Prueba de Hipótesis: Proporciones

Una prueba de hipótesis es una regla que especifica si se puede aceptar o rechazar una afirmación acerca de una población dependiendo de la evidencia proporcionada por una muestra de datos.

Una prueba de hipótesis examina dos hipótesis opuestas sobre una población: la hipótesis nula y la hipótesis alternativa. La hipótesis nula es el enunciado que se probará. Por lo general, la hipótesis nula es un enunciado de que "no hay efecto" o "no hay diferencia". La hipótesis alternativa es el enunciado que se desea poder concluir que es verdadero de acuerdo con la evidencia proporcionada por los datos de la muestra.

Prueba de Hipótesis: Proporciones

Los métodos de curva normal para muestra grande que se emplea para resolver problemas de estimación en el caso de p binomial se puede emplear también para probar hipótesis con respecto a p. Como resultado, las técnicas para probar la hipótesis de que p tiene un valor fijo o de que dos proporciones sean iguales, son muy similares para las medias.

Ejemplo 1. *Considérese el siguiente problema de genética. De acuerdo con la teoría de la herencia de Mendel, ciertos cruzamientos de chicharos producen chicharos amarillos y verdes en relación 3:1. En un experimento, se han obtenido 176 chicharos amarillos y 48 verdes. ¿Son estos números compatibles con la teoría mendeliana?*

Este problema se puede considerar como problema de prueba de hipótesis.

$$H_0: p = \frac{3}{4}$$

En la cual p designa la probabilidad de que un chícharo seleccionado al azar sea amarillo. Los 224 chícharos se pueden considerar como 224 intentos de un experimento para el cual $p = {}^{3}/_{4}$ es la probabilidad de éxito en un solo intento. De la fórmula (1) se infiere que $\widehat{p} = {}^{x}/_{n}$ se puede considerar como una variable normal cuya media $p = {}^{3}/_{4}$ y cuya desviación estándar está dada por

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

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{\frac{3}{4} * \frac{1}{4}}{224}} = 0.029$$

El problema ahora es muy similar al de prueba de una media normal. La región en las dos colas de la curva normal para p. Para α=0.05 la región critica consistirá en aquellos valores de $\widehat{p}$ que se encuentran fuera del intervalo dado por:

$$p - 1.96\sqrt{\frac{pq}{n}} \ |y|\ p + 1.96\sqrt{\frac{pq}{n}}$$

Puesto que $n = 224\ y\ p = \frac{3}{4}$ aquí, los cálculos arrojaran el intervalo (0.693, 0.807). La figura 6 muestra la distribución normal aproximada para $\widehat{p}$ y la región crítica que se acaba de determinar. Puesto que el valor de muestra $\widehat{p} = \frac{176}{224} = 0.79$ no se encuentra dentro de la región crítica, se acepta la hipótesis H_0. Así pues, a partir de estos datos no existe razón para dudar de que la ley de herencia mendeliana opera en este caso.

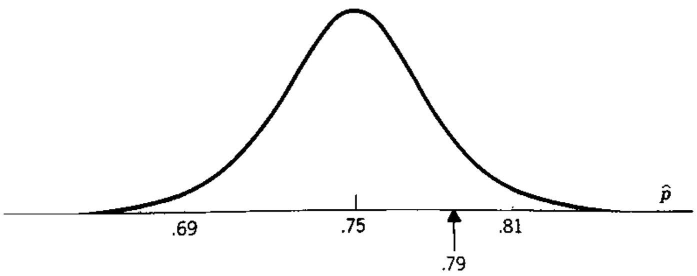

Figura 1. Distribución normal aproximada de $\hat{p}$

Ejemplo 2. *Un político ha pretendido que tiene un 60% de respaldo para una determinada ley y se ha tomado una muestra de 400 votantes para comprobar esta afirmación. Se presenta ahora la duda de qué tan pequeño necesitaría ser el porcentaje dado por la muestra para que la pretensión pudiera refutarse justificadamente.*

Este problema se puede considerar como problema de prueba de hipótesis.

$$H_0: p = 0.6$$

Puesto que el interés en este problema es probar la hipótesis de que $p = 0.6$, contra la posibilidad de que $p < 0.6$, las alternativas se encuentran todas hacia un lado del valor hipotético y, por lo tanto, la región crítica deberá encontrarse bajo una de las dos colas de la curva normal. La hipótesis alternativa natural aquí es

$$H_1: p < 0.6$$

La región crítica de tamaño $\alpha = 0.05$ para este problema debe, pues, seleccionarse abajo del 5% de la izquierda, en la cola de la curva normal cuya media es $p = 0.6$ y cuya desviación estándar está dada por

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.6 * 0.4}{400}} = 0.0245$$

Puesto que, de la Tabla IV, es 5% del área de una curva normal ordinaria se encuentra a la izquierda de $z = -1.64$, esto significa que la región critica deberá consistir en todos los valores de $\hat{p}$ que sean más pequeños de $\hat{p}$ que se encuentra a 1.64 desviaciones estándar a la izquierda de la media. Para este problema, la región critica consiste en todos aquellos valores de $\hat{p}$ inferiores a

$$p - 1.64\sqrt{\frac{pq}{n}} = 0.6 - 1.64(0.0245) = 0.56$$

Si el valor de muestra de $\hat{p}$ es menor que 0.56, la pretensión del político puede rechazarse.

Prueba de la diferencia de dos proporciones

Un problema de gran importancia y que se presenta frecuentemente en el trabajo estadístico es el de determinar si dos poblaciones difieren con respecto a cierta característica.

Ejemplo 3. *En un artículo aparecido en el Canadian Medical Association Journal de noviembre de 1972 se discuten los defectos de fuertes dosis de vitaminas C en el padecimiento común debido a un resfriado. En el experimento llevado a cabo había 407 individuos que tomaron fuertes dosis de vitamina C y 411 que ingirieron placebos. De los que ingirieron vitamina, 105 permanecieron sanos respecto a enfermedades respiratorias durante el estudio, mientras que este fue el caso para sólo 76 individuos de los cuales tomaron placebos.*

Haga la prueba de la hipótesis de que no hay diferencia entre las verdaderas proporciones de los dos grupos. Estos datos proporcionan los valores.

$$\hat{p}_1 = \frac{105}{407} = 0.26 \quad y \quad \hat{p}_2 = \frac{76}{411} = 0.18$$

Como la hipótesis $H_0: p_1 = p_2$, una estimación de $p = p_1 = p_2$ se obtienen combinando los dos resultados de muestra. Este nos da

$$\hat{p} = \frac{181}{818} = 0.22$$

Las siguiente es, pues, una estimación de $\sigma_{\widehat{p_1}-\widehat{p_2}}$

$$\sigma_{\widehat{p_1}-\widehat{p_2}} \doteqdot \sqrt{0.22 * 0.78 \left[\frac{1}{407} + \frac{1}{411}\right]} \doteqdot 0.029$$

Como la hipótesis alternativa $H_1: p_1 > p_2$, la cola derecha de la distribución de $p\hat{}_1 - p\hat{}_2$ es la que deberá tomarse como región crítica de esta prueba. Esto significa que la región crítica deberá consistir en los valores de $p\hat{}_1 - \hat{p}_2$ a la derecha del punto $1.64\sigma_{\widehat{p_1}-\widehat{p_2}} = 0.048$. Como $p\hat{}_1 - \hat{p}_2 = 0.26 - 0.18 = 0.08$, y este valor cae en la región crítica, deberá rechazar la hipótesis de que no habrá diferencia. La geometría del problema aparece claramente en la figura 9.

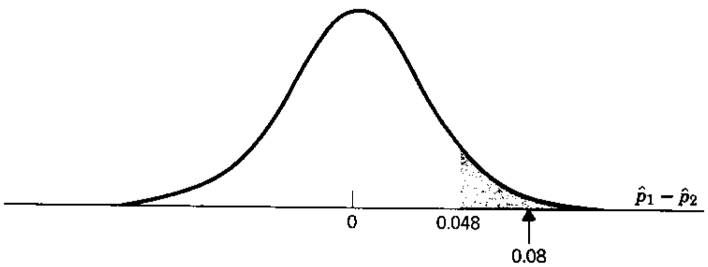

Figura 2. Distribución y región crítica para $\hat{p}_1 - \hat{p}_2$ en el problema de la vitamina C

Basándose en este experimento particular se concluye que fuertes dosis de vitamina C ayudan a evitar las enfermedades del sistema respiratorio.

Referencias:

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

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