Estimación por Intervalo: Proporción

En el siguiente ejemplo:

Ejemplo 1. *Una moneda tiene probabilidad p de ocurrir águila cuando se lanza y el resultado es aleatorio. En este caso el parámetro de interés es θ = p.*

si bien tenemos una aproximación (estimación) $\hat{p}$ del parámetro p no sabemos el nivel de precisión o margen de error respecto, esto es, que distancia $d(\hat{p}, p)$ hay entre los dos valores con cierta métrica definida.

Las estimaciones de las proporciones de preferencia entre hombres y mujeres son numéricamente distintas pero aproximadas, sin embargo, esta diferencia puedes ser atribuida a la variabilidad ya que proviene de una muestra aleatoria.

Estimación por intervalo
Sea $\hat{\theta}$ un estimador puntual del parámetro θ. Un intervalo de confianza para θ al nivel $1-\alpha$, $\alpha \in (0,1)$, es de la forma $$IC_\theta = \left(g_1(\hat{\theta}), g_2(\hat{\theta})\right)$$ de tal manera que $$P[\theta \in IC_\theta] \geq 1-\alpha$$

Las funciones $g_i(\hat{\theta})$ utilizadas en la construcción de intervalos de confianza dependen también de cierta distribución de probabilidad.

Intervalo de confianza para una proporción

Del teorema central del límite se desprende que:

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

Entonces, a partir de

$$P\left[Z_{\alpha/2} < Z < Z_{1-\alpha/2}\right] = 1 - \alpha$$

sustituyendo la expresión anterior se tiene

$$P\left[Z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-p)}{n}}} < Z_{1-\alpha/2}\right] = 1 - \alpha$$

y mediante un proceso algebraico se puede demostrarque esto equivale a

$$P\left[\hat{p} - Z_{1-\frac{\alpha}{2}} S_{\hat{p}} < p < \hat{p} + Z_{1-\alpha/2} S_{\hat{p}}\right] \geq 1 - \alpha$$

Donde

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Intervalo de confianza para p
Sea $\hat{p}$ un estimador puntual de la proporción p en una población infinita. Un intervalo de confianza para p al nivel $1-\alpha, \alpha \in (0,1)$, es tal que: $$P\left[\hat{p} - Z_{1-\frac{\alpha}{2}} S_p < p < \hat{p} + Z_{1-\alpha/2} S_p\right] \geq 1 - \alpha$$ donde $$S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$ o, de forma alternativa $$\begin{aligned} IC_p &= \hat{p} \pm Z_{1-\alpha/2} S_{\hat{p}} \\ &= \hat{p} \pm \hat{e} \end{aligned}$$

Donde $S_{\hat{p}}$ es el error estándar de $\hat{p}$ y $\hat{e}$ el margen de error estimado. Si la población es finita de tamaño N, entonces:

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\sqrt{\frac{N-n}{N-1}}$$

En la figura 2 se muestran los componentes del intervalo de confianza para la proporción, el cual se conforma por el estimador puntual y el margen de error, y este a su vez es el producto del cuantil de la distribución normal estándar y del error estándar. El error estándar de $p\hat{}$ es la desviación estándar del estimador de p.

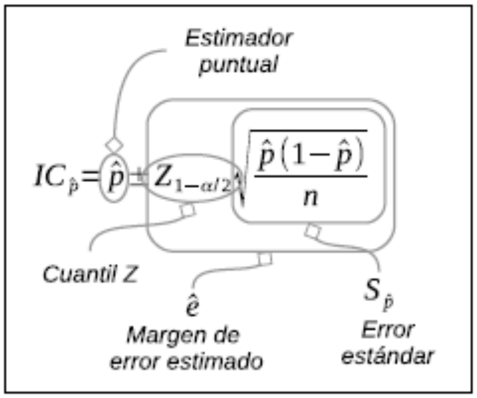

Figura 2 Estructura del IC de $\hat{p}$

Ejemplo 2 *En relación al ejemplo 1 tomando* 1 − α = 0.95, *el intervalo de confianza de la proporción de hombres a favor de que se construya la plaza es*

$$\begin{aligned} IC_{p_h} &= 0.8 \pm (1.96)\sqrt{\frac{(0.8)(1-0.8)}{80}} \\ &= 0.8 \pm (1.96)(0.0478) \\ &= 0.8 \pm 0.0937 \end{aligned}$$

Por tanto,

$$\begin{aligned} IC_{p_h} &= (0.8 - 0.0937, 0.8 + 0.0937) \\ &= (0.7063, 0.8937) \end{aligned}$$

Para mujeres

$$\begin{aligned} IC_{pm} &= 0.875 \pm (1.96)\sqrt{\frac{(0.875)(1-0.875)}{70}} \\ &= 0.875 \pm (1.96)(0.037 \\ &= 0.875 \pm 0.0725 \end{aligned}$$

Entonces,

$$\begin{aligned} IC_{p_m} &= (0.875 - 0.0725, 0.875 + 0.0725) \\ &= (0.8025, 0.9475) \end{aligned}$$

Si la población de hombres fuera finita de tamaño N = 250*, el intervalo de confianza para la proporción de hombres sería*

$$\begin{aligned} IC_{p_h}^{(N)} &= 0.8 \pm (1.96)\sqrt{\frac{(0.8)(1-0.8)}{80}}\sqrt{\frac{250-80}{250-1}} \\ &= 0.8 \pm (1.96)(0.0406) \\ &= 0.8 \pm 0.0797 \end{aligned}$$

por lo que

$$\begin{aligned} IC_{p_h}^{(N)} &= (0.8 - 0.0797, 0.8 + 0.0797) \\ &= (0.7203, 0.8797) \end{aligned}$$

Dado que el factor adicional de corrección por finitud es menor que 1, el nuevo intervalo es de menor longitud.

Se observa que $IC_ph \cap IC_pm$ /= ∅ con poblaciones infinitas. Este resultado será particularmente útil cuando se determine si se considera que hay diferencia significativa en la proporción de hombres y mujeres que están a favor de la construcción de la plaza.

Tamaño de muestra para una proporción

El margen de error en la estimación por intervalo de una proporción está asociado con el tamaño de muestra, puesto que

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

Entonces, si n crece entonces eˆ decrece y esto permite obtener una cantidad de muestra basada en un margen de error nominal e dado. Hay dos posibilidades: si se cuenta con una estimación previa o no.

La estimación del tamaño de muestra n, dado el error nominal e y cuando se cuenta con una estimación previa pˆ se obtiene como sigue:

$$e = Z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-p)}{n}}$$
$$\Rightarrow e^2 = Z^2_{1-\alpha/2}\frac{\hat{p}(1-p)}{n}$$
$$\Rightarrow n \approx Z^2_{1-\alpha/2}\frac{\hat{p}(1-p)}{e^2}$$

Si no se cuenta con una estimación previa, habrá que determinar el valor mínimo estimado de n que nos garantice error nominal e. Para tal efecto hay que maximizar $\hat{p}(1 - \hat{p})$. Para simplificar la notación se define $g(x) = x(1 - x)$ y se obtiene $x* = \text{argmax}g(x)$.

$$
\begin{aligned}
g(x) &= x(1-x) = x - x^2 \\
&\Rightarrow \frac{dg}{dx} = 1 - 2x
\end{aligned}
$$

por tanto,

$$
\begin{aligned}
&\frac{dg}{dx} = 0 \\
&\quad \Rightarrow 1 - 2x = 0 \\
&\quad \Rightarrow x^* = \frac{1}{2}
\end{aligned}
$$

Dado que

$$\frac{d^2g}{dx} = -2 < 0$$

se sigue que x* *corresponde al máximo de g. Sustituyendo este valor en lugar de* $\hat{p}$ *la estimación de n anterior se tiene:*

$$
\begin{aligned}
n &\cong Z_{1-\alpha/2}^2 \frac{\hat{p}(1-p)}{e^2} \\
&= Z_{1-\alpha/2}^2 \frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{e^2} \\
&= \frac{Z_{1-\alpha/2}^2}{4e^2}
\end{aligned}
$$

En ambos casos de la estimación de *n*, se toma el entero siguiente en cada una de las expresiones para garantizar al menos el error nominal *e*. El tamaño de muestra sin estimación previa siempre será mayor o igual a cuando se considera una estimación previa.

Tamaño de muestra para estimar p
Dado un margen de error nominal e en la estimación por intervalo de p el tamaño de muestra si se cuenta con una estimación previa $\hat{p}$ es: $$n \approx Z^2_{1-\alpha/2} \frac{\hat{p}(i-\hat{p})}{e^2}$$ Si no se tiene una estimación previa, el tamaño estimado es: $$n \approx \frac{Z^2_{1-\alpha/2}}{4e^2}$$

Ejemplo 3. *En el ejemplo* 3 *intervalo de confianza de la proporción de hombre se estimó como* $\hat{p}_h$ = 0.8*, el tamaño de muestra con estimación previa para e* = 0.06 *es:*

$$n \approx (0.96)^2 \frac{(0.8)(1-0.8)}{(0.06)^2} = 170.7315$$

Por tanto, se toma n =1

Sin estimación previa

$$n \approx \frac{(0.96)^2}{(4)(0.06)^2} = 266.768$$

Entonces, se toma n = 267*.*

En el ejemplo se observa que la estimación de n sin contar con $\hat{p}$ es considerablemente mayor a la obtenida cuando si se tiene $\hat{p}$.

Intervalo de confianza para diferencia de proporciones

Con anterioridad se observó que $IC_{ph} \cap IC_{pm} \;/\!= \varnothing$. Esto es equiparable a decir que la diferencia de proporciones entre hombres y mujeres no es estadísticamente significativa, o bien, que $p_h - p_m = 0$. En este sentido, es de interés obtener un intervalo de confianza para la diferencia de proporciones.

Intervalo de confianza para $p_2 - p_1$
Sean $\hat{p}_1$ y $\hat{p}_2$ estimadores puntual de las proporciones p_1 y p_2 respectivamente. Un intervalo de confianza para $p_2 - p_1$ al nivel $1-\alpha, \alpha \in (0,1)$, es de la forma $$\begin{aligned} IC_{p2} &= p_2 - p_1 \pm Z_{1-\alpha/2} S_{\hat{p}_2-\hat{p}_1} \\ &= \hat{p} \pm \hat{e} \end{aligned}$$ donde n_1 y n_2 son los tamaños de muestra para las poblaciones 1 y 2 respectivamente, y $S_{\hat{p}}$ es el error estándar de $p_2 - p_1$ dado como: $$S_{\hat{p}_2-\hat{p}_1} = \sqrt{\frac{\hat{p}_1\,(1-\hat{p}_1\,)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$ de $\hat{p}$ y $\hat{e}$ el margen de error estimado.

Ejemplo 4. *Tomando* $1 - \alpha = 0.95$*, el intervalo de confianza para la diferencia de proporciones de hombres y mujeres a favor de que se construya la plaza se obtiene como sigue:*

$$\begin{aligned} S_{\hat{p}_h-\hat{p}_m} &= \sqrt{\frac{(0.8)(1-0.8)}{80} + \frac{(0.875)(1-0.875)}{70}} \\ &= 0.0604 \end{aligned}$$

Se sigue

$$\begin{aligned} IC_{p_h - p_m} &= (0.8 - 0.875) \pm (1.96)(.0604) \\ &= -0.075 \pm 0.1185 \end{aligned}$$

Por tanto,

$$\begin{aligned} IC_{Ph-pm} &= (-0 - 075 - 0.1185, -0.075 + 0.1185) \\ &= (-0.1935, 0.00435) \end{aligned}$$

Es importante señalar que, en la diferencia de proporciones, los límites del intervalo no son proporciones, por lo que pueden tomar valores negativos.

El resultado final arroja que $0 \in IC_{p:h-pm}$, siendo consistente con el resultado donde se obtiene que $IC_{ph} \cap IC_{pm} \;/= \varnothing$.

Referencias:

Hoel, P. G. (1984). Elementary Statistics. John Wiley & Sons.

Valdés, J. R. (2019). Estadística. Notas de clase.