



Prueba chi cuadrada, χ^2

Inferencia estadística

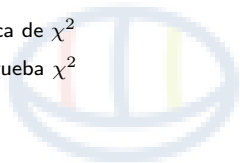
Instructor: Juan Luis Palacios Soto

palacios.s.j.l@gmail.com



scidata
Matemáticas para
la ciencia de datos

- 1 Gráfica de χ^2
- 2 La Prueba χ^2



scidata
matemáticas para
la ciencia de datos

Uso de la prueba χ^2

Tiene varios usos: se puede aplicar para determinar si es probable que una variable venga o no de una distribución dada (pueden ser datos categóricos); decidir si dos variables pueden o no estar relacionadas; decidir si una población tiene cierta varianza, etc.

- Nos ayuda cuando buscamos la dependencia entre características de nuestro fenómeno. Por ejemplo, determine si las ventas de diferentes colores de automóviles dependen de la ciudad donde se venden.
- Es una prueba de hipótesis que compara la distribución observada de los datos con una distribución esperada de los datos. Por ejemplo, puedes comprobar si un dado es justo lanzando el dado muchas veces y utilizando una prueba chi-cuadrada para determinar si los resultados siguen una distribución uniforme.
- En procesos de control se aplica frecuentemente para determinar si la desviación estándar es igual a cierto valor, por ejemplo: $H_0 : \sigma = 2$.

Null

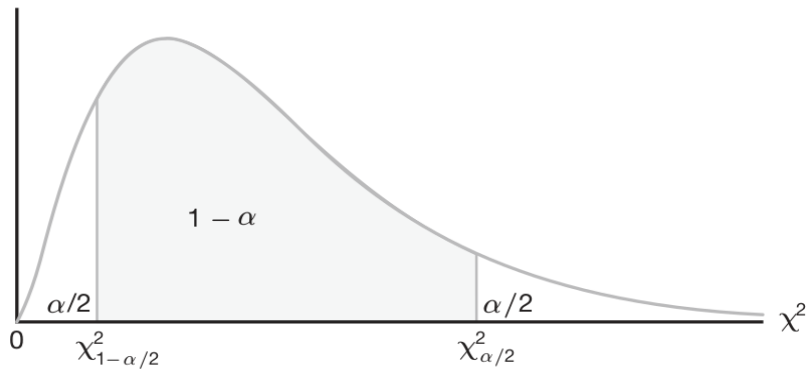
The data is
UNIFORMLY
distributed among
categories

Alternate

The data is
NOT UNIFORMLY
distributed among
categories

Los pasos para la aplicación de la prueba χ^2 son:

- 1 **Seleccionar el nivel de significación.** Este es denotado por α . Generalmente $\alpha = 0.01$, $\alpha = 0.05$ o $\alpha = 0.10$.
- 2 **Encontrar el valor crítico.** Este es denotado por $\chi^2_{n-1,\alpha}$. Antiguamente se usaba una tabla de χ^2 . Se seleccionaba la columna basado en α y la fila basado en los grados de libertad, que en este caso es $n - 1$ (siendo n el tamaño de la muestra).
- 3 **Calcular el valor χ^2 .** Se refiere a calcular el número (i) $\chi^2 = \sum \frac{(o - e)^2}{e}$, con los datos de la muestra (o es el dato observado y e el dato esperado), o bien (ii) $\chi^2 = \frac{(n - 1)s^2}{\sigma_0^2}$, donde σ_0^2 es la varianza hipotetizada bajo H_0 .
- 4 **Comparar y decidir.** Caso(i) Si $\chi^2 \geq \chi^2_{n-1,\alpha}$, rechazamos H_0 . En caso contrario, "aceptamos" H_0 . Caso(ii), tenemos varias opciones que dependen del tipo de prueba, por ejemplo para una prueba de dos colas, rechazamos H_0 si $\chi^2 \geq \chi^2_{n-1,\alpha/2}$ o bien $\chi^2 \leq \chi^2_{n-1,1-\alpha/2}$.



Proceso χ^2 en Excel

No hay una fórmula directa en Excel para hacer pruebas de χ^2 . Sin embargo, se pueden utilizar las funciones previamente vistas en excel, junto con la función INV.CHICUAD.CD que devuelve la función inversa de la probabilidad de una cola de la distribución chi cuadrada. La sintaxis de esta función es INV.CHICUAD.CD(α , n-1). Esto significa que

$$\chi_{n-1, \alpha}^2 = \text{INV.CHICUAD.CD}(\alpha, n - 1).$$



scidata
matemáticas para
la ciencia de datos

Ejemplo:

Los siguientes son los pesos, en decagramos, de 10 paquetes de semillas de pasto distribuidas por cierta empresa: 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2 y 46.0. Pruebe $H_0 : \mu = 46$, pero bajo el supuesto que $\sigma^2 = 0.15$. Suponga una población normal. Qué haría para verificar el supuesto de la varianza poblacional?



scidata
matemáticas para
la ciencia de datos