T5. Estimación por intervalos

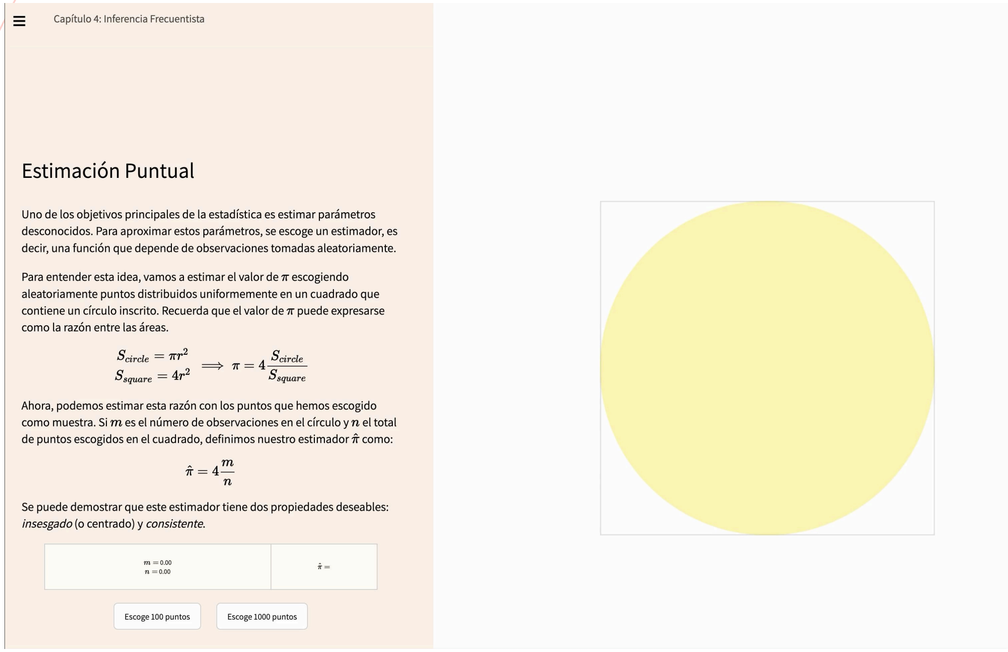
Estadística 2 – URJC

Grado de ADE y ECO

Índice

5.1. Introducción

- **Problemática:** En la estimación puntual (tema 3) <u>no sabemos si $\hat{\theta}$ es próximo θ</u> dada la muestra $\underline{X}$. ¿Cuánto nos equivocamos?



https://seeing-theory.brown.edu/es.html

5.1. Introducción

- **Problemática:** En la estimación puntual (tema 3) no sabemos si $\hat{\theta}$ es próximo θ dada la muestra $\underline{X}$. ¿Cuánto nos equivocamos?
- **Solución:** Acompañamos a la estimación puntual $\hat{\theta}$ de un intervalo que cumple:

$$\boxed{\left[\underline{\theta}(X) \leq \theta \leq \overline{\theta}(X)\right]}$$

$$P\left[\underline{\theta}(X) \leq \theta \leq \overline{\theta}(X)\right] = 1 - \alpha$$

- $\underline{\theta}(X)$: Límite inferior
- $\overline{\theta}(X)$:Límite superior
- $\gamma = 1 - \alpha$: nivel de confianza
- α : nivel de significación

5.1. Introducción

$$P\left[\underline{\theta}(X) \leq \theta \leq \overline{\theta}(X)\right] = 1 - \alpha$$

- $\left[\underline{\theta}(X) \leq \theta \leq \overline{\theta}(X)\right]$ Aleatorio: depende de la muestra X que ha surgido en el $m.a.s(n)$.
  - $\underline{\theta}(X)$ y $\overline{\theta}(X)$ son VARIABLES ALEATORIAS que dependen del resultado del $m.a.s.(n) \rightarrow X$
  - θ no es variable aleatoria, por tanto..
- $\gamma = 1 - \alpha$ ó nivel de confianza: probabilidad de que I.C. incluya a θ
- α ó nivel de significación : probabilidad de que I.C. no incluya a θ

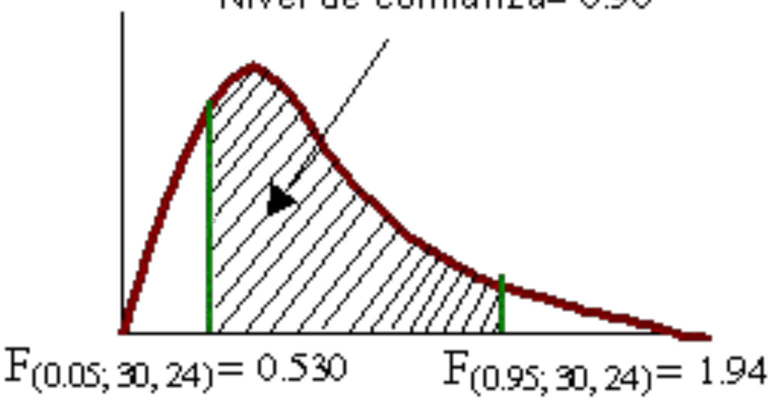
5.1. Introducción

(Martín Pliego,p.132)

$$P\{\theta \in [\underline{\theta}(\mathbf{X}); \overline{\theta}(\mathbf{X})]\} = 1 - \alpha .$$

Seleccionada una muestra **X,** la probabilidad de que el parámetro θ esté incluido en el intervalo $[\underline{\theta}(\mathbf{X}); \overline{\theta}(\mathbf{X})]$ es 1 o 0, dependiendo de que el parámetro θ esté o no esté entre los dos números en que se convierten $\underline{\theta}(\mathbf{X})$ y $\overline{\theta}(\mathbf{X})$, al particularizarlos para la muestra **X** concreta.

En esta situación no se puede hablar de probabilidad del intervalo al nivel $1-\alpha$ sino de **confianza** puesto que, una vez extraída la muestra **X**, la probabilidad será 1 o 0, y no la inicial $1-\alpha$ que se transforma en confianza.

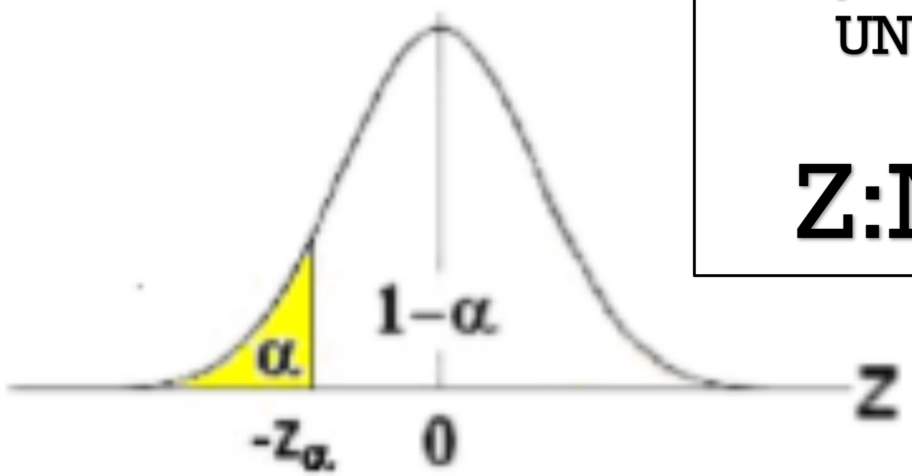


Ejemplo de I.C.
BILATERAL
en una
$F_{m,n}$

5.1. Introducción

En algunos casos específicos puede interesar construir intervalos de confianza **unilaterales**. Para ello se hace $\underline{\theta}(\mathbf{X}) = 0$, con lo que el intervalo de confianza adopta la forma $[0; \overline{\theta}(\mathbf{X})]$, o bien $\overline{\theta}(\mathbf{X}) = 0$ y el intervalo es $[\underline{\theta}(\mathbf{X}); 0]$.

(Martín Pliego,p.133)



Ejemplo de I.C.
UNILATERAL
en una
Z:N(0,1)

5.2. Métodos de construcción de I.C.

- X **obtenida por** $m.a.s.(n)$ **extraída de una población** $\xi : F[X, \theta]$ **donde** $\theta \in \Theta$**, siendo** $\Theta \in \mathbb{R}$**.**
- $T[X, \theta]$ **Pivote o cantidad pivotal :**
  - $T[X, \theta]$ **depende de** θ**,**
  - **Su distribución de probabilidad no depende de** θ**.**
  - $T[X, \theta]$ **es estadístico donde** θ **es un valor fijo.**

Ejemplo:

Sea una población $\xi, : N(\mu, \sigma)$, sabemos que $\boldsymbol{\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$ entonces el pivote será

$$\boxed{T[X, \theta] = \frac{\bar{x} - \mu}{\sigma} \rightarrow Z : N(0,1)}$$

5.2.1. Método de la cantidad pivotal

(Martín Pliego,p.134)

TEOREMA

Si la cantidad pivotal $T(\mathbf{X};\theta)$ es función monótona de θ, es posible determinar un intervalo de confianza para el parámetro θ.

En efecto, para un nivel de confianza $1-\alpha$ se pueden elegir un par de valores $K_1(\alpha)$ y $K_2(\alpha)$, pertenecientes al campo de variación de $T(\mathbf{X};\theta)$, tales que

$$P[K_1(\alpha) \leq T(\mathbf{X};\theta) \leq K_2(\alpha)] = 1-\alpha, \quad \text{para todo} \quad \theta \in \Theta.$$

Si $T(\mathbf{X};\theta)$ es monótona en θ, se pueden resolver las ecuaciones

$$T(\mathbf{X};\theta) = K_1(\alpha)$$
$$T(\mathbf{X};\theta) = K_2(\alpha)$$

obteniéndose los límites $\underline{\theta}(\mathbf{X})$ y $\overline{\theta}(\mathbf{X})$ al despejar θ, con lo que el intervalo buscado es

$$P[\underline{\theta}(\mathbf{X}) \leq \theta \leq \overline{\theta}(\mathbf{X})] = 1-\alpha.$$

La resolución a través del sistema de ecuaciones suele ser "compleja"… ver ejemplo, (Martín Pliego,p.134-136)… por eso se propone el método que describimos a continuación.

5.2.2. Método general de construcción de I.C.

- **Novedad:** Prescindimos de la exigencia de que la distribución de probabilidad del pivote dependa del parámetro.
- **Método:** Sea una población $\xi: f(x,\theta)$ en la que queremos construir un I.C. bilateral para θ
  1. Elegimos un estimador del parámetro θ, $\hat{\theta} = T(X)$ cuya función de probabilidad es $g(\hat{\theta};\theta)$

     $\boxed{\hat{\theta} = T(X): g(\hat{\theta};\theta)}$
  2. Dado un nivel de confianza $\gamma = 1 - \alpha$, se determinan los extremos de un intervalo $k_1(\alpha,\theta)$ y $k_2(\alpha,\theta)$, tales que

     $$\boxed{P\left[k_1(\alpha,\theta) \leq \hat{\theta} \leq k_2(\alpha,\theta)\right] = 1 - \alpha}$$
  3. Si ξ es continua

     $$\int_{k_1(\alpha,\theta)}^{k_2(\alpha,\theta)} g\left(\hat{\theta};\theta\right) d\,\hat{\theta} = 1 - \alpha$$

5.2.2. Método general de construcción de I.C.

$$\int_{k_1(\alpha,\theta)}^{k_2(\alpha,\theta)} g(\hat{\theta};\theta)d\,\hat{\theta} = 1-\alpha$$

(Martín Pliego,p.137)

O bien, siendo

$$\alpha = \alpha_1 + \alpha_2$$

$$\int_{k_1(\alpha,\theta)}^{\infty} g(\hat{\theta};\theta)d\,\hat{\theta} = \alpha_1$$

$$\int_{-\infty}^{k_2(\alpha,\theta)} g(\hat{\theta};\theta)d\,\hat{\theta} = \alpha_2$$

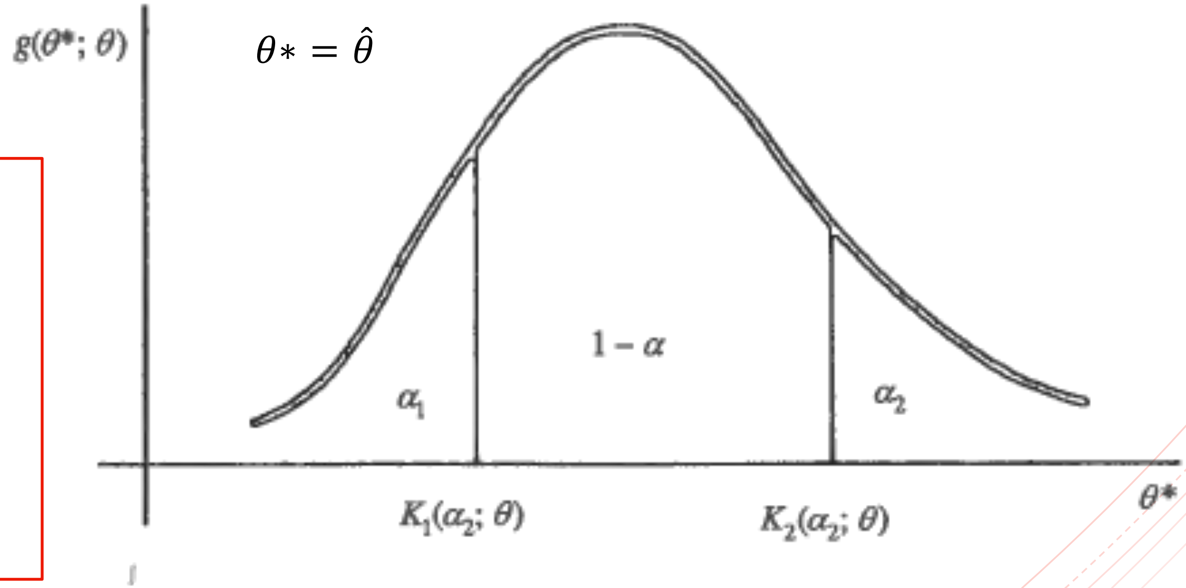


FIGURA 5.1

5.2.2. Método general de construcción de I.C.

$$\int_{k_1(\alpha,\theta)}^{\infty} g(\hat{\theta};\theta)d\,\hat{\theta} = \alpha_1$$

$$\int_{-\infty}^{k_2(\alpha,\theta)} g(\hat{\theta};\theta)d\,\hat{\theta} = \alpha_2$$

De aquí obtenemos los límites $k_1(\alpha,\theta)$ y $k_2(\alpha,\theta)$

y el sistema de ecuaciones integrales

$$\left.\begin{array}{l} k_1(\alpha,\theta) = \hat{\theta} \\ k_2(\alpha,\theta) = \hat{\theta} \end{array}\right\}$$

Y la solución al sistema anterior, nos da los extremos del intervalo de confianza

$$\left[\underline{\theta}(X) \leq \theta \leq \overline{\theta}(X)\right]$$

Si la población es discreta ξ: $P(x_i,\theta)$ no siempre será posible obtener un nivel de confianza exactamente igual a $1-\alpha$ y deberemos seleccionar uno aproximado.

5.2.3. I.C. de longitud mínima

- **Problema**: podemos obtener un número infinito de I.C. que verifiquen un nivel de confianza $1-\alpha$, pues los valores que cumplen $\alpha = \alpha_1 + \alpha_2$, son infinitos.

De todos es PREFERIBLE EL DE AMPLITUD MÍNIMA →Δ PRECISIÓN.

- Solución 1: **Minimizamos la longitud** del I.C. por el método de optimización condicionada de los multiplicadores de Lagrange:

$$\boxed{\boldsymbol{MIN\ D = \overline{\theta}(X) - \underline{\theta}(X)}\ s.a.\ P\left[\underline{\theta}(X) \leq \theta \leq \overline{\theta}(X)\right] = 1-\alpha}$$

- Pero, no siempre es posible (ej: distribuciones asimétricas), entonces, aplicamos el convenio de $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$

5.2.3. I.C. de longitud mínima

De todos los I.C. es PREFERIBLE EL DE AMPLITUD MÍNIMA →Δ PRECISIÓN.

- Solución 2: **Minimizamos la longitud esperada** del I.C. Por el método de optimización condicionada de los multiplicadores de Lagrange

$$\boxed{\boldsymbol{MIN\ E(D) = E\left[\overline{\theta}(X) - \underline{\theta}(X)\right]}\ s.a.\ P\left[\underline{\theta}(X) \leq \theta \leq \overline{\theta}(X)\right] = 1 - \alpha}$$

- Pero, tampoco resuelve todos los casos… y entonces, aplicamos el convenio de $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ que coincide en muchas ocasiones con el intervalo de longitud mínima (distribuciones simétricas).

5.3. I.C. en poblaciones normales.

- El método se simplifica: es relativamente sencillo encontrar pivotes.
- Se analizan los casos con una o dos muestras aleatorias simples, y se estudian los I.C. para μ y σ.
- Procedimiento:
  1. **Selección del pivote**: estadístico estimador del parámetro (que puede tener relación con los parámetros de la población).
  2. **Distribución del pivote**: Por las propiedades de la normal, este pivote tendrá una distribución de probabilidad conocida (aunque desconozcamos los parámetros).
  3. **Construimos el I.C**.
     1. Los Extremos K_1 y K_2 serán valores concretos de la V.A. con la que se relaciona el pivote (Z, t_n, $F_{m,n}$, …)
     2. Operamos con las relaciones entre el pivote y los extremos en intervalo K_1 y K_2, hasta dejar "en el centro de la inecuación" al parámetro poblacional.

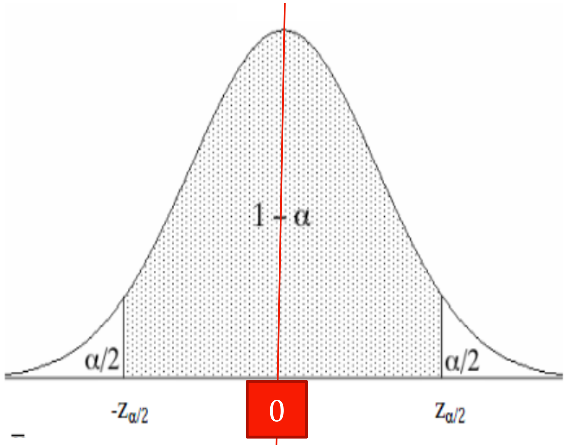
5.3.1. I.C. para μ siendo $\xi: N(\mu, \sigma)$ conocida σ^2 con m.a.s.(n)

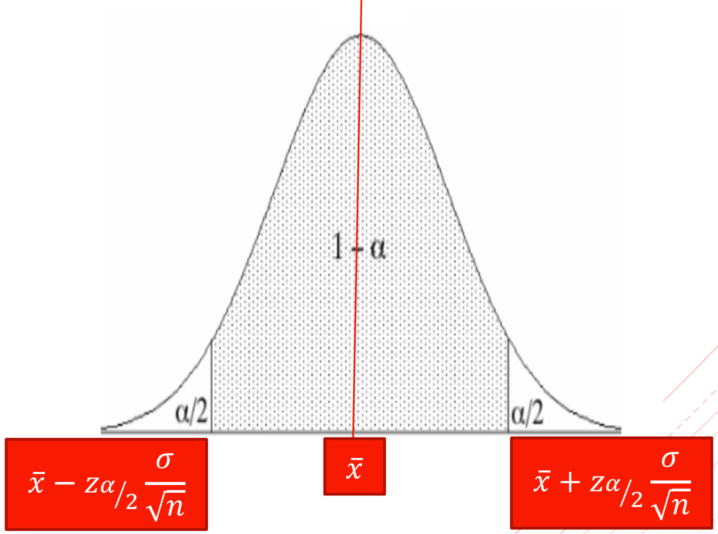
Et	Etapa	
1	**Selección del pivote:**	$T(X;\mu) = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = z$
2	**Distribución del pivote**	$\bar{x}: N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$
3. Construimos el I.C.	3.1. Intervalo teórico, K_1 y K_2 para $1-\alpha$	$P[k_1 \leq z \leq k_2] = 1-\alpha$ $k_1 = -z\alpha/_2; k_1 = z\alpha/_2$
	3.2 Despejamos μ en el I.C.	

$$P[k_1 \leq z \leq k_2] = P\left[k_1 \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq k_2\right] =$$

$$= P\left[-z\alpha/_2 \frac{\sigma}{\sqrt{n}} \leq \bar{x}-\mu \leq z\alpha/_2 \frac{\sigma}{\sqrt{n}}\right] = P\left[-z\alpha/_2 \frac{\sigma}{\sqrt{n}} - \bar{x} \leq -\mu \leq z\alpha/_2 \frac{\sigma}{\sqrt{n}} - \bar{x}\right] =$$

$$\boxed{= P\left[\bar{x} - z\alpha/_2 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z\alpha/_2 \frac{\sigma}{\sqrt{n}}\right] = 1-\alpha}$$





5.3.1. I.C. para μ siendo $\xi: N(\mu, \sigma)$ conocida σ^2 con m.a.s.(n)

EJEMPLO 4

De una población $N(\mu;\ 3)$ se extrae una muestra aleatoria simple de tamaño 25 siendo $\sum_{i=1}^{25} x_i = 60$.

Determínese el intervalo de confianza del 95% para la media de la población.

Tenemos que

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{60}{25} = 2{,}4$$

y, por otra parte, en una distribución $N(0;\ 1)$ si

$$P[-K \leq N(0;\ 1) \leq K] = 0{,}95$$

entonces $K = 1{,}96$.

El intervalo de confianza del 95% para la media μ resulta

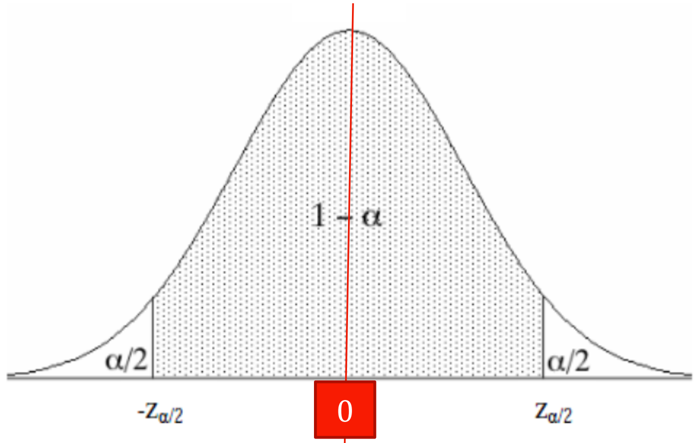
$$\left[2{,}4 - 1{,}96\,\frac{3}{\sqrt{25}};\ 2{,}4 + 1{,}96\,\frac{3}{\sqrt{25}}\right],$$

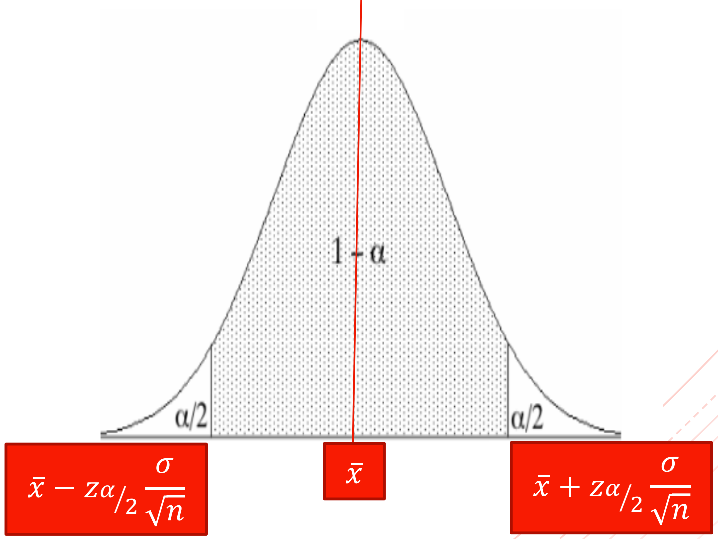
es decir, la media μ se encontrará incluida previsiblemente en el intervalo

$$[1{,}224;\ 3{,}576]$$

con una confianza del 95%.

(Martín Pliego, p. 144)





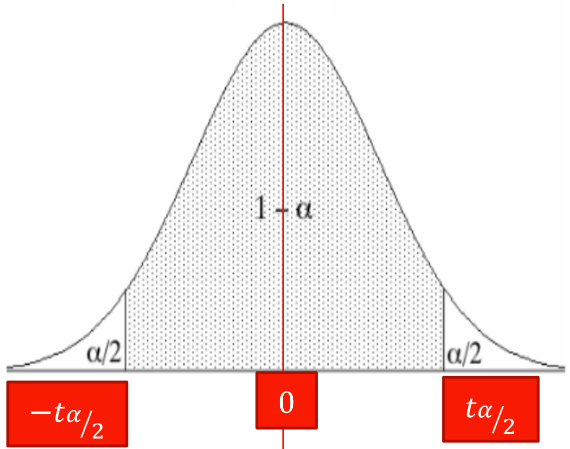
5.3.2. I.C. para μ siendo $\xi: N(\mu, \sigma)$ desconocida σ^2 con m.a.s.(n)

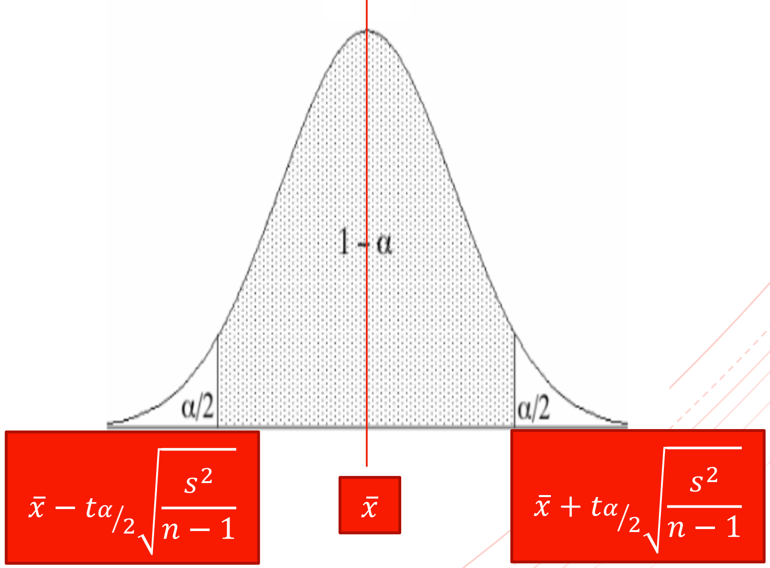
Et	Etapa	
1	**Selección del pivote:**	$T(X;\mu) = \dfrac{\bar{x} - \mu}{\sqrt{\dfrac{s^2}{n-1}}} = t_{n-1}$
2	**Distribución del pivote**	$\bar{x}: \mu + t_{n-1}\sqrt{\dfrac{s^2}{n-1}}$
3.	3.1. Intervalo teórico, K_1 y K_2 para $1-\alpha$	$P[k_1 \le t_{n-1} \le k_2] = 1 - \alpha$ $k_1 = -t\alpha/_2 ; k_1 = t\alpha/_2$

$$P[k_1 \le z \le k_2] = P\left[-t\alpha/_2 \le \frac{\bar{x} - \mu}{\sqrt{\dfrac{s^2}{n-1}}} \le t\alpha/_2\right] =$$

$$= P\left[-t\alpha/_2\sqrt{\frac{s^2}{n-1}} \le \bar{x} - \mu \le t\alpha/_2\sqrt{\frac{s^2}{n-1}}\right] = P\left[-t\alpha/_2\sqrt{\frac{s^2}{n-1}} \le -\mu \le t\alpha/_2\sqrt{\frac{s^2}{n-1}} - \bar{x}\right] =$$

$$\boxed{= P\left[\bar{x} - t\alpha/_2\sqrt{\frac{s^2}{n-1}} \le \mu \le \bar{x} + t\alpha/_2\sqrt{\frac{s^2}{n-1}}\right] = 1 - \alpha}$$





5.3.2. I.C. para μ siendo $\xi: N(\mu, \sigma)$ desconocida σ^2 con m.a.s.(n)

EJEMPLO 5

En una población $N(\mu; \sigma)$, donde se desconoce el valor del parámetro σ, se obtiene una muestra aleatoria simple de tamaño 10, en la que

$$\sum_{i=1}^{10} x_i = 41 \quad \text{y} \quad \sum_{i=1}^{10} x_i^2 = 229.$$

Hállese el intervalo de confianza del 90% para la media de la población.

Como

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{41}{10} = 4{,}1$$

$$s^2 = a_2 - \bar{x}^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - (4{,}1)^2 = \frac{229}{10} - (4{,}1)^2 = 6{,}09$$

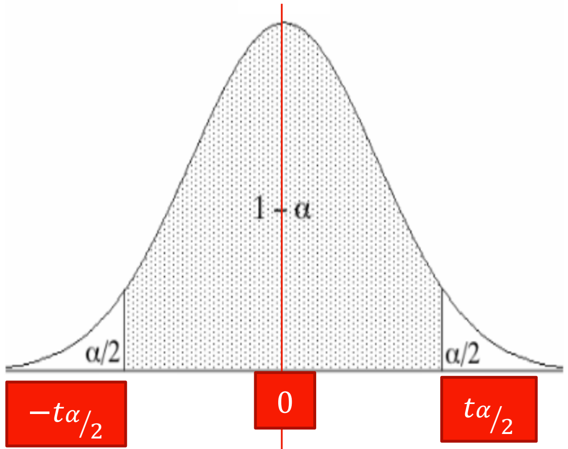
y dado que en una distribución $t(n-1) = t(9)$ el intervalo del 90% se obtiene para un valor t tal que

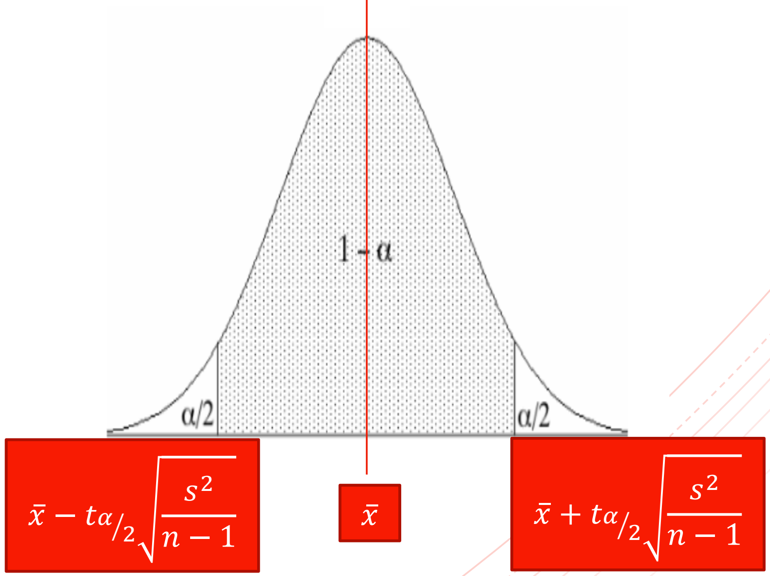
$$P[-t \leq t(9) \leq t] = 0{,}90$$
$$t = 1{,}8331.$$

El intervalo de confianza de longitud mínima para μ en este caso resulta

$$\left[4{,}1 - 1{,}8331\sqrt{\frac{6{,}09}{9}};\ 4{,}1 + 1{,}8331\sqrt{\frac{6{,}09}{9}}\right] = [2{,}592;\ 5{,}608].$$

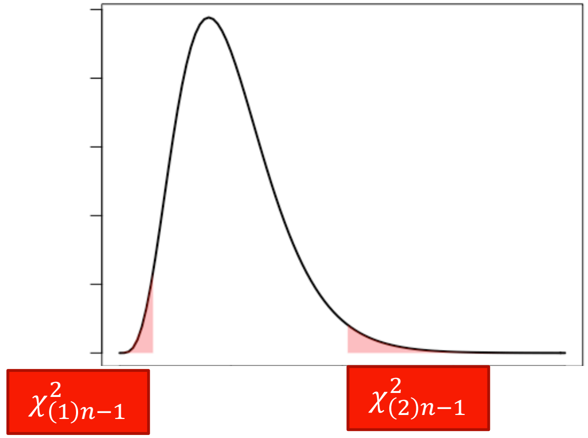
(Martín Pliego, p.146)





5.3.3. I.C. para σ^2 siendo $\xi: N(\mu, \sigma)$

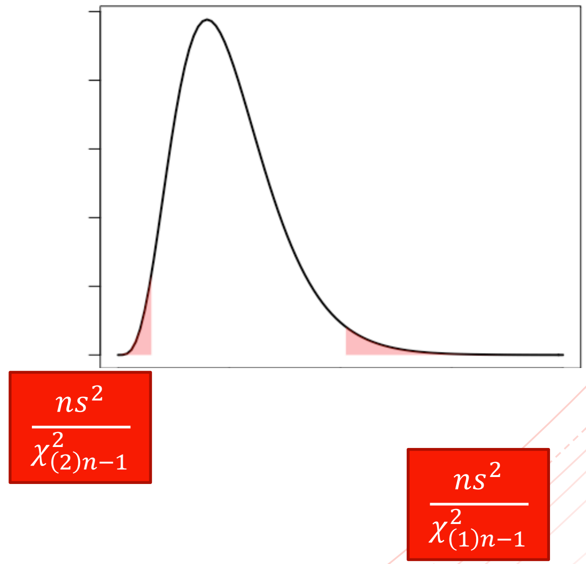
Et	Etapa	
1	**Selección del pivote:**	$T(X;\sigma^2) = \frac{ns^2}{\sigma^2} = \chi^2_{n-1}$
2	**Distribución del pivote**	$s^2 = \frac{\sigma^2}{n}\chi^2_{n-1}$
3.	3.1. Intervalo teórico, K_1 y K_2 para $1-\alpha$	$P[k_1 \leq z \leq k_2] = 1-\alpha$ $k_1 = \chi^2_{(1)n-1}; k_1 = \chi^2_{(2)n-1}$



$$P[k_1 \leq \chi^2_{n-1} \leq k_2] = P\left[\chi^2_{(1)n-1} \leq \frac{ns^2}{\sigma^2} \leq \chi^2_{(2)n-1}\right] =$$

$$= P\left[\frac{1}{ns^2}\chi^2_{(1)n-1} \leq \frac{1}{\sigma^2} \leq \frac{1}{ns^2}\chi^2_{(2)n-1}\right] =$$

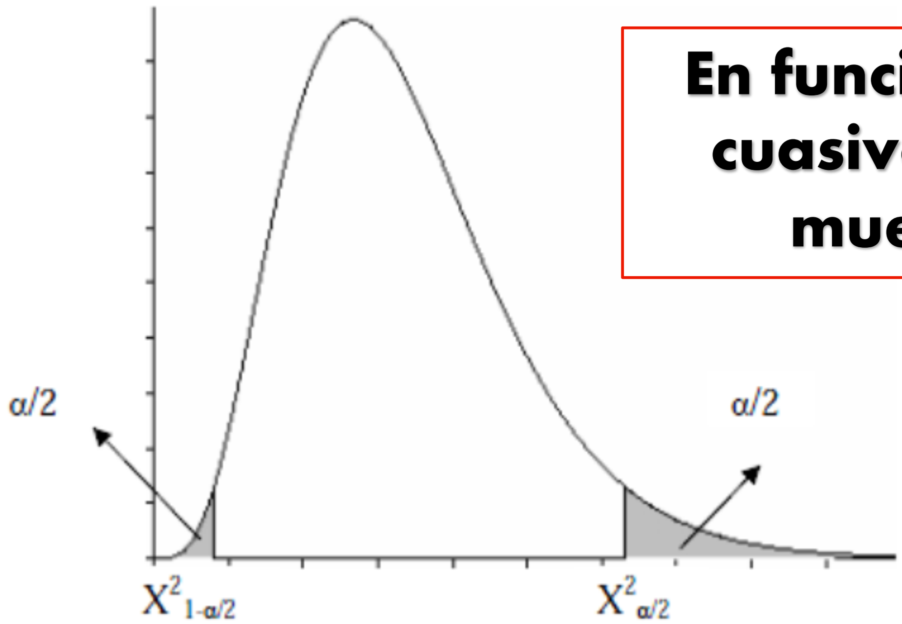
$$\boxed{= \mathrm{P}\left[\frac{ns^2}{\chi^2_{(2)n-1}} \leq \sigma^2 \leq \frac{ns^2}{\chi^2_{(1)n-1}}\right] = 1-\alpha}$$



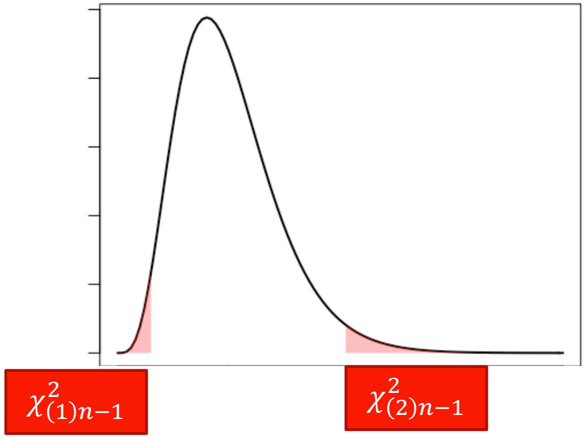
5.3.3. I.C. para σ^2 siendo $\xi: N(\mu, \sigma)$

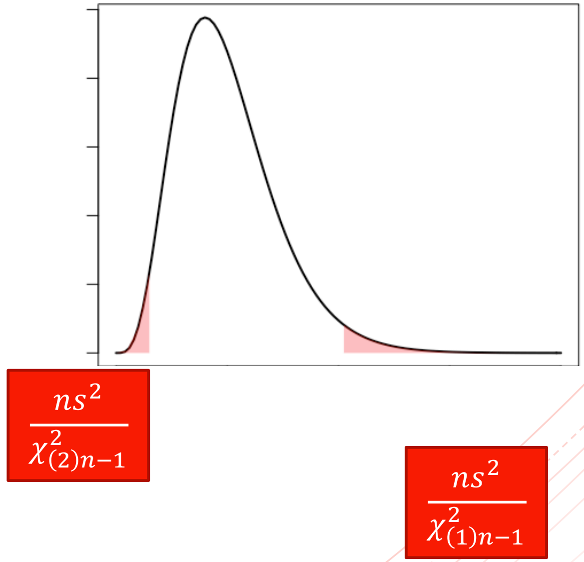
$$= \mathrm{P}\left[\frac{ns^2}{\chi^2_{(2)n-1}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{(1)n-1}}\right] = 1 - \alpha$$

$$\text{intervalo} = \left(\frac{(n-1)S_{n-1}{}^2}{\chi^2{}_{(\alpha/2)}} ; \frac{(n-1)S_{n-1}{}^2}{\chi^2{}_{(1-\alpha/2)}}\right)$$



En función de la cuasivarianza muestral





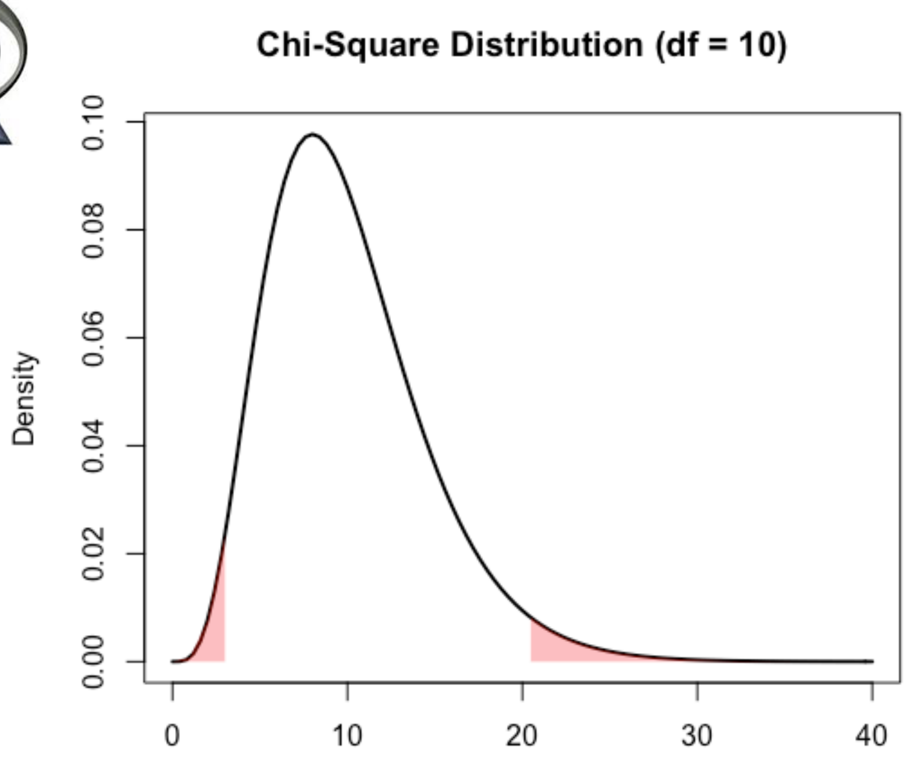
$$P(X^2_{1-\alpha/2} < \frac{(n-1)\,S^2{}_{n-1}}{\sigma^2} < X^2_{\alpha/2}) = 1-\alpha \quad \Rightarrow \quad P(\frac{(n-1)\,S^2{}_{n-1}}{X^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)\,S^2{}_{n-1}}{X^2_{1-\alpha/2}}) = 1-\alpha$$

5.3.3. I.C. para σ^2 siendo $\xi: N(\mu, \sigma)$

$$= \mathrm{P}\left[\frac{ns^2}{\chi^2_{(2)n-1}} \leq \sigma^2 \leq \frac{ns^2}{\chi^2_{(1)n-1}}\right] = 1 - \alpha$$

The following code illustrates how to fill in the portion of the density plot for the x values lying *outside* of the middle 95% of the distribution:

```
#create density curve
curve(dchisq(x, df = 10), from = 0, to = 40,
main = 'Chi-Square Distribution (df = 10)',
ylab = 'Density',
lwd = 2)

#find upper and lower values for middle 95% of distribution
lower95 <- qchisq(.025, 10)
upper95 <- qchisq(.975, 10)

#create vector of x values
x_lower95 <- seq(0, lower95)

#create vector of chi-square density values
p_lower95 <- dchisq(x_lower95, df = 10)

#fill in portion of the density plot from 0 to lower 95% value
polygon(c(x_lower95, rev(x_lower95)), c(p_lower95, rep(0, length(p_lower95))),
        col = adjustcolor('red', alpha=0.3), border = NA)

#create vector of x values
x_upper95 <- seq(upper95, 40)

#create vector of chi-square density values
p_upper95 <- dchisq(x_upper95, df = 10)

#fill in portion of the density plot for upper 95% value to end of plot
polygon(c(x_upper95, rev(x_upper95)), c(p_upper95, rep(0, length(p_upper95))),
        col = adjustcolor('red', alpha=0.3), border = NA)
```



https://www.statology.org/how-to-easily-plot-a-chi-square-distribution-in-r/

5.3.3. I.C. para σ^2 siendo $\xi: N(\mu, \sigma)$

$$= \mathrm{P}\left[\frac{ns^2}{\chi^2_{(2)n-1}} \leq \sigma^2 \leq \frac{ns^2}{\chi^2_{(1)n-1}}\right] = 1 - \alpha$$

EJEMPLO 6

Con la misma información muestral del ejemplo anterior, determínese el intervalo de confianza del 80% para la varianza σ^2 de una población normal.

Los valores K_1 y K_2, extremos del intervalo correspondiente a una distribución $\chi^2(n-1)$, $\chi^2(9)$, se obtienen sabiendo que

$$P[\chi^2(9) \geq K_1] = 0{,}90$$
$$P[\chi^2(9) \geq K_2] = 0{,}10$$

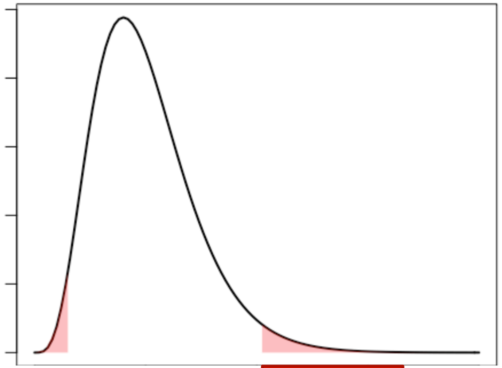
resultando, pues

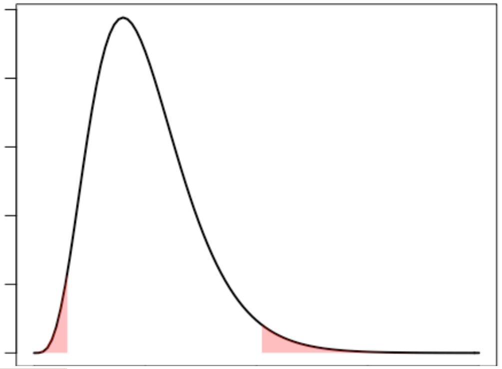
$$K_1 = 4{,}168$$
$$K_2 = 14{,}684$$

El intervalo deseado es

$$\left[\frac{10 \cdot 6{,}09}{14{,}684}; \frac{10 \cdot 6{,}09}{4{,}168}\right] = [4{,}147; 14{,}611].$$

(Martín Pliego, p.148)





5.4.1. I.C. para en poblaciones no normales (Caso General)

- Si conocemos $F[X,\theta]$ recurrimos a lo explicado en 5.2.
- Si no conocemos $F[X,\theta]$ recurrimos a la desigualdad de Chevychev (pero es necesario conocer la varianza) para construir un I.C. para μ.
- La expresión general de Chevychev es $P\left[|\bar{x}-\mu| \leq k\sqrt{V(\xi)}\right] \geq 1-\frac{1}{k^2}$
- Si usamos como ξ a $\bar{x}$ (buen estimador de μ), e igualamos $1-\alpha = 1-\frac{1}{k^2}$

$$P\left[|\bar{x}-\mu| \leq k\frac{\sigma}{\sqrt{n}}\right] \geq 1-\alpha$$

- Obtenemos el IC: $P\left[\bar{x}-k\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}+k\frac{\sigma}{\sqrt{n}}\right] \geq 1-\alpha$

con lo que $\mu \in \left[\bar{x}-k\frac{\sigma}{\sqrt{n}}, \bar{x}+k\frac{\sigma}{\sqrt{n}}\right]$ **con un nivel de confianza de al menos** $1-\alpha$

5.4.1. I.C. para en poblaciones no normales (Caso General)

EJEMPLO 7

De una población con $\sigma^2 = 11$, se extrae una muestra aleatoria simple de tamaño $n = 36$ siendo la suma de sus elementos

$$\sum_{i=1}^{36} x_i = 180.$$

Determínese el intervalo de confianza para la media poblacional μ de, por lo menos, el 90%.

Como

$$1 - \alpha = 0{,}90 = 1 - \frac{1}{K^2},$$

tendremos que $K = 3{,}162$, y como

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{180}{36} = 5$$

el intervalo de confianza para μ de, por lo menos, el 90% es

$$\left[5 - 3{,}162\,\frac{\sqrt{11}}{\sqrt{36}};\ 5 + 3{,}162\,\frac{\sqrt{11}}{\sqrt{36}}\right] = [3{,}255;\ 6{,}745].$$

$$\mu \in \left[\bar{x} - k\frac{\sigma}{\sqrt{n}}, \bar{x} + k\frac{\sigma}{\sqrt{n}}\right]$$

con un nivel de confianza de al menos $1 - \alpha$

5.4.2.A. I.C. para θ en poblaciones no normales : grandes muestras

- Si $n \to \infty$ entonces $\hat{\theta}_{MV}$ es al menos, asintóticamente normal, insesgado y eficiente (puede serlo extrictamente).

$$\hat{\theta}_{MV} \xrightarrow[d \leftrightarrow (n \to \infty)]{} N\left(\theta, \sqrt{V(\theta)}\right)$$

- Su varianza asintótica coincide con el inverso de la información de Fischer (CCR).

$$V(\hat{\theta}) = [I(\theta)]^{-1} = \frac{1}{nE\left[\frac{\partial \ln[f(x,\theta)]}{\partial \theta}\right]^2} = CCR$$

- Luego la cantidad pivotal

$$\boxed{T(X;\theta) = \frac{\hat{\theta} - \theta}{\sqrt{V(\hat{\theta})}} \xrightarrow[n \to \infty]{} Z: N(0,1)} \Rightarrow \boxed{P\left(\hat{\theta} - k\sqrt{V(\hat{\theta})} \leq \theta \leq \hat{\theta} - k\sqrt{V(\hat{\theta})}\right) = 1 - \alpha}$$

- Si no se conoce $V(\hat{\theta})$, puede sustituirse por su estimación $\widehat{V(\hat{\theta})}$, sin que ello afecte apreciablemente a la bondad de la aproximación.

5.4.2.A. I.C. para θ en poblaciones no normales : grandes muestras

$$\boxed{P\left(\hat{\theta} - k\sqrt{V(\hat{\theta})} \leq \theta \leq \hat{\theta} - k\sqrt{V(\hat{\theta})}\right) = 1 - \alpha}$$

EJEMPLO 8

En una población con distribución exponencial dada por

$$f(x;\theta) = \frac{1}{\theta} e^{-\frac{1}{\theta}x}; \quad x \geq 0, \quad \theta > 0.$$

Se extrae una muestra de tamaño $n = 100$ en la que

$$\sum_{i=1}^{100} x_i = 214 \quad \text{y} \quad \sum_{i=1}^{100} x_i^2 = 1.016.$$

Determínese el intervalo de confianza del 90% para el parámetro θ.

Puede comprobarse, fácilmente, que el estimador máximo verosímil del parámetro θ en esta población es

$$\hat{\theta} = \bar{x}$$

y su varianza

$$V(\hat{\theta}) = V(\bar{x}) = \frac{\sigma^2}{n} = \frac{\theta^2}{n}$$

siendo la varianza poblacional desconocida, puesto que $\sigma^2 = \theta^2$.[10]

Por la propiedad de invarianza de los estimadores máximo verosímiles, tendremos que

$$\widehat{V(\hat{\theta})} = \frac{(\hat{\theta}^2)}{n} = \frac{\bar{x}^2}{n}.$$

Para la muestra obtenida

$$\hat{\theta} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{214}{100} = 2{,}14$$

$$\widehat{V(\hat{\theta})} = \frac{\bar{x}^2}{n} = \frac{2{,}14^2}{100} = 0{,}045796.$$

Para un nivel de confianza del 90%, el valor tabular K se determina sabiendo que

$$P[-K \leq N(0;1) \leq K] = 0{,}90,$$

donde $K = 1{,}64$, y el intervalo de confianza para θ resulta

$$P\left[2{,}14 - 1{,}64\sqrt{0{,}045796};\ 2{,}14 + 1{,}64\sqrt{0{,}045796}\right] = [2{,}105;\ 2{,}491].$$

5.4.2.B. I.C. para π en poblaciones dicotómicas : grandes muestras

- **Si $n \to \infty$ entonces $\hat{\theta}_{MV}$ es al menos, asintóticamente normal, insesgado y eficiente (puede serlo extrictamente).**

$$\hat{\pi}_{MV} \xrightarrow[d \leftrightarrow (n \to \infty)]{} N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

- **Luego la cantidad pivotal**

$$\boxed{T(X;\pi) = \frac{p-\pi}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow[n \to \infty]{} Z : N(0,1)}$$

$$\boxed{P\left(p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le \pi \le p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha}$$

5.4.2.B. I.C. para π en poblaciones dicotómicas : grandes muestras

EJEMPLO 9

De una urna con bolas blancas y negras se obtiene una muestra aleatoria simple de tamaño $n = 140$, observándose 80 bolas blancas.

Hállese el intervalo de confianza de la proporción p de bolas blancas de la urna al nivel del 99,7%.

$$T(X;\pi) = \frac{p-\pi}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow[n\to\infty]{} Z: N(0,1)$$

Al repetir el experimento $B(1;\ p)$, extracción de bolas con reposición de la urna, un número de 140 veces con un resultado de 80 bolas blancas, el estimador máximo verosímil de la proporción p es

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{80}{140} = 0{,}57.$$

$$P\left(p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le \pi \le p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

El valor tabular K se determina teniendo en cuenta que

$$P[-K \le N(0;\ 1) \le K] = 0{,}997,$$

donde $K \simeq 3$.

Por tanto, el intervalo de confianza para p es

$$\left[0{,}57 - 3\sqrt{\frac{0{,}57\ (1-0{,}57)}{140}}\ ;\ 0{,}57 + 3\sqrt{\frac{0{,}57\ (1-0{,}57)}{140}}\right] = [0{,}445;\ 0{,}696].$$

(Martín Pliego,p.153)

5.5.A. n ?, I.C. para μ $\xi: N(\mu, \sigma)$ con σ^2 conocida,

- Recordemos: si $\xi: N(\mu, \sigma)$ con σ^2 conocida, entonces $P\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$, expresión equivalente a $P\left[|\bar{x} - \mu| \leq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$
- Error en la estimación $e = |\bar{x} - \mu|$. Si lo llevamos al límite aceptado $|\bar{x} - \mu| = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, es posible determinar el tamaño de la muestra necesario, despejando n.

$$\boxed{n = \frac{\left(z_{\alpha/2}\right)^2 \sigma^2}{e^2}}$$

5.5.A. n ?, I.C. para μ $\xi: N(\mu, \sigma)$ con σ^2 conocida, $\boxed{n = \frac{\left(z_{\alpha/2}\right)^2 \sigma^2}{e}}$

EJEMPLO 10

Determínese el tamaño de la muestra necesario para estimar la media μ de una población con $\sigma = 4{,}2$, si se quiere tener una confianza del 95% de que el error de estimación se sitúe, como mucho, entre $\pm 0{,}05$.

Como $1 - \alpha = 0{,}95$, en las tablas de la $N(0; 1)$ se obtiene $K = 1{,}96$, por tanto,

$$n = \frac{(1{,}96)^2 \cdot (4{,}2)^2}{(0{,}05)^2} = 27.106 \text{ elementos muestrales}$$

5.5.B. n ?, I.C. para μ $\xi: N(\mu, \sigma)$ con σ^2 desconocida,

- Recordemos: si $\xi: N(\mu, \sigma)$ con σ^2 conocida, entonces $P\left[\bar{x} - t_{\alpha/2}\sqrt{\frac{s^2}{n-1}} \leq \mu \leq \bar{x} + t_{\alpha/2}\sqrt{\frac{s^2}{n-1}}\right] = 1 - \alpha$, expresión equivalente a $P\left[|\bar{x} - \mu| \leq t_{\alpha/2}\sqrt{\frac{s^2}{n-1}}\right] = 1 - \alpha$
- Error en la estimación $e = |\bar{x} - \mu|$. Si lo llevamos al límite aceptado $|\bar{x} - \mu| = t_{\alpha/2}\sqrt{\frac{s^2}{n-1}}$, es posible determinar el tamaño de la muestra necesario, despejando n.

$$\boxed{n = \frac{\left(t_{\alpha/2}\right)^2 s^2}{e^2} - 1}$$

5.5.C. n ?, I.C. para π , $\xi : B(1,\pi)$

- **Recordemos: si** $\left.\begin{matrix} \xi : B(1,\pi) \\ \mathrm{p} = \hat{\pi}_{MV} \\ n \to \infty \end{matrix}\right\}$ $\mathrm{p} \underset{d}{\to} N\left(\pi, \sqrt{\frac{1-\pi}{\pi}}\right)$. **y,entonces** $P\left(p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le \pi \le p - Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) =$

  $1 - \alpha$, **expresión equivalente a** $P\left[|p - \pi| \le Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right] = 1 - \alpha$

- **Error en la estimación** $\mathrm{e} = |\bar{x} - \mu|$. **Si lo llevamos al límite aceptado** $|p - \pi| \le Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$, **es posible determinar el tamaño de la muestra necesario, despejando n.**

$$\boxed{n = \frac{\left(Z_{\alpha/2}\right)^2 p\,(1-p)}{e^2}}$$

- **Si no disponemos de una estimación previa** $p = \hat{\pi}_{MV}$, **entonces "nos ponemos en el peor de los casos de máxima varianza", cuando π=0,5, y entonces**

$$\boxed{n = \frac{\left(Z_{\alpha/2}\right)^2}{4e^2}}$$

5.5.C. n ?, I.C. para π , $\xi: B(1,\pi)$

$$n = \frac{\left(Z_{\alpha/2}\right)^2 p\,(1-p)}{e^2}$$

El peor de los casos

$$si\ V(\pi) = 0{,}5 \rightarrow n = \frac{\left(Z_{\alpha/2}\right)^2}{4e^2}$$

EJEMPLO 11

Determínese el tamaño muestral necesario para estimar la proporción p de una población, con una probabilidad del 99,7% de que el error cometido en la estimación sea como mucho del ±2%.

Al nivel $1-\alpha = 0{,}997$, le corresponde un valor tabular, en la $N(0;\,1)$, $K \simeq 3$, por tanto,

$$n = \frac{3^2}{4\cdot(0{,}02)^2} = 5.625.$$

(Martín Pliego,p.156)

Bibliografía

Martín Pliego, Fundamentos de Inferencia Estadística, Editorial AC, 3ªEd, 2005

Otros recursos en www.pacorabadan.com

y aula virtual en URJC

Universidad Rey Juan Carlos