

TEMA 6. ESTIMACIÓN. INTERVALOS DE CONFIANZA.

ÍNDICE

1. MUESTREO ESTADÍSTICO

- 1.1. POBLACIÓN Y MUESTRA
- 1.2. TIPOS DE MUESTREOS ALEATORIOS
- 1.3. TAMAÑO Y REPRESENTATIVIDAD DE UNA MUESTRA
- 1.4. DISTRIBUCIÓN DE LA MEDIA MUESTRAL
- 1.5. DISTRIBUCIÓN DE UNA PROPORCIÓN MUESTRAL

2. INTERVALOS DE CONFIANZA

- 2.1. ESTIMADORES PUNTUALES. PARÁMETROS DE UNA POBLACIÓN Y ESTADÍSTICOS OBTENIDOS A PARTIR DE UNA MUESTRA
- 2.2. INTERVALOS DE CONFIANZA
- 2.3. INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL CON DESVIACIÓN TÍPICA CONOCIDA
- 2.4. RELACIÓN ENTRE NIVEL DE CONFIANZA, ERROR ADMISIBLE Y TAMAÑO DE LA MUESTRA
- 2.5. INTERVALO DE CONFIANZA PARA LA PROPORCIÓN EN MUESTRAS GRANDES
- 2.6. DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA PARA UNA PROPORCIÓN

1. MUESTREO ESTADÍSTICO

Mediante la **inferencia estadística** se intenta conocer algo acerca de las características de la población en su conjunto mediante la generalización de lo obtenido en la muestra. Pero es necesario ser consciente de que, en la mayoría de los casos, la verdadera naturaleza y características exactas de la población van a ser desconocidas, y nunca van a poderse conocer con exactitud. Es por ello por lo que **la inferencia proporciona conclusiones** sin certeza total, sino **en términos de probabilidad o de nivel de confianza**.

Algunas de las características desconocidas de la población pueden ser su **distribución de probabilidad y**, en muchos casos, **el valor de los parámetros que definen dicha distribución**. Así, muchos de los procedimientos básicos de la inferencia estadística clásica están centrados alrededor del valor de dichos parámetros. A continuación desarrollamos la metodología de **estimación de parámetros**.

1.1. POBLACIÓN Y MUESTRA

Población estadística, colectivo o universo es el conjunto de todos los individuos (personas, objetos, animales, etc.) que contienen información sobre el fenómeno que se estudia.

Ejemplos:

Si estudiamos el **precio de la vivienda** en una ciudad, la **población** será el total de las **viviendas** de dicha ciudad.

Se va a realizar un estudio estadístico sobre el **porcentaje de personas casadas** en la península. Para ello no es factible estudiar a todos y cada uno de los habitantes por razones de coste y de rapidez en la obtención de la información. Por lo tanto, es necesario acudir a examinar sólo una parte de esta población. Esa parte es la muestra elegida.

Muestra es un subconjunto representativo que se selecciona de la población y sobre el que se va a realizar el análisis estadístico.

Muestreo es el proceso mediante el cual se selecciona la muestra de la población.

El **tamaño de la muestra** es el número de sus elementos.

Cuando la muestra comprende a todos los elementos de la población, se denomina **censo**.

Ejemplo:

Si se estudia el **precio de la vivienda** de una ciudad, lo normal será no recoger información sobre todas las viviendas de la ciudad (ya que sería una labor muy compleja y costosa), sino que **se suele seleccionar un subgrupo (muestra)** que se entienda que es suficientemente **representativo**.

En **control de calidad**, por ejemplo, si se estudia **la vida de un electrodoméstico**, y para ello deben funcionar hasta que se estropeen, es absurdo estudiar todos los electrodomésticos (población) pues nos quedamos sin fabricación, por lo que es imprescindible seleccionar una muestra que sea representativa de la población.

1.2. TIPOS DE MUESTREOS ALEATORIOS

La forma de seleccionar la muestra, **muestreo**, debe garantizar que sea representativa de la población. Para ello debe ser un muestreo **aleatorio**, es decir, al azar. Todos los individuos de la población deben tener las mismas posibilidades de ser seleccionados para la muestra.

Ejemplos:

1) *Se quiere estudiar el nivel adquisitivo de las personas de una ciudad, para lo que pasamos una encuesta a la puerta del Corte Inglés, ¿te parece un muestreo aleatorio?*

No lo es. Las personas que entran en un determinado establecimiento no representan a toda la población.

2) *Vas a hacer un estudio sobre los gustos musicales de los jóvenes, y para ello, preguntas a cinco de entre tus amistades, ¿te parece un muestreo aleatorio?*

No lo es. Tus amistades pueden tener unos gustos diferentes a los del resto de la población.

Si la muestra está mal elegida, no es representativa, se producen **sesgos**, errores en los resultados del estudio. Hay varios **tipos de muestreos** que garantizan la representatividad:

Muestreo aleatorio simple

Todos los individuos de la población tienen la misma probabilidad de ser elegidos en la muestra.

Muestreo aleatorio sistemático

Se ordenan los individuos de la población. Se elige al azar un individuo, y se selecciona la muestra tomando individuos mediante saltos igualmente espaciados.

Muestreo aleatorio estratificado

Se divide la población en grupos homogéneos de una determinada característica, *estratos*, por ejemplo edad, y se toma una muestra aleatoria simple en cada estrato.

Ejemplo: Se estudia el estado de los huesos de la población de un país, y se divide la población en “niños”, “jóvenes”, “edad media” y “tercera edad”. En cada grupo se hace un muestreo aleatorio simple.

Muestreo por conglomerados o áreas

Se divide la población en conglomerados o áreas, selecciona al azar uno o varios conglomerados y se estudia.

Ejemplo: Se estudia la incidencia de enfermedades cardíacas en la población rural española. Para ello se hace un censo de pueblos y se eligen varios al azar, donde se estudia a la población

Muestreo no aleatorio A veces también se usa.

Ejemplo: La estimación de voto que suele hacerse a pie de urna. Es cómodo, barato pero no es representativo.

1.3. TAMAÑO Y REPRESENTATIVIDAD DE UNA MUESTRA

Cuando se elige una muestra los dos aspectos que hay que tener en cuenta son, el tamaño y la representatividad de la muestra. Si la muestra es demasiado pequeña, aunque esté bien elegida, el resultado no será fiable.

[Tamaño muestral y tipos de muestreo por el IGE \(Instituto Galego de Estatística\)](#)

Ejemplo: Supongamos que de los 400 alumnos de un instituto, queremos escoger una **muestra de 78** para hacer un estudio sobre la edad.

a) Un **muestreo aleatorio simple** se podría hacer utilizando un censo del alumnado (en el que aparece ordenando alfabéticamente) y escoger al azar 78 números del 1 al 400.

b) Un **muestreo aleatorio sistemático** se podría hacer eligiendo al azar un número del 1 al 400 y tomar a los 77 restantes separados por intervalos de $400/78 \approx 5$.

Si, por ejemplo, sale en el sorteo el 57, elegiríamos a los alumnos que tengan los números 57, 62, 67, 72, 77, 82, 87, 92,....., 397, 2, 7, ...

c) Uno **estratificado** sería, por ejemplo, considerar los estratos:

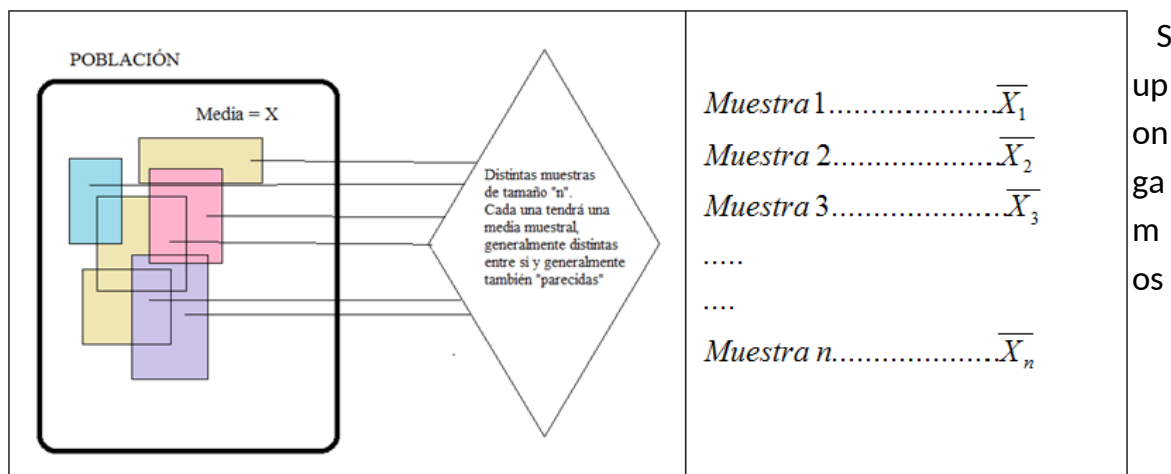
Curso	Nº alumnos	Nº alumnos en muestra (con afijación simple)	Nº alumnos en muestra (con afijación proporcional)
1º ciclo ESO	125	$78/3 = 26$	$\frac{125}{400} \cdot 78 = 24,375 \approx 24$
2º ciclo ESO	158	$78/3 = 26$	$\frac{158}{400} \cdot 78 = 30,81 \approx 31$
Bachillerato	117	$78/3 = 26$	$\frac{117}{400} \cdot 78 = 22,815 \approx 23$

d) Uno por **conglomerados** sería escoger, por ejemplo, los conglomerados de chicas y chicos y escoger uno de ellos al azar, por ejemplo el de las chicas. Después, se toman 30 chicas mediante muestreo aleatorio simple. Es algo similar a lo hecho en el apartado a pero sólo con las chicas.

Practica los tipos de muestreos en los escenarios de [Inferencia Estadística con Descartes](#).

1.4. DISTRIBUCIÓN DE LA MEDIA MUESTRAL

Supongamos que tenemos una **variable aleatoria X definida en una población** de la que se conoce la media μ y la desviación típica σ .



también que extraemos muestras de tamaño “n” de dicha población. Cada muestra proporcionará una determinada media (media muestral):

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

Si consideramos cada una de estas medias como valores de una variable aleatoria podemos estudiar su distribución, a lo que llamaremos **distribución muestral de medias** o distribución en el muestreo de las medias muestrales.

Las características principales de esta variable aleatoria son:

- La **media** es la misma que la de la población: μ
- La **desviación típica** es la misma que la de la población dividida entre $\sqrt{n}$: $\frac{\sigma}{\sqrt{n}}$
- Si la variable aleatoria X sigue una distribución $N(\mu, \sigma)$ entonces la distribución de la media muestral de tamaño n será: $N(\mu, \frac{\sigma}{\sqrt{n}})$
- En caso de que la variable aleatoria X no siga una distribución normal podemos aplicar el teorema central del límite:

Teorema Central del Límite:

Si X es una variable aleatoria de una población de **media μ finita y desviación típica σ finita**. Entonces, la distribución de la media muestral de tamaño n tiene de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$ y se aproxima a una distribución normal a medida que crece el tamaño de la muestra.

Es decir, para n suficientemente grande ($n \geq 30$) $\bar{X} \rightarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

Ejemplo: El peso de los libros de texto en el instituto se distribuye normalmente con un peso medio de $\mu = 400$ g y una desviación típica $= 50$ g. Si tomamos una muestra aleatoria de tamaño $n = 16$, halla la probabilidad de que el peso medio esté entre 375 y 425 g.

Solución: $X =$ peso de un libro de texto del instituto $X \sim N(400; 50)$

Al ser X una distribución normal, la media muestral $\bar{X}$ seguirá una normal sin importar el tamaño de la muestra:

$$n = 16 \Rightarrow \bar{X} \sim N\left(400; \frac{50}{\sqrt{16}}\right) = N(400; 12,5)$$

$$P(375 \leq \bar{X} \leq 425) \stackrel{\text{tipificamos}}{=} P\left(\frac{375 - 400}{12,5} \leq Z \leq \frac{425 - 400}{12,5}\right) = P(-2 \leq Z \leq 2) =$$

$$= 2P(Z \leq 2) - 1 = 2 \cdot 0,9772 - 1 = 0,9544$$

Ejercicios:

1. Las estaturas de 1200 estudiantes de un centro de enseñanza se distribuyen normalmente con una media de 1,72 m y desviación típica 0,09. Si se toma al azar una muestra de 36 estudiantes. Halla:

- La probabilidad de que la media sea inferior a 1,75 m
- La probabilidad de que la media esté entre 1,68 m y 1,73 m.

2. El gasto medio bimensual en electricidad por familia en España se distribuye según una ley normal de media 142,32 € y desviación típica de 8,5 €.

- Halla la probabilidad de que una muestra de 25 familias, elegidas al azar, tenga un gasto medio en electricidad superior a 144,6 €.
- Halla la probabilidad de que una muestra de 100 familias, elegidas al azar, tenga un gasto medio en electricidad superior a 144,6 €.

3. Los pesos (en gramos) de los tornillos que fabrica una máquina se distribuyen según una variable $N(20; 0,5)$. Se toman muestras de 25 tornillos. Calcula la probabilidad de que una muestra elegida al azar tenga un peso medio mayor de 19 gramos.

Soluciones:

1. a) 0,9772 b) 0,7448 2. a) 0,0901 b) 0,0037 3. 1

1.5. DISTRIBUCIÓN DE UNA PROPORCIÓN MUESTRAL

Si en una población conocemos la proporción p de los individuos que tienen cierta característica, podemos elegir aleatoriamente muestras de tamaño n y obtener la proporción en cada muestra

$$\hat{p} = \frac{n^{\circ} \text{ de éxitos}}{n}$$

El n° de éxitos de una muestra de tamaño n se distribuye de forma binomial $B(n;p)$, luego para muestras con n suficientemente grande, $n \geq 30$, siendo además $np \geq 5$ y $nq \geq 5$ se puede aproximar por la normal $N(np, \sqrt{npq})$.

Al dividir entre n obtenemos la distribución muestral de proporciones de la variable $\hat{p}$:

$$\hat{p} \rightarrow N\left(p, \sqrt{\frac{p \cdot q}{n}}\right)$$

- La **media o esperanza** de $\hat{p}$ coincide con la media poblacional p .

- La **desviación típica** de $\hat{p}$ coincide con $\sqrt{\frac{p \cdot q}{n}}$.

Ejemplo:

Una máquina fabrica piezas de precisión. En su producción habitual, fabrica un 3% de piezas defectuosas. Un cliente recibe una caja de 500 piezas procedentes de la fábrica. Calculemos la probabilidad de que:

a) Haya más de un 5% de piezas defectuosas en la caja.

De acuerdo con lo visto anteriormente, la distribución muestral de proporciones

sigue una distribución normal $\hat{p} \rightarrow N\left(0,03; \sqrt{\frac{0,03 \cdot 0,97}{500}}\right) = N(0,03; 0,0076)$.

Así pues, la probabilidad pedida será: $P(\hat{p} > 0,05) = P\left(\frac{\hat{p} - 0,03}{0,0076} > \frac{0,05 - 0,03}{0,0076}\right) =$
 $= P(Z > 2,63) = 1 - P(Z < 2,63) = 1 - 0,9957 = 0,0043$

b) Haya menos de 10 piezas defectuosas en la caja.

Como 10 piezas constituyen el 2% de las piezas ($10/500 = 0,02$), la probabilidad pedida es:

$$P(\hat{p} < 0,02) = P\left(\frac{\hat{p} - 0,03}{0,0076} < \frac{0,02 - 0,03}{0,0076}\right) = P(Z < -1,32) = 1 - P(Z < 1,32) = \\ = 1 - 0,9066 = 0,0934$$

Ejercicios:

1. De una población de 120 alumnos/as, hay 48 que tienen 2 o más hermanos. Si de dicha población se toma una muestra al azar de tamaño 40, ¿cuál es la probabilidad de que haya más de un 55% de alumnos con 2 o más hermanos?
2. En un saco mezclamos judías blancas y pintas en la relación de 14 blancas por cada pinta. Extraemos al azar un puñado de 100 judías, ¿cuál es la probabilidad de que haya entre 5 y 10 judías pintas en el puñado?

Soluciones:

1. 0,0262 2. 0,6585

2. INTERVALOS DE CONFIANZA

2.1. ESTIMADORES PUNTUALES. PARÁMETROS DE UNA POBLACIÓN Y ESTADÍSTICOS OBTENIDOS A PARTIR DE UNA MUESTRA

Es fundamental entender la diferencia entre parámetros y estadísticos. Los parámetros se refieren a la distribución de la población y los estadísticos a los datos de las muestras.

Ejemplo: Decimos que la duración media de las bombillas que fabrica una empresa es de 1600 horas, nos referimos a la población y la media la designamos por $\mu=1600$. Sin embargo si hallamos la duración media de una muestra de 20 bombillas y obtenemos 1580 horas, nos referimos a la muestra y la media la designamos por $\bar{X}=1580$.

- La **media μ** o la **proporción p** de la población es un **parámetro** y es constante.
- La **media $\bar{x}$** o la **proporción $\hat{p}$** de la muestra es un **estadístico** y es una **variable aleatoria**.

2.2. INTERVALOS DE CONFIANZA

Ahora queremos, a partir de una muestra de tamaño n , estimar el valor de un parámetro de la población dando un intervalo en el que confiamos que esté dicho parámetro.

Llamamos intervalo de confianza para un parámetro poblacional al intervalo que contiene a dicho parámetro con una probabilidad, también llamada nivel de confianza $N_c = 1 - \alpha$. Al valor de α se le llama nivel de significación.

•Cálculo de un intervalo centrado en la $N(0, 1)$ con nivel de confianza $1 - \alpha$

Antes de comenzar la explicación de intervalos de confianza vamos a aprender a determinar intervalos centrados en 0 $[-z_{1-\alpha/2}, z_{1-\alpha/2}]$ en la $N(0,1)$ que cumplan:

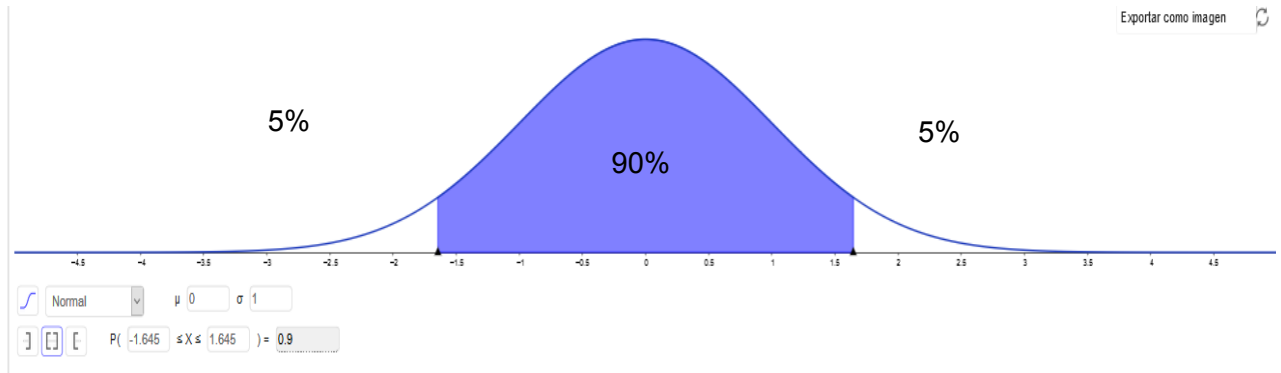
$$P\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Llamaremos valor crítico al valor $z_{1-\alpha/2}$ asociado al nivel de confianza $1 - \alpha$ que verifica

$$P\left(Z \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}$$

Ejemplo: Los valores más comunes del nivel de confianza son 90%, 95% y 99%:

$1 - \alpha = 0,9$ Buscamos un intervalo tal que $P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 0,9$ intervalo centrado en 0 que deja un 5% de probabilidad a izquierda y un 5% a derecha.



$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 0,9 \quad \rightarrow \quad 2P(Z < z_{1-\alpha/2}) - 1 = 0,9 \quad \rightarrow \quad P(Z < z_{1-\alpha/2}) = 0,95$$

$$\rightarrow \quad z_{1-\alpha/2} = 1,645$$

El intervalo $(-1,645, 1,645)$ encierra una probabilidad del 90%

$1 - \alpha = 0,95$ $\rightarrow \quad P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 0,95 \quad \rightarrow \quad 2P(Z < z_{1-\alpha/2}) - 1 = 0,95$

$$\rightarrow \quad P(Z < z_{1-\alpha/2}) = 0,975 \quad \rightarrow \quad z_{1-\alpha/2} = 1,96$$

El intervalo $(-1,96, 1,96)$ encierra una probabilidad del 95%

$1 - \alpha = 0,99$ $\rightarrow \quad P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = 0,99 \quad \rightarrow \quad 2P(Z < z_{1-\alpha/2}) - 1 = 0,99$

$$\rightarrow \quad P(Z < z_{1-\alpha/2}) = 0,995 \quad \rightarrow \quad z_{1-\alpha/2} = 2,575$$

El intervalo $(-2,575, 2,575)$ encierra una probabilidad del 99%

2.3. INTERVALO DE CONFIANZA PARA LA MEDIA POBLACIONAL CON DESVIACIÓN TÍPICA CONOCIDA

Bajo las condiciones vistas en el punto anterior la distribución de medias muestrales sigue una distribución normal:

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \text{tipificando} \quad Z = \frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Por tanto, podemos calcular un intervalo de confianza $1 - \alpha$:

$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = P(-z_{1-\alpha/2} < \frac{(\bar{X} - \mu)}{\frac{\sigma}{\sqrt{n}}} < z_{1-\alpha/2}) = 1 - \alpha \quad \Leftrightarrow$$

$P(-z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha \Leftrightarrow$ por simetría es equivalente: $P(-z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu - \bar{X} < z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$

$$P(\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Así pues, el intervalo de confianza para la media poblacional μ con un nivel de confianza $N_c=1-\alpha$, viene dado por:

$$I_c = (\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}})$$

Ejemplos:

1. Una muestra aleatoria de 100 alumnos que se presenta a las PAU revela que la media de edad es de 18,1 años. Halla un intervalo de confianza del 90% para la media de edad de todos los estudiantes, sabiendo que la desviación típica de la población es de 0,4.

DATOS POBLACIÓN: μ desconocida, $\sigma = 0,4$

DATOS MUESTRA: $n = 100$ $\bar{X} = 18,1$

NIVEL DE CONFIANZA: $1 - \alpha = 0,9$

Para este nivel de confianza, $z_{1-\alpha/2}$ es el valor que cumple $P(Z < z_{1-\alpha/2}) = 0,95$. Mirando la tabla de la normal: $z_{1-\alpha/2} = 1,645$.

Así pues el intervalo de confianza es:

$$\left(18,1 - 1,645 \cdot \frac{0,4}{\sqrt{100}}; 18,1 + 1,645 \cdot \frac{0,4}{\sqrt{100}} \right) = (18,0342; 18,1658)$$

Cuyo significado es que la media de la edad de todo el alumnado de la población se encuentra entre 18,0342 y 18,1658 con una probabilidad de 0,9.

2. Se ha extraído una muestra de 145 alumnos de una escuela de artes, a los que se les ha propuesto un test de habilidad. La media y la desviación típica obtenida de la muestra son 82 y 14 respectivamente. A partir de estos datos, calcula el intervalo de confianza en el cual se hallará la media de la población al nivel de confianza del 95%.

* (ESTE PROBLEMA NO ENTRARÁ EN EXAMEN, SÓLO EN PRÁCTICA DE CLASE)

DATOS POBLACIÓN: $\mu, \sigma \rightarrow$ desconocidos

DATOS MUESTRA: $n = 145$ $\bar{X} = 82$ $S = 14$ (desviación típica muestral)

* Aproximamos la desviación típica σ por la cuasidesviación típica muestral,

$$s = S \cdot \sqrt{\frac{n}{n-1}} = 14 \cdot \sqrt{\frac{145}{144}} = 14,0485 : \quad \bar{X} \rightarrow N\left(82, \frac{14,0485}{\sqrt{145}}\right) = N(82; 1,1667)$$

Calculamos el valor crítico $z_{1-\alpha/2}$ asociado al nivel de confianza del 95%:

$$P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 2P(Z \leq z_{1-\alpha/2}) - 1 = 0,95 \quad P(Z \leq z_{1-\alpha/2}) = 0,975 \quad z_{1-\alpha/2} = 1,96$$

Calculamos el intervalo de confianza:
$$z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 1,96 \cdot \frac{14,0485}{\sqrt{145}} = 2,2867$$

$$\rightarrow I_c = (82 - 2,2867, 82 + 2,2867) = (79,7133; 84,2867)$$

La media de la población, μ , se encontrará en el intervalo (79,7133; 84,2867) con una probabilidad del 95%.

2.4. RELACIÓN ENTRE NIVEL DE CONFIANZA, ERROR ADMISIBLE Y TAMAÑO DE LA MUESTRA

Error máximo admisible: es la diferencia máxima entre la media muestral y la poblacional

$$E = z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Amplitud del intervalo: es el doble del error máximo admisible

$$A = 2 \cdot z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Tamaño muestral para un error dado: de la expresión del error, podemos

despejar el tamaño muestral

$$n = \left(\frac{z_{1-\frac{\alpha}{2}} \cdot \sigma}{E} \right)^2$$

Comprueba como influyen en el cálculo del intervalo de confianza la desviación típica, el tamaño de la muestra y el nivel de confianza en:

<https://www.geogebra.org/m/kqFc2593>

Observa que el tamaño de la muestra debe ser más grande cuanto menor sea el error máximo admisible:

Para estimaciones más precisas se debe aumentar el tamaño de la muestra.

Al aumentar el nivel de confianza $1 - \alpha$ aumenta el tamaño de la muestra, luego:

Para aumentar el nivel de confianza se debe aumentar el tamaño de la muestra.

Ejemplos:

1. Un psicólogo quiere medir el tiempo de reacción de sus pacientes y para ello toma una muestra de 175 pacientes y realiza una estimación con un nivel de confianza del 99%. Sabiendo que la desviación típica es de 0,05 segundos, ¿qué error máximo ha cometido?

DATOS POBLACIÓN: $\mu \rightarrow$ desconocida, $\sigma = 0,05$

DATOS MUESTRA: $n = 175$ $\bar{X} = ?$

Distribución de la media muestral (no es imprescindible)

$$\bar{X} \rightarrow N\left(\mu, \frac{0,05}{\sqrt{175}}\right) = N(\mu, 0,0038)$$

Calculamos el valor crítico $z_{1-\alpha/2}$ asociado al nivel de confianza del 99%:

$$P(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 2P(Z \leq z_{1-\alpha/2}) - 1 = 0,99 \quad P(Z \leq z_{1-\alpha/2}) = 0,995 \quad z_{1-\alpha/2} = 2,575$$

Calculamos el error máximo:

$$E = z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$$E = z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 2,575 \cdot \frac{0,05}{\sqrt{175}} = 0,0097 \quad \text{Error} = 0,97\%$$

2. Se sabe que la dedicación media de los jóvenes al ocio sigue una distribución normal de desviación típica 63 minutos. Halla la media muestral y el tamaño mínimo de la muestra de jóvenes que garantiza con una probabilidad de 0,95 que el tiempo medio de ocio está entre 382 y 418 minutos.

DATOS POBLACIÓN: $\mu \rightarrow$ desconocida, $\sigma = 63$

DATOS MUESTRA: $n = ?$ $\bar{X} = ?$

NIVEL DE CONFIANZA: 95% $\rightarrow z_{1-\alpha/2} = 1,96$

INTERVALO DE CONFIANZA: (382, 418)

Como la media muestral es el punto medio del intervalo de confianza, tenemos que

$$\bar{X} = \frac{382+418}{2} = 400$$

Por otra parte, el error máximo admisible será: $E = 418 - 400 = 18$

Así pues, de la expresión del error máximo admisible, obtenemos la igualdad:

$$E = 1,96 \cdot \frac{63}{\sqrt{n}} = 18$$

Sin más que despejar n , obtenemos: $n = \left(\frac{1,96 \cdot 63}{18}\right)^2 = 6,86^2 = 47,0596$

Así pues, el tamaño muestral mínimo es de 48 jóvenes.

3. El otorrino conoce que la desviación típica del tiempo de respuesta a un sonido es de un segundo. Desea estudiar dicho tiempo de respuesta con un error máximo admisible de 0,1 haciendo un estudio con 100 pacientes: Determina con qué nivel de confianza obtendrá el intervalo de confianza.

$$E = z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow 0,1 = z_{1-\alpha/2} \cdot \frac{1}{\sqrt{100}} \rightarrow z_{1-\alpha/2} = 1 \rightarrow 1 - \alpha/2 = 0,8413$$

$$\alpha/2 = 1 - 0,8413 \quad \rightarrow \quad \alpha = 0,3174 \quad \rightarrow \quad 1 - \alpha = 0,6826$$

Es decir, el nivel de confianza es del 68,26%.

Otro método para razonar el nivel de confianza:

$$1 - \alpha = 2P(Z \leq z_{1-\alpha/2}) - 1 = 2 \cdot 0,8413 - 1 = 0,6826$$

Ejercicios: INTERVALOS DE CONFIANZA PARA LA MEDIA MUESTRAL

1. El tiempo de renovación de un teléfono móvil, expresado en años, se puede aproximar mediante una distribución normal con desviación típica 0,4 años. Se toma una muestra aleatoria simple de 100 usuarios y se obtiene una media muestral igual a 1,5 años. Determínese un intervalo de confianza al 95 % para el tiempo medio de renovación de un teléfono móvil.

2. Calcula para los datos del ejercicio anterior un intervalo de confianza del 99% y compáralo.

3. Se ha tomado una muestra aleatoria simple de 16 pacientes y se ha anotado el número de días que han recibido tratamiento para los trastornos del sueño que sufren. Los resultados han sido:

280; 285; 295; 330; 290; 350; 360; 320; 295; 310; 300; 305; 295; 280; 315; 305

Se sabe que la duración, en días, del tratamiento se puede aproximar por una variable aleatoria con distribución normal de media μ desconocida y desviación típica 34,5 días. Determina un intervalo de confianza con un nivel del 99% para la media poblacional

4. ¿Cuál es el número mínimo de estudiantes que debemos elegir de una población de $\sigma = 2$, para una muestra aleatoria simple si el error mínimo admisible es de 0,1, y el nivel de confianza del 95 %?

5. En una población de estudiantes de desviación típica $\sigma = 2$, se quiere pasar una prueba a 100 estudiantes para determinar sus conocimientos de Matemáticas con un error mínimo del 0,5. ¿Cuál es el nivel de confianza obtenido?

6. ¿Qué tamaño mínimo debe tener una muestra para que el error máximo cometido en la estimación de la media sea menor de 0,1 unidades, con un nivel de confianza del 95 %, sabiendo que la desviación típica poblacional es conocida y vale 4?

7. Determina el tamaño muestral mínimo necesario para que el valor absoluto de la diferencia entre la media muestral y la media poblacional sea menor o igual a 0,02 años con un nivel de confianza del 90% sabiendo que la población se distribuye según una normal de desviación típica 0,4.

8. En el estudio anterior se toma una muestra de 2401 individuos. Queremos que el error máximo admisible sea de 0,02. ¿Cuál será el nivel de confianza?

9. Las ventas mensuales en una tienda de electrodomésticos se distribuyen según una ley normal con desviación típica de 540 €. Se ha realizado un estudio en los últimos nueve meses, y se ha hallado el intervalo de confianza (2802; 3508)

a) ¿Cuál ha sido la media de las ventas en esos nueve meses?

b) ¿Cuál es el nivel de confianza para este intervalo?

10. Un fabricante de bombillas sabe que la desviación típica de la duración de las bombillas es de 100. Calcula el tamaño de la muestra que se ha de someter a prueba para tener una confianza del 95% de que el error de la duración media que se calcule sea menor de 10 h.

11. La longitud de los cables de los auriculares que fabrica una empresa es una variable aleatoria que sigue una ley normal con desviación típica 4,5 cm. Para estimar la longitud media se han medido los cables de una muestra aleatoria de 9 auriculares y se han obtenido las siguientes longitudes, en cm:

205, 198, 202, 204, 197, 195, 196, 201, 202

Halla un intervalo de confianza, al 97%, para la longitud media de los cables.

12. La vida media de un determinado modelo de bombilla sigue una distribución normal con desviación típica igual a 60 días. Elegida una muestra y con un nivel de confianza del 98%, se obtiene el intervalo (388,68; 407,32) para la vida media. Calcula la media y el tamaño de la muestra elegida.

13. Se sabe que el peso de los recién nacidos en una determinada población sigue una distribución normal de 3 600 g de media y 280 g de desviación típica. Se toma una muestra al azar de 196 de estos recién nacidos, y se calcula la media. ¿Cuál es la probabilidad de que esta media esté entre 3 580 y 3 620 g?

Soluciones:

1. (1,4216; 1,5784)

2. (1,397; 1,603)

3. (284,9781; 329,3969)

4. 1537 estudiantes

5. 98,76%

6. 6147

7. 1083

8. 98,58%

9. a) 3155€

b) 95%

10. 385 bombillas

11. (196,74; 203,26)

12. $\bar{X}$ = 398 días y n = 225 bombillas

13. 0,6826

2.5. INTERVALO DE CONFIANZA PARA LA PROPORCIÓN EN MUESTRAS GRANDES

Vamos a ver ahora cómo se construye el intervalo de confianza para la proporción. Hemos visto durante la unidad que, para poblaciones grandes ($n \geq 30$), la distribución muestral de proporciones sigue una normal $\hat{p} \rightarrow N\left(p, \sqrt{\frac{p \cdot q}{n}}\right)$

Como no disponemos de p , se estima puntualmente por la proporción muestral $\hat{p}$, tipificando $Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}} \rightarrow N(0,1)$.

Por tanto, podemos calcular un intervalo de confianza $1 - \alpha$:

$$P(-z_{1-\alpha/2} < Z < z_{1-\alpha/2}) = P\left(-z_{1-\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}} < z_{1-\alpha/2}\right) = 1 - \alpha \quad \Leftrightarrow$$

$$P\left(-z_{1-\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} < \hat{p} - p < z_{1-\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}\right) = 1 - \alpha \quad \Leftrightarrow$$

por simetría es equivalente:

$$P\left(-z_{1-\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} < p - \hat{p} < z_{1-\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}\right) = 1 - \alpha$$

$$P\left(\hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} < p < \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}\right) = 1 - \alpha$$

Así pues, el intervalo de confianza para la proporción p con un nivel de confianza

$N_c = 1 - \alpha$, viene dado por:

$$I_c = \left(\hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}} \right)$$

Ejemplo:

Se selecciona una muestra de 500 alumnos de ESO y se les pregunta si tienen smartphone con conexión de datos, contestando afirmativamente 225. Veamos cuál es el intervalo de confianza para la proporción de alumnos que tienen smartphone con conexión de datos con un nivel de confianza del 95%

DATOS MUESTRA: $n = 500$ $\hat{p} = \frac{225}{500} = 0,45$

Para un nivel de confianza $1 - \alpha = 0,95$, $z_{1-\alpha/2}$ es el valor que cumple $P(Z < z_{1-\alpha/2}) = 0,975$. Mirando la tabla de la normal: $z_{1-\alpha/2} = 1,96$.

Así pues el intervalo de confianza es: $\left(0,45 - 1,96 \cdot \sqrt{\frac{0,45 \cdot 0,55}{500}}; 0,45 + 1,96 \cdot \sqrt{\frac{0,45 \cdot 0,55}{500}} \right)$
 $= (0,45 - 0,0436; 0,45 + 0,0436) = (0,4064; 0,4936)$

2.6. DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA PARA UNA PROPORCIÓN

Error máximo admisible: es la diferencia máxima entre la proporción muestral y la poblacional

$$E = z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

Amplitud del intervalo: es el doble del error máximo admisible

$$A = 2 \cdot z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot \hat{q}}{n}}$$

Tamaño muestral para un error dado: de la expresión del error, podemos

despejar el tamaño muestral

$$n = \frac{z_{1-\alpha/2}^2 \cdot \hat{p} \cdot \hat{q}}{E^2}$$

ATENCIÓN:

En caso de que **no se conozca ningún estimador puntual previo de p** (caso que suele suceder casi siempre que se realizan encuestas de opinión), para calcular el tamaño de la muestra se utiliza el caso de máxima indeterminación (o caso más desfavorable) **p = 1 - p = 0,5**.

Ejemplo:

Deseamos conocer el número de niños que hay que incluir, como mínimo, en una encuesta sobre la adecuada alimentación en el desayuno, sabiendo que, con un nivel de confianza del 99,73%, ha resultado el intervalo de confianza para la proporción de niños que lleva a cabo una adecuada alimentación en el desayuno el siguiente: (0,42; 0,50)

Como la proporción muestral es el punto medio del intervalo de confianza, tenemos

$$\text{que: } \hat{p} = \frac{0,42 + 0,50}{2} = 0,46$$

Calculamos $z_{1-\alpha/2}$, el valor que cumple $P(Z < z_{1-\alpha/2}) = 0,9973 + 0,0027/2 = 0,9987$.

Mirando la tabla de la normal: $z_{1-\alpha/2} = 3$.

Por otra parte, el error máximo admisible será: $E = 0,50 - 0,46 = 0,04$

$$\text{Calculamos el tamaño muestral mínimo: } n = \frac{3^2 \cdot 0,46 \cdot 0,54}{0,04^2} = 1397,25$$

Así pues, el tamaño muestral mínimo es de 1398 niños.

1. Para estimar la proporción de estudiantes de una universidad a favor de la reinserción social del delincuente, se entrevistó aleatoriamente a 500 estudiantes. El 58% estaba a favor. Halla el intervalo de confianza, a un nivel del 95%, en el cual se hallará la población universitaria que se encuentra a favor.
2. Un psicólogo quiere medir el tiempo de reacción de sus pacientes y para ello toma una muestra de 175 pacientes y realiza la estimación con un nivel de confianza del 99%, obteniendo una proporción del 72%. ¿Qué error máximo ha cometido?
3. Tomada una muestra de 300 personas mayores de 15 años en una gran ciudad, se encontró que 104 de ellas leían algún periódico regularmente. Halla, con un nivel de confianza del 90%, un intervalo para estimar la proporción de lectores de periódicos entre los mayores de 15 años.
4. Se quiere estimar la proporción de hembras entre los peces de una piscifactoría; para ello se ha tomado una muestra aleatoria de 500 peces, y en ella hay 175 hembras.
 - a) Calcula un intervalo de confianza para la proporción de hembras en esta población de peces, con un nivel de confianza del 94%.
 - b) A la vista del resultado del muestreo se quiere repetir la experiencia para conseguir un intervalo de confianza con el mismo nivel y un error máximo de 0.02, ¿cuál es el tamaño mínimo que debe tener la nueva muestra?
5. Una empresa consultora quiere estudiar algunos aspectos de la vida laboral de los trabajadores de una ciudad. Para ello selecciona una muestra aleatoria de 500 trabajadores, de los que 118 afirman residir en otra ciudad. Con un nivel de confianza del 93%,
 - a) Calcula un intervalo de confianza para la proporción de trabajadores que residen fuera.
 - b) Calcula el error cometido en el intervalo anterior.
6. Como resultado de una encuesta en la que se utilizó el supuesto de máxima indeterminación ($p=1-p=1/2$) se afirma que, con un 97,56% de confianza, el porcentaje de individuos de una población que considera las redes sociales como causa principal de conflictos en la adolescencia, está entre el 57,5% y el 62,5%.
 - (a) Calcula el número de individuos de esa población a los que se les realizó la encuesta.
 - (b) De los que se les realizó la encuesta, ¿cuántos contestaron que la causa principal de los conflictos son las redes sociales?

SOLUCIONES:

1. (0.5408 , 0.6192) 2. 0.0876 3. (0.302 , 0.392)
4. a) (0.3099,0.3901) b) $n = 2011$ 5. a) (0.202,0.270) b) $E = 0,034$
6. a) 2025 individuos b) 1215 individuos