TEMA 8

Muestreo e Inferencia Estadística

Contenidos

1. Población y muestras.
2. Tipos de muestreo.
3. Parámetros de una muestra y parámetros de una población.
4. Distribución de las medias muestrales.
5. Distribución de las proporciones muestrales.
6. Intervalos característicos en las distribuciones normales.
7. Intervalos de confianza para la media.
8. Intervalos de confianza para una proporción.

Tiempo estimado

16 sesiones

Criterios de Evaluación

1. Conocer el vocabulario básico del muestreo.
2. Conocer la media y desviación típica de la distribución de la media muestral en poblaciones finitas o muestreo con reemplazo
3. Conocer la formulación básica del Teorema Central del Límite y su importancia, así como saber aplicarlo en poblaciones con media y desviación típica conocidas.
4. Conocer los elementos de un intervalo de confianza y el significado de dicho intervalo.
5. Saber determinar: un intervalo de confianza para la media, el nivel de confianza con el que se ha construido tal intervalo y el tamaño muestral mínimo necesario para acotar el error cometido.
6. Saber determinar: un intervalo de confianza para una proporción, el nivel de confianza con el que se ha construido tal intervalo y el tamaño muestral mínimo necesario para acotar el error cometido.

1.Población y Muestras

Ya manejamos estos dos conceptos en el curso pasado, cuando estudiamos Estadística. En este curso vamos a profundizar en su estudio.

Supongamos que deseamos conocer una característica en un un conjunto de personas, seres u objetos. Por ejemplo:

- En una fábrica de automóviles tenemos que inspeccionar las emisiones de CO/CO_2 de un determinado modelo.
- Necesitamos conocer cuántos españoles, en edad de trabajar, están ocupadas o están en paro.
- En una fábrica de bombillas queremos saber cuál es la duración media del modelo de 60 vatios.

La característica o propiedad que estudiamos se denomina variable aleatoria y el conjunto de individuos (que pueden ser personas, plantas, objetos, ...) en el que se estudia población.

Pero en muchas ocasiones no se puede estudiar esa característica en todos y cada uno de los individuos de la población. ¿Por qué? Puede ser debido a muchas causas. Veamos algunas:

- La población es muy numerosa y por ello se necesitaría un tiempo y un dinero del que no siempre se dispone.
- No es posible controlar la totalidad de los individuos, porque se trata de personas o animales en constante movimiento (imaginemos aves migratorias, animales de un ecosistema o los clientes de un centro comercial en un día concreto).
- El proceso de medición es destructivo (si encendemos todas las bombillas de 60w de una fábrica hasta que se fundan, para analizar su duración...)

Pensemos en un estudio que recoge datos de toda la población de un Estado, como es el Censo.

Este estudio precisa unos recursos enormes: elaboración de las preguntas, millones de formularios para completarlos, su distribución, cientos de miles de personas entregándolos en mano, comprobando de se rellenan correctamente los datos y entregándolos en su lugar, introducción de todos los millones de datos recogidos en sistemas informáticos, procesamiento de la información (separando poblaciones, edades, sexos, ...), análisis de resultados, etc, etc.

En estos casos, ¿que hacer? Pues no queda más remedio que elegir sólo una parte de la población, que se denomina muestra, y realizar el estudio sólo en los individuos de esa muestra.

Lógicamente, este procedimiento tiene un riesgo: las conclusiones que sacamos se refieren a los individuos de la población que están en la muestra. ¿Nos servirán para todos los de la población? Porque esta es la idea: que el estudio de la muestra sirva para inferir características de toda la población.

Concluyamos, pues:

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.

El estudio de la muestra sirve para **inferir** características de toda la población.

☞ Ejemplo: queremos conocer cuántos habitantes de nuestra ciudad, que tiene 13 000 habitantes fuman. Elegimos para averiguarlo una muestra de 100 individuos y obtenemos que el 31% de ellos fuman

Tenemos así que la población tiene un tamaño N = 13 000, y que la muestra tiene un tamaño n = 100. Como hemos obtenido que un 31% de los individuos de la muestra son fumadores, inferimos que también lo será idéntico porcentaje en la población.

Concluimos que hay 13 000 × 0,31 = 4 030 fumadores en nuestra ciudad.

2.Tipos de muestreo

En este epígrafe analizaremos los conceptos y el vocabulario, de forma resumida, en lo que respecta al procedimiento de elección de los individuos que constituirán una muestra.

Supongamos que queremos conocer una determinada característica en una población, pero que nos vemos obligados a restringirnos a su estudio en una muestra. 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.

De forma simple, podemos distinguir los siguientes tipos de muestreos aleatorios:

- Aleatorio **simple**: se enumeran los individuos y se eligen al azar (por sorteo) tantos como sea el tamaño de la muestra.
- Aleatorio **sistemático**: se enumeran los individuos y se elige al azar sólo uno de ellos, que se denomina **origen.** A partir de éste se toman los siguientes mediante saltos numéricos idénticos. Estos saltos están determinados por un número denominado **coeficiente de elevación**.

  Vamos a detallar un poco el procedimiento: partamos de que el tamaño de la población es N y el de la muestra es n.

  El coeficiente de elevación será $c=N/n$

  El origen se determinará eligiendo al azar un individuo x de la población de entre los c primeros.

  De esta forma quedan:

$$x, x+c, x+2c, \dots, x+(n-1)c$$

Para obtener muestras aleatorias se hace uso de los llamados **números aleatorios**.

En una población de tamaño N que está enumerada del 1 al N, un individuo elegido al azar es:

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

donde

$$0 \leq aleatorio < 1$$

es un número generado aleatoriamente.

- Aleatorio **estratificado**: la población se considera dividida en capas denominadas **estratos**.

  Al estudiar la muestra, elegimos aleatoriamente de cada uno de los estratos un número de individuos **proporcionalmente** al peso del estrato en la población.

☞ Ejemplo: Tenemos una asociación formada por 500 socios y hay que elegir al azar, para las elecciones del Equipo de Gobierno, a cinco de ellos para que compongan la mesa electoral.

Veamos un procedimiento simple para obtenerlos:

Primero enumeramos los socios del 1 al 500.

Y luego generamos con una calculadora cinco números aleatorios distintos:

0,75 , 0,263 , 0,052 , 0,21 , 0,835

Terminamos determinando los números de los socios elegidos:

$$N_1 = E(500 \times 0{,}75 + 1) = E(376) = 376$$

$$N_2 = E(500 \times 0{,}263 + 1) = E(132{,}5) = 132$$

$$N_3 = E(500 \times 0{,}052 + 1) = E(27) = 27$$

$$N_4 = E(500 \times 0{,}75 + 1) = E(106) = 106$$

$$N_5 = E(500 \times 0{,}835 + 1) = E(418{,}5) = 418$$

Resumiendo, la mesa electoral la formarán los socios cuyos números son los siguientes:

27 , 106 , 132 , 376 , 418

☞ Ejemplo: En una población hay censados 15 400 habitantes. Vamos a elegir una muestra de 200 individuos para cierto estudio.

Vamos a explicar cómo realizar un muestreo aleatorio sistemático.

En primer lugar suponemos que hemos enumerado todos los individuos de la población, del 1 al 15 400.

En segundo lugar, calculamos el coeficiente de elevación:

$$c = \frac{N}{n} = \frac{15400}{200} = 77$$

En tercer lugar, hemos de hallar el origen: será un individuo elegido al azar de entre los 77 primeros. Para ello hacemos uso de los números aleatorios:

$$x_0 = E(77 \times aleatorio + 1) = E(77 \times 0{,}242 + 1) = 19$$

Concluimos ya, eligiendo los 200 individuos:

19 ; 19 + 77 ; 19 + 2 · 77 ; ... ; 19 + 199 · 77 → 19 ; 96 ; 173 ; ... ; 15342

Observa que hemos generado un número aleatorio entre 0 y 1 para elegir el origen, tal y como se explicó anteriormente.

Una vez que tenemos el origen y el coeficiente de elevación, formamos la sucesión numérica que indica los números de los elegidos para la muestra.

Fíjate que el último se obtiene añadiendo al origen el coeficiente de elevación multiplicado por $n - 1$.

☞ Ejemplo: En un I.E.S. que cuenta con 1.000 alumnos se realiza un estudio sobre el consumo de drogas (tabaco, alcohol, estupefacientes, ...) por parte de los estudiantes. Para ello se eligen a 50 de ellos.

Deseamos diferenciar los estudiantes de E.S.O. de los de Bachillerato y separar la respuesta por sexos. En Secretaría nos dan esta información:

	ESO	***Bachillerato***
Chicas	250	225
Chicos	350	175

Seguiremos un muestreo aleatorio estratificado. La población se ha dividido en estratos: sexo y nivel de estudios. Debemos hacer lo mismo con la muestra.

Primero enumeramos la población del 1 al 1000.

Segundo calculemos la composición de la muestra. Debemos garantizar la representatividad respetando la proporción de cada estrato en la composición de la población:

$$Chicas\,ESO=\frac{250}{1000}\cdot 50=12{,}5 \qquad Chicas\,Bach=\frac{225}{1000}\cdot 50=11{,}25$$

$$Chicos\,ESO=\frac{350}{1000}\cdot 50=17{,}5 \qquad Chicos\,Bach=\frac{175}{1000}\cdot 50=8{,}75$$

Hay un empate a la hora de elegir un chica o un chico de la ESO: lo elegimos a cara o cruz y punto.

La composición de la muestra queda así:

	ESO	***Bachillerato***
Chicas	13	11
Chicos	17	9

Ahora sólo queda elegir de cada estrato de la población, por muestreo aleatorio simple, el número de individuos que indica la tabla anterior.

Ahora detalla tú el proceso de elección de las chicas de bachillerato de forma sistemática y de los chicos de bachillerato de forma simple.

☞ Ejemplo: En la población anterior, donde hay censados 15.400 habitantes, se desea ahora estudiar separadamente los sexos. Sabemos que es:

Hombres	***Mujeres***
7 124	8 276

¿Qué tipo de muestreo debe realizarse? ¿Cómo debemos proceder ahora para elegir la muestra de 200 individuos para cierto estudio.

3.Comparación de parámetros

Cuando estudiamos una característica en una población y recurrimos a la formación de muestras, es fundamental saber distinguir entre los parámetros de la población y los parámetros de una muestra.

En este epígrafe trataremos de comprender la diferencia entre ambos en dos casos: las medias y las proporciones.

❑ Media poblacional y medias muestrales.

En un grupo de 20 personas se ha estudiado la variable

X = "número de horas que pasa al día frente al televisor"

quedando distribuida así:

X = { 1 , 1 , 1 , 2 , 2 , 2 , 2 , 3 , 3 , 3 , 3 , 3 , 4 , 4 , 4 , 4 , 4 , 5 , 5 , 6 }

1. Hallemos la media (μ) y la desviación típica (σ) de la población:

$$\mu=\frac{\sum x_i}{N}=\frac{62}{20}=3{,}1$$

$$\sigma=\sqrt{\frac{\sum x_i^2}{N}-\mu^2}=\sqrt{\frac{230}{20}-3{,}1^2}=\sqrt{1{,}89}=1{,}3747\ldots$$

2. Formemos ahora, de forma aleatoria simple, una muestra de 10 individuos y anotemos el valor de la variable en ellos:

M = { 1 , 1 , 2 , 2 , 2 , 3 , 3 , 4 , 4 , 5 }

Hallemos la media de la muestra ($\bar{x}$) y la desviación típica (s) de la muestra:

$$\bar{x}=\frac{\sum x_i}{n}=\frac{27}{10}=2{,}7$$

$$s=\sqrt{\frac{\sum x_i^2}{n}-\mu^2}=\sqrt{\frac{89}{10}-2{,}7^2}=\sqrt{1{,}61}=1{,}2688\ldots$$

3. Si ahora formamos otra muestra **M'** de igual tamaño, obtendremos otra media muestral ($\bar{x}'$) y otra desviación típica muestral (s').

Así, los parámetros de la población son únicos, pero los parámetros muestrales no. Éstos dependen de la muestra elegida.

Precisamente esta relación entre la media poblacional y las distintas medias muestrales es la que vamos a estudiar.

Hazlo tú ahora formando otra muestra de tamaño $n = 10$ elegida mediante un muestreo aleatorio simple.

❑ Proporción poblacional y proporciones muestrales.

Una caja tiene chinchetas buenas y chinchetas defectuosas. Las hemos ido cogiendo de una en una y esto es lo que nos ha salido:

X = { *b* , *b* , *d* , *d* , *d* , *b* , *b* , *d* , *b* , *d* , *b* , *b* , *b* , *d* , *d* , *b* , *d* , *d* , *b* , *d* }

1. Hallemos la proporción (p) de chinchetas defectuosas que hay en la población:

$$p=\frac{10}{20}=0{,}5$$

José Álvarez Fajardo

2. Formemos ahora, de forma aleatoria, una muestra de tamaño $n = 10$:

$$\boldsymbol{M} = \{ b , b , d , d , b , d , b , b , d , d \}$$

la proporción ($\tilde{p}$) de chinchetas defectuosas que hay en esas 10 es una proporción muestral:

$$\tilde{p} = \frac{5}{10} = 0{,}5$$

3. Si ahora formamos otra muestra $\boldsymbol{M'}$ de igual tamaño, obtendremos otra proporción muestral ($\tilde{p}'$).

Así, la proporción poblacional de una característica es única, pero las proporciones muestrales no. Éstas dependen de la muestra elegida.

Bueno, en la muestra de la izquierda ha coincidido la proporción muestral con la poblacional, pero es evidente que eso no siempre será así. Hazlo tú ahora formando otra muestra de tamaño n = 10 elegida mediante un muestreo aleatorio simple.

Precisamente esta relación entre la proporción poblacional y las distintas proporciones muestrales es la que vamos a estudiar.

4. Distribución de las medias muestrales

❑ Distribuciones en poblaciones finitas.

Vamos a comenzar el estudio de la relación entre la media de la población y las medias muestrales con el caso de una población finita y una variable aleatoria discreta.

Consideremos una variable X en una población de tamaño $N = 4$ que toma los siguientes valores:

$$\boldsymbol{X} = \{ 1 , 2 , 3 , 4 \}$$

Vamos a ir dando estos pasos:

1. La media de la población y la desviación típica.
2. Formaremos todas las muestras de tamaño $n = 2$, con reemplazamiento. Y hallaremos la distribución $\overline{X}$ de las medias muestrales.
3. Calcularemos la media de la distribución de las medias muestrales, $\overline{X}$, así como su desviación típica.
4. Compararemos ambos y obtendremos las pertinentes consecuencias.

<u>Paso 1</u>: La media y la desviación típica poblaciones son, respectivamente:

$$\mu = \frac{\sum x_i}{N} = \frac{10}{4} = 2{,}5$$

$$\sigma^2 = \frac{\sum x_i^2}{N} - \mu^2 = \frac{30}{4} - 2{,}5^2 = 1{,}25 \quad \rightarrow \quad \sigma = \sqrt{1{,}25}$$

Paso 2: El conjunto de las muestras de tamaño dos que pueden formarse es:

{ 1-1 , 1-2 , 1-3 , 1-4 , 2-1 , 2-2 , 2-3 , 2-4 , 3-1 , 3-2 , 3-3 , 3-4 , 4-1 , 4-2 , 4-3 , 4-4}

Si hallamos las medias de todas las muestras obtendremos la denominada "distribución de las muestras muestrales":

$$\bar{X}=\{1, 1{,}5, 2, 2{,}5, 1{,}5, 2, 2{,}5, 3, 2, 2{,}5, 3, 3{,}5, 2{,}5, 3, 3{,}5, 4\}$$

Paso 3: Hallemos la media y la desviación típica de la distribución de las medias muestrales ($\bar{X}$):

Media: $\bar{\mu}=\frac{\sum \bar{x}_i}{16}=\frac{40}{16}=2{,}5$

Desviación típica: $\bar{\sigma}^2=\frac{\sum \bar{x}_i^2}{16}-\bar{\mu}^2=\frac{110}{16}-2{,}5^2=0{,}625 \rightarrow \bar{\sigma}=\sqrt{0{,}625}$

Observemos las barras que tienen los parámetros: fundamentales para no confundirlos con los

Paso 4: Comparemos ambas.

Sobre las medias, es evidente que ambas coinciden. Y eso ocurre en general:

$$\bar{\mu}=\mu$$

En cuanto a las varianzas, no coinciden; pero observamos que ésta es la mitad de la de la población:

$$\bar{\sigma}^2=\sigma^2/2 \rightarrow \bar{\sigma}=\sigma/\sqrt{2}$$

Ese dos aparece porque las muestras son de tamaño $n = 2$.

En general, concluimos que:

Sea X una v.a. en una población finita de tamaño N, y sea $\bar{X}$ la distribución de las medias muestrales, con reemplazamiento, de tamaño n.

Se cumple:

$$\bar{\mu}=\mu$$

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

Cuando las muestras son sin reemplazamiento es:

Sea X una v.a. en una población finita de tamaño N, y sea $\bar{X}$ la distribución de las medias muestrales, sin reemplazamiento, de tamaño n.

Se cumple:

$$\bar{\mu}=\mu$$

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

Observemos cómo se designan los parámetros de la población y los de la distribución de las medias muestrales. Cuidado para no confundirlos:

Distribución X en la población:

Media: μ

Desviación Típica: σ

Distribución $\bar{X}$:

Media $\bar{\mu}$

Desviación Típica: $\bar{\sigma}$

Comprueba tú esto con el ejemplo anterior, pero tomando ahora las muestras sin reemplazamiento (no podemos tomar dos veces el mismo dato).

❑ Distribuciones normales.

Las conclusiones a las que hemos llegado anteriormente pueden extrapolarse al caso de distribuciones normales:

> Sea X una v.a. en una población y designemos por $\bar{X}$ a la distribución de las medias muestrales para el tamaño n.
>
> Tenemos entonces que $\bar{X}$ también es normal con:
>
> $$\bar{\mu}=\mu$$
>
> $$\bar{\sigma}=\frac{\sigma}{\sqrt{n}}$$

☞ Ejemplo: en una Universidad se sabe que las tallas de los alumnos se distribuyen normalmente con una media de 175 cm, teniendo una desviación típica de 15 cm.

> La distribución X en la población es normal con:
>
> Media: $\mu=175$
>
> Desviación Típica: $\sigma=15$

Si consideramos todas las posibles muestras de 25 estudiantes, las medias muestrales $\bar{X}$ también se distribuyen normalmente, con:

$$\bar{\mu}=\mu=175 \text{ cm.}$$

$$\bar{\sigma}=\frac{\sigma}{\sqrt{25}}=\frac{15}{5}=3 \text{ cm.}$$

Calculemos ahora la probabilidad de que en una de las muestras la estatura media supere los 180 centímetros:

$$p(\bar{X}>180)\overset{*}{=}p(Z>1{,}67)=1-p(Z<1{,}67)=0{,}0475$$

$$(*)\,Z=\frac{\bar{X}-\bar{\mu}}{\bar{\sigma}}=\frac{180-175}{3}=\frac{5}{3}\approx 1{,}67$$

☞ Ejemplo: El peso de las naranjas de una cosecha se distribuye de forma normal con un peso medio de 200 gramos y una desviación típica de 20 gramos.

> La distribución X en la población es normal con:
>
> Media: $\mu=200$
>
> Desviación Típica: $\sigma=20$

Tomemos una naranja, al azar, de la cosecha: ¿cuál es la probabilidad de que pese menos de 190 gramos?

Veamos:

$$p(X<190)\overset{*}{=}p(Z<-0{,}5)=1-p(Z<0{,}5)=0{,}3085$$

$$(*)\,Z=\frac{X-\mu}{\sigma}=\frac{190-200}{20}=\frac{-1}{2}=-0'5$$

Tomemos ahora una muestra de 9 naranjas de la cosecha: ¿cuál es la probabilidad de que el peso medio de las naranjas de esa muestra esté por debajo de los 190 gramos?

Observemos ahora que nos referimos a la media en una muestra:

La distribución $\overline{X}$ es normal con $\left| \begin{array}{l} \overline{\mu}=\mu=200 \\ \overline{\sigma}=\dfrac{\sigma}{\sqrt{9}}=\dfrac{20}{3} \end{array} \right.$

Calculemos, pues:

$$p(\overline{X}<190)\overset{*}{=}p(Z<-1{,}5)=1-p(Z<1{,}5)=0{,}0668$$

$$(*)\,Z=\frac{\overline{X}-\overline{\mu}}{\overline{\sigma}}=\frac{190-200}{20/3}=-1{,}5$$

Es importante constatar que, cualquiera que sea la distribución de partida en la población, aunque no sea normal, la distribución de las medias muestrales sigue siendo prácticamente normal, simplemente con tomar el tamaño muestral suficientemente grande ($n > 30$):

> Sea X una v.a. en una población y designemos por $\overline{X}$ a la distribución de las medias muestrales de un tamaño n suficientemente grande ($n > 30$) .
>
> Tenemos entonces que $\overline{X}$ es prácticamente normal con:
>
> $$\overline{\mu}=\mu$$
>
> $$\overline{\sigma}=\frac{\sigma}{\sqrt{n}}$$

☞ Ejemplo: Una confitería prepara dulces con un peso medio de 100 gr. y una desviación típica de 10 gr. Esos dulces se empaquetan en bandejas de tres docenas. Vamos a calcular la probabilidad de que una bandeja pese menos de tres kilos y medio.

La distribución X en la población cumple:

Media: $\mu=100$

Desviación Típica: $\sigma=10$

Observemos que si la bandeja de 36 dulces pesa menos de 3500 gramos, el peso medio de los dulces será:

$$\overline{x}<\frac{3500}{36}\approx 97{,}22 \text{ gr.}$$

Luego queremos hallar $p(\overline{X}<97{,}22)$.

Como $n = 36$ es suf. grande, es $\overline{X}$ es normal con $\left| \begin{array}{l} \overline{\mu}=\mu=100 \\ \overline{\sigma}=\dfrac{\sigma}{\sqrt{36}}=\dfrac{5}{3} \end{array} \right.$

$$p(\overline{X}<97{,}22)\overset{*}{=}p(Z<-1{,}67)=1-p(Z<1{,}67)=0{,}0475$$

$$(*)\,Z=\frac{\overline{X}-\overline{\mu}}{\overline{\sigma}}=\frac{97{,}22-100}{5/3}=-1{,}67$$

5. Distribución de las proporciones muestrales

Estamos aquí frente al mismo estudio de antes, pero ahora referido a una proporción (o porcentaje) en lugar de a una media:

En una población la característica C aparece con una proporción p. Formemos todas las muestras de tamaño n que sea posible y en cada una de ellas calculemos la proporción $\tilde{p}$ con que aparece C.

¿Cómo se distribuyen ahora las proporciones muestrales $\tilde{p}$? He aquí la respuesta:

> Sea C una característica que se presenta en una población con una proporción p. Designemos $q = 1 - p$.
>
> Si $\tilde{P}$ es la distribución de las proporciones muestrales, para un tamaño n suficientemente grande ($np \geq 5$ y $nq \geq 5$), entonces:
>
> $$\tilde{P} \approx N\left(p, \sqrt{\frac{pq}{n}}\right)$$

☞ Ejemplo: uno de cada cinco fumadores que desarrolla problemas cardíacos continúa con el hábito tras sufrir alguna crisis importante y a pesar de estar informado sobre el efecto mortal del cigarrillo. En un hospital hay ingresados cincuenta fumadores que han desarrollado problemas cardíacos. ¿Cuál es probabilidad de que más la cuarta parte de ellos sigan fumando?

> Esta cifra es real y está referida a España.
>
> Recuerda que el tabaco provoca una fuerte adicción y es perjudicial tanto para el fumador como para los que aspiran sus humos.

La proporción aquí es:

$$p = \frac{1}{5} = 0{,}20 \quad \rightarrow \quad q = 1 - p = 0{,}80$$

Queremos hallar $p\left(\tilde{p} > \frac{1}{4}\right)$ para un tamaño muestral $n = 50$.

Como n es suficientemente grande, pues $np \geq 5$ y $nq \geq 5$:

$$\tilde{P} \text{ es normal con } \left| \begin{array}{l} \mu = p = 0{,}20 \\ \sigma = \sqrt{\frac{p \cdot q}{n}} = \sqrt{0{,}0032} \end{array} \right.$$

Así que la probabilidad pedida es:

$$p(\tilde{P} > 0{,}25) \overset{*}{=} p(Z > 0{,}88) = 1 - p(Z < 0{,}88) = 0{,}1894$$

$(*)\, Z = \frac{\tilde{P} - \mu}{\sigma} \approx 0{,}88$

6. Intervalos característicos en las distribuciones normales.

Como vemos, tanto las medias como las proporciones muestrales se distribuyen normalmente cuando los tamaños de las muestras son suficientemente grandes.

Son particularmente importantes en las distribuciones normales los denominados intervalos característicos, y aparecerán en los trabajos con muestras. Veámoslos brevemente.

Sea X una variable que se distribuye normalmente con media μ y desviación típica σ, y consideremos una probabilidad p.

Al intervalo (a, b) centrado en la media que cumple

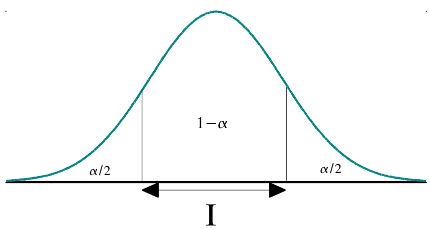
$$p(a < X < b) = p$$

se le llama intervalo característico correspondiente a la probabilidad p.

Observemos que:

1. El intervalo será de la forma $(\mu - z\cdot\sigma, \mu - z\cdot\sigma)$.
2. La probabilidad fuera del intervalo es $1 - p$. Si llamamos $1 - p = \alpha$ las probabilidades quedan repartidas como se muestra en el margen.

   Nótese que si $1 - p = \alpha \rightarrow p = 1 - \alpha$.



Es usual designar por $z_{\alpha/2}$ a ese valor z, pues

$$p\left(z > z_{\alpha/2}\right) = \frac{\alpha}{2}$$

y se le denomina valor crítico correspondiente a la probabilidad $p = 1 - \alpha$

En el margen tenemos cómo queda el gráfico correspondiente a un intervalo característico y a continuación el resumen:

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

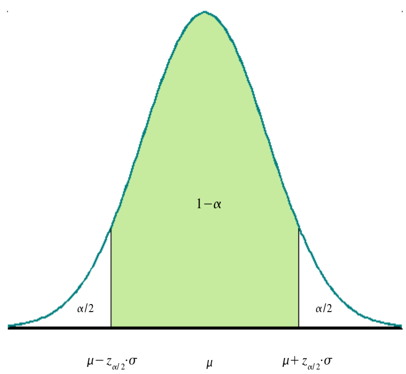
$$I = \left(\mu - z_{\alpha/2}\cdot\sigma\ ,\ \mu + z_{\alpha/2}\cdot\sigma\right)$$

donde el número $z_{\alpha/2}$, que es llamado valor crítico, cumple:

$$p\left(z > z_{\alpha/2}\right) = \frac{\alpha}{2}$$

Y por ello

$$p\left(\mu - z_{\alpha/2}\cdot\sigma < X < \mu + z_{\alpha/2}\cdot\sigma\right) = 1 - \alpha$$



☞ Nota: Los valores que más aparecen en la práctica son:

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

☞ Ejemplo: obtengamos el valor crítico asociado a $p=1-\alpha=0{,}95$.

Es $p=1-\alpha=0{,}95 \rightarrow \alpha=0{,}05 \rightarrow \frac{\alpha}{2}=0{,}025$

Así: $p\left(z>z_{\alpha/2}\right)=0{,}025 \rightarrow p\left(z<z_{\alpha/2}\right)=1-0{,}025=0{,}975$

Buscando en la tabla de la normal típica el valor correspondiente a esa probabilidad obtenemos:

$$z_{\alpha/2}=1{,}96$$

7.Intervalos de confianza para la media

❑ Intervalos de confianza.

Consideremos una variable X en una población, con una media μ y una desviación típica σ conocidas. En determinadas condiciones, tenemos que la distribución de las medias muestrales $\bar{X}$ es normal con:

Esas condiciones son:

- X es normal.

- El tamaño muestral es n > 30

$$\bar{\mu}=\mu \quad \text{y} \quad \bar{\sigma}=\frac{\sigma}{\sqrt{n}}$$

Así, un intervalo característico de $\bar{X}$ para la probabilidad $p=1-\alpha$ es:

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

Observemos que por ello, para el $p\cdot 100\%$ de las muestras la media muestral $\bar{x}$ está en el intervalo

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

Si tomamos una muestra en la que $\bar{x}$ esté en ese intervalo, entonces μ está en el intervalo

Los intervalos de confianza se usan para estimar la media de la población, cuando es desconocida, a partir de una muestra

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

A un intervalo así se le llama intervalo de confianza para la media poblacional, con un nivel de confianza $p=1-\alpha$.

☞ Nota importante: Si tomamos una muestra concreta y a partir de ella calculamos $\bar{x}$ y formamos un intervalo de confianza, la media de la población μ puede estar o no en él.

Lo único que podemos afirmar es que, fijado un nivel de confianza $p=1-\alpha$, si repetimos ese proceso de formación de muestras y de intervalos, en el $p\cdot 100\%$ de los intervalos de confianza calculados estará μ y en el resto no.

☞ Ejemplo: Una determinada variable aleatoria que sigue una ley normal de media μ y desviación típica 3. Una muestra de tamaño 100, seleccionada al azar, ha dado como media 9. Vamos a determinar un intervalo de confianza de μ si se fija como coeficiente de confianza 0'95.

Organicemos todo:

La variable aleatoria X sigue una distribución normal con $\left| \begin{array}{l} \mu=¿? \\ \sigma=3 \end{array} \right.$

Tamaño muestral: $n = 100$

Media muestral : $\bar{x}=9$

Nivel de confianza: $p=1-\alpha=0{,}95$ $\rightarrow$ Valor crítico: $z_{\alpha/2}=1{,}96$

El intervalo de confianza es:

$$I=\left(\bar{x}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\,,\,\bar{x}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)=\left(9-1{,}96\cdot\frac{3}{\sqrt{100}}\,,\,9+1{,}96\cdot\frac{3}{\sqrt{100}}\right)$$

Realizando los cálculos:

$$I=(8{,}412\,,\,9{,}588)$$

Observaciones:

1. No sabemos con certeza si μ está en ese intervalo. Por eso se les llama intervalos de confianza, porque se confía en que sí esté en ellos.
2. El nivel de confianza indica con qué probabilidad la muestra elegida nos dará un intervalo correcto.

Observemos que podemos aplicar la fórmula del intervalo de confianza por partida doble:

- Es X normal.
- Es $n = 100 > 30$

❑ Error.

Cuando formamos un intervalo de confianza, obviamente al número

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

se le llama error máximo admisible.

Observemos el error máximo depende del nivel de confianza y del tamaño muestral.

- Cuanto mayor es el nivel de confianza, mayor es el error y, por consiguiente, la amplitud del intervalo de confianza.
- Cuanto mayor es el tamaño muestral, menor es el error y, por consiguiente, la amplitud del intervalo de confianza.

Observemos que el intervalo de confianza

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

– Está centrado en la media muestral $\bar{x}$

– Su límite inferior es $\bar{x}-E$ y su límite superior es $\bar{x}+E$. Así, queda:

$$I=(\bar{x}-E,\bar{x}+E)$$

– Su longitud o amplitud es $L=2\cdot E$

☛ Ejemplo: Se sabe que las tallas de los alumnos de una Universidad se distribuye normalmente con una desviación típica de 6 cm.

Halle el tamaño muestral necesario para estimar la talla media de los alumnos de la Universidad con un nivel de confianza del 95% y un error máximo no superior a 1,2 cm.

Organicemos todo:

La v.a. X = "talla de los alumnos de la Univ." es normal con $\left|\begin{array}{l}\mu=¿?\\ \sigma=6\end{array}\right.$

Observemos que podemos aplicar la fórmula del error máximo porque sabemos que es X normal.

Error máximo: $E = 1{,}2$

Media muestral : $\bar{x}=9$

Nivel de confianza: $p=1-\alpha=0{,}95$ → Valor crítico: $z_{\alpha/2}=1{,}96$

Tamaño muestral: n = "*es lo que deseamos hallar*"

Despejamos de la fórmula del error:

$$E=z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} \rightarrow 1{,}2=1{,}96\cdot\frac{6}{\sqrt{n}} \rightarrow \sqrt{n}=1{,}96\cdot\frac{6}{1{,}2} \rightarrow \sqrt{n}=9{,}8$$

Elevando al cuadrado:

$$n=96$$

8.Intervalos de confianza para una proporción

Razonando de la misma forma que para la media en una población, llegamos a la siguiente conclusión en el caso de 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 de confianza**

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

donde $\tilde{q}=1-\tilde{p}$.

Se llama **error máximo** admisible a

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

Recordemos que debe ser

$np\geq 5$ y $nq\geq 5$

Y además debe ser

$n > 30$

para poder estimar p por $\tilde{p}$ en el error máximo admisible

☞ Ejemplo: Tomada al azar una muestra de 400 personas mayores de 18 años en una ciudad, se encontró que 130 de ellos acudían al cine con regularidad.

Obtengamos un intervalo para estimar la proporción de espectadores, con un nivel de confianza del 90%, que hay entre los adultos de esa ciudad.

Organicemos todo:

Estudiamos la característica

C = "ser mayor de 18 años que acude al cine con regularidad"

Tamaño muestral: $n = 400$

Proporción muestral: $\tilde{p}=\frac{140}{400}=0{,}35 \rightarrow \tilde{q}=1-p=0{,}65$

Nivel de confianza: $p=1-\alpha=0{,}90 \rightarrow$ Valor crítico: $z_{\alpha/2}=1{,}645$

El intervalo de confianza es:

$$I=\left(\tilde{p}-z_{\alpha/2}\cdot\sqrt{\frac{\tilde{p}\cdot\tilde{q}}{n}}\ ,\ \tilde{p}+z_{\alpha/2}\cdot\sqrt{\frac{\tilde{p}\cdot\tilde{q}}{n}}\right)=(0{,}31\ ,0{,}39)$$

Con un nivel de confianza del 95%, podemos afirmar que el porcentaje de adultos que va al cine con regularidad está comprendido entre el 31 y el 39 por ciento.

Ejercicios

1. [S/97] Los estudiantes de Bachillerato de una cierta comunidad duermen un número de horas que se distribuye según una ley normal de media μ desconocida y de desviación típica 3. A partir de una muestra de tamaño 30 se ha obtenido una media muestral igual a 7 horas.

Halla un intervalo de confianza al 96% para la media μ de horas de sueño.

2. [S/97] En un centro de Bachillerato se desea estimar el rendimiento medio de sus alumnos mediante un test. Se eligen, al azar, 100 estudiantes para los que se obtiene una puntuación media de 105 y una desviación típica de 16.

Supuesto que el rendimiento sigue una ley normal, encontrar al nivel de confianza del 99% un intervalo para la media de todos los estudiantes del centro.

3. [S/97]Un laboratorio farmacéutico fabrica un producto para la caída del cabello. Dicho producto es envasado en botes cuya etiqueta indica que su contenido aproximado es de 10 cc. Se eligen, al azar, siete de estos botes y se miden sus contenidos, dando el siguiente resultado:

9'7 , 10'1 , 10'2 , 9'9 , 9'8 , 10 , 10'3

Sabiendo que el contenido es una variable aleatoria normal con desviación típica 0'2, ¿podemos asegurar, con un nivel de confianza del 95%, que la capacidad media de los botes que se fabrican es la indicada en el bote?

4. [S/97] Una determinada variable aleatoria que sigue una ley normal de media μ y desviación típica 3. Una muestra de tamaño 100, seleccionada al azar, ha dado como media 9.

Determina un intervalo de confianza de μ si se fija como coeficiente de confianza 0'95.

5. [S/97] Una máquina automática fabrica piezas cuya longitud sigue una distribución normal de media desconocida y desviación típica 0'5 mm. Para estimar la longitud media se toma una muestra de 25 piezas, obteniéndose una media muestral de 50 m.

Calcula un intervalo de confianza del 95% para la longitud media de la población.

6. [S/97] Los siguientes datos corresponden a los salarios mensuales, en miles de pesetas, de un grupo de trabajadores de un hospital:

100, 110, 120, 150, 90, 80, 115, 110, 125, 600

a) Halla el porcentaje de salarios de esta muestra que están en el intervalo $(\bar{x}-\sigma, \bar{x}+\sigma)$.

b) Razona si se deben utilizar estos datos con el propósito de estimar la media salarial de todos los trabajadores españoles.

7. [S/97]

a) Determina un intervalo, con el 95% de confianza, para la media de una variable normal que tiene una desviación típica $\sigma=3$, teniendo en cuenta que en una muestra de taño 100 se ha obtenido una media muestral $\bar{x}=5$.

b) ¿Cuál debería haber sido el tamaño de la muestra si se quiere obtener un intervalo de confianza (con el mismo nivel de confianza) para la media con una longitud 0'4?

8. [S/98] Una variable aleatoria X sobre una población tiene de media 50 y de desviación típica 5.

Extraemos, aleatoriamente, de dicha población 1.000 muestras, todas ellas de tamaño 64. De cada muestra calculamos su media, y llamamos A al conjunto de números formados con esas medias.

a) Diga, de forma razonada, qué valores se pueden esperar para la media y la desviación típica de A.

b) ¿Cuál es la probabilidad de que una de esas muestras tenga una media comprendida entre 48'5 y 50'5?

9. [S/98] Se ha tomado una muestra de precios de un mismo producto alimenticio en 16 comercios, elegidos al azar en un barrio de la ciudad, y se han encontrado los siguientes precios:

95 , 108 , 97 , 112 , 99 , 106 , 105 , 100 , 99 , 98 ,
104 , 110 , 107 , 111 , 103 , 110

Suponiendo que los precios de este producto se distribuyen según una ley normal de varianza 25 y media desconocida:

a) ¿Cuál es la distribución de las medias muestrales?

b) Determine el intervalo de confianza, al 95% para la media poblacional.

10. [S/98] La variable altura de las alumnas que estudian en una escuela de idiomas sigue una distribución normal de media 1'62 m y desviación típica 0'12 m. ¿Cuál es la probabilidad de que la media de una muestra aleatoria de 100 alumnas sea mayor que 1'60 m?

11. [S/98] El peso de los individuos de una ciudad se distribuye según una ley normal de media desconocida y varianza $9\ kg^2$. Se ha seleccionado, en esa ciudad, una muestra aleatoria que ha dado un peso medio de $65\ kg$.

Con una confianza del 96% se ha construido un intervalo para la media poblacional cuyo límite inferior ha resultado ser $62'95\ kg$.

a) ¿Cuál ha sido el tamaño de la muestra seleccionada?

b) Determine el límite superior del intervalo.

12. [S/98] La variable X se distribuye según una ley normal de media 10 y desviación típica 3. Determine el tamaño de una muestra extraída de la población, de modo que la probabilidad de la media muestral esté por encima de 12 sea de $0'025$.

13. [S/98] Sea una población formada por sólo tres elementos, con valores 2 , 4 y 6. Consideremos todas las muestras, con reemplazamiento, de tamaño 2.

Calcule la media y desviación típica de la población así como de las medias muestrales. ¿Qué relación hay entre ambas medias?

14. [S/98] Se sabe que la desviación típica de las tallas de los alumnos de una Universidad es igual a 6 cm. Para estimar la talla media dichos alumnos se toma una muestra de 64 estudiantes, resultando una media muestral de 173 cm.

a) Determine un intervalo de confianza de la talla media de los alumnos con un nivel de confianza de $0'97$.

b) Halle el tamaño muestral necesario para estimar la talla media de los alumnos de la Universidad con un nivel de confianza del 95% y un error máximo no superior a 1'2 cm

15. [S/98] Un contable toma una muestra aleatoria de tamaño $n = 36$ de una población de 1.000 cuentas pro cobrar. El valor medio de las cuentas por cobrar es de 2.600 Pta. con una desviación típica de 450 Pta.

a) Calcule la probabilidad de que la media muestral sea inferior a 2.500 Pta.

b) Calcule la probabilidad de que la media muestral se encuentre a no más de 225 Pta. de la media de la población.

16. [S/98] Se dispone de una muestra aleatoria de 10 alumnos de una población de alumnos de 3º de ESO.

Se sabe, por experiencias anteriores, que la altura de los alumnos de ese curso se distribuye según una variable aleatoria normal de media 167 cm. y una desviación típica $3'2$ cm.

a) Determine la probabilidad de que la media muestral esté comprendida entre 166 cm. y la media poblacional.

b) ¿Cuál es la probabilidad de que la media muestral tenga un valor superior a 169 cm.?

17. [S/98] Un ascensor admite como peso máximo 300 Kg. La población de usuarios tiene un peso que se distribuye según una ley normal de media 70 Kg. y desviación típica 10 Kg.

Calcule la probabilidad de que 4 personas cualesquiera de dicha población, que suban al ascensor, superen el peso máximo.

18. [S/98] Una máquina fabrica clavos cuya longitud sigue una distribución normal con desviación típica 0'5 mm. Se toma una muestra de 25 y se obtiene una longitud media, para los mismos, de 50 mm.

Calcule un intervalo de confianza del 95% para la longitud media de la población.

19. [S/98] Si los alumnos de preescolar de Andalucía tienen una estatura que es una variable aleatoria de media 95 cm. y desviación típica 16 cm. y consideramos una muestra aleatoria de 36 de tales alumnos, ¿cuál es la probabilidad de que la media de esa muestra tome valores comprendidos entre 90 cm. y 10 cm.?

20. [S/99] En una población, una variable aleatoria sigue una ley normal de media desconocida y desviación típica 20.

a) Si de una muestra de tamaño 25 se ha observado que la media es 2.743, determine un intervalo, con el 90% de confianza, para la media de la población.

b) Elegida una muestra, su media ha sido 2.740 ; se ha construido un intervalo de confianza, al 95%, que ha resultado ser (2.736'08 , 2.743'92). ¿Cuál era el tamaño de la muestra?

21. [S/99] El tiempo de vida de un insecto sigue una distribución normal con media desconocida y desviación típica 25 días. Para estimar la vida media se hace un seguimiento a la duración de la vida de una muestra de n insectos. Calcule el valor de n para que el intervalo de confianza de esta media, con un nivel de confianza del 95%, tenga una amplitud como máximo de 5 días.

22. [S/99] Las ventas mensuales de una tienda de electrodomésticos se distribuye según una ley normal con $\sigma = 90\,000$ pta. . En un estudio estadístico de las ventas realizadas en los últimos nueve meses, se ha encontrado un intervalo de confianza para la media mensual de las ventas, cuyos extremos son 466.300 y 583.900 Pta.

a) ¿Cuál ha sido la media de las ventas en estos 9 meses?

b) ¿Cuál es el nivel de confianza de ese intervalo?

23. [S/99] La media de las estaturas de una muestra aleatoria de 400 personas de una ciudad es 1'75 metros. Se sabe que la estatura de las personas de esa ciudad es una variable aleatoria que sigue una distribución normal con varianza $\sigma^2 = 0'16\ m^2$.

a) Construya un intervalo, de un 95% de confianza, para la media de las estaturas de la población.

b) ¿Cuál sería el mínimo tamaño muestral necesario para que pueda decirse que la verdadera media de las estaturas está a menos de 2 cm. de la media muestral, con una confianza del 90%?

24. [S/99] Se ha tomado una muestra aleatoria de 100 individuos a los que se ha medido el nivel de glucosa en sangre, obteniéndose una media muestral de 110 mg/cc. Se sabe que la desviación típica de la población es de 20 mg/cc.

a) Obtenga un intervalo de confianza, al 90%, para que el nivel de glucosa en sangre de la población.

b) ¿Qué error máximo se comete con la estimación anterior?

25. [S/99] Un fabricante de bombillas sabe que la desviación típica de la duración de las bombillas es 90 horas. Tomada una muestra de tamaño 100 se ha encontrado que la media de la duración de las bombillas ha sido 1.200 horas. Determine un intervalo, con el 95% de confianza, para la duración media de las bombillas.

26. [S/99] La media de edad de los alumnos que se presentan alas pruebas de acceso a la Universidad es de 18'1 años y la desviación típica 0'6 años.

a) De los alumnos anteriores se elige, al alzar, una muestra de 100, ¿cuál es la probabilidad de que media de la edad de la muestra esté comprendida entre 17'9 y 18'2 años?

b) ¿Qué tamaño debe tener una muestra de dicha población para que su media esté comprendida entre 17'9 y 18'3 años, con una confianza del 99'5%?

27. [S/99] Sea un conjunto de 4 bolas, marcadas con los números 1, 3, 5 y 7.

a) Escriba todas las muestras de tamaño 2 que podrían formarse con esas bolas si el muestreo se hace sin reposición. Calcule también las medias de los números de cada muestra y halle la media de todas esas muestras.

b) Ídem pero suponiendo que el muestreo se hace con reemplazamiento.

28. [S/99] Al calificar los exámenes de un numeroso grupo de opositores, se ha visto que sus puntuaciones siguen una distribución normal con una media de 72 puntos y una desviación típica de 9 puntos. ¿Cuál es la probabilidad de que en una muestra de 16 de esos opositores, elegidos al azar, se obtenga una puntuación media superior a 78 puntos?

29. [S/99] El tiempo que permanece cada paciente en la consulta del médico es una variable aleatoria que sigue una distribución normal con una desviación típica de 4 minutos.

Se ha tomado una muestra aleatoria de 256 pacientes de este médico y se ha encontrado que su tiempo medio de consulta ha sido de 10 minutos. ¿Cuál es el intervalo de confianza, a un nivel del 95%, para el tiempo medio de consulta que se deduce que esta muestra?

30. [S/99] Se sabe que el tiempo de reacción a un determinado estímulo se distribuye según una ley normal de media desconocida y desviación típica 0'15 segundos.

Observada una muestra de tamaño 9, se ha obtenido una media muestral de 0'85 segundos.

a) Obtenga un intervalo de confianza para la media de la población, con un nivel de confianza del 99%.

b) ¿Con qué nivel de confianza se debería construir un intervalo para la media de manera que los límites de dicho intervalo fuesen 0'768 y 0'932?

31. [S/99] En un colegio hay 2.000 alumnos distribuidos en 5 cursos así: 400 en 1er curso, 380 en 2^{o}, 520 en 3^{o}, 360 en 4^{o} y 340 en 5^{o}.

Se quiere seleccionar una muestra de 100 alumnos, utilizando la técnica de muestreo aleatorio estratificado con afijación proporcional y considerando cada curso como un estrato. ¿Cómo se seleccionaría dicha muestra

32. [S/00] A 400 personas elegidas al azar se les ha preguntado su gasto anual en libros, obteniéndose una cantidad media de 22.000 ptas. Con independencia de esta muestra se sabe que la desviación típica de la inversión en libros de la población es de 4.000 ptas.

a) Halle un intervalo de confianza al 90% y centrado, para la media poblacional de esta inversión.

b) ¿Qué tamaño muestral sería necesario para que el correspondiente intervalo de confianza del apartado anterior fuese (21.904 , 22.096)?

33. [S/00] Sea la población {–1, –2, 3, 4}.

Forme todas las muestras sin reemplazamiento y de tamaño 2, y calcule la media y la varianza de las medias muestrales, comparando los resultados obtenidos con la media y varianza de la población.

34. [S/00] Una máquina que envasa aceite en garrafas de 5 litros está ajustada de manera que la cantidad que llena sigue una ley normal con desviación típica $\sigma = 0'15 \ l$.

a) Calcule un intervalo de confianza del 95% para la media del contenido de las garrafas que llena esta máquina sabiendo que una muestra aleatoria de 36 de ellas dio un contenido medio de 4'97 litros.

b) ¿Contiene las garrafas 5 litros de aceite?

35. [S/00] La altura de los jóvenes andaluces se distribuye según una ley normal de media desconocida y varianza 25 cm^2.

Se ha seleccionado una muestra aleatoria y con una confianza del 95% se ha construido un intervalo para la media poblacional cuya amplitud es de 2'45 cm.

a) ¿Cuál ha sido el tamaño de esa muestra?

b) Determine el límite superior y el inferior del intervalo de confianza si la muestra tomada dio una altura media de 170 cm.

36. [S/00] Se conoce que el número de días de permanencia de los enfermos de un hospital sigue una distribución normal de media 8'1 días y desviación típica 9 días. Se elige, al azar, una muestra de 100 enfermos.

a) Razone cuál es la distribución de la media muestral.

b) ¿Cuál es la probabilidad de que la media muestral esté comprendida entre 8 y 10 días?

37. [S/00] La duración de los matrimonios en un país se distribuye según una ley normal con desviación típica 4'8 años.

a) Si se toma una muestra de 64 matrimonios cuya media es 16 años, halle un intervalo de confianza al 95% para la media de la población.

b) Si sabemos que $\mu = 15$, ¿cuál es la probabilidad de que la media de una muestra de tamaño 100 sea superior a 16'35 años?

38.[S/00] Se sabe que (2'9 , 3'7) es un intervalo de confianza al 95% para el peso medio, en kilogramos, de los recién nacidos en el año 1999, elaborado a partir de una muestra de 200 de ellos.

a) Razone si se puede deducir del intervalo de confianza dado la siguiente afirmación: “el peso medio de los recién nacidos del año 1999 es seguro que está entre 2,9 y 3,7 kg”.

b) ¿Qué se podría hacer para tener un intervalo de confianza más pequeño?

39. [S/00] En una muestra aleatoria de 225 individuos se ha obtenido una media de edad de 16'5 años. Se sabe que la desviación típica de la población de la que procede esa muestra es de 0'7 años.

a) Obtenga un intervalo de confianza, al 95%, para la edad media de la población.

b) ¿Qué error se comete en la estimación anterior?

40.[S/00] Una población está formada por los 4 números siguientes: 3 , 7 , 11 , 15.

a) Encuentre todas las muestras posibles, con reemplazamiento, de tamaño 2.

b) Halle la media y la desviación típica de la distribución muestral de medias.

c) Halle la media y la desviación típica de la población.

41. [S/00] En una población, una variable aleatoria sigue una ley normal con desviación típica 12.

a) Si en una muestra de tamaño 100, tomada al azar, se ha observado que la media es 40, determine un intervalo, con el 95% de confianza, para la media de la población.

b) Con un nivel de confianza del 90% se ha construido un intervalo para la media poblacional cuyo límite inferior ha sido 36'71. ¿Qué tamaño de muestra se ha tomado en este caso?

42.[S/00] El tiempo de reacción de un automovilista ante un obstáculo inesperado sigue una distribución normal con desviación típica de 0'1 segundos.

Deduzca el tamaño con el que se ha de tomar una muestra para tener una confianza del 90% de que el error de estimación del tiempo medio de reacción no supere los 0'02 segundos.

43.[S/06] El gasto anual, en videojuegos, de los jóvenes de una ciudad sigue una ley Normal de media desconocida μ y desviación típica 18 euros. Elegida, al azar, una muestra de 144 jóvenes se ha obtenido un gasto medio de 120 euros.

a) Indique la distribución de las medias de las muestras de tamaño 144.

b) Determine un intervalo de confianza, al 99 %, para el gasto medio en videojuegos de los jóvenes de esa ciudad.

c) ¿Qué tamaño muestral mínimo deberíamos tomar para, con la misma confianza, obtener un error menor que 1.9?

44.[S/06]

a) Los valores { 52, 61, 58, 49, 53, 60, 68, 50, 53 } constituyen una muestra de una variable aleatoria Normal, con desviación típica 6. Obtenga un intervalo de confianza para la media de la población, con un nivel de confianza del 92%.

b) Se desea estimar la media poblacional de otra variable aleatoria Normal, con varianza 49, mediante la media de una muestra aleatoria. Obtenga el tamaño mínimo de la muestra para que el error máximo de la estimación, con una confianza al 97%, sea menor o igual que 2.

45.[S/06] En una muestra aleatoria de 1000 personas de una ciudad, 400 votan a un determinado partido político. Calcule un intervalo de confianza al 96 % para la proporción de votantes de ese partido en la ciudad.

46.[S/06] Se ha lanzado un dado 400 veces y se ha obtenido 80 veces el valor cinco.

Estime, mediante un intervalo de confianza al 95 %, el valor de la probabilidad de obtener un cinco.

47.[S/06]

a) Sea la población {1, 5, 7}. Escriba todas las muestras de tamaño 2, mediante muestreo aleatorio simple, y calcule la varianza de las medias muestrales.

b) De una población de 300 hombres y 200 mujeres se desea seleccionar, mediante muestreo aleatorio estratificado con afijación proporcional, una muestra de tamaño 30 distribuida en los dos estratos, ¿cuál será la composición de la muestra?

48. [S/06] Se han tomado las tallas de 16 bebés, elegidos al azar, de entre los nacidos en un cierto hospital, y se han obtenido los siguientes resultados, en centímetros:

51, 50, 53, 48, 49, 50, 51, 48, 50, 51, 50, 47, 51, 51, 49, 51

La talla de los bebés sigue una ley Normal de desviación típica 2 centímetros y media desconocida.

a) ¿Cuál es la distribución de las medias de las muestras de tamaño 16?

b) Determine un intervalo de confianza, al 97 %, para la media poblacional.

49. [S/06] Un fabricante produce tabletas de chocolate cuyo peso en gramos sigue una ley Normal de media 125 g y desviación típica 4 g.

a) Si las tabletas se empaquetan en lotes de 25, ¿cuál es la probabilidad de que el peso medio de las tabletas de un lote se encuentre entre 124 y 126 gramos?

b) Si los lotes fuesen de 64 tabletas, ¿cuál sería la probabilidad de que el peso medio de las tabletas del lote superase los 124 gramos?

50. [S/06] Una variable aleatoria sigue una ley Normal con media desconocida y desviación típica 2'4. Se quiere estimar la media poblacional, con un nivel de confianza del 93%, para lo que se toman dos muestras de distintos tamaños.

a) Si una de las muestras tiene tamaño 16 y su media es 10'3, ¿cuál es el intervalo de confianza correspondiente?

b) Si con la otra muestra el intervalo de confianza es (9'776, 11'224), ¿cuál es la media muestral? ¿Cuál es el tamaño de la muestra?

51. [S/06] De 500 encuestados en una población, 350 se mostraron favorables a la retransmisión de debates televisivos en tiempos de elecciones.

Calcule un intervalo de confianza, al 99'5 %, para la proporción de personas favorables a estas retransmisiones.

Ejercicios de la Ponencia para proporciones

52. Tomada una muestra de 300 personas mayores de edad en una gran ciudad, se obtuvo que 105 habían votado a un determinado partido X.

Halle, con un nivel de confianza del 90%, un intervalo de confianza que permita estimar la proporción de votantes del partido X en la ciudad.

Solución: $I=(3'305, 0'395)$

53. Se estima, mediante una muestra aleatoria, con un nivel de confianza del 96%, la proporción de hogares con conexión a Internet. Se obtiene una proporción estimada del 28% con un error máximo de 6%.

¿Cuál es el tamaño de la muestra?

Solución: $n=235$

54. Para estimar la proporción de consumidores que prefieren un determinado producto, se ha tomado una muestra al azar de 1075 consumidores, entre los que se ha encontrado a 516 que lo prefieren.

Determine una cota del error para la estimación de esa proporción a un nivel de confianza del 95%

Solución: $E=0'031$

55. Mediante una muestra aleatoria de tamaño 400 se estima la proporción de residentes en una ciudad que tienen intención de asistir a una exposición. Si para un nivel de confianza del 95% resulta un error máximo en la estimación del 3%, obtenga el valor de la estimación sabiendo que es inferior a 0'25.

Solución: $\tilde{p}=0'105=10'5\%$

56. Se estima la proporción de varones adultos, residentes en una población, con obesidad severa mediante una muestra aleatoria de tamaño 500. Se obtiene una estimación de la proporción del 18%, con un error máximo del 4%.

¿Con qué nivel de confianza se ha realizado dicha estimación?

Solución: $p=1-\alpha=0'98$

Cuestiones

1. Se realiza un estudio sobre la pobreza extrema en Andalucía y para facilitar la recogida de datos se considera la formación de una muestra usando la guía telefónica.
Comenta esta consideración.

2. En un país se decide formar un nuevo órgano para la toma de decisiones: la Asamblea de Municipios. Estará formada por todos los alcaldes electos. Todas aquellas decisiones que competan al ámbito municipal se tomarán por mayoría simple en la que cada uno de sus componentes tiene derecho a un voto.

¿Qué te parece esta idea?

3. En ese mismo país se decide formar un nuevo Censo Electoral: sólo podrán votar los mayores de 18 años de ese país que voluntariamente se inscriban en el "Censo del País", pudiendo hacerlo en el municipio en que residan. De esta forma sólo se tiene en cuenta la opinión sólo de aquellos que realmente estén interesados en votar.

¿Cuál es la población y la muestra? ¿Qué te parece esta idea?

4. En una fábrica de bombillas se realiza un estudio de calidad. Se toma una muestra de 100 bombillas y se dejan encendidas para estimar la duración media.

a) ¿Cuál es la población y cuál es la muestra?

b) ¿Por qué se recurre a una muestra en este caso?

5. En las Elecciones Generales se eligen los diputados y senadores que nos representarán en las Cámaras. Explica cuál es la población y cuál la muestra en este *estudio* de la opinión de los ciudadanos.

Un grupo de economistas propone que en las próximas elecciones generales sólo vote una parte convenientemente elegida del censo electoral. Con ello se conseguiría un importante ahorro y una gran facilidad en el recuento. ¿Qué opinión te merece esta idea?

6. Una determinada característica de cierta población sigue una distribución $N(10,2)$. Si formamos muestras de tamaño n = 20, ¿ cuál es la distribución de las medias muestrales?

7. El 30% de una determinada población presenta la característica C. Si formamos muestras de tamaño n = 50, ¿cuál es la distribución de las proporciones muestrales?

8. Consideremos una distribución $N(10,2)$. Si se toma una muestra de tamaño n = 16, obtén el intervalo característico para la distribución de las medias muestrales correspondiente a la probabilidad $p = 0'5$.

¿Qué significado tiene ese intervalo?

9. En una población se tiene para determinada característica p = 0'8. Si se toma una muestra de tamaño n = 50, obtén el intervalo característico para la distribución de las proporciones muestrales correspondiente a la probabilidad $p = 0'90$.

¿Qué significado tiene ese intervalo?

10. Tras extraer una muestra en una determinada población, obtenemos $(24'5, 28)$ como intervalo de confianza para estimar la media de una población . ¿Cuál es el valor de la media muestral?

11. Para estimar una proporción poblacional extraemos una muestra y obtenemos $(0'34, 0'46)$ como intervalo de confianza. ¿Cuál es el valor de la proporción muestral?

12. Para estimar la media poblacional hemos obtenido mediante una muestra de tamaño n = 30, con el nivel de confianza del 90%, el intervalo $(20, 30)$
Verdadero o falso:

a) La probabilidad de que la media poblacional se halle en ese intervalo es $p = 0'90$.

b) La media muestral es 25.

c) La media poblacional es 25, con un nivel de confianza del 90%.

d) Para el 90% de las muestras con dicho tamaño (n=30) el intervalo anterior contiene a la media poblacional.

e) Para el 90% de las muestras con dicho tamaño (n=30) el intervalo correspondiente contiene a la media poblacional.

13. Para estimar una proporción poblacional hemos tomado una muestra con tamaño $n = 30$. El intervalo $(0'4 , 0'5)$ es el obtenido para estimarla con un nivel de confianza del 95%. Verdadero o falso:

a) La proporción muestral es 0'45.

b) La proporción poblacional es 25.

c) La proporción poblacional se halla en ese intervalo con una probabilidad $1-\alpha=0'95$.

d) La probabilidad de que la proporción poblacional sea 0'45 es 0'95.

e) El intervalo anterior contiene, con una confianza del 95%, la proporción buscada.

f) Para el 95% de las muestras con dicho tamaño (n=30) el intervalo anterior contiene a la media poblacional.

Autoevaluación

1. Una biblioteca pública está organizada en cinco secciones, de modo que los libros existentes en cada una viene dada por:

Secc. 1	***Secc. 2***	***Secc. 3***	***Secc. 4***	***Secc. 5***
500	860	1.200	700	740

Con el objeto de estimar el porcentaje de libros de edición española, se quiere seleccionar una muestra del 5% del número total de libros, a través de muestreo estratificado aleatorio, considerando cada sección como un estrato. Determina el número de libros que habría que seleccionar en cada sección si:

a) Consideramos afijación igual.

b) Consideramos afijación proporcional.

En este último caso, explica detalladamente cómo se elegirían los libros de la sección 1.

2. Se supone que la estatura de los chicos de 18 años de cierta población sigue una normal de media 162 cm y una desviación típica 12 cm. Se toma una muestra al azar de 100 de esos chicos encuestados y se calcula su media.

¿Cuál es la probabilidad de que esta media esté entre 159 y 165 cm?

3. La fábrica de automóviles "Centauro" produjo el año pasado 120.000 unidades del modelo Σ . Se detectaron defectos en 4.800 de ellos.

En un concesionario de la marca, se vendieron 150 automóviles de ese modelo. Calcula la probabilidad de que este concesionario vendiera más de diez vehículos "Centauro Σ " que presentaran algún defecto.

4. Se desea estudiar el gasto semanal en fotocopias, medido en pesetas, de los estudiantes de bachillerato del Instituto. Para ello, se ha elegido una muestra aleatoria de nueve de éstos, obteniendo:

100 , 150 , 90 , 70 , 75 , 105 , 200 , 120 , 80

Se supone que la variable aleatoria objeto del estudio sigue una distribución normal de media desconocida y desviación típica igual a 12.

Determínese un intervalo de confianza al 95% para la media del gasto semanal en fotocopias por estudiante.

5. Se sabe que el contenido de fructosa de cierto alimento sigue una distribución normal cuya varianza es conocida, teniendo el valor de 0'25. Se desea estimar el valor de la media poblacional mediante el valor de la media de una muestra, admitiéndose un error máximo de 0'2 con una confianza del 95%.

¿Cuál ha de ser el tamaño de la muestra?

6. Tomada al azar una muestra de 1000 aparatos eléctricos, se encontró que 8 de ellos presentaban un defecto de fabricación.

Obtengamos un intervalo para estimar la proporción de aparatos defectuosos, con un nivel de confianza del 90%, en la producción.

Autoevaluación

1. En primer lugar, sumamos todos los volúmenes de las diferentes secciones para obtener el total de libros: hay 4.000 ejemplares.

El tamaño de la muestra será $n=4\,000\cdot 0'05=200$

a) Se nos indica que debemos tomar el mismo número en todas las secciones. Serán 200:5 = 40 libros en cada sección

b) Ahora debemos tomar en cada estrato un número proporcional al de su tamaño en relación con la población. En nuestro caso, el 5% de cada sección: 25, 43, 60, 35, 37 libros respectivamente.

En la sección 1 numeramos los volúmenes: desde 1 a al 500. A continuación calcularemos $Ent(500\cdot aleatorio+1)$ hasta obtener 25 números diferentes. Esos números indican los libros que debemos tomar para la muestra.

2. La distribución de las medias muestrales obtenida a partir de una población normal, de media μ y desviación típica σ, es también normal. Concretamente, si el tamaño de las muestras es n, la distribución es $N(\mu,\sigma/\sqrt{n})$.

Aquí: $\mu=162$, $\sigma=12$ y $n=100$

Así: $\bar{X}$ es $N(162,1'2)$.

Tenemos así:

$$p(159<\bar{X}<165)=p(-2'5<Z<1'5)=0'9876$$

3. La proporción de vehículos con defecto en la población (total fabricados) es:

$$p=4\,800:120\,000=0'04$$

En una muestra con n = 150 la proporción muestral debe cumplir:

$$\hat{p}>\frac{10}{150}=\frac{1}{15}$$

Como $np=150\cdot 0'04\geq 5$ y $nq=150\cdot 0'96\geq 5$:

$\hat{P}$ es normal con $\left|\begin{array}{l}\mu=p=0{,}04\\ \sigma=\sqrt{\dfrac{pq}{n}}=0{,}016\end{array}\right.$

Así:

$$p(\hat{p}>0'06)=p\left(z>\frac{0'06-0'04}{0'016}\right)=$$
$$=p(z>1'6)=1-\phi(1'67)=1-0'9525=0'0475$$

4. Anotemos todo:

n = 9 es el tamaño muestral

$\bar{x}=110$	es la media muestral
$\sigma=12$	es la desviación típica poblacional
$1-\alpha=0'95$	es el nivel de confianza
$z_{\alpha/2}=1'96$	es el correspondiente valor crítico

Realizando todos los cálculos y redondeando:

$$I=\left(\bar{x}-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\,,\,\bar{x}+z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)=(102\,,118)$$

5. Los datos del problema son:

n	es el tamaño muestral (¿?)
$\sigma^2=0'25$	es la varianza poblacional
$1-\alpha=0'95$	es el nivel de confianza
$z_{\alpha/2}=1'96$	es el valor crítico

Como E no debe ser superior a 0'2:

$$E=z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}=0'2\ \rightarrow\ n=\left(\frac{z_{\alpha/2}\cdot\sigma}{0'2}\right)^2=24'01$$

El tamaño de la muestra debe ser superior a 24.

6. Característica: C = “el aparato es defectuoso”

Tamaño muestral: n = 1000

Proporción muestral: $\tilde{p}=0{,}008\ \rightarrow\ \tilde{q}=0{,}992$

Nivel de confianza: $p=1-\alpha=0{,}90$

Valor crítico: $z_{\alpha/2}=1{,}645$

El intervalo de confianza es:

$$I=\left(\tilde{p}-z_{\alpha/2}\cdot\sqrt{\frac{\tilde{p}\cdot\tilde{q}}{n}}\ ,\ \tilde{p}+z_{\alpha/2}\cdot\sqrt{\frac{\tilde{p}\cdot\tilde{q}}{n}}\right)\approx(0{,}003\,,0{,}013)$$