1. Intervalos de confianza

Es lógico que para cada muestra el $\hat{\theta}$ cambie, entonces queremos determinar

$$P(\hat{\theta}_L < \theta < \hat{\theta}_U) = 1 - \alpha$$

Por lo tanto $\hat{\theta}_L < \theta < \hat{\theta}_U$ es un intervalo de confianza de $(1-\alpha)100\,\%$. $\hat{\theta}_L$ representa el límite inferior, $\hat{\theta}_U$ el límite superior del intervalo de confianza y $1-\alpha$ es el coeficiente o grado de confianza. Este también es conocido como intervalo de confianza bilateral.

Se pueden calcular intervalos de confianza unilaterales confianza

$$P(\hat{\theta}_L \leq \theta) = 1 - \alpha.$$

y

$$P(\theta \leq \hat{\theta}_U) = 1 - \alpha.$$

Un método que sirve para encontrar intervalos de confianza es el método del pivote. El pivote debe tener las siguientes características:

1. Que sea una función de las mediciones de la muestra y del parámetro desconocido θ, donde θ es la única cantidad desconocida.
2. Que su distribución de probabilidad no dependa del parámetro θ.

Conocida la distribución de probabilidad del pivote, entonces utilizamos operaciones de escala y/o traslación de la variable aleatoria para conseguir el intervalo deseado.

Ejemplo. Suponga que la variable aleatoria Y tiene una distribución gamma con parámetros $\alpha = 2$ y β desconocido. Aplicando funciones generadoras de momentos se puede demostrar que $2Y/\beta$ tiene una distribución χ^2 con 4 grados de libertad. Mediante $2Y/\beta$ como cantidad pivote, deduzca un intervalo de confianza de $90\,\%$ para β.

Solución.

Como nos dicen, $U = \dfrac{2Y}{\beta} \sim \chi^2$ con 4 grados de libertad. La distribución de χ^2 no depende de β, así que podemos usar U como nuestro pivote. Entonces queremos a y b tales que

$$P(a \leq U \leq b) = 0{,}90.$$

Eso lo podemos ver como

$$P(U \leq a) = 0{,}05 \Rightarrow P(U \geq a) = 0{,}95 \Rightarrow a = 0{,}711.$$

y

$$P(U \geq b) = 0{,}05 \Rightarrow b = 9{,}49.$$

Por lo tanto

$$0{,}90 = P(0{,}711 \leq U \leq 9{,}49) = P\left(0{,}711 \leq \frac{2Y}{\beta} \leq 9{,}49\right) = P\left(\frac{2Y}{9{,}49} \leq \beta \leq \frac{2Y}{0{,}711}\right).$$

Entonces nuestro intervalo de confianza es $\left(\dfrac{2Y}{9{,}49}, \dfrac{2Y}{0{,}711}\right)$.

2. Intervalos de confianza para muestras grandes

Si las muestras son grandes, entonces $Z = \dfrac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$ tiene distribución normal estándar. Entonces usemos dicha cantidad como pivote.

Utilicemos las colas de la distribución normal estándar, entonces

$$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha \Rightarrow P\left(-z_{\alpha/2} \leq \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} \leq z_{\alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow P\left(-z_{\alpha/2}\sigma_{\hat{\theta}} \leq \hat{\theta} - \theta \leq z_{\alpha/2}\sigma_{\hat{\theta}}\right) = 1 - \alpha$$

$$\Rightarrow P\left(-z_{\alpha/2}\sigma_{\hat{\theta}} - \hat{\theta} \leq -\theta \leq z_{\alpha/2}\sigma_{\hat{\theta}} - \hat{\theta}\right) = 1 - \alpha$$

$$\Rightarrow P\left(\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}\right) = 1 - \alpha$$

Por lo tanto el intervalo de confianza de $100(1-\alpha)\,\%$ es $[\hat{\theta} - z_{\alpha/2}\sigma_{\hat{\theta}}, \hat{\theta} + z_{\alpha/2}\sigma_{\hat{\theta}}]$.

En el caso de los intervalos unilaterales se puede demostrar, con argumentos similares, que son $(-\infty, \hat{\theta} + z_{\alpha}\sigma_{\hat{\theta}}]$ y $[\hat{\theta} - z_{\alpha}\sigma_{\hat{\theta}}, \infty)$.

<u>Ejercicio</u>. A veces las encuestas proporcionan información importante sobre temas que parecen ajenos a sus objetivos. El New York Times aplicó una encuesta para determinar las preferencias de los votantes en las elecciones presidenciales de 1992. Se realizaron 2374 entrevistas por vía telefónica con ciudadanos estadounidenses adultos, sin incluir Alaska y Hawaii. 1912 ciudadanos indicaron que contaban con registro de elector. La muestra de números de telefonos se obtuvo por computadora de una lista de centrales telefónicas del país. En el caso de cada central telefónica, los números telefónicos se formaron eligiendo dígitos aleatoriamente, con lo cual se tuvo acceso a los números que aparecían en la lista y a los que no aparecían. Con los datos anteriores, genere un intervalo de confianza del 99 % para la proporción de ciudadanos adultos que viven en la parte continental de Estados Unidos que tenían registro de elector en 1992.

SOLUCIÓN.
Como ya sabemos, la proporción, vista como una variable aleatoria que proviene de un experimento binomial, al ser dividido por n se aproxima a una normal cuando n es grande. De esta forma podemos usar el resultado antes deducido.

Nuestro estimador será:

$$\hat{p} = \frac{1912}{2374} = 0{,}8054.$$

Como hemos trabajado antes

$$\sigma_{\hat{p}} = \sqrt{\frac{(0{,}8054)(0{,}1946)}{2374}} = 0{,}0081.$$

Finalmente como queremos un intervalo al 99 %, entonces debemos buscar en la tabla para la normal estándar $z_{0{,}005} \backsimeq 2{,}58$.

Luego

$$\hat{p} \pm z_{0{,}005}\sigma_{\hat{p}} = 0{,}8054 \pm 2{,}58(0{,}0081) = 0{,}8054 \pm 0{,}0209.$$

Por lo tanto se tiene que la proporción debe estar entre $[0{,}7845; 0{,}8263]$ con una confianza de 99 %.

3. Tamaño de la muestra

La noción de calcular el tamaño de la muestra ya fue introducida con el teorema del límite central, y se relaciona con querer saber cuantas muestras se desean tener para obtener el resultado que se busca. Realizaremos un par de ejemplos de como conseguir esto.

EJEMPLO. Suponga que desea estimar el pH promedio de precipitaciones en un área con altos índices de contaminación debido a los gases que emite una planta de energía eléctrica. Suponga que σ está en la vecindad de 0.5 pH y que usted desea que su estimación difiera en a lo más 0.1 de μ con una probabilidad alrededor de 0.95. ¿Cuántas precipitaciones debe incluir aproximadamente en su muestra?¿Es válido que tome todas sus muestras de agua de una sola precipitación?

SOLUCIÓN.
Queremos n para que $P(|\bar{Y} - \mu| < 0{,}1) = 0{,}95$. Entonces

$$P\left(-\frac{0{,}1}{0{,}5}\sqrt{n} \leq \frac{(\bar{Y} - \mu)}{\sigma}\sqrt{n} \leq \frac{0{,}1}{0{,}5}\sqrt{n}\right) = 0{,}95.$$

De la tabla de la normal sabemos que $P(-1{,}96 \leq Z \leq 1{,}96) = 0{,}95$, luego

$$0{,}2\sqrt{n} = 1{,}96 \Rightarrow n = \left(\frac{1{,}96}{0{,}2}\right)^2 = 96{,}04 \backsimeq 96.$$

Las muestras no deben ser de una sola precipitación. (¿Por qué?)

<u>Ejemplo</u>. Suponga ahora que desea estimar la diferencia de acidez promedio del agua de lluvia en dos diferentes zonas, una sin contaminación en el océano y otra con altos índices de contaminación en el aire. Si desea que su estimación tenga un error máximo de 0.1 pH con una probabilidad de alrededor del 90 %, ¿aproximadamente cuántas precipitaciones (valores de pH) debe incluir en cada muestra? (Suponga que la varianza de las mediciones de pH es de alrededor de 0.25 en las dos zonas y que las muestras son del mismo tamaño).

<u>Solución</u>.
Ahora queremos que $P(|(\bar{Y_1} - \bar{Y_2}) - (\mu_1 - \mu_2)| < 0{,}2) = 0{,}90$, sabiendo que $\sigma_1^2 = \sigma_2^2 = \sigma_2$ y que $n_1 = n_2 = n$. Entonces

$$\sigma_{\text{est}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{2(0{,}25)}{n}}.$$

Por lo que

$$P\left(-\frac{0{,}1}{\sqrt{0{,}5}}\sqrt{n} \leq Z \leq \frac{0{,}1}{\sqrt{0{,}5}}\sqrt{n}\right) = 0{,}90.$$

De la tabla de la normal estándar sacamos que $P(-1{,}65 \leq Z \leq 1{,}65) \simeq 0{,}90$, por lo tanto

$$\frac{0{,}1}{\sqrt{0{,}5}}\sqrt{n} = 1{,}65 \Rightarrow \sqrt{n} = \frac{1{,}65\sqrt{0{,}5}}{0{,}1} \Rightarrow n = 136{,}125 \simeq 136.$$