UAX_Texto(01).dot

Estadística Empresarial

5. Estimación Paramétrica

DOCUMENTO1

ÍNDICE

MOTIVACIÓN

Cuando se estudia una población de la que se conoce su distribución de probabilidad excepto un parámetro o parámetros desconocidos, resulta fundamental asignar valores aproximados a esos parámetros poblacionales para poder, posteriormente, calcular probabilidades de interés para el análisis que se esté realizando y las medidas que caracterizan a la población. Al estudio de dicho proceso de asignación se dedicará esta Unidad Didáctica.

Dicha asignación se realiza a partir de la información muestral. Por tanto, esta Unidad Didáctica se basa en la Unidad Didáctica 4.

PROPÓSITOS

Los principales propósitos de esta Unidad Didáctica son:

- Entender el concepto de estimación.
- Distinguir entre estimación puntual y por intervalos.
- Comprender el concepto de estimador puntual y sus propiedades.
- Aprender los métodos de estimación puntual.
- Saber estimar por intervalos un parámetro.

PREPARACIÓN PARA LA UNIDAD

Se denomina estimación al proceso de asignación de valores a los parámetros desconocidos de una población a partir de la información muestral. Se distinguen dos tipos de estimación:

- Estimación puntual: se asigna a cada parámetro desconocido un cierto valor.
- Estimación por intervalos: se construyen intervalos en los que con un cierto grado de confianza se encuentra el verdadero valor del parámetro.

La primera parte de esta unidad didáctica se dedica a la estimación puntual. En primer lugar, se presenta el concepto de estimador. A continuación, se describen las propiedades de los estimadores. Finalmente, se presentan los métodos de estimación puntual.

La segunda parte se dedica a la estimación por intervalos. En primer lugar, se explica el método para generar intervalos de confianza y a continuación, se presentan algunos intervalos de confianza muy utilizados.

En esta Unidad Didáctica aprenderás a estimar un parámetro de forma puntual y por intervalos.

1. ESTIMACIÓN PUNTUAL

Se esta estudiando una población de la que se conoce su distribución de probabilidad excepto una serie de parámetros poblacionales que son desconocidos. Se toma una muestra y a partir de la información muestral se tratan de asignar valores aproximados a esos parámetros desconocidos.

Se denomina **estimación** al proceso de asignación de valores a los parámetros desconocidos a partir de la información muestral.

Se distinguen dos tipos de estimación:

1. **Estimación puntual**: se asigna a cada parámetro desconocido un cierto valor.
2. **Estimación por intervalos**: se construyen intervalos en los que con un cierto grado de confianza se encuentra el verdadero valor del parámetro.

En esta sección y en la siguiente se estudiará la estimación puntual.

En primer lugar, se va a definir el concepto de estimador puntual y de estimación. Se prestará especial atención a la diferencia que existe entre ambos conceptos. A continuación, se estudiarán las propiedades de los estimadores: insesgadez, eficiencia y consistencia. Finalmente, en la siguiente sección, se expondrán los métodos de estimación puntual.

CONCEPTO DE ESTIMADOR Y ESTIMACIÓN

Se considera una población de la que se conoce su distribución de probabilidad excepto un parámetro cuyo valor se desconoce, y que se representa de forma genérica por la letra θ.

Estimador puntual: un estimador puntual de un parámetro θ, que se representa por $\hat{\theta}$, es un estadístico que utilizamos para asignar valores al parámetro.

$$\hat{\theta} = T(X_1, X_2, ..., X_n)$$

Estimación: es el valor del estimador para una muestra concreta.

$$\hat{\theta} = T(x_1, x_2, ..., x_n)$$

Ejemplo 1. La esperanza de una población es desconocida. Se emplea la media muestral para asignarle un valor aproximado a la esperanza poblacional. Se toma una muestra aleatoria simple y se obtiene una media muestral de 25. Distinguir entre estimador y estimación.

El estimador es la media muestral, puesto que es el estadístico que se utiliza para asignar valores al parámetro.

La estimación es 25, porque es el valor que toma el estadístico en la muestra que se ha tomado.

Puede haber situaciones en las que existan diferentes estimadores de un parámetro. Por ejemplo, si la varianza de una población es desconocida, se podría utilizar como estimador la varianza muestral o la cuasivarianza muestral. En estos casos se plantea el siguiente problema: ¿Qué estimador se elegirá?

Se elegirá el que tenga mejores propiedades.

A continuación, se presentan las propiedades de los estimadores.

PROPIEDADES DE LOS ESTIMADORES

INSESGADEZ

Un estimador es insesgado si la esperanza del estimador coincide con el parámetro a estimar.

$$E(\hat{\theta}) = \theta$$

Sesgo: diferencia entre la esperanza del estimador y el parámetro a estimar.

$$sesgo = E(\hat{\theta}) - \theta$$

Obsérvese que si un estimador es insesgado la esperanza del estimador coincide con el parámetro a estimar. Por tanto, el sesgo es igual a cero.

Ejemplo 2. Una población viene representada por una variable aleatoria ξ. Se toma una muestra aleatoria simple de tamaño n. Se considera la media muestral como estimador de la esperanza poblacional. ¿Es un estimador insesgado?

Si su respuesta es negativa, determine el sesgo.

Se considera la media muestral como estimador de la esperanza poblacional.

$$\hat{\mu} = \overline{X}$$

Para comprobar si el estimador es insesgado, se determinará la esperanza del estimador.

$$E(\hat{\mu}) = E(\overline{X}) = \mu$$

Se ha tenido en cuenta que la esperanza de la media muestral es la media poblacional (véase la sección 2 de la Unidad Didáctica 4).

La esperanza del estimador coincide con el parámetro a estimar. Por tanto, el estimador es insesgado.

Conclusión: la media muestral es un estimador insesgado de la esperanza poblacional.

EFICIENCIA

Un estimador es eficiente si:

1. Es insesgado. $E(\hat{\theta}) = \theta$
2. Tiene mínima varianza.

Para determinar si la varianza de un estimador es mínima, se utiliza la cota de Cramer Rao. La cota es el valor mínimo que puede tener la varianza de un estimador insesgado.

COTA DE CRAMER RAO

Sea $\hat{\theta}$ un estimador insesgado de un parámetro θ entonces se verifica que:

$$V(\hat{\theta}) \geq \frac{1}{n\, E\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2}$$

Por tanto, un estimador es eficiente si:

1. Es insesgado. $E(\hat{\theta}) = \theta$
2. Su varianza coincide con la cota de Cramer Rao.

$$V(\hat{\theta}) = \frac{1}{n\, E\left[\frac{\partial \ln f(x,\theta)}{\partial \theta}\right]^2}$$

Ejemplo 3. Una población viene representada por una variable aleatoria ξ que sigue una distribución de Poisson. Se toma una muestra aleatoria simple de tamaño n. Se considera como estimador del parámetro λ la media muestral. ¿Es eficiente?

Tenga en cuenta que: Cota de Cramer Rao $= \dfrac{\lambda}{n}$

La población sigue una distribución de Poisson. Por tanto $E(\xi) = \mu = \lambda$ $V(\xi) = \sigma^2 = \lambda$ (véase la sección 4 de la Unidad Didáctica 3).

El estimador es: $\hat{\lambda} = \overline{X}$

A continuación, se analizarán las dos condiciones que debe cumplir el estimador para ser eficiente.

En primer lugar, se determinará la esperanza del estimador.

$$E(\hat{\lambda}) = E(\overline{X}) = \mu = \lambda$$

Para calcular la esperanza del estimador se ha tenido en cuenta que:

1. La esperanza de la media muestral es igual a la esperanza poblacional (véase la sección 2 de la Unidad Didáctica 4).
2. La esperanza poblacional es igual al parámetro λ, porque la población sigue una distribución de Poisson.

La esperanza del estimador coincide con el parámetro a estimar. Por tanto, el estimador es insesgado.

En segundo lugar, se determinará la varianza del estimador.

$$V(\hat{\lambda}) = V(\overline{X}) = \frac{\sigma^2}{n} = \frac{\lambda}{n}$$

Para calcular la varianza del estimador se ha tenido en cuenta que:

1. La varianza de la media muestral es igual a la varianza poblacional dividida entre el tamaño de la muestra (véase la sección 2 de la Unidad Didáctica 4).

2. La varianza poblacional es igual al parámetro λ, porque la población sigue una distribución de Poisson.

La varianza coincide con la Cota de Cramer Rao. Por consiguiente, es mínima.

El estimador verifica las dos condiciones: es insesgado y tiene mínima varianza. Por tanto, es eficiente.

CONSISTENCIA

Un estimador es consistente si se verifica que:

$$p\left[\left|\hat{\theta}-\theta\right|<\varepsilon\right]\to 1$$
$$n\to\infty$$

La diferencia entre el valor asignado al parámetro, que se representa por $\hat{\theta}$, y el valor verdadero, que se representa por θ, es el error cometido. El valor de ese error no se conoce, puesto que el verdadero valor del parámetro es desconocido.

Es importante que el error cometido sea pequeño, pero da igual que sea positivo (que el valor asignado esté por encima del valor verdadero o negativo (que el valor asignado esté por debajo del valor verdadero). Por ello se considera el valor absoluto del error.

Si el estimador es consistente se verifica que: cuando el tamaño muestral tiende a infinito, es decir, cuando la muestra es muy grande, se tiene casi certeza (la probabilidad tiende a 1) de que el error que se está cometiendo es infinitesimal.

Condición suficiente de consistencia

Si un estimador:

1. Es insesgado. $E(\hat{\theta})=\theta$

2. La varianza del estimador tiende a 0 cuando el tamaño muestral tiende a ∞.

$$V(\hat{\theta})\to 0$$
$$n\to\infty$$

entonces es consistente.

Ejemplo 4. Una población viene representada por una variable aleatoria ξ. Se toma una muestra aleatoria simple de tamaño n. Se considera la media muestral como estimador de la esperanza poblacional. ¿Es un estimador consistente?

Se considera la media muestral como estimador de la esperanza poblacional.

$$\hat{\mu} = \overline{X}$$

1. $E(\hat{\mu}) = E(\overline{X}) = \mu$.

   El estimador es insesgado (véase el ejemplo 2 de esta Unidad Didáctica).

2. $V(\hat{\mu}) = V\left(\overline{X}\right) = \dfrac{\sigma^2}{n}$

   Para calcular la varianza del estimador se ha tenido en cuenta que la varianza de la media muestral es igual a la varianza poblacional dividida entre el tamaño de la muestra (véase la sección 2 de la Unidad Didáctica 4).

A continuación, se analizará que ocurre cuando el tamaño muestral tiende a ∞.

$$V(\hat{\mu}) = V\left(\overline{X}\right) = \frac{\sigma^2}{n} \to 0$$
$$n \to \infty$$

Por tanto, el estimador es consistente.

Conclusión: la media muestral es un estimador consistente de la esperanza poblacional.

2. MÉTODOS DE ESTIMACIÓN PUNTUAL

Los métodos de estimación puntual son métodos que proporcionan estimadores puntuales de los parámetros. Se distinguen dos métodos:

- METODO DE MÁXIMA VEROSIMILITUD: consiste en considerar como estimación el valor más probable dada la muestra que se ha observado.
- METODO DE LOS MOMENTOS: se basa en un criterio de analogía. Consiste en igualar los momentos con respecto al origen de la población a los momentos con respecto al origen de la muestra.

A continuación se explicarán estos dos métodos.

MÉTODO DE MÁXIMA VEROSIMILITUD

En primer lugar, se definirá la función de verosimilitud de la muestra. Esta función es la función de probabilidad de la muestra, pero considerando que los parámetros desconocidos son las variables y las observaciones muestrales son conocidas.

Consideramos que el muestreo es aleatorio simple. Como se vio en la sección 1 de la Unidad Didáctica 4, este muestreo se caracteriza porque los elementos muestrales se seleccionan al azar y con reemplazamiento. Estas características tienen dos implicaciones:

1. Cada elemento muestral se distribuye como la población. Por tanto, la función de masa o de densidad de cada elemento muestral coinciden con las de la población.

2. Las observaciones muestrales son independientes entre sí, por que es un muestreo con reemplazamiento.

Si la variable es discreta, la función de verosimilitud es el producto de las funciones de masa de cada elemento muestral, porque los elementos muestrales son independientes entre sí.

$$L(x_1, x_2, ..., x_{n,} \theta) = p(x_1, \theta)\ p(x_2, \theta)\ ...\ p(x_n, \theta) = \prod_{i=1}^{n} p(x_i, \theta)$$

El operador $\prod$ representa un producto.

Si la variable es continua, la función de verosimilitud es el producto de las funciones de densidad de cada elemento muestral, porque los elementos muestrales son independientes entre sí.

$$L(x_1, x_2, ..., x_{n,} \theta) = f(x_1, \theta)\ f(x_2, \theta)\ ...\ f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

Estimadores máximo-verosímiles

Los estimadores máximo-verosímiles son los valores de los parámetros que maximizan la función de verosimilitud de la muestra, es decir, los que hacen máxima la probabilidad de obtener la muestra que se ha observado.

Se trata, por tanto, de encontrar el valor del parámetro que maximiza la función de verosimilitud de la muestra.

En lugar de maximizar la función de verosimilitud, maximizamos su logaritmo neperiano para simplificar el cálculo.

Esto se puede hacer porque la función logarítmica es una función creciente y, por tanto, los valores que maximizan la función de verosimilitud son los mismos que los que maximizan su logaritmo neperiano.

Ejemplo 5. Una población viene representada por una variable aleatoria ξ que sigue una distribución de Poisson. Se toma una muestra aleatoria simple de tamaño n. Determinar el estimador máximo verosímil del parámetro λ.

En primer lugar, construimos la función de verosimilitud de la muestra.

$$L(x_1, x_2,..., x_{n,}\lambda) = p(x_1,\lambda)\ p(x_2,\lambda)\ ...\ \ p(x_n,\lambda) = \prod_{i=1}^{n} p(x_i,\lambda) =$$

$$= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}...\frac{e^{-\lambda}\lambda^{x_n}}{x_n!} = \frac{e^{-\lambda n}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Para construir dicha función se ha tenido en cuenta que el muestreo es aleatorio simple (véase la sección 1 de la Unidad Didáctica 4) y por tanto:

1. Cada elemento muestral sigue una distribución de Poisson como la población. Su función de masa es (véase la sección 4 de la Unidad Didáctica 3):

$$p(x_i,\lambda) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \qquad i = 1,...,n$$

2. Las observaciones muestrales son independientes entre sí. Por consiguiente, la función de probabilidad de la muestra es el producto de las probabilidades de cada elemento muestral.

A continuación, se determina el logaritmo neperiano de la función de verosimilitud

$$\ln L(x_1, x_2,..., x_{n,}\lambda) = \ln\left(\frac{e^{-\lambda n}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}\right) = \ln\left(e^{-\lambda n}\lambda^{\sum_{i=1}^{n} x_i}\right) - \ln\left(\prod_{i=1}^{n} x_i!\right) =$$

$$= \ln\left(e^{-\lambda n}\right) + \ln\left(\lambda^{\sum_{i=1}^{n} x_i}\right) - \ln\left(\prod_{i=1}^{n} x_i!\right) = -\lambda n + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - \ln\left(\prod_{i=1}^{n} x_i!\right)$$

Por último, se encuentra el valor máximo.

$$Max \qquad \ln L(x_1, x_2,..., x_{n,}\lambda) = -\lambda\ n + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - -\ln\left(\prod_{i=1}^{n} x_i!\right)$$

Para ello, se iguala la derivada de la función de verosimilitud con respecto al parámetro a 0.

$$\frac{\partial \ln L}{\partial \lambda} = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda} = 0$$

Despejando el parámetro, se tiene que:

$$\hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

Por tanto, el estimador por el método de máxima verosimilitud del parámetro λ es la media muestral.

METODO DE LOS MOMENTOS

Consiste en igualar los momentos con respecto al origen de la población a los momentos con respecto al origen de la muestra.

Distinguiremos dos casos:

a) Hay un parámetro desconocido.

Como se tiene una sola incógnita, se necesita plantear una ecuación.

Se iguala el momento de orden 1 con respecto al origen de la población, que se representará por $\alpha_1(\theta)$ al momento de orden 1 con respecto al origen de la muestra, que se representará por a_1.

$$\alpha_1(\theta) = a_1$$

siendo:

$\alpha_1 = E(\xi) = \sum_{i=1}^{n} x_i \; p(x_i, \theta)$ si la variable es discreta o

$\alpha_1 = E(\xi) = \int_{-\infty}^{+\infty} x \; f(x, \theta) \; dx$ si la variable es continua.

$$a_1 = \frac{\sum_{i=1}^{n} x_i}{n}$$

Se resuelve la ecuación despejando el valor de θ y la solución es el estimador.

b) Hay dos parámetros desconocidos.

Como hay dos incógnitas, se plantea un sistema de dos ecuaciones.

- Se iguala el momento de orden 1 con respecto al origen de la población, que se representará por α_1, al momento de orden 1 con respecto al origen de la muestra, que se representará por a_1.

$$\alpha_1(\theta_1, \theta_2) = a_1$$

siendo:

$\alpha_1 = E(\xi) = \sum_{i=1}^{n} x_i \; p(x_i, \theta_1, \theta_2)$ si la variable es discreta o

$\alpha_1 = E(\xi) = \int_{-\infty}^{+\infty} x \; f(x, \theta_1, \theta_2) \; dx$ si la variable es continua.

$$a_1 = \frac{\sum_{i=1}^{n} x_i}{n}$$

- Se iguala el momento de orden 2 con respecto al origen de la población, que se representará por α_2, al momento de orden 2 con respecto al origen de la muestra, que se representará por a_2.

$$\alpha_2(\theta_1, \theta_2) = a_2$$

$\alpha_2 = \sum_{i=1}^{n} x_i^2 \; p(x_i, \theta_1, \theta_2)$ si la variable es discreta o $\alpha_2 = \int_{-\infty}^{+\infty} x^2 \; f(x, \theta_1, \theta_2) \; dx$ si la variable es continua.

$$a_2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

Se resuelve el sistema, y las soluciones son los estimadores de los parámetros por el método de los momentos.

Ejemplo 6. Una población viene representada por una variable aleatoria ξ que sigue una distribución de Poisson. Se toma una muestra aleatoria simple de tamaño n. Determinar el estimador por el método de los momentos del parámetro λ.

Se tiene un solo parámetro desconocido. Por tanto, se necesita plantear una sola ecuación.

Se iguala el momento de orden 1 con respecto al origen de la población al momento de orden 1 con respecto al origen de la muestra.

Como la población sigue una distribución de Poisson.

$$\alpha_1 = E(\xi) = \lambda$$

a_1 es siempre la media de la muestra.

Se igualan y se despeja el parámetro (en este caso ya está despejado).

$$\alpha_1 = a_1 \qquad \lambda = \overline{x} \qquad \hat{\lambda} = \overline{x}$$

Por tanto, el estimador por el método de los momentos del parámetro λ es la media muestral.

3. PLANTEAMIENTO GENERAL DE LA ESTIMACIÓN POR INTERVALOS

Se está analizando una población de la que se desconocen una serie de parámetros poblacionales. Como se vio en la sección 1, se denomina estimación al proceso de asignación de valores a los parámetros poblacionales desconocidos a partir de la información muestral. Se distinguen dos tipos de estimación:

1. **Estimación puntual:** se asigna a cada parámetro desconocido un cierto valor.
2. **Estimación por intervalos:** se construyen intervalos en los que con un cierto grado de confianza se encuentra el verdadero valor del parámetro.

La estimación puntual, que se ha desarrollado en las secciones anteriores, es más precisa. Lógicamente el grado de precisión es mayor si asignamos al parámetro un valor que si le asignamos un rango de valores, pero también más arriesgada.

Si se quiere estimar por intervalos un parámetro desconocido de una población, que se representa de forma genérica por la letra θ, se sigue el planteamiento que se presenta a continuación.

Se trata de encontrar dos estadísticos, es decir, dos funciones de la muestra, que se representarán por $T_1(X)$ y $T_2(X)$ respectivamente, tales que con una probabilidad γ el verdadero valor del parámetro $\theta \in [T_1(X), T_2(X)]$, es decir,

$$P[T_1(X) \leq \theta \leq T_2(X)] = \gamma = 1 - \alpha$$

γ se denomina grado de confianza y α nivel de significación.

El **grado de confianza** es la probabilidad de que el verdadero valor del parámetro se encuentre en el intervalo y el **nivel de significación** es la probabilidad de que el verdadero valor del parámetro no se encuentre en el intervalo.

Cuando particularizamos para una muestra obtenemos un intervalo numérico.

4. INTERVALOS DE CONFIANZA PARA LA ESPERANZA DE UNA POBLACIÓN NORMAL

Se considera una población que sigue una distribución normal siendo la esperanza poblacional desconocida. Se quiere estimar por intervalos dicho parámetro. Para deducir el intervalo nos basaremos en la distribución de probabilidad del estadístico media muestral.

Se distinguen dos casos dependiendo de si la varianza poblacional es conocida o desconocida.

INTERVALO DE CONFIANZA PARA LA ESPERANZA DE UNA POBLACIÓN NORMAL CON VARIANZA CONOCIDA

Una población sigue una distribución normal con varianza conocida. Se toma una muestra aleatoria simple de tamaño n. La construcción del intervalo de confianza para μ está basada en la distribución de probabilidad del estadístico media muestral que aparece en la Unidad Didáctica 4 (sección 3) y que se presenta a continuación.

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \cong N(0,1)$$

El estadístico sigue una distribución normal estándar.

Para deducir el intervalo se parte de la siguiente expresión:

$$P\left\{\left|\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \le z_{\alpha/2}\right\} = \gamma = 1-\alpha$$

En primer lugar, se elimina el valor absoluto de la desigualdad.

$$P\left\{-z_{\alpha/2} \le \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le z_{\alpha/2}\right\} = \gamma = 1-\alpha$$

A continuación, se despeja el parámetro μ .

$$P\left\{-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \overline{X}-\mu \le z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = \gamma = 1-\alpha$$

$$P\left\{-\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le -\mu \le -\overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = \gamma = 1-\alpha$$

Seguidamente, se multiplicamos la desigualdad por -1.

$$P\left\{\overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \ge \mu \ge \overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = \gamma = 1-\alpha$$

Por último, se cambia el sentido de la desigualdad.

$$P\left\{\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = \gamma = 1-\alpha$$

Por tanto, el intervalo de confianza para la esperanza de una población normal con varianza conocida es:

$$\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

siendo $z_{\alpha/2}$ es el valor que deja a la derecha un área de $\alpha/2$ debajo de la función de densidad de la distribución normal estándar.

Ejemplo 7. Una población sigue una distribución normal con varianza igual a 100. Se toma una muestra aleatoria simple de 144 observaciones obteniéndose a partir de dicha muestra una media muestral de 160. Construir un intervalo de confianza del 95% para la esperanza poblacional.

La población es normal y la varianza poblacional es conocida. Por tanto, el intervalo de confianza para la esperanza poblacional es:

$$\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

En primer lugar se determinará el radio del intervalo.

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

El grado de confianza fijado es del 95%. Por tanto, el nivel de significación será del 5%.

$$\gamma = 0{,}95 \quad \alpha = 1 - \gamma = 1 - 0{,}95 = 0{,}05 \quad \alpha/2 = 0{,}025$$

$z_{\alpha/2}$ es el valor que deja a la derecha un área de 0,025 debajo de la función de densidad de una distribución normal estándar.

Dicho valor se determina a partir de la tabla de la distribución normal estándar que se presenta en el anexo de la Unidad Didáctica 3.

Se busca dentro de la tabla 0,025. Se mira a que valor de la primera columna y de la primera fila corresponde. Dicha área corresponde a 1,9 y 0,06. Por tanto, $z_{\alpha/2} = 1{,}96$.

Sustituyendo se tiene que:

$$\varepsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 1{,}96 \frac{10}{\sqrt{144}} = 1{,}6\widehat{3}$$

El límite inferior del intervalo es la media muestral menos el radio y el límite superior la media muestral más el radio.

$$\overline{X} - \varepsilon \leq \mu \leq \overline{X} + \varepsilon$$

Sustituyendo por los valores obtenidos se tiene que:

$$160 - 1{,}6\widehat{3} \leq \mu \leq 160 + 1{,}6\widehat{3}$$

El intervalo de confianza del 95% para la esperanza poblacional es:

$$158{,}3\widehat{6} \leq \mu \leq 161{,}6\widehat{3}$$

INTERVALO DE CONFIANZA PARA LA ESPERANZA DE UNA POBLACIÓN NORMAL CON VARIANZA DESCONOCIDA

Una población sigue una distribución normal con varianza desconocida. Se toma una muestra aleatoria simple de tamaño n. La construcción del intervalo de confianza para μ está basado en la distribución de probabilidad del estadístico media muestral que se presento en la Unidad Didáctica 4 (sección 3). Se distinguen dos casos dependiendo del tamaño de la muestra.

a) Si el tamaño muestral es menor o igual a 30, la distribución de probabilidad del estadístico media muestral, viene dada por la siguiente expresión.

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \cong t_{n-1}$$

siendo t_{n-1} una distribución t de Student de $n-1$ grados de libertad.

Para deducir el intervalo se parte de la siguiente expresión:

$$P\left\{\left|\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\right| \leq t_{\alpha/2}\right\} = \gamma = 1-\alpha$$

En primer lugar, se elimina el valor absoluto de la desigualdad.

$$P\left\{-t_{\alpha/2} \leq \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \leq t_{\alpha/2}\right\} = \gamma = 1-\alpha$$

A continuación, se despeja el parámetro μ.

$$P\left\{-t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \overline{X}-\mu \leq t_{\alpha/2}\frac{S}{\sqrt{n}}\right\} = \gamma = 1-\alpha$$

$$P\left\{-\overline{X}-t_{\alpha/2}\frac{S}{\sqrt{n}} \leq -\mu \leq -\overline{X}+t_{\alpha/2}\frac{S}{\sqrt{n}}\right\} = \gamma = 1-\alpha$$

Seguidamente, se multiplicamos la desigualdad por -1.

$$P\left\{\overline{X}+t_{\alpha/2}\frac{S}{\sqrt{n}} \geq \mu \geq \overline{X}-t_{\alpha/2}\frac{S}{\sqrt{n}}\right\} = \gamma = 1-\alpha$$

Por último, se cambia el sentido de la desigualdad.

$$P\left\{\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right\} = \gamma = 1 - \alpha$$

Por tanto, el intervalo de confianza para la esperanza de una población normal con varianza desconocida y tamaño muestral menor o igual que 30 es:

$$\overline{X} - t_{\alpha/2, n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2, n-1}\frac{S}{\sqrt{n}}$$

siendo $t_{\alpha/2, n-1}$ es el valor que deja a la derecha un área de $\alpha / 2$ debajo de la función de densidad de una distribución t de Student de $n-1$ grados de libertad.

Ejemplo 8. Una población sigue una distribución normal. Se toma una muestra aleatoria simple de 20 observaciones obteniéndose a partir de dicha muestra una cuasidesviación típica muestral de 6,156. Construir un intervalo de confianza del 95% para la esperanza poblacional.

La población es normal, la desviación típica poblacional es desconocida y el tamaño muestral es menor que 30. Por tanto, el intervalo de confianza para la esperanza poblacional es:

$$\overline{X} - t_{\alpha/2, n-1}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2, n-1}\frac{S}{\sqrt{n}}$$

En primer lugar se determinará el radio del intervalo:

$$\varepsilon = t_{\alpha/2, n-1}\frac{S}{\sqrt{n}}$$

El grado de confianza fijado es del 95%. Por tanto, el nivel de significación será del 5%.

$$\gamma = 0{,}95 \quad \alpha = 1 - \gamma = 1 - 0{,}95 = 0{,}05 \quad \alpha/2 = 0{,}025$$

$t_{\alpha/2, n-1}$ es el valor que deja a la derecha un área de 0,025 debajo de la función de densidad de una distribución t de Student de 19 grados de libertad.

Dicho valor se determina a partir de la tabla de la distribución t de Student que se presenta en el anexo de la Unidad Didáctica 3.

Se busca 19 (grados de libertad de la distribución) en la primera columna y 0,025 (área a la derecha) en la primera fila. El valor situado dentro de la tabla que corresponde a la fila en la que está situado 19 y a la columna en la que está situado 0,025 es el que deja a la derecha esa área. Dicho valor es 2,093.

$$t_{\alpha/2, n-1} = 2{,}093$$

Sustituyendo se tiene que:

$$\varepsilon = t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} = 2{,}093 \frac{6{,}156}{\sqrt{20}} = 2{,}88$$

El límite inferior del intervalo es la media muestral menos el radio y el límite superior la media muestral más el radio.

$$\overline{X} - \varepsilon \leq \mu \leq \overline{X} + \varepsilon$$

Sustituyendo por los valores obtenidos se tiene que:

$$170 - 2{,}88 \leq \mu \leq 170 + 2{,}88$$

El intervalo de confianza del 95% para la esperanza poblacional es:

$$167{,}12 \leq \mu \leq 172{,}88$$

b) Si el tamaño muestral es mayor que 30, la distribución de probabilidad del estadístico media muestral viene dada por la siguiente expresión.

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \cong N(0,1)$$

El estadístico sigue una distribución normal estándar.

Para deducir el intervalo se parte de la siguiente expresión:

$$P\left\{\left|\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}\right| \leq z_{\alpha/2}\right\} = \gamma = 1-\alpha$$

En primer lugar, se elimina el valor absoluto de la desigualdad.

$$P\left\{-z_{\alpha/2} \leq \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \leq z_{\alpha/2}\right\} = \gamma = 1-\alpha$$

A continuación, se despeja el parámetro μ.

$$P\left\{-z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \overline{X}-\mu \leq z_{\alpha/2}\frac{S}{\sqrt{n}}\right\} = \gamma = 1-\alpha$$

$$P\left\{-\overline{X}-z_{\alpha/2}\frac{S}{\sqrt{n}} \leq -\mu \leq -\overline{X}+z_{\alpha/2}\frac{S}{\sqrt{n}}\right\} = \gamma = 1-\alpha$$

Seguidamente, se multiplicamos la desigualdad por -1.

$$P\left\{\overline{X}+z_{\alpha/2}\frac{S}{\sqrt{n}} \geq \mu \geq \overline{X}-z_{\alpha/2}\frac{S}{\sqrt{n}}\right\} = \gamma = 1-\alpha$$

Por último, se cambia el sentido de la desigualdad.

$$P\left\{\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right\} = \gamma = 1 - \alpha$$

Por tanto, el intervalo de confianza para la esperanza de una población normal con varianza desconocida y tamaño muestral mayor que 30 es:

$$\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}$$

siendo $z_{\alpha/2}$ es el valor que deja a la derecha un área de $\alpha/2$ debajo de la función de densidad de la distribución normal estándar.

5. INTERVALOS DE CONFIANZA PARA LA ESPERANZA DE UNA POBLACIÓN NO NORMAL

Se considera una población que no sigue una distribución normal siendo la esperanza poblacional desconocida. Se quiere estimar por intervalos dicho parámetro. Se toma una muestra aleatoria simple de tamaño n mayor que 30. Para deducir el intervalo nos basaremos en la distribución aproximada del estadístico media muestral.

Se distinguen dos casos dependiendo de si la varianza poblacional es conocida o desconocida.

INTERVALO DE CONFIANZA PARA LA ESPERANZA DE UNA POBLACIÓN NO NORMAL CON VARIANZA CONOCIDA

Se considera una población con varianza conocida. Se toma una muestra aleatoria simple de tamaño n mayor que 30. La construcción del intervalo de confianza para μ está basada en la distribución aproximada del estadístico media muestral que aparece en la Unidad Didáctica 4 (sección 4) y que se presenta a continuación.

$$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0,1)$$

El estadístico sigue una distribución normal estándar aproximada.

Si se procede de forma idéntica al caso de población normal con varianza conocida, se genera el intervalo de confianza para μ que viene dado por la siguiente expresión:

$$\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Obsérvese que en este caso la distribución del estadístico es aproximada y, por tanto, el intervalo es también aproximado.

INTERVALO DE CONFIANZA PARA LA ESPERANZA DE UNA POBLACIÓN NO NORMAL CON VARIANZA DESCONOCIDA

Se considera una población con varianza desconocida. Se toma una muestra aleatoria simple de tamaño n mayor que 30. La construcción del intervalo de confianza para μ está basada en la distribución aproximada del estadístico media muestral que aparece en la Unidad Didáctica 4 (sección 4).

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \approx N(0,1)$$

El estadístico sigue una distribución normal estándar aproximada.

Si se procede de forma idéntica al caso de población normal con varianza desconocida, se genera el intervalo de confianza para μ que viene dado por la siguiente expresión:

$$\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}$$

Obsérvese que en este caso la distribución del estadístico es aproximada y, por tanto, el intervalo es también aproximado.

6. INTERVALO DE CONFIANZA PARA LA VARIANZA DE UNA POBLACIÓN NORMAL

Se considera una población que sigue una distribución normal. Se toma una muestra aleatoria simple de tamaño n. La construcción del intervalo de confianza para la varianza poblacional está basada en la distribución de probabilidad del estadístico cuasivarianza muestral que aparece en la Unidad Didáctica 4 (sección 3) y que se presenta a continuación:

$$\frac{(n-1)S^2}{\sigma^2} \cong \chi^2_{n-1}$$

siendo χ^2_{n-1} una distribución χ^2 de Pearson con n-1 grados de libertad.

A partir de la distribución de probabilidad del estadístico, se deduce el intervalo de confianza de confianza que viene dado por la siguiente expresión:

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

siendo:

$\chi^2_{\alpha/2,n-1}$ es el valor que deja a la derecha un área de $\alpha/2$ debajo de la función de densidad de una distribución χ^2 de Pearson con $n-1$ grados de libertad.

$\chi^2_{1-\alpha/2,n-1}$ es el valor que deja a la derecha un área de $1-\alpha/2$ debajo de la función de densidad de una distribución χ^2 de Pearson con $n-1$ grados de libertad.

Ejemplo 9. Una población sigue una distribución normal. Se toma una muestra aleatoria simple de 20 observaciones obteniéndose a partir de dicha muestra una cuasivarianza muestral de 2,37. Construir intervalo de confianza del 90% para la varianza de la población.

El intervalo de confianza para la varianza de una población normal viene dado por la siguiente expresión:

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

En primer lugar, se determinara.

$$(n-1)S^2 = 19 \times 2{,}37 = 45$$

A continuación, se obtendrán los valores que aparecen en los denominadores.

El grado de confianza fijado es del 90%. Por tanto, el nivel de significación será del 10%.

$$\gamma = 0{,}9 \qquad \alpha = 1-\gamma = 1-0{,}9 = 0{,}1 \qquad \alpha/2 = 0{,}05$$

$\chi^2_{\alpha/2,n-1}$ es el valor que deja a la derecha un área de 0,05 debajo de la función de densidad de una distribución χ^2 de Pearson con 19 grados de libertad.

Dicho valor se busca en la tabla de la distribución χ^2 de Pearson que se presenta en el anexo de la Unidad Didáctica 3. Se busca 19 (grados de libertad de la distribución) en la primera columna y 0,05 (área a la derecha) en la primera fila. El valor situado dentro de la tabla que corresponde a la fila en la que está situado 19 y a la columna en la que está situado 0,025 es el que deja a la derecha esa área. Dicho valor es 30,144.

$$\chi^2_{\alpha/2,n-1} = 30{,}144$$

$\chi^2_{1-\alpha/2,n-1}$ el valor que deja a la derecha un área de 0,975. Se busca 19 (grados de libertad de la distribución) en la primera columna y 0,975 (área a la derecha) en la primera fila. El valor situado dentro de la tabla que corresponde a la fila en la que está situado 19 y a la columna en la que está situado 0,975 es el que deja a la derecha esa área. Dicho valor es 10,117.

$$\chi^2_{1-\alpha/2,n-1} = 10{,}117$$

Sustituyendo por los resultados obtenidos se tiene:

$$\frac{45}{30{,}144} \leq \sigma^2 \leq \frac{45}{10{,}117}$$

El intervalo de confianza del 90% para la varianza poblacional es:

$$1{,}493 \leq \sigma^2 \leq 4{,}448$$

7. INTERVALOS DE CONFIANZA PARA LA DIFERENCIAS DE ESPERANZAS DE DOS POBLACIONES

Se consideran dos poblaciones con varianzas conocidas. Se quiere construir un intervalo de confianza para la diferencia de esperanzas poblacionales. Se distinguen dos casos dependiendo de si las poblaciones siguen distribuciones normales o no.

INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE ESPERANZAS DE DOS POBLACIONES NORMALES CON VARIANZAS CONOCIDAS

Se consideran dos poblaciones que siguen distribuciones normales con varianzas conocidas. Sean μ_1 y μ_2 las esperanzas de la primera y de la segunda población respectivamente, y σ_1^2 y σ_2^2 las varianzas. Se toman dos muestras aleatorias simples de tamaños n_1 y n_2 respectivamente. La construcción del intervalo de confianza para la diferencia de esperanzas poblacionales está basada en la distribución de probabilidad del estadístico diferencia de medias muestrales que aparece en la Unidad Didáctica 4 (sección 3) y que se presenta a continuación:

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}} \cong N(0,1)$$

El estadístico sigue una distribución normal estándar.

Para deducir el intervalo se parte de la siguiente expresión:

$$P\left\{\left|\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}}\right|\leq z_{\alpha/2}\right\}=\gamma=1-\alpha$$

En primer lugar, se elimina el valor absoluto de la desigualdad.

$$P\left\{-z_{\alpha/2}\leq\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}}\leq z_{\alpha/2}\right\}=\gamma=1-\alpha$$

A continuación, se despeja $\mu_1-\mu_2$.

$$P\left\{-z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}\leq(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)\ \leq z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}\right\}=\gamma=1-\alpha$$

$$P\left\{-(\overline{X}_1-\overline{X}_2)-z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}\leq-(\mu_1-\mu_2)\ \leq-(\overline{X}_1-\overline{X}_2)+z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}\right\}=\gamma=1-\alpha$$

Seguidamente, se multiplicamos la desigualdad por -1.

$$P\left\{(\overline{X}_1-\overline{X}_2)+z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}\geq(\mu_1-\mu_2)\geq(\overline{X}_1-\overline{X}_2)-z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}\right\}=\gamma=1-\alpha$$

Por último, se cambia el sentido de la desigualdad.

$$P\left\{(\overline{X}_1-\overline{X}_2)-z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}\leq\mu_1-\mu_2\leq(\overline{X}_1-\overline{X}_2)+z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2)}{n_2}}\right\}=\gamma=1-\alpha$$

Por tanto, el intervalo de confianza es:

$$(\overline{X}_1-\overline{X}_2)-z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}\le\mu_1-\mu_2\le(\overline{X}_1-\overline{X}_2)+z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2)}{n_2}}$$

Ejemplo 10. Una empresa A produce un artículo cuya demanda sigue una distribución normal con desviación típica igual a 200, en tanto que la demanda de otro artículo similar producido por otra empresa B sigue una distribución normal con desviación típica igual a 100.

Se han observado 125 puestos de venta de cada uno de los artículos resultando ser la demanda media muestral para el artículo de la empresa A igual a 300 y 250 para la empresa B.

Elaborar un intervalo de confianza del 95% para la diferencia de demandas medias.

ξ_1= “Demanda del artículo de la empresa A”

$\sigma_1=200$ $\qquad\overline{X}_1=300$ $\qquad n_1=125$

ξ_2= “Demanda del artículo de la empresa B”

$\sigma_2=100$ $\qquad\overline{X}_2=250$ $\qquad n_2=125$

La poblaciones son normales y las varianzas poblacionales conocidas. Por tanto, el intervalo de confianza para la diferencia de esperanzas es:

$$(\overline{X}_1-\overline{X}_2)-z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}\le\mu_1-\mu_2\le(\overline{X}_1-\overline{X}_2)+z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2)}{n_2}}$$

En primer lugar, se determinará el radio del intervalo.

$$\varepsilon=z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1}+\frac{\sigma_2{}^2}{n_2}}$$

El grado de confianza fijado es del 95%. Por tanto, el nivel de significación será del 5%.

$$\gamma = 0{,}95 \quad \alpha = 1 - \gamma = 1 - 0{,}95 = 0{,}05 \quad \alpha/2 = 0{,}025$$

$z_{\alpha/2}$ es el valor que deja a la derecha un área de 0,025 debajo de la función de densidad de una distribución normal estandar.

Dicho valor se determina a partir de la tabla de la distribución normal estándar que se presenta en el anexo de la Unidad Didáctica 3.

Se busca dentro de la tabla 0,025. Se mira a que valor de la primera columna y de la primera fila corresponde. Dicha área corresponde a 1,9 y 0,06. Por tanto, $z_{\alpha/2} = 1{,}96$.

Sustituyendo se tiene que:

$$\varepsilon = z_{\alpha/2}\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}} = 1{,}96\sqrt{\frac{200^2}{125} + \frac{100^2}{125}} = 32{,}8$$

El límite inferior del intervalo es la diferencia de medias muestrales menos el radio y el límite superior la diferencia de medias muestrales más el radio.

$$(\overline{X}_1 - \overline{X}_2) - \varepsilon \leq \mu_1 - \mu_2 \leq (\overline{X}_1 - \overline{X}_2) + \varepsilon$$

Sustituyendo por los valores obtenidos se tiene que:

$$50 - 32{,}8 \leq \mu_1 - \mu_2 \leq 50 + 32{,}8$$

El intervalo de confianza del 95% para la diferencia de esperanzas es:

$$17{,}2 \leq \mu_1 - \mu_2 \leq 82{,}8$$

INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE ESPERANZAS DE DOS POBLACIONES CON VARIANZAS CONOCIDAS

Se consideran dos poblaciones con varianzas conocidas. Sean μ_1 y μ_2 las esperanzas de la primera y de la segunda población respectivamente, y σ_1^2 y σ_2^2 las varianzas. Se toman dos muestras aleatorias simples de tamaños n_1 y n_2 respectivamente ambos mayores que 30. La construcción del intervalo de confianza para la diferencia de esperanzas poblacionales está basada en la distribución aproximada del estadístico diferencia de medias muestrales que aparece en la Unidad Didáctica 4 (sección 4) y que se presenta a continuación:

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1{}^2}{n_1} + \dfrac{\sigma_2{}^2}{n_2}}} \approx N(0,1)$$

El estadístico sigue una distribución normal estándar aproximada.

Si se procede de forma idéntica al caso de poblaciones normales con varianzas conocidas, se genera el intervalo de confianza que viene dado por la siguiente expresión:

$$(\overline{X}_1 - \overline{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{X}_1 - \overline{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2)}{n_2}}$$

Obsérvese que en este caso la distribución del estadístico es aproximada y, por tanto, el intervalo es también aproximado.

CUADROS

CUADRO 1. INTERVALOS DE CONFIANZA (POBLACIONES NORMALES)

Intervalo de confianza para la esperanza. Varianza conocida.	$\overline{X} - z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$
Intervalo de confianza para la esperanza. Varianza desconocida y n<30.	$\overline{X} - t_{\alpha/2, n-1} \dfrac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{\alpha/2, n-1} \dfrac{S}{\sqrt{n}}$
Intervalo de confianza para la esperanza. Varianza desconocida y n>30.	$\overline{X} - z_{\alpha/2} \dfrac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2} \dfrac{S}{\sqrt{n}}$
Intervalo de confianza para la varianza.	$\dfrac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} \leq \sigma^2 \leq \dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}$
Intervalo de confianza para la diferencia de esperanzas	$(\overline{X}_1 - \overline{X}_2) - z_{\alpha/2} \sqrt{\dfrac{\sigma_1{}^2}{n_1} + \dfrac{\sigma_2{}^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{X}_1 - \overline{X}_2) + z_{\alpha/2} \sqrt{\dfrac{\sigma_1{}^2}{n_1} + \dfrac{\sigma_2{}^2}{n_2}})$

CUADRO 2. INTERVALOS DE CONFIANZA APROXIMADOS (POBLACIONES NO NORMALES)

Intervalo de confianza para la esperanza. Varianza conocida.	$\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
Intervalo de confianza para la esperanza. Varianza desconocida.	$\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}$
Intervalo de confianza para la diferencia de esperanzas.	$(\overline{X}_1 - \overline{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\overline{X}_1 - \overline{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}$

CONCLUSIONES

Cuando se estudia una población de la que se conoce su distribución de probabilidad excepto un parámetro o parámetros desconocidos, resulta fundamental asignar valores aproximados a esos parámetros para poder, posteriormente, calcular probabilidades de interés para el estudio que se esté realizando y las medidas que caracterizan a la población.

En dicho proceso se deben seguir las siguientes etapas:

En primer lugar, hay que tomar una muestra. La muestra debe ser representativa de la población para que los resultados sean fiables.

A continuación, se procede a estimar el parámetro o parámetros de forma puntual y por intervalos. Para realizar la estimación puntual se elige un buen estimador para el parámetro desconocido. Se sustituyen los valores muestrales en la expresión del estimador y se obtiene el valor numérico asignado al parámetro.

Para realizar la estimación por intervalos se elige el intervalo adecuado dependiendo del parámetro y de la situación en la que nos encontremos (distribución de la población, tamaño muestral, información conocida) y se fija un grado de confianza. A continuación, se sustituyen los valores muestrales en las expresiones de los límites inferior y superior del intervalo, y se genera el intervalo numérico.

Finalmente, se interpretan los resultados obtenidos.

RECAPITULACIÓN

La estimación es el proceso de asignación de valores a los parámetros poblacionales desconocidos a partir de la información muestral.

Se distinguen dos tipos de estimación:

- **Estimación puntual**: se asigna a cada parámetro desconocido un cierto valor.

  Un estimador puntual de un parámetro es un estadístico que se utiliza para asignar valores al parámetro.

  Los métodos de estimación puntual nos proporcionan estimadores puntuales de los parámetros.

- **Estimación por intervalos**: se construyen intervalos en los que con un cierto grado de confianza se encuentra el verdadero valor del parámetro.

AUTOCOMPROBACIÓN

1. Un estimador es:

   a) Un parámetro desconocido de una población.

   b) Un estadístico que se utiliza para asignar un valor a un parámetro desconocido de una población.

   c) Un valor numérico.

   d) Ninguna respuesta es correcta.

2. Considere que la esperanza de una población es desconocida. Se emplea la media muestral para asignarle un valor aproximado a la esperanza poblacional. Se toma una muestra aleatoria simple y se obtiene una media muestral de 150, entonces:

   a) El estimador es 150 y la estimación es la media muestral.

   b) El estimador es la media muestral y la estimación es 150.

   c) Con esta información no se puede determinar cual es la estimación y cual el estimador.

   d) Ninguna respuesta es correcta.

3. Un estimador es insesgado si:

   a) La varianza del estimador coincide con el parámetro a estimar.

   b) La esperanza del estimador coincide con el parámetro a estimar.

   c) Su varianza es mínima.

   d) Ninguna respuesta es correcta.

4. La media muestral es:

   a) Un estimador insesgado de la esperanza poblacional.

   b) Un estimador sesgado de la esperanza poblacional.

   c) Con esta información no se puede determinar si es un estimador insesgado o sesgado.

   d) Ninguna respuesta es correcta.

5. Si un estimador es insesgado y su varianza tiende a cero cuando el tamaño muestral tiende a infinito entonces:

   a) El estimador no es consistente.

   b) El estimador es consistente.

   c) Con esta información no se puede determinar si el estimador es consistente o no.

   d) Ninguna respuesta es correcta.

6. La media muestral:

   a) No es un estimador consistente de la esperanza poblacional.

   b) Es un estimador consistente de la esperanza poblacional.

   c) Con esta información no se puede determinar si es un estimador consistente o no.

   d) Ninguna respuesta es correcta.

7. La Cota de Cramer Rao es :

   a) El valor mínimo para la varianza de un estimador insesgado.

   b) El valor máximo para la varianza de un estimador insesgado.

   c) El valor de la varianza de un estimador insesgado.

   d) Ninguna respuesta es correcta.

8. Un estimador es eficiente si:

   a) Es insesgado y tiene mínima varianza.

   b) La esperanza del estimador coincide con el parámetro a estimar.

   c) Su varianza es mínima.

   d) Ninguna respuesta es correcta.

9. El grado de confianza fijado para construir un intervalo de confianza es:

   a) La probabilidad de que el verdadero valor del parámetro se encuentre en el intervalo.
   b) La probabilidad de que el verdadero valor del parámetro no se encuentre en el intervalo
   c) El verdadero valor del parámetro desconocido.
   d) Ninguna respuesta es correcta.

10. Si el grado de confianza fijado para construir un intervalo de confianza es del 90 %, el nivel de significación fijado será:

   a) 1%.
   b) 10%.
   c) Con esta información no es posible determinar el nivel de significación fijado.
   d) Ninguna respuesta es correcta.

SOLUCIONARIO

1.	b	2.	b	3.	b	4.	a	5.	b
6.	b	7.	a	8.	a	9.	a	10.	b

PROPUESTAS DE AMPLIACIÓN

Si estás interesado en ampliar los conocimientos sobre estimación paramétrica, puedes consultar los capítulos 5, 6 y 7 del siguiente libro:

Martín Pliego F. J. y Ruiz Maya L. (2001). Estadística II: Inferencia. Editorial Thomson.

BIBLIOGRAFÍA

Martín Pliego F. J. y Ruiz Maya L. (2005). Fundamentos de Inferencia Estadística. Editorial Thomson.

López de la Manzanara Barbero J. (1996). Problemas de Estadística. Editorial Pirámide.