Métodos de Inferencia Estadística con Proporciones y Frecuencias

Inferencias con una proporción poblacional

Una proporción p representa el número de unidades de análisis o individuos X que presentan una cierta característica de entre un total de n elementos de una muestra:

$$p = X/n$$

Por tanto una proporción es un porcentaje, por ejemplo 86%, expresado en tanto por uno, por ejemplo 0.86. Puesto que X es una variable aleatoria con distribución binomial, para un tamaño muestral n suficientemente grande, es decir si $n>30$, la proporción p se comportará como una variable aleatoria con distribución normal, y parámetros media y desviación típica igual a P (proporción poblacional) y $\sqrt{\frac{p.q}{n}}$, respectivamente. Obsérvese que p y q son parámetros estimados experimentalmente en la muestra, siendo q la proporción de individuos que no presenta la característica estudiada, cumpliéndose por tanto que $p + q = 1$.

Si tipificamos el parámetro p con distribución normal N(P, $\sqrt{\frac{p.q}{n}}$), tendremos un estadístico Z con distribución normal N(0, 1):

$$Z = \frac{p - P}{\sqrt{\frac{p.q}{n}}}$$

En la práctica q es obtenido a partir de p, esto es:

$$q = 1 - p$$

escribiéndose en algunas ocasiones el estadístico:

$$Z = \frac{p - P}{\sqrt{\frac{p.(1-p)}{n}}}$$

Este estadístico será de utilidad si realizamos un contraste acerca de la proporción poblacional *P* de sujetos que exhiben una cierta característica:

$$\text{H0: } P = p_0$$

$$\text{Ha: } P \neq p_0$$

siendo p_0 el valor, es decir la proporción de sujetos que efectivamente presenta una característica dada. Por consiguiente, si aceptamos provisionalmente la H0, y únicamente a efectos prácticos, tendremos que el estadístico que computaremos es:

$$Z = \frac{p - p_0}{\sqrt{\frac{p_0.(1-p_0)}{n}}}$$

Inferencias con dos proporciones poblacionales

Supóngase que en un experimento estuviéramos interesados una vez recopilados los datos en comparar dos proporciones de dos grupos experimentales o poblacionales, es decir P_1 y P_2. En primer lugar, y desde un punto de vista experimental, tomaremos dos muestras aleatorias n_1 y n_2, obteniendo los valores de las proporciones muestrales p_1 y p_2. A continuación, y en segundo lugar, si se cumple que los tamaños muestrales respectivos son grandes, es decir si n_1, n_2 > 30, entonces el estadístico tendrá distribución normal estándar Z:

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{p_1.q_1}{n_1} + \frac{p_2.q_2}{n_2}}}$$

Si comparamos dos grupos experimentales, este estadístico será de utilidad, siendo el contraste:

$$\text{H0: } P_1 = P_2$$
$$\text{Ha: } P_1 \neq P_2$$

En algunas ocasiones, y desde un punto de vista práctico, se estima un valor común p^* bajo el supuesto de aceptación de la H0 (y por tanto P_1-P_2 = 0), obteniéndose dicho valor con la expresión:

$$p^* = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

siendo q^* igual a 1- p^*. En tal caso el estadístico toma la forma:

$$Z = \frac{p_1 - p_2}{\sqrt{p^*(1 - p^*)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$