Distribución Muestral de Proporciones

Piensa que una población es infinita y que la probabilidad de ocurrencia de un evento (llamado su éxito) es p, mientras que la probabilidad de la no ocurrencia del evento es $q = 1 - p$. Por ejemplo, la población pueden ser todos los lanzamientos posibles de una moneda, donde la probabilidad del evento "caras" es $p = \frac{1}{2}$. Considera todas las muestras posibles, de tamaño *N,* obtenidas en esta población y determínese para cada muestra la proporción *P* de éxito. En el caso de la moda, *P,* sería la proporción de caras de *N* lanzamiento. Así, se obtiene una *distribución muestral de proporciones* cuya media μ_p y desviación estándar σ_p están dadas por

$$\mu_p \quad \text{y} \quad \sigma p = \sqrt{\frac{pq}{N}} = \sqrt{\frac{p(1-q)}{N}}$$

que suele obtenerse de las ecuaciones vistas en la distribución muestral de medias, haciendo μ=p y $\sigma=\sqrt{pq}$. Para valores grandes de *N (N≥30),* la distribución muestral se aproxima mucho a la distribución normal. Nótese que la población esta *binomialmente distribuida*.

Las ecuaciones μ_p y $\sigma p = \sqrt{\frac{pq}{N}} = \sqrt{\frac{p(1-q)}{N}}$ son válidas para una población finita, donde el muestreo se hace con reemplazamiento. Para poblaciones finitas donde el muestreo se realiza sin reemplazamiento, las ecuaciones μ_p y $\sigma p = \sqrt{\frac{pq}{N}} = \sqrt{\frac{p(1-q)}{N}}$ se sustituyen por las ecuaciones $\mu_x = \mu$ y $\sigma_x = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}}$, con μ=p y $\sigma=\sqrt{pq}$.

Observa que las ecuaciones μ_p y $\sigma p = \sqrt{\frac{pq}{N}} = \sqrt{\frac{p(1-q)}{N}}$ se obtiene más fácilmente al dividir la media y la desviación estándar$(Np \ y \ \sqrt{Npq})$, de la distribución binomial, entre *N.*

Ejemplo:

Calcula la probabilidad de que en 120 lanzamientos de una moneda a) entre 40% y 60% sean caras y b) $\frac{5}{8}$ o más sean caras.

Solución

Primer método

Los 120 lanzamientos son considerados una muestra de la población infinita de todos los posibles lanzamientos de dicha moneda. En esta población la probabilidad de caras es de $p = \frac{1}{2}$ y la probabilidad de cruces es de q= 1-p= $\frac{1}{2}$.

a) Se requiere la probabilidad de que el número de caras en 120 lanzamientos esté entre (40% de 120) = 48 y (60% de 120)= 72. Se procede aplicando la aproximación normal binomial. Dado que el número de caras es una variable discreta, se pide la probabilidad de que el número de caras esté entre 47.5 y 72.5

$$\mu = \textit{número esperado de caras} = Np = 120\ \frac{1}{2} = 60$$

$$\text{y } \sigma = \sqrt{Npq = \sqrt{(120)\left(\frac{1}{2}\right)}\left(\frac{1}{2}\right) = 5.48}$$

$$47.5\ \textit{en unidades estándar} = \frac{47.5 - 60}{5.48} = -2.28$$

$$47.5\ \textit{en unidades estándar} = \frac{72.5 - 60}{5.48} = 2.28$$

Probabilidad requerida = (área bajo la curva normal entre z= -2.28 y z=2.28)

=2(área entre z=0 y z= 2.28)

=2(0.4887)= 0.9774

Segundo método

$$\mu_p = p\frac{1}{2} = 0.50 \quad \sigma_p\sqrt{\frac{pq}{N}} = \sqrt{\frac{(\frac{1}{2})(\frac{1}{2})}{120}} = 0.0456$$

40% en unidades estándar $= \frac{0.40-0.50}{0.0456} = -2.19$

60% en unidades estándar $= \frac{0.40-0.50}{0.0456} = 2.19$

Probabilidad requerida = (área bajo la curva normal entre z= -2.19 y z= 2.19)

= 2(0.4857)= 0.9714

No obstante que este resultado es preciso con dos cifras, no coincide en forma exacta, ya que no se utilizó el hecho de que la proporción es en realidad una variable discreta. Para justificarlo, se resta 1/2N= ½ (120) de 0.40 y se suma 1/2N= ½ (120) a 0.60; entonces, dado que 1/240= 0.0047, las proporciones requeridas en unidades estándar son

$$\frac{0.40-0.00417-0.50}{0.0456} = -2.28 \quad \text{y} \quad \frac{0.60+0.00417-0.50}{0.0456} = 2.28$$

de tal manera que coincide con el primer método.

Observa que (0.40-0.00417) y (0.60+0.00417) corresponden a las proporciones 47.5/120 y 72.5/120, en el primer método

b) Usando el segundo método del inciso a), se encuentra que $\frac{5}{8}$ = 0.6250; entonces, (0.6250-0.00417) en unidades estándar $= \frac{0.6250-0.00417-0.50}{0.0456} = 2.65$

Probabilidad requerida = (área bajo la curva

= (área a la

-(área entre *z*=0 y *z*= 2.65)

= 0.5-0.4960=0.0040

Referencia:

Spiegel M. y Stephens L. (2002). Estadística. México: McGraw-Hill.