Inferencia Estadística

Sacar conclusiones sobre la población a partir de muestras

Matemáticas Aplicadas II

¿Qué es la inferencia estadística?

Tenemos información parcial que obtenemos **a partir de una muestra**

↓

Queremos sacar conclusiones sobre **toda la población**

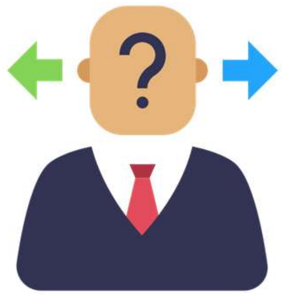
Ejemplo

Queremos saber qué porcentaje de jóvenes, que cursan estudios en A Coruña, usan el transporte público.

- ❌ No podemos preguntar a todos
- ✅ Así que tomamos una muestra representativa de 100 alumnos

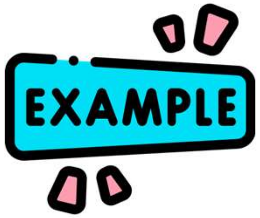
A partir de esa muestra ***estimamos*** **la realidad**

real vs estimación



- **POBLACIÓN** → toda la población sobre la que realizamos el estudio
- **MUESTRA** → *parte de esa población*
- **Parámetro** → valor real (desconocido) puede ser μ o p
- **Estadístico** → *valor calculado (estimado) con la muestra , puede ser $\bar{X}$ media muestral, o $\hat{p}$ proporción muestral*

real vs estimación



En el ejemplo anterior:

- **POBLACIÓN** → todos los jóvenes que cursan estudios en esta ciudad
- **MUESTRA** → 100 estudiantes
- **Parámetro,** proporción de jóvenes estudiantes que son usuarios de transporte público → p (valor real desconocido)
- **Estadístico** → $\hat{p}$ proporción muestral, valor calculado a partir de la muestra

IMPORTANTE

La estimación NUNCA es exacta

siempre hay error

Dos formas distintas:

Hay dos formas de hacerlo:

- **Estimación puntual:** la idea es obtener un único valor calculado a partir de las observaciones de las muestras. (Poco habitual)

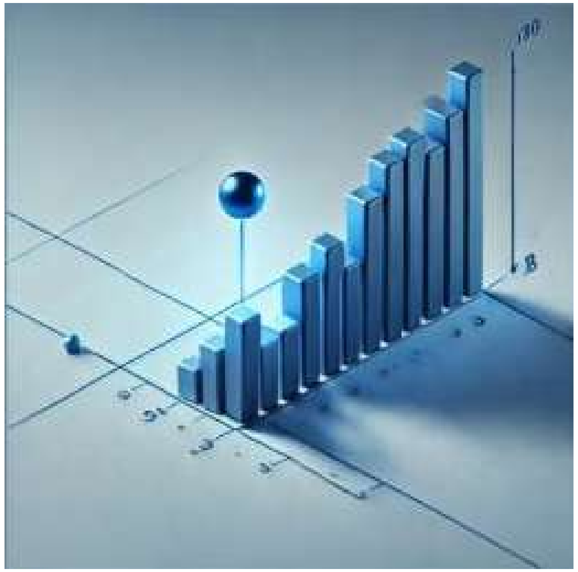
- **Intervalos:** utilizamos los datos de una muestra para obtener un intervalo dentro del cual se espera que se encuentre el parámetro con cierto nivel de confianza.

Estimación puntual

Buscamos un único valor para estimar un parámetro

Tipos de ejercicios:

- $\bar{X}$ **media muestral** → estima la media **μ**
- $\hat{p}$ **proporción muestral** → estima **p**



$\bar{X}$ Media muestral

Queremos saber la altura media de los alumnos de un instituto

¿Qué pasa si repetimos la muestra muchas veces?

- **Paso 1:** tomamos una primera muestra al azar de 10 alumnos y calculamos su media = 1,70 m
- **Paso 2:** repetimos el proceso, cogemos aleatoriamente otros 10, ahora la media = 1,68
- **Y seguimos…** otra muestra aleatoria → media = 1,72 m

¿Qué está pasando? Cada muestra da una media distinta, debido al azar

$\bar{X}$ Media muestral

Si tomamos MUCHAS muestras: iremos obteniendo diferentes medias: 1,65 1,68 1,70 1,72 1,69 ...

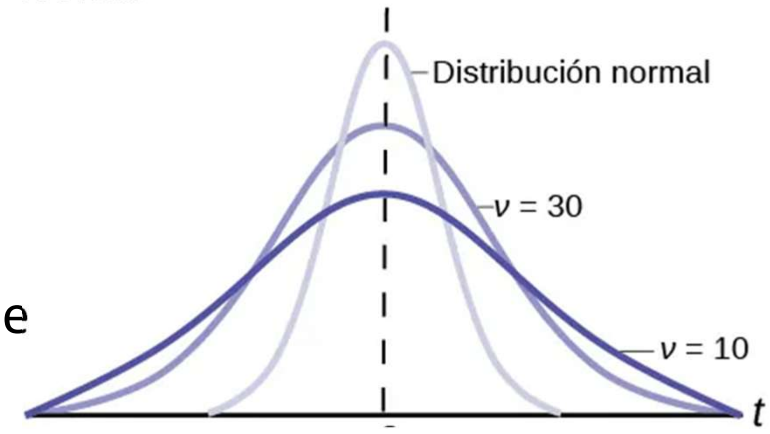
esas medias siguen una distribución

ESO es la distribución muestral, **la distribución de todas las medias posibles**

$\overline{X}$ Media muestral

¿Cómo es esa distribución?

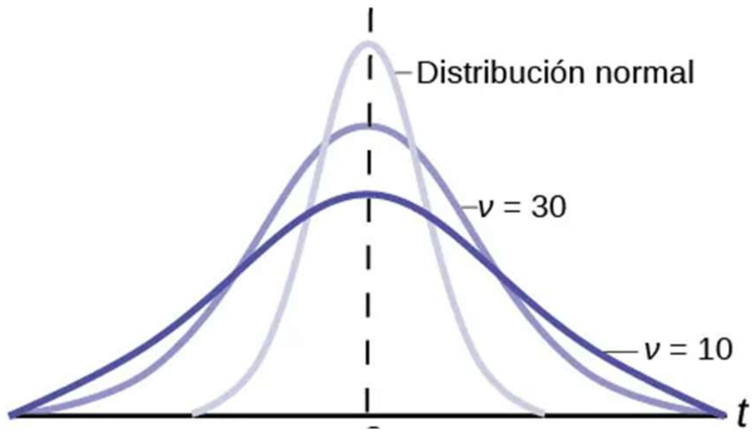
- Tiene forma de campana (campana normal o de Gauss)
- Está centrada en la media real (en el ejemplo anterior, la estatura media de todos los alumnos de este instituto)



$\bar{X}$ Media muestral

Las medias de las muestras:

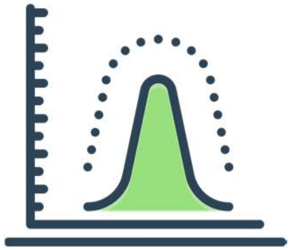
- no son todas iguales
- pero se agrupan alrededor del valor real de la media



$\bar{X}$ Media muestral

Observemos dos situaciones distintas:

- **Caso A:** Muestras de 5 alumnos, medias: 1,60 – 1,80 (muy dispersas)
- **Caso B:** Muestras de 50 alumnos, medias: 1,68 – 1,72 (muy concentradas)



Cuanto mayor es la muestra, más cerca estás de la realidad

$\bar{X}$ Media muestral

La media, es la misma

Pero **la desviación típica** (dispersión) **cambia**

desviación ↓ disminuye cuando n ↑ aumenta

- La media de las medias muestrales = μ (el centro correcto, la media real)
- Cuanto mayor es n → menos varían las medias
- más muestra → más precisión

$\bar{X}$ Media muestral

TEOREMA CENTRAL DEL LÍMITE

En el caso que la población no sea normal, si la muestra es lo bastante grande ($n \geq 30$), la distribución que sigue la **media muestral** $\bar{X}$ es **siempre** una distribución normal

$$\bar{X} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

$\bar{X}$ Media muestral

Si llamamos $\bar{X}$ a la media muestral

- la media de las medias muestrales $\mu_{\bar{X}} = \mu$ coincide con la media poblacional
- La desviación típica de la media muestral **NO** es la desviación típica poblacional, sino

$$\boxed{\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}}$$

Esquema

Repetimos muestras

↓

Obtenemos distintos resultados

↓

Esos resultados siguen una distribución

↓

Cuanto mayor es la muestra → menor variabilidad

RECUERDA

- ✓ no hay una única media
- ✓ hay muchas posibles
- ✓ pero siguen una ley (normal)
- ✓ y eso permite calcular probabilidades

Ejemplo

La altura de los estudiantes de una población se distribuye según una normal de media 167 cm y desviación típica de 3,2. Se toma una muestra de 10 estudiantes. Calcula la probabilidad de que la media muestral sea menor que 165 cm.

Claves:

- **Identificar** tipo de ejercicio: **media muestral** , estimación puntual (piden una probabilidad)
- Aplicar la **fórmula** correspondiente:

$$\bar{X} \sim N\left(167, \frac{3{,}2}{\sqrt{10}} = 1{,}012\right)$$

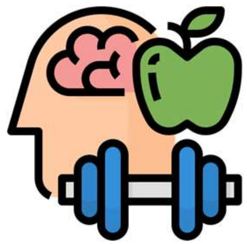
- **Tipificar**

$$Z = \frac{\bar{X} - 167}{1{,}012} \sim N(0{,}1)$$

- Usar **tabla** normal $P(\bar{X} < 165) = P\left(Z < \frac{165-1}{1{,}012}\right) = P(Z < -1{,}92) = P(Z > 1{,}92) = 1 - P(Z < 1{,}92) = 0{,}0244$

- **Interpretar** el resultado: La probabilidad de que la media de una muestra de **10** individuos sea menor que **167** cm es 0,0244, lo que indica que es un suceso **muy poco frecuente**.

Ejercicio 1



En una población, la altura media es de 170 cm y la desviación típica es de 10 cm. Se toman muestras de tamaño 25.

Preguntas:

a) ¿Cuál es la distribución de la media muestral?

b) ¿Cuál es la probabilidad de que la media muestral sea mayor que 172 cm?

Claves para hacer el ejercicio:

- Identificar tipo: **media muestral**
- **fórmula** correspondiente:

$$\bar{X} \sim N\left(170, \frac{10}{\sqrt{25}} = 2\right) \bar{X} \sim N(170, 2)$$

- **Tipificar** y usar **tabla normal:**

$$Z = \frac{\bar{X} - \mu}{\sigma_{\overline{\mathrm{X}}}} = \frac{\bar{X} - 170}{2} \sim \mathrm{N}(0,1)$$

$$\mathrm{P}(\overline{\mathrm{X}} > 172) = \mathrm{P}\left(\mathrm{Z} > \frac{172 - 170}{2}\right) = \mathrm{P}(\mathrm{Z} > 1) = 1 - \mathrm{P}(\mathrm{Z} < 1)$$

$$= 1 - 0{,}8413 = 0{,}1587$$

Interpretación:
La media real es 170, pero algunas muestras pueden dar valores mayores

La probabilidad de que la media de una muestra de **25** individuos sea superior a **172** cm es aproximadamente **0,1587** lo que indica que es un suceso **poco frecuente**.

cerca del 16% de las muestras tendrán una media mayor que 172

Ejercicio 2

PAU 2021 Extraordinaria

El peso de las naranjas para zumo recolectadas por un productor es una variable aleatoria que se distribuye normalmente con una media de μ=200 gramos y una desviación típica de σ=50 gramos.

> **a)** Si tomamos una muestra aleatoria de n=25 naranjas, ¿cuál es la probabilidad de que su peso medio está comprendido entre 175 y 215 gramos?
>
> **b)** ¿De qué tamaño se ha tomado otra muestra aleatoria si la probabilidad de que el peso medio sea inferior a 210 gramos es del 97,72%?

Errores típicos

- ❌ No dividir entre $\sqrt{n}$
- ❌ Usar σ en vez de $\frac{\sigma}{\sqrt{n}}$
- ❌ No hacer 1 - tabla

$\hat{p}$ proporción muestral

Distribución muestral de proporciones

El problema de origen es binomial y queremos aproximar la proporción muestral *p*

Gracias al teorema central del límite sabemos que si **n** es grande $(n \geq 30)$:

Ahora la fórmula es $\boxed{\hat{p} \sim N\left(p; \sqrt{\frac{p \cdot q}{n}}\right)}$

Muestras grandes → resultados más estables

$\hat{p}$ proporción muestral

Una industria de pasteles fabrica, en su producción habitual, un 3% de pasteles defectuosos. Un cliente recibe un pedido de 500 pasteles de la fábrica.

Calcular la probabilidad de que encuentre más del 4% de pasteles defectuosos.

$\widehat{p}$ proporción muestral

Claves para hacer el ejercicio:

- Identificar tipo de ejercicio: **proporción muestral $\widehat{p}$**
- Aplicar la fórmula correspondiente:

$$\hat{p} \sim N\left(0{,}03, \sqrt{\frac{0{,}03 \cdot 0{,}97}{500}} \approx 0{,}076\right) = N(0{,}03, 0{,}076)$$

- ¿Qué nos piden? $P(\hat{p} > 0{,}04)$
- Tipificar
- Usar tabla normal
- Interpretar el resultado

$\hat{p}$ proporción muestral

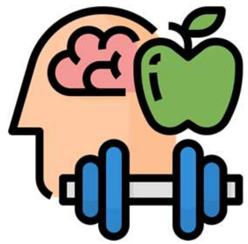
- Tipificamos $$Z = \frac{\hat{p} - 0{,}03}{0{,}076} \sim \mathrm{N}(0{,}1)$$

- $P(\hat{p} > 0{,}04) = P(Z > \frac{0{,}04-0{,}03}{0{,}076}) = P(Z > 1{,}32) = 1 - P(Z < 1{,}32)$

- Usando la tabla normal: $P(\hat{p} > 0{,}04) = 1 - \ 0{,}9066 = 0{,}0934$
$= 9{,}34\%$

la probabilidad de que encuentre más del 4% de pasteles defectuosos es baja, 0,0934

Ejercicio 3

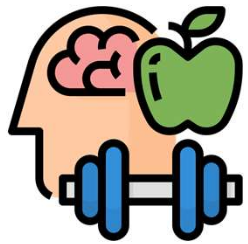


En una ciudad, el 40% de los jóvenes usa transporte público. Se toma una muestra de 100 jóvenes.

a) ¿Cuál es la distribución de la proporción muestral?

b) ¿Qué probabilidad hay de que más del 45% lo use?

Ejercicio 3



Claves para hacer el ejercicio:

- Identificar tipo de ejercicio: **proporción muestral $\hat{p}$**
- Aplicar la fórmula correspondiente:

$$\hat{p} \sim N\left(p; \sqrt{\frac{p \cdot q}{n}}\right) = N\left(0{,}4; \sqrt{\frac{0{,}4 \cdot 0{,}6}{100}}\right) = N(0{,}4; 0{,}049)$$

- ¿Qué nos piden? La distribución de $\hat{p}$ y $P(\hat{p} > 0{,}45)$
- Tipificar
- Usar tabla normal
- Interpretar el resultado

Ejercicio 3

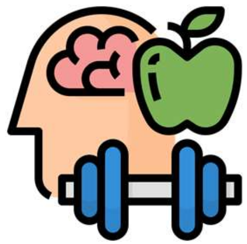
$$Z = \frac{\hat{p} - 0{,}4}{0{,}49} \sim \mathrm{N}(0; 1)$$

$$P(\hat{p} > 0{,}45) = P\left(Z > \frac{0{,}45 - 0{,}4}{0{,}049}\right) = P(Z > 1{,}0204) =$$

$$= 1 - P(Z < 1{,}02) = 1 - 0{,}8462 = 0{,}1538 = 15{,}38\%$$

Interpretación:
La probabilidad de que más del 45% use el transporte público, es 0,1538, lo que indica que es un suceso poco frecuente.
Sólo el 15% de las muestras de 100 jóvenes superan una proporción del 45%

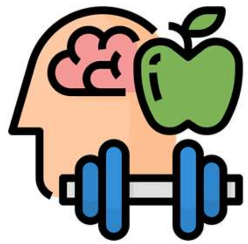
Ejercicio 4



Ordinaria 2025 (ejercicio 4.2.2)

Si se sabe que 8 de cada 10 individuos están satisfechos con el servicio de su compañía eléctrica y se toma una muestra de 100 individuos, ¿cuál es la probabilidad de que la proporción de individuos satisfechos con el servicio de su compañía eléctrica sea superior al 87%?

Ejercicio 4



- Datos: 8 de cada 10 = 0,8 = p, n=100
- tipo de ejercicio: **proporción muestral $\widehat{\boldsymbol{p}}$**, probabilidad
- fórmula correspondiente:

$$\hat{p} \sim N\left(p; \sqrt{\frac{p \cdot q}{n}}\right) = N\left(0{,}8; \sqrt{\frac{0{,}8 \cdot 0{,}2}{100}}\right) = N(0{,}8; 0{,}04)$$

- ¿Qué nos piden? $P(\hat{p} > 0{,}87)$
- Tipificar y usar tabla normal:

$$P(\hat{p} > 0{,}87) = P\left(Z > \frac{0{,}87 - 0{,}8}{0{,}04}\right) = P(Z > 1{,}75) = 1 - P(Z < 1{,}75)$$

$$= 1 - 0{,}9599 = 0{,}0401$$

Ejercicio 4

La probabilidad de que más del 87% estén satisfechos con su compañía eléctrica, es 0,0401, lo que indica que es un suceso muy poco frecuente.

Ejercicio 5

El 30% de los clientes compra online. Se eligen 200 al azar. Probabilidad de que la proporción esté entre 0,25 y 0,35

Recuerda

1. Identifica el tipo de problema
2. Elige la fórmula correcta
3. Tipifica (si hay normal)
4. Calcula
5. Interpreta SIEMPRE