Homogeneidad de muestras. Proporciones

En el tema anterior se había puesto de manifiesto la necesidad absoluta de la homogeneidad de las muestras. Si la muestra correspondiente al grupo de control no se ha extraído de la misma población que la muestra para el grupo de intervención, no se podrá posteriormente concluir que una determinada intervención farmacéutica implementada sobre una población concreta presenta o no una cierta eficacia, puesto que cada muestra proviene de dos subpoblaciones diferentes o, al menos, no son homogéneas.

En este tema se analizó cómo evaluar esta característica con las variables cuantitativas, expresadas como valores medios, mientras que en el presente se detallará cómo realizarlo cuando la variable es de tipo cualitativo, por lo que se expresará como proporciones.

Análisis de valores como proporciones

La coexistencia o no de un factor de riesgo tiene, obviamente, una influencia en la prevalencia de la enfermedad resultante. Por este motivo, por ejemplo, la presencia de hipertensión entre los individuos que entran a formar parte de un estudio de intervención farmacéutica en diabetes, tiene un componente negativo en su efectividad. Es obvio, entonces, que la proporción de hipertensos en el grupo de control y en el de intervención tiene que ser similar con el fin de que las dos muestras sean homogéneas. De otra manera, no se podría concluir si la mayor o menor efectividad evaluada es consecuencia de la intervención o de la ausencia o no de presión arterial elevada, que actúa como un factor que altera el resultado.

La metodología básica es similar a la que se ha explicado en el tema anterior; consiste en demostrar la hipótesis nula que indica que la diferencia entre los porcentajes de una y otra muestra de la variable sería 0 de promedio, es decir:

$$H_0 : [p_i - p_m] = d = 0$$

La hipótesis alternativa sería:

$$H_1 : [p_i - p_m] = d \neq 0$$

Donde p_i es el porcentaje estimado para la variable cualitativa (p. ej., prevalencia de un factor de riesgo) en la muestra del grupo de intervención y p_m el correspondiente del grupo control.

Si se extrajeran N muestras de control y N muestras de intervención, el valor promedio de las N diferencias (d_N) tendría que ser 0 con un rango de más/menos el doble del error estándar. Si la diferencia d_N fuera mayor del rango definido, significaría que ésta no puede ser explicada únicamente por el azar, por lo que las muestras no serían homogéneas, al presentar una diferencia excesiva de sus porcentajes, lo que indicaría que no proceden de una misma población homogénea. Por el contrario, si esta diferencia d_N fuera menor de estos límites, la diferencia entre p_i y p_m puede justificarse estrictamente por el azar, lo que confirmaría que las dos muestras son homogéneas.

La tabla de contingencia

Si se considera que las muestras del grupo de intervención y control tienen un tamaño de n_i y n_c individuos, respectivamente, y en ellas hay a y b individuos, respectivamente, que presentan la variable cualitativa analizada, es natural que la proporción esperada (p_e) de prevalencia de esta variable (que se corresponde con la media ponderada) sea similar a la estimada para cada grupo.

La idea inicial consiste en considerar, e intentar demostrar, que como la dos muestras son homogéneas, cualquiera de ellas presentaría una proporción esperada (p_g) similar. Más exactamente, las diferencias entre los valores esperados y los reales podrían explicarse sola–

Tabla 1. Tabla de contingencia con las frecuencias obtenidas

	Variable cualitativa Sí	No	
Muestra de intervención	A	b	n_i
Muestra de control	C	d	n_c
Total	a + c	b + d	$n_i + n_c$

Tabla 2. Tabla de contingencia de valores obtenidos

Valores obtenidos	HTA (Sí)	HTA (No)	
Muestra de intervención	12	20	32
Muestra de control	10	21	31
Total	22	41	63

Tabla 3. Tabla de contingencia de valores esperados

Valores esperados	HTA (Sí)	HTA (No)	
Muestra de intervención	32*0,35 = 11,17	20,83	32
Muestra de control	31*0,35 = 10,83	20,17	31
Total	22	41	63

mente por el azar, si efectivamente hay homogeneidad de muestras.

Para ello se dibuja la tabla de contingencia, donde se representan numéricamente los datos obtenidos (tabla 1):

A continuación se representa la tabla de contingencia con los valores que cabrían esperarse a partir de los datos iniciales. Después se estima el estadístico c^2.

Para considerar las dos muestras homogéneas, el valor de éste para un grado de libertad (el usado cuando se emplean dos muestras solamente) y para un valor habitual de confianza del 95%, no debe superar a 3,84. En caso de que c^2 fuera mayor de este valor, las muestras no serían homogéneas. El programa Excel ofrece directamente el valor de p, por lo que se puede calcular después el valor del estadístico c^2.

Cálculo práctico

Lo explicado se verá mejor mediante un ejemplo resuelto. Supongamos que se implementa una determinada intervención en una muestra de pacientes diabéticos; se instaura un cuidado usual en otra que actúa como control. La presencia de hipertensión arterial se considera como una variable que puede afectar al resultado final, por lo que las dos muestras tienen que ser homogéneas respecto de esta variable cualitativa. Analizada ésta, se hallan los datos representados en la tabla de contingencia (tabla 2), que se anota en una hoja de Excel (haciendo coincidir la celda A1 con «Valores obtenidos»).

A continuación, en la misma hoja de Excel, pero haciendo coincidir la celda A8 con «Valores esperados», se estima su tabla de contingencia (tabla 3). En la celda B6 se calcula la proporción esperada de pacientes hipertensos uniendo las dos muestras (B4/D4), que es de:

22/63 = 0,349

de modo que en el grupo de intervención tendría que haber:

32 ¥ 0,349 = 11,17 pacientes hipertensos

y en el control:

31 ¥ 0,349 = 10,83 pacientes hipertensos

aplicando la misma proporción a ambos grupos, al suponer que ambos son homogéneos (hipótesis nula). Obviamente, la proporción esperada de pacientes no hipertensos será la diferencia hasta el tamaño de cada muestra:

32 – 11,17 = 20,83 en el de intervención y
31 – 10,83 = 20,17 en el de control.

A partir de las dos tablas de contingencia se calcula el valor de «p» (mediante la función PRUEBA.CHI) y la de «c^2» (mediante la función PRUEBA.CHI.INV).

En la celda C13 se estima p:

[=PRUEBA.CHI(B2:C3;B9:C10)]

que ofrece un valor de 0,6626 a partir del que se estima en C15 c^2 :

[=PRUEBA.CHI.INV(C13;1)]

que da finalmente el valor de 0,1904. Dado que este valor es menor de 3,84 —mencionado previamente para un valor de significación del 95%— se puede aceptar la hipótesis nula si se considera las muestras homogéneas.

En el caso de que los resultados del grupo control hubieran sido de 21 hipertensos y 10 no hipertensos, con los mismos valores mantenidos que antes en el grupo de intervención, los valores de p y c^2 estimados hubieran sido de 0,0163 y 5,773. En este último caso, al ser mayor de 3,84, se hubiera rechazado la hipótesis nula, con lo que se hubiese estimado que las muestras no son homo-

PDF
Consultora científica.