Distribución Ji-Cuadrada: Teoria y Aplicaciones.

Héctor Badillo Sánchez[a], Fernando Velasco Luna[b], Francisco Solano Tajonar Sanabria[c], Hugo Cruz Suárez[d].
[a,b,c,d] *Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias Físico Matemáticas, Puebla, Puebla, México.*
[a]*hectorbs.2417@gmail.com*, [b]*fvelasco@fcfm.buap.mx*, [c]*ftajonar@fcfm.buap.mx*, [d]*hcs@fcfm.buap.mx*.

Resumen

La distribución Ji-Cuadrada es una familia de distribuciones, donde cada distribución se define por los *grados de libertad*. La distribución Ji-cuadrada puede derivarse de la distribución normal. Un uso de la distribución Ji-cuadrada se da en las pruebas de hipótesis en las que los datos disponibles para el análisis se encuentran en forma de frecuencias. Las pruebas de Ji-cuadrada se pueden considerar como pruebas de bondad de ajuste, ya que prueban la bondad de ajuste de las frecuencias observadas a las frecuencias que se esperaría si los datos fueron generados bajo alguna hipótesis particular. La estadística Ji-cuadrada es más apropiada para usar con variables categóricas, como el nivel socioeconómico, cuyos valores son las categorías bajo, medio y alto. En este trabajo se da una breve introducción a la distribución Ji-Cuadrada y se presentan algunas de las aplicaciones de la estadística de prueba Ji-Cuadrada.

Palabra claves: Frecuencia Observada, Prueba de Hipótesis. Prueba de bondad de ajuste.

Introducción

En ocasiones es de interés conocer si los datos observados en una muestra aleatoria se ajustan a una distribución de probabilidad, hay pruebas de hipótesis que permiten llevar a cabo esta tarea, tales pruebas de hipotesis se denominan pruebas de bondad de ajuste. La distribuciones de probabilidad que se usan más comúnmente son la uniforme, exponencial, normal, poisson,

Las pruebas se requieren con frecuencia en la práctica de investigación y desarrollo en diversas áreas donde existen conjuntos de datos reales que tienen comportamiento aleatorio. Una de las pruebas de bondad de ajuste más común en la práctica es la Ji-Cuadrada.

En ocasiones dado un conjunto de datos se tienen la necesidad de establecer un modelo probabilístico adecuado para tales datos, con el objetivo posterior de estimar probabilidades de eventos relacionados a tales

Los datos de la muestra se examinan para ver si esta distribución es coherente con la distribución hipotética de la población o no. Una forma en que se puede usar la prueba de bondad de Ji-cuadrada es examinar qué tan cerca una muestra coincide con una población.

Un uso de la distribución Ji-cuadrada se da en las pruebas de hipótesis en las que los datos disponibles para el análisis se encuentran en forma de frecuencias. Tales procedimientos de prueba de hipótesis se encuentran en la literatura bajo los nombres:

Distribución Ji-Cuadrada

La distribución Ji-Cuadrada es una familia de distribuciones. Cada distribución se define por los *grados de libertad*. La distribución Ji-cuadrada puede derivarse de la distribución normal. Suponga que a partir de una variable aleatoria distribuida normalmente, se selecciona aleatoria e independientemente muestras de tamaño 1. Cada valor seleccionado puede transformarse en la variable normal estándar mediante la fórmula

$$z_i = \frac{y_i - \mu}{\sigma}$$

Si se seleccionara una muestra aleatoria de tamaño *n* de la distribución normal, y cada una se eleva al cuadrado y se obtiene la suma, es decir,

$$\chi^2_{(n)} = z_1^2 + z_2^2 + \cdots + z_n^2$$

Se tiene que esta nueva variable se distribuye como una Ji-cuadrada con *n* grados de libertad.

Si *U* es una variable aleatoria con distribución Ji-cuadrada con *k* grados de libertad su función de densidad está dada por

$$f(u) = \frac{1}{\left(\frac{k}{2}-1\right)!}\frac{1}{2^{k/2}}u^{(k/2)-1}e^{-(u/2)}, \quad u > 0$$

La grafica de la función de densidad está dada por medio de

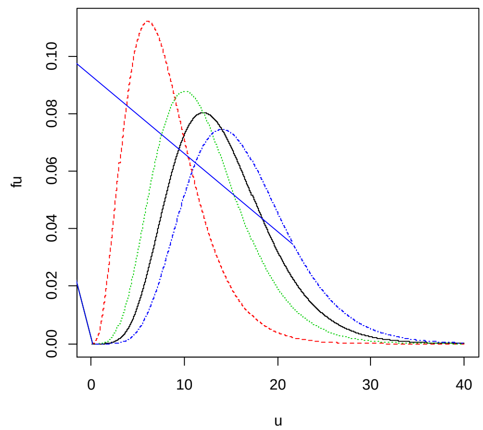

Figura 1.- Grafica de la funcion de densidad para una variable aleatoria Ji-Cuadrada.

Tipos de prueba Ji-Cuadrada

Un uso de la distribución Ji-cuadrada se da en las pruebas de hipótesis en las que los datos disponibles para el análisis se encuentran en forma de frecuencias. Tales procedimientos de prueba de hipótesis se encuentran en la literatura bajo los nombres:

- Pruebas de bondad de ajuste,
- Pruebas de independencia

Las pruebas de Ji-cuadrada se pueden considerar como pruebas de bondad de ajuste, ya que prueban la bondad de ajuste de las frecuencias observadas a las frecuencias que se esperaría si los datos fueron generados bajo alguna hipótesis particular.

Bondad de ajuste se usa para hacer referencia a una comparación de una distribución muestral con alguna distribución teórica que se supone que describe la población de la que proviene la muestra.

La justificación de nuestro uso de la distribución en estas situaciones se debe a Karl Pearson (1), quien demostró que la distribución Ji-cuadrada puede usarse como una prueba de la concordancia entre observación e hipótesis siempre que los datos estén en forma de frecuencias.

Frecuencias observadas versus esperadas

La estadística Ji-cuadrada es más apropiada para usar con variables categóricas, como el nivel socioeconómico, cuyos valores son las categorías bajo, medio y alto.

Los datos cuantitativos son las frecuencias observadas O_i y las frecuencias esperadas E_i. Las frecuencias observadas son el número de individuos en la muestra que caen en las diversas categorías de la variable de interés. Por ejemplo, si tenemos una muestra de 1000 clientes, podemos observar que 200 tiene nivel socioeconómico alto, 300 medio y 500 bajo. Las frecuencias esperadas son el número de individuos en la muestra que esperaríamos observar si alguna hipótesis nula sobre la variable fuera cierta. Por ejemplo, nuestra hipótesis nula podría ser que las tres categorías de nivel socioeconómico están igualmente representadas en la población de la muestra.

La estadística de prueba de Ji-cuadrada

La estadística de prueba para las pruebas de Ji-cuadrada está dada por

$$X^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

Cuando la hipótesis nula es verdadera, X^2 se distribuye aproximadamente como una Ji-Cuadrada con *k-r* grados de libertad. Al determinar los grados de libertad, *k* es igual al número de grupos para los cuales están disponibles las frecuencias observadas y esperadas, y *r* es el número de restricciones o restricciones impuestas en la comparación dada. Se impone una restricción cuando forzamos la suma de las frecuencias esperadas para que sea igual a la suma de las frecuencias observadas, y se impone una restricción adicional para cada parámetro que se estima a partir de la muestra.

La cantidad X^2 es una medida del grado en que, en una situación dada, coinciden los pares de frecuencias observadas y esperadas. Como veremos, la naturaleza de X^2 es tal que cuando existe una estrecha concordancia entre las frecuencias observadas y esperadas, es pequeña, y cuando la concordancia es pobre, es grande. En consecuencia, solo un valor suficientemente grande de X^2 provocará el rechazo de la hipótesis nula.

Si existe una concordancia perfecta entre las frecuencias observadas y las frecuencias esperadas, dado que la hipótesis nula es verdadera, el término O_i-E_i será igual a cero para cada par de frecuencias observadas y esperadas. Tal resultado produciría un valor de X2 igual a cero, y no podríamos rechazar la hipótesis nula. Cuando hay desacuerdo entre las frecuencias observadas y las frecuencias que uno esperaría dado que la hipótesis nula es verdadera, podremos rechazar la hipótesis nula.

La regla de decisión

La cantidad será pequeña si las frecuencias observadas y esperadas están muy juntas y será grande si las diferencias son grandes. El valor calculado de X^2 se compara con el valor tabulado con *k-r* grados de libertad. La regla de decisión, entonces, es: Rechazar la hipótesis nula si X^2 es mayor o igual que el valor tabulado para el valor elegido de alpha.

Pruebas de bondad de ajuste

Una prueba de bondad de ajuste es apropiada cuando se desea decidir si una distribución de frecuencias observada es incompatible con alguna distribución hipotética.

Pruebas de independencia

Por ejemplo, podemos desear determinar si una muestra de valores observados de alguna variable aleatoria es compatible con la hipótesis de que se extrajo de una población de valores que se distribuye normalmente.

Otro uso de la distribución Ji-cuadrada es para probar la hipótesis nula es la de independencia, es decir, de que dos criterios de clasificación, cuando se aplican al mismo conjunto de entidades, son independientes. Decimos que dos criterios de clasificación son independientes si la distribución de un criterio es la misma sin importar cuál sea la distribución del otro criterio.

Referencias

Anderson, D.R., Sweeney, D.J. y Williams, T.A. (2015) **Estadística papa administración y Economía**. 10ª Ed. CENGAGE Learning.

Wayne, D. and Croos Ch. (2013) **Biostatistics: An Foundation for Analysis in the Health Sciences**. 10ª Ed. Wiley.

https://files.sld.cu/prevemi/files/2018/02/Prueba-Ji-cuadrado-de-Bondad-de-ajuste.-Ejemplo.pdf

http://asesorias.cuautitlan2.unam.mx/Laboratoriovirtualdeestadistica/CARPETA%203%20INFERENCIA_ESTADISTICA/DOC_%20INFERENCIA/TEMA%204/08%20PRUEBA%20DE%20CHICUADRADA.pdf