

Estimación por Intervalo

Hasta ahora los estimadores estudiados son puntuales, es decir, exhiben un valor como estimación del parámetro de interés. En muchos casos esto no es suficiente; se requiere de un rango de posibles valores donde se cree el parámetro de interés estará con una alta confianza. Sea θ un parámetro de interés y $\hat{\theta}$ un estimador puntual de θ .

Una estimación de θ por intervalos es un intervalo real de la forma:

$$(l, u) \quad (l < \theta < u)$$

donde l y u dependen de $\hat{\theta}$ y de la distribución de $\hat{\theta}$

Usando $\hat{\theta}$ y la distribución de $\hat{\theta}$ se pueden determinar L y U tales que $P(L < \theta < U) = 1 - \alpha$, $\alpha \in (0, 1)$ dado.

estos extremos se convierten en v.a

Se denomina

INTERVALO ALEATORIO


Para una m.a. dada el intervalo (l, u) es llamado **intervalo de confianza** al $100(1-\alpha)\%$ para θ . l y u son llamados límites de confianza (inferior y superior respectivamente) y $1 - \alpha$ es llamado coeficiente de confianza. Este intervalo se conoce como IC Bilateral. pero también es posible calcular IC unilaterales:


a) Por su extremo o cota se clasifican en:

I.C inferior para θ al $100(1 - \alpha)\%$: $l \leq \theta$ $(l, +\infty)$

I.C superior para θ al $100(1 - \alpha)\%$: $\theta \leq u$ $(-\infty, u)$

b) Por la región que estiman se clasifican en:

I.C hacia la derecha para θ al $100(1 - \alpha)\%$: $l \leq \theta$ $(l, +\infty)$ 

I.C hacia la izquierda para θ al $100(1 - \alpha)\%$: $\theta \leq u$ $(-\infty, u)$ 

En un IC bilateral la longitud $u - l$ es una medida de la calidad de la información obtenida. El semiintervalo $\theta - l$ ó $u - \theta$ se conoce como **Precisión del Estimador**. Lo ideal es tener IC angostos con una alta confianza.

Método general para obtener intervalos de confianza

Sea X_1, X_2, \dots, X_n una m.a. de una distribución que depende de un parámetro θ . Supongamos que existe una función $T(X_1, X_2, \dots, X_n, \theta)$ (es decir, una función de la muestra y del parámetro denominado pivote) cuya distribución no depende de θ ni de ningún otro parámetro desconocido. Entonces, existen dos valores a y b tales que:

$$P(a \leq T(X_1, X_2, \dots, X_n, \theta) \leq b) = 1 - \alpha$$

a partir de esta expresión, es posible obtener un intervalo de confianza para θ .

$$P(l(X_1, \dots, X_n) < \theta < u(X_1, \dots, X_n)) = 1 - \alpha$$

$l(X_1, \dots, X_n)$ y $u(X_1, \dots, X_n)$ son los límites de confianza, y para una muestra en particular se obtiene el I. C. (l, u) $100(1 - \alpha)\%$ para θ .

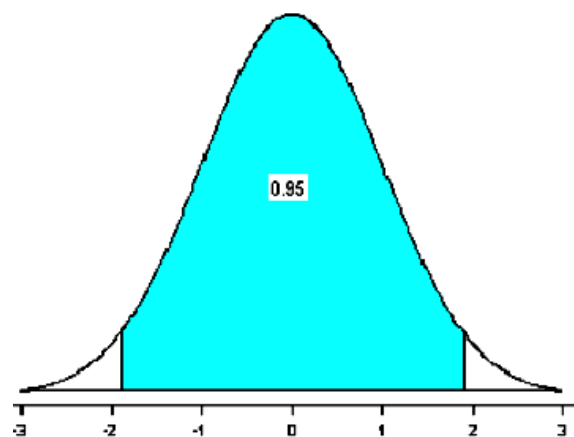
Ejemplo: Supongamos que tenemos una m.a. X_1, X_2, \dots, X_n de una distribución $N(\mu, \sigma_o^2)$ con varianza σ_o^2 conocida. Por ser los datos normales, sabemos que

$$\bar{X} \sim N\left(\mu, \frac{\sigma_o^2}{n}\right) \Leftrightarrow \frac{\bar{X} - \mu}{\frac{\sigma_o}{\sqrt{n}}} \sim N(0,1)$$

pivote \rightarrow Distribución muestral del pivote

y, por lo tanto, sabemos que la probabilidad de que $\sqrt{n} \frac{\bar{X} - \mu}{\sigma_o}$ se encuentre entre -1.96 y 1.96 es 0.95 , es decir

$$P\left(-1.96 \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma_o} \leq 1.96\right) = 0.95$$



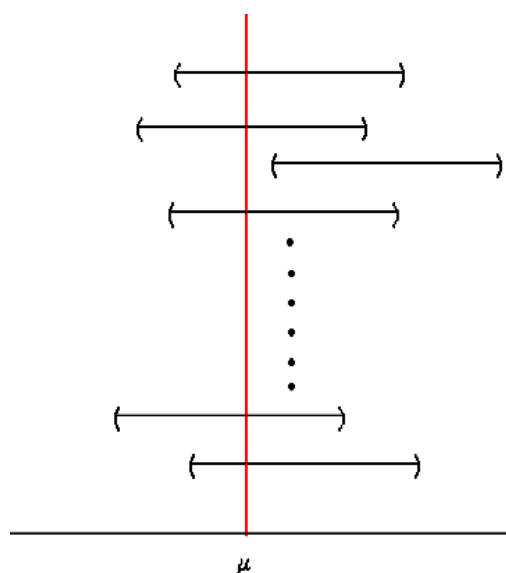
A partir de esta expresión obtenemos

$$P\left(-1.96 \frac{\sigma_o}{\sqrt{n}} \leq \bar{X} - \mu \leq 1.96 \frac{\sigma_o}{\sqrt{n}}\right) = 0.95 \quad \Leftrightarrow \quad P\left(\bar{X} - 1.96 \frac{\sigma_o}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma_o}{\sqrt{n}}\right) = 0.95$$

Es decir, que la probabilidad de que el intervalo

$$\left[\bar{X} - 1.96 \frac{\sigma_o}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma_o}{\sqrt{n}} \right]$$

contenga al verdadero valor del parámetro μ es 0.95. Este intervalo se denomina intervalo de confianza para μ de nivel de confianza 0.95.



1. Estimación de la Media

Suponga que X_1, X_2, \dots, X_n es una muestra aleatoria de una población Normal con media μ y varianza σ^2 y que se quiere obtener un intervalo de confianza del $(1-\alpha) 100\%$ para μ . El estimador de μ es:

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

1.1. σ^2 conocido

Pivote	Intervalo
$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$	$IC(\mu) = \bar{X} \pm \frac{\sigma}{\sqrt{n}} \cdot Z_{1-\alpha/2}$

usar

$$\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}; \text{ si la población es finita}$$

Nota:

Teorema del Limite Central: Si X_1, X_2, \dots, X_n es una muestra aleatoria de una población X con media μ y varianza finita σ^2 , entonces:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0, 1) \text{ cuando } n \rightarrow \infty$$

Debido a este Teorema, el intervalo anterior es valido para estimar la media de cualquier distribución especialmente cuando $n \rightarrow \infty$.

Ejemplo:

Una muestra aleatoria de 36 cigarrillos de una marca determinada dio un contenido promedio de nicotina de 3 miligramos. Suponga que el contenido de nicotina de estos cigarrillos sigue una distribución normal con una desviación estándar de 1 miligramo.

- a) Obtenga e interprete un intervalo de confianza del 95% para el verdadero contenido promedio de nicotina en estos cigarrillos.
- b) El fabricante garantiza que el contenido promedio de nicotina es de 2,9 miligramos, ¿qué puede decirse de acuerdo con el intervalo hallado?

Solución

a) *INTERVALO*: $\bar{x} \pm z_{(1-\alpha/2)} \frac{\sigma_x}{\sqrt{n}}$

reemplazando los datos: $3 \pm (1.96) \left(\frac{1}{\sqrt{36}} \right)$

$$[2.67, 3.33]$$

Interpretación:

Tenemos una certeza del 95% de que el verdadero contenido promedio de nicotina se halla entre 2.67 y 3.33 miligramos

b) Como 2.9 se encuentra en el intervalo hallado no podemos descartarlo como valor posible del parámetro

Ejemplo:

Se debe estimar el contenido promedio de grasa en leche “Larga Vida”, cuyo contenido nominal es de 30 gr por litro. Se selecciona una m.a. de 64 cajas (de 1

litro) obteniéndose $\sum_{i=1}^{64} X_i = 1895,3$ y $\sum_{i=1}^{64} X_i^2 = 57403,0$. Se sabe que

la desviación estándar poblacional es de 1 gr.

- Obtenga un valor con el cual se tenga 97,5% de confianza de que la media poblacional excederá a dicho valor.
- Dé una cota superior de confianza del 95% para el contenido medio de grasa en la leche, ¿diría usted que el valor nominal difiere exageradamente a la luz de lo observado?

Solución

De acuerdo a la información dada, tenemos

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{1895,3}{64} = 29,61406, \quad \sigma = 1, \quad \gamma = 1 - \alpha = 0,975, \quad \alpha = 0,025, \quad Z_{\alpha} = 1,96$$

Solución

a) El intervalo de confianza unilateral hacia la derecha para μ con coeficiente de confianza 97,5% es

$$I = < \bar{X} - Z_{\alpha} \cdot \frac{\sigma}{\sqrt{n}} ; +\infty > = < 29,61406 - (1,96) \frac{1}{\sqrt{64}} ; +\infty > = < 29,36906 ; +\infty >$$

Por consiguiente, el valor buscado es 29,36906

b) Para el coeficiente de confianza $\gamma = 1 - \alpha = 0,95$, tenemos $Z_{\alpha} = 1,645$. Luego, el intervalo de confianza unilateral hacia la izquierda para μ con coeficiente de confianza del 95% es:

$$I = < -\infty ; \bar{X} + Z_{\alpha} \frac{\sigma}{\sqrt{n}} > = < -\infty ; 29,61406 + (1,645) \frac{1}{\sqrt{64}} > = < -\infty ; 29,819685 >$$

Por tanto el valor nominal 30 (mayor que 29,819685) difiere exageradamente del valor observado \bar{X} .

1.2. σ^2 desconocido

Se reemplaza σ^2 por su estimador insesgado

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Pivote	Intervalo
$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$	$IC(\mu) = \bar{X} \pm \frac{S}{\sqrt{n}} \cdot t_{n-1, 1-\alpha/2}$

usar

$$\frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}; \text{ si la población es finita}$$

Notas:

a) Si $n > 30$ se puede aproximar $t_{n-1, 1-\alpha/2}$ por $Z_{1-\alpha/2}$ pues la distribución t-Student converge a la distribución Normal.

Ejemplo:

El tiempo (en minutos) que tardaron 15 operarios para familiarizarse con el manejo de una máquina moderna adquirida por la empresa fue: 3.4, 2.8, 4.4, 2.5, 3.3, 4, 4.8, 2.9, 5.6, 5.2, 3.7, 3, 3.6, 2.8, 4.8. Suponga que los tiempos se distribuyen normalmente.

- a) Determine e interprete un intervalo del 95% de confianza para el verdadero tiempo promedio.
- b) El instructor considera que el tiempo promedio requerido por la población de trabajadores que recibe instrucción sobre esta máquina es superior a 5 minutos, ¿qué se puede decir de acuerdo con el intervalo hallado?

Solución

a) *INTERVALO*: $\bar{x} \pm t_{(n-1, 1-\alpha/2)} \frac{s}{\sqrt{n}}$

reemplazando los datos: $3'8 \pm (2145) \left(\frac{0'971}{\sqrt{15}} \right)$
 $[3'26, 4'34]$

Interpretación:

Tenemos un 95% de certeza de que el verdadero tiempo promedio que requieren los operarios para familiarizarse con la máquina está entre 3'26 y 4'34 minutos.

b) La apreciación del instructor no parece ser correcta ya que el promedio 5 minutos está fuera del intervalo hallado (aunque debería analizarse con un I.C. unilateral)

2. Comparación de Dos Medias

Sean X_1, X_2, \dots, X_{n_1} muestras aleatorias de una población Normal $N(\mu_1, \sigma_1^2)$ e Y_1, Y_2, \dots, Y_{n_2} muestras aleatorias de una población Normal $N(\mu_2, \sigma_2^2)$. Ambas muestras independientes. Las medias poblacionales μ_1 y μ_2 pueden ser comparadas usando la diferencia $\mu_1 - \mu_2$, cuyo estimador es $\hat{\mu}_1 - \hat{\mu}_2 = \bar{X} - \bar{Y}$.

2.1. σ_1^2 y σ_2^2 conocidas

Pivote	Intervalo
$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$	$IC(\mu_1 - \mu_2) = (\bar{X} - \bar{Y}) \pm \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \cdot Z_{1-\alpha/2}$

2.2. σ_1^2 y σ_2^2 desconocidas pero iguales ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

Se usa el estimador de σ^2 dado por:

$$\hat{\sigma}^2 = S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$$

$$S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$$

Pivote	Intervalo
$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$	$IC(\mu_1 - \mu_2) = (\bar{X} - \bar{Y}) \pm S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot t_{n_1 + n_2 - 2, 1 - \alpha/2}$

Ejemplo:

Se registraron los siguientes datos, en minutos, que tardan algunos hombres y mujeres en realizar cierta actividad en una empresa, los cuales fueron seleccionados aleatoriamente

HOMBRES MUJERES

$n_1=14$ $n_2=25$

Media=17 Media=19

Varianza=1,5 Varianza=1,8

Suponga que los tiempos para los dos grupos se distribuyen normalmente y que las varianzas son iguales, aunque desconocidas.

a) Calcule e interprete un intervalo de confianza del 99% para la verdadera diferencia de medias.

b) De acuerdo con el intervalo hallado, ¿hay evidencia de que los dos tiempos promedio son iguales?

Solución

$$a) \text{ INTERVALO : } (\bar{x} - \bar{y}) \pm t_{(k, 1-\alpha/2)} S_p \sqrt{1/n_x + 1/n_y}$$

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

$$\text{reemplazando los datos : } (19 - 17) \pm (2.704)(1.3) \sqrt{1/25 + 1/14}$$
$$[1.61, 2.39]$$

Interpretación:

Tenemos una certeza del 99% de que la verdadera diferencia promedio de tiempo se encuentra entre 1'61 y 2'39 minutos.

b) Como el cero no está contenido en el intervalo, estos datos no evidencian una igualdad entre las medias.

2.3. σ_1^2 y σ_2^2 desconocidas y distintas

Pivote	Intervalo
$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_v$	$IC(\mu_1 - \mu_2) = (\bar{X} - \bar{Y}) \pm \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \cdot t_{v, 1-\alpha/2}$

$$\text{con } v = (S_1^2/n_1 + S_2^2/n_2)^2 / \left[\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1} \right]$$

3. Comparación de Dos Medias para datos Pareados

Sean los pares $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ donde X_1, X_2, \dots, X_n puede ser considerada una muestra aleatoria de una población normal $N(\mu_1, \sigma_1^2)$ e Y_1, Y_2, \dots, Y_n una muestra aleatoria de una población normal $N(\mu_2, \sigma_2^2)$. Se supone que X_i e Y_i no son independientes pero si lo son (X_i, Y_i) de (X_j, Y_j) , $i \neq j$. Se trata de hacer inferencia respecto a $\mu_1 - \mu_2$.

Pivote	Intervalo
$F = \frac{\bar{d} - \mu_D}{S_D/\sqrt{n}} \sim t_{n-1}$	$IC(\mu_1 - \mu_2) = \bar{d} \pm \frac{S_D}{\sqrt{n}} \cdot t_{n-1, 1-\alpha/2}$

$$\begin{aligned}
 \mu_D &= \mu_1 - \mu_2 \\
 d_i &= X_i - Y_i \\
 \bar{d} &= \frac{1}{n} \sum_{i=1}^n d_i = \bar{X} - \bar{Y} \\
 S_D^2 &= \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2
 \end{aligned}$$

Ejercicio

10 personas fueron sometidas a un test antes y después de recibir cierta instrucción los resultados fueron como sigue:

Individuo	1	2	3	4	5	6	7	8	9	10
Antes	70	84	88	110	105	100	110	67	79	86
Después	115	148	176	191	158	178	179	140	161	157

¿Proporcionan estos datos evidencia suficiente para decir que la instrucción fue efectiva? Tome un nivel de confianza del 99%.

4. Estimación de la Varianza

Suponga que X_1, X_2, \dots, X_n es una muestra aleatoria de una población normal $N(\mu, \sigma^2)$ y que se quiere obtener un intervalo de confianza del $(1 - \alpha)100\%$ para σ^2 cuyo estimador es:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

4.1. μ desconocido

Pivote	Intervalo
$X = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi_{n-1}^2$	$IC(\sigma^2) = \left[\frac{(n-1)S^2}{\chi_{n-1, 1-\alpha/2}^2} ; \frac{(n-1)S^2}{\chi_{n-1, \alpha/2}^2} \right]$

4.2. μ conocido

Pivote	Intervalo
$X = \frac{\sum (X_i - \mu)^2}{\sigma^2} \sim \chi_n^2$	$IC(\sigma^2) = \left[\frac{\sum (X_i - \mu)^2}{\chi_{n, 1-\alpha/2}^2} ; \frac{\sum (X_i - \mu)^2}{\chi_{n, \alpha/2}^2} \right]$

Ejemplo:

Un ingeniero de control de la calidad midió el espesor de la pared de 25 botellas de vidrio de 2 litros. La media muestral resultó 4.05 m.m. y la desviación típica 0.08 mm. Obtén un intervalo de confianza al 90% para la variabilidad del espesor de la pared de las botellas

Solución

$X = \text{Espesor} \rightarrow N(\mu, \sigma)$, con μ, σ desconocidas

El I.C. para σ^2 al 90 % es:

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}; n-1}}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}; n-1}} \right] = \left[\frac{24(0,08)^2}{30,15}, \frac{24(0,08)^2}{10,11} \right] = [0,005, 0,015]$$

5. Comparación de Dos Varianzas

Sea X_1, X_2, \dots, X_{n_1} una muestra aleatoria de una población normal $N(\mu_1, \sigma_1^2)$ e Y_1, Y_2, \dots, Y_{n_2} otra muestra aleatoria de una población normal $N(\mu_2, \sigma_2^2)$. Ambas muestras independientes con μ_1 y μ_2 desconocidas. Las varianzas poblacionales σ_1^2 y σ_2^2 pueden ser comparadas usando el cociente cuyo estimador es:

$$\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} = \frac{S_1^2}{S_2^2}$$

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$$

$$S_2^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$$

Pivote	Intervalo
$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F_{n_1-1, n_2-1}$	$IC(\sigma_2^2/\sigma_1^2) = \left[\frac{S_2^2}{S_1^2} \cdot F_{n_1-1, n_2-1, \alpha/2} ; \frac{S_2^2}{S_1^2} \cdot F_{n_1-1, n_2-1, 1-\alpha/2} \right]$

Ejemplo:

Se piensa que la concentración del ingrediente activo de un detergente líquido para ropa está afectado por el tipo de catalizador utilizado en el proceso de fabricación. Por experiencias anteriores se supone que la desviación estándar de la concentración activa es de 3 g/l, sin importar el tipo de catalizador utilizado. Se toman 10 observaciones con cada catalizador y se obtienen los siguientes datos:

Cat.1	57.9	66.2	65.4	65.4	65.2	62.6	67.6	63.7	67.2	71
Cat.2	66.4	71.7	70.3	69.3	64.8	69.6	68.6	69.4	65.3	68.8

Obtén un intervalo de confianza al 90% para el cociente de varianzas?. ¿Puede suponerse la misma variabilidad en la concentración con el empleo de ambos catalizadores?.

Solución

X = Concentración con catalizador 1 $\rightarrow N(\mu_1, \sigma_1)$

Y = Concentración con catalizador 2 $\rightarrow N(\mu_2, \sigma_2)$,

son independientes y todos los parámetros se desconocen.

El I.C. para $\frac{\sigma_2^2}{\sigma_1^2}$ al 90 % es:

$$\left[\frac{S_2^2 F_{0,05;9,9}}{S_1^2}, \frac{S_2^2 F_{0,95;9,9}}{S_1^2} \right] = \left[\frac{4,946(0,314)}{13,343}, \frac{4,946(3,18)}{13,343} \right] = [0,116, 1,180],$$

Al estar el 1 contenido en el intervalo, las varianzas podrían considerarse iguales.

6. Estimación de una Proporción

Sea X_1, X_2, \dots, X_n una muestra aleatoria de una población Bernoulli de parámetro p cuyo estimador es:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i, \quad X_i = 0, 1, \quad 0 \leq p \leq 1$$

Usando el Teorema del Limite Central se puede obtener un intervalo aproximado para p si n es grande ($n > 30$).

$$IC(p) = \hat{p} \pm \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \cdot Z_{1-\alpha/2}$$

reemplazando en el denominador del pivote p por su estimador.

Observación

Usar $\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \cdot \sqrt{\frac{N - n}{N - 1}}$; si la población es finita

$$P \left(-Z_{\alpha/2} \leq \frac{\frac{X}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq Z_{\alpha/2} \right) \cong 1 - \alpha$$

Intervalos de confianza unilaterales

1.- El intervalo de confianza unilateral hacia la izquierda del $(1-\alpha)\times 100\%$ para el parámetro p es

$$I = < 0; \hat{p} + Z_{\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} >$$

2.- Análogamente, el intervalo de confianza unilateral hacia la derecha del $(1-\alpha)\times 100\%$ para el parámetro p es

$$I = < \hat{p} - Z_{\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; 1 >$$

Ejemplo:

Para poder controlar la fabricación de un producto se toman 85 muestras de un determinado componente y se concluye que 10 de ellos no cumplen las especificaciones. Calcula un intervalo de confianza al 95% para la proporción de defectuosos.

Solución

$$X = \text{N}^{\circ} \text{ de defectuosos} \rightarrow B(85, p)$$

El I.C. para p al 95 % es:

$$\left[\hat{p} \pm z_{0,975} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right] = \left[0,118 \pm 1,96 \sqrt{\frac{0,118(1 - 0,118)}{85}} \right] = [0,05, 0,186]$$

Ejemplo:

La reacción de un pequeño inversionista con respecto a un cambio de política que modifique las reglas de juego de la economía puede adoptar una de dos formas: retirar su inversión (A) o continuar con la inversión iniciada (B). Un investigador quiere estimar la probabilidad p de que un inversionista reaccione de manera A.

- a) Suponiendo que el investigador estará satisfecho cuando su error de estimación sea menor que 0,04 con una probabilidad igual a 0,90, ¿cuántos pequeños inversionistas debe incluir en la encuesta?. Considérese también que el investigador espera que p tenga el valor 0,6.
- b) El investigador realiza la encuesta con una m.a. de pequeños inversionistas de tamaño igual al que obtuvo en a) y obtiene que 60% de dichas personas reaccionan de manera A.
Obtenga un valor con el cual el investigador tenga el 95% de confianza de que la proporción poblacional no exceda de dicho valor.

Solución

- a) Para $\gamma = 1 - \alpha = 0,90$ y $\alpha / 2 = 0,05$, en la Tabla encontramos $Z_{\alpha/2} = 1,645$

Al utilizar $\hat{p} = 0,60$ como el estimador inicial de p y el error de estimación $e = 0,04$, el tamaño requerido de la muestra es

$$n = \frac{(Z_{\alpha/2})^2 \hat{p}(1 - \hat{p})}{e^2} = \frac{(1,645)^2 (0,6)(0,4)}{(0,04)^2} = 406$$

Por tanto, el investigador debe incluir en la encuesta a 406 pequeños inversionistas.

- b) Ahora, para el coeficiente de confianza fijado $\gamma = 1 - \alpha = 0,95$, en la Tabla encontramos $Z_{\alpha} = 1,645$

Luego, el intervalo de confianza del 95% para p es

$$I = \left(\hat{p} - Z_{\alpha} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + Z_{\alpha} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right) = \left(0,6 - (1,645) \sqrt{\frac{(0,6)(0,4)}{406}}, 0,6 + (1,645) \sqrt{\frac{(0,6)(0,4)}{406}} \right) = (0,53999953, 0,63999953)$$

Por tanto, el valor pedido es 0,63999953

7. Comparación de Proporciones

Sean X_1, X_2, \dots, X_{n1} e Y_1, Y_2, \dots, Y_{n2} muestras aleatorias independientes de poblaciones Bernoulli de parámetros p_1 y p_2 respectivamente. Las proporciones poblacionales pueden ser comparadas a través de un intervalo de confianza aproximado para muestras grandes, dado por:

$$IC(p_1 - p_2) = (\hat{p}_1 - \hat{p}_2) \pm \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \cdot Z_{1-\alpha/2}$$

Ejemplo

En Lima se toma una muestra aleatoria de 100 votantes y se encuentra que 25 de ellos están a favor de un candidato al Congreso. En Arequipa se toma una muestra aleatoria de 100 votantes y se encuentra que 45 de ellos están a favor del candidato. Estimar con un intervalo de 95% de confianza la diferencia de proporciones.

Solución

Por conveniencia especificamos:

En ciudad 1 (Arequipa) se tiene: $n_1 = 100$; $\hat{p}_1 = 0,45$; $\hat{q}_1 = 0,55$; $z_{\text{tabla}} = 1,96$;

En ciudad 2 (Lima) se tiene: $n_2 = 100$; $\hat{p}_2 = 0,25$; $\hat{q}_2 = 0,75$ y

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1 \cdot \hat{q}_1}{n_1} + \frac{\hat{p}_2 \cdot \hat{q}_2}{n_2}} = \sqrt{\frac{0,45 \cdot 0,55}{100} + \frac{0,25 \cdot 0,75}{100}} = 0,06595; \text{ por lo tanto:}$$

$$\text{Límite inferior: } (\hat{p}_1 - \hat{p}_2) - z_{\text{tabla}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} = 0,20 - 0,1293 = 0,0707$$

$$\text{Límite superior: } (\hat{p}_1 - \hat{p}_2) + z_{\text{tabla}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} = 0,20 + 0,1293 = 0,3293$$

Se estima con un 95% de confianza que en Arequipa el % a favor del candidato es mayor que en Lima, en un % que va entre 7,07% y 32,93%.

Si se diera que el intervalo contiene al cero (0), entonces se dice que no hay diferencias en los % a favor en las dos ciudades.

8. Estimación de Cualquier Parámetro

Si X_1, X_2, \dots, X_n es una muestra aleatoria de una población X con función de densidad $f(x, \theta)$ y $\hat{\theta}_{EMV}$ es el estimador máximo verosímil de θ entonces se pueden usar las propiedades de normalidad asintótica de estos estimadores para obtener un intervalo aproximado para θ .

Como $\frac{\hat{\theta}_{EMV} - \theta}{1/\sqrt{nI_I(\theta)}}$ tiende a $N(0, 1)$ cuando $n \rightarrow \infty$. Con $I_I(\theta) = -E\left(\frac{\partial^2}{\partial \theta^2} \ln(f(x, \theta))\right)$, entonces el intervalo de confianza asintótico es:

$$IC(\theta) = \hat{\theta}_{EMV} \pm \frac{1}{\sqrt{nI_I(\hat{\theta})}} \cdot Z_{1-\alpha/2}$$

Ejemplo

Un ingeniero analiza la resistencia a la compresión del concreto. De una m.a de 49 especímenes se obtuvo una resistencia promedio de 3250 psi y una desviación estándar de 31.62 psi. Construya un I.C al 95% para la resistencia media a la compresión de este concreto.

Solución

Suponga que X_1, \dots, X_{49} es una m.a que representa las resistencias a la compresión de 49 especímenes de este concreto y suponga que $E[X_i] = \mu$ y $V[X_i] = \sigma^2$, $i = 1, \dots, 49$. Un I. C. aproximado al $100(1-\alpha)\%$ para μ es:

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Como σ^2 es desconocida, y los estadísticos son:

$$\bar{X} = 3250, \quad S^2 = (31.62)^2 \quad n = 49$$

Si $\alpha = 0.05 \Rightarrow Z_{\alpha/2} = 1.96$ y así, un IC al 95% para μ es $\bar{X} \pm \frac{(1.96)(31.62)}{\sqrt{49}}$

$$3250 \pm 8.854 \Leftrightarrow (3241.146, 3258.854)$$

Determinación del tamaño de muestra n para estimar la media poblacional (N infinito)

$E = Z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ es la mitad del ancho del intervalo de confianza (producto del coeficiente y el error estándar) y se denomina error máximo de estimación E .

Dado un valor de error y un cierto nivel de confianza, podemos estimar cuál sería el tamaño de la muestra

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

Ejemplo:

Se desea estudiar la variable *altura de los individuos de una población*, considerando que ésta es una variable que se distribuye de modo gaussiana. Para ello se tomó una muestra de 25 individuos (que podemos considerar piloto), que ofreció los siguientes resultados:

$$\bar{x} = 170 \text{ cm}$$

$$S = 10,206 \text{ cm}$$

Calcular el tamaño que debería tener una muestra para que se obtuviese un intervalo de confianza para la media poblacional con un nivel de confianza al 99% y con una precisión de $E=1$ cm.

Solución

En este caso se obtiene:

$$n \approx \frac{z_{0,995}^2 \cdot 10,206^2}{1^2} = 2,58^2 \cdot 10,206^2 \approx 694$$

Por tanto, si queremos realizar un estudio con toda la precisión requerida en el enunciado se debería tomar una muestra de 694 individuos. Esto es una indicación de gran utilidad antes de comenzar el estudio. Una vez que el muestreo haya sido realizado, debemos confirmar que el error para el nivel de significación dado es inferior o igual a 1 cm, utilizando la muestra obtenida.

Determinación del tamaño de muestra n para estimar la media poblacional (N finito)

$E = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ es la mitad del ancho del intervalo de confianza (producto del coeficiente y el error estándar) por el factor de corrección para población finita y se denomina error máximo de estimación E.

Dado un valor de error y un cierto nivel de confianza, podemos estimar cuál sería el tamaño de la muestra

$$n = \frac{z_{1-\alpha/2}^2 \sigma^2 N}{z_{1-\alpha/2}^2 \sigma^2 + E^2 (N - 1)}$$

Ejemplo:

Interés: Conocer el salario semanal promedio de trabajadores en una zona de cierto distrito de Lima (se asume que son 1800 trabajadores).

El estudio queremos realizarlo mediante una muestra y necesitamos calcular el tamaño de muestra n , considerando un grado de confianza del 95%. Los resultados de un estudio preliminar proporciona un promedio de 210 nuevos soles y una desviación estándar de 30 nuevos soles semanal. Con un error relativo del 6%, obtener el tamaño de n .

Solución

Definición de error relativo:

$$Er = (E/\text{media}) * 100 \Rightarrow E = (Er * \text{media}) / 100$$

Por consiguiente:

$$E = (6 * 210) / 100 = 12.6$$

Grado de confianza 95%, nos indica que $Z=1.96$.

Reemplazando valores en la formula se tiene:

$$n = [(1.96)^2(30)^2(1800)] / [1.96^2(30)^2 + (12.6)^2(1799)] = 22$$

Por tanto, $n=22$ se requiere como mínimo 22 trabajadores.

Determinación del tamaño de muestra n para estimar la proporción poblacional (N infinito)

$E = z_{1-\alpha/2} \sqrt{\frac{PQ}{n}}$ es la mitad del ancho del intervalo de confianza (producto del coeficiente y el error estándar) y se denomina error máximo de estimación E.

Dado un valor de error y un cierto nivel de confianza, podemos estimar cuál sería el tamaño de la muestra

$$n = \frac{z_{1-\alpha/2}^2 PQ}{E^2}$$

Ejemplo:

Para poder controlar la fabricación de un producto se toman 85 muestras de un determinado componente y se concluye que 10 de ellos no cumplen las especificaciones. ¿Cuál debería ser el tamaño de la muestra si se quiere que el error cometido al estimar la proporción sea menor de 0.05 con una probabilidad 0.95?.

Solución

Puesto que \hat{p} es el estimador puntual de p , puede definirse el error cometido al estimar p por \hat{p} como $E = |p - \hat{p}|$. Si el I.C. al $(1-\alpha)100\%$ para p es $\left[\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$,

eso significa que el error de estimación E es menor o igual que $z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ con una probabilidad de $(1-\alpha)$. En consecuencia, el tamaño de muestra n para obtener un error en la estimación inferior o igual a E con una probabilidad $(1-\alpha)$ debe ser:

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

En este caso:

$$n = \left(\frac{1,96}{0,05} \right)^2 0,118(1-0,118) = 160$$

Determinación del tamaño de muestra n para estimar la proporción poblacional (N finito)

$E = z_{1-\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \sqrt{\frac{N-n}{N-1}}$ es la mitad del ancho del intervalo de confianza (producto del coeficiente y el error estándar) y se denomina error máximo de estimación E.

Dado un valor de error y un cierto nivel de confianza, podemos estimar cuál sería el tamaño de la muestra

$$n = \frac{z_{1-\alpha/2}^2 N \hat{p} \hat{q}}{z_{1-\alpha/2}^2 \hat{p} \hat{q} + E^2 (N - 1)}$$

Ejemplo:

El Director de un Colegio Nacional desea calcular la proporción de los 1000 alumnos de último año que piensan seguir estudios en la universidad.

¿Qué tamaño debe tener la muestra que necesita tomar el director si su estimación debe estar a 0,04 del valor verdadero, con 99% de confianza? . El año anterior, el 70% de los alumnos encuestados dijeron que tenían planeado seguir estudios en la universidad.

Solución:

Para $\gamma = 1 - \alpha = 0,99$ y $\frac{\alpha}{2} = 0,005$, en la Tabla se encuentra $Z_{\alpha/2} = 2,5757$

Al utilizar $\hat{p} = 0,70$ como un estimado inicial de p , el tamaño requerido de la muestra es:

$$n = \frac{Z_{\alpha/2}^2 \hat{p}(1 - \hat{p})N}{Z_{\alpha/2}^2 \hat{p}(1 - \hat{p}) + E^2(N - 1)} = \frac{(2,5757)^2 (0,7)(0,3)(1000)}{(2,5757)^2 (0,7)(0,3) + (0,04)^2 (999)} = 465,70$$

Así pues, el director del colegio debe tomar una muestra de tamaño $n = 466$

Ejercicio

Para calibrar un nuevo aparato de medida, un investigador realizó 15 mediciones, en las que obtuvo los siguientes errores:

-0.10 -0.15 0.00 0.50 0.10 -0.20 -0.15 0.20 0.25 0.30 -0.45 -0.40
0.35 0.25 -0.50

Suponiendo que los errores se distribuyen según una ley normal,

- (a) calcula un intervalo de confianza del error medio de medida con coeficiente de confianza 0,99.
- (b) Sabiendo que la desviación típica del error es menor o igual a 0,4, ¿cuántas medidas serán necesarias para obtener con probabilidad 99% una estimación del error medio de medida con un error máximo menor o igual a 0,1?
- (c) Suponiendo que un aparato de medida se considera admisible si la desviación típica de los errores de medida que produce es menor o igual que 0,3, determina si dicho aparato puede ser admisible al nivel de confianza del 99%.