# INFERENCIA ESTADÍSTICA

mariel*mates*

La inferencia estadística trata de conocer características de una población a través del estudio de una muestra. Esta información de la muestra se calcula en términos de probabilidad o de nivel de confianza.

Las características de la población desconocidas suelen ser los parámetros que definen la distribución.

Vamos a ver como ESTIMAR dichos paramentros desconocidos.

## ESTIMACIÓN PUNTUAL

Es el resultado de la estimación con un valor numérico con un cierto grado de confianza.

## ESTIMACIÓN POR INTERVALOS DE CONFIANZA

Es el resultado de la estimación con un conjunto de valores con un cierto grado de confianza.

Llamamos $\bar{x}$ a la media de la muestra

$$\bar{x} = \sum x_i \cdot p_i$$

Llamamos **s** a la desviación típica de la muestra

$$s = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

Llamamos $\hat{p}$ a la proporción muestral

Llamamos $\mu$ a la media de la población, $\sigma$ a la desviación típica de la población y p la proporción poblacional. Estos serán los PARÁMETROS de la población que vamos a estimar.

La media muestral, la desviación típica muestral y la proporción muestral serán los ESTADÍSTICOS (valores de la muestra) que vamos a utilizar.

## TEOREMA CENTRAL DEL LIMITE

Si X es una variable aleatoria con media $\mu$ y desviación típica $\sigma$, entonces se cumple que la DISTRIBUCIÓN DE LA MEDIA MUESTRAL $\bar{X}$ de tamaño n, se aproxima a una distribución normal con media $\mu$ y desviación típica $\frac{\sigma}{\sqrt{n}}$, siempre que el tamaño de la muestra sea grande (n>30).

$$\text{Si } X \sim (\mu, \sigma) \xrightarrow[n>30]{} \boxed{\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)} \rightarrow Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \boxed{\sum X \underset{n>30}{\sim} N(n\mu, \sigma\sqrt{n})}$$

OBSERVACIÓN: $\text{Si } X \sim N(\mu, \sigma) \rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \rightarrow Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \sum X \sim N(n\mu, \sigma\sqrt{n})$

Aplicando el teorema central del límite podemos averiguar la DISTRIBUCIÓN DE LA PROPORCIÓN MUESTRAL:

$$X \sim Bin(n,p) \xrightarrow[n\geq 30,\ np>5,\ n(1-p)>5]{} X \sim AN\left(np, \sqrt{np(1-p)}\right) \rightarrow \boxed{\hat{p} = \frac{\sum X_i}{n} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)} \rightarrow Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

# INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL $\mu$



UN **INTERVALO DE CONFIANZA** ES UN RANGO DE VALORES ENTRE LOS CUALES SE ESTIMA QUE ESTARÁ EL VALOR VERDADERO DE UN PARÁMETRO CON UN DETERMINADO NIVEL DE CONFIANZA.

Suponemos que disponemos de una muestra aleatoria simple de n observaciones. La población sigue una distribución normal y conocemos la varianza poblacional $\sigma^2$.

**OBJETIVO**: CONSTRUIR UN INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL $\mu$ DESCONOCIDA.

Sea $X_1, \ldots, X_n$ la muestra aleatoria simple y $\bar{X}$ su media muestral (nuestro estimador puntual)

$$\text{Si } X \sim N(\mu, \sigma) \rightarrow \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \rightarrow Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

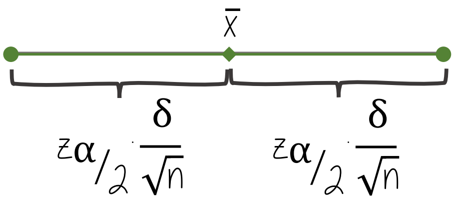
$$P\left(-z_{\alpha/2} \leq Z \leq -z_{\alpha/2}\right) = 1-\alpha$$

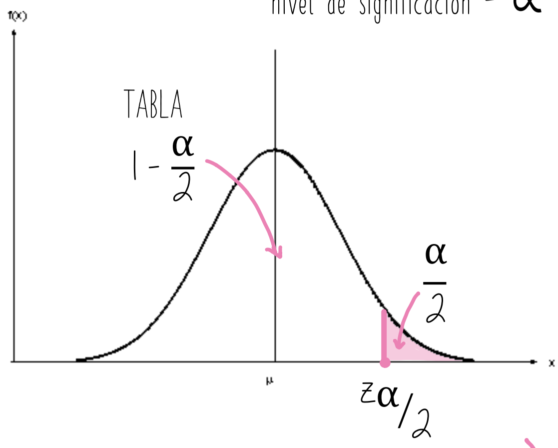
nivel de confianza $= 1-\alpha$

nivel de significación $= \alpha$

$$IC(\mu) = \left(\bar{x} - z_{\alpha/2} \cdot \frac{\delta}{\sqrt{n}},\ \bar{x} + z_{\alpha/2} \cdot \frac{\delta}{\sqrt{n}}\right)$$

$$P(\mu \in IC(\mu)) = 1-\alpha$$




($z_{\alpha/2}$: Valor de la distribución normal que deja a la derecha una probabilidad de $\frac{\alpha}{2}$)

ERROR EN LA ESTIMACIÓN $= \boxed{z_{\alpha/2} \cdot \frac{\delta}{\sqrt{n}}}$

Error máximo cometido en la estimación de la media poblacional por la media muestral (semilongitud del intervalo)

LA LONGITUD DEL INTERVALO DEPENDE DE
- TAMAÑO MUESTRA
- NIVEL DE CONFIANZA
- VARIANZA POBLACIONAL

CUANTO MAYOR LA DESVIACIÓN TÍPICA DE LA POBLACIÓN $\rightarrow$ MENOS FIABLE SERÁ LA ESTIMACIÓN (MAYOR ERROR)

CUANTO MAYOR SEA EL TAMAÑO DE LA MUESTRA $\rightarrow$ MÁS FIABLE SERÁ LA ESTIMACIÓN (MENOR ERROR)

CUANTO MAYOR EL NIVEL DE CONFIANZA $\rightarrow$ MAYOR SERÁ EL TAMAÑO DEL INTERVALO (MAYOR ERROR)

# INTERVALO DE CONFIANZA PARA LA PROPORCIÓN EN MUESTRAS GRANDES

La estimación de proporciones es un caso particular del caso anterior con datos no normales. Nuestro estimador en este caso será la proporción muestral.

$$X \sim Bin(n,p) \rightarrow X \sim AN\left(np, \sqrt{np(1-p)}\right) \rightarrow \hat{p} = \frac{\sum X_i}{n} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) \rightarrow Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

$n \geq 30,\ np>5\ n(1-p)>5$

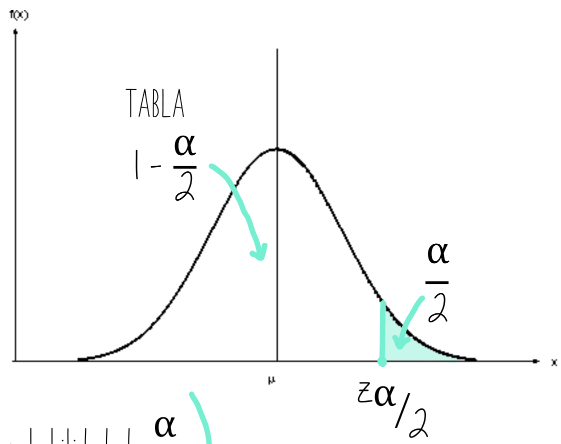
$$IC(p) = \left(\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}},\ \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \qquad q = 1-p$$

$$P(p \in IC(p)) = 1-\alpha$$

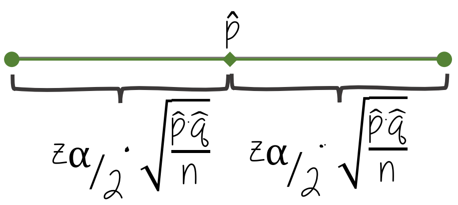
$$P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1-\alpha$$

nivel de confianza $= 1-\alpha$

nivel de significación $= \alpha$




$\left( z_{\alpha/2}\right.$: Valor de la distribución normal que deja a la derecha una probabilidad de $\left.\frac{\alpha}{2}\right)$




ERROR EN LA ESTIMACIÓN $= \boxed{z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}}$

Error máximo cometido en la estimación de la proporción poblacional por la proporción muestral (semilongitud del intervalo)

CUANTO MAYOR SEA EL NIVEL DE CONFIANZA $\longrightarrow$ MAYOR SERÁ EL TAMAÑO DEL INTERVALO (MAYOR ERROR)

CUANTO MAYOR SEA EL TAMAÑO DE LA MUESTRA $\longrightarrow$ MÁS FIABLE SERÁ LA ESTIMACIÓN (MENOR ERROR)