Asociación de variables cualitativas: El test exacto de Fisher y el test de McNemar

Pértega Díaz, S.[1]; Pita Fernández, S.[2]

1. Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario Universitario Juan Canalejo. 2. Centro de Salud de Cambre (A Coruña)

CAD. ATEN. PRIMARIA 2004; 11: 304-308

Desde que Pearson[1,2] introdujo el test de la χ^2 en 1900, ésta se ha convertido en una herramienta de uso general para conocer si existe o no relación entre variables de tipo cualitativo. Sin embargo, su aplicación exige de ciertos requerimientos acerca del tamaño muestral que no siempre son tenidos en cuenta[3]. La prueba χ^2 es aplicable a los datos de una tabla de contingencia solamente si las frecuencias esperadas son suficientemente grandes. Del mismo modo, cuando los datos exhiben algún grado de dependencia, el test χ^2 no será el método apropiado para contrastar la hipótesis nula de independencia. En este trabajo se introducirán la prueba exacta de Fisher y el test de McNemar como alternativa estadística al test χ^2 cuando no se verifiquen las condiciones necesarias para su utilización[4-7].

LA PRUEBA DE LA PROBABILIDAD EXACTA DE FISHER

El test exacto de Fisher permite analizar si dos variables dicotómicas están asociadas cuando la muestra a estudiar es demasiado pequeña y no se cumplen las condiciones necesarias para que la aplicación del test χ^2 sea adecuada. Estas condiciones exigen que los valores esperados de al menos el 80% de las celdas en una tabla de contingencia sean mayores de 5. Así, en una tabla 2x2 será necesario que todas las celdas verifiquen esta condición, si bien en la práctica suele permitirse que una de ellas muestre frecuencias esperadas ligeramente por debajo de este valor.

En situaciones como esta, una forma de plantear los resultados es su disposición en una tabla de contingencia de dos vías. Si las dos variables que se están considerando son dicotómicas, nos encontraremos con el caso de una tabla 2 x 2 como la que se muestra en la Tabla 1. El test exacto de Fisher se basa en evaluar la probabilidad asociada a cada una de las tablas 2 x 2 que se pueden formar manteniendo los mismos totales de filas y columnas que los de la tabla observada. Cada una de estas probabilidades se obtiene bajo la hipótesis nula de independencia de las dos variables que se están considerando.

La probabilidad exacta de observar un conjunto concreto de frecuencias a, b, c y d en una tabla 2 x 2 cuando se asume independencia y los totales de filas y columnas se consideran fijos viene dada por la distribución hipergeométrica:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} \qquad (1)$$

TABLA 1

Tabla de contingencia general para la comparación de dos variables dicotómicas en el caso de grupos independientes.

	Característica A		
Característica B	Presente	Ausente	Total
Presente	a	b	a+b
Ausente	c	d	c+d
Total	a+c	b+d	n

Esta fórmula se obtiene calculando todas las posibles formas en las que podemos disponer n sujetos en una tabla 2 x 2 de modo que los totales de filas y columnas sean siempre los mismos, (a+b), (c+d), (a+c) y (b+d).

TABLA 2

Tabla de contingencia para estudiar las diferencias en la prevalencia de obesidad entre sexos. Estudio de prevalencia sobre 14 sujetos.

	Obesidad		
	Sí	No	Total
Mujeres	1 (a)	4 (b)	5 (a+b)
Hombres	7 (c)	2 (d)	9 (c+d)
Total	8 (a+c)	6 (b+d)	14 (n)

La probabilidad anterior deberá calcularse para todas las tablas de contingencia que puedan formarse con los mismos totales marginales que la tabla observada. Posteriormente, estas probabilidades se usan para calcular valor de la p asociado al test exacto de Fisher. Este valor de p indicará la probabilidad de obtener una diferencia entre los grupos mayor o igual a la observada, bajo la hipótesis nula de independencia. Si esta probabilidad es pequeña ($p<0.05$) se deberá rechazar la hipótesis de partida y deberemos asumir que las dos variables no son independientes, sino que están asociadas. En caso contrario, se dirá que no existe evidencia estadística de asociación entre ambas variables.

TABLA 3

Posibles combinaciones de frecuencias con los mismos totales marginales de filas y columnas que en la Tabla 2.

		Obesidad Si	Obesidad No	
(i)	Mujeres	0	5	5
	Hombres	8	1	9
		8	6	14
(ii)	Mujeres	1	4	5
	Hombres	7	2	9
		8	6	14
(iii)	Mujeres	2	3	5
	Hombres	6	3	9
		8	6	14
(iv)	Mujeres	3	2	5
	Hombres	5	4	9
		8	6	14
(v)	Mujeres	4	1	5
	Hombres	4	5	9
		8	6	14
(vi)	Mujeres	5	0	5
	Hombres	3	6	9
		8	6	14

En la literatura estadística, suelen proponerse dos métodos para el cómputo del valor de la p asociado al test exacto de Fisher. En primer lugar, podremos calcularlo sumando las probabilidades de aquellas tablas con una probabilidad asociada menor o igual a la correspondiente a los datos observados. La otra posibilidad consiste en sumar las probabilidades asociadas a resultados al menos tan favorables a la hipótesis alternativa como los datos reales. Este cálculo proporcionaría el valor de p correspondiente al test en el caso de un planteamiento unilateral. Duplicando este valor se obtendría el p-valor correspondiente a un test bilateral.

Para ilustrar la explicación anterior, supongamos que en una determinada población se desea averiguar si existen diferencias en la prevalencia de obesidad entre hombres y mujeres o si, por el contrario, el porcentaje de obesos no varía entre sexos. Tras ser observada una muestra de 14 sujetos se obtuvieron los resultados que se muestran en la Tabla 2. En esta tabla a=1, b=4, c=7 y d=2. Los totales marginales son así a+b=5, c+d= 9, a+c=8 y b+d=6. La frecuencia esperada en tres de las cuatro celdas es menor de 5, por lo que no resulta adecuado aplicar el test χ^2, aunque sí el test exacto de Fisher. Si las variables sexo y obesidad fuesen independientes, la probabilidad asociada a los datos que han sido observados vendría dada por:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} = \frac{5!9!8!6!}{14!1!4!7!2!} = 0{,}0599$$

La Tabla 3 muestra todas las posibles combinaciones de frecuencias que se podrían obtener con los mismos totales marginales que en la Tabla 2. Para cada una de estas tablas, se ha calculado la probabilidad exacta de ocurrencia bajo la hipótesis nula, según la expresión (1). Los resultados obtenidos se muestran en la Tabla 4. El valor de la p asociado al test exacto de Fisher puede entonces calcularse sumando las probabilidades de las tablas que resultan ser menores o iguales a la probabilidad de la tabla que ha sido observada:

$$p = 0{,}0030 + 0{,}0599 + 0{,}0280 = 0{,}0909$$

Otro modo de calcular el valor de p correspondiente consistiría en sumar las probabilidades asociadas a aquellas tablas que fuesen más favorables a la hipótesis alternativa que los datos observados. Es decir, aquellas situaciones en las que la diferencia en la prevalencia de obesidad entre hombres y mujeres fuese mayor que la observada en la realidad. En el ejemplo, sólo existe una tabla más extrema que la correspondiente a los datos observados (aquella en la que no se observa ninguna mujer obesa), de forma que:

$$p = 0{,}0030 + 0{,}0599 = 0{,}0629 \qquad (2)$$

Este sería el valor de la p correspondiente a un planteamiento unilateral. En este caso la hipótesis a contrastar sería que la prevalencia de obesidad es igual en hombres y mujeres, frente a la alternativa de que fuese mayor en los varones. Cuando el planteamiento se hace con una perspectiva bilateral, la hipótesis alternativa consiste en asumir que existen diferencias en la prevalencia de obesidad entre sexos, pero sin especificar de antemano en qué sentido se producen dichas diferencias. Para obtener el valor de la p correspondiente a la alternativa bilateral deberíamos multiplicar el valor obtenido en (2) por dos:

TABLA 4

Probabilidad exacta asociada con cada una de las disposiciones de frecuencias de la Tabla 3.

	a	b	c	d	p
(i)	0	5	8	1	0,0030
(ii)	1	4	7	2	0,0599
(iii)	2	3	6	3	0,2797
(iv)	3	2	5	4	0,4196
(v)	4	1	4	5	0,2098
(vi)	5	0	3	6	0,0280

TABLA 5

Frecuencia de cada una de las posibles combinaciones en un estudio de datos pareados

	Observación 1	Observación 2	Número de pares
Tipo	Característica	Característica	
1	Presente	Presente	a
2	Presente	Ausente	b
3	Ausente	Presente	c
4	Ausente	Ausente	d
	Total		n

$$p = 2 \times 0{,}0629 = 0{,}1258$$

Como se puede observar, las dos formas de cálculo propuestas no tienen por qué proporcionar necesariamente los mismos resultados. El primer método siempre resultará en un valor de p menor o igual al del segundo método. Si recurrimos a un programa estadístico como el SPSS para el cómputo del test, éste utilizará la primera vía para obtener el p-valor correspondiente a la alternativa bilateral y el segundo método de cálculo para el valor de p asociado a un planteamiento unilateral. En cualquier caso, y a la vista de los resultados, no existe evidencia estadística de asociación entre el sexo y el hecho de ser obeso en la población de estudio.

TABLA 6

Datos de 20 pacientes intervenidos quirúrgicamente en los que se valoró el dolor tras la cirugía y al cabo de 1 hora tras la administración de un analgésico.

Individuo	Dolor tras la intervención	Dolor 1 hora después del tratamiento
1	No	No
2	Sí	No
3	No	No
4	No	No
5	Sí	No
6	Sí	No
7	No	No
8	Sí	Sí
9	No	Sí
10	No	No
11	Sí	No
12	Sí	No
13	Sí	No
14	Sí	No
15	Sí	No
16	No	Sí
17	No	Sí
18	Sí	No
19	Sí	No
20	Sí	No

EL TEST DE MCNEMAR

En otras ocasiones, una misma característica se mide en más de una ocasión para cada uno de los individuos que se incluyen en una investigación. En estos casos, el interés se centra en comparar si las mediciones efectuadas en dos momentos diferentes (normalmente antes y después de alguna intervención) son iguales o si, por el contrario, se produce algún cambio significativo. Por ejemplo, puede interesarnos estudiar, a distintos tiempos, el porcentaje de sujetos que se mantienen con fiebre tras la aplicación de un antitérmico o comparar la proporción de enfermos con un determinado síntoma antes y después de un tratamiento.

Para el caso de datos pareados, existen claramente cuatro tipos de pares de observaciones, según cada individuo presente o no la característica de interés en los dos momentos en los que se efectúa la evaluación (Tabla 5). Así, los resultados obtenidos pueden mostrarse igualmente en una tabla 2 x 2 como en la Tabla 1, con la salvedad de que aquí los datos son dependientes y por lo tanto no resultará adecuada la utilización del test χ^2.

Con esta notación, las proporciones de individuos con la característica de interés en los dos momentos en los que se efectúa la medición son:

$$p_1 = \frac{a+b}{n} \quad \text{y} \quad p_2 = \frac{a+c}{n}$$

respectivamente. Estamos interesados por lo tanto en la diferencia entre estas dos proporciones:

$$d = p_1 - p_2 = \frac{a+b}{n} - \frac{a+c}{n} = \frac{b-c}{n}$$

La hipótesis nula que se quiere contrastar es que el valor esperado para esta diferencia es cero, frente a la hipótesis alternativa de que las dos proporciones p_1 y p_2 sean efectivamente diferentes. Esto se puede contrastar centrando nuestra atención en las casillas b y c que son las que muestran discordancia en los dos momentos en los que se efectuó la medición. La prueba de McNemar contrasta así si el número de individuos que han dejado de presentar la

TABLA 7

Tabla de contingencia con los datos de 20 pacientes intervenidos quirúrgicamente en los que se valoró el dolor tras la cirugía y al cabo de 1 hora tras la administración de un analgésico.

		Dolor 1 hora después del tratamiento		
		Sí	No	Total
Dolor	Si	1 (a)	11 (b)	12 (a+b)
tras la	No	2 (c)	6 (d)	8 (c+d)
intervención	Total	3 (a+c)	17 (b+d)	20 (n)

característica de interés (b) es el mismo que el número de individuos que han realizado el cambio inverso (c).
El error estándar para la diferencia entre dos proporciones viene dado por:

$$SE(p_1 - p_2) = \frac{1}{n}\sqrt{b + c - \frac{(b-c)^2}{n}} \qquad (3)$$

De modo que, bajo la hipótesis nula de que no existe diferencia entre ambas *(b-c=0)*, la ecuación (3) se reduce a:

$$SE(p_1 - p_2) = \frac{1}{n}\sqrt{b+c}$$

El estadístico de contraste se construye así de la forma siguiente:

$$z = \frac{p_1 - p_2}{SE(p_1 - p_2)} = \frac{\frac{b-c}{n}}{\frac{1}{n}\sqrt{b+c}} = \frac{b-c}{\sqrt{b+c}} \qquad (4)$$

que sigue una distribución normal *N(0,1)*.
Alternativamente, se puede considerar el estadístico de contraste:

$$\chi^2 = \frac{(b-c)^2}{b+c}$$

que sigue una distribución chi-cuadrado con un grado de libertad y proporciona el mismo valor de la p asociado.
A su vez, se puede aplicar una corrección de continuidad para trabajar sobre muestras pequeñas:

$$z = \frac{|b-c| - 1}{\sqrt{b+c}}$$

refiriendo el valor de dicho estadístico al de una distribución normal *N(0,1)* ó, equivalentemente, a una distribución chi-cuadrado con un grado de libertad si se trabaja con su valor al cuadrado:

$$\chi^2 = \frac{(|b-c| - 1)^2}{b+c}$$

De modo análogo, es posible obtener un intervalo de confianza para la diferencia de proporciones como:

$$p_1 - p_2 \pm 1.96\, SE(p_1 - p_2)$$

Para ilustrar los cálculos anteriores, se dispone de información acerca de 20 pacientes a los que se les administró un determinado tratamiento para tratar el dolor tras una intervención quirúrgica. En cada individuo, se realizó una valoración del dolor inmediatamente después de la operación y al cabo de 1 hora tras la administración del analgésico. Los datos observados se muestran en la Tabla 6. En primer lugar se construye la tabla 2 x 2 con las frecuencias observadas en el estudio (Tabla 7). Según estos datos, el porcentaje de pacientes que manifiestan dolor inicialmente es de 12/20=60%, frente al *3/20=15%* de los enfermos que dicen tener dolor una vez administrado el analgésico. El estadístico de contraste se construye según la expresión (4) como:

$$z = \frac{b-c}{\sqrt{b+c}} = \frac{11-2}{\sqrt{11+2}} = \frac{9}{\sqrt{13}} = 2.49$$

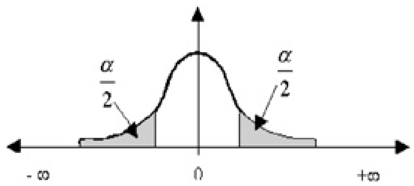
El valor obtenido del estadístico *(z=2.49)* se compara con los valores de una distribución normal estándar (Tabla 8). El valor crítico correspondiente para α=0.01 es de z=2.576 y para α=0.02 es de 2.326. Como quiera que en el cálculo del test de McNemar en el ejemplo obtuvimos un valor de 2.49, que supera al valor para α=0.02, podremos concluir que las dos variables no son independientes, sino que están asociadas *(p<0.02)*. Aplicando la corrección de continuidad proporciona un resultado de:

$$z = \frac{|b-c| - 1}{\sqrt{b+c}} = \frac{|11-2| - 1}{\sqrt{11+2}} = \frac{8}{\sqrt{13}} = 2{,}21,$$

que sigue siendo un resultado significativo *(p<0.03)*.
Otro modo de obtener esta misma información es mediante el cálculo de intervalos de confianza para la diferencia de proporciones en los dos momentos de observación. A

TABLA 8

Tabla de valores de la distribución normal. La tabla muestra los valores de z para los que la probabilidad de observar un valor mayor o igual (en valor absoluto) es igual a α. La cifra entera y el primer decimal de α se buscan en la primera columna, y la segunda cifra decimal en la cabecera de la tabla.



α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	∞	2.576	2.326	2.170	2.054	1.960	1.881	1.812	1.751	1.695
0.1	1.645	1.598	1.555	1.514	1.476	1.440	1.405	1.372	1.341	1.311
0.2	1.282	1.254	1.227	1.200	1.175	1.150	1.126	1.103	1.080	1.058
0.3	1.036	1.015	0.994	0.974	0.954	0.935	0.915	0.896	0.878	0.860
0.4	0.842	0.824	0.806	0.789	0.772	0.755	0.739	0.722	0.706	0.690
0.5	0.674	0.659	0.643	0.628	0.613	0.598	0.583	0.568	0.553	0.539
0.6	0.524	0.510	0.496	0.482	0.468	0.454	0.440	0.426	0.412	0.399
0.7	0.385	0.372	0.358	0.345	0.332	0.319	0.305	0.292	0.279	0.266
0.8	0.253	0.240	0.228	0.215	0.202	0.189	0.176	0.164	0.151	0.138
0.9	0.126	0.113	0.100	0.088	0.075	0.063	0.050	0.038	0.025	0.013

mayores, el intervalo de confianza constituye una medida de la incertidumbre con la que se estima esa diferencia a partir de la muestra, permitiendo valorar tanto la significación estadística como la magnitud clínica de esa diferencia. En el caso que nos ocupa, el intervalo de confianza vendrá dado como:

$$\frac{12}{20}-\frac{3}{20}\pm 1.96\,SE(p_1-p_2)\Rightarrow 0.6-0.15\pm 1.96\times\frac{1}{20}\sqrt{11+2}\Rightarrow 0.45\pm 1.96\times 0.18$$

$$0.45\pm 0.3528\Rightarrow(0.0972;0.8028)$$

Es decir, podemos asegurar (con una seguridad del 95%) de que la diferencia real en el porcentaje de pacientes que manifiestan dolor antes y después de recibir el tratamiento analgésico se mueve entre un 9.72% y un 80.28%.

En definitiva, el uso generalizado de la metodología estadística ha contribuido a dotar de un mayor rigor a la investigación clínico-epidemiológica en los últimos años. Sin embargo, también ha hecho que estas técnicas se apliquen en ocasiones de una manera un tanto superficial. Es extremadamente importante tener en cuenta las asunciones subyacentes a los distintos métodos estadísticos, como en el caso del test χ^2, para comprender cuándo es adecuado o no su uso y disponer de las técnicas estadísticas alternativas que deben utilizarse en cada ocasión.

BIBLIOGRAFÍA

1. Pearson, K. On a criterion that a given system of deviations fromthe probable in the case of a correlated system of variables is Duch thatit can be reasonably supposed to have arisen from random sampling. Philos. Mag.1900, Series 5, 50: 157-175
2. Pearson, K. On the χ^2 test of goodness of fit. Biometrika 1922; 14: 186-191
3. Pita Fernández S, Pértega Díaz S. Asociación de variables cualitativas: Test χ^2. Cad Aten Primaria 2004; 11: 236-239.
4. Altman DG. Practical statistics for medical research. Chapman & Hall, London, 1991
5. Armitage P, Berry G. Estadística para la investigación biomédica. Doyma, Barcelona, 1992
6. Juez Martel P. Herramientas estadísticas para la investigación en Medicina y Economía de la Salud. Ed. Centro de Estudios Ramón Areces, Madrid, 2001
7. Agresti A. Categoriacl Data Análisis. John Wiley & Sons, New York, 1990.