Tópicos en investigación clínica y epidemiológica

PRUEBA CHI-CUADRADO DE INDEPENDENCIA APLICADA A TABLAS $2_X N$

Fredy Mendivelso [1], Milena Rodríguez [2]

[1] *MD. MPH. MSc. FETP. Clínica Reina Sofía*
[2] *MD. Veterinaria. Epidemióloga. MSc. FETP. Fundación Universitaria Sanitas*

RESUMEN

La prueba ji-cuadrado (X^2) de Pearson es una de las técnicas estadísticas más usadas en la evaluación de datos de conteo o frecuencias, principalmente en los análisis de tablas de contingencia ($r_X c$) donde se resumen datos categóricos.

Palabras clave: Distribución de Chi-Cuadrado, Pruebas de Hipótesis, Interpretación Estadística de Datos, Investigación biomédica.
DOI: 10.26852/01234250.6

INDEPENDENCE CHI-SQUARE TEST APPLIED TO $2_X N$ TABLES

ABSTRACT

Pearson chi-square test (X^2) is one of the most used statistical techniques in the assessment of data counting or frequencies, mainly in the analysis of contingency tables ($r_X c$) where categorical data are summarized.

Keywords: Chi-Square Distribution; Hypothesis-Testing, Data Interpretation, Statistical; Biomedical Research

Recibido: 1 de junio de 2018
Aceptado: 6 de junio de 2018
Correspondencia: fmendivelso@colsanitas.com

INTRODUCCIÓN

La X^2 es una prueba de libre distribución (no paramétrica) que mide la discrepancia entre una distribución de frecuencias observadas y esperadas. Dentro de sus características generales, la prueba X^2 toma valores entre cero e infinito y no tiene valores negativos porque es la suma de valores elevados al cuadrado (1).

Existen tres usos relevantes de la prueba X^2:

- Prueba de bondad de ajuste (una variable)
- Prueba de independencia (dos variables)
- Prueba de homogeneidad (dos variables)

En esta publicación, vamos a realizar mayor énfasis al uso que se da a la X^2 como prueba de independencia (2).

USO DE LA PRUEBA

Investigar la diferencia en valores de frecuencias cuando se clasifica una muestra "n" por un atributo "A" y después se realiza una nueva clasificación de "A" por un segundo atributo "B"(3).

SUPUESTOS

- Se considera que los datos provienen de una muestra aleatoria extraída de la población de interés.
- La muestra debe ser lo suficientemente grande

HIPÓTESIS

H_0: **No** hay asociación entre las variables A|B *(Las variables son independientes)*

H_1: **Si** hay asociación entre las variables A|B *(Las variables no son independientes)*

PRUEBA ESTADÍSTICA

$$X^2 = \sum_{i=1}^{k} \left[\frac{(0_i - E_i)^2}{E_i} \right]$$

Donde:

Oi: Valor observado
Ei: Valor esperado

DISTRIBUCIÓN DE LA PRUEBA ESTADÍSTICA

Cuando H_0 es verdadera, sigue una distribución X^2 con (r-1)(c-1) grados de libertad. (r: número de filas y c: número de columnas en la tabla de contingencia)

LIMITACIONES

La muestra debe ser lo suficientemente grande. Si menos del 20% de las celdas de la tabla de contingencia, presentan valores esperados ≤5 no se recomienda aplicar la prueba X^2 y optar por la alternativa del test exacto de Fisher(4).

EJEMPLO

Un investigador recolecta información sobre los patrones de actividad física (AF) de los niños de quinto grado de primaria de una escuela pública. Define tres categorías de AF (1, Baja; 2. Media; 3. Alta). También indaga sobre consumo regular de bebidas azucaradas en la escuela y define dos categorías (1. Si consume; 0. No consume). Su interés es evaluar si existe una asociación entre los patrones de AF y el consumo de bebidas azucaradas en los niños de esta institución escolar con un nivel de significancia del 5%.

Los resultados se muestran en el siguiente tabla:

TABLA 1. VALORES OBSERVADOS

		Bebidas azucaradas		Total
		Si	No	
AF	Baja	32	12	44
	Media	14	22	36
	Alta	6	9	15
	Total	52	43	95

Paso 1: Defina la hipótesis de trabajo:

H_0: **No** hay asociación entre la práctica de actividad física y el consumo de bebidas azucaradas en este grupo de escolares

H_1: **Si** hay asociación entre la práctica de actividad física y el consumo de bebidas azucaradas en este grupo de escolares

Paso 2: Defina el nivel de significancia para la prueba estadística

$$\propto\ = 0{,}05$$

Paso 3: Calcule los grados de libertad (gl) para la prueba (r = # filas, c = # columnas)

gl = (r-1) (c-1)
gl = (3-1) (2-1)
gl = (2) (1)
gl = 2

Paso 4: Establezca el valor de critico (rechazo) de la H_0 para la distribución X^2

Con los valores calculados de alfa y grados de libertad, se consulta en una tabla de distribución de probabilidad X^2 su valor crítico. Para nuestro ejemplo (consultar una tabla de distribución de probabilidad X^2 con los valores gl = 3 y alfa = 0,05 este valor es igual a 7,815)

$$X^2_{2;0,05} = 7{,}815$$

Paso 5: Calcule el valor para el estadístico de contraste (X^2 para los datos del ejemplo:

Ya cuenta con los datos de los valores observados en el cuadro uno. Los valores esperados en cada celda se calculan como el producto aritmético entre sus valores marginales (color gris) dividido por el total de observaciones así:

TABLA 2. CÁLCULO DE LOS VALORES ESPERADOS

		Bebidas azucaradas		Total
		Si	No	
AF	Baja	(52x44)/95	(43x44)/95	44
	Media	(52x36)/95	(43x36)/95	36
	Alta	(52x15)/95	(43x15)/95	15
	Total	52	43	95

Luego tendremos que los valores esperados son:

TABLA 3. VALORES ESPERADOS PARA EL EJEMPLO

		Bebidas azucaradas		
		Si	No	
AF	Baja	24,1	19,9	
	Media	19,7	16,3	
	Alta	8,2	6,8	

Paso 6. Valide el supuesto de que menos del 20% de las celdas en la tabla tiene valores esperados ≤5. En este caso se cumple el supuesto y no es necesario acudir a una prueba exacta como el test exacto de Fisher.

Paso 7. Calcule el valor de X^2 para el ejemplo usando la formula descrita inicialmente

$$X^2_{2;0,05} = \left(\frac{(32-24{,}1)^2}{24{,}1}\right) + \left(\frac{(14-19{,}7)^2}{19{,}7}\right) + \left(\frac{(6-8{,}2)^2}{8{,}2}\right) + \left(\frac{(12-19{,}9)^2}{19{,}9}\right) + \left(\frac{(22-16{,}3)^2}{16{,}3}\right) + \left(\frac{(9-6{,}8)^2}{6{,}8}\right)$$

$$X^2_{2;0,05} = 10{,}7$$

Paso 8. Regla de decisión

Como el valor calculado de X^2 para el ejemplo es 10,7 y previamente establecimos que el punto crítico para la distribución X^2 con un alfa de 0,05 y 3 gl es igual a 7,8 podemos afirmar que nuestro valor del estadístico de prueba está dentro de la zona de rechazo de H_0 formulada a priori.

Paso 9. Interpretación

Con los datos de nuestro estudio, tenemos suficiente evidencia para rechazar la H_0 de que

No hay asociación entre la práctica de actividad física y el consumo de bebidas azucaradas en este grupo de escolares. Cuidado; un error frecuente es afirmar que en consecuencia, se acepta la H_1. Es necesario recordar que el análisis de pruebas estadísticas y su interpretación se hacen siempre sobre la H_0.

CONCLUSIÓN

La prueba X^2 es muy usada en el análisis de datos de investigación biomédica, su aplicación es relativamente sencilla al igual que su interpretación. Múltiples estudios en ciencias sociales e investigación clínica hacen uso de esta prueba en tablas de 2_x2 (n_xn); sin embargo, esta puede ser aplicada en tablas relativamente más complejas (n_xN) siempre y cuando se cumplan los supuestos necesarios para su aplicación

REFERENCIAS

1. McHugh ML. The chi-square test of independence. Biochemia medica. 2013;23(2):143-9.
2. Betensky RA, Rabinowitz D. Maximally selected chi2 statistics for k x 2 tables. Biometrics. 1999;55(1):317-20.
3. Gibbons JD, Chakraborti S. Nonparametric statistical inference. International encyclopedia of statistical science: Springer; 2011. p. 977-9.
4. Agresti A. Categorical data analysis: John Wiley & Sons; 2003.