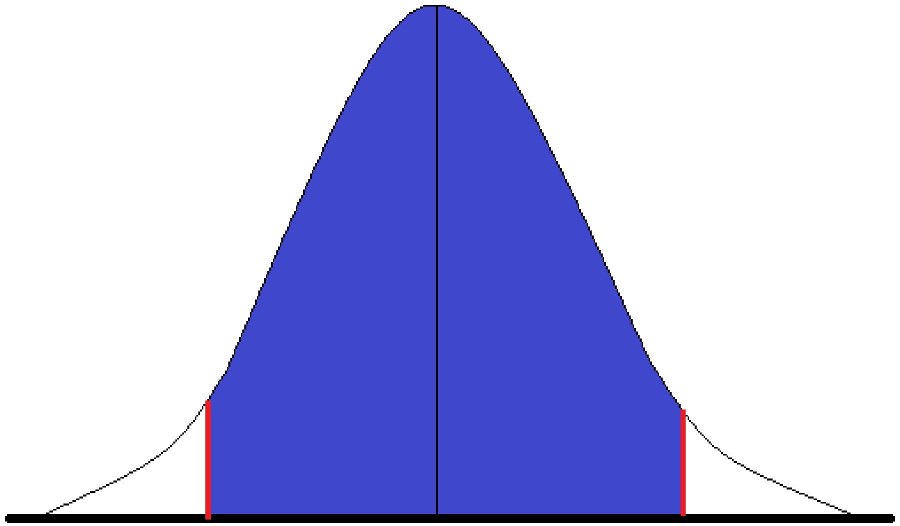
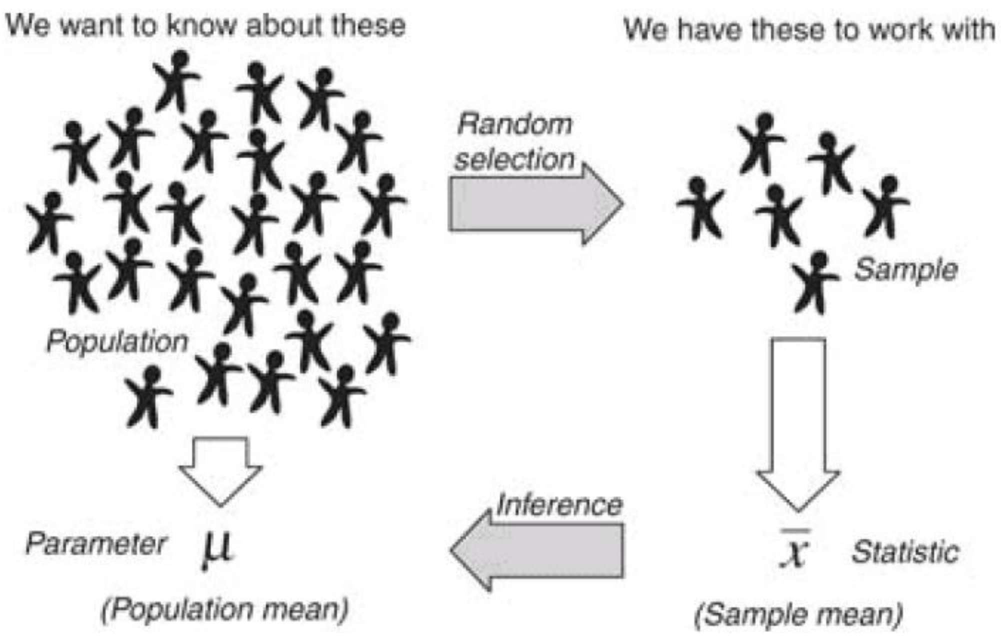
Intervalos de confianza



La estadística trata de obtener conclusiones a partir de la observación de datos. Generalmente estos datos se obtienen a partir del **muestreo** de una **población**.



Un **estadístico S** es **una VARIABLE ALEATORIA calculada a partir de los valores de una muestra**

$$S = f(X_1, X_2, \ldots, X_n)$$

CEU Universidad San Pablo

Ejemplo

Sea X una variable aleatoria con esperanza μ y varianza $0 < \sigma^2 < \infty$. Si $(X_1, \dots, X_n)$ es una muestra aleatoria simple, para n suficientemente grande entonces el TLC asegura que:

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Definición

Variable tipificada:

$$Z = \frac{X - \mu_X}{\sigma_X}$$

El TLC se aplica a cualquier variable aleatoria.

¡ X puede ser CUALQUIER variable aleatoria !

(Si X es normal entonces no hace falta TLC)

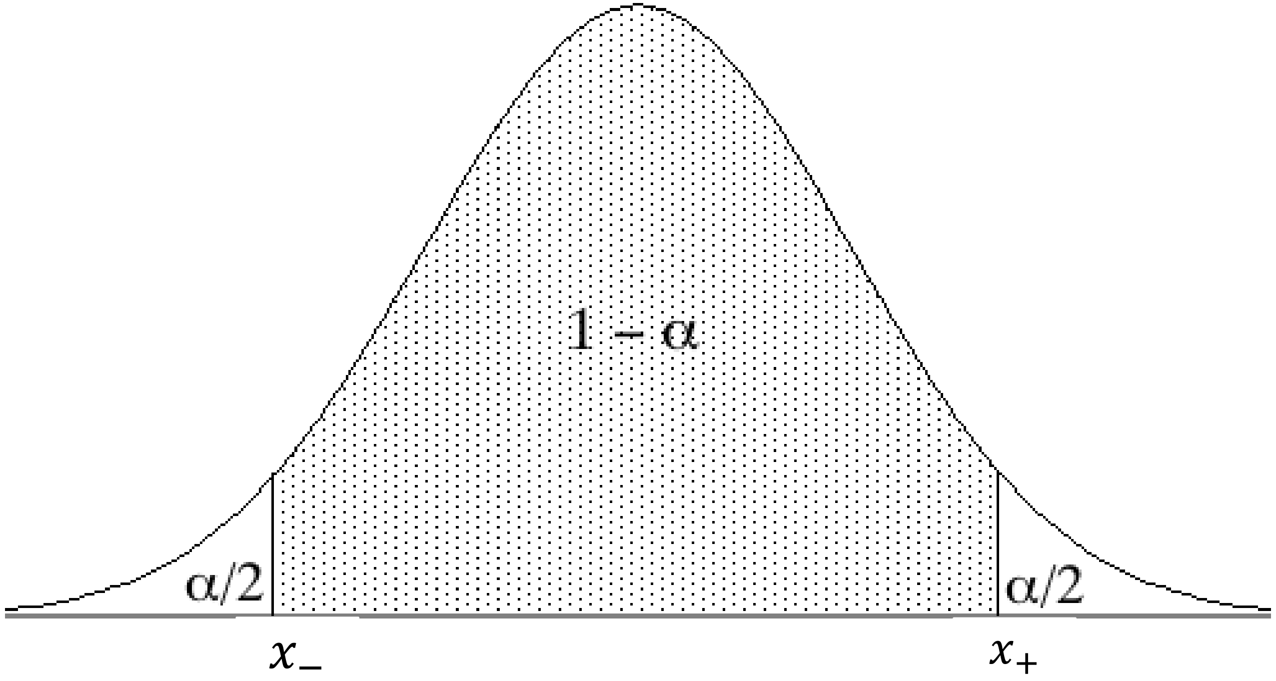
CEU Universidad San Pablo

Intervalo bilateral (dos colas)

Sea X una variable aleatoria con esperanza μ y varianza $0 < \sigma^2 < \infty$. Entonces podemos afirmar que $\mu \in (x_-, x_+)$ con un nivel de confianza $1 - \alpha$ siendo:

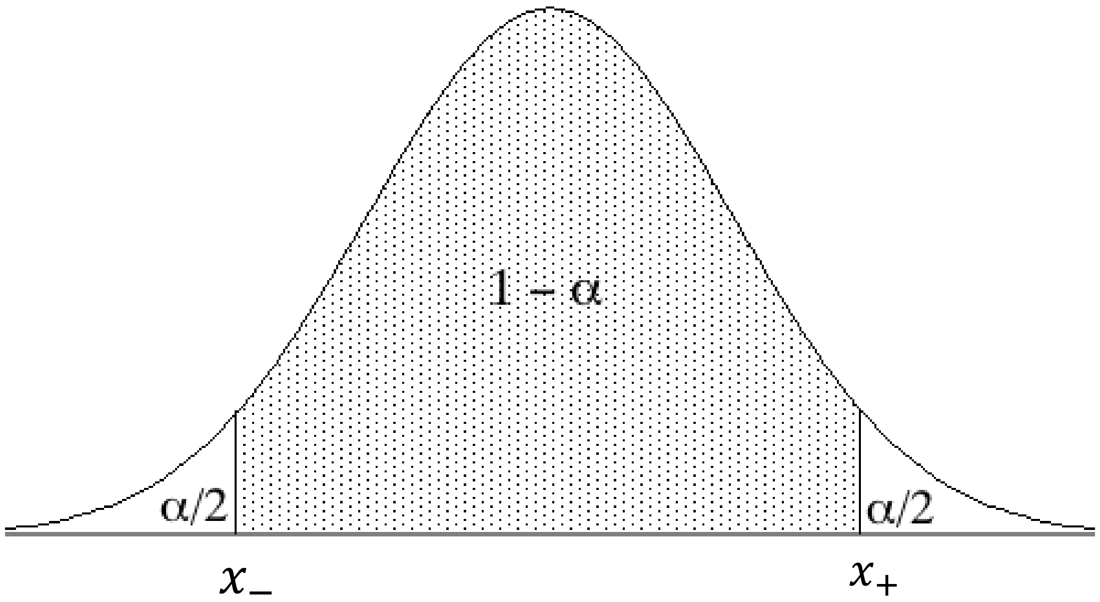
$$x_\pm = \bar{x} \mp \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}$$

Donde $z_{\frac{\alpha}{2}}$ es el $\frac{\alpha}{2}$ −cuantil de $N(0,1)$.



Ejemplo

Una variable aleatoria tiene media desconocida μ y varianza conocida $\sigma^2 = 156{,}25$. Si tomamos una muestra de $n = 50$ con media muestral $\bar{x} = 72{,}4$. Calcula el intervalo de confianza bilateral con una significancia del 5%.



El 95% de las veces el intervalo es correcto.
El 5% de las veces el intervalo es INCORRECTO.

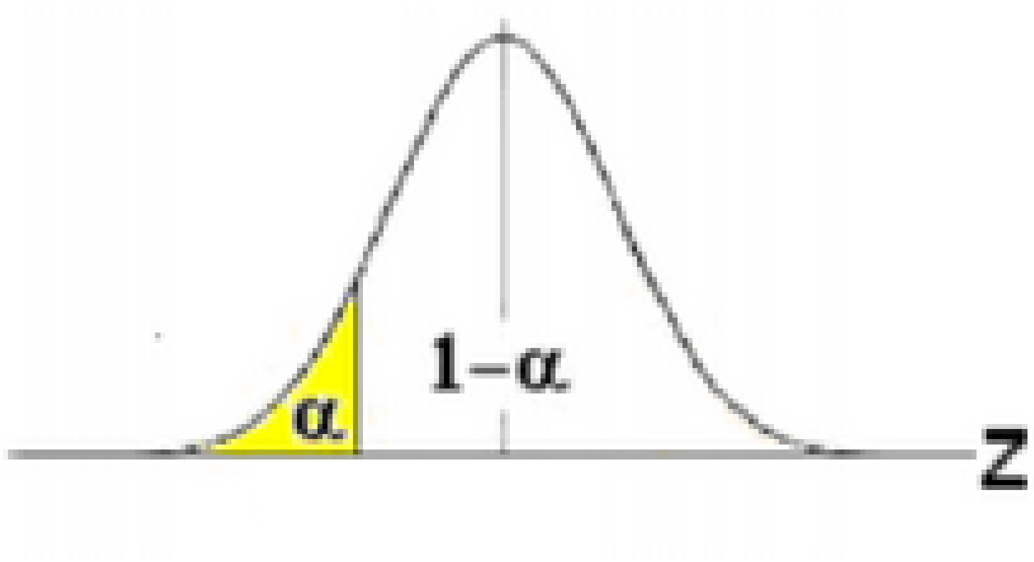
$$\mu \in (68.94, 75.86)$$

Intervalo unilateral (una cola)

Sea X una variable aleatoria con esperanza μ y varianza $0 < \sigma^2 < \infty$. Entonces podemos afirmar que $\mu \in (x_{-}, \infty)$ con un nivel de confianza $1 - \alpha$ siendo:

$$x_{-} = \bar{x} + \frac{\sigma}{\sqrt{n}} z_{\alpha}$$

Donde z_{α} es el α −cuantil de $N(0,1)$.

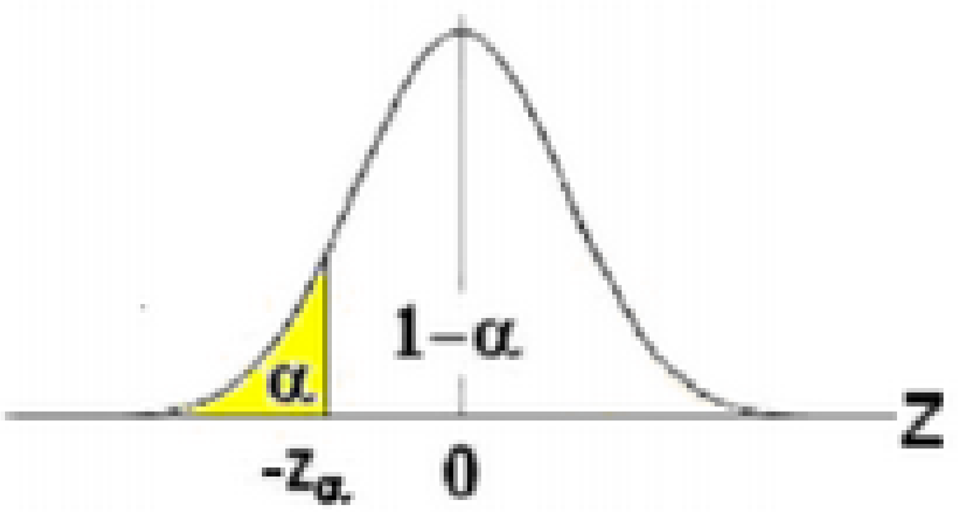


Ejemplo

En una universidad se ha tomado una muestra de alumnos a los cuales se les ha medido el CI.

¿Intervalo para el CI medio de la universidad?

Supón la desviación poblacional $\sigma = 15$.



$$\mu \in (98.77, \infty)$$

CEU Universidad San Pablo

Ejemplo

Sea X una variable aleatoria con esperanza μ y varianza **desconocida** $\sigma^2 < \infty$. Si $(X_1, \dots, X_n)$ es una muestra aleatoria simple.

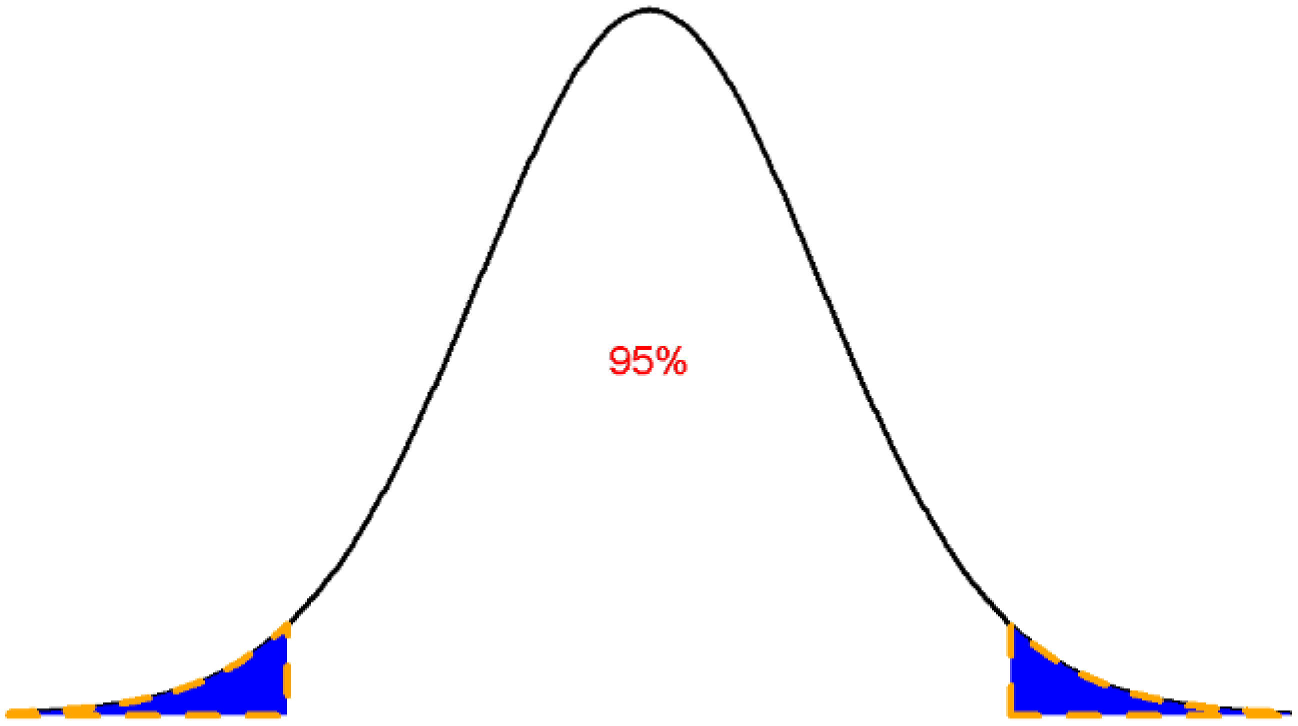
¿Cómo podemos estimar μ?

Para buscar un intervalo de confianza necesitamos que X sea **normal**.

Teorema

Si $X \sim N(\mu, \sigma^2)$, entonces:

$$\frac{\bar{X} - \mu}{\hat{S}/\sqrt{n}} \sim T_{n-1}$$

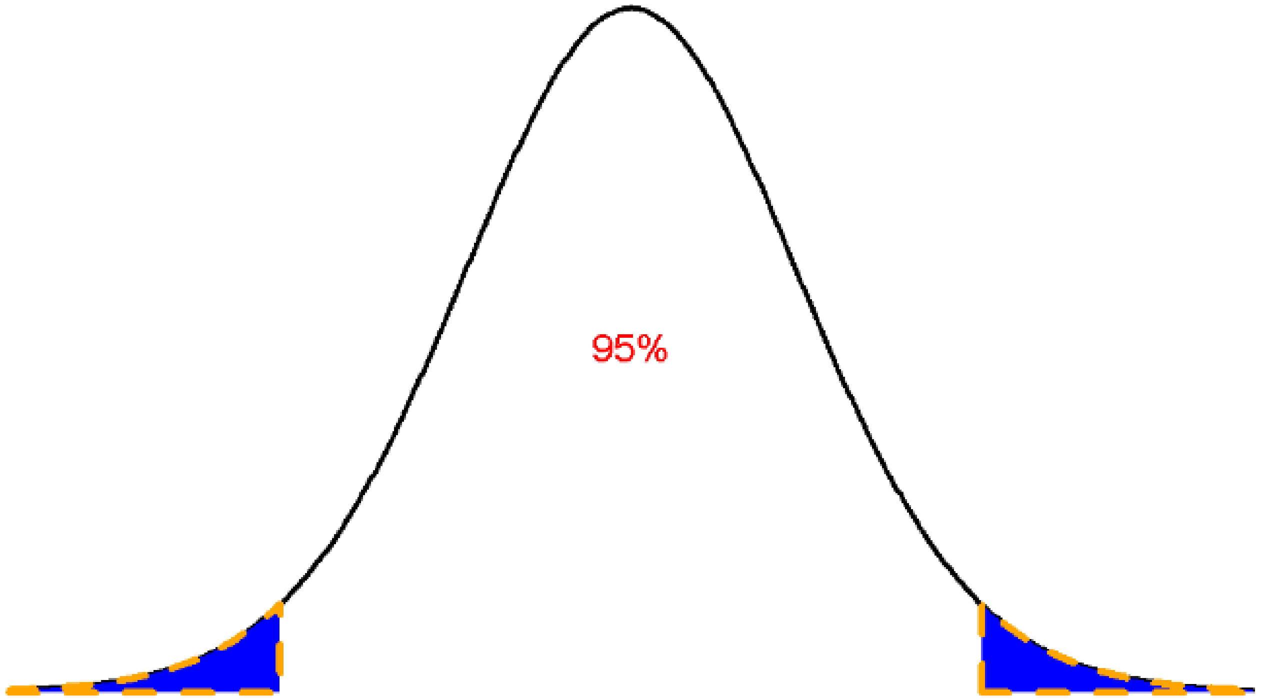


Intervalo bilateral (dos colas)

Sea $X \sim N(\mu, \sigma^2)$. Entonces podemos afirmar que $\mu \in (x_-, x_+)$ con un nivel de confianza $1 - \alpha$ siendo:

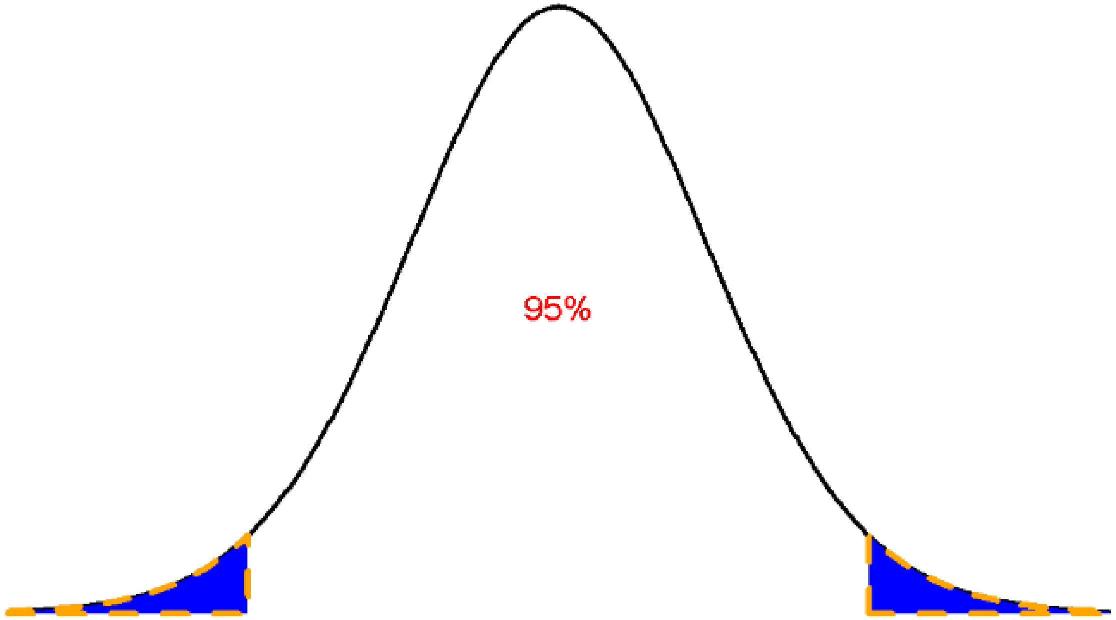
$$x_{\pm} = \bar{x} \mp \frac{\hat{s}}{\sqrt{n}} t_{\frac{\alpha}{2}}$$

Donde $t_{\frac{\alpha}{2}}$ es el $\frac{\alpha}{2}$ −cuantil de T_{n-1}.



Ejemplo

Tomamos una muestra de $n = 50$ en una población normal con media muestral $\bar{x} = 72{,}4$ y desviación muestral $\hat{s} = 12{,}5$. Calcula el intervalo de confianza bilateral con una significancia del 5%.



$$\mu \in (68.85, 75.95)$$

Ejemplo

Sea X una variable aleatoria con varianza $\sigma^2 < \infty$. Si $(X_1, \dots, X_n)$ es una muestra aleatoria simple.

¿Cómo podemos estimar σ^2?

Para buscar un intervalo de confianza necesitamos que X sea **normal**.

Definición

$$\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

ESTIMADOR SESGADO

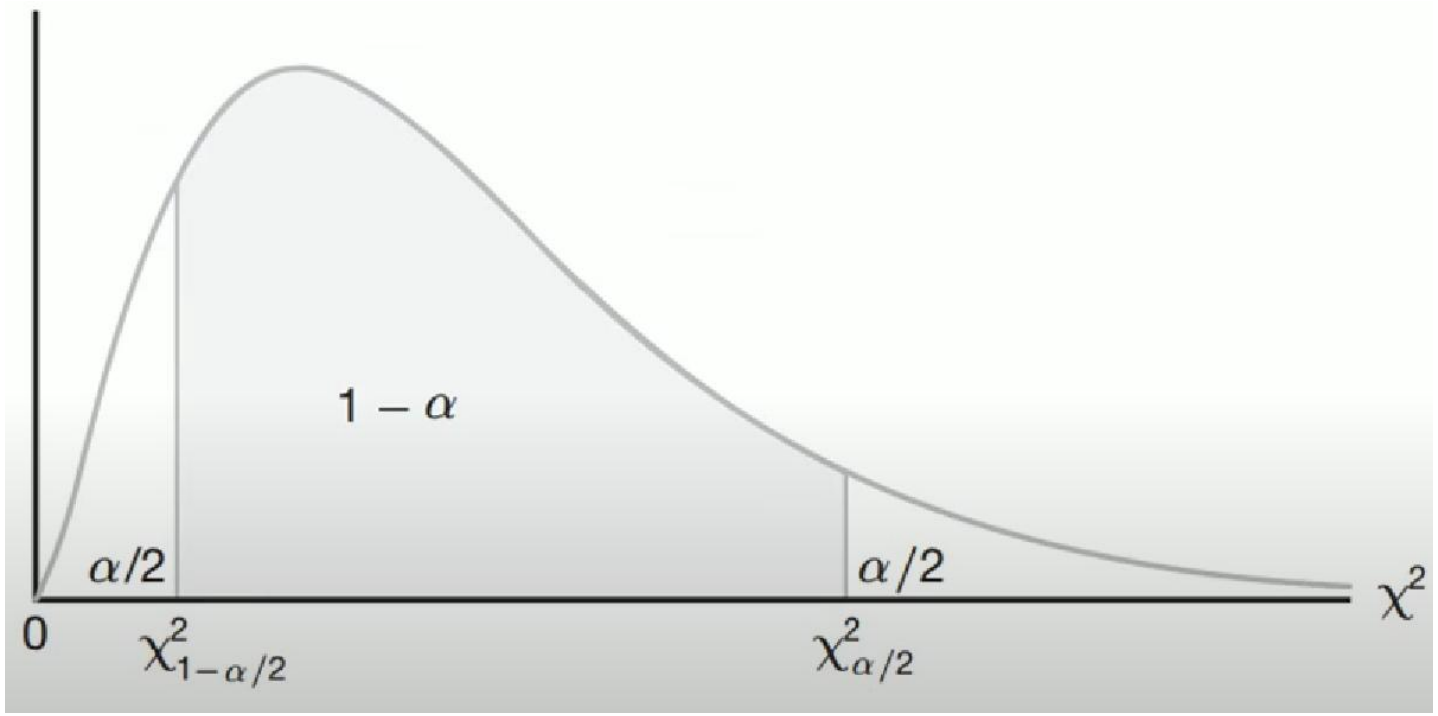
$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \Rightarrow E[S^2] = \frac{n-1}{n}\sigma^2$$

Teorema

Si $X \sim N(\mu, \sigma^2)$, entonces:

1. $(n-1)\dfrac{\hat{S}^2}{\sigma^2} \sim \chi^2_{n-1}$

2. $E\left[\hat{S}^2\right] = \sigma^2$

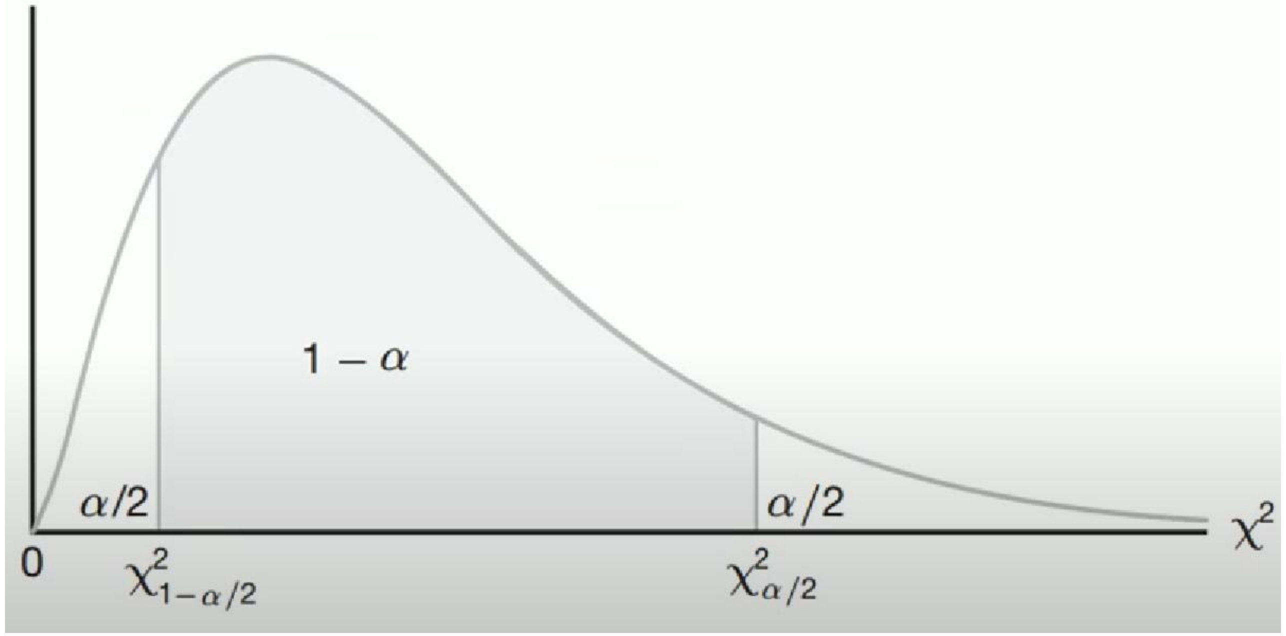


Intervalo bilateral (dos colas)

Sea $X \sim N(\mu, \sigma^2)$. Entonces podemos afirmar con un nivel de confianza $1-\alpha$:

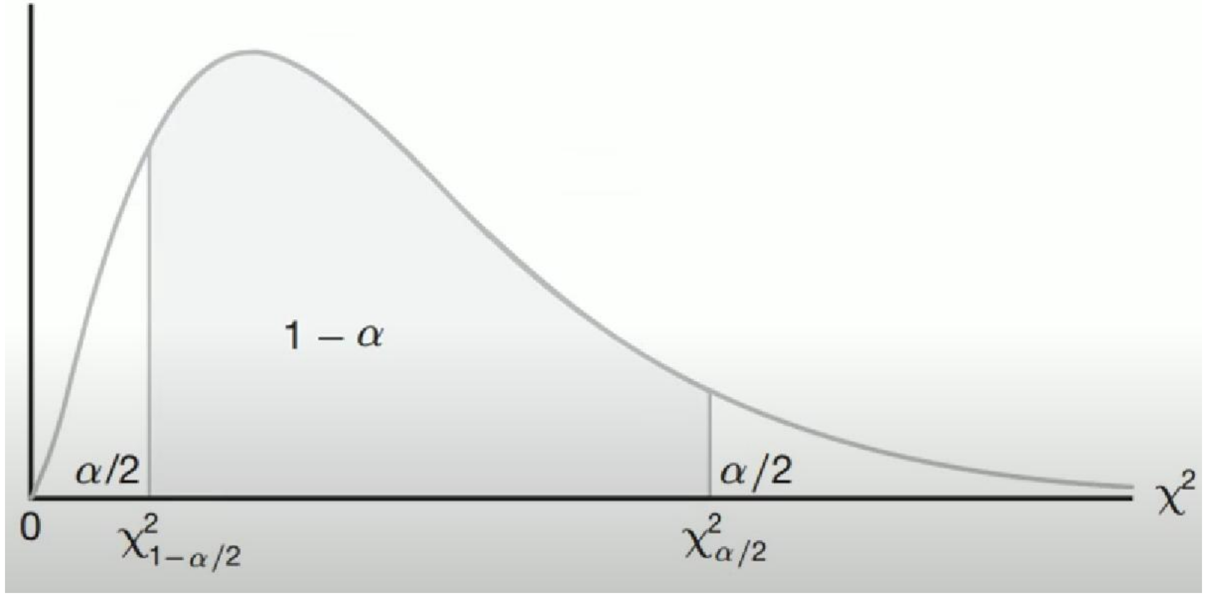
$$(n-1)\frac{\hat{s}^2}{x_{1-\frac{\alpha}{2}}} \leq \sigma^2 \leq (n-1)\frac{\hat{s}^2}{x_{\frac{\alpha}{2}}}$$

Donde x_γ es el γ −cuantil de χ^2_{n-1}.



Ejemplo

En una escuela se toma una muestra de 16 estudiantes. La desviación típica muestral de sus alturas es $\hat{s} = 2{,}4$. Calcula el intervalo de confianza bilateral con una significancia del 5%.



$$\sigma^2 \in (3.14, 13.80)$$

Ejemplo

Sean X, Y dos variables aleatorias con esperanzas μ_X, μ_Y y varianzas **conocidas** $0 < \sigma_X^2, \sigma_Y^2 < \infty$. Si $(X_1, \ldots, X_n)$ e $(Y_1, \ldots, Y_n)$ son dos muestras aleatorias simples.

¿Cómo podemos estimar $\mu_1 - \mu_2$?

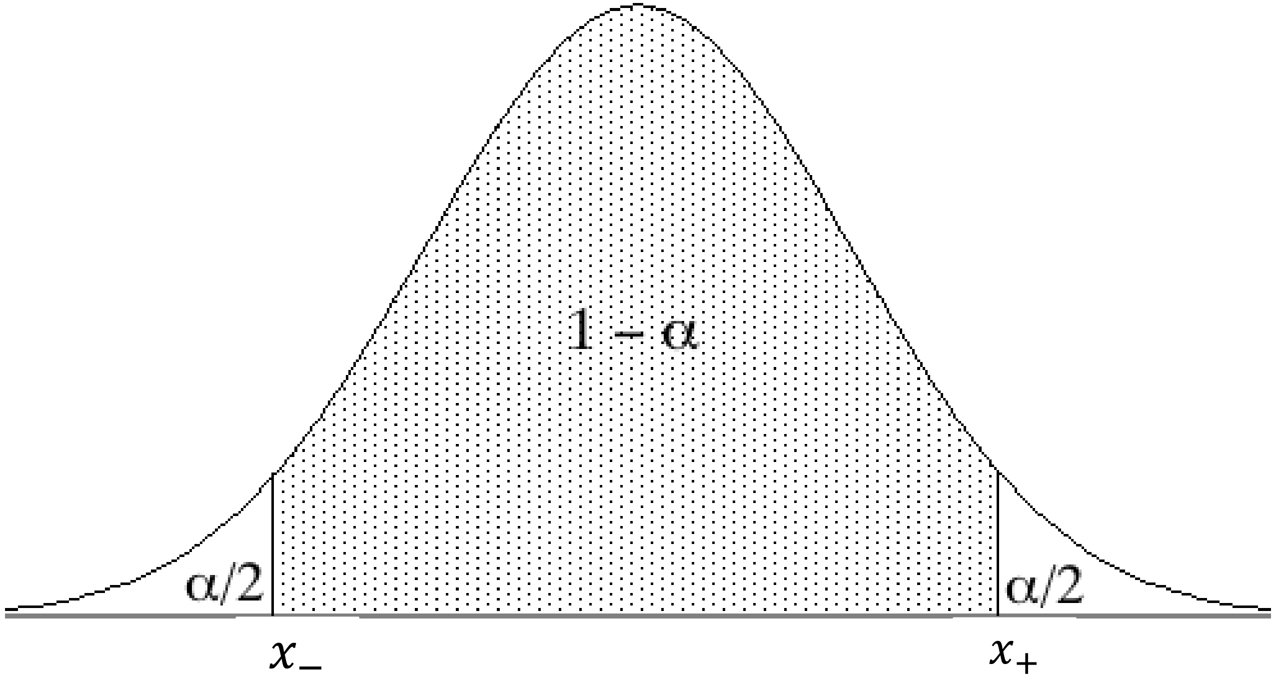
Puntualmente podemos estimarlo con la ley de los grandes números. Para un intervalo de confianza, igual que en el caso de una sola variable aleatoria.

Intervalo bilateral (dos colas)

Sean X, Y dos variables aleatorias con esperanzas μ_X, μ_Y y varianzas **conocidas** $0 < \sigma_X^2, \sigma_Y^2 < \infty$. Entonces podemos afirmar que $\mu_X - \mu_Y \in (x_-, x_+)$ con un nivel de confianza $1 - \alpha$ siendo:

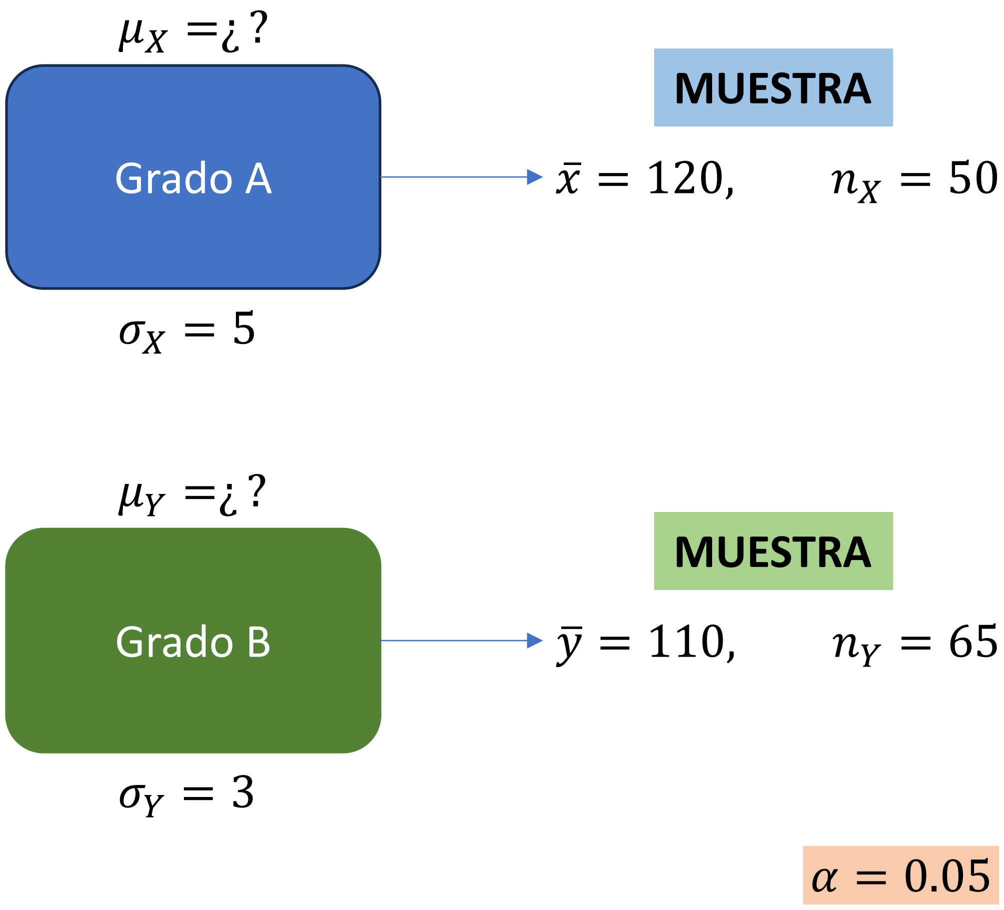
$$x_\pm = \bar{x}_X - \bar{x}_Y \mp z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

Donde $z_{\frac{\alpha}{2}}$ es el $\frac{\alpha}{2}$−cuantil de $N(0,1)$.



Ejemplo

En una universidad se estudia el CI de dos grados distintos. Para ello se ha tomado una muestra de 50 personas en el grado A obteniendo una media muestral de 120 y, otra muestra de 65 personas en el grado B resulta en una media muestral de 110. Teniendo en cuenta que la desviación poblacional del grado A es de 5 y en el grado B es de 3. Calcula el intervalo de confianza bilateral para la diferencia de medias con una significancia del 5%.



$\mu_X = ¿?$

Grado A

$\sigma_X = 5$

MUESTRA

$\bar{x} = 120, \quad n_X = 50$

$\mu_Y = ¿?$

Grado B

$\sigma_Y = 3$

MUESTRA

$\bar{y} = 110, \quad n_Y = 65$

$\alpha = 0.05$

Intervalo

$$\mu_X - \mu_Y \in (x_-, x_+)$$

$$x_\pm = \bar{x}_X \pm \bar{x}_Y \mp z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}$$

$(8.43, 11.57)$

$\mu_X > \mu_Y$
al 95%

Ejemplo

Sean X, Y dos variables aleatorias con esperanzas μ_X, μ_Y y varianzas **desconocidas** $0 < \sigma_X^2, \sigma_Y^2 < \infty$. Si $(X_1, \dots, X_n)$ e $(Y_1, \dots, Y_n)$ son dos muestras aleatorias simples.

¿Cómo podemos estimar $\mu_1 - \mu_2$?

Para un intervalo de confianza, necesitamos **normalidad** e **igualdad en las varianzas**.

Definición

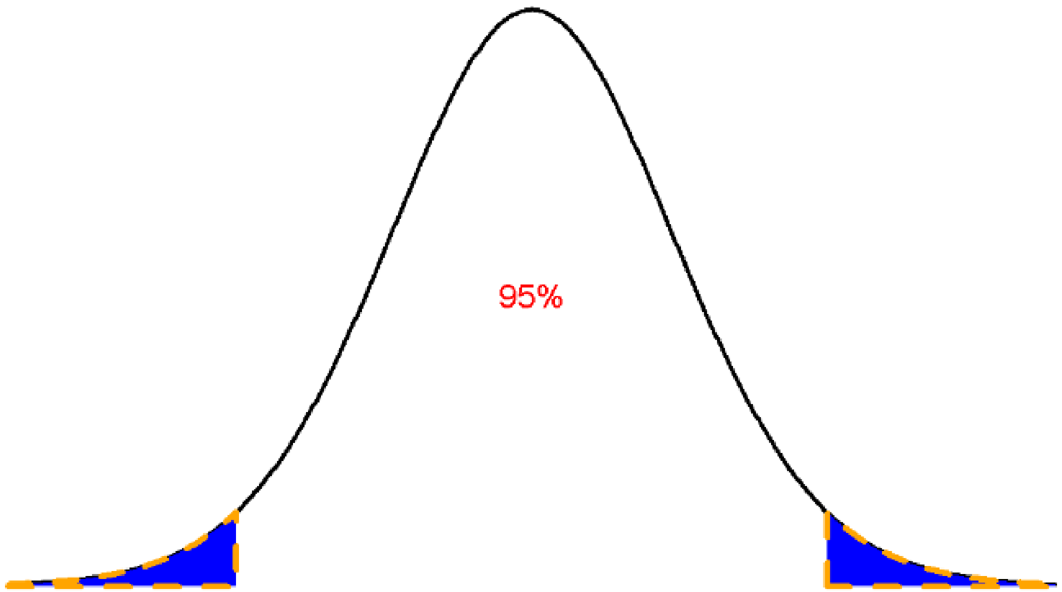
$$\hat{S}^2_{pooled} = \frac{(n_X - 1)\hat{S}_X^2 + (n_Y - 1)\hat{S}_Y^2}{n_X + n_Y - 2}$$

Intervalo bilateral (dos colas)

Sean $X \sim N(\mu_X, \sigma^2)$ e $Y \sim N(\mu_Y, \sigma^2)$ dos variables aleatorias normales. Entonces podemos afirmar que $\mu_X - \mu_Y \in (x_-, x_+)$ con un nivel de confianza $1 - \alpha$ siendo:

$$x_\pm = \bar{x}_X \pm \bar{x}_Y \mp t_{\frac{\alpha}{2}} \cdot \hat{s}_{pooled} \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$$

Donde $t_{\frac{\alpha}{2}}$ es el $\frac{\alpha}{2}$ −cuantil de $T_{n_X+n_Y-2}$.



Ejemplo

En una universidad se estudia el CI de dos grados distintos. Para ello se ha tomado una muestra de 50 personas en el grado A obteniendo una media muestral de 120 y una desviación muestral de 5 y, otra muestra de 65 personas en el grado B resulta en una media muestral de 110 y una desviación muestral de 3. Teniendo en cuenta que la desviación poblacional de ambos grados es la misma. Calcula el intervalo de confianza bilateral para la diferencia de medias con una significancia del 5%.

Asumimos normalidad.

$\mu_X =¿?$

Grado A

$\sigma_X = \sigma_Y$

MUESTRA

$\bar{x} = 120, \hat{s}_X = 5, n_X = 50$

$\mu_Y =¿?$

Grado B

$\sigma_Y = \sigma_X$

MUESTRA

$\bar{y} = 110, \hat{s}_Y = 3, n_Y = 65$

$\alpha = 0.05$

$\mu_X = ¿?$

Grado A

$\sigma_X = \sigma_Y$

MUESTRA

$\bar{x} = 120, \hat{s}_X = 5, n_X = 50$

$\mu_Y = ¿?$

Grado B

$\sigma_Y = \sigma_X$

MUESTRA

$\bar{y} = 110, \hat{s}_Y = 3, n_Y = 65$

$\alpha = 0.05$

Intervalo

$$\mu_X - \mu_Y \in (x_-, x_+)$$

$$x_\pm = \bar{x}_X \pm \bar{x}_Y \mp t_{\frac{\alpha}{2}} \cdot \hat{s}_{pooled} \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}$$

$(8.51, 11.49)$

$\mu_X > \mu_Y$

al 95%

Ejemplo

Sean X, Y dos variables aleatorias con esperanzas μ_X, μ_Y y varianzas $0 < \sigma_X^2, \sigma_Y^2 < \infty$. Si $(X_1, \ldots, X_n)$ e $(Y_1, \ldots, Y_n)$ son dos muestras aleatorias simples.

¿Cómo podemos estimar $\frac{\sigma_X^2}{\sigma_Y^2}$?

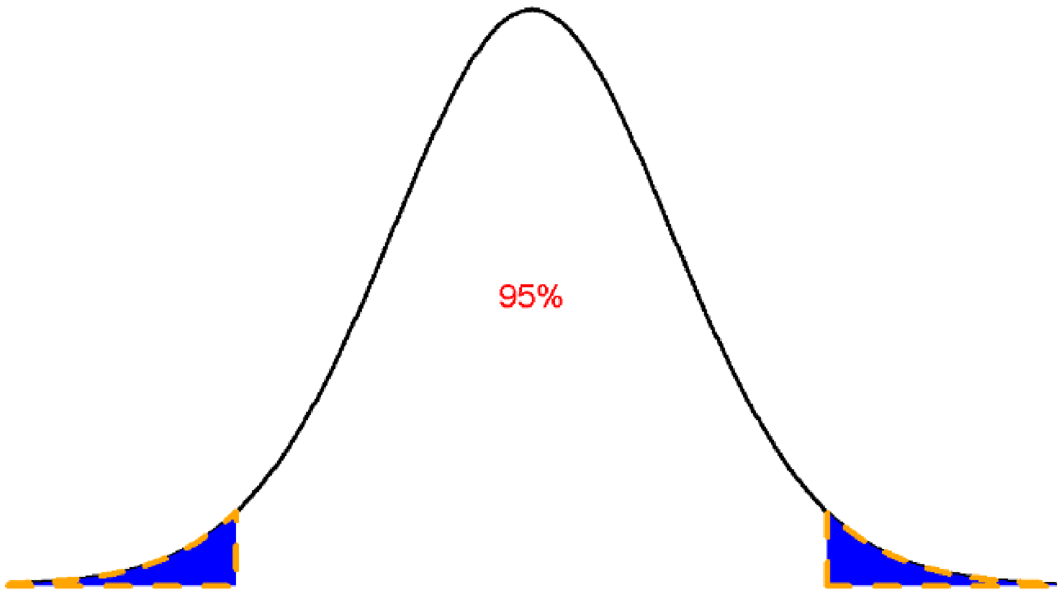
Para un intervalo de confianza, necesitamos **normalidad**.

Intervalo bilateral (dos colas)

Sean $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$ dos variables aleatorias normales. Entonces, con un nivel de confianza $1-\alpha$:

$$\frac{1}{f_{1-\frac{\alpha}{2}}} \frac{\hat{s}_X^2}{\hat{s}_Y^2} \leq \frac{\sigma_X^2}{\sigma_Y^2} \leq \frac{1}{f_{\frac{\alpha}{2}}} \frac{\hat{s}_X^2}{\hat{s}_Y^2}$$

Donde f_γ es el γ −cuantil de F_{n_X-1, n_Y-1}.



Asumimos normalidad.

Ejemplo

En una empresa, se están comparando dos métodos de producción de cierto chip (A, mucho más barato, y B). La potencia media consumida por ambos chips es idéntica, si bien los dos métodos tienen distinta variabilidad. Se obtienen dos muestras de tamaño 16 y 10, sus varianzas muestrales son 24 y 18. Calcula un intervalo de confianza bilateral para la ratio de las varianzas con una significancia del 5%.

$\mu_X =¿?$

Grado A → MUESTRA: $\hat{s}_X^2 = 24, \quad n_X = 16$

$\sigma_X =¿?$

$\mu_Y =¿?$

Grado B → MUESTRA: $\hat{s}_Y^2 = 18, \quad n_Y = 10$

$\sigma_Y =¿?$

$\alpha = 0.05$

$\mu_X = ¿?$

Grado A

$\sigma_X = ¿?$

MUESTRA

$\hat{s}_X^2 = 24, \quad n_X = 16$

$\mu_Y = ¿?$

Grado B

$\sigma_Y = ¿?$

MUESTRA

$\hat{s}_Y^2 = 18, \quad n_Y = 10$

$\alpha = 0.05$

Intervalo

$$\frac{1}{f_{1-\frac{\alpha}{2}}} \frac{\hat{s}_X^2}{\hat{s}_Y^2} \leq \frac{\sigma_X^2}{\sigma_Y^2} \leq \frac{1}{f_{\frac{\alpha}{2}}} \frac{\hat{s}_X^2}{\hat{s}_Y^2}$$

$$\frac{\sigma_X^2}{\sigma_Y^2} \in (0.35, 4.16)$$

Ejemplo

Sea $X \sim Pois(\lambda)$. Si $(X_1, \dots, X_n)$ es una muestra aleatoria simple.

¿Cómo podemos estimar λ?

Para un intervalo de confianza, haremos uso de la chi-cuadrado y la suma muestral $n\bar{X}$ o una aproximación mediante la normal usando el TLC.

CEU Universidad San Pablo

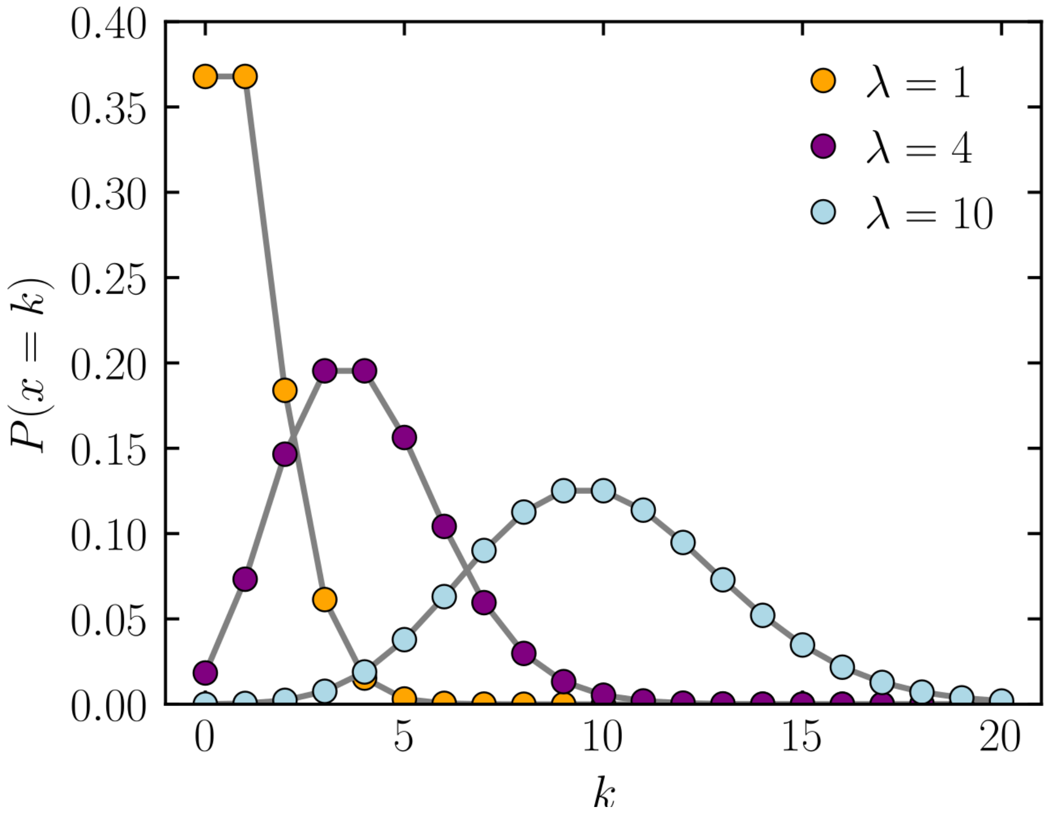
Intervalo bilateral (dos colas)

Sea $X \sim Pois(\lambda)$ y $(X_1, \dots, X_n)$ una m.a.s. Entonces $\lambda \in (L_1, L_2)$ con un nivel de confianza $1-\alpha$ siendo:

$$L_1 = \min\{\lambda_1, \lambda_2\}, \qquad L_2 = \max\{\lambda_1, \lambda_2\}$$

$$\lambda_1 = \frac{1}{2n}\chi^2_{\frac{\alpha}{2};2n\bar{x}}, \qquad \lambda_2 = \frac{1}{2n}\chi^2_{1-\frac{\alpha}{2};2n\bar{x}+2}$$

Donde $\chi^2_{\frac{\alpha}{2};2n}$ es el $\frac{\alpha}{2}$ −cuantil de χ^2_{2n} y $\chi^2_{1-\frac{\alpha}{2};2n+2}$ es el $\left(1-\frac{\alpha}{2}\right)$ −cuantil deχ^2_{2n+2}

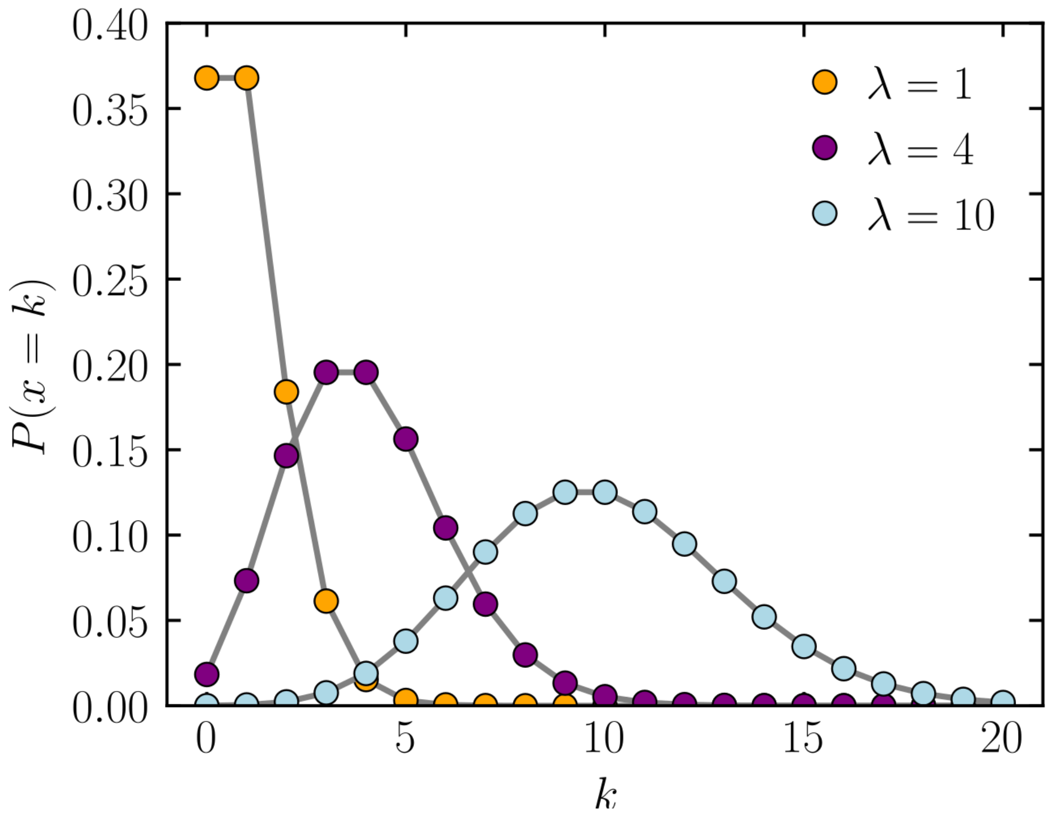


Ejemplo

En la liga de fútbol española se han recogido datos de goles marcados por un equipo.

¿Intervalo para el nº goles medio?

Significancia de $\alpha = 0.05$.



$$\lambda \in (1.13, 1.45)$$

Ejemplo

Sea $X \sim Ber(p)$. Si $(X_1, \dots, X_n)$ es una muestra aleatoria simple.

¿Cómo podemos estimar p?

Para un intervalo de confianza, haremos uso de la F Snedecor y la suma muestral $n\bar{X}$ o una aproximación mediante la normal usando el TLC.

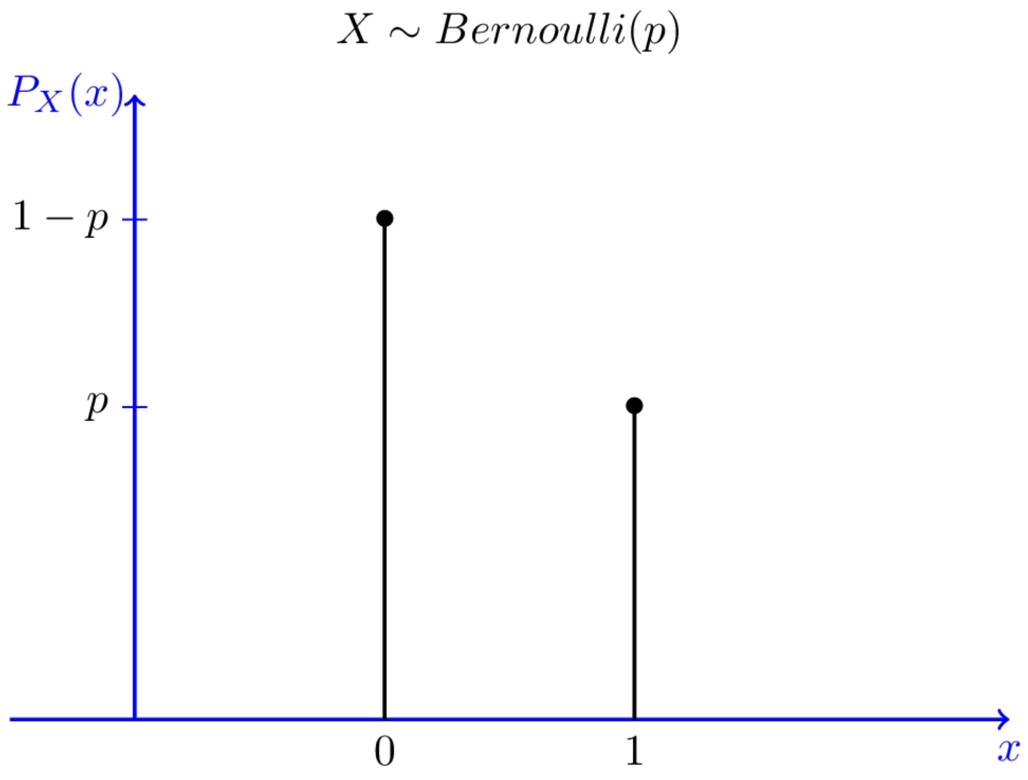
Intervalo bilateral (dos colas)

Sea $X \sim Ber(p)$ y $(X_1, \ldots, X_n)$ una m.a.s. Entonces $p \in (L_1, L_2)$ con un nivel de confianza $1 - \alpha$ siendo:

$$L_1 = \min\{p_1, p_2\}, \qquad L_2 = \max\{p_1, p_2\}$$

$$p_1 = \frac{n\bar{x}F_1}{(n - n\bar{x} + 1) + n\bar{x}F_1}, p_2 = \frac{(n\bar{x} + 1)F_2}{(n - n\bar{x}) + (n\bar{x} + 1)F_2}$$

Donde F_1 es el $\frac{\alpha}{2}$ −cuantil de $F_{2n\bar{x},2(n-n\bar{x}+1)}$ y F_2 es el $\left(1 - \frac{\alpha}{2}\right)$ −cuantil de $F_{2(n\bar{x}+1),2(n-n\bar{x})}$.

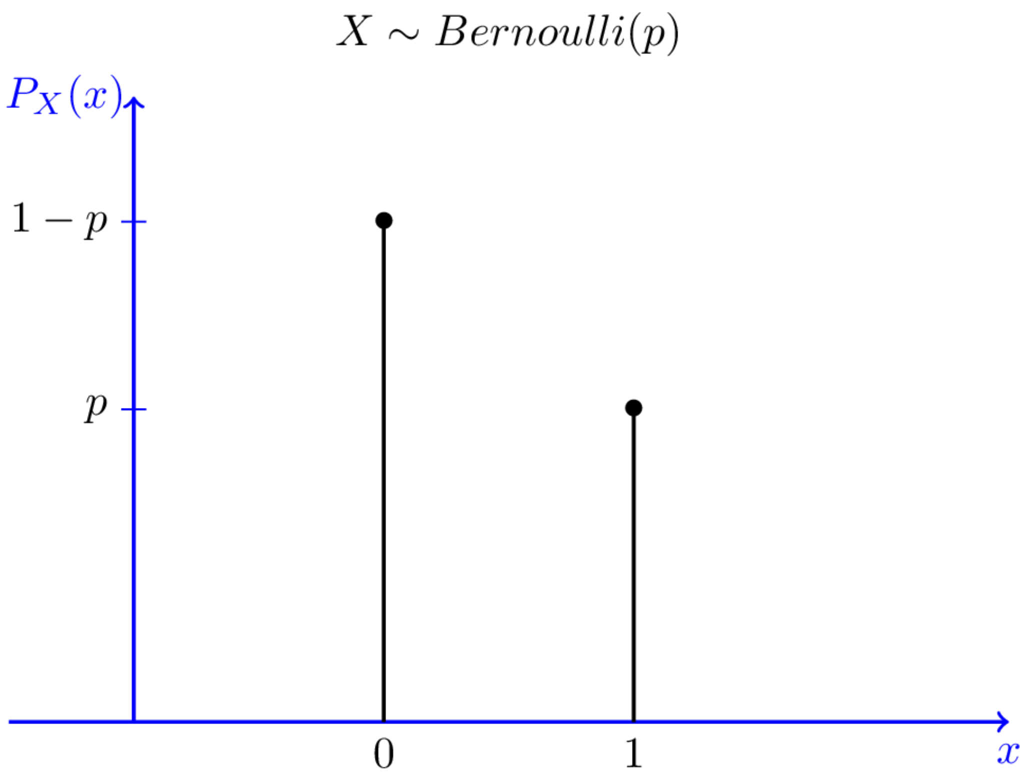


Ejemplo

Un amigo tuyo dice tener una moneda "trucada" pero no está seguro de conocer la probabilidad de cara en la moneda. Te decides a realizar 200 experimentos para estimar su probabilidad.

¿Intervalo para p?

Significancia de $\alpha = 0.05$.



$$p \in (0.68, 0.81)$$

CEU Universidad San Pablo

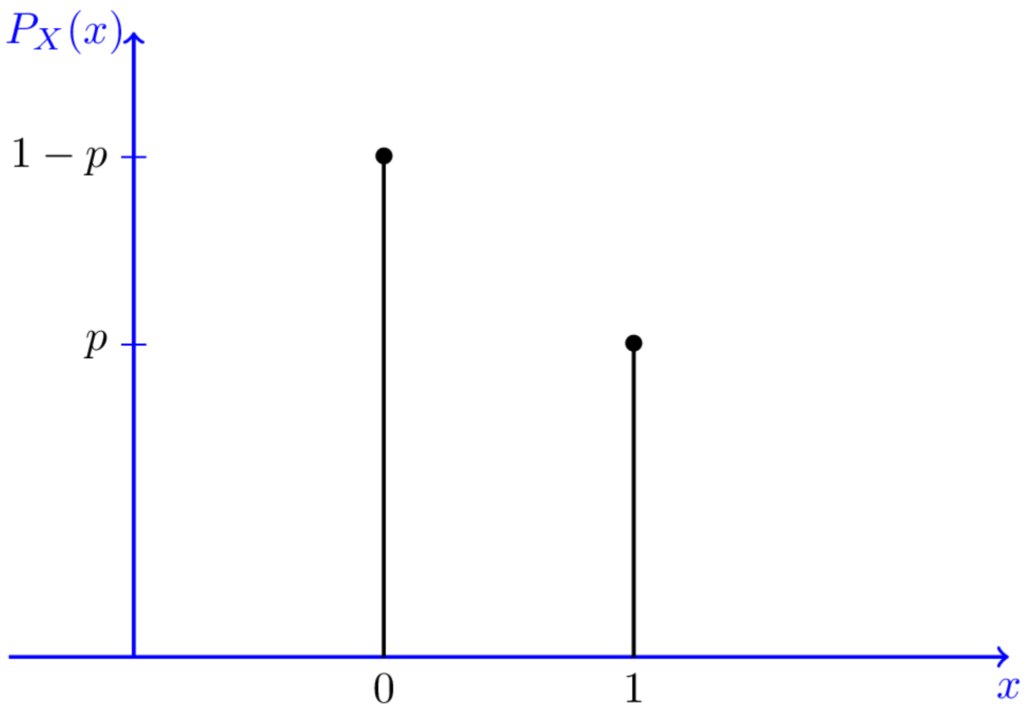
Ejemplo

Durante los 60s-70s, se dieron casos de racismo en la elección de jurados populares. Supuestamente, la elección es al azar entre un listado de todos los ciudadanos. Sin embargo, se daban situaciones como que en una preselección de 80 posibles jurados solo 4 fuesen afroamericanos (de una población con un 50 % de afroamericanos). Las autoridades se defendían diciendo que era pura casualidad.

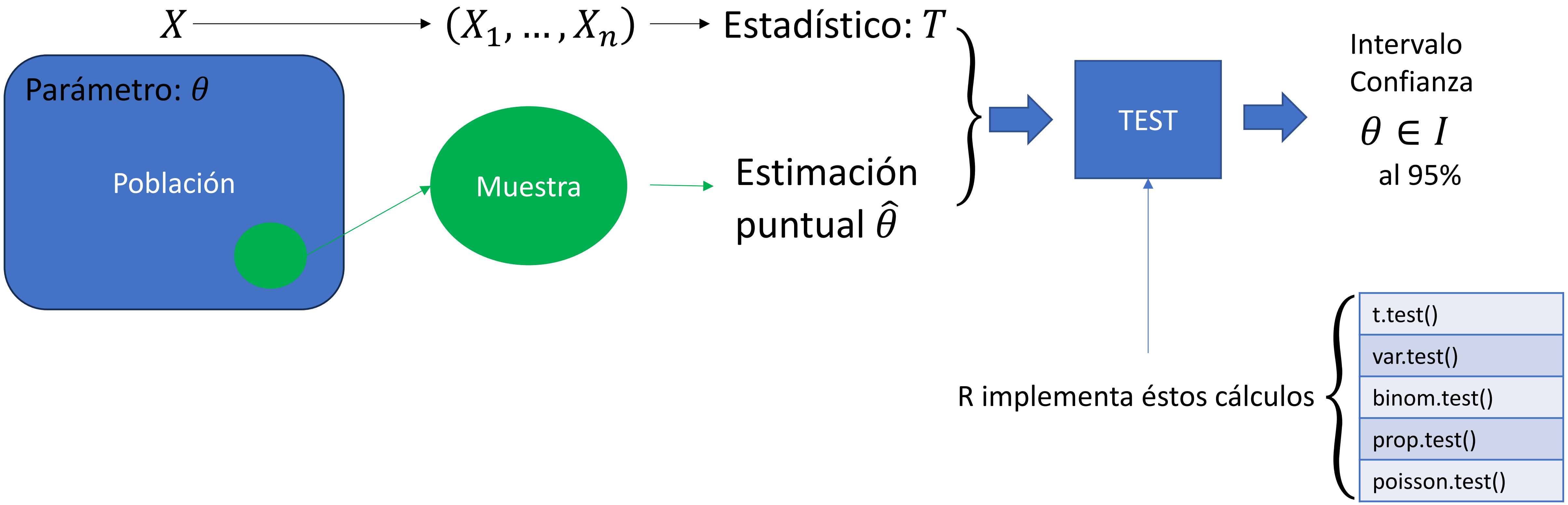
¿Intervalo para $p =$“prob. afroamericano”?

Significancia de $\alpha = 0.05$.

$X \sim Bernoulli(p)$



$p \in (0.01, 0.12)$


X
$(X_1, \dots, X_n)$
Estadístico: T
Parámetro: θ
Población
Muestra
Estimación puntual $\hat{\theta}$
TEST
Intervalo Confianza
$\theta \in I$
al 95%
R implementa éstos cálculos
t.test()
var.test()
binom.test()
prop.test()
poisson.test()