INFERENCIA ESTADÍSTICA

La *Inferencia Estadística* es la parte de la estadística matemática que se encarga del estudio de los métodos para la obtención del modelo de probabilidad (forma funcional y parámetros que determinan la función de distribución) que sigue una variable aleatoria de una determinada población, a través de una muestra (parte de la población) obtenida de la misma

DISTRIBUCIÓN DE LA MEDIA MUESTRAL

Media de la población: μ   Desviación típica de la población: σ
Medida de la muestra: $\bar{X}$ Desviación típica de la muestra: s_n

Media muestral:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Varianza y desviación típica:

Varianza muestral = $\frac{1}{n}\sum(x_i - \bar{x})^2$; s_n = $\sqrt{Varianza\ muestral}$

$$s_n = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$$

Como la distribución de la población es N(μ, σ), la media muestral $\bar{X}$ sigue una distribución normal de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$. Es decir:

$\bar{X} = \rightarrow N(\mu, \frac{\sigma}{\sqrt{n}})$

Luego, la variable tipificada sigue una distribución normal N(0, 1):

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0,1)$$

DISTRIBUCIÓN DE LA PROPORCIÓN MUESTRAL

Media y desviación típica:

$$\boldsymbol{\mu = np; \ \sigma = \sqrt{npq}}$$

Supongamos que sea X la variable que mide el número de éxitos. Ya sabes que los posibles valores de X son 0,1,2,...,n. Si utilizáramos la nueva variable, (**proporción muestral**):

$$\hat{p} = \frac{x}{n}$$

$$N\left(p, \sqrt{\frac{pq}{n}}\right)$$

ESTIMACIÓN POR INTERVALO

El **nivel de confianza** se designa mediante **1 - α**.
El **nivel de significación** se designa mediante **α.**
El valor crítico (k) como zα/2 .

Si la muestra aleatoria es de tamaño n y media $\bar{x}$; para saber el intervalo de confianza con nivel 1 – α, tenemos que hallar los valores:

$-z_{\frac{\alpha}{2}}\, y\, z_{\frac{\alpha}{2}} \;\rightarrow\; p(-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}})$

De modo que $\mathrm{z} = \frac{\bar{\mathrm{x}} - \mu}{\frac{\sigma}{\sqrt{\mathrm{n}}}} \;\rightarrow\; \mathrm{N}(0{,}1)$

$-z_{\frac{\alpha}{2}}\, y\, z_{\frac{\alpha}{2}}$ se expresan mediante:

$$p\left(z < -z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \quad y\; p\left(z > z_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

El intervalo para la media μ a un nivel de confianza 1 – α es:

$$\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

Intervalos característicos

1 – α	**α/2**	$\mathbf{z_{\alpha/2}}$	**intervalo**
0’90	0’05	1’645	(μ – 1’645 · σ, μ + 1’645 · σ)
0’95	0’025	1’96	(μ – 1’96 · σ, μ + 1’96 · σ)
0’99	0’005	2’575	(μ – 2’575 · σ, μ + 2’575 · σ)

Cuando **σ es desconocida**, lo lógico es sustituirla por su estimador *s*, obteniendo así que

$$\boldsymbol{\mu \in \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}}$$

$$\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

Intervalo de confianza para una población no normal.

Tendremos en cuenta que partimos de una población no normal con n > 30; por tanto:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow n(0'1)$$

En este caso el intervalo es el mismo que para una distribución normal:

$$\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$$

Intervalo de confianza para una población no normal, con σ desconocida

Cuando la población es superior a 30 se utiliza:

$$\left(\bar{x} - z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

Intervalo de confianza para la proporción p de una población

Consideramos una distribución binomial B(1,p) donde:

μ = p; σ = $\sqrt{\frac{pq}{n}}$ n >30 y $\hat{p}$ = estimación proporcional de p

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

TAMAÑO DE LA MUESTRA

Si conocemos σ:

E $= z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$, y despejando n queda:

$$n = \frac{z_{\alpha/2}^2 \cdot \sigma^2}{E^2}$$

Tamaño de la muestra con σ desconocida

$n = \frac{z_{\alpha/2}^2 \cdot s^2}{E^2}$ donde s es la desviación típica muestral.

Tamaño de la muestra para estimar la proporción p de una población
Partimos de:

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

El error de la estimación será:

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

De donde se saca n:

$$n = \frac{z_{\alpha/2}^2 \cdot \hat{p}\hat{q}}{E^2}$$

EJEMPLO:

Se ha extraído una muestra de 150 familias de residentes en un barrio obteniéndose que la renta familiar media de la misma asciende a 20000 €. Se supone que la renta familiar de los residentes en el barrio sigue una distribución normal de desviación típica 1500 €.
a) A partir de estos datos, calcule un intervalo de confianza para la renta familiar media con un nivel de confianza del 93%.
b) ¿Qué tamaño muestral mínimo es necesario para conseguir, con un nivel de confianza del 90%, un error en la estimación de la renta familiar media no superior a +;- 42 €?

a) Sacamos los datos: n = 150; σ = 1500; $\bar{x} = 20.000$; Nivel de confianza = 93%

1 – α = 0'93 → α = 0'07 → α/2 = 0'035

1- α/2 = 1 – 0'035 = 0'965. Buscamos en la tabla y vemos que es un valor que está entre 1'81 y 1'82, entonces hacemos la media y nos da que $z_{\alpha/2}$ = 1'815

Los extremos del intervalo son: $\mu = \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 20.000 \pm 1'815 \cdot \frac{1500}{\sqrt{150}} = 20.000 \pm 222'30$

Y el intervalo es (19777'7; 20222'30)

b) nivel de confianza = 90%; Error = ± 42 €

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 42$$

1 – α = 0'90 → α = 0'10 → α/2 = 0'05 → 1- α/2 = 1 – 0'05 = 0'95 que en las tablas nos da el valor de $z_{\alpha/2}$ = 1'645

$$E = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq 42 \quad \rightarrow \quad 1'645 \cdot \frac{1500}{\sqrt{n}} \leq 42$$

$$n \geq \left(1'645 \cdot \frac{1500}{42}\right)^2 \approx 3452$$