TEMA 10

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 de confianza para la media.
7. 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 el 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: intervalos de confianza (bien para la media poblacional bien para la proporción poblacional), 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

$$c = \frac{N}{n}$$

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

  De esta forma quedan:

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

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:

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

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:

$n_1 = E(500 \cdot 0.75 + 1) = E(376) = 376$

$n_2 = E(500 \cdot 0.264 + 1) = E(123.5) = 132$

$n_3 = E(500 \cdot 0.052) = E(27) = 27$

$n_4 = E(500 \cdot 0.21 + 1) = E(106) = 106$

$n_5 = E(500 \cdot 0.835 + 1) = E(418.5) = 418$

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:

$$x_0 = E(77 \cdot 0.242 + 1) = 19$$

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:

$$\mu = \frac{\sum x_i}{N} = \frac{62}{30} = 3.1$$

$$\sigma^2 = \frac{\sum {x_i}^2}{N} - \mu^2 = \frac{230}{20} - 3.1^2 = 1.89 \to \sigma = \sqrt{1.89} = 1.3747\ldots$$

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 }

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

$$\mu = \frac{\sum x_i}{N} = \frac{27}{10} = 2.7$$

$$s^2 = \frac{\sum {x_i}^2}{N} - \bar{x}^2 = \frac{89}{10} - 2.7^2 = 1.61 \to \sigma = \sqrt{1.61} = 1.2688\ldots$$

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 proporción 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 defectuosas que hay en la población:

$$p = \frac{10}{20} = 0.5$$

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

M = { *b* , *b* , *d* , *d* , *b* , *d* , *b* , *b* , *d* , *d* }

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

$$\widetilde{p} = \frac{5}{10} = 0.5$$

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.

3. Si ahora formamos otra muestra ***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.

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.

Comenzamos 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. Tomemos, como ejemplo:

Observemos que es **X** una variable aletoria discreta en una población de tamaño *N* = 4 que toma esos valores

X = { 1 , 2 , 3 , 4 }

Vamos primero a calcular los parámetros μ y σ en la población, luego formaremos todas las muestras de tamaño $n = 2$ (con reemplazamiento) obteniendo la distribución $\overline{\mathbf{X}}$ de las medias muestrales, después calcularemos la media y la desviación típica de la distribución de las medias muestrales y, finalmente, compararemos ambos con los de la población.

Paso 1: La media y la desviación típica poblaciones son, respectivamente:

$$\mu = \frac{\sum x_i}{N} = \frac{10}{4} = 2.5$$

$$\sigma^2 = \frac{\sum {x_i}^2}{N} - \mu^2 = \frac{30}{4} - 2.5^2 = 1.125 \rightarrow \sigma = \sqrt{1.125}$$

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":

$$\overline{\mathbf{X}} = \{1\,, 1.5\,, 2\,, 2.5\,, 1.5\,, 2\,, 2.5\,, 3\,, 2\,, 2.5\,, 3\,, 3.5\,, 2.5\,, 3\,, 3.5\,, 4\}$$

Paso 3: Hallemos la media y la desviación típica de $\overline{\mathbf{X}}$.

$$\overline{\mu} = \frac{\sum \bar{x}_i}{16} = \frac{40}{16} = 2.5$$

$$\overline{\sigma}^2 = \frac{\sum \bar{x}_i^2}{16} - \overline{\mu}^2 = \frac{110}{16} - 2.5^2 = 0.625 \rightarrow \overline{\sigma} = \sqrt{0.625}$$

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: $\bar{\sigma}$

Paso 4: Comparemos ambas.

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

$$\overline{\mu} = \mu$$

En cuanto a las desviaciones, no coinciden; pero se cumple esta relación:

$$\overline{\sigma} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{1.125}}{\sqrt{2}} = \sqrt{0.625}$$

En general, concluimos que:

Sea **X** una variable aleatoria discreta en una población finita de tamaño N, y sea $\overline{\mathbf{X}}$ la distribución de las medias muestrales, con reemplazamiento, de tamaño n. Entonces los parámetros de ésta son:

$$\overline{\mu} = \mu \quad , \quad \overline{\sigma} = \frac{\sigma}{\sqrt{n}}$$

Cuando tomamos muestras sin reemplazamiento es:

Sea **X** una variable aleatoria discreta en una población finita de tamaño N, y sea $\overline{\mathbf{X}}$ la distribución de las medias muestrales, sin reemplazamiento, de tamaño n. Entonces los parámetros de ésta son:

$$\overline{\mu} = \mu \quad , \quad \overline{\sigma} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

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

❑ Distribuciones normales.

Podemos extrapolar lo anterior a las distribuciones normales:

Si es **X** una variable aleatoria normal, de parámetros poblacionales μ y σ, la distribución $\overline{\mathbf{X}}$ de las medias muestrales de tamaño n es también normal y sus parámetros son:

$$\overline{\mu} = \mu \quad , \quad \overline{\sigma} = \frac{\sigma}{\sqrt{n}}$$

☞ 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. Consideremos todas las posibles muestras de 25 estudiantes.

¿Cómo es la distribución de las medias muestrales? ¿Qué probabilidad hay de que la estatura media en una muestra supere los 180 centímetros?

Primero indicamos cómo es la distribución de las medias muestrales:

La distribución de las medias muestrales $\overline{\mathbf{X}}$ es normal con $\begin{vmatrix} \overline{\mu} = \mu = 175 \\ \overline{\sigma} = \dfrac{15}{\sqrt{25}} = 3 \end{vmatrix}$

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

Podemos calcular la probabilidad pedida normalizando y usando la tabla:

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

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

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

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

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

Veamos:

$$\mathrm{p}\left[X < 190\right] = \mathrm{p}\left[z < -0.5\right] = 1 - \mathrm{p}\left[z > 0.5\right] = 0.3085$$

$[*]\ z = \dfrac{x-\mu}{\sigma} = \dfrac{190-200}{20} = -0.5$

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

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

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

Calculemos, pues:

$$\mathrm{p}\left[\overline{X} < 190\right] = \mathrm{p}\left[z < -1.5\right] = \mathrm{p}\left[z > 1.5\right] = 1 - 0.9332 = 0.0668$$

$[*]\ z = \dfrac{\bar{x}-\bar{\mu}}{\bar{\sigma}} = \dfrac{190-200}{20/3} = -1.5$

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 ($n > 30$):

> Si es $\mathbf{X}$ una variable aleatoria, de parámetros poblacionales μ y σ, la distribución $\overline{\mathbf{X}}$ de las medias muestrales de tamaño $n > 30$ es casi normal y sus parámetros son:
>
> $$\overline{\mu} = \mu \quad , \quad \overline{\sigma} = \frac{\sigma}{\sqrt{n}}$$

☞ 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:

$$\bar{x} < \frac{3500}{36} \approx 97.22 \text{ gr}$$

Como es $n = 36 > 30$ suficientemente grande, entonces:

La distribución de las medias muestrales $\overline{\mathbf{X}}$ es normal con $\left|\begin{array}{l} \overline{\mu} = \mu = 105 \\ \overline{\sigma} = \dfrac{10}{\sqrt{36}} = \dfrac{5}{3} \end{array}\right.$

Calculamos la probabilidad pedida tipificando y usando la tabla:

$$\mathrm{p}\left[\overline{X} < 97.22\right] = \mathrm{p}\left[z < -1.67\right] = \mathrm{p}\left[z > 1.67\right] = 1 - 0.9525 = 0.0475$$

$[*]\ z = \dfrac{\bar{x}-\bar{\mu}}{\bar{\sigma}} = \dfrac{97.22-100}{5/3} \approx -1.67$

5.Distribución de 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 $\widetilde{p}$ con que aparece **C** .

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

> Sea **C** una característica que se presenta en una población con una proporción p. Designemos $q = 1 - p$.
>
> Consideremos todas las muestras de tamaño n de forma que sea $n\,p > 5$ y $n\,q > 5$. Entonces la distribución $\widetilde{P}$ de las proporciones muestrales, es normal con parámetros
>
> $$\mu = p \quad , \quad \sigma = \sqrt{\frac{p \cdot q}{n}}$$

☞ 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 de 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:

$$p = \frac{1}{5} = 0.20 \;\rightarrow\; q = 1 - p = 0.80$$

Queremos hallar $\mathrm{p}\left[\widetilde{P} > 0.25\right]$ para un tamaño muestral $n = 50$.

Como n es suficientemente grande, pues $n\,p > 5$ y $n\,q > 5$, la distribución de las proporciones muestrales $\widetilde{P}$ cumple:

$$\widetilde{P} \text{ es normal con } \left| \begin{array}{l} \mu = p = 0.20 \\ \sigma = \sqrt{\dfrac{p \cdot q}{n}} = \sqrt{0.0032} \end{array} \right.$$

Así que la probabilidad pedida es:

$$\mathrm{p}\left[\widetilde{P} > 0.25\right] = \mathrm{p}\left[z > 0.88\right] = 1 - 0.8106$$

$$[*] \; z = \frac{\widetilde{p} - \mu}{\sigma} \approx 0.88$$

6.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 $\overline{\mathbf{X}}$ es normal con $\begin{cases} \overline{\mu} = \mu \\ \overline{\sigma} = \dfrac{\sigma}{\sqrt{n}} \end{cases}$

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

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

Donde $z_{\alpha/2}$ es el valor crítico bilateral asociado. Observemos que por ello, para el $p \cdot 100\%$ de las muestras la media muestral $\overline{x}$ está en dicho intervalo

Si para una muestra está $\overline{x}$ en ese intervalo, entonces μ está en el intervalo

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

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

Consideremos en una población una v.a. **X** de media desconocida y una desviación típica σ.

Consideremos una muestra de tamaño n en la que obtenemos una media muestral $\overline{x}$.

El intervalo de confianza para la media poblacional, correspondiente a un nivel de confianza $p = 1 - \alpha$ es:

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

☞ 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 para μ si se fija como coeficiente de confianza 0.95.

Organicemos todo:

La variable aleatoria **X** sigue una distribución normal con $\begin{cases} \mu = ? \\ \sigma = 3 \end{cases}$

Tamaño muestral: $n = 100$

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

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

El intervalo de confianza es:

$$I = \left(9 - 1,96 \cdot \frac{3}{\sqrt{100}}\,,9 + 1,96 \cdot \frac{3}{\sqrt{100}}\right) = (8,412\,,9,588)$$

Esas condiciones son:
- X es normal.
- El tamaño muestral es n > 30

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

Fijado un nivel de confianza, si tomamos una muestra concreta y a partir de ella formamos un nivel de confianza, la media poblacional puede o no estar en él.

Ahora bien, si **X** es normal o si, en cualquier caso, es n > 30 la media poblacional estará en el intervalo de confianza asociado al p·100% de las muestras elegidas.

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

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

<u>Observaciones:</u>

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 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.

❑ Error.

Al formar un intervalo de confianza, se llama error máximo admisible a

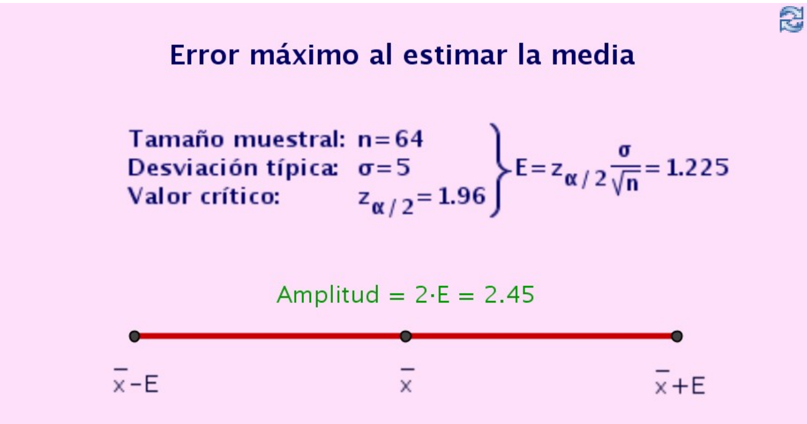
$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Observemos que así el intervalo de confianza es

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

donde el centro es la media muestral $\overline{x}$ y la longitud o amplitud es $L = 2\,E$

He aquí un ejemplo:



☞ <u>Ejemplo</u>: Las tallas de los alumnos de una Universidad se distribuyen normalmente con una desviación típica de 6 cm. Hallemos el tamaño muestral necesario para estimar la talla media con un nivel de confianza del 95% y un error máximo no superior a 1.2 cm.

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 : $\overline{x} = 6$

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

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

Despejamos de la fórmula del error:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \to 1.2 = 1.96 \cdot \frac{6}{\sqrt{n}} \to \sqrt{n} = 1.96 \cdot \frac{6}{1.2} \to n = 96$$

7.Intervalos de confianza proporciones

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 $\widetilde{p}$.

Para estimar p, con un nivel de confianza $(1-\alpha)\cdot 100\%$, se utiliza el intervalo de confianza

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

donde $\widetilde{q} = 1-\widetilde{p}$. Y se llama error máximo admisible a:

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

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: $\widetilde{p} = \dfrac{140}{400} = 0.35 \to \widetilde{q} = 0.65$

Nivel de confianza: $1-\alpha = 0.90$ $\to$ Valor crítico: $z_{\alpha/2} = 1.645$

El intervalo de confianza es:

$$I = \left(\widetilde{p} - z_{\alpha/2}\cdot\sqrt{\frac{\widetilde{p}\cdot\widetilde{q}}{n}}\,,\widetilde{p} + z_{\alpha/2}\cdot\sqrt{\frac{\widetilde{p}\cdot\widetilde{q}}{n}}\right) = (0.31\,,0.39)$$

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/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?

2. [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.

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

4. [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.

5. [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?

6. [S/06] Se han tomado las tallas de 16 neonatos, elegidos al azar en 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 cm 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.

7. [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?

8. [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?

9. [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.

10. [S/08] Se desea estimar la proporción de individuos zurdos en una determinada ciudad. Para ello se toma una muestra aleatoria de 300 individuos resultando que 45 de ellos son zurdos.

a) Calcule, usando un nivel de confianza del 97%, el correspondiente intervalo de confianza para la proporción de individuos zurdos de la población.

b) ¿Sería mayor o menor el error de estimación si se usara un nivel de confianza del 95%?

11.[S/08] Una variable aleatoria sigue una ley Normal con desviación típica 6. ¿De qué tamaño, como mínimo, se debe elegir una muestra que nos permita estimar la media de esa variable con un error máximo de 2 y una confianza del 99%?

12.[S/08] 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.

a) Halle un intervalo de confianza, al 97%, para la longitud media de los cables.

b) Determine el tamaño mínimo que debe tener una muestra de estos auriculares para que el error de estimación de la longitud media sea inferior a 1 cm, con el mismo nivel de confianza.

13.[S/08] Se ha aplicado un medicamento a una muestra de 200 enfermos y se ha observado una respuesta positiva en 140 de ellos. Estímese, mediante un intervalo de confianza del 99%, la proporción de enfermos que responderían positivamente si este medicamento se aplicase a la población de la que se ha extraído la muestra.

14.[S/08] El número de días de permanencia de los enfermos en un hospital sigue una ley Normal de media μ días y desviación típica 3 días.

a) Determine un intervalo de confianza para estimar μ, a un nivel del 97 %, con una muestra aleatoria de 100 enfermos cuya media es 8.1 días.

b) ¿Qué tamaño mínimo debe tener una muestra para poder estimar μ con un error máximo de 1 día y un nivel de confianza del 92%?

15.[S/08] Sea la población { 1 , 2 , 3 , 4 }.

a) Construya todas las muestras posibles de tamaño 2, mediante muestreo aleatorio simple.

b) Calcule la varianza de las medias muestrales.

16.[S/08] Tomada al azar una muestra de 90 alumnos de un Instituto se encontró que un tercio habla inglés.

Halle, con un nivel de confianza del 97%, un intervalo de confianza para estimar la proporción de alumnos de ese Instituto que habla inglés.

17.[S/08] El peso, en kg, de los alumnos de primaria de un colegio sigue una distribución Normal de media 28 kg y desviación típica 2.7 kg.

Consideremos muestras aleatorias de 9 alumnos.

a) ¿Qué distribución sigue la media de las muestras?

b) Si elegimos, al azar, una de esas muestras, ¿cuál es la probabilidad de que su media esté comprendida entre 26 y 29 kg?

18.[S/08] En un centro de anillamiento de aves se ha detectado que en una muestra de 250 ejemplares de una especie, 60 son portadoras de una bacteria.

Obtenga un intervalo de confianza, al 97%, para la proporción de aves de esa especie que son portadoras de la bacteria.

19.[S/08] En una muestra representativa de 1200 residentes de una ciudad, 450 utilizan habitualmente el transporte público. Obtenga el intervalo de confianza, al 90%, de la proporción de residentes en la ciudad que utilizan habitualmente el transporte público.

20.[S/08] El consumo, en gramos, de un cierto producto sigue una ley Normal con varianza 225 g^2 .

a) A partir de una muestra de tamaño 25 se ha obtenido una media muestral igual a 175 g. Halle un intervalo de confianza, al 90%, para la media del consumo.

b) ¿Cuál debe ser el tamaño mínimo de la muestra para que el correspondiente intervalo de confianza, al 95%, tenga una amplitud máxima de 5?

21.[S/08] El tiempo de utilización diaria de ordenador entre los empleados de una empresa sigue una distribución Normal de media μ y desviación típica 1.2 horas.

a) Una muestra aleatoria de 40 empleados tiene una media del tiempo de utilización de 2.85 horas diarias. Determine un intervalo de confianza, al 96%, para la media del tiempo de utilización diaria de ordenador.

b) Calcule el tamaño mínimo que debería tener una muestra para estimar la media del tiempo de utilización diaria del ordenador con un error no superior a 0.75 horas y el mismo nivel de confianza del apartado anterior.

22.[S/11]

a) Una población de tamaño 1000 se ha dividido en 4 estratos de tamaño 150, 400, 250 y 200. Utilizando muestreo aleatorio estratificado con afijación proporcional se han seleccionado 10 individuos del tercer estrato, ¿cuál es el tamaño de la muestra?

b) El peso de los individuos de una población se distribuye según una ley Normal de desviación típica 6 kg. Calcule el tamaño mínimo de la muestra para estimar, con un nivel de confianza del 95%, el peso medio en la población con un error no superior a 1 kg.

23.[S/11] El peso neto de las tabletas de chocolate de una determinada marca es una variable aleatoria Normal con media μ y desviación típica 7 gramos. Se sabe que 36 tabletas, elegidas al azar, han dado un peso total de 5274 gramos.

a) Calcule un intervalo con un nivel de confianza del 94% para la media μ .

b) Con el mismo nivel de confianza, ¿cuántas tabletas, como mínimo, habrá que tomar como muestra para que la amplitud del intervalo que se obtenga sea, como máximo, de 3 gramos?

24.[S/11] La estatura de las personas de una población es una variable aleatoria que sigue una distribución Normal cuya desviación típica es de 0.04 m. Para estimar la media de esta variable se ha tomado una muestra aleatoria de 60 personas de esa población y se ha encontrado una estatura media de 1.73 m.

a) Obtenga un intervalo de confianza, con un nivel del 97%, para la media de la distribución de estaturas.

b) Halle el tamaño mínimo que debe tener una muestra de esta población, para que la amplitud de un intervalo de la media con este nivel de confianza sea inferior a 0.08 m.

25.[S/11] Sea X una variable aleatoria Normal de media 50 y desviación típica 4. Se toman muestras de tamaño 16.

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

b) ¿Cuál es la probabilidad de que la media muestral esté comprendida entre 47.5 y 52.5?

26.[S/11] En un distrito universitario, la calificación de los alumnos sigue una distribución Normal de media 6.2 puntos y desviación típica de 1 punto.

Se selecciona, al azar, una muestra de tamaño 25.

a) Indique la distribución de la media de las muestras de tamaño 25.

b) ¿Cuál es la probabilidad de que la media de las calificaciones de los alumnos de una de esas muestras esté comprendida entre 6 y 6.6 puntos?

27.[S/11] El peso de los adultos de una determinada población sigue una distribución Normal de media 70 kg y desviación típica 16 kg. Si elegimos, al azar, muestras de tamaño 4:

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

b) ¿Cuál es la probabilidad de que el peso medio de una de esas muestras esté comprendido entre 65 y 72 kg?

c) ¿Cuál es la probabilidad de que ese peso medio sea menor que 70 kg?

28.[S/11] Con el fin de estudiar el peso medio de los perros recién nacidos de una determinada raza, se tomó una muestra en una clínica veterinaria y se obtuvieron los siguientes pesos, medidos en kg:

1.2 , 0.9 , 1 , 1.2 , 1.1 , 1 , 0.8 , 1.1

Se sabe que el peso de los cachorros de esta raza se distribuye según una ley Normal con desviación típica 0.25 kg.

a) Obtenga un intervalo de confianza para estimar la media poblacional, al 95%.

b) Halle el error máximo que se cometería usando el intervalo anterior.

c) Razone cómo variaría la amplitud del intervalo de confianza si, manteniendo el mismo nivel de confianza, aumentásemos el tamaño de la muestra.

29.[S/19]

a) Una población de 25 000 personas se ha dividido en cuatro estratos con tamaños 15000, 5000, 3000 y 2000 personas respectivamente. En esa población se ha realizado un muestreo estratificado con afijación proporcional, en el que se han elegido al azar 36 personas del tercer estrato. Determine el tamaño de la muestra total obtenida con este muestreo y su composición.

b) Dada la población P = {2, 4, 6}, construya todas las muestras posibles de tamaño 2 mediante muestreo aleatorio simple y halle la desviación típica de las medias muestrales.

30.[S/19] Se ha tomado una muestra de 16 pacientes tratados por un especialista y se ha observado que el tiempo de espera en su consulta, en minutos, ha sido de:

8 9.2 10 8.5 12 9 11.3 7 8.5 8.3 7.6 9 9.4 10.5 8.9 6.8

Supongamos que el tiempo de espera en esta consulta se distribuye según una ley Normal de varianza 4 y media desconocida.

a) [1.5] Halle un intervalo de confianza al 97.5% para estimar el tiempo medio de espera de los pacientes tratados por este especialista.

a) [1] ¿Cuál debería ser el tamaño mínimo de la muestra para asegurar, con un nivel de confianza del 90%, que el error cometido sea, a lo sumo, de 0.3 minutos?

31.[S/19]El precio de venta al público del kilogramo de frambuesas sigue una ley Normal de media desconocida y varianza 9. En una localidad se eligen 10 comercios de manera aleatoria, obteniéndose los siguientes precios en euros:

12.3 10 9.1 11 10.5 11.8 9.9 11.5 10.9 13

a) ¿Qué distribución siguen las medias de las muestras de tamaño 10?

b) Con los datos obtenidos de la muestra, determine un intervalo de confianza al 97 % para el precio medio del kilogramo de frambuesas.

c) Con el mismo nivel de confianza, calcule el tamaño mínimo que debe tener una muestra para que el error cometido al estimar el precio medio del kilogramo de frambuesas sea menor a 1.5 €.

32.[S/19]Se sabe que la longitud, en centímetros, de una especie de estrella de mar en una determinada zona sigue una ley Normal con desviación típica 3. Para estimar la longitud media de esa especie de estrella de mar, se extrae una muestra de tamaño 36 y se obtiene el intervalo de confianza (6.04, 8) al 95 %. Se pide:

a) Calcule la media muestral.

b) Calcule el error de estimación máximo cometido.

c) Si aumentamos el tamaño muestral a 49, ¿qué efecto produce sobre el error máximo cometido? Calcule este error.

d) Si aumentamos el nivel de confianza, ¿qué efecto produce sobre el error de estimación máximo?

33.[S/19] La vida útil, en años, de las lavadoras de un determinado modelo, se distribuye según una ley Normal de varianza 7.84. En una muestra de 12 lavadoras, la vida útil en años ha sido:

9.5 9 10.2 8.6 11.4 10.8 12.6 11 11.8 14.5 10.4 9.8

a) Con estos datos, determine un intervalo de confianza al 93.5% para estimar la vida útil media de estas lavadoras.

b) Calcule el error máximo que se puede cometer al estimar la vida útil media de este modelo de lavadoras, si se toma una muestra de 50 lavadoras y asumimos un nivel de confianza del 99%.

34.[S/19] La renta anual de los hogares andaluces, en miles de euros, se distribuye según una ley Normal con desviación típica 5 y media desconocida μ.

a) Si se desea que en el 99 % de las posibles muestras del mismo tamaño, elegidas de entre los hogares andaluces, la media muestral no difiera de la renta media anual poblacional de dichos hogares en más de una unidad, ¿cuál debe ser el tamaño mínimo de las muestras?

b) Si se consideran muestras de hogares andaluces de tamaño 100, ¿qué distribución de probabilidad sigue la variable aleatoria "Renta media anual muestral"?

c) Suponiendo que la renta media anual poblacional de los hogares andaluces es $\mu = 24$, ¿cuál es la probabilidad de que en una muestra de tamaño 100 la renta media anual muestral sea superior a 25?

35.[S/19]La distancia en kilómetros recorrida al día por los vehículos de una empresa de coches de alquiler sigue una distribución Normal de media desconocida y varianza 225. Se toma una muestra aleatoria simple de 36 coches y se obtiene el intervalo de confianza $(153.65\,,162.35)$ para la media poblacional.

a) Calcule la media muestral y el error máximo de estimación para ese intervalo de confianza.

b) Si con el mismo nivel de confianza, aumentamos el tamaño muestral, ¿cómo se vería afectado el error?

c) Con un nivel de confianza del 95%, ¿cuál debería ser el tamaño mínimo de la muestra para que el error cometido sea inferior a 3 km?

36.[S/19]El tiempo de espera para ser atendido en un servicio hospitalario es una variable aleatoria que sigue una distribución Normal con desviación típica de 2 meses. Tomada una muestra al azar de 9 pacientes que han utilizado ese servicio, se han registrado los siguientes tiempos de espera en meses:

8.5 3.7 4.3 3.6 5.6 4.8 1.0 1.4 6.0

a) Determine un intervalo de confianza al 95% para el tiempo de espera medio poblacional.

b) Con un nivel de confianza del 97%, ¿qué tamaño muestral mínimo se ha de tomar para que el error máximo cometido en la estimación del tiempo de espera medio poblacional no exceda de un mes?

37.[S/19]El tiempo de desfase, en minutos, entre la hora de paso programada de un autobús por cierta parada y la hora real a la que pasa, sigue una distribución Normal de media desconocida y varianza 4. Se observa el paso del autobús por la parada en 10 ocasiones elegidas al azar, registrándose los siguientes desfases:

4.7 2.1 3.6 5.4 0.0 4.2 4.0 -0.2 1.9 5.2

a) Obtenga un intervalo de confianza al 97% para el desfase medio en la hora de paso del autobús.

b) ¿Qué tamaño muestral mínimo sería necesario para estimar el desfase medio con un error inferior a 30 segundos y un nivel de confianza del 95%? ¿Cómo variaría dicho tamaño muestral si se aumentara el nivel de confianza?

38.[S/19] Una tienda de ropa quiere estudiar la aceptación de un nuevo sistema de pago a través del teléfono móvil. Para ello realiza una encuesta entre 200 de sus clientes elegidos al azar, resultando que 150 de ellos sí estarían dispuestos a usar el nuevo sistema de pago.

a) Determine un intervalo de confianza al 97% para estimar la proporción de clientes de esa tienda que estarían dispuestos a usar el nuevo sistema de pago.

b) Mediante una nueva encuesta se quiere estimar la proporción de clientes de esa tienda que usarían el nuevo sistema de pago, con un error máximo del 3% y un nivel de confianza del 94%. Suponiendo que se mantiene la proporción muestral del apartado anterior, ¿a cuántos clientes como mínimo habría que realizar la encuesta?

39.[S/19] Tomada al azar una muestra de 600 alumnos de una universidad española, se encontró que 2/3 de los mismos podían expresarse en inglés con fluidez.

a) [1.5] Calcule un intervalo de confianza al 98% para estimar la proporción de alumnos de esa universidad que pueden expresarse en inglés con fluidez. ¿Se podría admitir a ese nivel de confianza que la proporción de alumnos de esa universidad que pueden expresarse en inglés con fluidez es 13/20?

b) [0.25] Teniendo en cuenta el intervalo anterior, ¿qué error máximo se cometería en dicha estimación?

c) [0.75] Si se mantienen la misma proporción muestral y la misma confianza, ¿cuántos alumnos como mínimo habría de tener una muestra para que el error de estimación sea inferior al 2%?

40.[S/19] La cantidad de café por taza que suministra una máquina de café sigue una distribución Normal con media desconocida y desviación típica 0.8 cm^3. En una muestra de 45 tazas suministradas por esa máquina, se ha medido un total de 5400 cm^3 de café.

a) Calcule el estimador puntual para la cantidad media de café por taza que suministra la máquina.

b) Calcule un intervalo de confianza al 97% para estimar la cantidad media de café por taza que suministra la máquina.

c) Calcule, con el mismo nivel de confianza, el tamaño muestral mínimo que se ha de tomar para que, al estimar la cantidad media de café por taza, el error cometido sea inferior a 0.2 cm^3 .

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 las redes sociales.

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 opinas sobre esto?

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

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

8. Consideremos una distribución $\mathcal{N}(10\,,2)$. 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 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 el intervalo anterior contiene a la media poblacional.

e) El 90% de las muestras con $n = 30$, el intervalo que se calcule contiene a la media poblacional.

12. 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?

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 0.45.

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.

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. 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.

4. Se sabe que el contenido de fructosa de cierto alimento sigue una distribución normal cuya varianza es 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?

5. 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 $n = 4000 \cdot 0.05 = 200$.

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 $E\,(500 \cdot aleatorio + 1)$ 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 $\mathcal{N}\left(\mu\,,\sigma/\sqrt{n}\right)$.

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

Así: $\overline{\mathbf{X}} \sim \mathcal{N}\,(162\,, 1.2)$

Tenemos entonces:

$$\mathrm{p}\left[159 < \overline{X} < 165\right] = \mathrm{p}\left[-2.5 < z < 1.5\right] = 0.9876$$

3. 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:

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

4. Los datos del problema son:

$n = ?$ es el tamaño muestral

$\sigma^2 = 0.25$ es la varianzapoblacional

$1 - \alpha = 0.95$ es el nivel de confianza

$z_{\alpha/2} = 1.96$ es el correspondiente valor crítico

$E = 0.2$ es el error máximo

Así:

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2 = 24.01$$

El tamaño de la muestra debe ser superior a 24.

5. Característica: C = "el aparato es defectuoso"

Tamaño muestral: $n = 1000$

Proporción muestral: $\widetilde{p} = 0.08 \rightarrow \widetilde{q} = 0.992$

Nivel de confianza: $1 - \alpha = 0.90$

Valor crítico: $z_{\alpha/2} = 1.645$

El intervalo de confianza es:

$$\left(\widetilde{p} - z_{\alpha/2} \cdot \sqrt{\frac{\widetilde{p} \cdot \widetilde{q}}{n}}\,, \widetilde{p} + z_{\alpha/2} \cdot \sqrt{\frac{\widetilde{p} \cdot \widetilde{q}}{n}}\right) = (0.003\,, 0, 013)$$

Ejercicios ponencia

Proporciones

1. 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: $I=(3.305, 0.395)$

2. 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: $n=235$

3. 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: $E=0.031$

4. 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: $\tilde{p}=0.105$

5. 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: $1-\alpha=0.98$