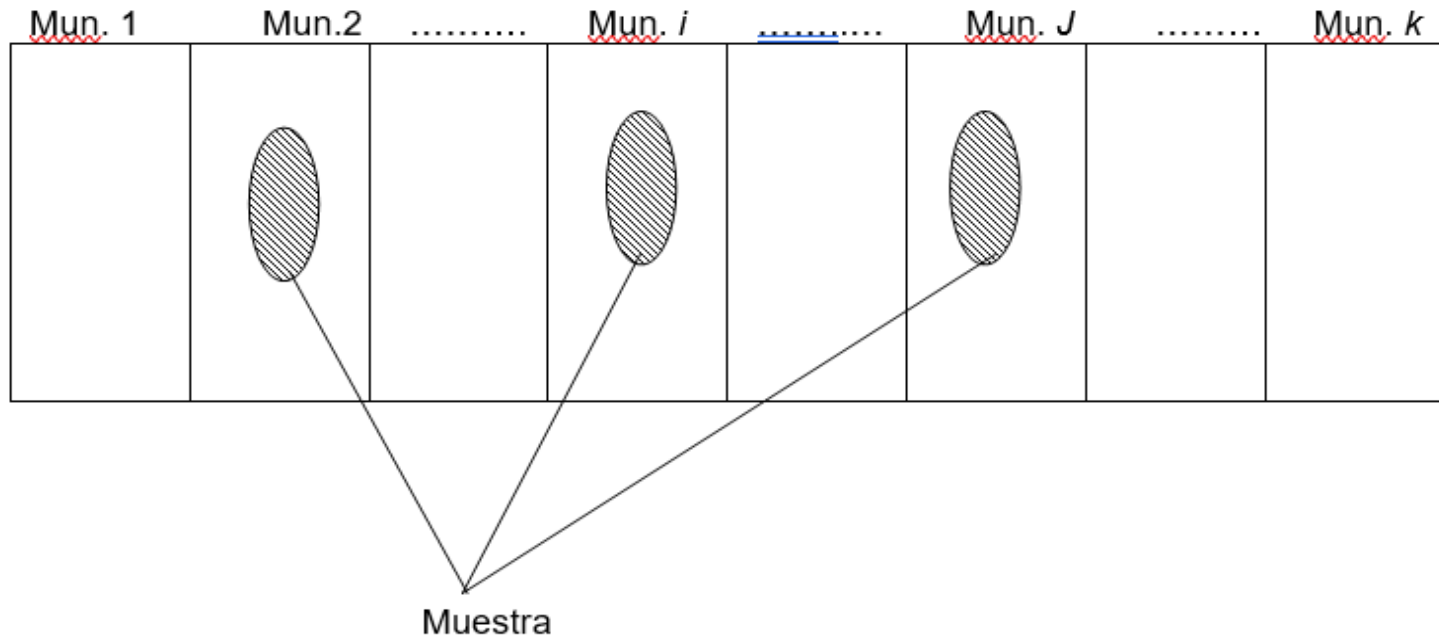
Se emplea cuando la población está dividida (o se la puede dividir) en grupos “semejantes en su composición”.

Los conglomerados son heterogéneos “dentro” y homogéneos “entre”.

Consiste en obtener una M. A. S. de algunos grupos (seleccionar una muestra de “K” grupos o conglomerados) y luego censar “dentro de” cada uno de estos.

Hay dos razones principales para la extensa aplicación de estos planes de muestreo: falta de una lista confiable de elementos en la población y consideraciones del tipo económico.

Muestreo por Conglomerados



Muestreo en varias etapas

La muestra se toma en varios pasos:

- **en el primero se selecciona la muestra de unidades primarias y**
- **en la segunda se selecciona una muestra de elementos a partir de cada unidad primaria escogida**
- **y ...**

¿QUÉ ES UN “EXPERIMENTO”?

- **Experimento:** es una acción mediante la cuál se obtiene un resultado y que implica la observación de éste.
- **Experimento aleatorio:** es aquel en que el resultado no se puede predecir con exactitud.
- **Experimento determinista:** es aquel en el que el resultado puede predecirse con total exactitud.

ESTADÍSTICA

DESCRIPTIVA

trabaja con

**Poblaciones
Muestras**

e incluye las siguientes etapas

- ◆ **R**ecopilación u obtención
- ◆ **O**rganización y reducción
- ◆ **P**resentación
- ◆ **A**nálisis e interpretaciones

a partir de las muestras se calculan
Estimadores

y con aporte de las

- ◆ **T**eoría de probabilidades
- ◆ **D**istribuciones de probabilidad
- ◆ **D**istribuciones de muestreo

a partir de lo informado por la muestra,
llevará a hacer dos tipos de generalización

INFERENCIAL

- ◆ **E**stimación
- ◆ **P**ruueba de hipótesis

sobre los **Parámetros(*)**

(*) También pueden ser no paramétricos

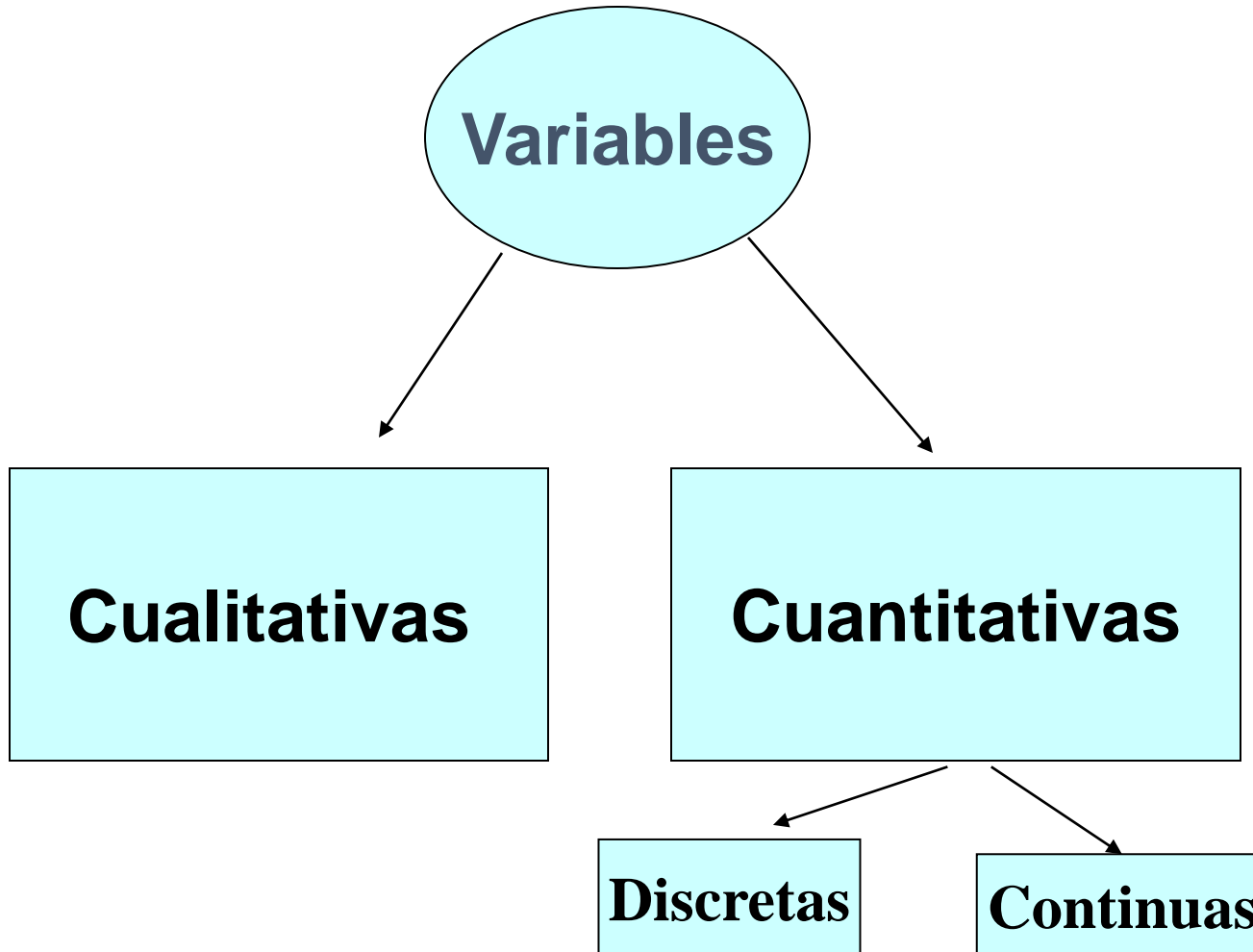
Variable

A cada característica de los elementos de una población se le llama **variable**.

En general son de naturaleza diversa. La consideración de estas diferencias es esencial para decidir el método de análisis estadístico adecuado.

Las expresiones: variable, variable aleatoria o variable al azar, son características que pueden cambiar su valor en las distintas unidades de observación.

Clasificación de variables por el tipo de resultado



Clasificación de variables

Las variables **cualitativas** son aquellas que se refieren a categorías o atributos de los elementos (unidades experimentales) estudiados.

- Son no numéricas.

Ejemplos: Estado de salud (enfermo-sano), sexo, color del cabello, grupo sanguíneo, grado tumoral.

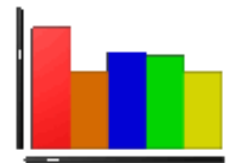
Las variables **cuantitativas** son aquellas cuyos datos son de tipo numérico.

- Se subdividen en:
- **Continuas:** número infinito no numerable de elementos. Tiene asociado el concepto de medida.

Ejemplo: presión arterial, edad, peso.

- **Discretas:** número finito o infinito numerable de elementos. Se asocia con el concepto de conteo.

Ejemplo: # de hijos por familia, # del calzado.



Según la escala de medición se clasifican en:

- **nominales**: tienen dos o más categorías y no hay orden entre ellas.

Ejemplos: Nacionalidad, raza, color de pelaje, marca de autos.

- **ordinales**: tienen varias categorías o valores y hay orden natural entre ellas.

Ejemplos: Nivel Socioeconómico, Grado en la Fuerzas Armadas, grado de mejoría.

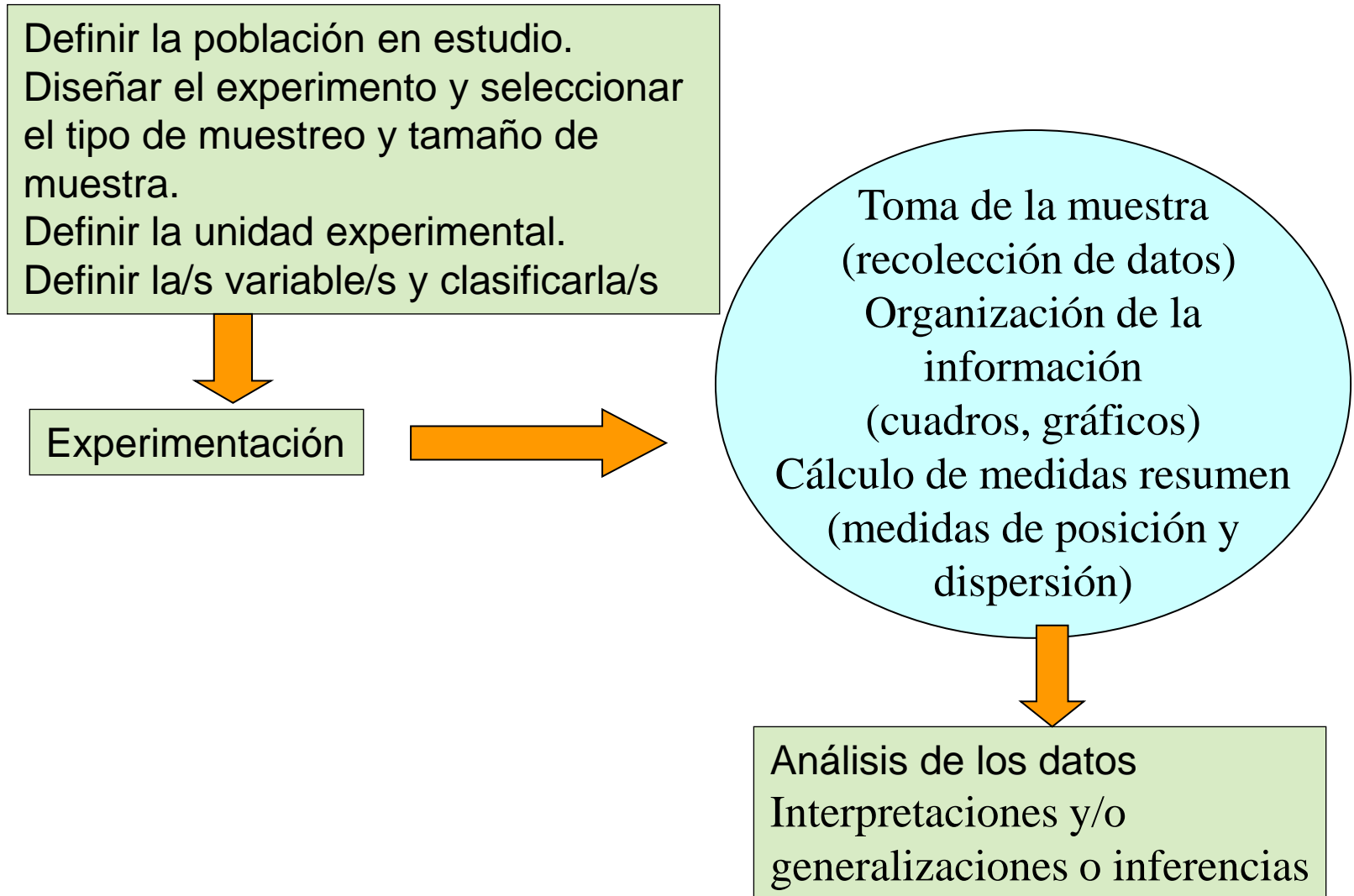
- **intervalo**: se pueden realizar operaciones de suma y resta. El cero es puesto por el hombre.

Ejemplos: temperatura ambiente, altitud de una ciudad.

- **razón**: se pueden realizar las cuatro operaciones básicas. El cero es real.

Ejemplos: altura de los árboles, peso de una persona.

Pasos del análisis de datos



Ejemplo

Se desea conocer el tiempo de uso del celular para buscar información. Para ese fin se seleccionó un grupo de 30 estudiantes universitarios lo más homogéneo posible en cuanto a la edad y la carrera que cursan. El aumento del uso del celular para buscar información depende del tipo de carrera.

Se solicita indicar:

POBLACIÓN DE ESTUDIO **Estudiantes universitarios**

b) VARIABLE EN ESTUDIO **Tiempo de uso del celular**

c) CLASIFICACION DE LA VARIABLE **Cuantitativa continua**

d) MUESTRA DE INDIVIDUOS **indicando el tamaño de la misma**

30 estudiantes universitarios de similar edad y carrera

e) MUESTRA DE OBSERVACIONES **no están los datos**

Los tiempos de uso del celular luego de un cierto lapso de tiempo en una determinada clase universitaria.

f) UNIDAD EXPERIMENTAL o DE OBSERVACION **1 estudiante universitario**

Estadística Descriptiva

- Suponemos un conjunto de datos obtenidos empíricamente a partir de una medición concreta de una o varias variables aleatorias en un conjunto de Unidades de Análisis de una Población o de una Muestra.

- **Objetivo:** Caracterizar (organizar, sintetizar y describir) los datos de forma que la información que se obtenga sea útil.

- **Herramientas:** Técnicas analíticas (cálculo de medidas estadísticas) y gráficas. El proceso de elaboración e interpretación de este conjunto de medidas y gráficos es lo que se llama **Estadística Descriptiva**.

Tablas de Frecuencias

Ejemplo 1: Supongamos que hemos preguntado a un conjunto de n alumnos cuantas faltas a clase han tenido el último mes

El resultado de la medición es el siguiente:

7 5 6 8 6 5 9 5 8 6 5 7 5 5 4 5 8 5 4 2 6 6 4 6 4 8 4 3 4 3 3 1 4 5 6 5 8 5 4 7 4 3 5 3 4 9 4 2 3 4 2 4 1 3 6 3 1 2 4 4 6 2 4 7 4 2 4 6 4 4 6 7 5 8 5 7 6 5 6 5 7 5 6 4 5 4 1 6 5 6 5 5 5 4 6 2 5 5 6 5 4 4 3 5 5 9 4 3 6 5 7 3 2 4 4 7 4 2 1 8 2 7 4 5 5 7 5 5 1 5 8 5 7 6 6 7 7 5 2 5 6 5 8 5 3 6 5 5
--

Tabla 1: *Conjunto original de datos*

Intenten responder a las siguientes preguntas:

- ¿Cuántos alumnos fueron encuestados?
- ¿Cuál fue la respuesta más frecuente?
- ¿Cuántos alumnos tienen, como máximo, 4 faltas en el último mes?

¿Qué hacer? La primera acción a realizar es ordenar los datos desde el que posee el valor más pequeño hasta el que cuenta con el valor mayor.

El resultado es:

7 5 6 8 6 5 9 5 8 6 5 7 5 5 4 5 8 5 4 2 6 6 4 6 4 8 4 3 4 3 3 1 4 5 6 5 8 5 4 7 4 3 5 3 4 9 4 2 3 4 2 4 1 3 6 3 1 2 4 4 6 2 4 7 4 2 4 6 4 4 6 7 5 8 5 7 6 5 6 5 7 5 6 4 5 4 1 6 5 6 5 5 5 4 6 2 5 5 6 5 4 4 3 5 5 9 4 3 6 5 7 3 2 4 4 7 4 2 1 8 2 7 4 5 5 7 5 5 1 5 8 5 7 6 6 7 7 5 2 5 6 5 8 5 3 6 5 5
--

Tabla 1: *Conjunto original de datos*



1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 4 5 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 9 9 9
--

Tabla 2: *Conjunto ordenado de datos*

Sin perder información se puede expresar:

1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4
4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5
5 5 5 5 5 5 5 5 5 5 5 6 7 7 7 7 7 7 7 7 7
7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 9 9 9
Tabla 2: <i>Conjunto ordenado de datos</i>



1 (6), 2 (11), 3 (12), 4 (30), 5 (40), 6 (25), 7 (14), 8 (9), 9 (3)
Tabla 3: <i>Conjunto ordenado de "valores" y "frecuencias"</i>

Para disponer la información de manera óptima, vamos a generar una tabla que tenga dos columnas. En la columna primera se presentarán los diferentes **valores** de la variable, que representaremos con la letra **X** mientras que en la segunda columna se dispondrán las **frecuencias**, que representaremos con la letra **f**. Obtenemos así la tabla 4:

1 (6), 2 (11), 3 (12),
4 (30), 5 (40), 6 (25),
7 (14), 8 (9), 9 (3)

Tabla 3: *Conjunto ordenado de "valores" y "frecuencias"*



X_i	f_i
1	6
2	11
3	12
4	30
5	40
6	25
7	14
8	9
9	3
Total	150

Tabla 4: *Tabla de frecuencias*

Como se lee en el título de la tabla 4, se trata de una *tabla de frecuencias*.

Ahora sí, la tabla de frecuencias nos permite responder a las preguntas planteadas con facilidad:

X_i	f_i
1	6
2	11
3	12
4	30
5	40
6	25
7	14
8	9
9	3
Total	150
Tabla 4: <i>Tabla de frecuencias</i>	

- ¿Cuántos alumnos fueron encuestados? Solución: **150**
- ¿Cuál fue la respuesta más frecuente? Solución: **5** (40 datos)
- ¿Cuántos alumnos tienen, como máximo, 4 faltas en el último mes?
Solución: **59** (6+11+12+30)

No todas las respuestas han exigido el mismo esfuerzo.

Para la tercera pregunta se necesitó hacer algunas operaciones.

Hemos sumado todas las frecuencias comprendidas entre el primer valor de la tabla y el valor que nos interesa, ambos inclusive.

Esta cantidad final recibe el nombre de *frecuencia acumulada*

X_i	f_i	F_i
1	6	6
2	11	17
3	12	29
4	30	59
5	40	99
6	25	124
7	14	138
8	9	147
9	3	150
Total	150	
<i>Tabla 5: Tabla de frecuencias</i>		

Organización de los Datos

Variable Numérica

Diagrama de Tallos y Hojas

Se ubican en una columna, en orden ascendente, todos los números que forman los datos una vez que se ha eliminado la última cifra (de las unidades o de los décimos) . Estos valores son los “tallos”.

Se separa con una línea vertical estos números y a la derecha de cada uno de ellos se localizan en filas las cifras de las unidades (o de sus décimos) de cada dato (hojas) que comienza con ese número.

En cada fila, las hojas se ordena también de menor a mayor.
(en un segundo paso)

Ejemplo: El número de inscriptos por día en una universidad fueron:

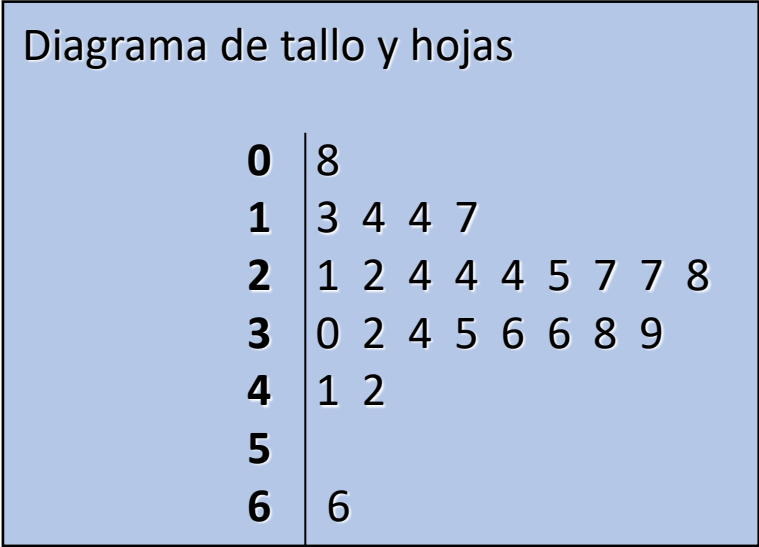
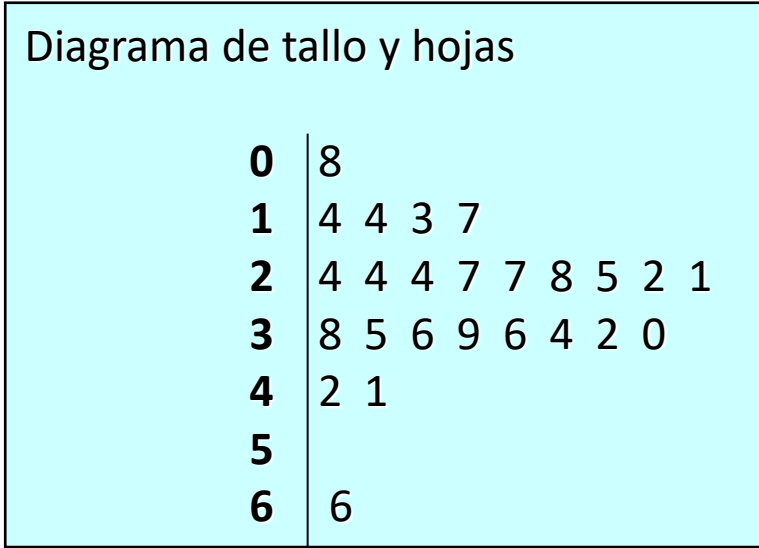
8- 14- 14- 24- 66- 24- 13- 24- 27- 38- 35- 36- 39- 42- 41- 36-
27- 28- 25- 17- 34- 32- 22- 21- 30

Construir un diagrama de tallo y hojas

Ordenar el diagrama de tallo y hojas



El número de inscriptos por día en una universidad fueron:
 8- 14- 14- 24- 66- 24- 13- 24- 27- 38- 35- 36- 39- 42-
 41-36- 27- 28- 25- 17- 34- 32- 22- 21- 30



Xmin:

Xmax:

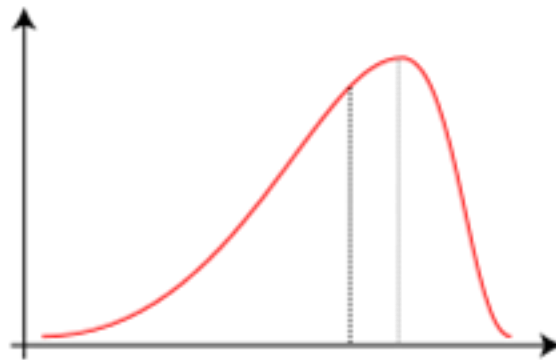
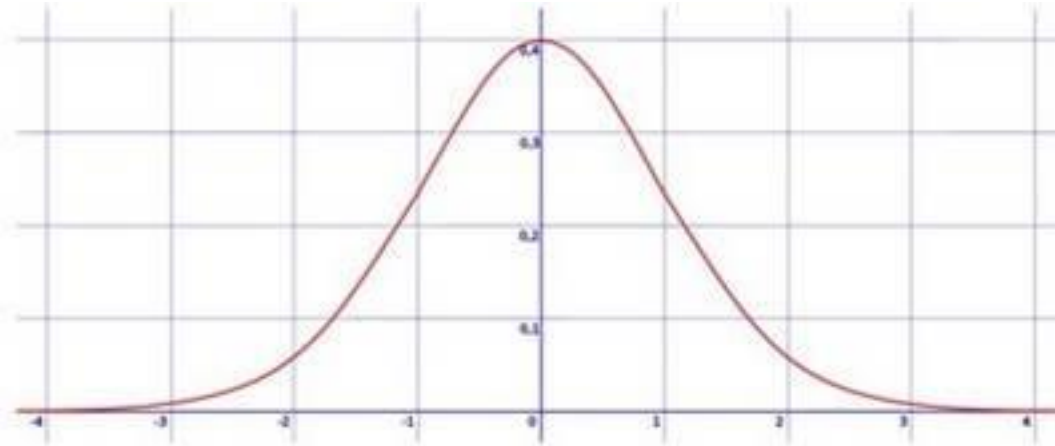
¿qué hay de raro?

¿cuántos datos hay? n=?

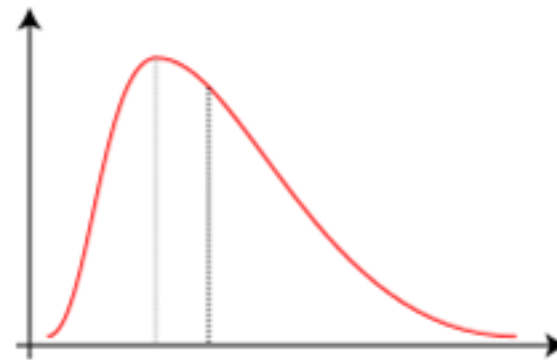
¿cuál es el valor que está en la mitad?

¿la distribución es simétrica?

SIMETRÍA Y ASIMETRÍA



Asimetría negativa



Asimetría positiva

Ejemplo: 52 – 57 – 62 – 62 – 62 – 62 – 65 – 66 – 67 – 68 – 68 – 68 –
69 – 69 – 69 – 71 – 71 – 72 – 72 – 73 – 74 – 74 – 75 – 75 – 75 – 76 – 77
– 78 – 79 – 79 – 79 – 80 – 80 – 82 – 83 – 85 – 88 – 89 – 91 – 93 – 97 –
97 – 97 – 98 – 99 – 101 – 104 – 105 – 105 – 109

Diagrama de tallo y hojas

5	2 7
6	2 2 2 2 5 6 7 8 8 8 9 9 9
7	1 1 2 2 3 4 4 5 5 5 6 7 8 9 9 9
8	0 0 2 3 5 8 9
9	1 3 7 7 7 8 9
10	1 4 5 5 9

Si consideramos que el diagrama original de tallo y hojas ha condensado mucho los datos, podemos ampliar el diagrama utilizando dos o mas renglones por cada uno o mas dígitos

Cuando un valor de tallo se indica dos veces, el primer valor corresponde a las hojas de 0-4 (.) y el segundo valor corresponde a los valores de 5-9 (*).

5.	2
5*	7
6.	2 2 2 2
6*	5 6 7 8 8 8 9 9 9
7.	1 1 2 2 3 4 4
7*	5 5 5 6 7 8 9 9 9
8.	0 0 2 3
8*	5 8 9
9.	1 3
9*	7 7 7 8 9
10.	1 4
10*	5 5 9

Unidades hojas

- Un dígito simple es usado para definir cada hoja.
- En el ejemplo precedente la unidad de hoja fue 1.
- Las unidades de las hojas pueden ser 100, 10, 1, 0.1, y así sucesivamente.
- Cuando la hoja no se muestra, se asume que es igual a 1.

Ejemplo: unidad de hoja = 0.1

Si tenemos los datos: 8.6 11.7 9.4 9.1 10.2 11.0 8.8

Un diagrama de tallo y hojas de estos datos será:

Unidad de Hoja = 0,1

8		6 8
9		1 4
10		2
11		0 7

Ejemplo 2: Imaginen ahora que hemos preguntado a 25 argentinos su provincia de nacimiento, obteniendo los siguientes resultados:

Salta, Córdoba, Tucumán, Neuquén, Neuquén, Córdoba, Tucumán, Misiones, Salta, San Juan, Chubut, Tucumán, Misiones, Salta, Córdoba, Misiones, San Juan, Córdoba, San Juan, Misiones, San Juan, Salta, Córdoba, Chubut, Córdoba

Con la información sobre las provincias de nacimiento, la tabla de frecuencias es:

Provincia	f_i
Córdoba	6
Chubut	2
Misiones	4
Neuquén	2
Salta	4
San Juan	4
Tucumán	3

¿Frecuencias acumuladas?

Quizá la respuesta de alguno de Ustedes ha sido ésta:

Provincia	f	F
Córdoba	6	6
Chubut	2	8
Misiones	4	12
Neuquén	2	14
Salta	4	18
San Juan	4	22
Tucumán	3	25

Tabla 6: Distribución por provincias

Si es así ... ¡Error!

¿Qué sentido tiene acumular frecuencias en el problema que se ha planteado sobre las provincias?. Ej.: ¿Qué significado tiene la cantidad 12 que acompaña al valor (Misiones)? Sólo se puede hacer una lectura: hay doce personas que *han nacido en Misiones o en menos que Misiones*. No podemos afirmar que Córdoba o Chubut sean *menos provincia de nacimiento* que Misiones.

La **diferencia** esencial entre el problema de las provincias de nacimiento y el de las respuestas a la escala de opinión, se encuentra en el **tipo de variable**. En el caso de las provincias, éstas **no** pueden ordenarse en función de ser más o ser menos "provincia de nacimiento" (se pueden ordenar según número de habitantes, extensión, altitud, etc. Pero no en función de ser más o ser menos provincia de nacimiento).

Luego, *la acumulación de frecuencias sólo puede hacerse si los valores de la variable se pueden ordenar*. Así, la respuesta correcta al problema es la primera que se presentó.

La tabla de frecuencias no termina donde la hemos dejado. Se puede añadir más información útil en la que basar respuestas para otras preguntas.

Por ejemplo ¿Cuántos estudiantes han respondido que tuvieron 5 faltas (valor 5)? Solución: 40.

Observemos la siguiente tabla para responder la misma pregunta.

X	f_i
1	200
2	170
3	120
4	60
5	40
6	60
7	120
8	170
9	200
Total	1140

Tabla 7: Nueva tabla de frecuencias

En la tabla 7 ha cambiado el conjunto de datos. Ahora son 1140, frente a los 150 anteriores. Una misma frecuencia, en este caso $f=40$, no tiene la misma interpretación en ambas tablas. ¿Qué ha cambiado?

La importancia relativa de la frecuencia. $f=40$ frente a $n=150$ es diferente de $f=40$ frente a $n=1140$. El valor 5 pasa de ser el más frecuente a ser el menos frecuente. La solución es expresar las frecuencias en términos relativos en vez de absolutos.

Así, añadimos la columna de las **frecuencias relativas** (h) que surgen de hacer la operación $h_i = f_i / n$.

Comparamos el resultado en las tablas afectadas al problema (4 y 7):

	Nuevos datos		Datos anteriores	
X	f_i	h_i	f_i	h_i
1	200	$200/1140=0,1754$	6	$6/150=0,0400$
2	170	$170/1140=0,1491$	11	$11/150=0,0733$
3	120	$120/1140=0,1053$	12	$12/150=0,0800$
4	60	0,0526	30	$30/150=0,2000$
5	40	$40/1140=0,0351$	40	$40/150=0,2667$
6	60	0,0526	25	$25/150=0,1667$
7	120	0,1053	14	$14/150=0,0933$
8	170	0,1491	9	$9/150=0,0600$
9	200	0,1754	3	$3/150=0,0200$
Total	1140	1,0000	150	1,0000

Tabla 8: *Comparación entre dos tablas de frecuencias*

Podemos completar también la tabla que se refiere a las provincias de nacimiento:

Provincia		f_i	h_i	%
Córdoba		6	$6/25=0,24$	24
Chubut		2	$2/25=0,08$	8
Misiones		4	$4/25=0,16$	16
Neuquén		2	$2/25=0,08$	8
Salta		4	$4/25=0,16$	16
San Juan		4	$4/25=0,16$	16
Tucumán		3	$3/25=0,12$	12

Tabla 10: Distribución por provincias

Y, por último...

Como sabemos, hay variables como “*nivel de estudios alcanzado*” que admite orden entre sus valores.

Luego, para completar la tabla de la variable número de faltas a clase han tenido el último mes bastará con acumular sus frecuencias.

X_i	f_i	h_i	h_i (%)	F_i	H_i	H_i (%)
1	6	$6/150=0,0400$	4,00	6	$6/150=0,0400$	4,00
2	11	$11/150=0,0733$	7,33	17	$17/150=0,1133$	11,33
3	12	$12/150=0,0800$	8,00	29	$29/150=0,1933$	19,33
4	30	$30/150=0,2000$	20,00	59	$59/150=0,3933$	39,33
5	40	$40/150=0,2667$	26,67	99	$99/150=0,6600$	66,00
6	25	$25/150=0,1667$	16,67	124	$124/150=0,8267$	82,67
7	14	$14/150=0,0933$	9,33	138	$138/150=0,9200$	92,00
8	9	$9/150=0,0600$	6,00	147	$147/150=0,9800$	98,00
9	3	$3/150=0,0200$	2,00	150	$150/150=1,0000$	100,00
Total	150	1,0000	100,00			

Tabla 11: *Tabla de frecuencias completa*

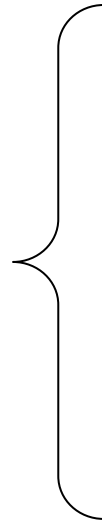
Ejemplo 3: Si tomamos la variable tiempo en realizar un servicio.
(cuantitativa)

Pocos valores de la variable serán repetidos por lo que se presentan los **datos** en forma agrupada en **Intervalos o clases**.

- Ventajas
- Desventajas

Tiempo en realizar 100 servicios

C
l
a
s
e
s



Tiempo (min)	No. de servicios
[60-63)	5
[63-66)	18
[66-69)	42
[69-72)	27
[72-75)	8
Total	100

Marca de Clase: Se denota por x'_i , y se determina por: $x'_i = \frac{LS_i + LI_i}{2}$

Donde LS: límite superior; LI: límite inferior

En el ej.: $x'_1 = (60 + 63)/2 = 61,5$; $x'_4 = (69 + 72)/2 = 70,5$

Podemos completar la tabla

Tiempo (min)	x'_i	f_i	h_i	F_i	H_i
[60-63)	61,5	5	$5/100=0,05$	5	$5/100=0,05$
[63-66)	64,5	18	$18/100=0,18$	23	$23/100=0,23$
[66-69)	67,5	42	$42/100=0,42$	65	$65/100=0,65$
[69-72)	70,5	27	$27/100=0,27$	92	$92/100=0,92$
[72-75)	73,5	8	$8/100=0,08$	100	$100/100=1,00$
Total		100	1,00		

Los intervalos son del tipo [LI -LS)

Para propósitos prácticos todos los datos se consideran que están representados por la marca de clase:

- La tabla dice que se tienen 5 servicios entre 60 y menos de 63 min. Para los cálculos se considera que los 5 datos del intervalo están representados por 61,5 min.
- el 65% de los servicios es inferior a 69 min
- el 42% de los servicios se encuentra entre [66 min; 69min]

¿Qué hemos visto?

TABLAS DE FRECUENCIAS

¿Para qué se construyen las tablas de frecuencias ?

1. ORDENAR
2. AGRUPAR
3. RESUMIR información

TIPOS DE FRECUENCIAS

a) Frecuencia Absoluta simples

Es el número de veces que se presenta un valor o categoría de una variable. Se representa por f_i .

b) Frecuencia Relativa Simple

La frecuencia relativa se puede expresar en términos de porcentaje o de proporción y se representan por h_i

En general, es claro e informativo indicar que parte del total de los datos representa cada frecuencia f_i . Así, si hay n datos en total, para una frecuencia absoluta f_i la frecuencia relativa se obtiene $h_i = f_i/n$

c) Frecuencia Absoluta Acumulada

Se representa por F_i .

donde $F_1 = f_1$

$$F_2 = f_1 + f_2$$

$$F_3 = f_1 + f_2 + f_3$$

.

$$F_k = f_1 + f_2 + f_3 + \dots + f_k = n$$

K es la cantidad de categorías

F_i es el número de veces que se presentan valores (o categorías) de una variable menores o iguales al i-ésimo

Nota: sólo tiene sentido cuando la variable es al menos ordinal.

d) Frecuencia Relativa Acumulada

Se representa por H_i

Donde: $H_1 = h_1$

$$H_2 = h_1 + h_2$$

$$H_3 = h_1 + h_2 + h_3$$

$$\cdot$$
$$H_k = h_1 + h_2 + h_3 + \dots + h_k = 1 \text{ ó } 100\%$$

H_i representa la proporción de valores acumulados hasta el valor de la variable i -ésimo

Nota: sólo tiene sentido cuando la variable es al menos ordinal.

Según el tipo de variables

Organización y resumen de Datos

Tabla de *Distribución de Frecuencias* tiene el formato siguiente:

Nominales (Categorías)

<i>Variable</i> <i>X</i>	<i>Frecuencias</i> <i>f_i</i>	<i>Porcentaje</i> <i>h_i%</i>
<i>Categoría 1</i>	<i>f₁</i>	<i>h₁%</i>
<i>Categoría 2</i>	<i>f₂</i>	<i>h₂%</i>
...
<i>Categoría k</i>	<i>f_k</i>	<i>h_k%</i>
<i>Total</i>	<i>n</i>	<i>100</i>

Si la variable ***categoría*** se mide con una escala ***ordinal***, entonces tiene sentido obtener las *frecuencias acumuladas*:

<i>Variable</i> <i>X</i>	<i>Frecuencias</i> <i>f_i</i>	<i>Porcentaje</i> <i>h_i%</i>	<i>Frec.</i> <i>Acumuladas F_i%</i>	<i>Porcentajes</i> <i>Acumulados H_i%</i>
<i>Categoría 1</i>	<i>f₁</i>	<i>h₁%</i>	<i>F₁%</i>	<i>H₁%</i>
<i>Categoría 2</i>	<i>f₂</i>	<i>h₂%</i>	<i>F₂%</i>	<i>H₂%</i>
...
<i>Categoría k</i>	<i>f_k</i>	<i>h_k%</i>	<i>F_k% = n</i>	<i>H_k% = 100%</i>
<i>Total</i>	<i>n</i>	<i>100</i>		

Según el tipo de variables

Organización de los Datos

Variable cuantitativa que toma pocos valores distintos y en general números enteros, conviene construir una Distribución de Frecuencias Simple:

<i>Variable</i> <i>X</i>	<i>Frecuencias</i> <i>f_i</i>	<i>Porcentajes</i>	<i>Frecuencias</i> <i>Acumuladas</i>	<i>Porcentajes</i> <i>Acumulados</i>
<i>x₁</i>	<i>f₁</i>	<i>h₁%</i>	<i>F₁ = f₁</i>	<i>H₁ = h₁</i>
<i>x₂</i>	<i>f₂</i>	<i>h₂%</i>	<i>F₂ = f₁ + f₂</i>	<i>H₂ = h₁ + h₂</i>
...
<i>x_k</i>	<i>f_k</i>	<i>h_k%</i>	<i>F_k = n</i>	<i>H_k = 100</i>
Total	n	100		

Variable cuantitativa que toma muchos valores distintos o es de naturaleza continua, conviene agrupar los datos en intervalos de clase y construir una Distribución de Frecuencias Agrupadas :

<i>Variable</i> <i>X</i>	<i>Frecuencias</i> <i>f_i</i>	<i>Porcentajes</i>	<i>Frecuencias</i> <i>Acumuladas</i>	<i>Porcentajes</i> <i>Acumulados</i>
<i>[LI₁ ; LS₁)</i>	<i>f₁</i>	<i>h₁%</i>	<i>F₁ = f₁</i>	<i>H₁ = h₁</i>
<i>[LI₂ ; LS₂)</i>	<i>f₂</i>	<i>h₂%</i>	<i>F₂ = f₁ + f₂</i>	<i>H₂ = h₁ + h₂</i>
...
<i>[LI_k ; LS_k)</i>	<i>f_k</i>	<i>h_k%</i>	<i>F_k = n</i>	<i>H_k = 100</i>
Total	n	100		

Representaciones Gráficas

Hemos visto que la tabla estadística resume los datos que disponemos de una población o muestra, de forma que ésta se puede analizar de una manera más sistemática y resumida. Para darnos cuenta de un sólo vistazo de las características de un grupo de datos resulta aún más esclarecedor el uso de gráficos y diagramas.



Variable Categórica

<i>caso</i>	<i>satisfacción</i>	<i>caso</i>	<i>satisfacción</i>
1	Compl. Satisfecho	16	Satisfecho
2	Satisfecho	17	Satisfecho
3	Satisfecho	18	Compl. Satisfecho
4	Compl. Satisfecho	19	Insatisfecho
5	Compl. Satisfecho	20	Satisfecho
6	Compl. Satisfecho	21	Satisfecho
7	Insatisfecho	22	Compl. Satisfecho
8	Satisfecho	23	Satisfecho
9	Compl. Satisfecho	24	Compl. Insatisfecho
10	Compl. Satisfecho	25	Satisfecho
11	Insatisfecho	26	Satisfecho
12	Compl. Satisfecho	27	Compl. Satisfecho
13	Compl. Satisfecho	28	Compl. Satisfecho
14	Compl. Insatisfecho	29	Satisfecho
15	Satisfecho	30	Compl. Satisfecho

<i>Satisfacción del Cliente</i>	<i>Nº de clientes</i>	<i>Porcentajes</i>	<i>Frecuencias Acumuladas</i>	<i>Porcentajes Acumulados</i>
<i>Completamente Satisfecho</i>	13	43,3	13	43,3
<i>Satisfecho</i>	12	40,0	25	83,3
<i>Insatisfecho</i>	3	10,0	28	93,3
<i>Completamente Insatisfecho</i>	2	6,7	30	100
<i>Total</i>	30	100,0		

Distribución de Frecuencias de la variable "Satisfacción del Cliente"

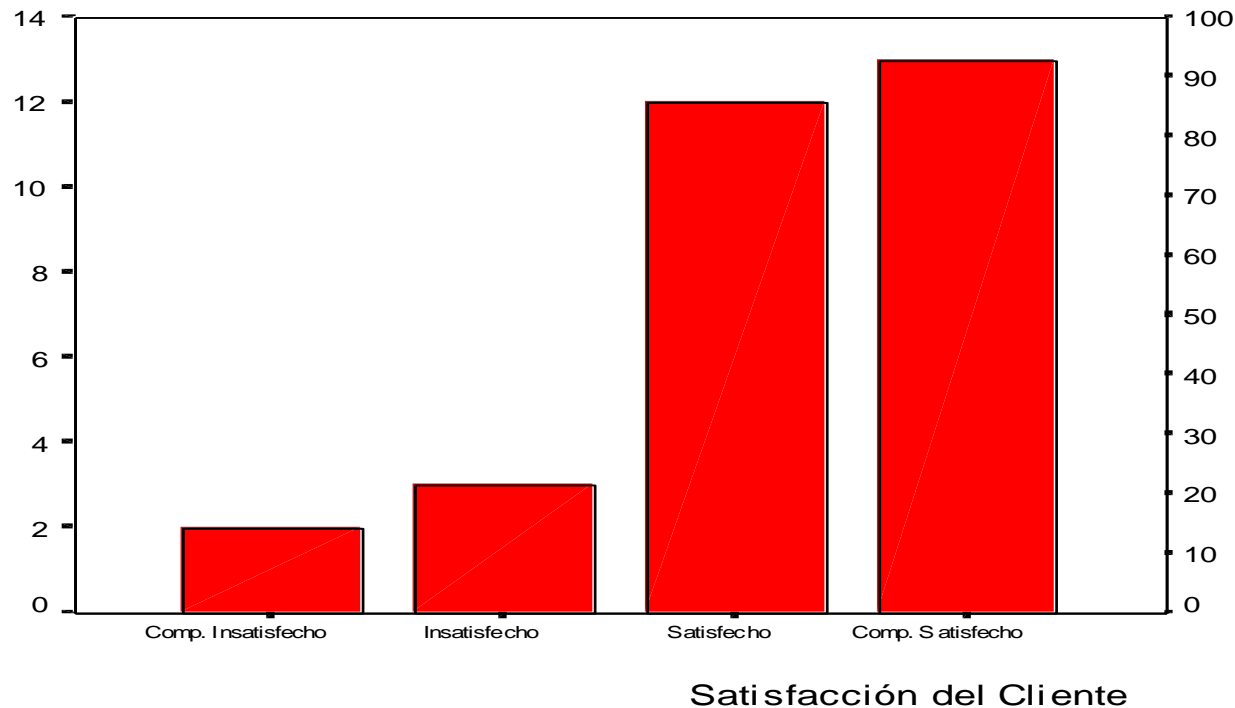
Representaciones Gráficas

Variable Categórica

Gráfico de Barras

En el eje horizontal se representan las categorías de la variable y sobre el eje vertical las frecuencias respectivas de cada categoría.

Gráfico de Barras de la variable “Satisfacción del Cliente”



Representaciones Gráficas

Variable Categórica

Gráfico de Pastel o torta (Pie Chart)

Consta de un círculo que se divide en tantos sectores circulares como categorías tenga la variable, cuyas áreas son proporcionales a las frecuencias correspondientes

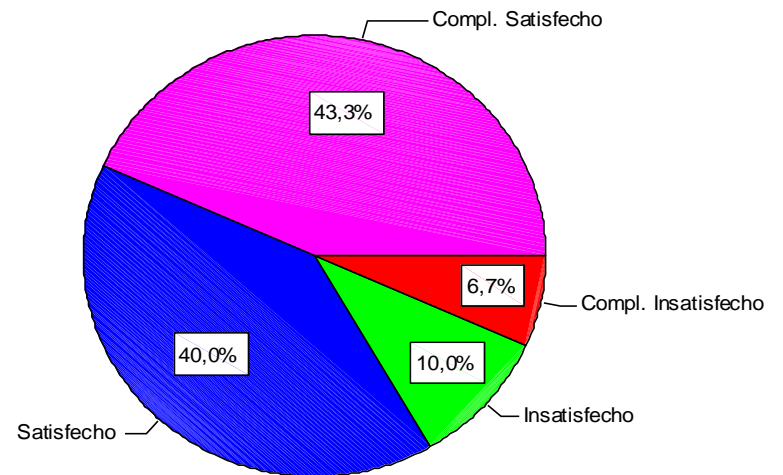
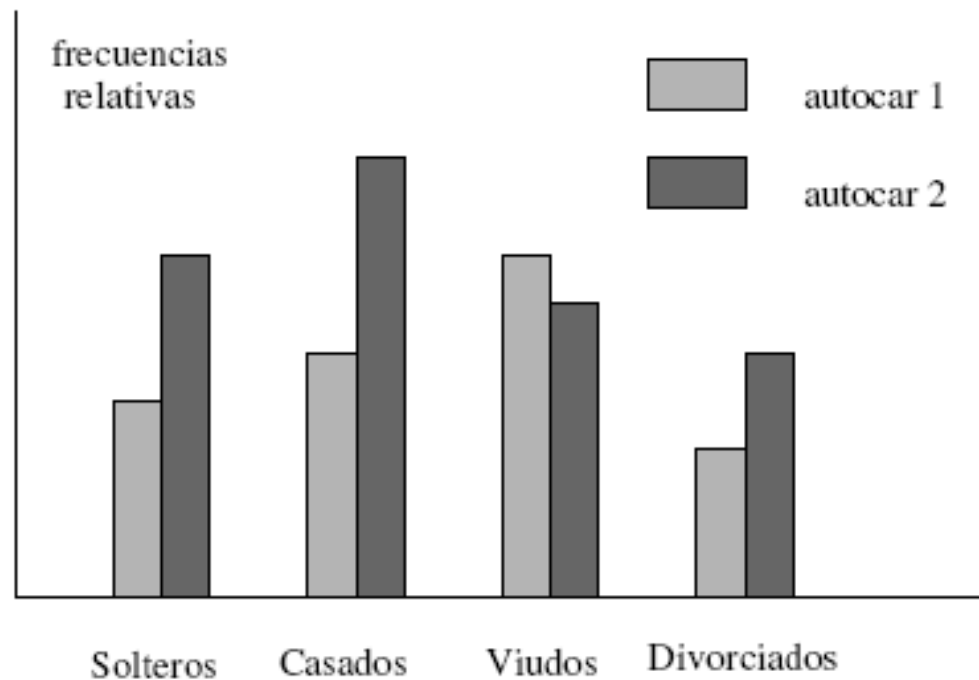


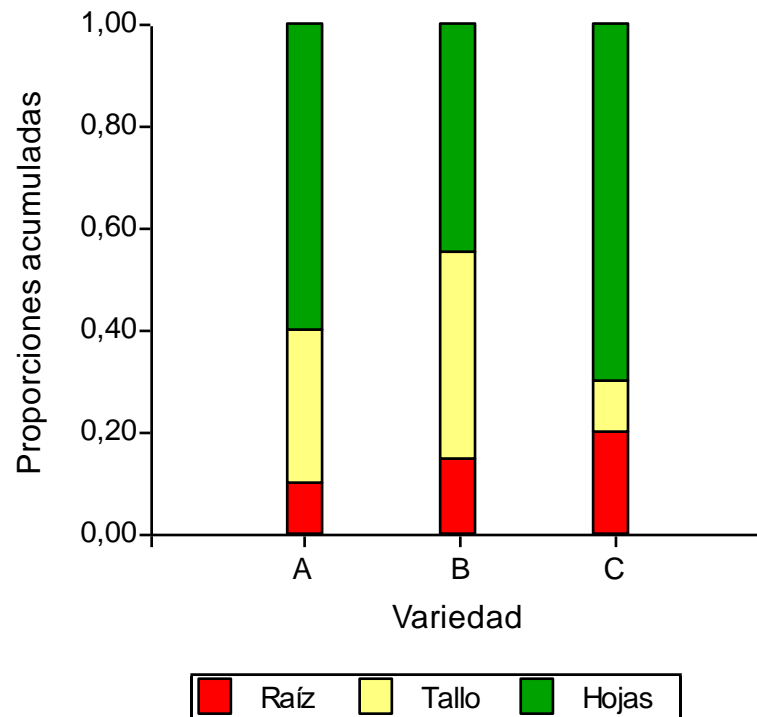
Gráfico de Pastel de la variable "Satisfacción del Cliente"

Otros gráficos que no veremos son:

Si, mediante el gráfico, se intenta comparar varias poblaciones entre sí, existen otras modalidades, como el [diagrama de barras comparativas](#). Cuando los tamaños de las dos poblaciones son diferentes, es conveniente utilizar las frecuencias relativas, ya que en otro caso podrían resultar engaños.



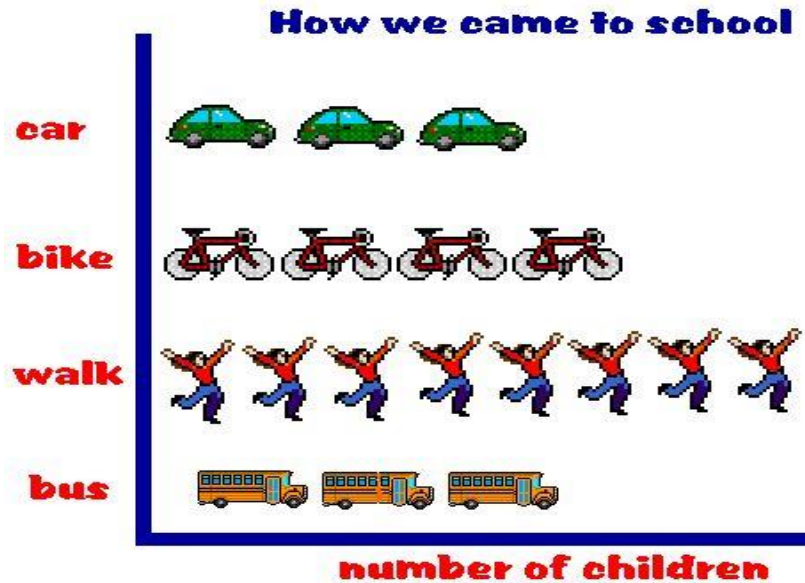
El **gráfico de barras apiladas** se aplica cuando se quiere representar comparativamente la contribución que distintas componentes hacen a un total. Por ejemplo, si el peso de una planta se divide en el peso de raíz, tallo y hojas, entonces, la contribución de cada una de estas particiones al peso total se puede representar como los segmentos de una barra con altura 1.



Gráficos para variables cualitativas

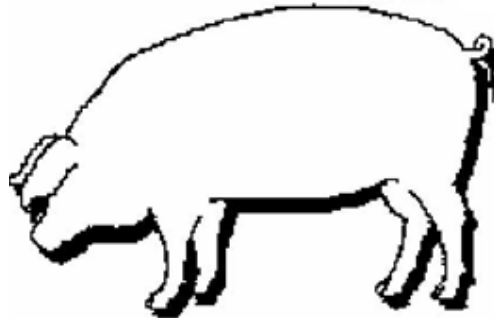
• Pictogramas

- Fáciles de entender.
- Cada modalidad debe ser proporcional a la frecuencia.

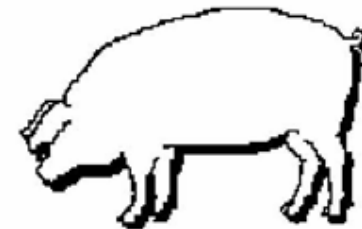


De los dos pictogramas, ¿cuál dirías que es incorrecto?

Botellas recogidas en un fin de semana



100 Kg
Ciudad A



50Kg
Ciudad B

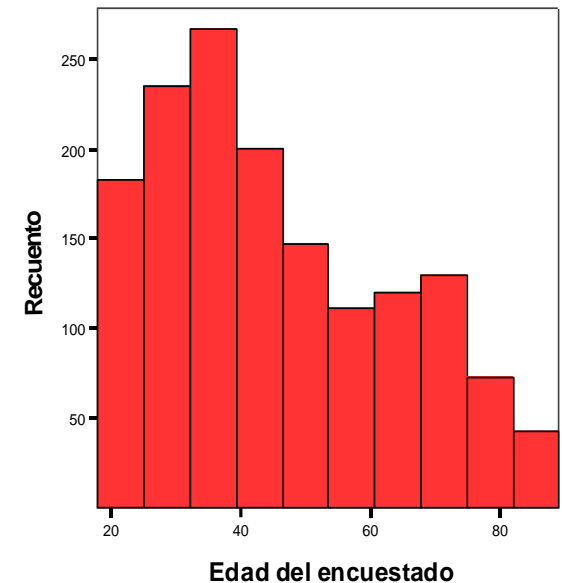
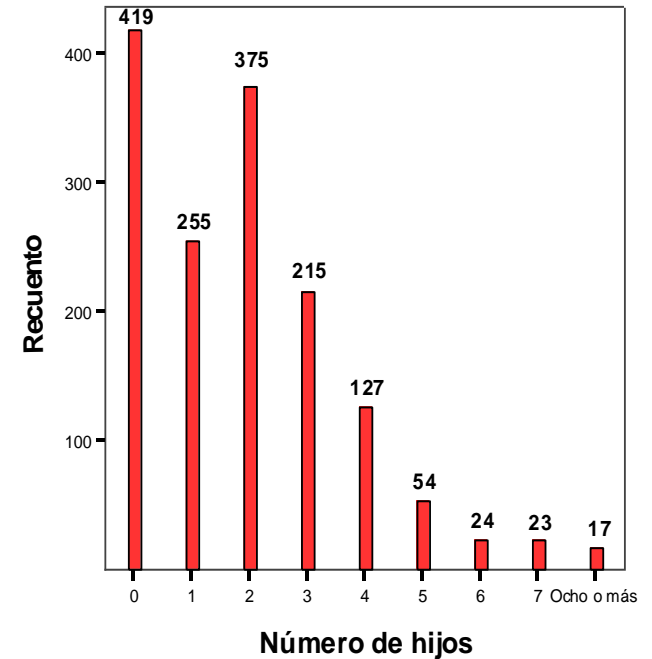
Gráficos para variables cuantitativas

Son diferentes en función de que las variables sean **discretas** o **continuas**.

Valen con frec. absolutas o relativas.

- **Diagramas bastones para v. discretas**
 - Se deja un hueco entre barras para indicar los valores que no son posibles

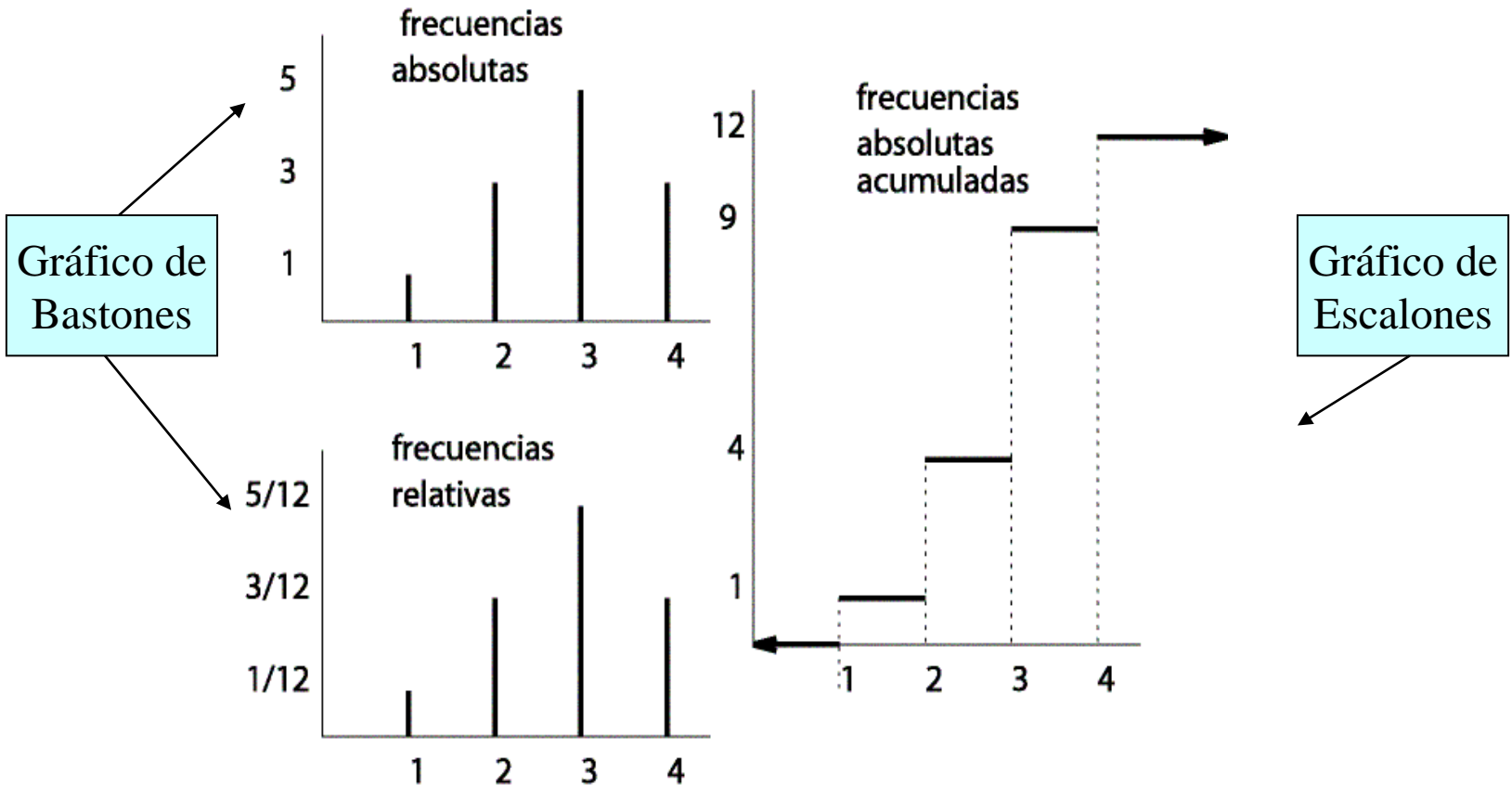
- **Histogramas para v. continuas**
 - El área que hay bajo el histograma entre dos puntos cualesquiera indica la cantidad (porcentaje o frecuencia) de individuos en el intervalo.



Gráficos Acumulados (v. cuantitativas)

- Cada uno de los anteriores diagramas tiene su correspondiente **grafico acumulado** que se realizan a partir de las **frecuencias acumuladas**. Indican, para cada valor de la variable, **la cantidad (frecuencia) de individuos que poseen un valor inferior o igual al mismo**.

V. A. Discreta



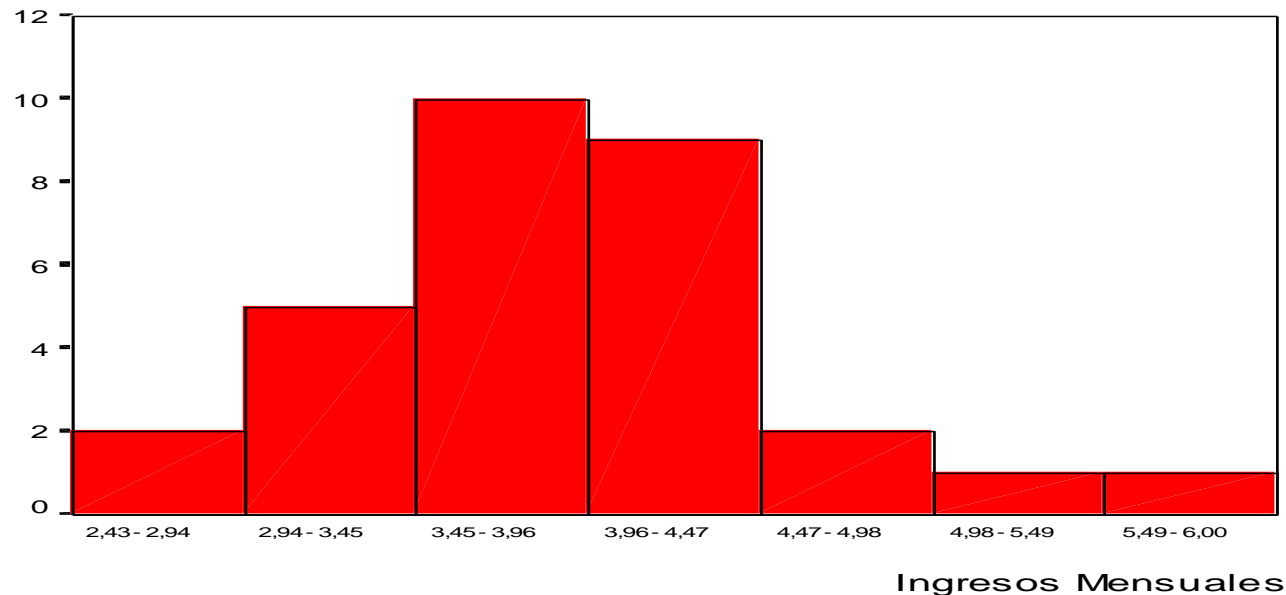
Representaciones Gráficas

Variable Numérica

Histograma

Consta de un conjunto de rectángulos con bases en el eje X, centros en las marcas de clase, longitudes iguales a los tamaños de los intervalos y alturas iguales a las frecuencias de cada intervalo:

Histograma de la variable "Ingresos Mensuales"



Representaciones Gráficas

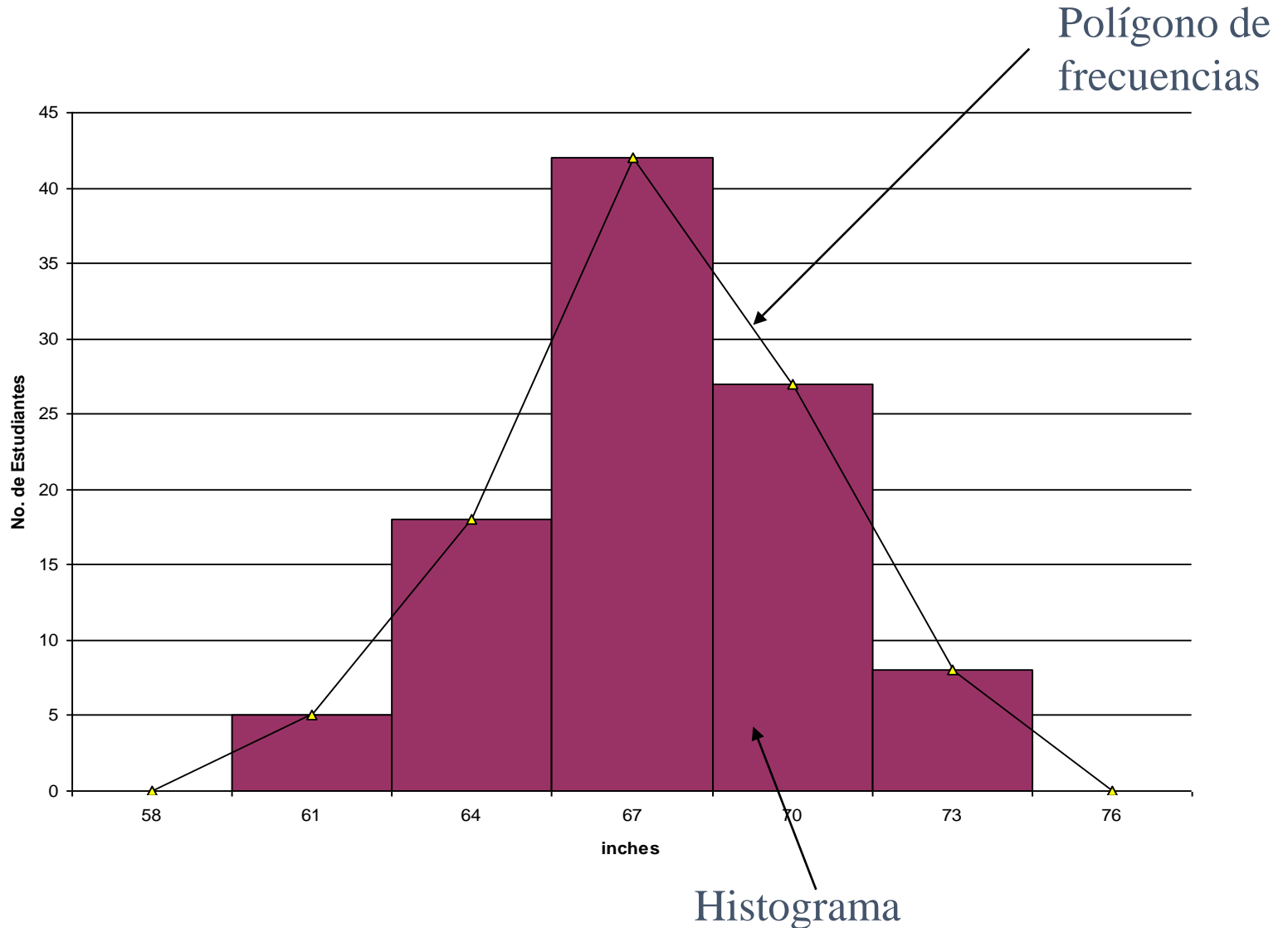
Variable Numérica

Polígono de Frecuencias

Es una línea de trazos (poligonal) que se obtiene uniendo las frecuencias correspondientes a los puntos medios de los intervalos de clase del histograma:

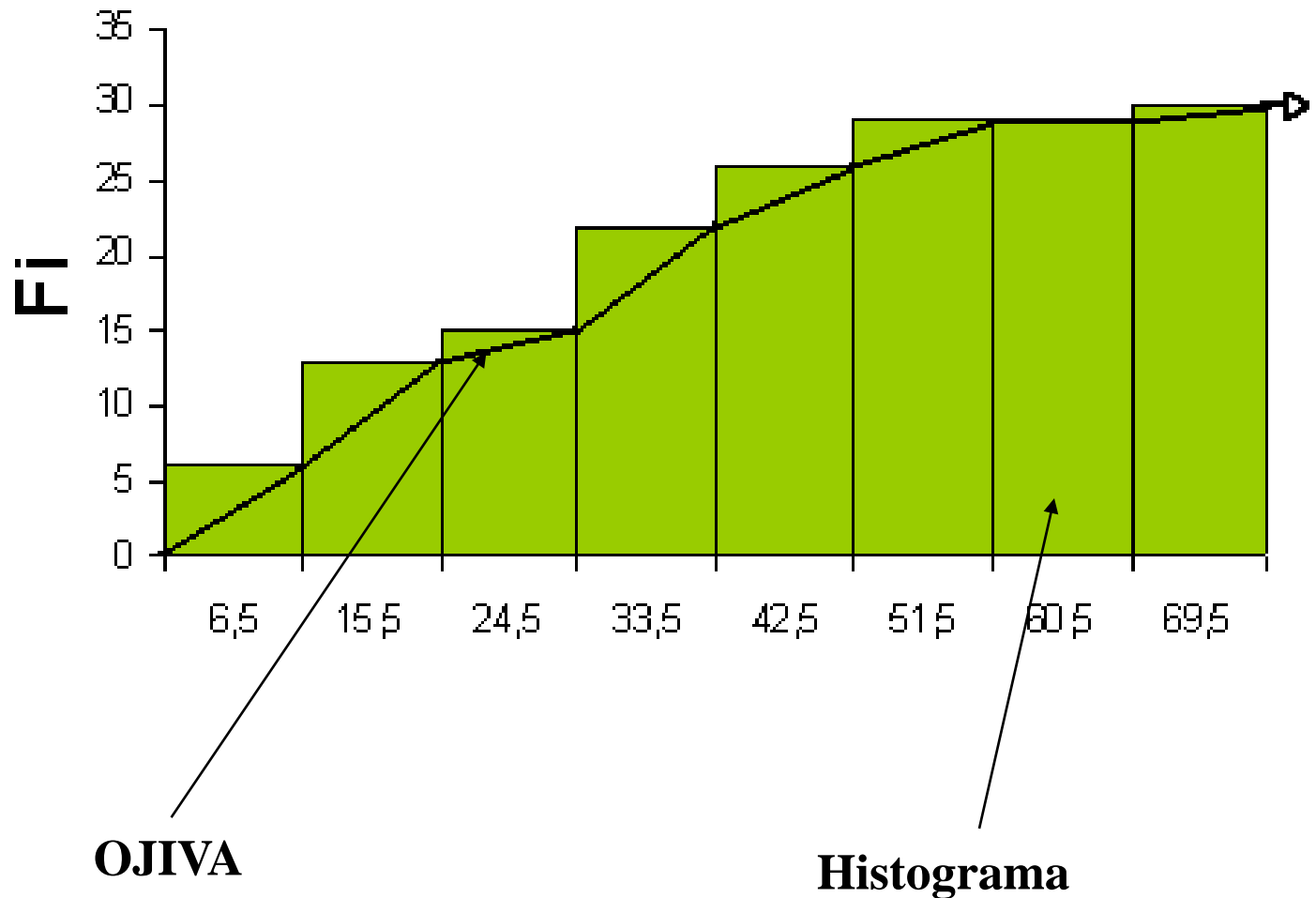
Gráficos Acumulados (v. cuantitativas)

Variable aleatoria continua



Gráficos Acumulados (v. cuantitativas)

Variable aleatoria continua

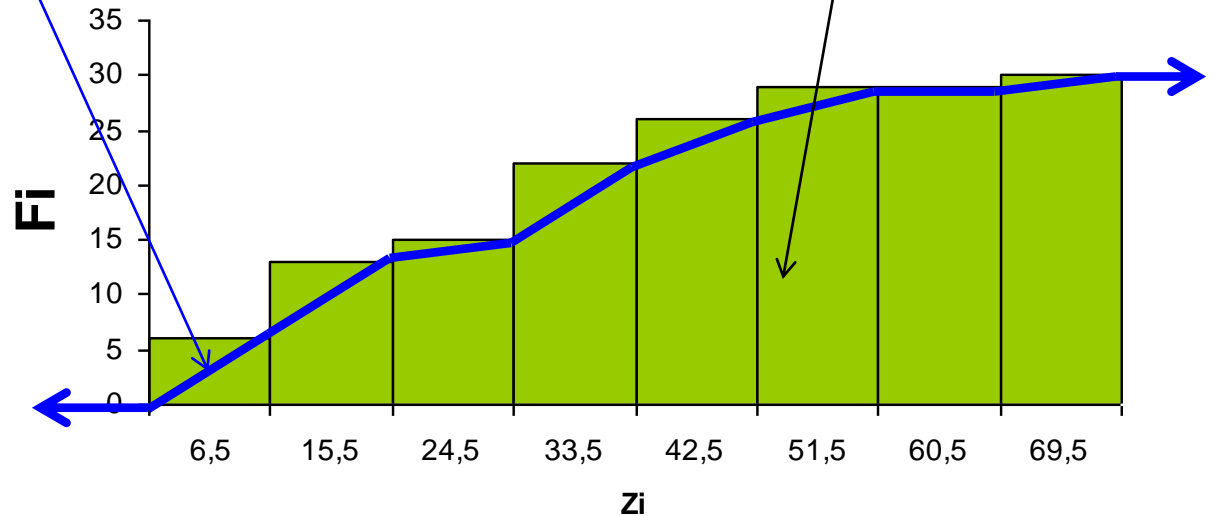


Gráficos Acumulados (v. cuantitativas)

Variable aleatoria continua

OJIVA

HISTOGRAMAS



**Representación de
datos cuantitativos**

Resumen

V. A. Discreta

Diagrama de Bastones

Diagrama de Escalones

V. A. Continua

Histograma

Polígono - Ojiva

**Representación de
datos cualitativos**

Gráfico de Barras

Diagrama de Torta



Gracias

