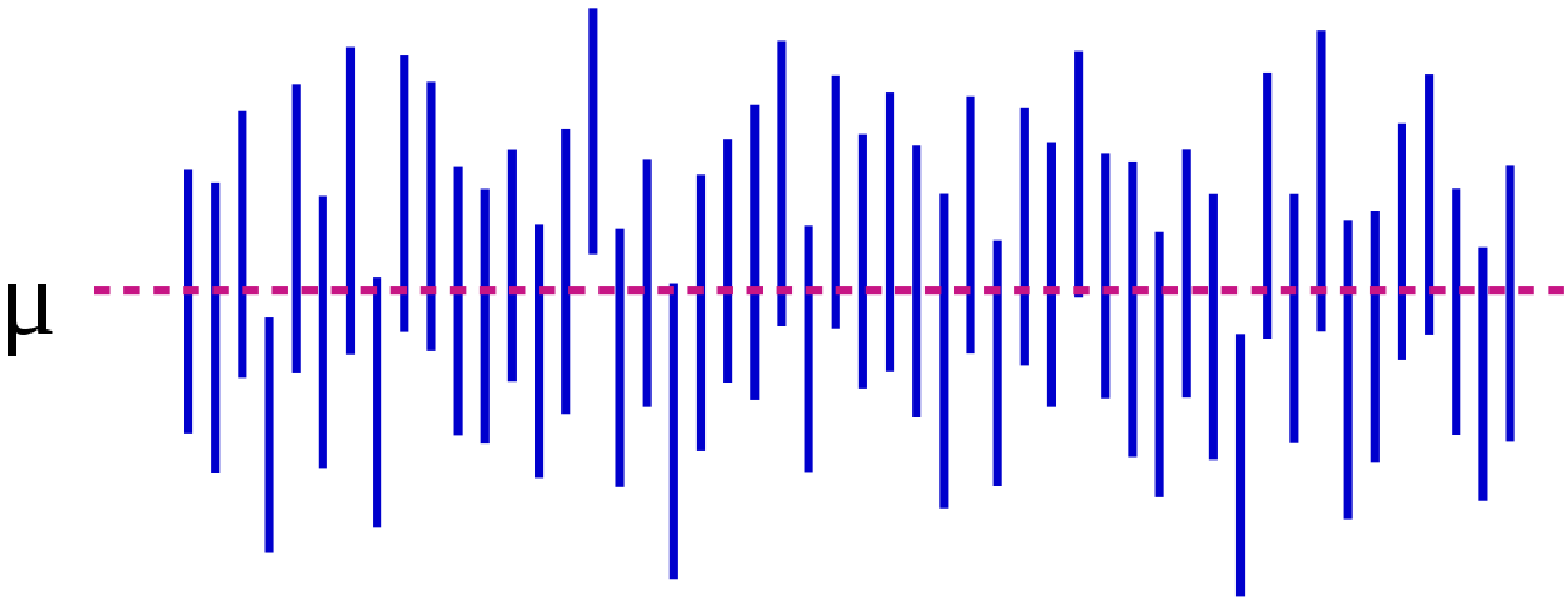
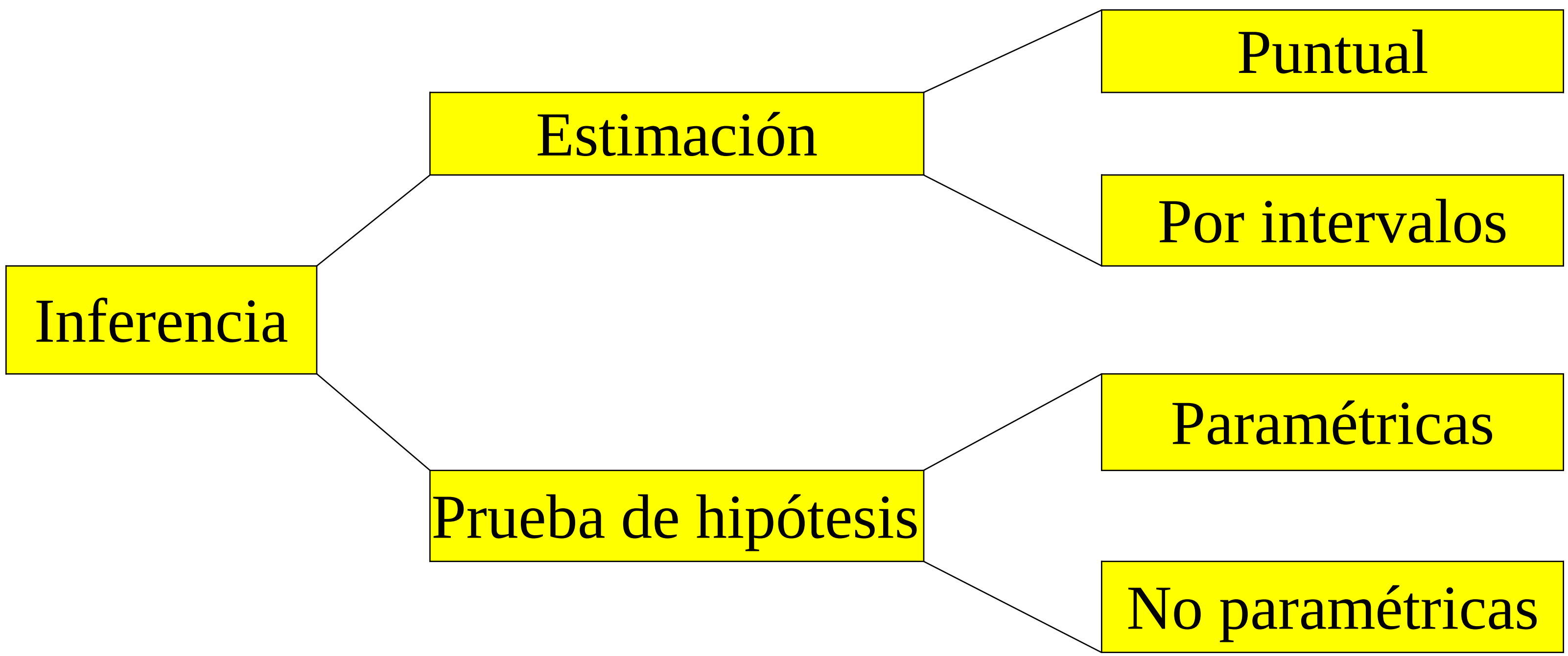
Intervalos de confianza



INFERENCIA


Inferencia
Estimación
Puntual
Por intervalos
Prueba de hipótesis
Paramétricas
No paramétricas

Recordemos

Parámetro θ	Estimador $\hat{\theta}$
μ	$\bar{X}$
σ^2	S^2
π	p

Un estimador es un valor que puede calcularse a partir de los datos muestrales y que proporciona información sobre el valor del parámetro. Un estimador es una variable aleatoria.

$\hat{\theta}$ estimador de θ

PROPIEDADES DE LOS ESTIMADORES

1.-Se dice que un estimador es insesgado si $E[\hat{\theta}] = \theta$

2.-El estimador $\hat{\theta}_1$ **es más eficiente que** $\hat{\theta}_2$ **si** $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$

3.-Un estimador es consistente si $lim_{n \to \infty} \hat{\theta} = \theta$

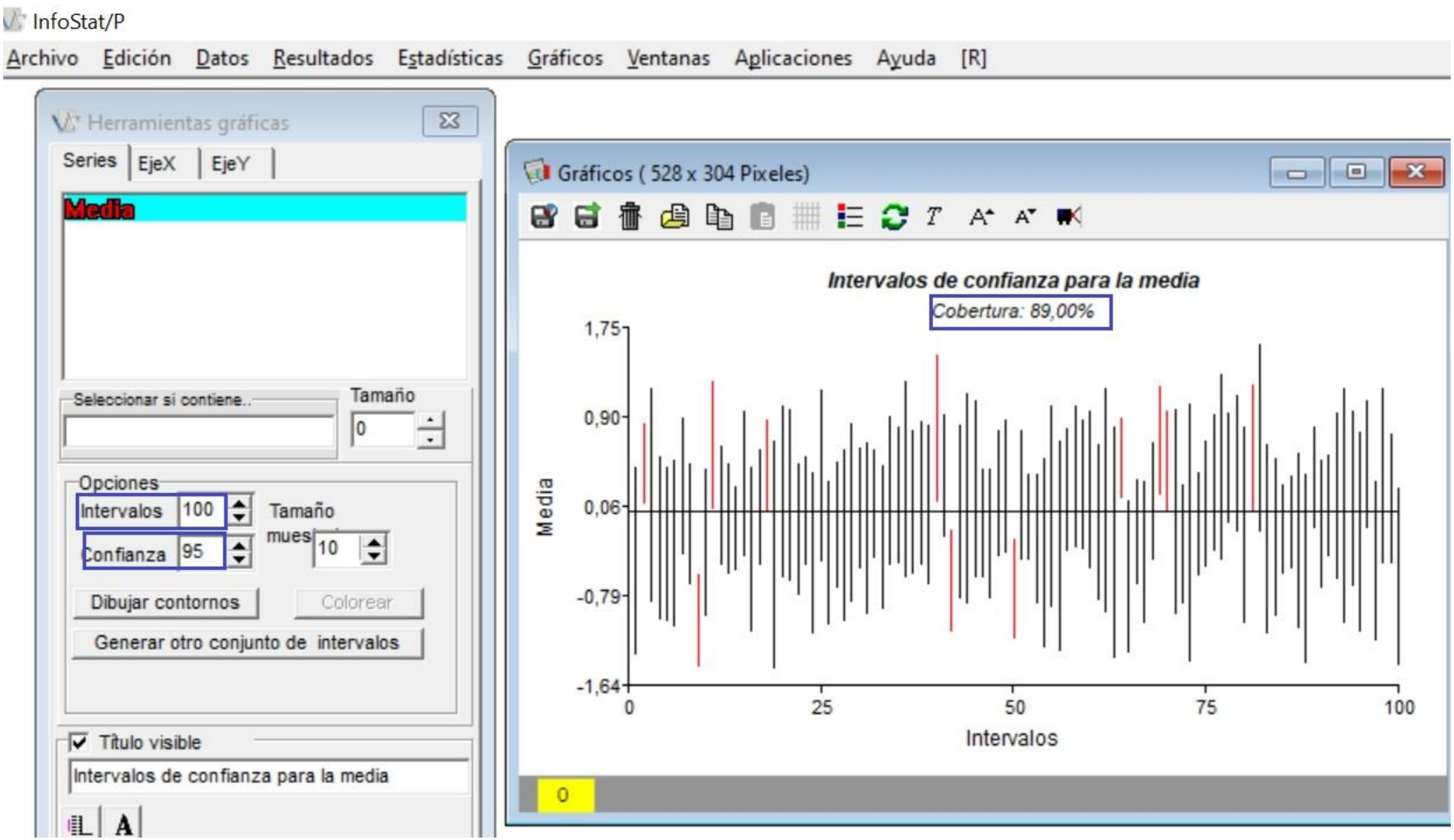
4.-Un estimador suficiente para un parámetro es aquel capaz de recoger o resumir toda la información que la muestra de una Variable aleatoria X contiene.

En la práctica, interesa no sólo dar una estimación de un parámetro, sino que además, un intervalo que permita precisar la incertidumbre existente en la estimación:

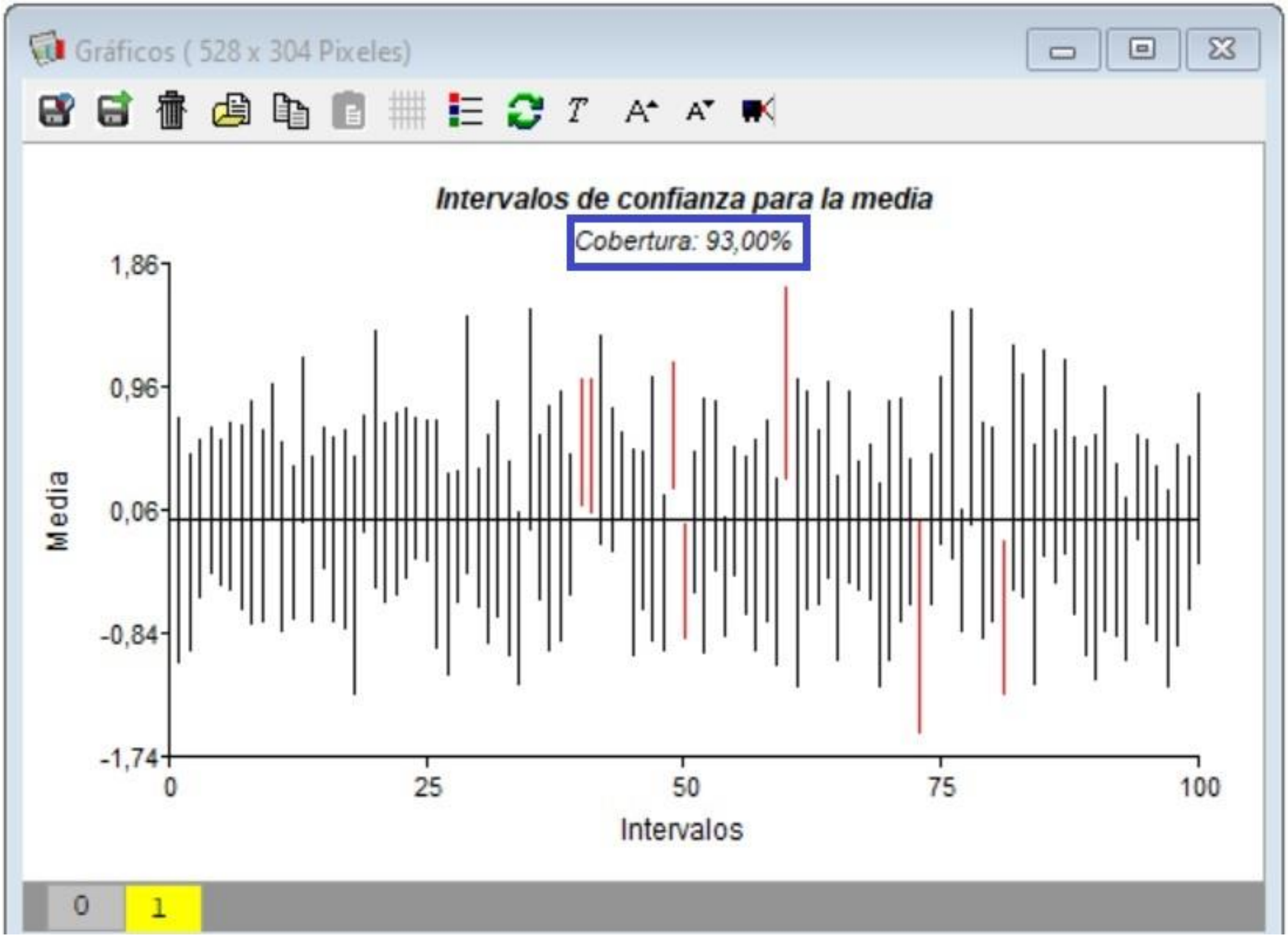
INTERVALO DE CONFIANZA

SIMULACIÓN DE INTERVALOS DE CONFIANZA

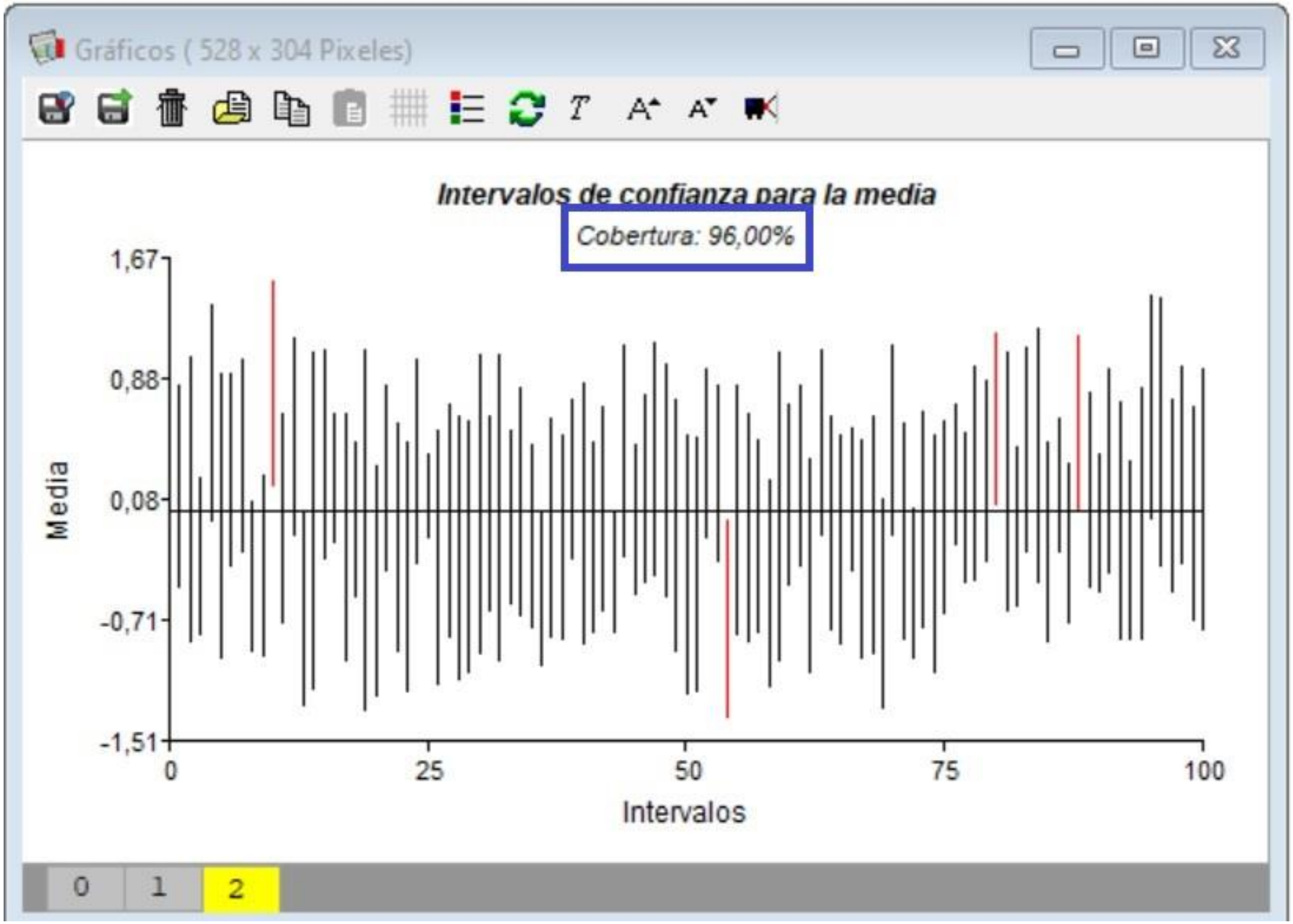
Miren en la tabla. Son 100 intervalos de n=10 calculados con una confianza del 95%. En este ejemplo 89 contienen a la media poblacional y los señalados en rojo (11) no la contienen.



Intervalos calculados con una confianza del 95% y sin embargo sólo 93 de los 100 intervalos contienen a la media poblacional (7 no la contienen)



Intervalos calculados con una confianza del 95% y sin embargo 96 de los 100 intervalos contienen a la media poblacional (4 no la contienen).



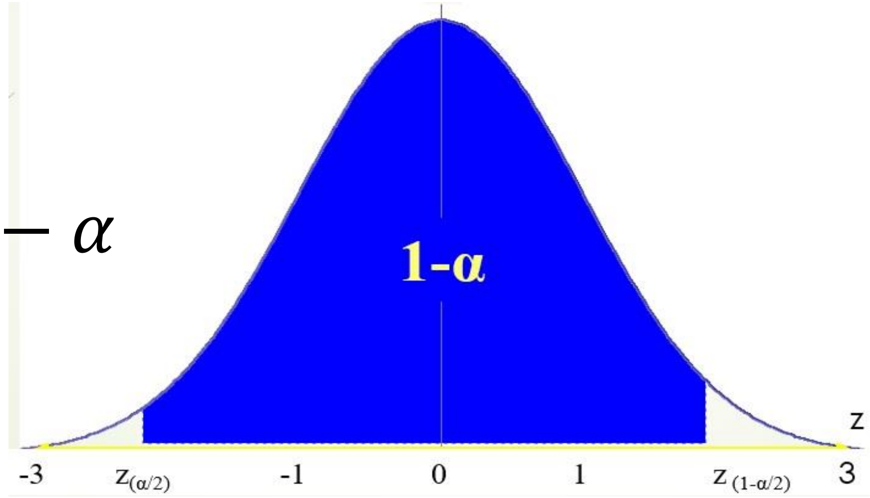
El parámetro es un valor desconocido que tiene su ubicación, si se extrae una muestra y se construye un intervalo ese intervalo puede o no contener al parámetro.

La confianza del 95% no indica que de 100 intervalos, 95 de ellos contendrían a la media poblacional.

La confianza es la probabilidad de contener con un intervalo un parámetro. Es un valor que se logra a la larga, luego de realizar muchos grupos de 100 intervalos (como en el experimento de la moneda, se llega a que P(cara)=P(cruz)=0,5 luego de cientos de tiradas)

A continuación veremos como se obtienen los límites de los intervalos para los parámetros μ, π, y σ^2

INTERVALO DE CONFIANZA PARA μ CON VARIANZA CONOCIDA



$P(LI \leq \mu \leq LS) = 1 - \alpha$

$$P\left(z_{(\alpha/2)} \leq z \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

Reemplazamos $z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$$P\left(z_{(\alpha/2)} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

$$P\left(z_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Despejando μ desde el término entre paréntesis de la igualdad anterior se obtienen las expresiones de los límites inferior y superior del IC.

$$\left(z_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\frac{\alpha}{2}} \right)$$ Multiplicamos a todos los miembros de la desigualdad por $\frac{\sigma}{\sqrt{n}}$

$$z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$ Restamos a todos los miembros de la desigualdad $\bar{X}$

$$-\bar{X}+ z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X}+z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Multiplicamos por (-1): $\bar{X}- z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{X}-z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

Por ser simétrica emplearemos $Z_{1-\alpha/2}$

Reordenamos:

$$\mathrm{P}(\bar{X}- z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}+z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}})=1-\alpha$$

Otra forma de escribir $\bar{X} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

CONSIDERACIONES

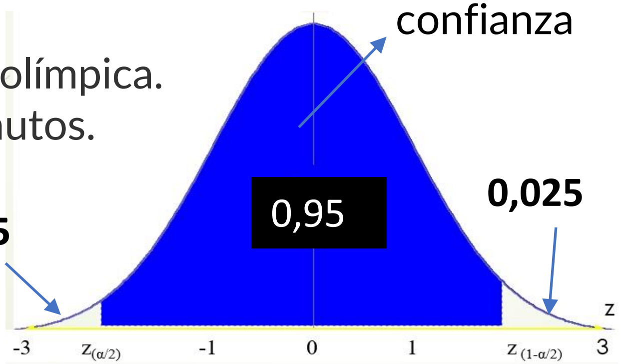
Ojo: No se puede poner $P(566{,}2 \leq \mu \leq 683{,}8) = 0{,}95$ Lo único que se puede hacer es confiar que el intervalo (566,2 ; 683,8) sea uno de los del 95% que contienen al parámetro μ. Solamente podemos hablar de probabilidad antes de tomar la muestra.

Por tanto, un intervalo de confianza viene determinado por:

a) El estimador del parámetro que se haya escogido.
b) El nivel de confianza **$1 - \alpha$** (generalmente del 90%, 95% o 99%)
c) La amplitud del intervalo L.
d) El tamaño de la muestra n.

Ejemplo: Se desea estimar la media del tiempo empleado por un nadador en una prueba olímpica, para lo cual se cronometran 10 pruebas, obteniéndose una media de 41,5 minutos. Sabiendo que esta variable tiene distribución normal con desvío 0,3 minutos. Obtener un intervalo de confianza con un 95% de confianza.

La variable X: tiempo empleado por un nadador en una prueba olímpica.
Se sabe que la distribución es normal y que el desvío es 0,3 minutos.
Pero no se conoce la media μ:



$X \sim N(\mu=?;\sigma^2=(0,3)^2 min^2)$ $\bar{X}=41,5min$ n=10

Si el nivel de confianza es $1-\alpha=0,95$ debemos hallar $1-\alpha/2$

Si $\alpha=0,05$; $\alpha/2=0,025$ entonces $1-\alpha/2=0,975$

buscamos el valor de la variable normal estándar que acumula un área de 0,975 a su izquierda:
$\Rightarrow z_{0,975}=1,96$, ahora se puede reemplazar en:

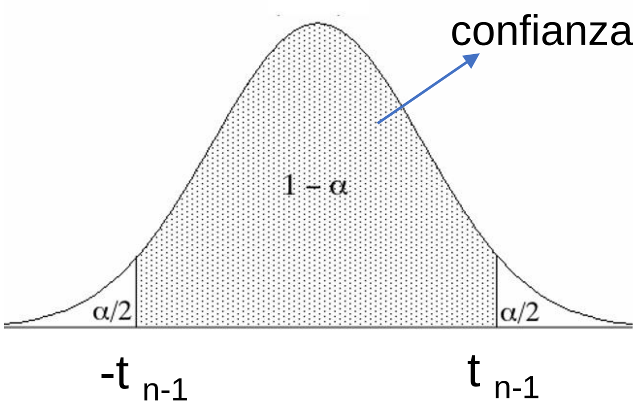
$$P(\bar{X} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}})=1-\alpha \qquad (41,5 - 1,96 \cdot \frac{0,3}{\sqrt{10}} \leq \mu \leq 41,5 + 1,96 \cdot \frac{0,3}{\sqrt{10}})$$

(41,31;41,69)min

Interpretación: Con un nivel de confianza del 95%, se espera que el intervalo [41,31; 41,69]min contenga al verdadero valor del tiempo medio poblacional empleado por un nadador en una prueba olímpica.

INTERVALO DE CONFIANZA PARA μ CON VARIANZA DESCONOCIDA

Del mismo modo que se trabajó para encontrar el intervalo de confianza para la media μ con varianza σ^2 conocida, se realiza para encontrar un intervalo de confianza para la media μ cuando la varianza es desconocida. Para ello necesitamos utilizar la distribución t.



$$P\left(t_{n-1;\alpha/2} \leq t \leq t_{n-1;\,1-\alpha/2}\right) = 1-\alpha$$

$$t = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

Reemplazamos y despejamos y obtenemos

$$P\left(\bar{X} - t_{n-1;\,1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1;\,1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\right) = 1-\alpha$$

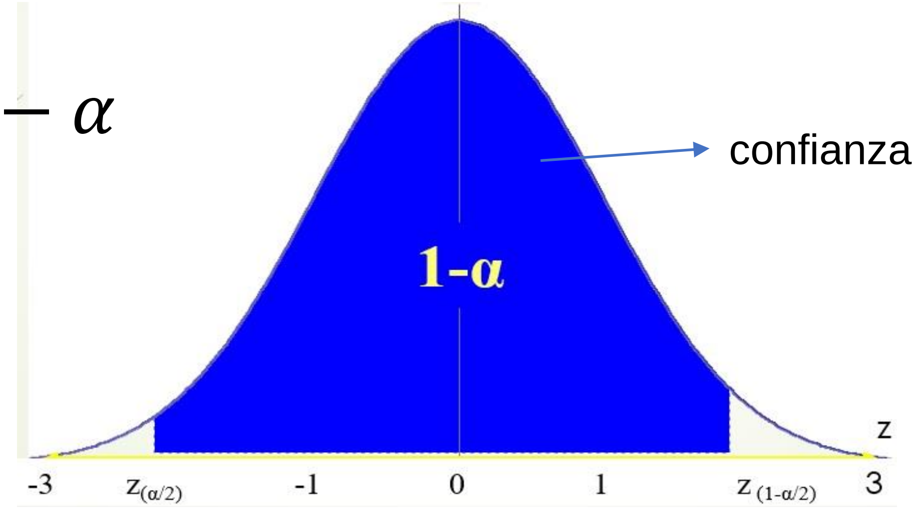
O escrito de este modo $\quad \bar{X} \pm t_{n-1;\,1-\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}$

INTERVALO DE CONFIANZA PARA LA PROPORCIÓN

Para realizar un intervalo para la proporción π, se utiliza la distribución normal que se muestra a continuación:

$$P\left(z_{(\alpha/2)} \leq z \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

$$z = \frac{p - \pi}{\sqrt{\frac{p(1-p)}{n}}} \approx \mathrm{N}(0{,}1)$$



Se reemplaza y despeja para obtener:

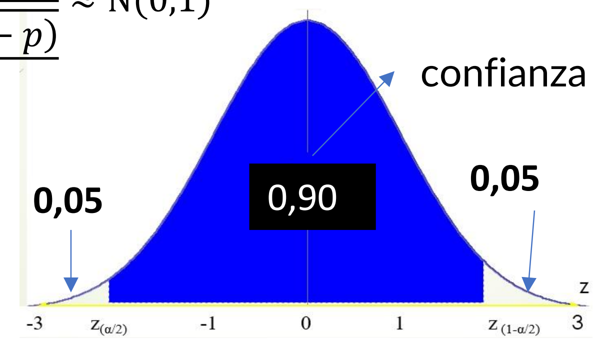
$$\mathrm{P}\left(p - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha$$

O escrito de este modo

$$p \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}$$

EJEMPLO: Un fabricante de material de laboratorio desea saber qué proporción de sus productos **no contienen fallas**. Estimar puntualmente y mediante un intervalo de confianza la proporción de material **sin fallas** si se sabe que en una muestra de 100 unidades producidas se contó sólo 4 unidades con fallas. Usar un nivel de confianza del 90% (1 - α = 0,90).

En este caso, la variable pivotal a emplear es $z = \dfrac{p-\pi}{\sqrt{\dfrac{p(1-p)}{n}}} \approx \mathrm{N}(0,1)$



Datos: n = 100; unidades con fallas = 4; 1 - α = 0,90

$$p = \frac{\#casos favorables}{tamaño de muestra} = \frac{96}{100} = 0{,}96$$

Si el nivel de confianza es 1–α=0,90 debemos hallar 1–α/2

Si α=0,1 ; α/2=0,05 entonces 1-α/2=0,95≈0,9495

Z para P= 0,9495 es 1,64

$$\mathrm{P}\left(p - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \leq \pi \leq p + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}}\right) = 1-\alpha \quad \text{ó} \quad p \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \quad (1)$$

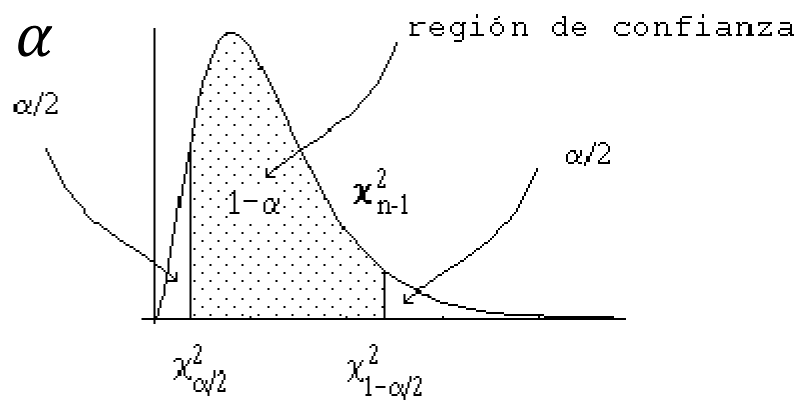
Reemplazando en (1) $0{,}96 \pm 1{,}64 \cdot \sqrt{\dfrac{0{,}96 \cdot 0{,}04}{100}} \equiv 0{,}96 \pm 0{,}032$ ***[0,928; 0,992]***.

<u>Interpretación</u>: Con un nivel de confianza del 90%, se espera que el intervalo [0,928; 0,992] contenga o cubra al verdadero valor de la proporción de unidades producidas sin fallas correspondiente al fabricante de material de laboratorio.

INTERVALO DE CONFIANZA PARA LA VARIANZA

Para realizar un intervalo para la varianza poblacional σ^2 , se utiliza la distribución chi cuadrado (χ^2 que se muestra a continuación:

$$P\left(\chi^2_{\frac{\alpha}{2}} < \chi^2 < \chi^2_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$



Se reemplaza :

$$\chi^2 = \frac{(n-1)\cdot S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$P\left(\chi^2_{n-1;\,\frac{\alpha}{2}} < \frac{(n-1)\cdot S^2}{\sigma^2} < \chi^2_{n-1;\,1-\frac{\alpha}{2}}\right) = 1-\alpha$$

Como la varianza poblacional debe quedar en el numerador, se aplican propiedades de las desigualdades

$$P\left(\frac{1}{\chi^2_{n-1;\,\frac{\alpha}{2}}} > \frac{\sigma^2}{(n-1)\cdot S^2} > \frac{1}{\chi^2_{n-1;\,1-\frac{\alpha}{2}}}\right) = 1-\alpha$$

$$P\left(\frac{1}{\chi^2_{n-1;\,\frac{\alpha}{2}}} > \frac{\sigma^2}{(n-1)\cdot S^2} > \frac{1}{\chi^2_{n-1;\,1-\frac{\alpha}{2}}} \right) = 1-\alpha$$

Multiplicamos a todos los miembros por (n-1).S^2

$$P\left(\frac{(n-1)S^2}{\chi^2_{n-1;\,\frac{\alpha}{2}}} > \sigma^2 > \frac{(n-1)S^2}{\chi^2_{n-1;\,1-\frac{\alpha}{2}}} \right) = 1-\alpha$$

Colocamos los signos correctamente, cambiando de lugar los extremos

$$\mathrm{P}\left(\frac{(n-1)S^2}{\chi^2_{n-1;\,1-\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;\,\frac{\alpha}{2}}} \right) = 1-\alpha$$

Otra forma de escribir los límites:

$$\text{Limite Inferior: } \frac{(n-1)S^2}{\chi^2_{n-1;\,1-\frac{\alpha}{2}}}$$

$$\text{Limite Superior: } \frac{(n-1)S^2}{\chi^2_{n-1;\,\frac{\alpha}{2}}}$$

Ejemplo: Se hicieron determinaciones de la concentración de amilasa en suero en individuos de una muestra aleatoria, de tamaño 15, tomada de una población con individuos aparentemente normales. La muestra proporcionó una media de 96 UI/ml y una desviación estándar de 35 UI/ml. Suponiendo que la población se distribuye normalmente, estimar la varianza poblacional mediante un intervalo de confianza el 95%.

UI (Unidad internacional): es una unidad de medida de la cantidad de una sustancia, basada en su actividad biológica mediada (o sus efectos).

Para estimar la varianza poblacional (σ^2) mediante un intervalo de confianza, primero debemos observar que la variable aleatoria es la indicada para ser usada como variable pivotal

$$\chi^2 = \frac{(n-1)\cdot S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Como en el caso anterior, para obtener las expresiones de los límites del intervalo de confianza, se utiliza la expresión:

$$\mathrm{P}\left(\frac{(n-1)S^2}{\chi^2_{n-1;\ 1-\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;\ \frac{\alpha}{2}}} \right) = 1-\alpha$$

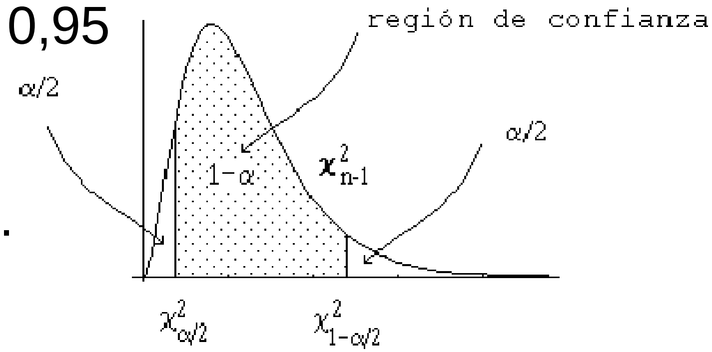
Utilizamos esta forma de escribir los límites:

$$\text{Limite Inferior (LI): } \frac{(n-1)S^2}{\chi^2_{n-1;\ 1-\frac{\alpha}{2}}}$$

$$\text{Limite Superior (LS): } \frac{(n-1)S^2}{\chi^2_{n-1;\ \frac{\alpha}{2}}}$$

Datos: n = 15; $\bar{X}$= 96 UI/ml; s = 35 UI/ml ; y 1-α = 0,95

Se necesita buscar dos valores de la tabla chi cuadrado.



$$\chi^2_{n-1;\ \frac{\alpha}{2}} = \chi^2_{14;\ 0{,}025} = 5{,}629$$

$$\chi^2_{n-1;\ 1-\frac{\alpha}{2}} = \chi^2_{14;\ 0{,}975} = 26{,}12$$

Reemplazando nos queda:

$$\text{LI: } \frac{(14)\cdot(35^2)}{26{,}15} = 656{,}58$$

$$\text{LS: } \frac{14\cdot(35^2)}{5{,}629} = 3046{,}72$$

<u>INTERPRETACIÓN</u>: Se espera con una confianza del 95% que el intervalo [656,58 (UI/ml)^2; 3046,72 (UI/ml)^2] cubra o abarque al verdadero valor de la varianza de la concentración de amilasa en suero en la población de individuos aparentemente normales.

ERROR EN LA ESTIMACIÓN (Sinónimo de margen de error)

$$P(\bar{X} - \underbrace{z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}_{\text{Error}} \leq \mu \leq \bar{X} + \underbrace{z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}}_{\text{Error}}) = 1-\alpha$$

Amplitud del intervalo: $(\bar{X} - Error\, ;\, \bar{X} + Error)$

Siendo E, la semi-amplitud

$$E = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \cong z_{1-\alpha/2} \frac{s}{\sqrt{n}}$$

Despejando n, obtenemos:

$$\boldsymbol{n = \left(\frac{z_{1-\alpha/2} \,.\, S}{E}\right)^2}$$

Gracias