2.2 INTERVALOS DE CONFIANZA PARA PROPORCIONES

De acuerdo al teorema central del límite, x tiene distribución aproximadamente normal cuando n es grande. Como n es una constante, se puede pensar que p es también aproximadamente normal cuando n es grande y esto es realmente así.

Dado que X es una variable binomial, tenemos:

$$E(X) = nP \quad \text{y} \quad V(X) = nPQ$$

y como:

$$p = \frac{x}{n}$$
$$E(p) = \frac{E(X)}{n} = \frac{nP}{n} = P$$
$$V(p) = \frac{V(X)}{n^2} = \frac{nPQ}{n^2} = \frac{PQ}{n} = \frac{P(1 - P)}{n}$$

Cuando n es suficientemente grande (n > 30) podemos decir que p se distribuye normal con media P y varianza PQ/n.

Luego,

$$p \sim N(P, \sqrt{\frac{PQ}{n}})$$

Como hemos especificado que p tiene distribución normal y además conocemos su esperanza y varianza, podemos estandarizar la variable aleatoria proporción muestral a efectos de obtener una nueva variable z con distribución N (0,1).

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

Esta transformación es importante por cuanto de este modo podemos utilizar la distribución de probabilidad normal para establecer una estimación por intervalos del parámetro proporción poblacional.

Así:

$$P\left(z_{\alpha/2} \leq \frac{(p - P)}{\sqrt{\frac{p(1 - p)}{n}}} \leq z_{1 - \alpha/2}\right) = 1 - \alpha$$

Despejando convenientemente, tendremos:

$$P\left(p - z_{1 - \alpha/2}\sqrt{\frac{p(1 - p)}{n}} \leq P \leq p - z_{\alpha/2}\sqrt{\frac{p(1 - p)}{n}}\right) = 1 - \alpha$$

Un problema que ya deben haber detectado es que, en la desviación estándar del estimador p aparece P que es un parámetro desconocido y que, precisamente, es lo que se quiere estimar. La única manera de solucionar esta situación consiste en reemplazarlo por su correspondiente estimador muestral p. Esta sustitución produce muy poco error cuando la muestra es grande.

Veremos ahora un ejemplo en el que calcularemos un intervalo de confianza para estimar al parámetro binomial o proporción poblacional P.

Una empresa dedicada a sondeos de opinión pública realiza una encuesta para averiguar la intención de voto hacia un candidato A en una elección futura.

Para ello toma una muestra al azar de 100 posibles votantes, de los cuales 59 opinan que se inclinan por el candidato A.

Para estimar la proporción de simpatizantes del candidato A en la población, la empresa decide construir un intervalo estableciendo un coeficiente de confianza del 95%.

Los datos con que cuenta para realizar la estimación son:

De acuerdo al coeficiente de confianza establecido:

$$z_{\alpha/2} = -1{,}96 \text{ y } z_{1-\alpha/2} = 1{,}96$$

El intervalo quedará luego:

$$P\left(0{,}59 - 1{,}96 \bullet \sqrt{\frac{0{,}59 \bullet 0{,}41}{100}} \leq P \leq 0{,}59 + 1{,}96 \bullet \sqrt{\frac{0{,}59 \bullet 0{,}41}{100}}\right) = 0{,}95$$

$$P(0{,}59 - 0{,}10 \leq P \leq 0{,}59 + 0{,}10) = 0{,}95$$

$$P(0{,}49 \leq P \leq 0{,}69) = 0{,}95$$

De acuerdo al resultado obtenido, la empresa puede asegurar que la proporción de simpatizantes del candidato A en la población será un valor comprendido entre el 49% y el 69%, con una confianza del 95%.

Adjuntamos un resumen de los conceptos desarrollados hasta aquí. Le sugerimos su lectura, antes de pasar a resolver las actividades propuestas.