8 Muestreo, Inferencia e Hipótesis

Muestras

En el estudio de una característica:

- **Población** o **Universo** es el conjunto de todos los individuos objeto del estudio.
- **Muestra** es cualquier subconjunto extraído de la población.

Se recurre a las muestras por causas diversas: porque la población es numerosa, debido a que es imposible controlar la totalidad de los individuos, porque el proceso de medición es destructivo o porque consultar a toda la población sería lento y/o costoso.

El estudio de la muestra sirve para **inferir** características de toda la población.

Muestreos

Al elegir una muestra tendremos en cuenta su **tamaño** y la selección de sus individuos.

Se dice que una muestra es **representativa** cuando está bien elegida, pudiendo obtenerse conclusiones razonables para toda la población a partir de ella.

En cualquier caso, se producirán errores imprevistos e incontrolables, que se denominan **sesgos**.

Muestreo es el proceso de formación de la muestra.

Se dice que es **aleatorio** cuando todos los individuos se eligen al azar, de modo que todos tienen, a priori, la misma probabilidad de ser elegidos. Pueden ser:

- Aleatorios **simples**: se enumeran los individuos y se sortean los elegidos.
- Aleatorio **estratificado**: la población se divide previamente en **estratos**. Al estudiar la muestra, elegimos aleatoriamente de cada uno de los estratos un nº de individuos **proporcionalmente** al peso del estrato en la población.

Para obtener muestras aleatorias se hace uso de los denominados **números aleatorios**.

En una población de tamaño N y enumerada, un individuo elegido al azar es:

$$E\left(N \cdot aleatorio + 1\right)$$

donde $0 \leq aleatorio < 1$ es un número generado aleatoriamente.

Medias Muestrales Finitas

Sea **X** una v.a. con media μ y desviación típica σ en una población de tamaño N , y $\overline{\mathbf{X}}$ la distribución de las medias muestrales de tamaño n. Se cumple:

- Las medias son iguales: $\overline{\mu} = \mu$
- Las desviaciones típicas (con reemplazamiento son)

$$\overline{\sigma} = \frac{\sigma}{\sqrt{n}}$$

- Si las muestras se toman sin reemplazamiento es:

$$\overline{\sigma} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Medias Muestrales

Sea **X** una v.a. con media μ y desviación típica σ .

Si es normal o si n es suficientemente grande $(n \geq 30)$ entonces :

$$\overline{\mathbf{X}} \sim \mathcal{N}\left(\mu\,, \frac{\sigma}{\sqrt{n}}\right)$$

Esto se conoce como **Teorema Central del Límite**.

Proporciones Muestrales.

Sea una población en la que la característica **C** aparece con proporción p. Si $\widetilde{p}$ es la proporción de **C** en una muestra de tamaño n suficientemente grande($np > 5$ y $nq > 5$), entonces:

$$\widetilde{\mathbf{P}} \sim \mathcal{N}\left(p\,, \sqrt{\frac{p \cdot q}{n}}\right)$$

Intervalo de Confianza para la Media

Sea **X** una v.a. en una población de la que desconocemos μ y conocemos σ.

Extraemos una muestra de tamaño n, que será suficientemente grande $(n \geq 30)$ en caso de no ser **X** normal, y obtenemos la media muestral $\overline{x}$.

Para **estimar** μ mediante una muestra de tamaño n, con un **nivel de confianza** $p = -\alpha$, se utiliza un **intervalo de confianza**

$$I = \left(\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\,, \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Esto significa que en el $(1-\alpha) \cdot 100\%$ de los casos la media poblacional está en un intervalo así obtenido.

8 Muestreo, Inferencia e Hipótesis

Error al Estimar la Media

Se llama **error máximo** admisible al número

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

Observemos así que el intervalo de confianza es:

$$I = (\overline{x} - E\,, \overline{x} + E)$$

y su **amplitud** o longitud es

$$L = 2E$$

Nota: si σ fuese desconocida, se sustituye por $\overline{s}$, donde $\overline{s}^2$ es la cuasivarianza muestral.

Intervalo de Confianza para una Proporción

En una población desconocemos la proporción p de individuos que cumple una determinada condición.

Extraemos una muestra de tamaño n suficientemente grande y obtenemos la proporción muestral $\tilde{p}$.

Para **estimar** p con un **nivel de confianza** $p = 1 - \alpha$, se utiliza el **intervalo**

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

Error al Estimar la Proporción

Se llama **error máximo** admisible al número

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

Observemos así que el intervalo de confianza es:

$$I = (\widetilde{p} - E\,, \widetilde{p} + E)$$

y su **amplitud** o longitud es

$$L = 2E$$

Contraste de Hipótesis: Planteamiento

1. Se plantea la hipótesis, sobre la media o sobre la proporción poblacional:

   H_0: hipótesis nula (emitida)

   H_1: hipótesis alternativa (contraria a la emitida)

   Si la hipótesis es una igualdad (=) se dice que es bilateral; si es una desigualdad (≤ ó ≥) se dice que es unilateral.

2. Se elige una muestra aleatoria significativa (tamaño n) y se obtiene de ella el parámetro estadístico de contraste (media muestral ó proporción muestral).

3. Se fija un nivel de significación $\alpha(\%)$. Inmediatamente se calcula el valor crítico asociado:

   Valor crítico bilateral $z_{\alpha/2}$: $\quad \mathrm{p}\left(z > z_{\alpha/2}\right) = \dfrac{\alpha}{2}$

   Valor cŕitico unilateral z_{α} : $\quad \mathrm{p}\left(z > z_{\alpha}\right) = \alpha$

4. Hallamos el intervalo de aceptación (I) y comprobamos si el estadístico está en él:

   Si está en el I → Se acepta H_0

   Si no está en I → Se rechaza H_1

   Se denomina zona de rechazo o región crítica al complementario del intervalo de aceptación.

5. Hay dos posibles errores:

   Error de tipo I: rechazar una hipótesis verdadera.

   La probabilidad de cometerlo es α y es independiente de n .

   Error de tipo II: aceptar una hipótesis falsa.

   La probabilidad de cometerlo depende del verdadero valor del parámetro. Disminuye al aumentar n .

Contraste de Hipótesis: Casos

HIPÓTESIS	TIPO	ESTADÍSTICO	NIVEL SIGNIFICACIÓN	INTERVALO DE ACEPTACIÓN
$\mu = \mu_0$	Bilateral	$\overline{x}$	$\alpha(\%) \to z_{\alpha/2}$	$\left(\mu_0 - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\,, \mu_0 + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$
$\mu \geq \mu_0$	Unilateral	$\overline{x}$	$\alpha(\%) \to z_{\alpha}$	$\left(\mu_0 - z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}}\,, +\infty\right)$
$\mu \leq \mu_0$	Unilateral	$\overline{x}$	$\alpha(\%) \to z_{\alpha}$	$\left(-\infty\,, \mu_0 + z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}}\right)$
$p = p_0$	Bilateral	$\widetilde{p}$	$\alpha(\%) \to z_{\alpha/2}$	$\left(p_0 - z_{\alpha/2} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}\,, p_0 + z_{\alpha/2} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}\right)$
$p \geq p_0$	Unilateral	$\widetilde{p}$	$\alpha(\%) \to z_{\alpha}$	$\left(p_0 - z_{\alpha} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}\,, +\infty\right)$
$p \leq p_0$	Unilateral	$\widetilde{p}$	$\alpha(\%) \to z_{\alpha}$	$\left(-\infty\,, p_0 + z_{\alpha} \cdot \sqrt{\frac{p_0 \cdot q_0}{n}}\right)$