Intervalos de Confianza

INTERVALOS DE CONFIANZA

Cuando se obtiene una estimación puntual de un parámetro, es conveniente acompañar dicha estimación por una **medida** de la precisión de la estimación.

Un modo de hacerlo es informar el estimador y su error standard.

Otro modo es reemplazar la estimación puntual por un intervalo de valores posibles para el parámetro.

Ejemplo: Supongamos que tenemos una m.a. $X_1, X_2, \ldots, X_n$

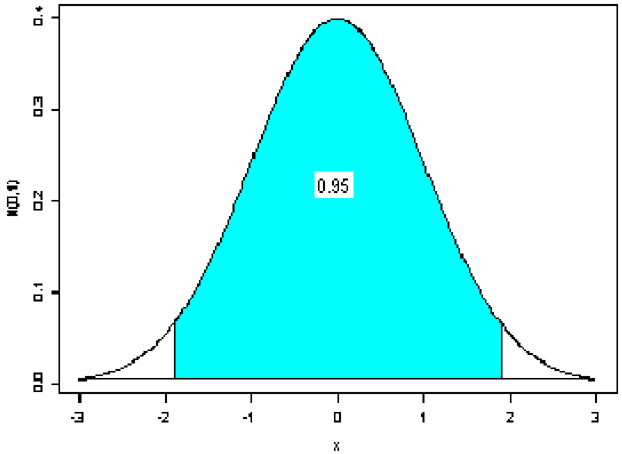
de una distribución $N(\mu, \sigma_o^2)$ con varianza σ_o^2 conocida.

Por ser los datos normales, sabemos que

$$\overline{X} \sim N\left(\mu, \frac{\sigma_o^2}{n}\right) \quad \Leftrightarrow \quad \frac{\overline{X} - \mu}{\frac{\sigma_o}{\sqrt{n}}} \sim N(0,1)$$

y, por lo tanto,

$$P\left(-1.96 \leq \sqrt{n}\, \frac{\overline{X} - \mu}{\sigma_o} \leq 1.96\right) = 0.95$$



A partir de esta expresión obtenemos

$$P\left(-1.96\frac{\sigma_o}{\sqrt{n}} \leq \overline{X} - \mu \leq 1.96\frac{\sigma_o}{\sqrt{n}}\right) = 0.95 \quad \Leftrightarrow \quad P\left(\overline{X} - 1.96\frac{\sigma_o}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma_o}{\sqrt{n}}\right) = 0.95$$

Es decir, que la probabilidad de que el intervalo

$$\left[\overline{X} - 1.96\frac{\sigma_o}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma_o}{\sqrt{n}}\right]$$

contenga al verdadero valor del parámetro μ es 0.95. Este intervalo se denomina **intervalo de confianza para μ de nivel de confianza 0.95**.

A partir de esta expresión obtenemos

$$P\left(-1.96\frac{\sigma_o}{\sqrt{n}} \leq \overline{X} - \mu \leq 1.96\frac{\sigma_o}{\sqrt{n}}\right) = 0.95 \quad \Leftrightarrow \quad P\left(\overline{X} - 1.96\frac{\sigma_o}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma_o}{\sqrt{n}}\right) = 0.95$$

Es decir, que la probabilidad de que el intervalo

$$\left[\overline{X} - 1.96\frac{\sigma_o}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma_o}{\sqrt{n}}\right]$$

contenga al verdadero valor del parámetro μ es 0.95. Este intervalo se denomina **intervalo de confianza para μ de nivel de confianza 0.95**.

En general, tendremos

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

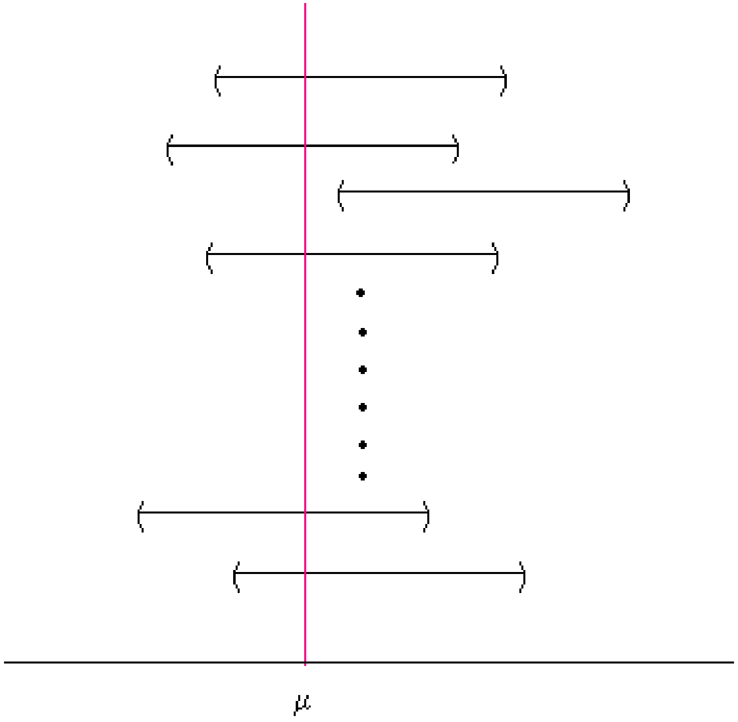
luego el siguiente **intervalo de confianza es de nivel 1 - α para μ**

$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma_o}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma_o}{\sqrt{n}}\right]$$

Interpretación:

Supongamos que, en base a diferentes muestras calculamos los correspondientes intervalos de confianza para μ siguiendo el método anterior.

Entonces el $(1 - \alpha)$ 100% de ellos contendrán al verdadero valor μ.



Ejemplo

Sea $X_1, X_2, \ldots, X_{49}$ una m.a., $X_i \sim N(\mu, \sigma^2)$.

Supongamos que el verdadero valor del desvío standard es $\sigma_o = 35$ y que se observa $\bar{x} = 160$ y construyamos un intervalo de confianza para la media de nivel 0.95.

Como las v.a. son normales y la varianza es conocida, el intervalo para μ será de la forma

$$\left(\overline{X} - z_{\alpha/2} \frac{\sigma_o}{\sqrt{n}}, \overline{X} + z_{\alpha/2} \frac{\sigma_o}{\sqrt{n}} \right)$$

con $z_{\alpha/2} = z_{0.025} = 1.96$, $\sigma_o = 35, n = 49$ y valor observado de $\overline{X}$ igual a 160. Obtenemos

$$\left(160 - 1.96 \frac{35}{\sqrt{49}}, 160 + 1.96 \frac{35}{\sqrt{49}} \right) = (160 - 9.8, 160 + 9.8) = (150.2, 169.8)$$

Determinación del tamaño de muestra

Consideremos el intervalo de confianza para μ con varianza conocida en el caso de una m.a. normal. La longitud del intervalo obtenido es

$$L = 2z_{\alpha/2}\frac{\sigma_o}{\sqrt{n}}$$

y depende de

- nivel de confianza (α)
- varianza o desvío standard de las observaciones (σ_o)
- tamaño de la muestra (n)

Un modo de obtener mayor precisión, es decir un intervalo más angosto, es aumentando el tamaño de la muestra. Si se desea una longitud menor o igual que L_o, entonces

$$\boxed{L = 2z_{\alpha/2}\frac{\sigma_o}{\sqrt{n}} \leq L_o \Leftrightarrow \sqrt{n} \geq \frac{2z_{\alpha/2}\sigma_o}{L_o} \Leftrightarrow n \geq \left(\frac{2z_{\alpha/2}\sigma_o}{L_o}\right)^2}$$

Ejemplo

Supongamos que σ_o = 35, ¿qué tamaño de muestra se requiere como mínimo para obtener un intervalo de nivel 0.95 de longitud menor o igual que 10?

Por lo tanto, $L_o = 10, \sigma_o = 35$ y $z_{0.025} = 1.96$, entonces

$$n \geq \left(\frac{2 \cdot 1.96 \cdot 35}{10} \right)^2 = 188.23 \quad \Rightarrow \quad n \geq 189$$

En el caso de varianza desconocida el problema es más complejo porque el percentil *t* también depende del tamaño de muestra.

INTERVALOS DE CONFIANZA PARA LOS PARAMETROS DE LA DISTRIBUCION NORMAL

Distribución t:

Sean dos v.a. Z ~N(0,1) y U $\sim \chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ independientes, entonces

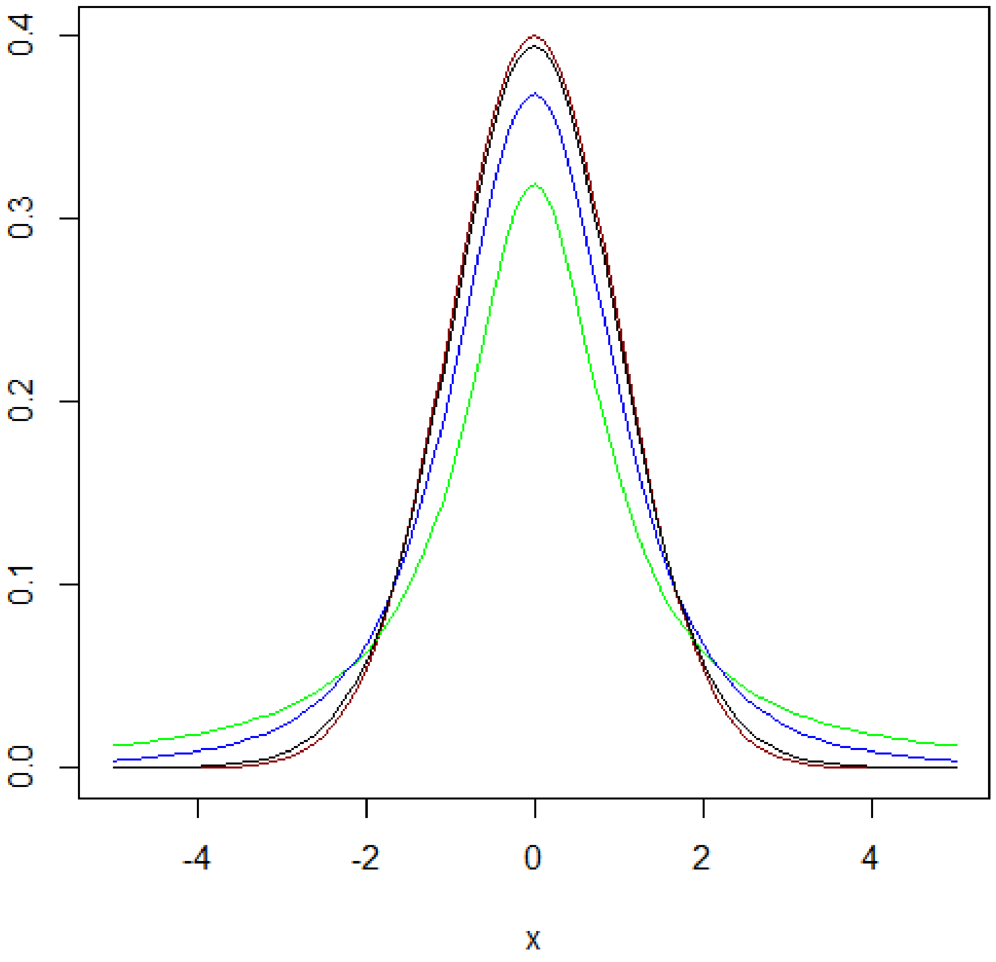
$$T = \frac{Z}{\sqrt{U/n}} \sim t_n$$

Se dice que *T* tiene distribución ***t* de Student con *n* grados de libertad**.

Esta distribución está tabulada para diferentes valores de *n*. Su densidad es simétrica respecto al 0 y tiene forma de campana, pero tiene colas más pesadas que la distribución normal standard.

Cuando n tiende a infinito, la distribución de Student tiende a la distribución normal standard.

```
x=-seq(-5,5,0.1)
y=dnorm(x)
y1=dt(x,1)
y3=dt(x,3)
y20=dt(x,20)
plot(x,y,type="n",ylab=" ")
lines(x,y,col="dark red")
lines(x,y1,col="green")
lines(x,y3,col="blue")
lines(x,y20,col="black")
```



Proposición: Sea $X_1, X_2, \ldots, X_n$ una m.a. de una distribución N(μ, σ²), entonces

a) $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right) \quad \Leftrightarrow \quad \sqrt{n}\,\dfrac{\overline{X}-\mu}{\sigma} \sim N(0,1)$

b) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ con $S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$

c) $\overline{X}$ y S^2 son independientes

d) $\sqrt{n}\,\dfrac{\overline{X}-\mu}{S} \sim t_{n-1}$

Intervalo de confianza para la media de la distribución normal con varianza desconocida:

Sea $X_1, X_2, \ldots, X_n$ una m.a. de una distribución N(μ, σ^2), entonces

$$\sqrt{n}\,\frac{\overline{X}-\mu}{S} \sim t_{n-1}$$

$$P\left(-t_{n-1,\alpha/2} \leq \sqrt{n}\,\frac{\overline{X}-\mu}{S} \leq t_{n-1,\alpha/2}\right) = 1-\alpha$$

de donde se deduce el siguiente **intervalo de confianza de nivel 1 - α para μ,**

$$\left[\overline{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \overline{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]$$

Volviendo al ejemplo:

Supongamos ahora que la varianza es desconocida, pero que el valor observado de S es s=35.

El correspondiente intervalo de confianza para μ será de la forma

$$\left(\bar{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right)$$

con $t_{n-1,\alpha/2} = t_{48,0.025} = 2.01$. Obtenemos

$$\left(160 - 2.01\frac{35}{\sqrt{49}}, 160 + 2.01\frac{35}{\sqrt{49}}\right) = (160 - 10.05, 160 + 10.05) = (149.95, 170.05)$$

y como es lógico, resulta un intervalo más largo.

Intervalo de confianza para la varianza de la distribución normal con media conocida

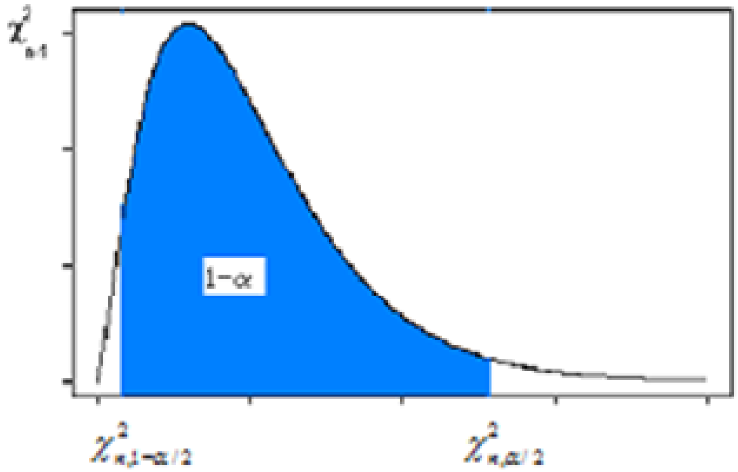
Sea $X_1, X_2, \ldots, X_n$ una m.a. de una distribución N(μ_o, σ^2), con media μ_o conocida, entonces

$$\frac{X_i - \mu_o}{\sigma} \sim N(0,1) \quad \forall 1 \leq i \leq n \quad \Rightarrow \quad \left(\frac{X_i - \mu_o}{\sigma}\right)^2 \sim \chi_1^2 = \Gamma\left(\frac{1}{2}, \frac{1}{2}\right) \quad \forall 1 \leq i \leq n$$

Como además las v.a. son independientes

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu_o}{\sigma}\right)^2 \sim \chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

¿Cómo elegimos los percentiles de la distribución χ^2 que encierran un área igual a 1 - α?



Los elegimos de manera tal que quede un área igual a α/2 en cada extremo. Entonces,

$$P\left(\chi^2_{n,1-\alpha/2} \leq \frac{\sum_{i=1}^{n}(X_i-\mu_o)^2}{\sigma^2} \leq \chi^2_{n,\alpha/2} \right) = 1-\alpha$$

Se obtiene el siguiente intervalo

$$\left[\frac{\sum_{i=1}^{n}(X_i-\mu_o)^2}{\chi^2_{n,\alpha/2}} , \frac{\sum_{i=1}^{n}(X_i-\mu_o)^2}{\chi^2_{n,1-\alpha/2}} \right]$$

Intervalo de confianza para la varianza de la distribución normal con media desconocida

Sea $X_1, X_2, \dots, X_n$ una m.a. de una distribución N(μ, σ^2), entonces

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

Por lo tanto,

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

Se obtiene el siguiente intervalo

$$\left[\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right]$$

Ahora un intervalo para σ

Suponiendo como antes que observamos $\bar{x} = 160$ y $s = 35$, hallemos un intervalo de confianza para σ de nivel 0.95.

Por tratarse de una muestra normal con media desconocida, el intervalo para σ^2 será de la forma

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} \right)$$

con $\chi^2_{n-1,\alpha/2} = \chi^2_{48,0.025} = 69.02$ y $\chi^2_{n-1,1-\alpha/2} = \chi^2_{48,0.975} = 30.75$ Obtenemos

$$\left(\frac{48 \cdot 35^2}{69.02}, \frac{48 \cdot 35^2}{30.75} \right) = (851.93, 1912.20)$$

y tomando raíz cuadrada, un intervalo de confianza para σ de nivel 0.95 será

$$\left(\sqrt{\frac{48 \cdot 35^2}{69.02}}, \sqrt{\frac{48 \cdot 35^2}{30.75}} \right) = \left(\sqrt{851.93}, \sqrt{1912.20} \right) = (29.19, 43.73)$$