

Chi-cuadrado (χ^2): distribución, cómo se calcula, ejemplos

La prueba **chi cuadrado** o *ji-cuadrado* (χ^2 , donde χ es la letra griega llamada "chi") se utiliza para determinar el comportamiento de cierta variable y también cuando se quiere saber si dos o más variables son independientes estadísticamente.

Para comprobar el comportamiento de una variable, la prueba a realizar se llama *prueba chi cuadrado de ajuste*. Para saber si dos o más variables son independientes estadísticamente la prueba se llama *chi cuadrado de independencia*, también llamada *de contingencia*.

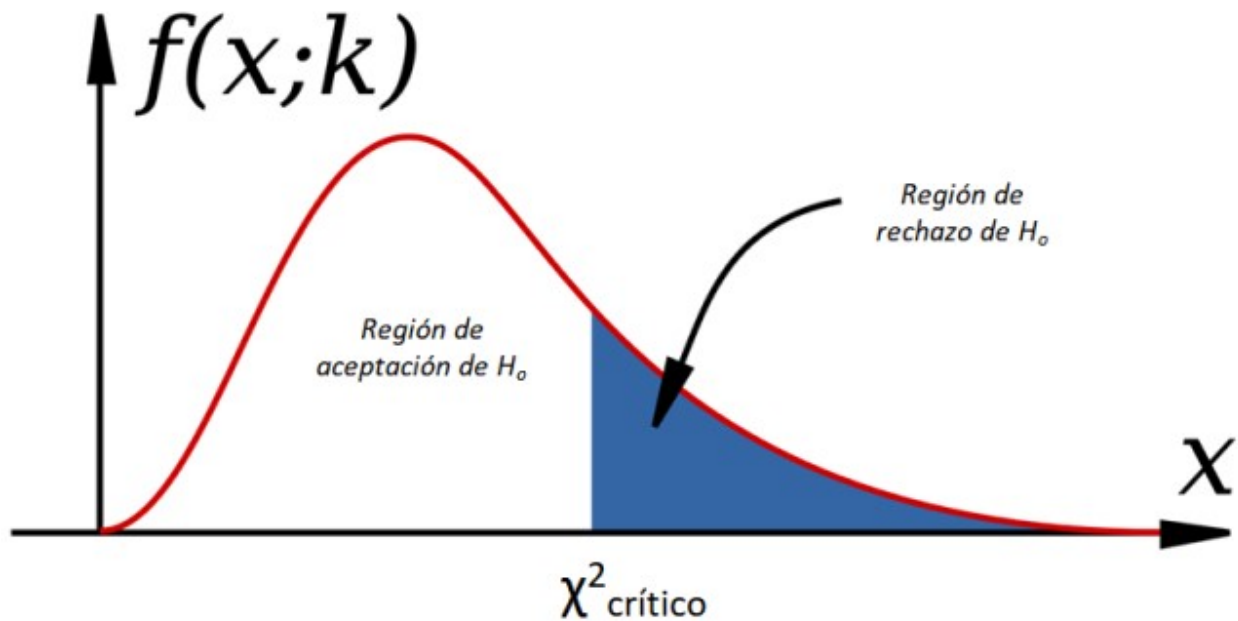


Figura 1. Pruebas de hipótesis mediante chi cuadrado

Estas pruebas forman parte de la teoría estadística de decisiones, en la cual se estudia una población y se toman decisiones acerca de esta, analizando una o varias muestras extraídas de ella. Para ello es preciso hacer ciertas suposiciones con respecto a las variables, llamadas *hipótesis*, las cuales pueden ser o no ciertas.

Hay algunas pruebas para contrastar estas conjeturas y determinar cuáles son válidas, dentro de un cierto margen de confianza, entre ellas la prueba chi-cuadrado, que puede aplicarse para comparar dos y más poblaciones.

Como veremos, suelen plantearse dos tipos de hipótesis acerca de algún parámetro poblacional en dos muestras: la hipótesis nula, llamada H_0 (las muestras son independientes), y la hipótesis alternativa, denotada como H_1 , (las muestras están correlacionadas) que es contraria de aquella.

[toc]

¿Cuándo se usa la prueba chi-cuadrado?

La prueba chi cuadrado se aplica a variables que describen [cualidades](#), como por ejemplo sexo, estado civil, grupo sanguíneo, color de ojos y preferencias de diversos tipos.

La prueba está pensada cuando se desea:

-Comprobar si una distribución es apropiada para describir una variable, lo cual se denomina *bondad del ajuste*. Mediante la prueba de chi cuadrado se puede saber si hay diferencias significativas entre la distribución teórica seleccionada y la distribución de frecuencias observada.

-Conocer si dos variables X e Y son independientes desde el punto de vista estadístico. Esto se conoce como *prueba de independencia*.

Dado que se aplica a variables cualitativas o categóricas, la prueba chi cuadrado se utiliza ampliamente en ciencias sociales, administración y medicina.

Condiciones para aplicarla

Hay dos requerimientos importantes para aplicarla correctamente:

-Los datos deben estar agrupados en frecuencias.

-La muestra tiene que ser lo suficientemente grande para que sea válida la distribución chi cuadrado, de lo contrario su valor se sobreestima y da lugar al rechazo de la hipótesis nula cuando no debiera ser así.

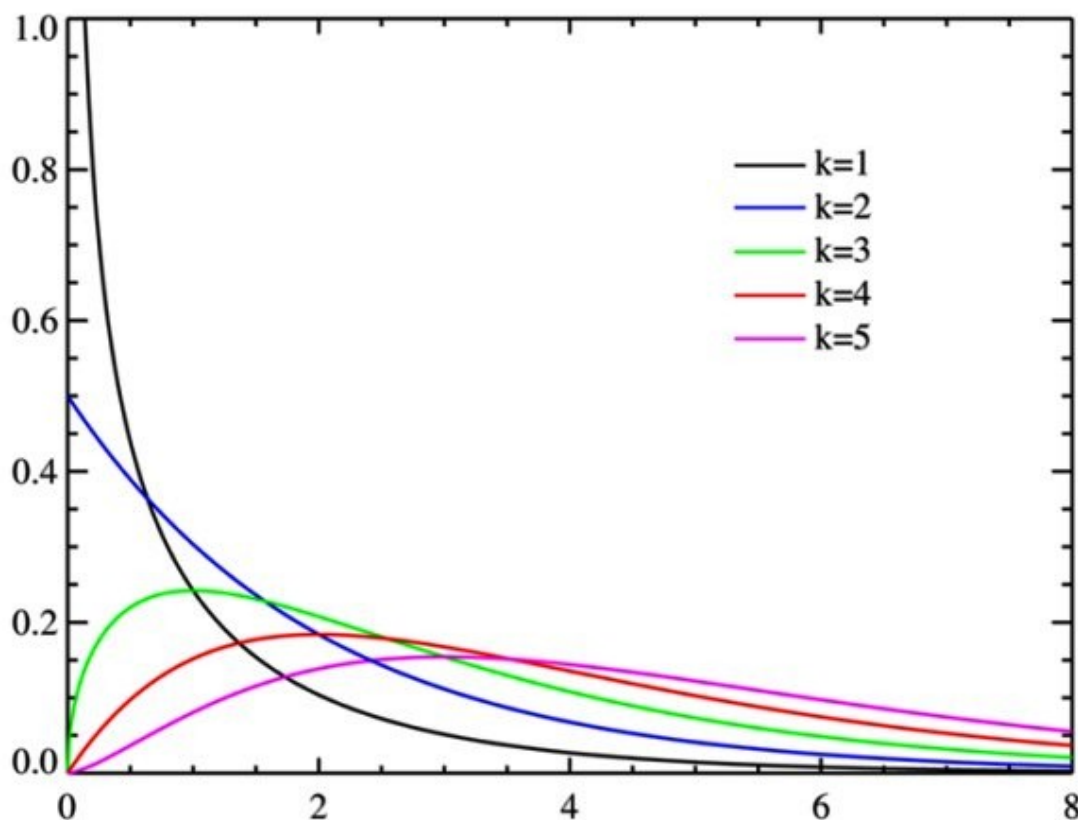
La regla general es que si en los datos agrupados aparece una frecuencia con valor menor a 5, esta no se use. Si hay más de una frecuencia menor a 5, entonces deben combinarse en una para obtener una frecuencia con valor numérico mayor a 5.

Distribución chi cuadrado

χ^2 es una distribución continua de probabilidades. En realidad hay diferentes curvas, dependiendo de un parámetro k llamado *grados de libertad* de la variable aleatoria.

Sus propiedades son:

- El área bajo la curva es igual a 1.
- Los valores de χ^2 son positivos.
- La distribución es asimétrica, es decir, posee sesgo.



Figura

2. Distribución de chi cuadrado para varios grados de libertad. Fuente: Wikimedia Commons.

Grados de libertad

A medida que aumentan los grados de libertad, la distribución chi-cuadrado tiende a la normalidad, como se aprecia de la figura.

Para una distribución dada, los grados de libertad se determinan a través de la *tabla de contingencia*, que es la tabla donde se registran las frecuencias observadas de las variables.

Si una tabla tiene f filas y c columnas, el valor de k es:

$$k = (f - 1) \cdot (c - 1)$$

Formulación de las hipótesis

Cuando la prueba chi cuadrado es de ajuste, se formulan las siguientes hipótesis:

- H_0 : la variable X tiene distribución de probabilidad $f(x)$ con los parámetros específicos y_1, y_2, \dots, y_p

- H_1 : X tiene otra distribución de probabilidad.

La distribución de probabilidad que se supone en la hipótesis nula puede ser, por ejemplo, la conocida distribución normal, y los parámetros serían la media μ y la desviación estándar σ .

Además, la hipótesis nula es evaluada con un cierto nivel de significación, esto es, una medida del error que se cometería al rechazarla siendo cierta.

Por lo general este nivel se establece de 1%, 5 % o 10 % y cuanto menor sea, más fiable el resultado de la prueba.

Y si se utiliza la prueba chi cuadrado de contingencia, que como hemos dicho sirve para comprobar la independencia entre dos variables X e Y , las hipótesis son:

- H_0 : las variables X e Y son independientes.

- H_1 : X e Y son dependientes.

De nuevo se requiere especificar un nivel de significación para conocer la medida del error al tomar la decisión.

¿Cómo se calcula la estadística chi-cuadrado?

La estadística chi cuadrado se calcula de la siguiente manera:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

El símbolo \sum significa “sumatoria”, la cual debemos realizar sobre la expresión fraccionaria indicada.

La sumatoria se lleva a cabo desde la primera clase $i = 1$ hasta la última, que es $i = k$.

Además:

– f_o es una frecuencia observada (proviene de los datos obtenidos).

– f_e es la frecuencia esperada o teórica (es necesario calcularla a partir de los datos).