DISTRIBUCIÓN DE PEARSON (1)

Teoría y aplicación en la prueba de bondad de ajuste

Definición

La distribución de Pearson, con símbolo χ^2 (letra griega chi, pronunciada “ji”) es una distribución de probabilidad continua.

Posee un parámetro k que representa los grados de libertad de la variable aleatoria

Se obtiene de la siguiente manera:

$$\chi = z_1^2 + \cdots + z_k^2$$

Función de densidad

$$f(x;k) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} & \text{para } x > 0, \\ 0 & \text{para } x \leq 0 \end{cases}$$

$$f(x;k) = \begin{cases} \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2} & \text{para } x > 0, \\ 0 & \text{para } x \leq 0 \end{cases}$$

Características

- El resultado de la función es siempre positivo por se una suma de cuadrados
- El dominio va de 0 a infinito +
- Su parámetro k son los grados de libertad
- Media = k
- Varianza = 2k
- Moda = k – 2 (Si k es mayor o igual a 2) o 0 (Si k=1)
- Tiene sesgo positivo
- Con k tendiendo a infinito se aproxima a una distribución normal por el teorema del límite central

Usos en inferencia estadística

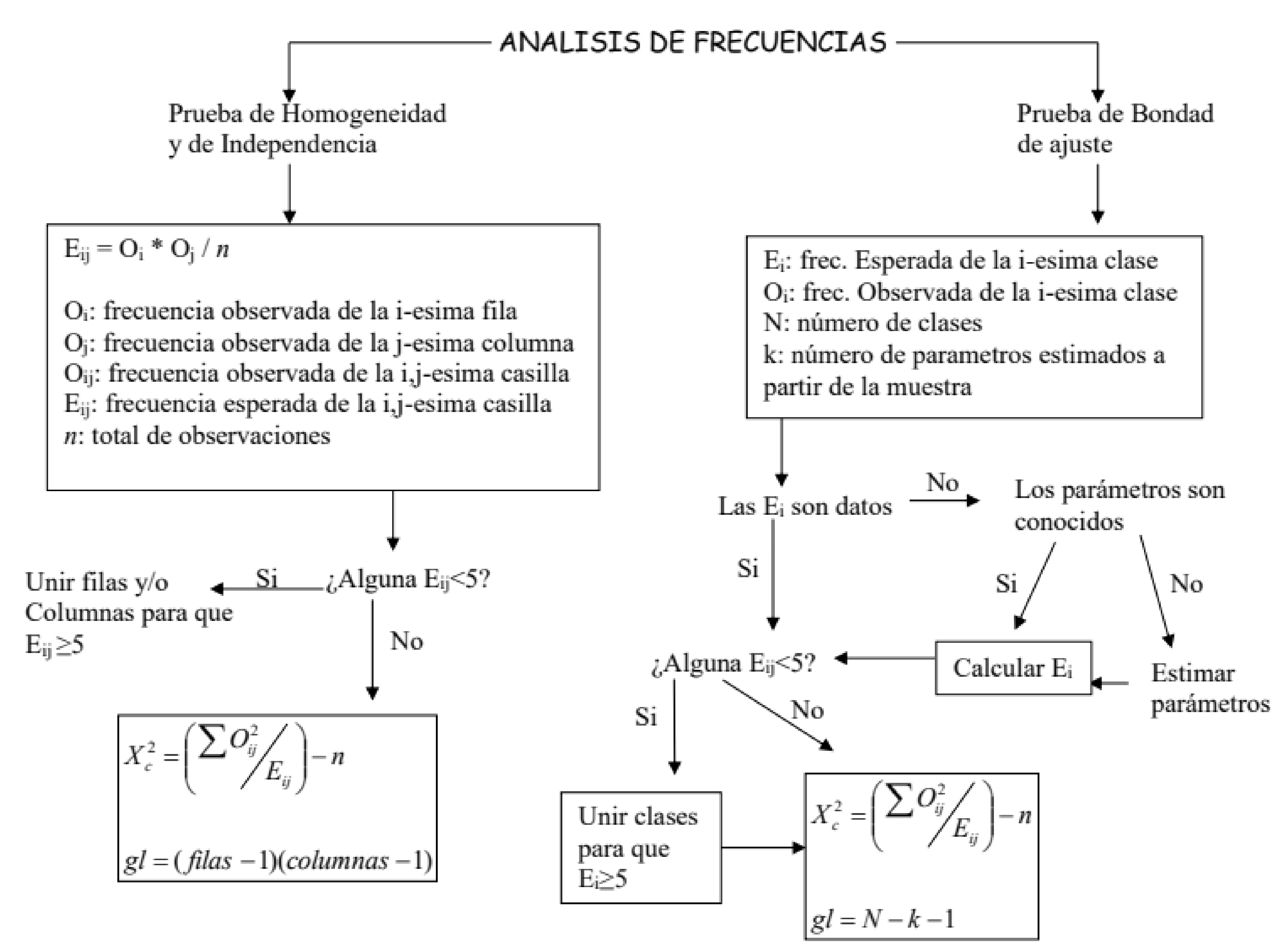
- El uso mas extendido es la prueba de Chi cuadrado, utilizada para prueba de independencia, como bondad de ajuste y en estimación de varianzas.
- También está relacionada con el problema de estimar la media de una población normalmente distribuida.
- Se usa en la estimación de la pendiente de una recta de regresión lineal por su relación con la distribución t de Student.
- Por su relación con la distribución F también puede aparecer en los problemas de análisis de varianza

Bondad de ajuste

- A lo largo del estudio de inferencia se han visto estudios estadísticos sobre hipótesis acerca de parámetros de una población (media, varianza, proporciones).
- La prueba de bondad de ajuste sin embargo se utiliza para determinar si la distribución de una muestra presenta resultados acordes con una distribución específica hipotética.
- Se basa en ver que tan buen ajuste hay entre la frecuencia de las observaciones en la muestra y la frecuencia que se espera que ocurra según la distribución hipotética.

Supuestos para la prueba

- Experimento multinomial
- Tamaño de muestra suficiente (La frecuencia esperada en todas las categorías debe de ser igual o mayor a 5)
- Si no se cumple lo anterior se pueden colapsar categorías


ANALISIS DE FRECUENCIAS
Prueba de Homogeneidad y de Independencia
Prueba de Bondad de ajuste
$E_{ij} = O_i * O_j / n$
O_i: frecuencia observada de la i-esima fila
O_j: frecuencia observada de la j-esima columna
O_{ij}: frecuencia observada de la i,j-esima casilla
E_{ij}: frecuencia esperada de la i,j-esima casilla
n: total de observaciones
¿Alguna E_{ij}<5?
Si
Unir filas y/o Columnas para que $E_{ij} \geq 5$
No
$X_c^2 = \left(\sum O_{ij}^2 / E_{ij} \right) - n$
$gl = (filas - 1)(columnas - 1)$
E_i: frec. Esperada de la i-esima clase
O_i: frec. Observada de la i-esima clase
N: número de clases
k: número de parametros estimados a partir de la muestra
Las E_i son datos
No
Los parámetros son conocidos
Si
Si
No
Calcular E_i
Estimar parámetros
¿Alguna E_{ij}<5?
Si
No
Unir clases para que $E_i \geq 5$
$X_c^2 = \left(\sum O_{ij}^2 / E_{ij} \right) - n$
$gl = N - k - 1$

Ejercicio

- Una moneda fue lanzada al aire 1000 series, de 5 veces cada serie y se observó el número de caras de cada serie. El número de series en los que se presentaron 0, 1, 2, 3, 4 y 5 caras se muestra en la siguiente tabla.
- ¿Los datos se ajustan a una distribución binomial (95% de confianza)?

Número de caras	Número de series (frecuencia observada)
0	38
1	144
2	342
3	287
4	164
5	25
Total	1000