TEMA 10

Estimación por intervalos

Vniversitat de València
OpenCourseWare

Resumen Tema 9 ESTIMADORES PUNTUALES

Parámetro poblacional a estimar θ	*Estimador* $\hat{\theta}$	*valor esperado* $E(\hat{\theta})$	*Sesgo* $b(\hat{\theta}) = E(\hat{\theta}) - \theta$	*Varianza* $Var(\hat{\theta})$	*ECM* $Var(\hat{\theta}) + b^2(\hat{\theta})$
μ *Normal*	$\hat{\mu} = \bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$	$E(\bar{X}_n) = \mu$	insesgado 0	$\frac{\sigma^2}{n}$	EMV
p *Bernoulli*	$\hat{p} = \frac{\sum_{i=1}^n X_i}{n}$	$E(\hat{p}) = p$	insesgado 0	$\frac{p(1-p)}{n}$	EMV
σ^2 *Normal* μ *desconocida*	$\hat{\sigma}^2 =$ $S_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{x}_n)^2}{n}$	$E(S_n^2) = \frac{n-1}{n}\sigma^2$	*sesgado* $b(S_n^2) = -\frac{\sigma^2}{n}$ Asintóticamente insesgado	$Var(S_n^2)$	EMV
σ^2 *Normal* μ *conocida*	$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$	$E(\hat{\sigma}^2) = \sigma^2$	Insesgado 0	$Var(\hat{\sigma}^2)$	EMV
λ *Poisson*	$\hat{\lambda} = \bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$	$E(\bar{X}_n) = \lambda$	insesgado 0	$\frac{\lambda}{n}$	EMV

10. Estimación por intervalos.

La estimación puntual de un parámetro poblacional no coincidirá con el verdadero valor del parámetro pero se espera que esté próximo a él.
La estimación por intervalos construye intervalos, alrededor del estimador puntual, en los que será altamente probable que se encuentre el valor del parámetro poblacional.

A un intervalo de este tipo se le llamará: *intervalo aleatorio o estimador por intervalo*.

Nivel de confianza $1-\alpha$: probabilidad de que el intervalo aleatorio contenga al parámetro.
Es decir, si tomamos muestras de tamaño "n", el nivel de confianza indicará que el $(1-\alpha)\cdot 100$ de las muestras de ese tamaño proporcionará un intervalo que contendrá al parámetro.

Niveles de confianza más habituales: 0,99; 0,95; 0,955 y 0,90.

10.1. Intervalos de confianza para la MEDIA POBLACIONAL μ

10. 1.1. En una población NORMAL con varianza σ^2 ***conocida.***

PARA ***n cualquier tamaño***

Nivel de confianza $1-\alpha \Rightarrow$ Intervalo aleatorio: $\overline{X}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

(Reproductividad de la Normal)

Ejemplo 10.1 Una empresa automovilística quiere estimar la media de las emisiones de CO_2 del último modelo de coche que quiere sacar al mercado. De una muestra aleatoria de n = 9 coches ha resultado una media de emisiones de 170 gr de CO_2 por Km recorrido. Se supondrá que la variable X, que mide las emisiones de CO_2 de esos coches, tiene una distribución Normal de varianza $\sigma^2 = 48^2$. Nivel de confianza $1-\alpha = 0{,}95$.

Solución: para una m.a.s. de n = 9 coches $\bar{x}_n = 170$ *gr por Km (estimación puntual).*
La v.a. X: emisiones de CO_2 *del coche (en gr por Km) tienen una distribución Normal con varianza conocida:* $X \sim Normal(\mu; \sigma^2 = 48^2)$*. Por tanto, aplicando el resultado anterior y teniendo en cuenta que para un nivel de confianza de* $1-\alpha = 0{,}95 \rightarrow \frac{\alpha}{2} = 0{,}025 \rightarrow P(Z < z_{\alpha/2}) = 0{,}9750 \leftrightarrow z_{\alpha/2} = 1{,}96$*, queda:*
$\overline{X}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \rightarrow 170 \pm 1{,}96\frac{48}{\sqrt{9}} \rightarrow 170 \pm 31{,}36$*. Por tanto* $\boldsymbol{IC(\mu)} = [\mathbf{138{,}64};\ \mathbf{201{,}36}]$ *(estimación por intervalo). La media de las emisiones de* CO_2 *de este modelo de coche está entre 138,64 y 201,36 gr por Km con una confianza del 95%*

<u>Solución desarrollada</u>:

Sea la v.a. X: emisiones de CO_2 del coche (en gr por Km). $X \sim Normal(\mu; \sigma^2 = 48^2)$*(población normal)*

Se toman m.a.s. de n = 9 coches: el vector aleatorio $(X_1, X_2, \cdots, X_9)$ *representa todas las muestras de 9 coches que se pueden tomar.*

Las v.a. X_i: *emisiones del* $i-$*ésimo coche* ; $X_i \sim Normal(\mu; \sigma^2 = 48^2)$ *con* $i = 1,2,\cdots,9$*, es decir, iid.*

Sabemos que, en una población normal con varianza conocida, por la propiedad de reproductividad del modelo normal, la media muestral tiene una distribución Normal de media μ y desviació típica $\frac{\sigma}{\sqrt{n}}$*, es decir:* $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \sim Normal\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

En nuestro caso $\bar{X}_n = \frac{\sum_{i=1}^{9} X_i}{9} \sim Normal\left(\mu,\ \frac{\sigma}{\sqrt{n}} = \frac{48}{\sqrt{9}} = 16\right)$

Por tanto la v.a. $Z = \frac{\bar{X}_n - \mu}{\frac{48}{\sqrt{9}}} \sim N(0;1)$*, es decir, Z se distribuye como una Normal tipificada.*

Utilizaremos esta expresión para obtener una estimación por intervalo de la media poblacional μ al 95% de confianza.

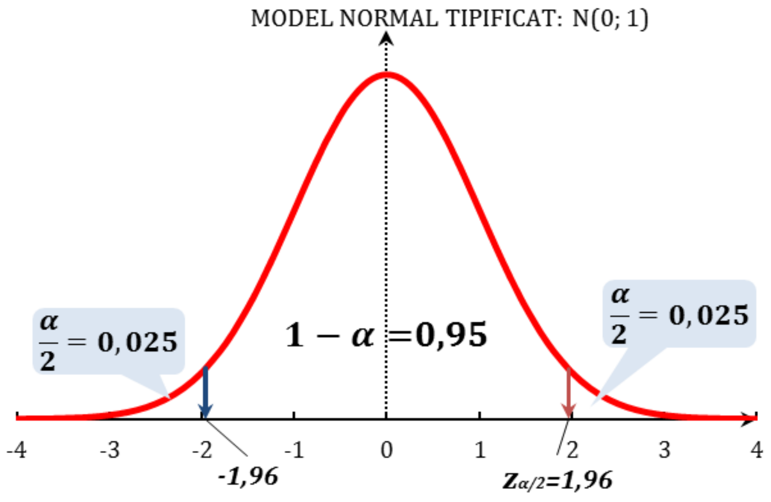
Para un nivel de confianza $1-\alpha = 0{,}95$ *buscamos en la normal tipificada el intervalo* $\left[-z_{\alpha/2};\ z_{\alpha/2}\right]$ *tal que* $P\left(-z_{\alpha/2} < Z < z_{\alpha/2}\right) = 0{,}95 \leftrightarrow 2F\left(z_{\alpha/2}\right) - 1 = 0{,}95 \leftrightarrow F\left(z_{\alpha/2}\right) = 0{,}9750 \leftrightarrow z_{\alpha/2} = 1{,}96$

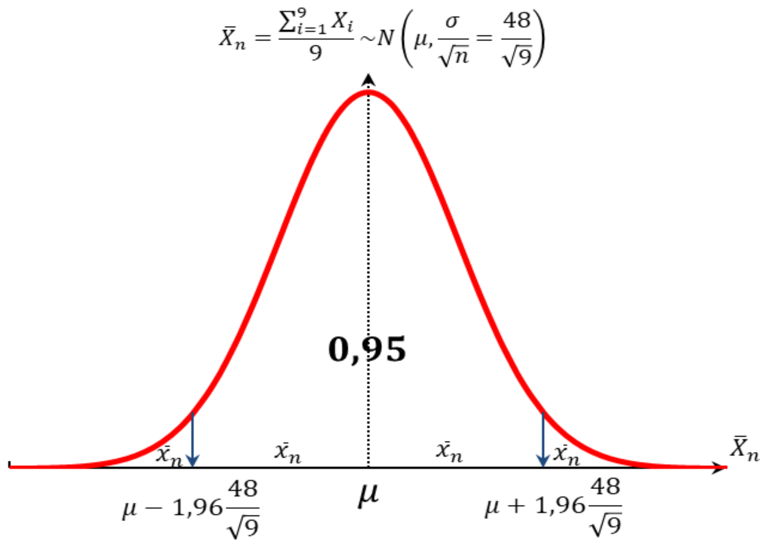
Así queda: $0{,}95 = P(-1{,}96 < Z < 1{,}96) = P\left(-1{,}96 < \frac{\bar{X}_n - \mu}{\frac{48}{\sqrt{9}}} < 1{,}96\right) =$

$= P\left(-\bar{X}_n - 1{,}96\frac{48}{\sqrt{9}} < -\mu < -\bar{X}_n + 1{,}96\frac{48}{\sqrt{9}}\right) = P\left(\bar{X}_n - 1{,}96\frac{48}{\sqrt{9}} < \mu < \bar{X}_n + 1{,}96\frac{48}{\sqrt{9}}\right)$

Para la muestra concreta de n = 9 coches se ha obtenido una media de $\bar{x}_n = 170$ *y el intervalo de confianza para la media poblacional* μ *es:* $IC(\mu) = \left[170 - 1{,}96\frac{48}{\sqrt{9}};\ 170 + 1{,}96\frac{48}{\sqrt{9}}\right] = [138{,}64; 201{,}36]$
Por tanto, la media de las emisiones de CO_2 *de este modelo de coche está entre 138,64 y 201,36 gr por Km con una confianza del 95%.*

GRÁFICAMENTE
Nótese que: $0{,}95 = P\left(\bar{X}_n - 1{,}96\frac{48}{\sqrt{9}} < \mu < \bar{X}_n + 1{,}96\frac{48}{\sqrt{9}}\right) = P\left(\mu - 1{,}96\frac{48}{\sqrt{9}} < \bar{X}_n < \mu + 1{,}96\frac{48}{\sqrt{9}}\right)$ *para todas las muestras de tamaño n = 9* $(X_1, X_2, \cdots, X_9)$.





10.1.2. En una población NORMAL con varianza σ^2 desconocida.

PARA ***n cualquier tamaño (muestra pequeña $n \leq 30$)***

Variable aleatoria $\Rightarrow T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n-1}} = \frac{\bar{X}_n - \mu}{S_n} \cdot \sqrt{n-1} \sim t_{n-1}$

t de Student con (n-1) grados de libertad.

Nivel de confianza $1-\alpha \Rightarrow$ Intervalo aleatorio: $\overline{X}_n \pm t_{\alpha/2} \frac{S_n}{\sqrt{n-1}}$

- Si la población es NORMAL y la varianza σ^2 desconocida y la MUESTRA es suficientemente GRANDE se puede aplicar el mismo intervalo aleatorio que el caso 2.1.1 pero sustituyendo la varianza poblacional por la muestral (la t – Student converge a la Normal):

Población NORMAL, varianza σ^2 desconocida y PARA ***n GRANDE*** $n > 30$

Nivel de confianza $1-\alpha \Rightarrow$ Intervalo aleatorio: $\overline{X}_n \pm z_{\alpha/2} \frac{S_n}{\sqrt{n}}$

Ejemplo 10.2 Una empresa automovilística quiere estimar la media de las emisiones de CO_2 del último modelo de coche que quiere sacar al mercado. De una muestra aleatoria de n = 9 coches ha resultado una media de emisiones de 170 gr de CO_2 por Km recorrido con una desviación típica de 45 gr por Km. Se supondrá que la variable X, que mide las emisiones de CO_2 de esos coches, tiene una distribución Normal de varianza poblacional σ^2 desconocida. Nivel de confianza $1-\alpha=0{,}95$.

Solución: $IC(\mu)=[135{,}41;\ 204{,}59]$

<u>Solución</u>: para una m.a.s. de n = 9 coches $\bar{x}_n=170$ y $s_n=45$ gr por Km (<u>estimación puntual</u>)

La v.a. X: emisiones de CO_2 del coche (en gr por Km) tienen una distribución Normal con varianza desconocida: $X \sim Normal(\mu;\sigma^2)$.

Por tanto, aplicando el resultado anterior en 2.1.2, la v.a. $T=\frac{\bar{X}_n-\mu}{S_n/\sqrt{n-1}} =\sim t_{n-1} \to T=\frac{\bar{X}_n-\mu}{45/\sqrt{8}} =\sim t_8$, es decir, T se distribuye como una t de Student con 8 grados de libertad.

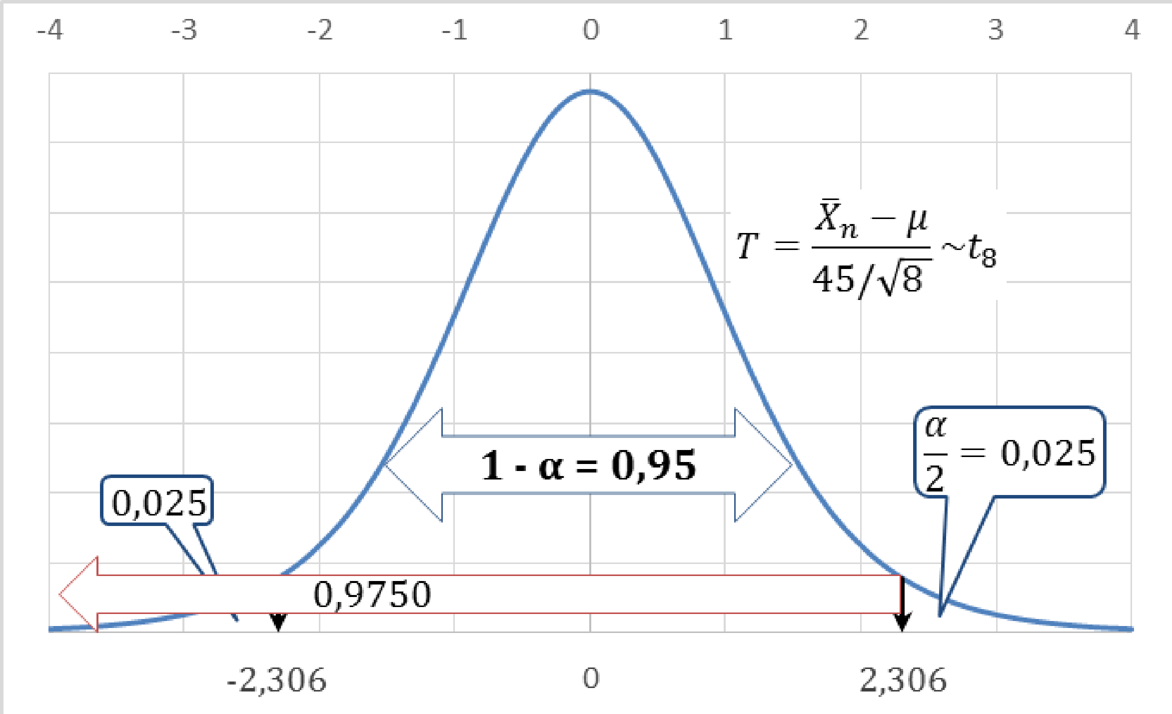
El intervalo es $\overline{X}_n \pm t_{\frac{\alpha}{2}}\frac{S_n}{\sqrt{n-1}}$. Para un nivel de confianza de $1-\alpha=0{,}95 \to \frac{\alpha}{2}=0{,}025$ y $t_{\alpha/2}$ es un valor de la t_8 tal que $P\left(t_8>t_{\alpha/2}\right)=0{,}025 \to P\left(t_8<t_{\alpha/2}\right)=0{,}9750 \leftrightarrow t_{\alpha/2}=2{,}306$,

Así, el intervalo de confianza queda:

$\overline{X}_n \pm t_{\frac{\alpha}{2}}\frac{S_n}{\sqrt{n-1}} \to 170 \pm 2{,}306\frac{45}{\sqrt{8}} \to 170 \pm 36{,}69$. Por tanto $\boldsymbol{IC(\mu)=[135{,}41;\ 204{,}59]}$ (<u>estimación por intervalo</u>). La media de las emisiones de CO_2 de este modelo de coche estaría entre 135,41 y 204,59 gr por Km con una confianza del 95%.

GRÁFICAMENTE:

$$0{,}95 = P\left(-2{,}306 < \frac{\bar{X}_n-\mu}{45/\sqrt{8}} < 2{,}306\right) = P\left(\bar{X}_n - 2{,}306\frac{45}{\sqrt{8}} < \mu < \bar{X}_n + 2{,}306\frac{45}{\sqrt{8}}\right)$$



10.1.3. En una población no normal con varianza σ^2 conocida.

PARA ***n GRANDE***

Variable aleatoria $\Rightarrow Z = \dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \dfrac{\bar{X}_n - \mu}{\sigma} \cdot \sqrt{n} \sim Normal(0,1)$ (TCL)

Nivel de confianza $1-\alpha \Longrightarrow$ Intervalo aleatorio: $\overline{X}_n \pm z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

- Si la distribución poblacional es desconocida o es un modelo conocido distinto del normal, la varianza ***σ^2 es desconocida*** y la MUESTRA es suficientemente GRANDE, por el TCL, se puede aplicar también el mismo criterio, sustituyendo la varianza poblacional por una estimación adecuada.

$$\overline{X}_n \pm z_{\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}$$

$$Desconocida(\mu, \sigma) \rightarrow \hat{\sigma} = s_n$$

$$Bernoulli(p) \rightarrow \hat{\sigma} = \sqrt{\hat{p}(1-\hat{p})}$$

$$Poisson(\lambda) \rightarrow \hat{\sigma} = \sqrt{\hat{\lambda}} = \sqrt{\bar{x}}$$

Ejemplo 10.3 Una empresa automovilística quiere estimar la media de las emisiones de CO_2 del último modelo de coche que quiere sacar al mercado. De una muestra aleatoria de n = 100 coches ha resultado una media de emisiones de 170 gr de CO_2 por Km recorrido con una desviación típica de 45 gr por Km. $X \sim \mathcal{D}(\mu;\ \sigma)$ Nivel de confianza $1-\alpha = 0{,}95$.

Solución: $\boldsymbol{IC(\mu) = [161,18;\ 178,82]}$

<u>Solución</u>: para una m.a.s. de n = 100 coches $\bar{x}_n = 170$ *y* $s_n = 45$ *gr por Km (<u>estimación puntual</u>)*

La v.a. X: emisiones de CO_2 del coche (en gr por Km) tienen una distribución desconocida de parámetros desconocidos: $X \sim \mathcal{D}(\mu; \sigma^2)$.
Las v.a. X_i: *emisiones del* $i-$*ésimo coche* ; $X_i \sim \mathcal{D}(\mu; \sigma^2)$ *con* $i = 1,2,\cdots,100$, *es decir, iid.*

Por el TCL, la media muestral tiene una distribución aproximadamente Normal de media μ y desviació típica $\frac{\sigma}{\sqrt{n}}$, *es decir:* $\bar{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} \sim Normal\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. *Como σ es desconocida sustituimos su valor su estimador puntual:* $\hat{\sigma} = s_n = 45$.
Por tanto, aplicando el resultado anterior en 2.1.3 y teniendo en cuenta que para un nivel de confianza de $1-\alpha = 0{,}95 \rightarrow \frac{\alpha}{2} = 0{,}025 \rightarrow P\left(Z < z_{\alpha/2}\right) = 0{,}9750 \leftrightarrow z_{\alpha/2} = 1{,}96$, *queda:*
$\overline{X}_n \pm z_{\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}} \rightarrow \overline{X}_n \pm z_{\alpha/2}\frac{s_n}{\sqrt{n}} \rightarrow 170 \pm 1{,}96\frac{45}{\sqrt{100}} \rightarrow 170 \pm 8{,}82$. *Por tanto* $\boldsymbol{IC(\mu) = [161,18;\ 178,82]}$.
La media de las emisiones de CO_2 de este modelo de coche está entre 161,18 y 178,82 gr por Km con una confianza del 95%

10.1.4. Estimación por intervalo de una PROPORCIÓN POBLACIONAL p.

PARA ***n muy GRANDE*** $n \geq 100$

Por el TCL, la v.a. $\hat{p} \sim Normal\left(p; \sqrt{\frac{p(1-p)}{n}}\right)$

Por tanto, la variable aleatoria $\Rightarrow Z = \dfrac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim Normal(0,1)$

Nivel de confianza $1-\alpha \Rightarrow$ Intervalo aleatorio: $\hat{p} \pm z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Siendo $\hat{p} = \dfrac{X}{n} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ proporción con que se observa el fenómeno en la muestra.

Ejemplo 10.4 Supongamos que se pretende estimar el porcentaje de coches híbridos o eléctricos que hay en la Comunitat Valenciana (CV). Si de una muestra aleatoria simple de 500 propietarios de automóviles, 50 son eléctricos o híbridos, obtener un intervalo de confianza para dicho porcentaje al 90%.

Solución: $IC(p) = [7{,}37\%;\ 12{,}63\%]$

<u>*Solución*</u>:

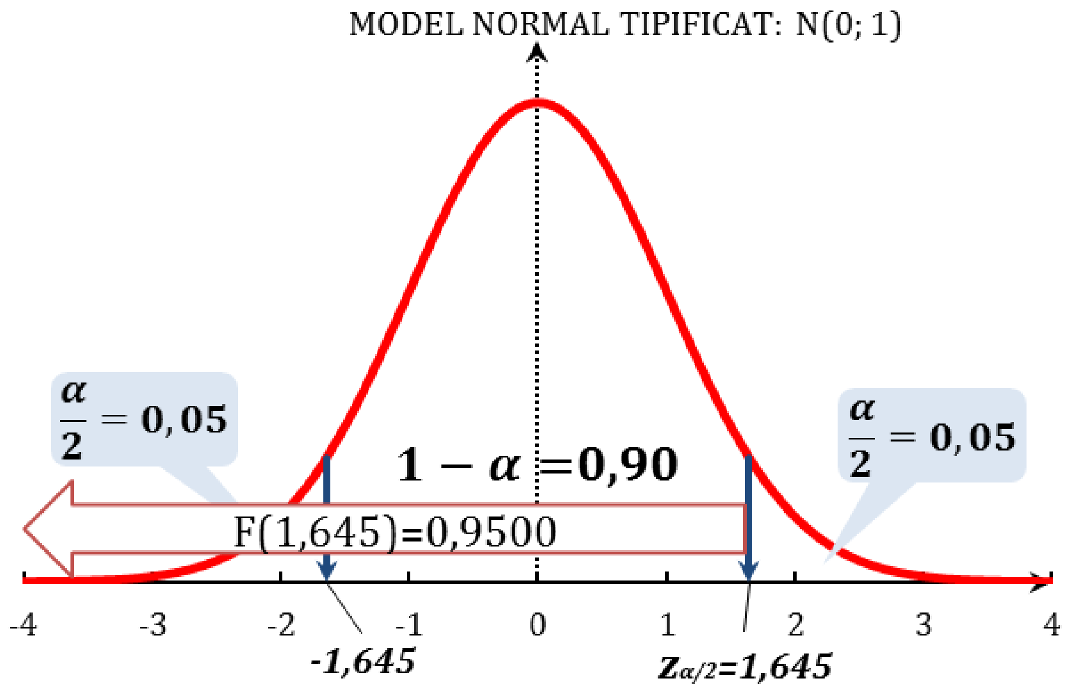
El modelo de probabilidad poblacional es una Bernoulli.

La variable aleatoria $X = \begin{cases} 0 \text{ No híbrido o eléctrico} \\ 1 \textit{ Sí} \text{ híbrido o eléctrico} \end{cases}$ *tiene una distribución Bernoulli de parámetro* $\boldsymbol{p}$ *desconocido. Si tenemos en cuenta los 500* $\left(x_i = \begin{cases} 0\ no \\ 1\ sí \end{cases}\right)$*, la proporción muestral de propietarios de híbridos o eléctricos es.* $\hat{p} = \frac{\sum_{i=1}^{500} x_i}{n} = \frac{50}{500} = 0{,}10.$

Por tanto, aplicando el resultado anterior en 2.1.4 y teniendo en cuenta que para un nivel de confianza de $1 - \alpha = 0{,}90 \rightarrow \frac{\alpha}{2} = 0{,}05 \rightarrow P\left(Z > z_{\alpha/2}\right) = 0{,}05 \rightarrow P\left(Z \leq z_{\alpha/2}\right) = F\left(z_{\alpha/2}\right) = 0{,}95 \leftrightarrow z_{\alpha/2} = 1{,}645,$ *el intervalo queda:*

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \equiv 0{,}10 \pm 1{,}645\sqrt{\frac{0{,}1 \cdot 0{,}9}{500}} \equiv 0{,}10 \pm 0{,}0263 \rightarrow IC(p) = [0{,}0737;\ 0{,}1263].$$

En conclusión, el porcentaje de propietarios de coches híbridos o eléctricos en la CV está entre el 7,37 y el 12,63% con una confianza del 90%.


MODEL NORMAL TIPIFICAT: N(0; 1)
$\frac{\alpha}{2} = 0,05$
$1 - \alpha = 0,90$
$\frac{\alpha}{2} = 0,05$
F(1,645)=0,9500
-4
-3
-2
-1
0
1
2
3
4
-1,645
$z_{\alpha/2}=1,645$

10.2. Intervalo de confianza para la VARIANZA POBLACIONAL σ^2

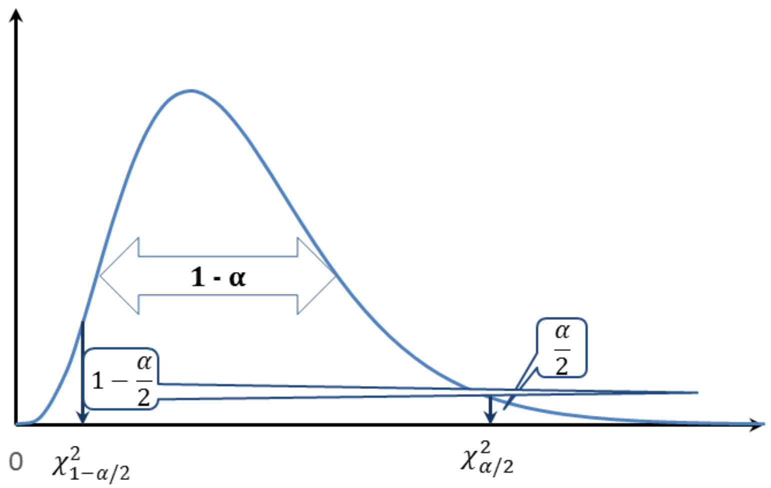
<u>Población NORMAL con media μ desconocida</u>.

PARA ***n cualquier tamaño***

Variable aleatoria $Y = \frac{nS_n^2}{\sigma^2} \sim \chi^2_{n-1}$ ***n−1*** grados de libertad

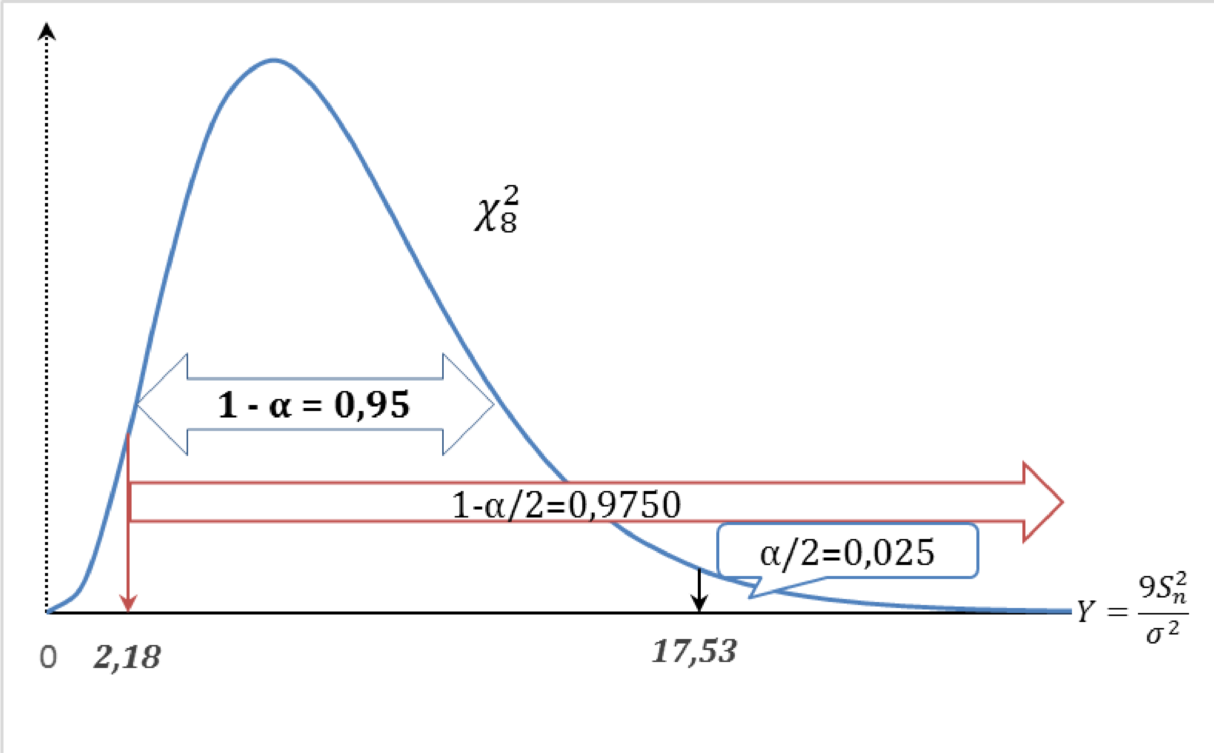
Nivel de confianza $1-\alpha \Longrightarrow$ Intervalo aleatorio: $\left[\frac{nS_n^2}{\chi^2_{\alpha/2}} ; \frac{nS_n^2}{\chi^2_{1-\alpha/2}}\right]$

Donde $P\left(\chi^2_{n-1} > \chi^2_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$ y $P\left(\chi^2_{n-1} > \chi^2_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$



Ejemplo 10.5 Una empresa automovilística quiere estimar la varianza de las emisiones de CO_2 del último modelo de coche que quiere sacar al mercado. De una muestra aleatoria de n = 9 coches ha resultado una media de emisiones de 170 gr de CO_2 por Km recorrido con una desviación típica de 45 gr por Km. Se supondrá que la variable X, que mide las emisiones de CO_2 de esos coches, tiene una distribución Normal de varianza poblacional σ^2 desconocida. Nivel de confianza $1-\alpha = 0{,}95$.

Solución: para una m.a.s. de n = 9 coches $\bar{x}_n = 170$ *y* $s_n = 45$ *gr por Km*
Para obtener un intervalo de confianza para la varianza poblacional σ^2 *en un contexto de población normal con media μ desconocida,* $X \sim Normal(\mu;\sigma^2)$*, se ha de utilizar la v.a.* $Y = \frac{nS_n^2}{\sigma^2} \sim \chi_{n-1}^2$ *que se distribuye como una ji – cuadrado con n – 1 grados de libertad. En este caso:* $Y = \frac{9S_n^2}{\sigma^2} \sim \chi_8^2$.
El intervalo aleatorio es: $\left[\frac{nS_n^2}{\chi_{\alpha/2}^2};\ \frac{nS_n^2}{\chi_{1-\alpha/2}^2}\right]$
Para un nivel de confianza de $1-\alpha = 0{,}95 \rightarrow \frac{\alpha}{2} = 0{,}025$. *Por tanto, los valores del denominador para una* χ_8^2 *son:* $P\left(\chi_8^2 > \chi_{1-\frac{\alpha}{2}}^2\right) = 1-\frac{\alpha}{2} = 0{,}9750 \leftrightarrow F\left(\chi_{1-\frac{\alpha}{2}}^2\right) = 0{,}025 \rightarrow \chi_{1-\frac{\alpha}{2}}^2 = 2{,}18$ *y*
$P\left(\chi_8^2 > \chi_{\frac{\alpha}{2}}^2\right) = \frac{\alpha}{2} = 0{,}025 \leftrightarrow F\left(\chi_{\frac{\alpha}{2}}^2\right) = 0{,}9750 \rightarrow \chi_{\frac{\alpha}{2}}^2 = 17{,}53$.
El intervalo queda: $IC(\sigma^2) = \left[\frac{9\cdot 45^2}{17{,}53};\ \frac{9\cdot 45^2}{2{,}18}\right] = [1039{,}646; 8360{,}09] \rightarrow IC(\sigma) = [32{,}24;\ 91{,}43]$.
La variabilidad en las emisiones de CO_2 *está entre 32,24 y 91,43 gramos por Km recorrido.*



χ^2_8
1 - α = 0,95
1-α/2=0,9750
α/2=0,025
$Y = \frac{9S_n^2}{\sigma^2}$
0
2,18
17,53