



## Investigación clínica XVII

# Prueba $\chi^2$ , de lo esperado a lo observado

Rodolfo Rivas-Ruiz,<sup>a</sup> Osvaldo D. Castelán-Martínez,<sup>a</sup> Marcela Pérez,<sup>a</sup> Juan O. Talavera<sup>a</sup>

### Clinical research XVII. $\chi^2$ test, from the expected to the observed

When you want to show if there is a statistical association or differences between categorical variables, it is recommended to use the  $\chi^2$  test. This nonparametric test is one of the most used in clinical research; it contrasts nominal or ordinal qualitative variables that are observed in clinical practice. This test calculates the p value that determines whether differences between groups are real or due to chance. The  $\chi^2$  test is the basis of other tests to analyze qualitative ordinal variables as  $\chi^2$  for linear trend, which compares three groups with two outcomes or McNemar test which contrasts two related samples (a before and afterward comparison) or Mantel-Haenszel  $\chi^2$ , which controls for potential confounding variables. When using small samples, where the expected is less than 5, Fisher's exact test should be used. These tests are the most widely used in the medical literature, however, do not give us the magnitude or the direction of the event and a proper interpretation that requires clinical judgment is needed.

#### Key words

chi-square  
qualitative variables  
statistics

En la parte XVI de esta serie mostramos que para analizar variables cuantitativas de libre distribución se utiliza la *U* de Mann-Whitney. En este artículo, al igual que en el anterior, presentamos cómo se estudian las variables categóricas (cualitativas) de libre distribución, ya sea ordinales o nominales. Este tipo de pruebas están incluidas en la llamada *estadística no paramétrica*. De este grupo, la más característica es la  $\chi^2$  de Pearson, la cual contrasta la asociación entre dos grupos cuyo resultado es una variable dicotómica. Casi todas las investigaciones médicas utilizan esta prueba estadística, por el tipo de variable que contrasta. En la mayoría de los artículos, en el cuadro estadístico en el que se muestra que el estado basal de los sujetos es similar, se emplea la prueba de  $\chi^2$  para las variables categóricas. También se utiliza en los cuadros estadísticos de los resultados del estudio, para mostrar las diferencias entre los grupos analizados.

Esta prueba se basa en la diferencia entre lo observado en un experimento y lo que podría suceder por simple azar. La fórmula de  $\chi^2$  expresa la sumatoria de lo observado con lo esperado:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Donde:

O = eventos observados en el experimento.  
E = eventos esperados por el azar.

Antes de iniciar el uso de  $\chi^2$  deben considerarse dos requisitos indispensables:

- Las frecuencias que se usan en este estadístico representan cuentas individuales y no rangos ni medidas de resumen.
- Las categorías son exhaustivas y mutuamente excluyente entre ellas, por lo tanto, cada individuo puede ser asignado a una categoría, pero únicamente a una.

Con esto en mente, pensemos en el siguiente ejemplo: si lanzamos 100 veces una moneda al aire, la hipótesis nula o de igualdad de proporciones (dado que en un principio debe asumirse que la moneda no tiene trampa) es que 50 veces caerá un lado de la moneda hacia arriba (cara) y 50 veces, el otro (cruz). El resultado es que el lado de la cara cae 47 veces y el lado de la cruz, 53 veces. Por lo tanto, las opciones posibles son dos: existe truco o el resultado se debe al azar. Para verificar qué es lo que sucede, realizamos la prueba de  $\chi^2$  en la que sustituimos de la siguiente forma:

Cuando se busca demostrar que existe relación o diferencias estadísticamente significativas, entre las variables categóricas se utiliza la prueba de  $\chi^2$ . Esta prueba no paramétrica es una de las más usadas en la investigación clínica, ya que contrasta variables cualitativas nominales u ordinales observadas en la práctica clínica. Con esta prueba se calcula el valor de  $p$  para decidir si las diferencias entre los grupos son reales o se deben al azar. Esta prueba genéricamente recibe el nombre de  $\chi^2$  o chi cuadrada (dependiendo de las preferencias idiomáticas) y es la base de otras pruebas con las que se analizan las variables cualitativas ordinales: la  $\chi^2$  de tendencia lineal, con la que se comparan tres grupos que pueden tener dos desenlaces; la prueba de McNemar, que sirve para

contrastar dos muestras relacionadas (antes y después de una maniobra); o la  $\chi^2$  de Mantel-Haenszel, que controla el efecto de variables potencialmente confusoras. Para muestras pequeñas, en las que el resultado esperado es menor de 5, se debe utilizar la prueba exacta de Fisher. Como se puede ver, este grupo de pruebas es el de mayor uso en los artículos médicos, sin embargo, no dan cuenta de la magnitud ni de la dirección del evento, por lo que su adecuada interpretación requiere del juicio clínico.

## Resumen

### Palabras clave

chi-cuadrada  
variables cualitativas  
estadística

$$\text{Cara} = (O - E)^2/E = (47 - 50)^2/50 = (3)^2/(50) = 0.18$$

$$\text{Cruz} = (O - E)^2/E = (53 - 50)^2/50 = (3)^2/(50) = 0.18$$

Finalmente se realiza la sumatoria ( $\Sigma$ ) de estas dos posibilidades ( $0.18 + 0.18$ ), cuyo resultado es igual a 0.36, que corresponde al valor de  $\chi^2$ .

Para determinar si es o no significativo, este valor se busca en la tabla de valores críticos de  $\chi^2$  (también llamada de contingencias), en el valor de  $p$  con

1 grado de libertad (gl): 0.36 se encuentra entre los valores críticos  $p = 0.9$  y  $p = 0.1$  y como está más cerca del primero se asume que le corresponde este. Dado que el valor de alfa (con significación) por lo general es  $\leq 0.05$ , entonces el valor de  $\chi^2$  (al que le corresponde una  $p = 0.9$ ) no es significativo (figura 1).

El modelo de  $\chi^2$  tiene múltiples usos y particularidades, no solo funciona con variables cualitativas para dos grupos, si no que puede ser usado por tres grupos y con muestras pequeñas.

gl	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.0	0	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.01	0.02	0.051	0.103	0.216	4.605	5.991	7.378	9.21	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.948	11.45	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.86
5	0.412	0.554	0.831	1.145	1.61	9.236	11.071	12.833	15.086	16.75

Valor de  $p$

Localizamos el valor de  $p$  que le corresponde a nuestra  $\chi^2$

Figura 1 Tabla de valores críticos de  $\chi^2$

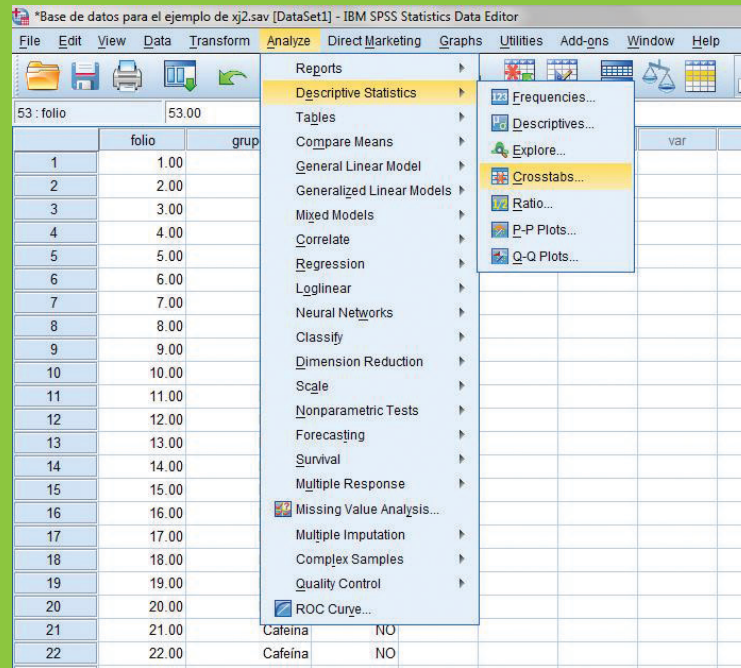
## Ejemplo de un análisis con el programa estadístico SPSS

Se desea determinar si la aplicación de cafeína comparada con la de teofilina es eficaz para evitar una nueva intubación en los recién nacidos prematuros con dificultad para respirar. Se diseña un ensayo clínico aleatorizado con dos grupos:

- Uno al que se le administra cafeína.
- Otro al que se le administra teofilina.

1

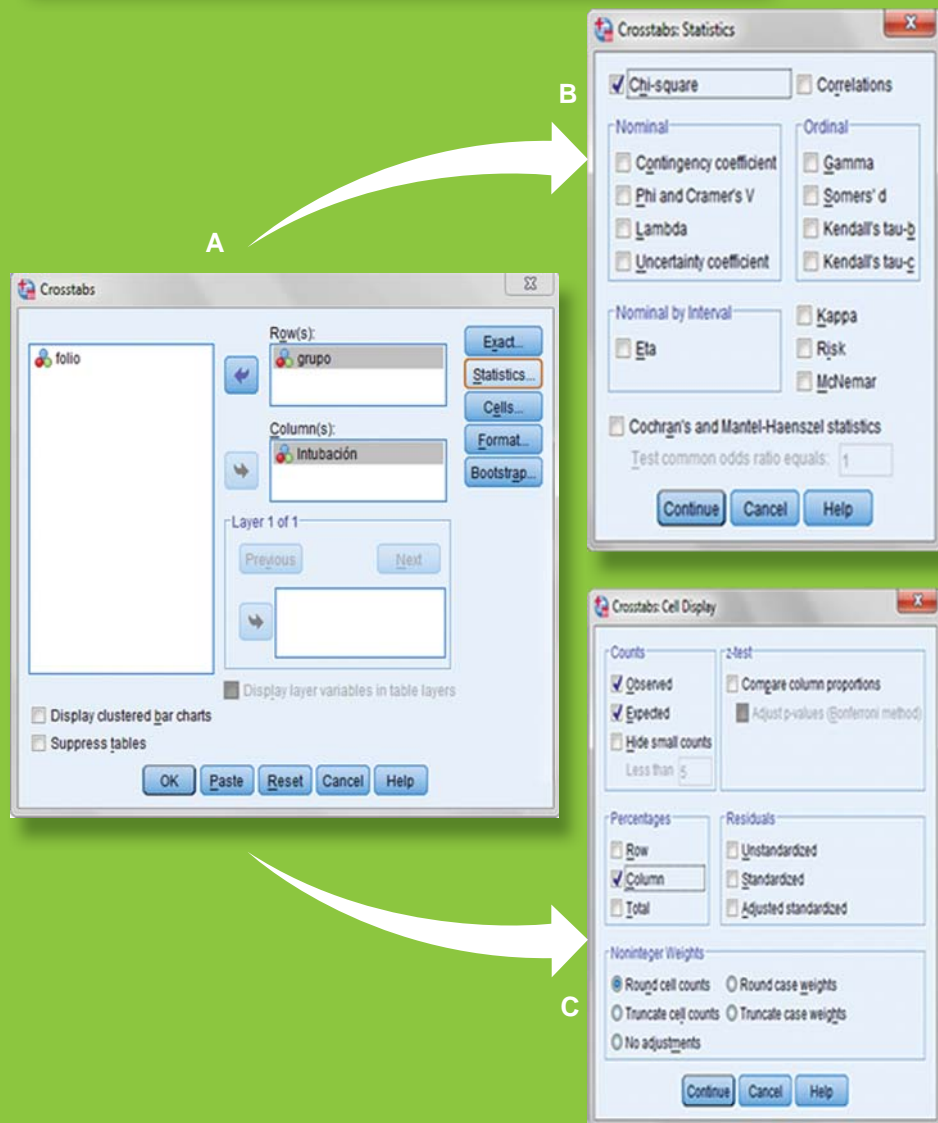
En el menú “Analizar” se selecciona la opción “estadísticos descriptivos” y en ésta, las tablas de  $2 \times 2$  (*crosstabs*).



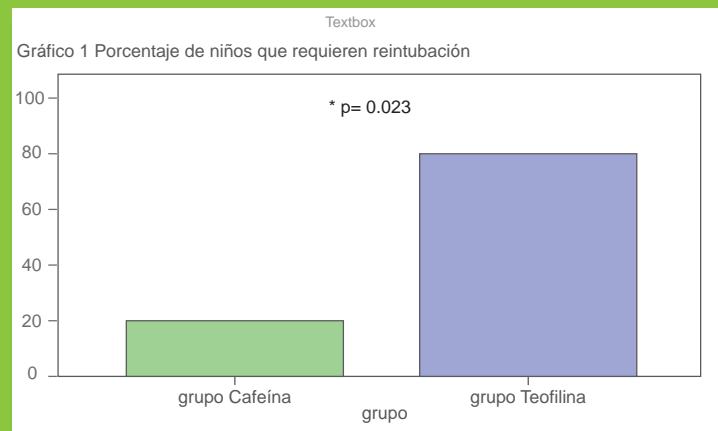
2

Al desplegar el menú de *crosstabs* (A), se selecciona la prueba estadísticas y después la prueba de  $\chi^2$  (B).

Una buena opción es seleccionar celdas o *cells* en el menú de *crosstabs* (A) y al desplegarse (C), activar las casillas de los valores observados y los esperados, estos últimos con la finalidad de seleccionar la prueba apropiada, ya que si en alguna de las celdas el valor esperado es menor de 5 se debe seleccionar la prueba exacta de Fisher (el programa la calcula automáticamente al seleccionar la prueba de  $\chi^2$ ).



Finalmente, una manera de saber la magnitud y dirección de esta comparación es elegir los porcentajes de respuesta por columna, de esta manera será posible determinar qué tanto funciona la cafeína en los niños prematuros al compararla con la teofilina.



### 3

En una tabla de  $2 \times 2$  se puede ver el resultado de comparar la proporción de reintubación en el grupo con cafeína y en el grupo con teofilina. Por convención, en las columnas se tabula la variable dependiente, en este caso, la reintubación.

		Reintubación		Total
		Sí	No	
Cafeína	Recuento	2	38	40
	Frecuencia esperada	5.5	34.5	40.0
	% dentro de reintubación	18.2	55.1	50.0
Teofilina	Recuento	9	31	40
	Frecuencia esperada	5.5	34.5	40.0
	% dentro de reintubación	81.8	44.9	50.0
Total	Recuento	11	69	80
	Frecuencia esperada	11.0	69.0	80.0
	% dentro de reintubación	100.0	100.0	100.0

### 4

Abajo de la tabla de  $2 \times 2$  aparecen las pruebas de  $\chi^2$ . La dificultad que tiene esta salida es la correcta identificación del valor  $p$  correspondiente. Debido a que nuestro ejemplo se trata de una comparación de grupos independientes y en las celdas el valor esperado más pequeño es de 5.5, se usa la prueba de  $\chi^2$  de Pearson ( $p = 0.023$ ). Si en alguna celda hubiera un valor  $< 5$  se usaría el estadístico exacto de Fisher con significación bilateral ( $p = 0.048$ ). Si hubiera tres grupos (si quisiéramos comparar un grupo con cafeína, otro con teofilina y un grupo placebo), se usaría la asociación lineal por lineal ( $p = 0.024$ ).

Pruebas de chi-cuadrado					
	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi cuadrado de Pearson	5.165 <sup>a</sup>	1	0.023		
Corrección por continuidad <sup>b</sup>	3.794	1	0.051		
Razón de verosimilitudes	5.530	1	0.019		
Estadístico exacto de Fisher				0.048	0.024
Asociación lineal por lineal	5.100	1	0.024		
N de casos válidos	80				

<sup>a</sup>0 casillas (0.0 %) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 5.50

<sup>b</sup>Calculado solo para una tabla de  $2 \times 2$

Para conocer la magnitud, se debe ver la tabla de  $2 \times 2$  ( **3** ), en la parte donde se calculan los porcentajes. En este ejemplo, el grupo de niños tratados con cafeína tuvo una incidencia de reintubación de 18.2 %, mientras que la incidencia de intubación del grupo de teofilina fue de 81.8 %. La prueba de  $\chi^2$  muestra que la diferencia entre estos dos grupos es estadísticamente significativa. ( $p = 0.023$ ) ( **4** ).

### Prueba exacta de Fisher

Es una prueba estadística alternativa a la  $\chi^2$  para comparar dos proporciones binomiales independientes en una muestra pequeña, en la que en alguna de las casillas el valor esperado es  $< 5$ . Calcula las probabilidades exactas de obtener un resultado determinado cuando la hipótesis nula es cierta. Es muy útil cuando la aproximación normal a la distribución binomial podría no ser aplicable, por ejemplo, cuando se presentan pocas observaciones en una de las casillas de la tabla de  $2 \times 2$  o en proporciones extremas.

### Prueba de McNemar (antes y después)

La prueba de McNemar es un caso especial de la prueba binomial que se usa para comparar dos proporciones de muestras pareadas, por ejemplo en un estudio de antes y después en el que un paciente o un grupo es su propio control. Con frecuencia se emplea en los ensayos clínicos cuyos resultados son variables dicotómicas que se registran dos veces para cada paciente bajo diferentes condiciones y se busca evaluar los cambios ocurridos al transcurrir un periodo. Las condiciones pueden representar diferentes tratamientos o diferentes tiempos de medición. El objetivo es comparar las tasas de respuesta en los dos conjuntos de condiciones. Debido a que las mediciones provienen de los mismos pacientes, no es posible asumir que los grupos sean independientes, por tal motivo no es posible utilizar la  $\chi^2$  o la prueba exacta de Fisher.

Estudios típicos, más específicos, en los que se aplica la prueba de McNemar son los que comparan las tasas de respuesta de dos tratamientos oftálmicos cuando ambos se proporcionan a cada paciente, uno en cada ojo.

### $\chi^2$ de Mantel-Haenszel

Cuando se reconoce que la asociación observada puede ser influida por una variable externa es pertinente utilizar la  $\chi^2$  de Mantel-Haenszel (variante de la  $\chi^2$  de Pearson), cuya función es controlar el efecto de una variable potencialmente confusora sobre la variable de resultado, por ejemplo: al estudiar la asociación entre el consumo de café y el desarrollo de gastritis, se

identifica que el consumo de café está asociado con el hábito tabáquico, que, a su vez, se asocia con el riesgo de gastritis, por lo tanto, se debe investigar si la asociación entre consumo de café y gastritis es independiente del hábito tabáquico. Es importante calcular el estimador de riesgo (*odds ratio* o riesgo relativo) de gastritis, después calcular el estimador de riesgo por estratos de la variable potencialmente confusora, es decir, el riesgo de gastritis en quienes fuman y en quienes no fuman. Después se calcula el estimador de riesgo ajustado de Mantel-Haenszel para el análisis estratificado, con base en los riesgos interestrato. Para saber si la asociación entre consumo de café y la gastritis es independiente del hábito tabáquico se calcula la  $\chi^2$  de Mantel-Haenszel. El valor que arroja esta prueba se busca en la tablas de  $\chi^2$ : si es mayor al valor crítico para un grado de libertad, la asociación es estadísticamente significativa independientemente del hábito tabáquico.

### $\chi^2$ de tendencia lineal

También denominada asociación lineal por lineal, es la prueba más apropiada para analizar situaciones en las que una variable de exposición ordenada tiene tres o más categorías (como en el ejemplo anterior) y la variable de resultado es de tipo binario.

### Conclusiones

La  $\chi^2$  es la base de la mayoría de las pruebas estadísticas que se usan en la práctica clínica. Se basa en la diferencia de lo observado menos lo esperado, fenómeno que más se asocia con la práctica clínica, en la que el resultado es si un paciente mejora o no o si una maniobra previene o no una condición. Esta prueba solo ayuda a saber si el resultado se debe a la maniobra y no al azar. Para determinar la magnitud del evento se requiere observar los porcentajes de respuesta por grupo, graficar esta o aplicar otras pruebas de asociación, como la razón de momios y el riesgo relativo.

**Declaración de conflicto de interés:** los autores ha completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno en relación con este artículo.

<sup>a</sup>Centro de Adiestramiento en Investigación Clínica, Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Distrito Federal, México

Comunicación con: Rodolfo Rivas-Ruiz.  
Correo electrónico: rivasrodolfo@gmail.com



**Bibliografía**

1. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959;22(4):719-48.
2. Henderson-Smart DJ, Davis PG. Prophylactic methylxanthines for endotracheal extubation in preterm infants. *Cochrane Database Syst Rev.* 2010;(12):CD000139.
3. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. New Jersey: Pearson/Prentice Hall; 2009.
4. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. Fourth edition. Massachusetts: Blackwell Science; 2002.
5. Peat JK, Barton B. Medical statistics: a guide to data analysis and critical appraisal. Massachusetts: Blackwell 2005.
6. Le Chap T. Introductory biostatistics. New Jersey: Wiley-Interscience; 2003.
7. Bland M. Introduction to medical statistics. Third edition. US: Oxford University Press; 2003.
8. Dawson B, Trapp R. Basic and clinical biostatistics. Fourth edition. US: Lange Medical Books-McGrawHill; 2004.
9. Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. *Rev Med Inst Mex Seguro Soc.* 2011;49(3):289-94. Texto libre en [http://revistamedica.imss.gob.mx/index.php?option=com\\_multicategories&view=article&id=1410:investigacion-clinica-iii-estudios-de-causalidad&catid=501:temas-de-actualidad&Itemid=664](http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=1410:investigacion-clinica-iii-estudios-de-causalidad&catid=501:temas-de-actualidad&Itemid=664)
10. Talavera JO, Rivas-Ruiz R. Clinical research IV. Relevancy of the statistical test chosen. *Rev Med Inst Mex Seguro Soc.* 2011;49(4):401-5. Texto libre en [http://revistamedica.imss.gob.mx/index.php?option=com\\_content&view=article&id=1432:investigacion-clinica-iv-pertinencia-de-la-prueba-estadistica&catid=139:2011&Itemid=511](http://revistamedica.imss.gob.mx/index.php?option=com_content&view=article&id=1432:investigacion-clinica-iv-pertinencia-de-la-prueba-estadistica&catid=139:2011&Itemid=511)
11. Talavera JO, Rivas-Ruiz R. Clinical research V. Sample size. *Rev Med Inst Mex Seguro Soc.* 2011;49(5):517-22. Texto libre en [http://revistamedica.imss.gob.mx/index.php?option=com\\_content&view=article&id=1452:investigacion-clinica-v-tamano-de-muestra&catid=139:2011&Itemid=511](http://revistamedica.imss.gob.mx/index.php?option=com_content&view=article&id=1452:investigacion-clinica-v-tamano-de-muestra&catid=139:2011&Itemid=511)
12. Rivas-Ruiz R, Pérez-Rodríguez M, Talavera JO. Del juicio clínico al modelo estadístico. Diferencia de medias. Prueba *t* de Student. *Rev Med Inst Mex Seguro Soc.* 2013;51(3):300-3. Texto libre en [http://revistamedica.imss.gob.mx/index.php?option=com\\_multicategories&view=article&id=2077:51-3-11-investigacion&catid=785:temas-de-actualidad&Itemid=775](http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=2077:51-3-11-investigacion&catid=785:temas-de-actualidad&Itemid=775)
13. Rivas-Ruiz R, Moreno-Palacios J, Talavera JO. Investigación clínica XVI. Diferencia de medianas con la U de Mann-Whitney. *Rev Med Inst Mex Seguro Soc* 2013;51(4):414-9.