Tema 7. Estimación por intervalos de confianza

Eva Romero Ramos
evarom03@ucm.es

Universidad Complutense de Madrid

Outline

Outline

Intervalo de confianza

- Cuando se particulariza el estimador θ^* para una muestra X concreta, es decir, cuando se obtiene la estimación puntual para una muestra determinada, no se sabe si la estimación obtenida está o no próxima al verdadero valor del parámetro θ, debido a la aleatoriedad del muestreo.
- Por ello, lo que se hace es acompañar toda estimación puntual, de un intervalo, donde confiamos que esté incluido el verdadero valor de θ.
- Este intervalo, denominado **intervalo de confianza** presenta dos límites, ambos funciones de la información muestral y tales que:

$$\boxed{P\left[\underline{\theta}(X) \leq \theta \leq \overline{\theta}(X)\right] = 1 - \alpha}$$

- $1 - \alpha$ se llama nivel de confianza

Interpretación

- La interpretación correcta de la expresión es que $1-\alpha$ es la probabilidad de que el intervalo aleatorio

$$\left[\underline{\theta}\left(X\right);\overline{\theta}\left(X\right)\right]$$

  incluya el verdadero valor del parámetro θ antes de extraer la muestra, es decir:

$$P\left\{\theta \in \left[\underline{\theta}\left(X\right);\overline{\theta}\left(X\right)\right]\right\} = 1-\alpha$$

- Seleccionada una muestra concreta X, la probabilidad de que el parámetro θ esté incluido en el intervalo es 1 ó 0, según si el parámetro está o no entre los dos números en que se convierten $\underline{\theta}$ y $\overline{\theta}$ para esta muestra X concreta
- Otra forma de interpretarlo: Si se consideraran todas las muestras posibles X, en $100\left(1-\alpha\right)$ de cada 100 casos, el parámetro θ pertenecerá al intervalo.

Ejemplo

Un estudiante se examina de un programa de 100 lecciones, de esas lecciones hay sólo una que no sabe, no recuerda exactamente el número de esa lección, tendrá que mirarlo en el programa. Se examina sacando una bola y desarrollando la lección que le ha correspondido. Si la sabe, aprueba y si no, le suspenden. Este estudiante antes de sacar la bola tiene una probabilidad de 0.99 de aprobar. Supongamos que comparece ante el tribunal y saca una bola, la probabilidad que tiene ya en ese momento de aprobar no es de 0.99, es de uno si la lección que ha salido es de las que se sabe y de cero si es la que no se sabe, ahora bien, ***mientras consulta el programa él tiene una confianza del 0.99 de que aprobará****, ya que antes de sacar la bola tenía una probabilidad de 0.99 de obtener una bola cuya lección correspondiente se sabía.*

Outline

IC para la media de población normal con varianza conocida

- Recordemos que en este caso:

$$\overline{X} \sim \mathcal{N}\left(\mu; \frac{\sigma^2}{n}\right)$$

- Por tanto, el intervalo de confianza para el parámetro μ de nivel de confianza $1-\alpha$ vendrá dado por la expresión:

$$\boxed{\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}$$

donde $z_{\alpha/2}$ es el valor de la tabla $\mathcal{N}(0;1)$ tal que:

$$P\left(\mathcal{N}(0;1) > z_{\alpha/2}\right) = \frac{\alpha}{2}$$

Ejemplo

De una población $\mathcal{N}(\mu;9)$ *se extrae una muestra aleatoria simple de tamaño* 25, *siendo* $\sum_{i=1}^{25} x_i = 60$. *Determénese el intervalo de confianza del* 95% *para la media de la población.*

Ejemplo

De una población $\mathcal{N}(\mu; 9)$ *se extrae una muestra aleatoria simple de tamaño 25, siendo* $\sum_{i=1}^{25} x_i = 60$. *Determénese el intervalo de confianza del* 95% *para la media de la población.*

Solución

La media muestral es $\overline{X} = \frac{1}{n}\sum_{i=1}^{25} x_i = \frac{1}{25}60 = 2,4$ *y* $\alpha = 1 - 0,95 = 0,05$

$$P\left(\mathcal{N}(0;1) > z_{0,025}\right) = \frac{0,05}{2} = 0,025$$

$\Longrightarrow z_{0,025} = 1,96$

El intervalo de confianza para el parámetro μ *de nivel de confianza* 95% *es*

$$\left[\overline{X} - z_{0,025}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{0,025}\frac{\sigma}{\sqrt{n}}\right] = \left[2,4 - 1,96\frac{3}{5}; 2,4 + 1,96\frac{3}{5}\right] =$$

$$= \left[1,224; 3,576\right]$$

IC para la media de una población normal con varianza desconocida

- Recordemos que en este caso:

$$\frac{\sqrt{n}\left(\overline{X}-\mu\right)}{S'} \sim t\left(n-1\right)$$

- El intervalo de confianza para el parámetro μ de nivel de confianza $1-\alpha$ es

$$\boxed{\left[\overline{X}-t_{n-1,\alpha/2}\frac{S'}{\sqrt{n}};\overline{X}+t_{n-1,\alpha/2}\frac{S'}{\sqrt{n}}\right]}$$

siendo $t_{n-1,\alpha/2}$ tal que:

$$P\left(-t_{n-1,\alpha/2}<t\left(n-1\right)<t_{n-1,\alpha/2}\right)=1-\alpha$$

Ejemplo

En una población $\mathcal{N}\left(\mu;\sigma^2\right)$*, donde se desconoce el valor del parámetro* σ^2*, se obtiene una muestra aleatoria simple de tamaño* 10*, en la que* $\sum_{i=1}^{10} x_i = 41$ *y* $\sum_{i=1}^{10} x_i^2 = 229$*. Hállese el intervalo de confianza del* 90% *para la media de la población.*

Ejemplo

En una población $\mathcal{N}\left(\mu;\sigma^2\right)$*, donde se desconoce el valor del parámetro* σ^2*, se obtiene una muestra aleatoria simple de tamaño* 10*, en la que* $\sum_{i=1}^{10} x_i = 41$ *y* $\sum_{i=1}^{10} x_i^2 = 229$*. Hállese el intervalo de confianza del* 90% *para la media de la población.*

Solución

La media muestral es $\overline{X} = \frac{1}{10}\sum_{i=1}^{10} x_i = 4,1$*, la varianza muestral es* $S^2 = \frac{1}{10}\sum_{i=1}^{10} x_i^2 - \overline{X}^2 = 6,09 \rightarrow S'^2 = 10 \cdot 6,09/9 = 6,77$ *y* $\alpha = 1 - 0,90 = 0,10$

$$P\Big(-t_{9;0,05} < t\,(9) < t_{9;0,05}\Big) = 0,9 \implies P\Big(t\,(9) > t_{9;0,05}\Big) = 0,05$$

$\implies t_{9;0,05} = 1,833$

El intervalo de confianza para el parámetro μ *de nivel de confianza* 90% *es*

$$\left[4,1 - 1,833\frac{\sqrt{6,77}}{\sqrt{10}}; 4,1 + 1,833\frac{\sqrt{6,77}}{\sqrt{10}}\right] = \left[2,59; 5,61\right]$$

Outline

IC para la diferencia de medias en poblaciones normales con σ^2 conocida

- Sabemos que:

$$\overline{X}_1 \sim \mathcal{N}\left(\mu_1; \frac{\sigma_1^2}{n}\right) \quad \& \quad \overline{X}_2 \sim \mathcal{N}\left(\mu_2; \frac{\sigma_2^2}{m}\right)$$

- Entonces:

$$\overline{X}_1 - \overline{X}_2 \sim \mathcal{N}\left(\mu_1 - \mu_2; \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

- El intervalo de confianza para $\mu_1 - \mu_2$ a nivel de confianza $1-\alpha$ es

$$\boxed{\left[\overline{X}_1 - \overline{X}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}; \overline{X}_1 - \overline{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right]}$$

donde $z_{\alpha/2}$ es el valor de la tabla $\mathcal{N}(0;1)$ tal que:

$$P\left(\mathcal{N}(0;1) > z_{\alpha/2}\right) = \frac{\alpha}{2}$$

Ejemplo

Se desea comparar el tiempo que tardan en obtener el título dos grupos de estudiantes universitarios y para ello se cuenta con un grupo de 40 estudiantes de la Facultad de Ingeniería y un grupo de 50 estudiantes de la Facultad de Ciencias. Se sabe que la varianza del tiempo que tardan en obtener el título en la Facultad de Ingeniería es de 2 años y la varianza para la Facultad de Ciencias es de 3,5 años.

Los datos muestrales indican que la media del tiempo que tardan en obtener el título de los estudiantes de Ingeniería es de 5 años y la media para los estudiantes de la Facultad de Ciencias es de 4 años.

Se desea construir un intervalo de confianza al 99% para la diferencia entre las medias del tiempo que tardan en obtener el título ambos grupos de estudiantes.

Solución

Se puede calcular el intervalo de confianza para la diferencia de medias en poblaciones normales con varianzas conocidas mediante:

$$\left(\bar{X}_1 - \bar{X}_2 - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ \bar{X}_1 - \bar{X}_2 + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Teniendo en cuenta que $z_{\alpha/2} = z_{0,005} = 2,58$ *y sustituyendo los valores correspondientes tenemos que:*

$$\left(5 - 4 - 2.58\sqrt{\frac{2}{40} + \frac{3,5}{50}};\ 5 - 4 + 2,58\sqrt{\frac{2}{40} + \frac{3,5}{50}}\right)$$

Luego el intervalo de confianza para la diferencia de medias del tiempo que tardan en obtener el título los estudiantes de ingeriería y los de ciencias, a nivel de confianza 99% es:

$$\boxed{(0,1062;\ 1,8937)}$$

IC para $\mu_1 - \mu_2$ en poblaciones normales con varianzas desconocidas pero iguales $\sigma_1^2 = \sigma_1^2$

- En este caso usaremos el estadístico:

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_c\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t\left(n+m-2\right)$$

- Siendo:

$$S_c = \sqrt{\frac{(n-1)\cdot S_1'^2 + (m-1)\cdot S_2'^2}{n+m-2}}$$

- El intervalo de confianza para $\mu_1 - \mu_2$ a nivel de confianza $1-\alpha$ es

$$\boxed{\left[\overline{X}_1 - \overline{X}_2 \pm t_{n+m-2,\alpha/2} \cdot S_c\sqrt{\frac{1}{n} + \frac{1}{m}}\right]}$$

siendo $t_{n+m-2,\alpha/2}$ tal que:

$$P\Big(t\left(n+m-2\right) > t_{n+m-2,\alpha/2}\Big) = \alpha/2$$

Ejemplo

Supongamos que se quiere comparar el tiempo de respuesta de dos sistemas de almacenamiento de datos. Se toma una muestra aleatoria de 30 consultas para el sistema 1 y se obtiene una media muestral de $\bar{X}_1 = 4,8$ *segundos y una cuasivarianza de* $S_1'^2 = 1,2^2$. *Asimismo, se toma una muestra aleatoria de 25 consultas para el sistema 2 y se obtiene una media muestral de* $\bar{X}_2 = 5,2$ *segundos y una cuasivarianza de* $S_2'^2 = 1,5^2$. *Se asume que las varianzas poblacionales son desconocidas pero iguales. Obtenga un intervalo de confianza del 99% para la diferencia de medias poblacionales.*

Solución

Como las varianzas poblaciones son desconocidas pero se asumen iguales, usaremos las siguiente expresión para calcular el intervalo de confianza para $\mu_1 - \mu_2$*:*

$$\left[\overline{X}_1 - \overline{X}_2 \pm t_{n+m-2,\alpha/2} \cdot S_c \sqrt{\frac{1}{n} + \frac{1}{m}}\right]$$

Donde:

$$S_c = \sqrt{\frac{(30-1)\cdot 1,2^2 + (25-1)\cdot 1,5^2}{30+25-2}} = 1,34$$

Luego:

$$\left[4,8-5,2-2,675\cdot 1,34\sqrt{\frac{1}{30}+\frac{1}{25}}; 4,8-5,2+2,675\cdot 1,34\sqrt{\frac{1}{30}+\frac{1}{25}}\right]$$

$$\boxed{(-1,3707;\ 0,5707)}$$

IC para $\mu_1 - \mu_2$ en poblaciones normales con varianzas desconocidas y distintas $\sigma_1^2 \neq \sigma_1^2$

- En este caso usaremos el estadístico:

$$\frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1'^2}{n} + \frac{S_2'^2}{m}}} \sim t\,(k)$$

- Siendo:

$$k = \frac{\left(S_1'^2/n + S_2'^2/m\right)^2}{\left[\frac{(S_1'^2/n)^2}{n-1}\right] + \left[\frac{(S_2'^2/m)^2}{m-1}\right]} - 2$$

- El intervalo de confianza para $\mu_1 - \mu_2$ a nivel de confianza $1 - \alpha$ es

$$\boxed{\left[\overline{X}_1 - \overline{X}_2 - t_{k,\alpha/2} \cdot \sqrt{\frac{S_1'^2}{n} + \frac{S_2'^2}{m}};\overline{X}_1 - \overline{X}_2 + t_{k,\alpha/2} \cdot \sqrt{\frac{S_1'^2}{n} + \frac{S_2'^2}{m}}\right]}$$

siendo $t_{k,\alpha/2}$ tal que:

$$P\left(t\,(k) > t_{k,\alpha/2}\right) = \alpha/2$$

Ejemplo

Se pretende estudiar el tiempo de computación de dos procesos de simulación que utilizan algoritmos diferentes. Para ello se lanzan ambos procesos 5 veces obteniendose los siguientes tiempos medidos en segundos:

Proceso A	*52*	*53*	*50*	*48*	*55*
Proceso B	*54*	*57*	*51*	*53*	*59*

Construir un intervalo de confianza para la diferencia entre los tiempos de ambos procesos, a nivel confianza 95%.

Ejemplo

Se pretende estudiar el tiempo de computación de dos procesos de simulación que utilizan algoritmos diferentes. Para ello se lanzan ambos procesos 5 veces obteniendose los siguientes tiempos medidos en segundos:

Proceso A	*52*	*53*	*50*	*48*	*55*
Proceso B	*54*	*57*	*51*	*53*	*59*

Construir un intervalo de confianza para la diferencia entre los tiempos de ambos procesos, a nivel confianza 95%.

Solución

$$\bar{X}_1 = \frac{\sum_{i=1}^{5} X_{1i}}{5} = 51,6 \qquad \bar{X}_2 = \frac{\sum_{i=1}^{5} X_{2i}}{5} = 54,8$$

$$S_1'^2 = \frac{\sum_{i=1}^{5} X_{1i} - \bar{X}_1}{4} = 7,3 \qquad S_2'^2 = \frac{\sum_{i=1}^{5} X_{2i} - \bar{X}_2}{4} = 10,2$$

Solución

Como las varianzas poblaciones son desconocidas y distintas, usaremos las siguiente expresión para calcular el intervalo de confianza para $\mu_1 - \mu_2$*:*

$$\left[\overline{X}_1 - \overline{X}_2 \pm t_{k,\alpha/2} \cdot \sqrt{\frac{S_1'^2}{n} + \frac{S_2'^2}{m}}\right]$$

Siendo:

$$k = \frac{\left(S_1'^2/n + S_2'^2/m\right)^2}{\left[\frac{\left(S_1'^2/n\right)^2}{n-1}\right] + \left[\frac{\left(S_2'^2/m\right)^2}{m-1}\right]} - 2 = \frac{(7,3/5 + 10,2/5)^2}{\left[\frac{(7,3/5)^2}{4}\right] + \left[\frac{(10,2/5)^2}{4}\right]} - 2 = 5,78 \approx 6$$

Luego:

$$\left[51,6 - 54,8 - 2,44 \cdot \sqrt{\frac{7,3}{5} + \frac{10,2}{5}}; 51,6 - 54,8 + 2,44 \cdot \sqrt{\frac{7,3}{5} + \frac{10,2}{5}}\right]$$

$$\boxed{(-7,7648;\ 1,3648)}$$

Outline

IC para la varianza de una población normal

- Recordemos que:

$$\frac{nS^2}{\sigma^2} \sim \chi^2\left(n-1\right)$$

- Por tanto, el intervalo de confianza para el parámetro σ^2 de nivel de confianza $1-\alpha$ queda:

$$\boxed{\left[\frac{nS^2}{k_2}; \frac{nS^2}{k_1}\right]}$$

- k_1 y k_2 se calculan así:

$$\begin{aligned} P\left[\chi^2\left(n-1\right) > k_1\right] &= 1-\frac{\alpha}{2} \\ P\left[\chi^2\left(n-1\right) > k_2\right] &= \frac{\alpha}{2} \end{aligned}$$

de forma que:

$$P\left[k_1 < \frac{nS^2}{\sigma^2} < k_2\right] = 1-\alpha$$

Ejemplo

En una población $\mathcal{N}\left(\mu;\sigma^2\right)$*, se obtiene una muestra aleatoria simple de tamaño* 10*, en la que* $\sum_{i=1}^{10} x_i = 41$ *y* $\sum_{i=1}^{10} x_i^2 = 229$*. Hállese el intervalo de confianza del* 80% *para la varianza de la población.*

Ejemplo

En una población $\mathcal{N}\left(\mu;\sigma^2\right)$*, se obtiene una muestra aleatoria simple de tamaño 10, en la que* $\sum_{i=1}^{10} x_i = 41$ *y* $\sum_{i=1}^{10} x_i^2 = 229$*. Hállese el intervalo de confianza del 80% para la varianza de la población.*

Solución

La varianza muestral es $S^2 = \frac{1}{10}\sum_{i=1}^{10} x_i^2 - \overline{X}^2 = 6,09$ *y* $\alpha = 0,20$

$$\begin{aligned} P\left[\chi^2(9) > k_1\right] &= 1 - \frac{\alpha}{2} = 0,9 \\ P\left[\chi^2(9) > k_2\right] &= \frac{\alpha}{2} = 0,1 \end{aligned}$$

$\Longrightarrow k_1 = 4,17$ *y* $k_2 = 14,68$

El intervalo de confianza para el parámetro σ^2 *de nivel de confianza 80% es*

$$\left[\frac{nS^2}{k_2}; \frac{nS^2}{k_1}\right] = \left[\frac{10\cdot 6,09}{14,68}; \frac{10\cdot 6,09}{4,17}\right] = \left[4,15; 14,60\right]$$

Outline

IC para el cociente de varianzas en poblaciones normales

- Para obtener un IC para el cociente de varianzas usaremos el siguiente estadístico:

$$\frac{\frac{n-1}{n-1}\frac{S_1'^2}{\sigma_1^2}}{\frac{m-1}{m-1}\frac{S_2'^2}{\sigma_2^2}} \sim F\left(n-1, m-1\right)$$

- Por tanto, el intervalo de confianza para $\frac{\sigma_1^2}{\sigma_2^2}$ de nivel de confianza $1-\alpha$ será:

$$\boxed{\left[\frac{S_1'^2/S_2'^2}{F_{n-1,m-1,\alpha/2}};\ \frac{S_1'^2/S_2'^2}{F_{n-1,m-1,1-\alpha/2}}\right]}$$

Donde:

$$P\left[F\left(n-1, m-1\right) > F_{n-1,m-1,\alpha/2}\right] = \frac{\alpha}{2}$$

$$P\left[F\left(n-1, m-1\right) > F_{n-1,m-1,1-\alpha/2}\right] = 1-\frac{\alpha}{2}$$

Ejemplo

Se desea comparar la variabilidad de dos grupos de estudiantes en términos de sus calificaciones en un examen. Se selecciona una muestra aleatoria de 20 estudiantes de cada grupo y se registran sus calificaciones. Los resultados se resumen en la siguiente tabla:

Grupo	*Tamaño de la muestra*	*Media muestral*	*Cuasivarianza*
Grupo 1	*20*	*75*	*100*
Grupo 2	*20*	*85*	*225*

Construye un intervalo de confianza al 90% *para el cociente de varianzas entre los grupos.*

Solución

El intervalo de confianza para $\frac{\sigma_1^2}{\sigma_2^2}$ *de nivel de confianza* $1-\alpha$ *es:*

$$\left[\frac{S_1'^2/S_2'^2}{F_{n-1,m-1,\alpha/2}};\ \frac{S_1'^2/S_2'^2}{F_{n-1,m-1,1-\alpha/2}}\right]$$

Luego:

$$\left[\frac{100/225}{F_{19;19;0,05}};\ \frac{100/225}{F_{19;19;0,95}}\right]=\left[\frac{100/225}{2,16};\ \frac{100/225}{0,466}\right]$$

$$\boxed{(0,2049;\ 0,4612)}$$

Outline

IC para la media con datos pareados en poblaciones normales

Diremos que X e Y son variables pareadas cuando son dependientes o están relacionadas.

Ejemplo.- Para ver si un tratamiento es eficaz, podemos observar la medida de interés antes y después del tratamiento para los mismos individuos.

Tendrá sentido entonces obtener una nueva variable $D = X - Y$, con la diferencia entre ambas medidas de la variable para cada individuo.

IC para la media con datos pareados en poblaciones normales

En esta situación, si ambos datos provienen de distribuciones normales, D también seguirá una normal, de forma que:

$$\frac{\sqrt{n}\left(\overline{D}-\mu_p\right)}{S'_d} \sim t\left(n-1\right)$$

El intervalo de confianza para el parámetro μ_d de nivel de confianza $1-\alpha$ es

$$\boxed{\left[\overline{D}-t_{n-1,\alpha/2}\frac{S'_d}{\sqrt{n}};\overline{D}+t_{n-1,\alpha/2}\frac{S'_d}{\sqrt{n}}\right]}$$

siendo $t_{n-1,\alpha/2}$ tal que:

$$P\Big(t\left(n-1\right)>t_{n-1,\alpha/2}\Big)=\alpha/2$$

Ejemplo

Un investigador desea determinar si un nuevo tratamiento para la hipertensión tiene un efecto significativo sobre la presión arterial. Para ello, dispone de una m.a.s con $n = 10$ *pacientes con hipertensión y a los que se les mide su presión arterial antes y después de recibir el tratamiento. Los resultados, en mmHg, se presentan a continuación:*

Paciente	Antes	Después
1	152	145
2	150	142
3	148	141
4	146	139
5	154	144
6	152	141
7	148	139
8	154	147
9	148	142
10	150	139

Utilizando un nivel de confianza del 95%, *calcula el intervalo de confianza para la diferencia media de la presión arterial antes y después del tratamiento.*

Solución

Primero, calculamos la diferencia de las mediciones antes y después del tratamiento para cada paciente:

Paciente	Antes	Después	Diferencia Después - Antes
1	152	145	-7
2	150	142	-8
3	148	141	-7
4	146	139	-7
5	154	144	-10
6	152	141	-11
7	148	139	-9
8	154	147	-7
9	148	142	-6
10	150	139	-11
		Media	-8,3
		Cuasivarianza	3,44

Solución

El intervalo de confianza para el parámetro μ_d *de nivel de confianza* $1-\alpha$ *es:*

$$\left[\overline{D} - t_{n-1,\alpha/2}\frac{S'_d}{\sqrt{n}}; \overline{D} + t_{n-1,\alpha/2}\frac{S'_d}{\sqrt{n}}\right]$$

En nuestro caso:

$$\left[-8,3 - t_{9;0,025}\frac{1,828}{\sqrt{10}}; -8,3 + t_{9;0,025}\frac{1,828}{\sqrt{10}}\right]$$

$$\left[-8,3 - 2,262\frac{1,828}{\sqrt{10}}; -8,3 + 2,262\frac{1,828}{\sqrt{10}}\right]$$

$$\boxed{\left[-9,6081; -6,9918\right]}$$

Outline

IC para la proporción

Se desea estimar ahora la proporción con la que se da una característica dicotómica que aparece en la muestra con probabilidad p.

Tenemos que,

$$\hat{p} \sim N\left(p, \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

Luego,

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

El intervalo de confianza para el parámetro p de nivel de confianza $1-\alpha$ es

$$\boxed{\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]}$$

Ejemplo

Se pretende estudiar la frecuencia de uso de una determinada red social, para ello se realiza una encuesta y se obtiene que el 15% *de los encuestados tiene cuenta activa en la plataforma. Si la muestra es de tamaño* 50*, se pide:*

a. *Obtenga un intervalo de confianza a nivel de confianza* 99% *para la proporción de individuos con cuenta activa en la plataforma.*
b. *Si se pretende obtener un intevalo de confianza a nivel de significación* $0,05$ *con una amplitud máxima de* $0,14$*, ¿Qué tamaño muestral será necesario?*

Solución

El intervalo de confianza para la proporción, para un nivel de confianza $1-\alpha$ *es:*

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

En nuestro caso:

$$\left[0,15 - 2,57\sqrt{\frac{0,15(1-0,15)}{50}}; 0,15 + 2,57\sqrt{\frac{0,15(1-0,15)}{50}}\right]$$

$$\boxed{\left[0,0202; 0,2797\right]}$$

Solución

Si queremos garantizarnos una amplitud máxima para el intervalo de $0,14$, *a nivel de significación* $0,05$ *tendrá que cumlirse que:*

$$\left(\hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)-\left(\hat{p}-z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)\leq 014$$

Entonces:

$$\left(0,15+1,96\sqrt{\frac{0,15(1-0,15)}{n}}\right)-\left(0,15-1,96\sqrt{\frac{0,15(1-0,15)}{n}}\right)\leq 014$$

$$\frac{0,9668}{\sqrt{n}}+\frac{0,9668}{\sqrt{n}}\leq 014\rightarrow\sqrt{n}\geq\frac{1,3997}{0,14}\rightarrow n\geq 99,979\approx 100$$

Outline

IC para la diferencia de proporciones

Si consideramos ahora la diferencia de proporciones, tendremos que,

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right)$$

Luego,

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0,1)$$

El intervalo de confianza para la diferencia de proporciones, $(p_1 - p_2)$ de nivel de confianza $1-\alpha$ es

$$\boxed{\left[(\hat{p}_1 - \hat{p}_2) \mp z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right]}$$

Ejemplo

En una encuesta a 100 estudiantes de la facultad de informática, se encontró que 60 utilizaban habitualmente Chrome como navegador. Se realiza la misma escuesta en la facultad de matemáticas y los resultados indican que 90 de los 115 encuestados utilizan el citado navegador de forma habitual.

Encuentre un intervalo de confianza al 95% para la diferencia de proporciones de uso de Chrome en ambas facultades.

Solución

El intervalo de confianza para la diferencia de proporciones a nivel de confianza $1-\alpha$ *es:*

$$\left[(\hat{p}_1 - \hat{p}_2) \mp z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right]$$

En nuestro caso:

$$\left[\left(\frac{60}{100} - \frac{90}{115}\right) \mp 1,96 \sqrt{\frac{\frac{60}{100}(1-\frac{60}{100})}{100} + \frac{\frac{90}{115}(1-\frac{90}{115})}{115}} \right]$$

$$\boxed{\left[-0,3047; \ -0,0605 \right]}$$