San Cristóbal de las Casas, Chiapas

Alumno: BENITO DE JESUS Pérez TRUJILLO.

Docente: MARIA GISELLE VILLATORO VALENZUELA.

Trabajo: glosario.

Carrera: Psicología general.

Grupo: D

Asignatura: “ESTADISTICA INFERENCIAL.”

Cuatrimestre: IV

Intervalos de confianza para medias: Supongamos que la estatura de los niños de 2 años está distribuida normalmente con una media de 90 cm y una desviación estándar de 36 cm. ¿Cuál sería la distribución muestral de la media para una muestra de tamaño 9? Recordemos que la media de una distribución muestral de medias es igual a μ: Y el error estándar es: $\mu = \mu_{\overline{x}} \; \sigma_m = \frac{\sigma}{\sqrt{n}}$

Intervalos de confianza para diferencia entre medias:

Supongamos que tenemos dos variables aleatorias ***independientes*** normalmente distribuidas:

$$\begin{cases} X_1 \sim N(\mu_1, \sigma_1^2) \\ X_2 \sim N(\mu_2, \sigma_2^2) \end{cases}$$ y suponemos que las varianzas σ_1^2 y σ_2^2 son conocidas.

Sean además

$(X_{11}, X_{12}, ..., X_{1n_1})$ una muestra aleatoria de tamaño n_1 de X_1

$(X_{21}, X_{22}, ..., X_{2n_2})$ una muestra aleatoria de tamaño n_2 de X_2.

Deseamos construir un intervalo al nivel de confianza $1-\alpha$ para la diferencia de esperanzas $\mu_1 - \mu_2$.

Ya sabemos cuál es la distribución del promedio de variables aleatorias normales independientes:

$$\begin{cases} \overline{X}_1 = \dfrac{1}{n_1}\displaystyle\sum_{i=1}^{n_1} X_{1i} \sim N\left(\mu_1, \dfrac{\sigma_1^2}{n_1}\right) \\ \overline{X}_2 = \dfrac{1}{n_2}\displaystyle\sum_{i=1}^{n_2} X_{2i} \sim N\left(\mu_2, \dfrac{\sigma_2^2}{n_2}\right) \end{cases}$$

Consideremos ahora la diferencia $\overline{Y} = \overline{X}_1 - \overline{X}_2$. Si $\overline{X}_1$ y $\overline{X}_2$ tienen distribución normal y son independientes, su diferencia también es normal, con esperanza igual a la diferencia de las esperanzas y la varianza es la suma de las varianzas:

$$\overline{X}_1 - \overline{X}_2 \sim \mathrm{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

Por lo tanto

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim \mathrm{N}(0,1),$$ es decir, tiene distribución normal estandarizada.

La v.a. Z cumple con toda las condiciones para servir de pivote y construiremos nuestro intervalo en forma análoga a cómo hicimos en los casos anteriores:

Comenzamos por plantear la ecuación

$$P(-z \leq Z \leq z) = 1\text{-}\alpha\,,$$

donde la incógnita es el número real z.

Reemplazamos la v.a. Z por su expresión y tenemos sucesivamente (multiplicando por $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$ y restando $\overline{X}_1 - \overline{X}_2$):

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

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

Multiplicando todos los miembros de la desigualdad por -1 (el orden de los miembros se invierte) llegamos a:

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

Evidentemente, si definimos

$$\begin{cases} \hat{\Theta}_1 = \overline{X}_1 - \overline{X}_2 - z\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \\ \hat{\Theta}_2 = \overline{X}_1 - \overline{X}_2 + z\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \end{cases},$$

habremos construido dos estadísticos $\hat{\Theta}_1$ y $\hat{\Theta}_2$ tales que $P(\hat{\Theta}_1 \leq (\mu_1 - \mu_2) \leq \hat{\Theta}_2) = 1\text{-}\alpha$, es decir habremos construido el intervalo de confianza bilateral deseado $[\hat{A}_1, \hat{A}_2]$. Todos los elementos que forman los estadísticos $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son conocidos ya que el número z verifica la ecuación anterior, es decir:

$P(-z \leq Z \leq z) = \Phi(z) - \Phi(-z) = 1\text{-}\alpha$ donde $\Phi(z)$ es la Fda para la v.a. $Z \sim N(0,1)$

o bien, según vimos:

$\Phi(z) = 1 - \dfrac{\alpha}{2}$ que anotamos $z_{\frac{\alpha}{2}}$

En consecuencia, el intervalo de confianza bilateral al nivel de significación $1\text{-}\alpha$ queda:

$$\left[\overline{X}_1 - \overline{X}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\quad \overline{X}_1 - \overline{X}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$

Por lo tanto

Si X_1 y X_2 son dos variables aleatorias ***independientes*** normalmente distribuidas: $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ y suponemos que las varianzas σ_1^2 y σ_2^2 son conocidas. Un intervalo de confianza para la diferencia $\mu_1 - \mu_2$ de nivel $1-\alpha$ es

$$\left[\bar{X}_1 - \bar{X}_2 - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\quad \bar{X}_1 - \bar{X}_2 + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right] \qquad (8.3)$$

Intervalos de confianza para proporciones entre proporciones.

Dada una variable aleatoria con **distribución Binomial** B(*n*, *p*), el objetivo es la construcción de un intervalo de confianza para el parámetro *p*, basada en una observación de la variable que ha dado como valor *x*. El mismo caso se aplica si estudiamos una Binomial B(1, *p*) y consideramos el número de veces que ocurre el suceso que define la variable al repetir el experimento *n* veces en condiciones de **independencia**.

Existen dos alternativas a la hora de construir un intervalo de confianza para *p*:

- Considerar la **aproximación asintótica** de la distribución Binomial en la distribución Normal.
- Utilizar un método exacto.

Aproximación asintótica

Tiene la ventaja de la simplicidad en la expresión y en los cálculos, y es la más referenciada en la mayoría de textos de estadística. Se basa en la aproximación que, trasladada a la frecuencia relativa, resulta Tomando como estadístico pivote. que sigue una distribución N(0, 1), y añadiendo una **corrección por continuidad** al pasar de una variable discreta a una continua, se obtiene el intervalo de confianza asintótico:

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}+\frac{1}{2n}}$$

$$X \sim B(n,p) \rightarrow N(np, \sqrt{npq})$$

$$\hat{p} = X/n \rightarrow N\left(p, \sqrt{pq/n}\right)$$

$$Z = \frac{\hat{p}-p}{\sqrt{\hat{p}\hat{q}/n}}$$

Donde $z_{\alpha/2}$ es el valor de una distribución Normal estándar que deja a su derecha una probabilidad de $\alpha/2$ para un intervalo de confianza de $(1 - \alpha) \cdot 100$ %. Las condiciones generalmente aceptadas para considerar válida la aproximación asintótica anterior son:

$$n \geq 30 \quad ; \quad n\hat{p} \geq 5 \quad ; \quad n\hat{q} \geq 5$$

Intervalos de confianza para diferencias.

Como

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \quad \text{tiene distribución en el muestreo} \quad N(0,1)$$

resulta directamente que

$$\left(\bar{x} - \bar{y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}, \bar{x} - \bar{y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$$

es un intervalo de confianza para $\mu_1 - \mu_2$ de nivel de confianza $1 - \alpha$.

Análogo resultado se obtiene, reemplazando σ_1^2 y σ_2^2 por S_1^2 y S_2^2, en el caso en que las varianzas poblacionales sean desconocidas pero los tamaños de ambas muestras sean los suficientemente grandes $(n, m \geq 15)$. En el caso de tamaños muestrales grandes, también se puede prescindir de la hipótesis de normalidad y utilizar el Teorema Central del Límite, como después se comentará en más detalle.

8.2.3. Intervalo de confianza para la varianza

Un intervalo de confianza al nivel $1 - \alpha$ para la varianza de una distribución gaussiana (cuyos parámetros desconocemos) lo obtenemos como

$$\boxed{\sigma^2 \in \left[\frac{(n-1)\hat{\mathcal{S}}^2}{\chi^2_{n-1,1-\alpha/2}} \quad , \quad \frac{(n-1)\hat{\mathcal{S}}^2}{\chi^2_{n-1,\alpha/2}}\right]}$$

Ejemplo

Se estudia la altura de los individuos de una ciudad, obteniéndose en una muestra de tamaño 25 los siguientes valores:

Ejemplo

Se estudia la altura de los individuos de una ciudad, obteniéndose en una muestra de tamaño 25 los siguientes valores:

$$\begin{aligned} \bar{x} &= 170 \text{ cm} \\ \mathcal{S} &= 10 \text{ cm} \end{aligned}$$

Calcular un intervalo de confianza con $\alpha = 0,05$ para la varianza σ^2 de la altura de los individuos de la ciudad.

Solución:

$$\sigma^2 \in [63,45\,;\,201,60]$$

Intervalos de confianza para razones de dos varianzas.

Sea s_1^2 y s_2^2 las varianzas muestrales de dos muestras aleatorias e independientes de tamaño n_1 y n_2 seleccionadas desde dos poblaciones normales con varianzas σ_1^2 y σ_2^2, respectivamente.

El estadístico $\hat{\theta} = \frac{s_1^2}{s_2^2}$ es un estimador de la razón de varianzas $\theta = \frac{\sigma_1^2}{\sigma_2^2}$

Puesto que nuestro interés consiste en encontrar un intervalo de confianza del 100(1-α)% para esta razón de varianzas poblacionales, entonces se debe cumplir que $P(|\hat{\theta} - \theta| < \varepsilon) = 1 - \alpha$

Reemplazando valores tenemos: $P\left(\left|\frac{s_1^2}{s_2^2} - \frac{\sigma_1^2}{\sigma_2^2}\right| < \varepsilon\right) = 1 - \alpha$

Factorizando $\frac{\sigma_1^2}{\sigma_2^2}$ y aplicando el valor absoluto $-\varepsilon < \frac{\sigma_1^2}{\sigma_2^2}\left(\frac{s_1^2\sigma_2^2}{s_2^2\sigma_1^2} - 1\right) < \varepsilon$

Despejando tenemos: $1 - \frac{\sigma_2^2}{\sigma_1^2}\varepsilon < \frac{s_1^2\sigma_2^2}{s_2^2\sigma_1^2} < 1 + \frac{\sigma_2^2}{\sigma_1^2}\varepsilon$

La expresión central es una variable muestral $T = \frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}}$ es tal que T → $F(n_1-1, n_2-1)$

Por lo tanto el intervalo buscado deberá tener sus extremos valores F tal que

$$F_{\alpha/2}(n_1-1, n_2-1) \quad < \quad T \quad < \quad F_{1-\alpha/2}(n_1-1, n_2-1)$$

Ahora bien, si reemplazamos T por su valor y despejamos $\frac{\sigma_1^2}{\sigma_2^2}$ obtenemos

$$\frac{s_2^2}{s_1^2}\, F_{\alpha/2}(n_1-1, n_2-1) \quad < \quad \frac{\sigma_2^2}{\sigma_1^2} \quad < \quad \frac{s_2^2}{s_1^2}\, F_{1-\alpha/2}(n_1-1, n_2-1)$$

Observación

Del mismo modo, invirtiendo la razón y tomando en cuenta que $F_{\alpha/2} = 1/F_{1-\alpha/2}$ se puede tener

$$\frac{s_1^2}{s_2^2}\, F_{\alpha/2}(n_2-1, n_1-1) \quad < \quad \frac{\sigma_1^2}{\sigma_2^2} \quad < \quad \frac{s_1^2}{s_2^2}\, F_{1-\alpha/2}(n_2-1, n_1-1)$$

<u>Recuerde que:</u>

No siendo simétrica esta distribución los valores de F son diferentes. El F de lado izquierdo de intervalo debe producir un F menor que el de la derecha.

Observe también cómo se deben tomar los grados de libertad.

Bibliografias.

Estadistica infereencial, instituto tecnologico de chiuhuahua

Devore, Jay L. Probabilidad y estadística para ingeniería y ciencias. Internacional Thompson

Hildebrand, David K. & Ott, Lyman R. Estadística aplicada a la administración y la economía.