INFERENCIA ESTADÍSTICA: Intervalo de confianza para la media poblacional.

Teorema Central del Límite

Dada una población de media μ y desviación típica σ, no necesariamente normal, la distribución de las medias de las muestras de tamaño *n*:

- Tiene la misma media, μ, que la población.
- Su desviación típica es $\sigma / \sqrt{n}$ y, por consiguiente, disminuye al aumentar *n*.
- Cuando $n \geq 30$ es prácticamente normal.

Es decir:

$$X \rightarrow \mu,\ \sigma \quad \text{, entonces} \quad \overline{X} \rightarrow N\left(\mu,\ \sigma/\sqrt{n}\right) \quad (\text{si } n \geq 30)$$

Intervalo de confianza para la media

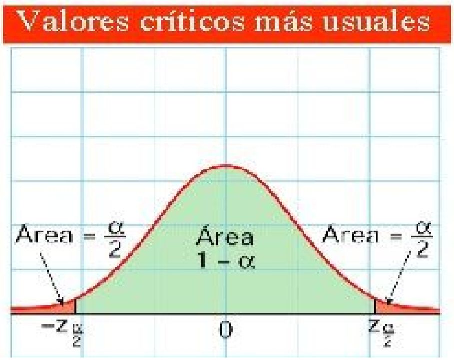
Se desea estimar la media μ, de una población cuya desviación típica, σ es conocida.

Para ello tomamos una muestra de tamaño *n* en la cual se obtiene una media muestral, $\overline{x}$.

Si $n \geq 30$, por el Teorema Central del Límite, el intervalo de confianza para μ con **un nivel de confianza de (1-α).100%** es:

$$\left(\overline{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},\ \ \overline{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

- $z_{1-\alpha/2}$ es el **valor crítico**: valor de la abscisa de una distribución N(0, 1) que deja a su izquierda una probabilidad de $1 - \alpha/2$.
- El valor $E = z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ se llama **error máximo admisible**.



$$\Pr\left(-z_{\alpha/2} < Z \leq z_{\alpha/2}\right) = 1 - \alpha$$

$1-\alpha$	0,8	0,9	0,95	0,99
α	0,2	0,1	0,05	0,01
$\alpha/2$	0,1	0,05	0,025	0,005
$z_{\alpha/2}$	1,28	1,64	1,96	2,58

TEOREMA CENTRAL DEL LIMITE

Se sabe que el cociente intelectual de los alumnos de una universidad se distribuye según la ley normal de media 100 y varianza 729.

a) Halla la probabilidad de que la muestra de 81 alumnos tenga un cociente intelectual
b) medio inferior a 109.

b)Halla la probabilidad de que la muestra de 36 alumnos tenga un cociente intelectual medio superior a 109.

Hemos de tener muy presente que nos piden probabilidades correspondientes a distribuciones de medias muestrales. Por tanto, si una distribución es *N(μ, σ)*, las medias de las muestras de tamaño *n* se distribuyen: $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

a) Las medias de las muestras de tamaño *81* se distribuyen: *N (100, 3)*

En esta distribución, $P(x < 109) = P\left(Z < \frac{109-100}{3}\right) = P(z < 3) = 0{,}9987$

b) Análogamente, las muestras de tamaño *36* se distribuyen: *N(100, 4.5)*

Obtenemos, $P(x > 109) = P(z > 2) = 1 - P(z < 2) = 1 - 0{,}9772 = 0{,}0228$

INTERVALO DE CONFIANZA: EJEMPLO BÁSICO.

Estamos interesados en conocer el consumo diario medio de cigarrillos entre los alumnos de Centros de Bachillerato de nuestra localidad. Seleccionada una muestra aleatoria de 100 alumnos se observó que fumaban una media de 8 cigarrillos diarios. Si admitimos que la varianza de dicho consumo es de 16 cigarrillos en el colectivo total, estime dicho consumo medio con un nivel de confianza del 90 %.

Pretendemos estimar una media, suponiendo que se conoce la varianza de la población.
Para el nivel de confianza indicado 90 % se tiene:
$1-\alpha = 0.90 \Rightarrow \alpha = 0.10 \Rightarrow \alpha/2 = 0.05$. Podemos buscar $z_{1-\alpha/2}$ que deja en N(0,1) un área a la izquierda igual a 0.95; dicho valor es z = 1.64 .El intervalo de confianza será:

$$\left(8 - 1.64\left(\frac{\sqrt{16}}{\sqrt{100}}\right), 8 + 1.64\left(\frac{\sqrt{16}}{\sqrt{100}}\right) \right) \Rightarrow (7.34,\ 8.66)$$

Es decir: *Los alumnos de Bachillerato de nuestra localidad fuman entre 7.34 y 8.66 cigarrillos diarios,* ***con un margen de error (α)*** *del 10%.*

INTERVALO DE CONFIANZA: EJEMPLO INVERSO

Estamos interesados en conocer el consumo diario medio de cigarrillos entre los alumnos de Centros de Bachillerato de nuestra localidad . Seleccionada una muestra aleatoria de 100 alumnos se observó que fumaban una media de 8 cigarrillos diarios. Si admitimos que la varianza de dicho consumo es de 16 cigarrillos en el colectivo total, ¿qué margen de error cometeríamos si afirmamos que la media en la población está comprendida entre 7.5 y 8.5 cigarrillos/día?.

El intervalo de confianza es:

$$\left(\bar{x} - z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right),\ \bar{x} + z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right) = (\ 7.5,\ 8.5\ )$$

. Sustituyendo los valores conocidos del enunciado en uno de los valores extremos del intervalo de confianza, obtenemos el valor $z_{1-\alpha/2}$:

$$\bar{x} + z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) = 8.5 \Rightarrow 8 + z_{\alpha/2}\left(\frac{\sqrt{16}}{\sqrt{100}}\right) = 8.5 \Rightarrow 8 + 0.4\, z_{1-\alpha/2} = 8.5 \Rightarrow z_{1-\alpha/2} = 1.25$$

Mediante la tabla de la N(0,1) obtenemos el área que queda a la izquierda de 1.25:

$1-\alpha/2 = 0.8944 \Rightarrow \alpha/2 = 0.1056 \quad \alpha = 0.2112$ (21.12 %)

Es decir: *Trabajamos con un margen de error del 21.12 % si afirmamos que los alumnos de Bachillerato de nuestra localidad fuman entre 7.5 y 8.5 cigarrillos diarios.*

TAMAÑO DE LAS MUESTRAS Y EL ERROR.

Se desea hacer una estimación sobre la edad media de una determinada población. Calcula el tamaño de la muestra necesario para poder realizar dicha estimación con un error menor de medio año a un nivel de confianza del 99,73%. Se conoce de estudios previos que la edad media de dicha población tiene una desviación típica de $\sigma = 3$.

A un nivel de confianza NC = 99,73% le corresponde un valor crítico $z_{1-\alpha/2} = 3$.
Además, $\sigma = 3$ y **el error (distancia desde la media a uno de los extremos del intervalo de confianza)** $E = z \cdot \frac{\sigma}{\sqrt{n}} = 0{,}5$.

Con estos datos llevados a la fórmula , se obtiene $n = \frac{3^2 . 2^2}{0.5^2} = 324$

De 324 personas, al menos, debe estar compuesta la muestra.

EJERCICIOS PROPUESTOS

Ejercicio 1.
Un fabricante de electrodomésticos sabe que la vida media de éstos sigue una distribución normal con media μ = 100 meses y desviación típica σ = 12 meses. Determina el mínimo tamaño muestral que garantiza, con una probabilidad de 0,98, que la vida media de los electrodomésticos en dicha muestra se encuentra entre 90 y 110 meses.

Ejercicio 2. Una variable tiene **distribución normal**, con desviación típica **poblacional conocida**, σ = 5. Se toma una muestra de tamaño (**n**) = 10, cuya media ($\boldsymbol{x}$) es 20
a.- Determine los límites del intervalo de confianza a nivel 99%, para la media poblacional
b.- Determine los límites del intervalo de confianza a nivel 95%, para la media poblacional
c.- Calcule el tamaño de muestra necesario para estimar por intervalo la media poblacional a nivel de un 90% de confianza con un error menor que 1

Ejercicio 3.
Se sabe que el peso en kilogramos de los alumnos de bachillerato de Madrid, es una variable aleatoria X que sigue una distribución normal de desviación típica igual a 5 kg.

a) En el caso de considerar muestras de 25 alumnos, ¿qué distribución tiene la variable aleatoria media muestral $\overline{X}$?
b) Si se desea que la media de la muestra no difiera en más de 1 kg de la media de población, con probabilidad 0,95; ¿cuántos alumnos se deberían tomar en la muestra?

Solución:

a) $\overline{X} \sim N\left(\mu, \frac{5}{\sqrt{25}}\right) = N(\mu, 1)$

b) $P\{|\overline{X} - \mu| \leq 1\} = 0{,}95$

Tipificando: $P\left\{|Z| \leq \frac{1}{\sigma/\sqrt{n}}\right\} = 0{,}95$; $P\left\{Z \leq \frac{\sqrt{n}}{\sigma}\right\} = 0{,}975$

Entonces: $1{,}96 = \frac{\sqrt{n}}{5} \Rightarrow n = (5 \cdot 1{,}96)^2 \approx 96$

Ejercicio 4. (Puntuación máxima: 2 puntos)
Estamos realizando una encuesta sobre el nivel de conocimientos generales de los estudiantes de Bachillerato de los diferentes centros de Madrid. Para ello, hemos elegido una muestra aleatoria de 9 de estos estudiantes, a los que le hemos realizado un examen. Las calificaciones obtenidas han sido las siguientes:

7,8 6,5 5,4 7,1 5,0 8,3 5,6 6,6 6,2

Se supone que la variable aleatoria objeto de estudio sigue una distribución normal de desviación típica conocida e igual a 1.
Se pide:

a) Un intervalo de confianza al 98% para la media de las calificaciones en el examen.
b) El tamaño mínimo que debería tener la muestra, en el caso de admitir un error máximo de 0,5 puntos, con un nivel de confianza del 95%