

Investigación enfermera: herramientas y conceptos

La prueba chi-cuadrado o ji-cuadrado (χ^2)

Romero Saldaña, M. Enfermero Especialista en Enfermería del Trabajo.
Ayuntamiento de Córdoba

I. Introducción

Dentro de la investigación sanitaria encontramos frecuentemente estudios que utilizan variables cualitativas, como por ejemplo, la variable sexo, puesto de trabajo, estado civil, nivel de formación académica, estar presente ausente un determinado factor de riesgo laboral o enfermedad, etc.

A su vez, estas variables cualitativas, están formadas por categorías que cumplen la condición de exhaustividad (representan todas las posibilidades de la variable) y son mutuamente excluyentes (un dato no puede estar ubicado simultáneamente en más de una categoría).

Generalmente, la forma de expresar relativamente estos datos es mediante el uso de porcentajes. Así por ejemplo, decimos que el 35,3% de la muestra de estudio presenta hipertensión arterial, el 24,8% son fumadores, un 8,2% tiene estudios universitarios, o el 65,7% son mujeres, etc.

En otras ocasiones, el/la investigador/a suele utilizar tablas de varias entradas en las que representa las frecuencias absolutas y los porcentajes obtenidos para cada una de las categorías. Nos referimos a las conocidas tablas de contingencia.

Tabla 1. Frecuencia de consumo de tabaco según la variable sexo.

	FUMADORES/AS	NO FUMADORES/AS	TOTAL
HOMBRES	55 (40,7%)	80 (59,3%)	135 (52,1%)
MUJERES	48 (38,7%)	76 (61,3%)	124 (47,9%)
TOTAL	103 (39,8%)	156 (60,2%)	259

Tabla 2. Prevalencia de hipertensión arterial según grupos de edad.

	HIPERTENSIÓN	NO HIPERTENSIÓN	TOTAL
Población Joven	5 (2,5%)	195 (97,5%)	200 (43,5%)
Población Madura	31 (20,7%)	119 (79,3%)	150 (32,6%)
Población Mayor	47 (42,7)	63 (57,3%)	110 (23,9%)
TOTAL	83 (18%)	377 (82%)	460

Si observamos la tabla 1, podremos afirmar que los hombres fuman más que las mujeres, al menos, porcentualmente. La tabla 2, muestra como la prevalencia de hipertensión aumenta a medida que lo hace el grupo de edad.

Sin embargo, las diferencias existentes pueden llegar a ser significativas o por el contrario, dichas diferencias son tan pequeñas que podrían ser explicadas por el azar, sesgos, etc.

Para responder a esta cuestión, el/la investigador/a debe recurrir a la prueba de chi-cuadrado.

2. La prueba chi-cuadrado

Posee tres aplicaciones básicas:

1. Test de ajuste de distribuciones: Es un contraste de significación para saber si los datos de una muestra son conformes a una ley de distribución teórica que consideramos como correcta.
2. Homogeneidad de varias muestras cualitativas: Sirve para contrastar la igualdad de procedencia de un conjunto de muestras de tipo cualitativo.
3. Test para tablas de contingencia: Es un contraste para determinar la dependencia o independencia de caracteres cualitativos.

Como vemos en las tablas 1 y 2, nos centraremos en esta tercera aplicación.

Así por tanto, la prueba χ^2 permite determinar si dos variables cualitativas están o no asociadas. Si al final del estudio concluimos que las variables no están relacionadas podremos decir con un determinado nivel de confianza, previamente fijado, que ambas son independientes.

El test χ^2 consiste en tomar una muestra y observar si hay diferencias significativas entre las frecuencias observadas y las especificadas por la ley teórica del modelo que se contrasta, también denominadas "frecuencias esperadas". O sea, el test o prueba χ^2 , compara dos tablas, una observada (de frecuencias obtenidas en nuestro estudio), con otra teórica (de frecuencias esperadas y calculadas teóricamente).

Fue Karl Pearson quien propuso el siguiente estadístico:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Donde:

O_i : Frecuencias observadas. Es el número de casos observados clasificados en una determinada celda de la tabla de contingencia.

E_i : Frecuencias esperadas o teóricas. Es el número de casos esperados correspondientes a cada celda de la tabla de contingencia. Se puede definir como aquella frecuencia que se observaría si ambas variables fuesen independientes.

Como en cualquier prueba de contraste estadístico, se intenta rechazar la hipótesis nula y aceptar en consecuencia, la hipótesis alternativa.

La hipótesis nula se corresponde con la independencia de las variables, o lo que es lo mismo, que las diferencias entre las frecuencias observadas y esperadas son muy pequeñas, y en consecuencia, el estadístico χ^2 también obtendrá un valor muy pequeño.

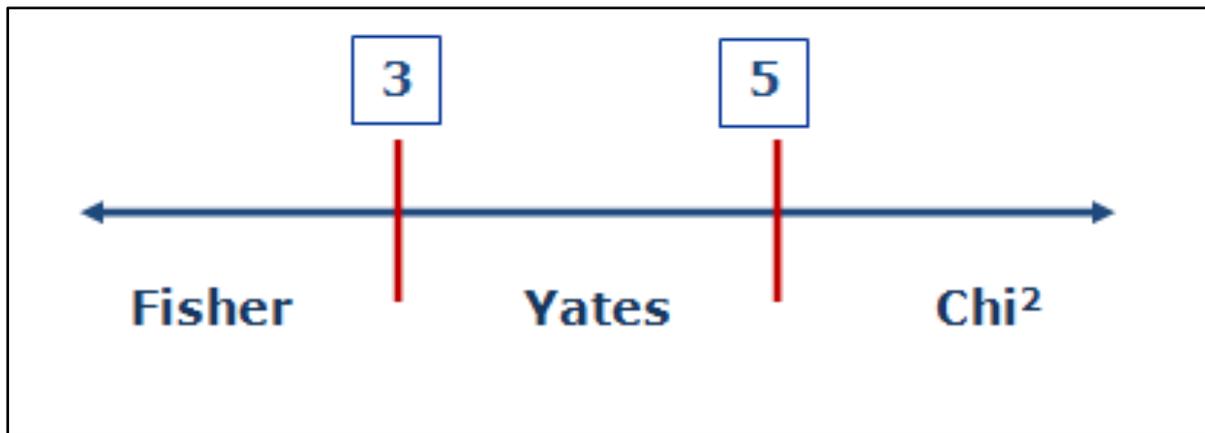
Al contrario, cuando el valor de χ^2 sea muy grande, pensaremos que las diferencias entre nuestros valores observados en el estudio y los esperados teóricamente son tan grandes que el azar no podría explicarlas, y por tanto, rechazaremos la hipótesis nula y aceptaremos la alternativa.

Existen dos consideraciones especiales en la aplicación de la prueba χ^2 cuando el tamaño muestral es pequeño, y se puedan encontrar celdas con frecuencias esperadas cuyo valor sea inferior a 5.

a) Si al menos el 25% de las frecuencias esperadas es inferior a 5 pero superior a 3, se ha de emplear la prueba χ^2 aplicando la corrección Yates.

b) Por el contrario, si al menos el 25% de las frecuencias esperadas es inferior a 3, se ha de emplear el test exacto de Fisher como alternativa a la prueba χ^2 .

Figura 1. Indicación del test exacto de Fisher, corrección de Yates o Chi2, según el valor de las frecuencias esperadas.

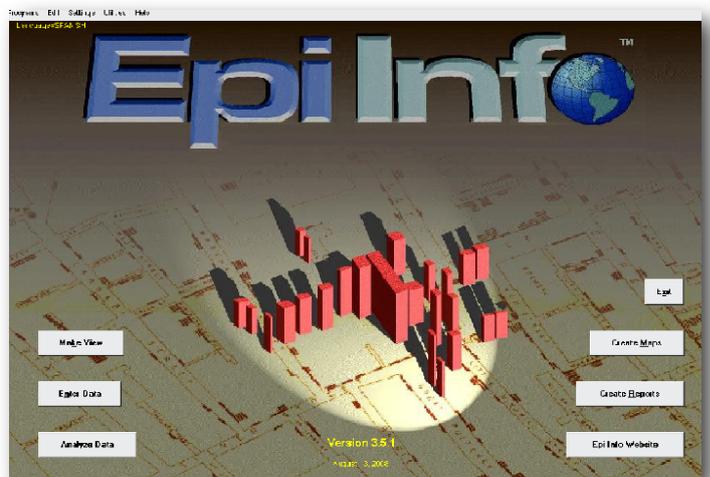
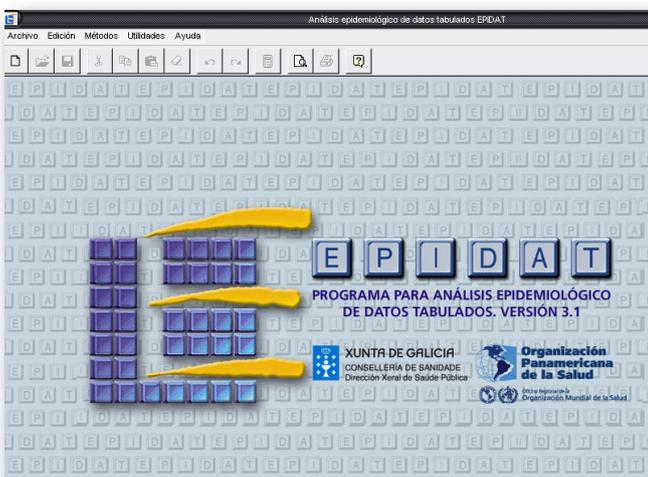


3. Herramientas de cálculo.

Si sabemos la utilidad de la prueba chi-cuadrado, el siguiente paso será conocer las herramientas informáticas-estadísticas a nuestro alcance para aplicarlas y ahorrarnos los cálculos matemáticos manuales.

Existen paquetes estadísticos gratuitos que podemos descargarlos fácilmente desde internet. Entre ellos, las más utilizadas son los programas EPIDAT y EPIINFO.

EPIDAT. Se puede descargar desde la página de la dirección general de salud pública del Servicio Gallego de Salud <http://dxsp.sergas.es/ApliEdatos/Epidat/cas/default.asp>, o también a través de la página organización panamericana de salud www.paho.org/spanish/sha/epifiles.htm.



EPIINFO. Se recomienda descargar desde la página oficial del CDC (Center for Disease Control) de Atlanta (EE.UU.) www.cdc.gov/epiinfo

4. Ejercicios prácticos

a) Ejercicio de la Tabla 1

Como se observa, el tamaño muestral es grande y todas las frecuencias esperadas son superiores a 5, por tanto, no hay que aplicar ni la corrección de Yates, ni el teste exacto de Fisher, respectivamente.

Utilizando el programa EPIDAT, se obtiene el siguiente resultado:



Tablas de contingencia: Tablas 2x2 simples

Origen de datos | Resultados

Tipo de estudio

Transversal

Cohortes

Caso-control

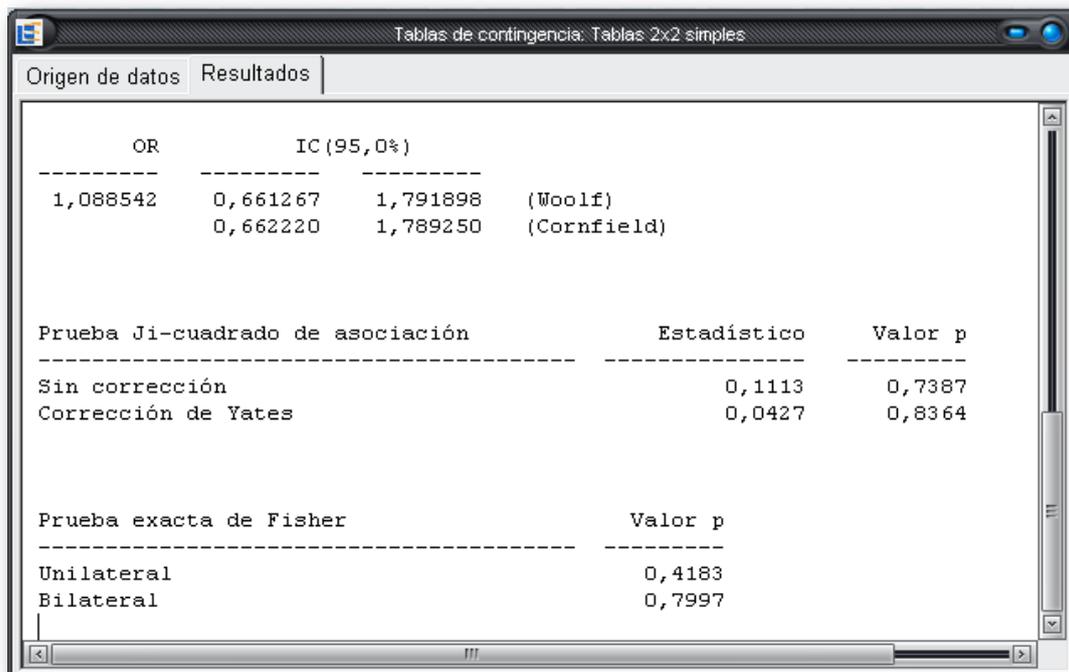
Caso-control emparejado

Nivel de confianza (%)

Sumar 0,5 a todas las frecuencias si hay ceros

Enfermedad

		Enfermos	Sanos	Total
Factor de riesgo	Expuestos	55	80	135
	No expuestos	48	76	124
Total		103	156	259



OR		IC (95,0%)	
1,088542	0,661267	1,791898	(Woolf)
	0,662220	1,789250	(Cornfield)

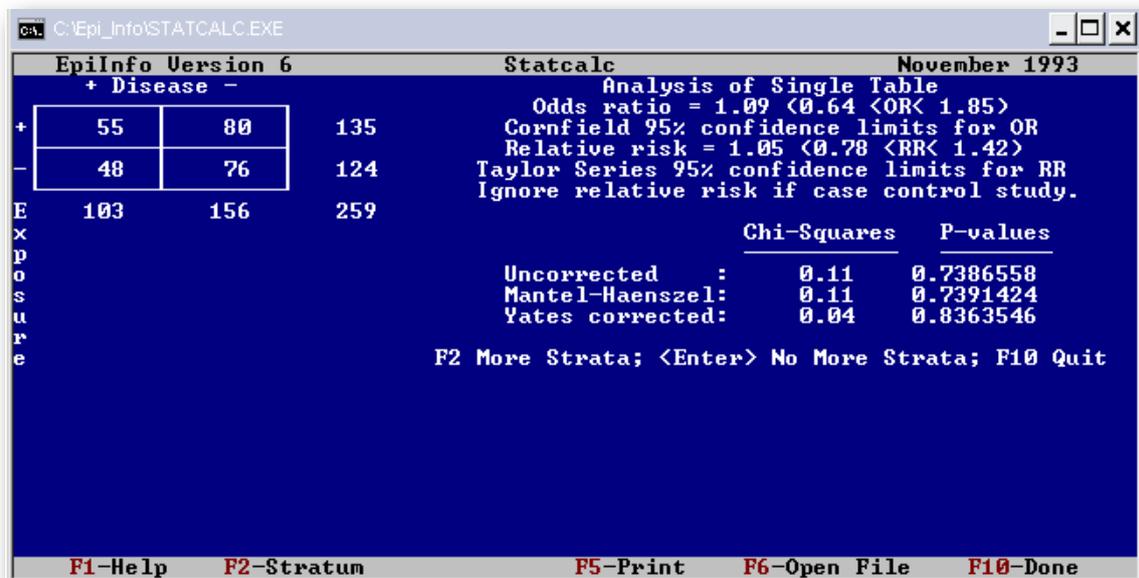
Prueba Ji-cuadrado de asociación	Estadístico	Valor p
Sin corrección	0,1113	0,7387
Corrección de Yates	0,0427	0,8364

Prueba exacta de Fisher	Valor p
Unilateral	0,4183
Bilateral	0,7997

El valor del estadístico chi-cuadrado es muy pequeño (0,113), y la significación estadística (valor de p) es claramente superior al 0,05. Por tanto, las diferencias entre los porcentajes de fumar para hombres y mujeres no son significativas, y en consecuencia, se acepta la hipótesis nula, concluyendo que dichas diferencias se deben al azar, y que no existe asociación entre las variables sexo y fumar, o sea, que son independientes.

Utilizando el programa EPIINFO, se obtendría el mismo valor.





b) Ejercicio de la Tabla 2

En este ejercicio, el tamaño muestral es grande, y no existen frecuencias teóricas inferiores a 5, por tanto, se procede al cálculo del estadístico chi-cuadrado. En primer lugar utilizamos el programa EPIDAT.

The screenshot shows the 'Tablas de contingencia: Tablas MXN' window in EPIDAT. It displays a 3x2 contingency table with the following data:

	1	2
1	5	195
2	31	119
3	47	63

Control settings: N° de filas = 3, N° de columnas = 2, Tipo de datos = Nominales, Porcentajes = Ninguno.

Statistical results:

```

% de celdas con frecuencia esperada <5: 0,0%

Prueba Ji-cuadrado de Pearson

Ji-cuadrado      gl      Valor p
-----
78,6959          2          0,0000
    
```

Ahora, el valor de χ^2 es muy grande, y el valor de p es inferior a 0,05. Las diferencias observadas entre los tres grupos sí difieren significativamente, y se rechaza la hipótesis nula, aceptando la hipótesis alternativa. Se concluye indicando que la prevalencia de hipertensión aumenta con la edad, mostrando diferencias significativas, o sea, existe asociación entre las variables de estudio.

Vemos qué ocurre a continuación con el programa EPIINFO:

The screenshot shows the 'EpiInfo Version 6' window with the 'Statcalc' menu open. It displays a 3x2 contingency table and its statistical analysis results:

+ Disease -		
5	195	200
31	119	150
47	63	110
83	377	460

Statistical results:

```

Chi square = 78.70
2 degrees of freedom.
p value = 0.00000000 <----

<Enter> for Another Table; F10 to Quit
    
```

Footer: F1-Help, F5-Print, F6-Open File, F10-Done

5. Referencias bibliográficas.

- Carrasco J.L. El método estadístico en la investigación médica. Ed. Ciencia 3. Madrid. 1989
- Pita Fernández, S., Pértega Díaz, S. Asociación de variables cualitativas: test de Chi-cuadrado. Fisterra. www.fisterra.com
- Polit D.F., Hungler B.P. Investigación científica en ciencias de la salud. Ed. McGraw-Hill Interamericana. México. 1997
- Juez Martel, P. Herramientas estadísticas para la investigación en Medicina y Economía de la salud. Ed. Centro de estudios Ramón Areces. S.A. Madrid. 2000
- Mora, M^a A. Estadística para Enfermería. Ed. Pirámide. Madrid. 1984

Agenda Científica

IV Jornadas Nacionales de Enfermería en Cuidados Paliativos

Valencia. 10 y 11 de marzo de 2011

Organiza: Asociación Española de Enfermería en Cuidados Paliativos

<http://www.4enfermeriacuidadospaliativos.com/>

XXIII Congreso Neumosur Enfermería y Fisioterapia

Maspalomas, Gran Canaria. 24-26 de marzo de 2011

Organiza: Asociación de Enfermería en Neumología y Cirugía torácica del Sur

<http://www.neumosurenfermeria.org/AAA/ne11w/index.htm>

VII Congreso Internacional de Nutrición, Alimentación y Dietética

XV Jornadas Nacionales de Nutrición Práctica
Madrid. 30 de marzo - 1 de abril de 2011

Organizan:

- Sociedad Española de Dietética y Ciencias de la Alimentación
- Fundación Alimentación Saludable
- SPRIM Health Marketing & Communication

<http://www.nutricionpractica.org/>

XV Congreso Nacional de la Sociedad Española de Enfermería Radiológica

Castellón. 1-3 de junio de 2011

Organiza: Sociedad Española de Enfermería Radiológica (SEER)

<http://radiologia2011.coecs.com/>

XXXVII Congreso Nacional de la SEEIUC

Bilbao. 12-15 de junio de 2011

Organiza: Sociedad Española de Enfermería Intensiva y Unidades Coronarias (SEEIUC)

<http://www.seeiuc.com/>