8 MUESTREO E INFERENCIA

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.
- Aleatorios **sistemáticos**: se enumeran los individuos y, a partir de uno de ellos –**origen**– elegido al azar, se toman los siguientes mediante saltos numéricos –**coeficiente de elevación**– iguales.
- 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 $\overline{X}$ 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 Y DISTRIBUCIÓN NORMAL

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

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

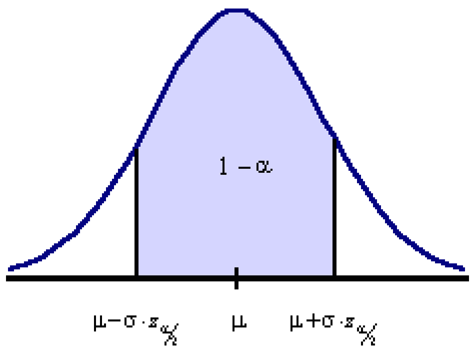
Esto se conoce como **Teorema Central del Límite**.

8 MUESTREO E INFERENCIA

INTERVALO CARACTERÍSTICO

En una distribución $N(\mu,\sigma)$ el **intervalo característico** correspondiente a la probabilidad $p=1-\alpha$ es:

$$I=\left(\mu-z_{\alpha/2}\cdot\sigma\ ,\ \mu+z_{\alpha/2}\cdot\sigma\right)$$



donde $p\left(-z_{\alpha/2}<z<z_{\alpha/2}\right)=1-\alpha$.

De ahí se deduce que $p\left(z>z_{\alpha/2}\right)=\dfrac{\alpha}{2}$

A $p=1-\alpha$ se la llama **nivel de confianza**, y a $z_{\alpha/2}$ se le denomina **valor crítico**.

Los valores que más aparecen son:

- $90\% \rightarrow 1-\alpha=0'90 \rightarrow z_{\alpha/2}=1'645$
- $95\% \rightarrow 1-\alpha=0'95 \rightarrow z_{\alpha/2}=1'96$
- $99\% \rightarrow 1-\alpha=0'99 \rightarrow z_{\alpha/2}=2'575$

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 $\overline{x}$.

Para **estimar** μ mediante una muestra de tamaño *n*, con un **nivel de confianza** $p=1-\alpha$, se utiliza un **intervalo de confianza**

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

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

Se llama **error máximo** admisible a

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

Si σ es desconocido, se sustituye por $\overline{s}$, donde es $\overline{s}^2$ la cuasivarianza muestral.

PROPORCIONES MUESTRALES.

Sea una población en la que la característica ***C*** aparece con proporción *p*.

Si $\tilde{p}$ es la proporción de ***C*** en una muestra de tamaño *n*, suf. grande ($np\geq 5$ y $nq\geq 5$), entonces:

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

ESTIMACIÓN DE UNA PROPORCIÓN.

En una población desconocemos la proporción p de individuos de un colectivo 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**

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

Se llama **error máximo** admisible a

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