Algunas aplicaciones estadísticas en la esfera bibliotecológica y su tratamiento computarizado V

Emilio Setién Quesada

La prueba de χ2

La prueba de χ^2 es una técnica del tipo de la bondad de ajuste que se emplea para comparar un grupo de frecuencias observadas con uno de frecuencias esperadas a fin de comprobar si existe una diferencia significativa entre el número observado de objetos o respuestas y el número esperado. Se aplica tomando como base la hipótesis de nulidad H_o.

H_o establece que no existen diferencias significativas entre lo observado y lo esperado, y se contrapone a la hipótesis alterna H_1 que sí acepta esas diferencias.

La prueba se efectúa mediante el empleo de la fórmula:

$$\chi^2 = \sum_{ij1}^{k} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

donde:

oij = frecuencia observada en la iésima fila, jésima columna
eij = frecuencia esperada en la iésima fila, jésima columna
k = número de categorías en que se clasifican los datos
o mediante la fórmula

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

donde
Oij = frecuencia observada en la iésima fila, jésima columna
Eij = frecuencia esperada en la iésima fila, jésima columna
k = número de categorías en que se clasifican los datos
c = número de grupos (muestras) o columnas

Si el acuerdo entre las frecuencias observadas y esperadas es grande, las diferencias entre $o_i - e_j$ serán pequeñas y, consecuentemente, χ^2 será también pequeña. Sin embargo, si el desacuerdo o divergencia es grande, su valor también será grande.

La distribución depende de una cantidad que se denomina grados de libertad (gl) y estos dependen a, su vez, del número de categorías o filas (k) y de grupos o columnas (c) en que se clasifiquen los datos observados y los calculados o esperados. Los grados de libertad están dados por la multiplicación del número de categorías k y de grupos c a los que se le resta 1, esto es, equivalen a (k - 1)(c -1).

El nivel de significación para decidir si ji cuadrado es lo suficientemente grande, como para rechazar o aceptar el criterio de una estrecha concordancia, se basa en distribuciones que aparecen en tablas de valores críticos de chi cuadrada. Esas tablas permiten identificar valores críticos de chi cuadrado según el número de grados de libertad que corresponden a la operación en cada caso.

Si el valor calculado de es igual o mayor que el valor crítico de chi cuadrado dado en la tabla para un nivel particular de significación y en un grado de libertad particular, se rechaza H_o para el nivel de significación dado, esto es, se rechaza que no haya diferencias.

Nótese que las expresiones χ^2 y chi cuadrado se emplean para designar, con χ^2 los resultados de las operaciones, y con chi cuadrado los valores que ofrecen las tablas para cada nivel de significación en cada grado de libertad.

En síntesis:

$$\chi^2 \geq \text{chi cuadrado}$$

Si $\chi^2 \geq$ chi cuadrado se rechaza la no existencia de diferencias significativas entre o y e y se acepta que sí las hay porque no existe concordancia entre ellas

$$\chi^2 \leq \text{chi cuadrado}$$

Si $\chi^2 \leq$ chi cuadrado se acepta la existencia de concordancia o acuerdo entre o y e y, por tanto, que sus diferencias no son significativas

Nótese que

Si no hay diferencias significativas, entonces hay acuerdo o concordancia y χ^2 es pequeña.

Si existen diferencias significativas, entonces no hay acuerdo o concordancia y χ^2 es grande.

Luego las inferencias de la prueba dependerán de cómo se enuncien las hipótesis:

H_0 No existen diferencias entre o y e

falsa si χ^2 es grande $\geq$ chi

verdadera si χ^2 es pequeña $<$ chi

H_0 No existe concordancia entre o y e

falsa si χ^2 es pequeña $<$ chi

verdadera si χ^2es grande $\geq$ chi

H_o No existen **diferencias** entre o y e
falsa si χ^2 es grande □ chi
verdadera si χ^2 es pequeña < chi
H_o No existe **concordancia** entre o y e
falsa si χ^2 es pequeña < chi
verdadera si χ^2 es grande □ chi

Si en caso de la sección anterior se hubiese aplicado como técnica de bondad de ajuste la prueba estadística de chi cuadrado, tomando como valores esperados (e) los correspondientes a los rangos que representan la secuencia de las fases del fenómeno bibliotecario a nivel mundial y como valores observados (o) aquellas que corresponden al país analizado, se tendría lo siguiente:

o	e
Genética 1	1Genética
Institucional 3	2 Profesional
Profesional 2	3 Institucional
Tecnológica 4	4 Tecnológica
Científica 5	5 Científica

$$\chi^2 = \frac{(1-1)^2}{1} + \frac{(3-2)^2}{2} + \frac{(2-3)^2}{3} + \frac{(4-4)^2}{4} + \frac{(5-5)^2}{5}$$

$\chi^2 = 0 + 0,5 + 0,33 + 0 + 0$
$\chi^2 = 0,83$

H_o No existen diferencias significativas entre lo observado y lo esperado
En las tablas de valores críticos de χ^2, para □ =0,05 con 4 gl, chi cuadrado = 9,49. Entonces, como 0,83 < 9,49, se acepta H_o: no hay diferencias significativas entre ambas secuencias.

Para trabajar χ^2 utilizando la hoja de cálculo de Excel es necesario aplicar dos fórmulas: la PRUEBA.CHI y la PRUEBA.CHI.INV. A continuación se explica cada una de ellas.
Prueba.Chi Devuelve la prueba de independencia mediante el valor de la distribución chi cuadrado para la estadística y los grados de libertad apropiados en forma de probabilidad. Esto es, devuelve la probabilidad de una estadística chi cuadrado con determinados grados de libertad.

Prueba.Chi.Inv. Devuelve para una probabilidad dada, de una sola cola, el valor de la variable aleatoria siguiendo una distribución chi cuadrado.

Función estadística PRUEBA.CHI
Aparece la pantalla 16

Asistente de funciones paso 2 de 2

PRUEBA.CHI **Valor** ____

Devuelve la prueba de independencia

Rango __ actual requerido

es el rango de datos que contiene observaciones
para probar frente a valores esperados

rango actual *fx* ____
rango esperado *fx* ____

Ayuda Cancelar Atrás Siguiente Terminar

Pantalla 17

Asistente de funciones paso 2 de 2

PRUEBA.CHI **Valor** __0,9339___

Devuelve la prueba de independence

Rango __ actual requerido

es el rango de datos que contiene observaciones
para probar frente a valores esperados

rango actual *fx* _{1;3;2;4;5}_
rango esperado *fx* {1;2.3.4.5}___

Ayuda Cancelar Atrás Siguiente Terminar

Función estadística PRUEBA.CHI.INV
Aparece la pantalla 18

Asistente de funciones paso 2 de 2

PRUEBA.CHI.INV **Valor** _____

Devuelve para una probabilidad dada, de una sola
cola, el valor de la variable aleatoria siguiendo una distribución chi cuadrado

Probabilidad (requerido)

es una probabilidad asociada con la distribución chi cuadrado

probabilidad *fx* _______

grados de libertad *fx* ______

Ayuda Cancelar Atrás Siguiente Terminar

Función estadística PRUEBA.CHI.INV
Aparece la pantalla 19

Asistente de funciones paso 2 de 2

PRUEBA.CHI.INV Valor _0,83_

Devuelve para una probabilidad dada, de una sola
cola, el valor de la variable aleatoria siguiendo
una distribución chi cuadrado

Probabilidad (requerido)

es una probabilidad asociada con la distribución
chi cuadrado

probabilidad fx __0,9339_____

grados de libertad fx ___4___

Ayuda Cancelar Atrás Siguiente Terminar

En el otro ejemplo, si se supusiera que las frecuencias observadas de libros leídos por el sexo femenino no fueran diferentes que las del masculino, tomando estas últimas como frecuencias esperadas, se tendría:

	o	e
Niños	13	15
Adultos	7	14
Tercera edad	10	9
Jóvenes	10	7
Adolescentes	9	6

$$\chi^2 = \frac{(13-15)^2}{15}+\frac{(7-14)^2}{14}+\frac{(10-9)^2}{9}+\frac{(10-7)^2}{7}+\frac{(9-6)^2}{6}$$

$$\chi^2 = 6,66$$

χ^2 = 6,66 tiene una probabilidad asociada que está entre □ =0,20 (chi 5,99) y □ =0,10 (chi 7,78), mientras que □ =0,05 (chi 9,49) lo que indica que H_o no puede ser rechazada en □ =0,05, pero si puede serlo en □ =0,20. Parece que se necesitan más datos antes de hacer cualquier conclusión definitiva sobre H_o a partir de los resultados de esta prueba, para que sean consecuentes con los anteriores. Nótese que χ^2 es una prueba no paramétrica mientras que la media, la aproximación a la normal y el coeficiente de correlación son paramétricas.

Pantalla 20

Asistente de funciones paso 2 de 2

PRUEBA.CHI **Valor** __0,1547___

Devuelve la prueba de independencia

Rango __ actual requerido

es el rango de datos que contiene observaciones para probar frente a valores esperados

rango actual *fx* _{13;7;10;10;9}_

rango esperado *fx* __{15;14;9;7;6}___

Ayuda Cancelar Atrás Siguiente Terminar

Pantalla 21

Asistente de funciones paso 2 de 2

PRUEBA.CHI.INV **Valor** _6,66

Devuelve para una probabilidad dada, de una sola
cola, el valor de la variable aleatoria siguiendo una distribución chi cuadrado

Probabilidad (requerido)

es una probabilidad asociada con la distribución chi cuadrado

probabilidad *fx* __0,1547_____

grados de libertad *fx* ___4___

Ayuda Cancelar Atrás Siguiente Terminar

Recapitulación de pruebas anteriores sobre las muestras B y C

La media de B no es segura, la de C si lo es. La distribución de B no se aproxima a la normal, la de C sí . No hay correlación entre los valores de ambas muestras . Todo hace suponer, entonces, que existan diferencias significativas entre lo observado y lo esperado, o sea que hay diferencias entre la lectura del sexo masculino y el femenino