UNIDAD 3

ESTIMACIÓN DE PARÁMETROS E INTERVALOS DE CONFIANZA

Introducción a la unidad

En el momento de tomar decisiones el conocimiento de los parámetros de población es de vital importancia, tal conocimiento generalmente solo se puede tener al estimar el valor de dichos parámetros, sin embargo, la estimación es mejor cuando se da un margen de confianza y uno de error, siendo importante la correcta estimación de dichos parámetros a través de la construcción de intervalos de confianza que puedan sustentar la toma de decisiones de manera eficiente.

En el momento de tomar decisiones, es de vital importancia tener el conocimiento de los parámetros de población aunque éstos solo pueden ser estimados sus valores, sin embargo, la estimación es mejor cuando se tiene un margen de confianza y uno de error, para ello se debe tener una correcta estimación de los parámetros por medio de la construcción de intervalos de confianza que puedan sustentar la toma de decisiones de manera más eficiente.

Objetivo particular de la unidad

Analizar los conceptos fundamentales de estimación de parámetros e intervalos de confianza y los aplicará en la práctica a problemas de su área profesional laborable.

Lo que sé

Elige la respuesta correcta a las siguientes preguntas.

1. La media de la distribución de las medias de las muestras $\mu_{\bar{x}}$ siempre es…

- a) mayor que la de la población
- b) menor que la de la población
- c) igual a la de la población

2. La formula para la desviación estándar de la distribución de las medias de las muestras para una población suficientemente grande es:

- a) $\sigma^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{i=n}(x_i - x)^2$
- b) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- c) $\sigma_{\bar{x}} = \sigma$

Temas de la unidad III

1. Definición de estimador y estimación
2. Propiedades de los estimadores
3. Estimación de media, varianza y proporciones
4. Intervalo de confianza para la media y para proporciones
5. Determinación del tamaño de la muestra

Resumen de la unidad

Las inferencias acerca de una población que se obtienen del estudio de una muestra pueden ser tan buenas como lo sean las estimaciones obtenidas, aquí, el cuidado va evidentemente sobre la recolección de los datos, pues existe una gran variedad de estimadores que pueden ser utilizados dependiendo del contexto pero el éxito de la aplicación de un estimador (estimación) dependerá necesariamente de la calidad de los datos mismos, resulta evidente que esto es extendible a los intervalos de confianza tanto para la media como para proporciones.

En el presente análisis únicamente nos restringimos a la aplicación de estimadores, considerando que los datos son de una calidad suficientemente buena para que obtengamos buenas estimaciones de los diferentes parámetros de interés así como también del tamaño de la muestra necesario tanto para la media como para la proporción.

Tema 1. Definición de estimador y estimación

Objetivo del tema

Identificar los conceptos de estimador como una función matemática y de estimación.

Desarrollo

Para realizar un análisis requerimos de una definición técnica. Utilicemos "a" como un símbolo genérico de un parámetro poblacional y, "â" para indicar una estimación de "a" basada en datos de la muestra. Una vez acordado esto podemos decir que un estimador "â" de un parámetro "a" es una función de los **valores muestrales aleatorios**, que proporciona una estimación puntual de "a". Un estimador es en sí una variable aleatoria y por consiguiente tiene una distribución muestral teórica.

Se llama **estimador puntual**[1] al número (punto sobre la recta real o recta de los números reales), que se calcula a partir de una muestra dada y que sirve como una aproximación (estimación) del valor exacto desconocido del parámetro de la población; es decir, es un valor que se calcula a partir de la información de la muestra, y que se usa para estimar el parámetro de la población.

Existe una distinción técnica entre un **estimador** como una función de variables aleatorias y una **estimación** como un único número. Tal distinción se refiere al proceso en sí (estimador) y el resultado de dicho proceso (la estimación.) Lo que en realidad importa de esta definición es que nosotros sólo podemos definir buenos procesos (estimadores), mas no garantizar buenos resultados (estimaciones).

[1] Erwin. Kreyszig, *Matemáticas avanzadas para ingeniería*, vol. 2, p. 958.

Por ejemplo, la media muestral es el mejor estimador de una población normal (); sin embargo, no podemos garantizar que el resultado sea óptimo todas las veces. Es decir, no podemos garantizar que para cada muestra la media muestral esté siempre más cerca de la media poblacional, que, digamos, la mediana muestral (es decir, puede darse el caso en el que la mediana muestral esté más próxima a la media poblacional que la media muestral). Así, lo más que podemos hacer es encontrar estimadores que den buenos resultados en el límite.

Como una aproximación[2] de la media μ de una población, puede tomarse la media $\bar{x}$ de una muestra correspondiente, lo cual da la estimación: $\hat{u}=\bar{x}$, para μ, es decir:

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{i=n} x_i \qquad \text{----------------(1)}$$

Donde n= tamaño de la muestra.

Del mismo modo, una estimación para la varianza de una población es la varianza de una muestra correspondiente; es decir:

$$\sigma^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{i=n}(x_i - x)^2 \qquad \text{-------------(2)}$$

Evidentemente, (1) y (2) son estimaciones de los parámetros para distribuciones en las que tanto la media como la varianza aparecen explícitamente como parámetros, tales como las distribuciones normal y de Poisson. Aquí, podemos mencionar que (1) es un caso muy especial del llamado **método de los momentos**, en la que los parámetros que van a estimarse se expresan en términos de los momentos de la distribución[3] en las fórmulas resultantes; esos momentos se reemplazan por los

[2] Kreyszig. Erwin. *"Matemáticas avanzadas para ingeniería vol. 2"*. pp 958
[3] Para mayor información consulte la sección 19.8 del libro: "Matemáticas avanzadas para ingeniería Vol. 2." de Erwin Kreyszig.

momentos correspondientes de la muestra, lo cual proporciona las estimaciones deseadas. Aquí, el k-ésimo momento de una muestra $x_1, x_2, \ldots x_n$, es:

$$m_k = \frac{1}{n}\sum_{i=1}^{i=n}(x_i)^k$$

ACTIVIDAD 1

Elabora un cuadro comparativo sobre las ventajas y desventajas de la estimación puntual y la estimación intervalo.

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar,** localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Autoevaluación

Selecciona si las siguientes aseveraciones son verdaderas (V) o falsas (F) las siguientes aseveraciones. Una vez que concluyas, obtendrás tu calificación de manera automática.

	Verdadera	Falsa
1. El mejor estimador de la media poblacional μ es la media muestral $\bar{x}$.	( )	( )
2. El mejor estimador para la varianza poblacional σ^2 es la varianza de la muestra.	( )	( )
3. El método de los momentos es un caso muy especial en el que los parámetros que van a estimarse se expresan en términos de los momentos de la distribución en las fórmulas resultantes	( )	( )
4. Se llama estimación puntual al número calculado a partir de una muestra dada y que sirve como aproximación del valor exacto desconocido del parámetro de la población.	( )	( )
5. Un estimador es un número que se obtiene de la aplicación de una formula	( )	( )
6. Una estimación es una formula estadística	( )	( )
7. Un estimador es una función de variables aleatorias	( )	( )

8. Una estimación es un número obtenido de la aplicación de un estimador?	( )	( )

Tema 2. Propiedades de los estimadores

Objetivos del tema

Identificará las propiedades de los estimadores

Desarrollo

Estimador insesgado

Un estimador **â**, que es una función de datos muestrales, se conoce como **estimador insesgado del parámetro poblacional** a si su valor esperado o esperanza matemática es igual a **e**. Dicho de otra manera, â es un estimador insesgado del parámetro **a** sí:

E(â)=a

La condición de que el estimador â es insesgado supone que el valor promedio de â es exactamente correcto.

Cuando el estimador es sesgado, la magnitud del sesgo está dada por la siguiente fórmula:

Sesgo (â)=E(â)–a

Si la media de las distribuciones de muestreo de un estadístico es igual que la del correspondiente parámetro de la población, el estadístico se llama un **estimador sin sesgo del parámetro**; en caso contrario, se llama un estimador sesgado. Los correspondientes valores de tales estadísticos se llaman **estimaciones sin sesgo** y **sesgadas**, respectivamente.

Por ejemplo, la media de las distribuciones de muestreo de medias $\mu_{\bar{x}}$ y μ, la media de la población. Por tanto, la media muestral $\mu_{\bar{x}}$ es una estimación sin sesgo de la media de la población μ.

En términos de esperanza matemática, podríamos decir que un estadístico es insesgado si su esperanza es igual al correspondiente parámetro de población.

Estimador eficiente

Se dice que un **estimador** es **el más eficiente** para un problema particular cuando tiene el error estándar más pequeño de todos los estimadores insesgados posibles.

Se utiliza la palabra eficiente porque, en una situación dada, el estimador hace el mejor uso posible de los datos muestrales. De acuerdo con la teoría estadística clásica, en términos generales se debe preferir el estimador insesgado más eficiente sobre cualquier otro. Más adelante veremos que las **hipótesis** nos dicen cuál es el estimador más eficiente de un cierto parámetro en un momento dado.

Así, por ejemplo, si las distribuciones de muestreo de dos estadísticos tienen la misma media (o **esperanza**), el de menor varianza se llama un **estimador eficiente de la media**, mientras que el otro se llama un estimador ineficiente. Los valores correspondientes de los estadísticos se llaman estimación eficiente y estimación ineficiente, respectivamente.

Si consideramos todos los posibles estadísticos cuyas distribuciones de muestreo tienen la misma media, aquel de varianza mínima se llama a veces el **estimador de máxima eficiencia**, o sea, el mejor estimador.

Por ejemplo, las distribuciones de muestreo de media y mediana tienen ambas la misma media, a saber, la media de la población. Sin embargo, la varianza de la distribución de muestreo de medias es menor que la varianza de la distribución de muestreo de medianas. Por tanto, la media muestral da una estimación eficiente de la media de la población, mientras la mediana de la muestra da una estimación ineficiente de ella.

De todos los estadísticos que estiman la media de la población, la media muestral proporciona la mejor (la más eficiente) estimación.

En la práctica, las estimaciones ineficientes se usan con frecuencia a causa de la relativa sencillez con que se obtienen algunas de ellas.

De manera desafortunada, las declaraciones de eficiencia dependen fuertemente de algunos supuestos. Por ejemplo, cuando la distribución de la población no es normal, la media muestral no es siempre el estimador más eficiente. Por lo anterior, surge un tema de investigación en la teoría estadística: el de los llamados **estimadores robustos**, estadísticos casi insesgados y casi eficientes para una gran variedad de distribuciones poblacionales.

Estimador consistente

Un estimador es consistente si se aproxima al parámetro poblacional con probabilidad uno a medida que el tamaño de la muestra tiende a infinito.

Por ejemplo: la media muestral $\mu_{\bar{x}}$ de una muestra aleatoria tiene valor esperado μ y un error estándar que se aproxima a cero a medida que “n” tiende a infinito. Por lo tanto, cuando el tamaño de la muestra tiende a infinito, la media muestral $\mu_{\bar{x}}$ se aproxima a μ tanto como se quiera. De acuerdo con la definición, la media muestral $\mu_{\bar{x}}$ es consistente.

Un estimador inconsistente es evidentemente un mal estimador y no es aconsejable dar una estimación imprecisa basada en una infinidad de datos, lo cual puede suceder si el sesgo de un estimador se aproxima a cero a medida que "n" tiende a infinito. Por ejemplo, utilizar el percentil 25 para estimar la mediana poblacional produciría un estimador inconsistente. También habría inconsistencia si el error estándar de un estimador no tiende a cero a medida que el tamaño muestral crece.

Por lo general, los estimadores inconsistentes son el resultado de alguna equivocación o, lo que es más probable, resultan del fracaso de una hipótesis clave.

Estimaciones de intervalo y fiabilidad

Una estimación de un parámetro de la población dada por un solo número se llama una estimación de punto del parámetro. No obstante[4], un estimador puntual sólo refiere una parte de la historia. Si bien se espera que el estimador puntual esté próximo al parámetro de la población, se desearía expresar qué tan cerca está. Un intervalo de confianza sirve para este propósito.

[4] Douglas A. Lind et al, *Estadística para administración y economía*, pp. 242.

ACTIVIDAD 1

Completa el siguiente cuadro sobre los tipos de estimadores.

	Ventajas	Desventajas
Estimadores sesgados		
Estimadores insesgados		
Estimadores consistentes		
Estimadores inconsistentes		

Descarga el siguiente cuadro para completarlo, una vez que lo tengas listo presione el botón **Examinar**. Localice el archivo, ya seleccionado, presione **Subir este archivo** para guardarlo en la plataforma.

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Autoevaluación

Elige la respuesta correcta a las siguientes preguntas.

1. En este estimador su esperanza matemática es igual a parámetro en cuestión

   - a) robusto
   - b) insesgado
   - c) sesgado

2. Es un estimador para un problema particular y tiene el error estándar más pequeño de todos los estimadores insesgados posibles.

   - a) el más eficiente
   - b) sesgado
   - c) ineficiente

3. Este tipo de estimaciones se usan con frecuencia a causa de la relativa sencillez con que se obtienen algunas de ellas

   - a) consistentes
   - b robustas
   - c ineficientes

4. Este tipo de estimadores son estadísticos casi insesgados y casi eficientes para una gran variedad de distribuciones poblacionales

   - a) consistentes
   - b) robustos
   - c) eficientes

5. Este tipo de estimador se aproxima al parámetro poblacional con probabilidad uno a medida que el tamaño de la muestra tiende a infinito

- a) consistente
- b) robusto
- c) inconsistente

Tema 3. Estimación de media, varianza y proporciones

Objetivos del tema

Aplicar los estimadores de la media, varianza y proporciones en los intervalos de confianza.

Desarrollo

Intervalo de confianza

Un rango de valores que se construye a partir de datos de la muestra de modo que el parámetro ocurre dentro de dicho rango con una probabilidad específica. La probabilidad específica se conoce como nivel de confianza.

Es decir, una estimación de un parámetro de la población dada por dos números, entre los cuales se puede considerar encajado al parámetro, se llama una **estimación de intervalo del parámetro**.

Las estimaciones de intervalo indican la precisión de una estimación y son por tanto preferibles a las estimaciones de punto.

Por ejemplo: si decimos que una distancia se ha medido como 5.28 metros (m), estamos dando una **estimación de punto.** Por otra parte, si decimos que la distancia es 5.28±0.03 m (o sea, que esta entre 5.25 y 5.31 m) estamos dando una **estimación de intervalo**.

El margen de error (o la precisión) de una estimación nos informa de su fiabilidad.

En estadística, numerosos problemas están relacionados con la estimación de la media o la desviación estándar de una población dada a partir del estudio de una muestra de tamaño “n”.

Así, por ejemplo:

- A una empresa le puede interesar el número promedio de piezas defectuosas producidas por una cierta máquina.
- A un ingeniero especialista en vehículos le puede interesar la **variabilidad** en el funcionamiento de un tipo vehículo.

En las secciones anteriores se vio que si se supone que cada muestra de tamaño “n” tiene la misma probabilidad de ser seleccionada, entonces la media de la distribución de las medias de la muestra es la misma que la de la población original, $\mu_{\bar{x}} = \mu$. Aún más, para poblaciones suficientemente grandes, o para muestreos con reemplazo, la desviación estándar de la distribución de las medias de la muestra, $\sigma_{\bar{x}}$, está relacionada con la desviación estándar de la población σ por la ecuación:

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

Si en una aplicación particular fuera práctico seleccionar todas las posibles muestras de tamaño “n”, para determinar la media de cada una de ellas y, después, calcular la media y la desviación estándar de la distribución de las medias de las muestras, las fórmulas anteriores permitirían calcular μ y σ directamente. Por lo general, este procedimiento no es práctico. Lo que comúnmente se hace es no estudiar todas las muestras de tamaño “n” sino únicamente una de ellas. La media $\bar{x}$ y la desviación estándar “s” de esa muestra únicamente se toman como estimaciones de μ y σ, es decir, de la media y la desviación estándar que corresponden a la población original. Puesto que $\mu_{\bar{x}} = \mu$ y $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, las estimaciones para $\mu_{\bar{x}}$ y $\sigma_{\bar{x}}$, son $\bar{x}$ y $\frac{s}{\sqrt{n}}$ respectivamente.

Enseguida se ilustra el procedimiento de estimación con un ejemplo: se escoge una muestra aleatoria de 36 recién egresados en la carrera de contaduría de cierta universidad; al aplicarles un examen de aptitudes, se obtuvieron las siguientes puntuaciones:

63	**64**	**64**	**65**	**65**	**66**
66	**66**	**67**	**67**	**67**	**67**
67	**68**	**68**	**68**	**69**	**69**
69	**69**	**69**	**70**	**70**	**70**
71	**72**	**72**	**72**	**72**	**73**
73	**74**	**74**	**76**	**76**	**77**

La media de la muestra $\bar{x}$ es de 69, (al punto más próximo), y la desviación estándar "s", es de 3.5. Utilizando $\bar{x}$ y "s" como estimaciones de μ y σ podemos afirmar que la puntuación media de todos los recién egresados de dicha universidad es de alrededor de 69 puntos. Aún más, podemos decir que la desviación estándar de las puntuaciones de los recién egresados respecto a la media es, aproximadamente, 3.5 puntos.

El procedimiento anterior es satisfactorio tal como se ha presentado. El problema estriba en el contenido de las palabras alrededor de y aproximadamente.

Por supuesto, la exactitud de nuestra estimación depende de la muestra escogida. Afortunadamente, en el caso de muestras aleatorias, es posible dar apoyo probabilístico al significado de las palabras alrededor de y aproximadamente.

Un hecho importante que se debe tener en cuenta en la distribución de las medias de las muestras, cuando ésta es grande y se selecciona aleatoriamente, es que se puede aproximar a una distribución normal que tenga la misma media $\mu_{\bar{x}}$ y la misma desviación estándar $\sigma_{\bar{x}}$.

Puesto que la distribución de las medias de las muestras es aproximadamente normal, se puede utilizar de manera efectiva el conocimiento sobre este tipo de distribución.

ACTIVIDAD 1

Completa el siguiente cuadro sobre los tipos de estimadores.

Estimador	Situación
Media	
Varianza	
Proporción	

Descarga el siguiente cuadro para completarlo, una vez que lo tengas listo presione el botón **Examinar**. Localice el archivo, ya seleccionado, presione **Subir este archivo** para guardarlo en la plataforma.

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Autoevaluación

Elige la respuesta correcta a las siguientes preguntas.

1. Esta estimación es dada por dos números, entre los cuales se puede considerar encajado al parámetro de estudio.

- a) estimación de intervalo
- b) estimación puntual
- c) estimación de la media

2. En una muestra de tamaño "n" seleccionada de manera aleatoria sabemos que la media de la distribución de las medias siempre es igual a:

- a) la media de cualquier población
- b) la media de cualquier otra muestra
- c) la media de la población

3. Para muestreos llevados a cabo con reemplazo o para poblaciones suficientemente grandes, la desviación estándar de la distribución de las medias de la muestra se relaciona con la desviación estándar de la población por medio de la ecuación:

- a) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
- b) $\mu_{\bar{x}} = \mu$
- c) $-1 < z < +1$

4. Con este símbolo se denota a la media de la distribución de las medias de las muestras obtenidas de la misma población:

   - a) Z
   - b) $\mu_{\bar{x}}$
   - c) φ

5. Con este símbolo se denota a la desviación estándar de la distribución de las medias de las muestras obtenidas de la misma población:

   - a) σ
   - b) $\sigma_{\bar{x}}$
   - c) $\sigma^2_{\bar{x}}$

Tema 4. Intervalo de confianza para la media y para proporciones

Objetivos del tema

Aplicar intervalos de confianza de diferentes porcentajes para la media y para la proporción.

Desarrollo

Una estimación de un parámetro de la población dada por un solo número se llama una **estimación de punto del parámetro**. No obstante[5], un estimador puntual sólo refiere una parte de la historia. Si bien se espera que el estimador puntual esté próximo al parámetro de la población, se desearía expresar qué tan cerca está. Un intervalo de confianza sirve a este propósito.

Intervalo de confianza

Un rango de valores que se construye a partir de datos de la muestra de modo que el parámetro ocurre dentro de dicho rango con una probabilidad específica se conoce como nivel de confianza.

Es decir, una estimación de un parámetro de la población dada por dos números, entre los cuales se puede considerar encajado al parámetro, se llama una estimación de intervalo del parámetro.

Las **estimaciones de intervalo** indican la precisión de una estimación y son, por tanto, preferibles a las estimaciones puntuales.

Por ejemplo: si decimos que el porcentaje de productos defectuosos que produce una máquina es del 6%, entonces el nivel se ha medido en 0.06 y estamos dando una **estimación de punto**. Por otra parte, si decimos que el porcentaje es

[5] Douglas A. Lind et al., Estadística para administración y economía, pp. 242

0.05±0.03 m (o sea, que esta entre 2% y 8%), estamos dando una **estimación de intervalo.**

El **margen de error** (o la precisión) de una estimación nos informa de su fiabilidad.

Intervalo para estimar la media

De acuerdo con tablas de la distribución normal estándar el área bajo la curva entre z=-1 y z=+1 es 0.6826; por consiguiente, y de acuerdo con la definición de la función normal estándar de probabilidad, las desigualdades siguientes se cumplen con probabilidad de 0.6826

$$-1 < z < +1$$

Como la distribución de las medias de las muestras (con media $\mu_{\bar{x}}$ y desviación estándar $\sigma_{\bar{x}}$) es normal, entonces:

si reemplazamos z por $\frac{x-\mu_x}{\sigma_{\bar{x}}}$ en las desigualdades anteriores, se deberá cumplir:

$$-1 < \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} < +1$$

Con probabilidad 0.6826. Esto es equivalente a que las desigualdades:

$$\bar{X} - \sigma_{\bar{x}} < \mu_{\bar{x}} < \bar{X} + \sigma_{\bar{x}}$$

se cumplan también con probabilidad 0.6826; sustituyendo ahora:

$\sigma_{\bar{x}}$ por $\frac{s}{\sqrt{n}}$ y $\mu_{\bar{x}}$ por μ_x

se tiene que:

$$\overline{X} - \frac{s}{\sqrt{n}} < \mu_x < \overline{X} + \frac{s}{\sqrt{n}}$$

Se cumple con la misma probabilidad.

Podemos esperar entonces que con una probabilidad de 0.68 que μ_x se encuentre dentro del intervalo:
(69– 0.58, 69+0.58)

Es decir: $68.42 < \hat{\mu}_x < 69.58$ aquí, la media aritmética de la población lleva un acento circunflejo debido a que se trata de una estimación.

Se dice que éste es un **intervalo de confianza** de 0.68 o 68%, ya que se tiene una confianza de 68% de que el intervalo contenga la media de la población.

Si una confianza de 68% fuese insuficiente se pueden construir otros intervalos con porcentajes de confianza que sean más útiles.

Por ejemplo: si se deseara encontrar un intervalo de confianza de 0.95 para se requeriría determinar “k” de tal manera que las desigualdades siguientes se cumplieran con probabilidad de 0.95

$$-k < z < +k \quad \text{------------------------------------1}$$

En términos generales, para encontrar un intervalo de cualquier porcentaje de confianza, se hace lo siguiente:

1º. Se divide el porcentaje de confianza requerido entre 100

2º. El resultado del punto anterior se divide entre 2

3º. El valor así obtenido se busca en las tablas de la curva de distribución normal

4º. El valor encontrado en las tablas se sustituye en 1 y comenzamos el proceso nuevamente.

Es decir, en nuestro caso el valor resultante es de 0.475; por lo tanto, el valor en las tablas que se encuentra junto a éste último es "1.96". Es decir, el área bajo la curva normal estándar entre –1.96 y +1.96 es 0.9544, o sea, aproximadamente 0.95. Así, la probabilidad de que z se encuentre dentro del intervalo:

(-1.96, +1.96)

es, aproximadamente 0.95 o, en otra forma, las desigualdades:

-1.96 <z<+1.96

se cumplen con probabilidad 0.95;

y puesto que se sabe que la distribución de las medias de las muestras es normal,

se puede reemplazar z por $\frac{x - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$

expresión que aproximada a:

$$\frac{\bar{X} - \mu_x}{\frac{s}{\sqrt{n}}}$$

en las desigualdades anteriores, se obtiene:

$$-1.96 < \frac{\bar{X} - \mu_x}{\frac{s}{\sqrt{n}}} < +1.96$$

Resolviendo estas desigualdades para μ , se tiene que:

$$\bar{X} - \frac{1.96s}{\sqrt{n}} < \mu_x < \bar{X} + \frac{1.96s}{\sqrt{n}} \qquad \text{------------------2}$$

Como un intervalo con 0.95 de confianza para μ . Por lo tanto, se puede afirmar con 95% de confianza que μ se encuentra dentro del intervalo:

$$\bar{X} - \frac{1.96s}{\sqrt{n}} \qquad \text{y} \qquad \bar{X} + \frac{1.96s}{\sqrt{n}}$$

Por lo tanto, sustituyendo los valores de la media y de la desviación estándar, así como del tamaño de la muestra para el ejercicio anterior (media 69, desviación estándar 3.5 y tamaño de muestra 36) en 2 se tiene que el intervalo con 95% de confianza es:

$$69 - \frac{1.96(3.5)}{\sqrt{36}} < \mu_x < 69 + \frac{1.96(3.5)}{\sqrt{36}}$$

$$67.8 < \mu_x < 70.1$$

(67.8 , 70.1)

(67.9

Intervalo para estimar la varianza

Propiedades de los estimadores, sabemos que el estimador para varianza poblacional (σ^2) es S^2; sin embargo, para estimar un intervalo de confianza para σ^2 es necesario conocer la distribución del estadístico; más aún, la metodología implica que es necesario tener un estadístico que involucre el parámetro desconocido y que además tenga distribución perfectamente conocida. Por lo cual, en este caso el estadístico es:

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

Que de acuerdo con lo estudiado en el tema 2 tiene una distribución Chi-cuadrada con n-1 grados de libertad. Así que para una muestra particular, dicho estadístico tiene una probabilidad de estar en un rango dado.

Ejemplo:

Considera el caso de estimar si no hay deficiencias en una máquina que llena envases con capacidad de 500 ml.; para ello, se extrae una muestra periódicamente; si la muestra indica que hay una variación de ±5 ml. alrededor de los 500 y con un nivel de confianza del 95%, entonces se puede decir que el proceso está bajo control.

En este caso lo que importa es la variación en el llenado, pues el nivel promedio de llenado se puede controlar programando la máquina. Por ello, si la muestra arroja una variación arriba de 5 unidades, entonces el proceso no estará bajo control.

Suponga que la muestra de tamaño 41 arroja una varianza de 13 unidades (desviación estándar de 3.60 ml). Entonces, de acuerdo con la estimación por intervalos de confianza, se tendrá que:

$$X^2_{0.025} < \frac{(n-1)S^2}{\sigma^2} < X^2_{0.975}$$

El resultado anterior de acuerdo con tablas de Chi-cuadrada con 40 grados de libertad $X^2_{0.025}$=24.433 y $X^2_{0.09750}$ = 59.342.

(Recuerda que el uso de las tablas y de los grados de libertad se encuentra en el apartado 3.2)

Entonces el intervalo es:

$$24.433 < \frac{(n-1)S^2}{\sigma^2} < 59.342$$

Sustituyendo los resultados de la muestra se tiene:

$$24.433 < \frac{(40-1)(13)}{\sigma^2} < 59.342$$

Al obtener inversos multiplicativos tenemos:

$$\frac{1}{24.433} > \frac{\sigma^2}{(40-1)(13)} > \frac{1}{59.342}$$

Despejando todas las constantes y dejar solo σ^2 se tiene el intervalo:

$$\frac{1}{24.433} > \frac{\sigma^2}{(40-1)(13)} > \frac{1}{59.342}$$

$$20.75 > \sigma^2 > 8.54$$

Obteniendo raíz cuadrada, se tiene:

$$4.555 > \sigma > 2.92$$

Por lo cual se puede decir que el proceso está bajo control.

Intervalo para estimar la proporción

En el caso de la proporción, el estadístico por utilizar es:

$$\frac{\bar{p} - \mu_{\bar{p}}}{\sigma_{\bar{p}}} = \frac{\bar{p} - P}{\sqrt{P(1-P)/n}}$$

El cual, de acuerdo con el teorema del límite central, tendrá distribución normal estándar. En este caso, P es la proporción de la población con una característica dada y que se puede estimar por medio de $\overline{p}$, que es la proporción de la muestra con la característica.

Ejemplo:

Considera el caso de la Bolsa Mexicana de Valores; se desea estimar la proporción de las 250 acciones que tendrán una baja en precio al cierre del día. Para ello se observa una muestra de las primeras 4 horas sobre 50 acciones operadas y se observó que la proporción que bajo de precio son el 0.10 (10%). En el día se estima que no se presenten turbulencias por información importante o privilegiada. Se pide determinar el intervalo de confianza para la proporción total de acciones a la baja con un nivel de confianza del 90%.

De acuerdo con la metodología indicada el intervalo estará determinado por:

$$Z_{\alpha/2} < \frac{\overline{p} - P}{\sqrt{\overline{p}(1-\overline{p})/n}} < Z_{1-\alpha/2}$$

Pero de acuerdo con tablas normal estándar $Z_{\alpha/2} = Z_{0.05} = -1.64$ y $Z_{0.95} = 1.64$ y como $\overline{p} = 0.10$ entonces el intervalo se deduce de:

$$-1.64 < \frac{0.10 - P}{\sqrt{0.10(1-0.10)/50}} < 1.64$$

$que\, equivale\, a:$

$$-1.64(0.0424264) < 0.10 - P < 1.64(0.0424264)$$

$y\, despejando\, P\, se\, tiene:$

$$-1.64(0.04242064) - 0.10 < -P < 1.64(0.0424264) - 0.10$$

$igual\, a:$

$$1.64(0.0424264) + 0.10 > P > -1.64(0.0424264) + 0.10$$

$Por\, lo\, cual\, el\, intervalo\, es:$

$$0.169 > P > 0.0304$$

Es decir aproximadamente entre el 3% y 17%.

ACTIVIDAD 1

1. Construye un intervalo de confianza para la varianza de forma general.

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar**, localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

2. Construye un intervalo de confianza para la proporción de forma general.

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar,** localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

Autoevaluación

Elige la respuesta correcta a las siguientes preguntas.

1. Un intervalo de confianza es:

- a) Un estimador puntual
- b) Un rango de valores que se construye a partir de datos de una muestra de modo que el parámetro en cuestión ocurre dentro de dicho rango con una probabilidad específica conocida como nivel de confianza.
- c) Una estimación puntual

2. Este tipo de estimación se prefiere sobre las estimaciones puntuales debido a que indican la precisión de la estimación:

- a) estimación de intervalo
- b) estimación subjetiva
- c) estimación sesgada

3. La formula para estimar la media de la población a través de un intervalo de confianza es:

- a) $\overline{X} - \frac{s}{\sqrt{n}} < \mu_x < \overline{X} + \frac{s}{\sqrt{n}}$
- b) $P(y) = e^{-\mu} \frac{\mu^y}{y!}$
- c) $\frac{dl(1,4,\mu)}{d\mu} = -2 + \frac{5}{\mu} = 0$

4. La fórmula más utilizada para realizar una estimación de intervalo para la proporción es:

❍ a) $\dfrac{\bar{p}-\mu_{\bar{p}}}{\sigma_{\bar{p}}}=\dfrac{\bar{p}-P}{\sqrt{P(1-P)/n}}$

❍ b) $\bar{X}-\dfrac{s}{\sqrt{n}}<\mu_x<\bar{X}+\dfrac{s}{\sqrt{n}}$

❍ c) $Z_{\alpha/2}<\dfrac{\bar{p}-P}{\sqrt{\bar{p}(1-\bar{p})/n}}<Z_{1-\alpha/2}$

5. La fórmula para estimar la varianza de la población utilizando un intervalo de confianza es:

❍ a) $s^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$

❍ b) $\dfrac{s^2(n-1)}{\chi^2_{\alpha/2}}\leq\sigma^2\leq\dfrac{s^2(n-1)}{\chi^2_{1-\alpha/2}}$

❍ c) $\dfrac{s^2(n-1)}{\chi^2_{\alpha/2}}$

Tema 5. Determinación del tamaño de la muestra

Objetivos del tema

Calcular el tamaño de la muestra tanto para la media como para la proporción.

Desarrollo

Tamaño de muestra para la media

Hemos visto que para estimar por intervalos la media, el ancho del intervalo está dado por:

$$Z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Lo anterior representa el número de desviaciones estándar alrededor de la media μ dado el nivel de confianza 1-α, por lo que si quisiéramos estimar μ con un nivel de confianza dado y obtener un error en la estimación de a lo más B, tenemos que despejar n de la ecuación:

$$B = Z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Despejando n, tenemos:

$$\mathrm{B}\sqrt{\mathrm{n}} = Z_{\alpha/2}S$$

o bien:

$$\mathrm{n} = \left(\frac{Z_{\alpha/2}S}{B}\right)^2$$

Observa que la fórmula involucra el valor S de una muestra, por lo cual el muestreo se puede hacer en dos etapas: en una primera prueba piloto se muestrea con un número reducido de elementos y con ello se calcula el tamaño de n; posteriormente, se muestrea en una segunda etapa y se completa la muestra dada por el valor de n.

Como ejemplo supongamos que una empresa comercializa soya texturizada (tipo carne) y deseamos estimar el consumo promedio semestral de una población de consumidores potenciales. Suponga que una muestra piloto de 15 personas arroja que S=12.2 kg.; así, si deseamos un nivel de confianza del 95% y un error en la estimación de B=2 Kg., entonces el tamaño de muestra en este caso se obtiene como:

$$n = \left(\frac{Z_{\alpha/2} S}{B}\right)^2 = \left(\frac{1.96(12.2)}{2}\right)^2 = 142.9459$$

Es decir, se deben muestrear aproximadamente 143 (128 adicionales a los 15 ya muestreados).

Tamaño de muestra para la proporción

En este caso el error en la estimación está dado por:

$$B = Z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}$$

La fórmula anterior representa el número de desviaciones estándar alrededor de la media P dado el nivel de confianza 1-α; así, si quisiéramos estimar P con un nivel de confianza dado y obtener un error en la estimación de a lo más B, tenemos que despejar n de la ecuación:

$$B = Z_{\alpha/2}\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

Despejando n, tenemos:

$$\mathrm{B}^2 = Z^2{}_{\alpha/2}\,\frac{\overline{p}(1-\overline{p})}{n} \qquad \text{o bien:}$$

$$\mathrm{n} = \left(\frac{\mathrm{Z}^2{}_{\alpha/2}\,\overline{p}(1-\overline{p})}{B^2}\right)$$

Suponga que se desea estimar la proporción de acciones que tendrán una baja en el día, para lo cual se observa una muestra de 20 acciones en las que el promedio de las que bajaron son $\overline{p} = 0.17$; si se desea tener un nivel del 95% de confianza de cometer un error de cuando mucho B=0.09 (9%) en la estimación, determinar el tamaño de muestra.

$$\mathrm{n} = \left(\frac{\mathrm{Z}^2{}_{\alpha/2}\,\overline{p}(1-\overline{p})}{B^2}\right)^2 = \left(\frac{(1.96)^2(0.17)(0.83)}{0.09^2}\right)^2 = 66.91$$

Es decir se deben muestrear aproximadamente 67 (47 adicionales a los 20 ya muestreados).

Como podemos observar el método estadístico nos permite realizar estudios tales que nos permite autocorregir en un momento dado nuestra apreciación no solo en cuanto al tamaño de una muestra, sino también a la hora de dar una confianza el momento de emitir nuestros resultados.

ACTIVIDAD 1

Resuelve los siguientes problemas, escribe tu respuesta en el espacio en blanco.

1. Considera una empresa que comercializa productos genéricos de limpieza y deseamos estimar el consumo promedio anual de una población potencia. Si obtenemos una muestra piloto de 15 personas en donde S=28.9 l y queremos un nivel de confianza de 95% con un error en la estimación de B=2 l. determina el tamaño de la muestra que debe evaluarse.

2. El día de hoy la bolsa mexicana de valores estimo con una muestra de 20 acciones que el promedio de las que estuvieron a la alza fue de $\overline{p}=0.10$; con un nivel de confianza de 90% y un error a lo más de 5%. Determina el tamaño de la muestra que debe ser estudiada.

Pulsa el botón **Comenzar** para contestar las preguntas, una vez concluyas pulsa el botón **Enviar todo y terminar**

Bibliografía básica

Autor	Capítulo	Páginas

Sitios electrónicos

Sitio	Descripción

Autoevaluación

Elige la respuesta correcta a las siguientes preguntas, una vez que concluyas, obtendrás de manera automática tu calificación.

1. Cuando se define un intervalo de valores de la muestra y se menciona que dentro del mismo es muy probable que se encuentre el parámetro poblacional, se dice que se está realizando:

- a) Un análisis estadístico
- b) Una estimación de punto
- c) Una estimación de intervalo
- d) Una prueba de hipótesis
- e) Un estudio inferencial

2. Suponga que estamos realizando la estimación por intervalo del valor de la media poblacional considerando un nivel de confianza del 99%, ¿cuál de los intervalos siguientes expresa nuestra intención?

- a) $\mu_S \pm \sigma_S$
- b) $\mu_S \pm 1.96\sigma_S$
- c) $\mu_S \pm 2\sigma_S$
- d) $\mu_S \pm 2.58\sigma_S$
- e) $\mu_S \pm 3\sigma_S$

3. Determine el intervalo de valores correspondiente a un nivel de confianza del 99% para el valor de la media poblacional si una muestra de 200 datos dieron una media de 0.824 pulgadas con una desviación estándar de 0.042 pulgadas:

- a) 0.824 ± 0.005
- b) 0.824 ± 0.006
- c) 0.824 ± 0.009
- d) 0.824 ± 0.008
- e) 0.824 ± 0.003

4. La corrección que se realiza al valor de la desviación estándar por la consideración de población finita depende de la relación que guardan los tamaños de la población y de la muestra, ¿cuál es la relación por considerar?

- a) n/N >5%
- b) n/N < 5%
- c) n/N =5%
- d) n/N = 10%
- e) n/N >10%

5. Una muestra al azar de 50 calificaciones de proyectos de inversión de un total de 200 arrojó una media de 75 y una desviación estándar de 10. ¿Con qué nivel de confianza podrá decirse que la media de las 200 calificaciones es de 75±1?

- a) 73.2%
- b) 46.9%
- c) 63.4%
- d) 56.52%
- e) 81.0%

6. Durante el envase de mermeladas se obtuvo en un envase un peso de 216.48 gramos. Si se sabe que el error probable es de 0.272 gramos. ¿Cuáles son los límites de confianza del 95% (en gramos) para dicho peso?

- a) 216.48 ± 0.56
- b) 216.48 ± 0.57
- c) 216.48 ± 0.55
- d) 216.48 ± 0.53
- e) 216.48 ± 0.54

7. Si las distribuciones muestrales de dos estadísticos tienen la misma media, entonces el estadístico más eficiente es el que tenga:

- a) Solo una frecuencia modal
- b) Menor varianza
- c) Sesgo hacia la derecha
- d) Mediana y media más parecidas
- e) Más intervalos de clase

8. La precisión o margen de error de una estima se conoce como:

- a) Seguridad
- b) Variación
- c) Aproximación
- d) Desviación estándar
- e) Varianza

9. Si en una muestra grande con distribución normal se toma un estadístico S, ¿qué porcentaje de la muestra se encuentra en el intervalo $\mu_S \pm \sigma_S$?:

- a) 60%
- b) 68.27%
- c) 75%
- d) 95.45%
- e) 90%.

10. Si en una muestra grande con distribución normal se toma un estadístico S, ¿con qué intervalo se obtiene el 99% de nivel de confianza?:

- a) $\mu_S \pm \sigma_S$
- b) $\mu_S \pm 1.96\sigma_S$
- c) $\mu_S \pm 2\sigma_S$
- d) $\mu_S \pm 2.58\sigma_S$
- e) $\mu_S \pm 3\sigma_S$

LO QUE APRENDÍ DE LA UNIDAD

Construya un intervalo de confianza de 95% para la vida media de los neumáticos muestreados en la tabla mostrada a continuación. (Nota. Los datos están dados en miles de kilómetros.)

85,000	90,000	100,000	105,000
90,000	95,000	92,300	97,200
91,000	98,000	97,000	97,500
88,000	89,900	99,600	99,500
97,890	99,870	95,490	94,789
90,890	99,810	98,900	
97,870	97,980	99,950	
96,190	96,710	95,498	
98,990	97,900	95,267	
96,876	96,930	99,900	

Realiza esta actividad en un procesador de textos, guárdala en tu computadora y, una vez concluida, presiona el botón **Examinar,** localiza el archivo, selecciónalo y haz clic en **Subir este archivo** para guardarlo en la plataforma.

Glosario de la unidad

Estimación de intervalo

Estimación de un parámetro de la población que define un intervalo dentro del que se cree está contenido el valor del parámetro. Tiene la forma de: Estimación puntual $\pm$ margen de error.

Margen de error

Es el valor $\pm$ sumado y restado a una estimación puntual a fin de determinar un intervalo de confianza de un parámetro poblacional.

Error muestral

Es el valor absoluto de la diferencia entre el valor de un estimador puntual insesgado, tal como la media de la muestra $\bar{x}$ y el valor del parámetro poblacional que estima, (en este caso, la media de la población μ); es decir, en este caso el error muestral es: $\left|\bar{x}-\mu\right|$

Nivel de confianza

Es la confianza asociada con una estimación de intervalo. Por ejemplo si en un proceso de estimación de intervalo, el 90% de los intervalos formados con este procedimiento contienen el valor del parámetro buscado, se dice que éste es un intervalo de 90% de confianza.

Distribución t

Es en realidad una familia de distribuciones de probabilidad que se emplea para construir un intervalo de confianza para la media poblacional, siempre que la desviación estándar σ se estime mediante la desviación estándar muestral "s" y la población tenga una distribución de probabilidad normal o casi normal.

Grados de libertad

Es el número de observaciones independientes para una fuente de variación menos el número de parámetros independientes estimado al calcular la variación.

MESOGRAFÍA

Bibliografía básica