5 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:

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

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

PARÁMETROS

Variable en una población:	X
Distribución de medias muestrales:	$\overline{X}$
Nivel de significación:	α.
Nivel de confianza:	$(1-\alpha)\cdot 100\%$.
Valor crítico unilateral:	z_{α}.
Valor crítico bilateral:	$z_{\alpha/2}$.
Tamaño población:	N.
Tamaño muestral:	n.
Media población:	μ.
Media muestral:	$\overline{x}$.
Media de las medias muestrales:	$\overline{\mu}$.
Desv. típica de las medias muestrales:	$\overline{\sigma}$.
Proporción poblacional:	p.
Proporción muestral:	$\widetilde{p}$.

MEDIAS MUESTRALES FINITAS

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

- Ambas tienen la misma media: $\overline{\mu}=\mu$.
- Si las muestras se toman con reemplazamiento es:

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

- Si las muestras se toman sin reemplazamiento es:

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

MEDIAS MUESTRALES

Si X es $N(\mu,\sigma)$, entonces:

$$\overline{X} \text{ es } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Para n suficientemente grande ($n \geq 30$), siempre

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

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

5 MUESTREO, INFERENCIA E HIPÓTESIS

VALOR CRÍTICO

Nivel de significación α (confianza = $(1-\alpha)\cdot 100\%$)

El valor crítico bilateral cumple: $p(z>z_{\alpha/2})=\frac{\alpha}{2}$.

Valor crítico unilateral cumple: $p(z>z_{\alpha})=\alpha$.

ESTIMACIÓN DE 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 $\bar{x}$.

Para **estimar** μ , con un **nivel de confianza** $(1-\alpha)\cdot 100\,\%$, se utiliza el **intervalo de confianza**

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

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

Se llama **error máximo** admisible a

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

La amplitud del intervalo de confianza es igual al doble del error máximo.

ESTIMACIÓN DE 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 de confianza**

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

Se llama **error máximo** admisible a

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

La amplitud del intervalo de confianza es igual al doble del error máximo.

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** α (%). Inmediatamente se calcula el valor crítico asociado.

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 de del verdadero valor del parámetro. Disminuye al aumentar n .

cc BY NC SA

CONTRASTE DE HIPÓTESIS: CASOS

HIPÓTESIS	TIPO	ESTADÍSTICO	NIVEL SIGNIFICACIÓN	INTERVALO DE ACEPTACIÓN
$\mu=\mu_0$	Bilateral	$\overline{x}$	$\alpha(\%) \rightarrow z_{\alpha/2}$	$\left(\mu_0-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\ ,\ \mu_0+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)$
$\mu\geq\mu_0$	Unilateral	$\overline{x}$	$\alpha(\%) \rightarrow z_{\alpha}$	$\left(\mu_0-z_{\alpha}\cdot\frac{\sigma}{\sqrt{n}}\ ,\ +\infty\right)$
$\mu\leq\mu_0$	Unilateral	$\overline{x}$	$\alpha(\%) \rightarrow z_{\alpha}$	$\left(-\infty\ ,\ \mu_0+z_{\alpha}\cdot\frac{\sigma}{\sqrt{n}}\right)$
$p=p_0$	Bilateral	$\tilde{p}$	$\alpha(\%) \rightarrow z_{\alpha/2}$	$\left(p_0-z_{\alpha/2}\cdot\sqrt{\frac{p_0q_0}{n}}\ ,\ p_0+z_{\alpha/2}\cdot\sqrt{\frac{p_0q_0}{n}}\right)$
$p\geq p_0$	Unilateral	$\tilde{p}$	$\alpha(\%) \rightarrow z_{\alpha}$	$\left(p_0-z_{\alpha}\cdot\sqrt{\frac{p_0q_0}{n}}\ ,\ +\infty\right)$
$p\leq p_0$	Unilateral	$\tilde{p}$	$\alpha(\%) \rightarrow z_{\alpha}$	$\left(-\infty\ ,\ p_0+z_{\alpha}\cdot\sqrt{\frac{p_0q_0}{n}}\right)$

cc BY NC SA