Distribución de Pearson (2)

Teoría y aplicación en las pruebas de homogeneidad e independencia

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

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

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

Características

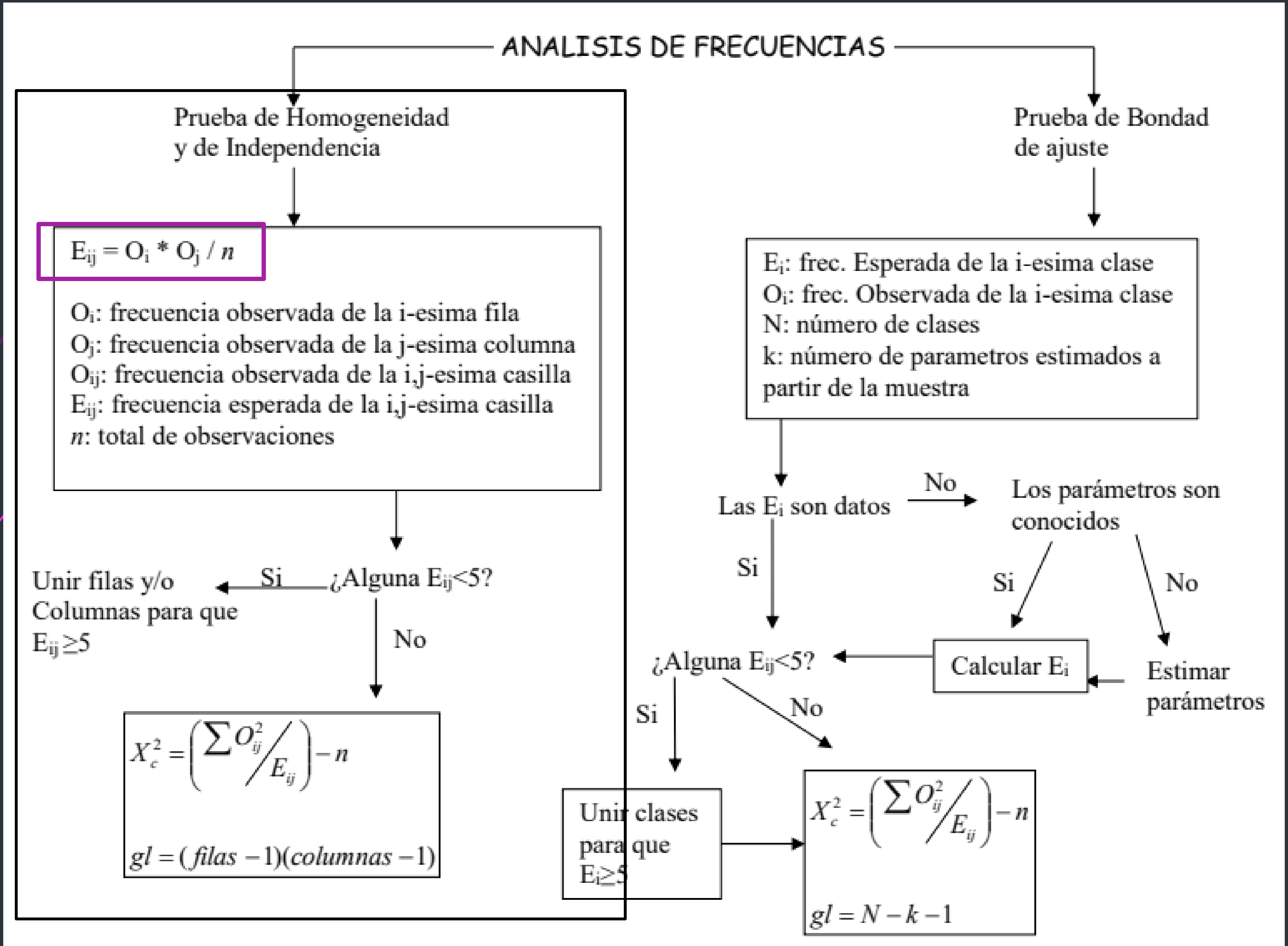
- El resultado de la función es siempre positivo por se una suma de cuadrados
- El dominio va de 0 a infinito +
- Su parámetro k son los grados de libertad
- Media = k
- Varianza = 2k
- Moda = k – 2 (Si k es mayor o igual a 2) o 0 (Si k=1)
- Tiene sesgo positivo
- Con k tendiendo a infinito se aproxima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

Tablas de contingencia
Pruebas de independencia

- En muchas ocasiones, los *n* elementos de una muestra tomada de una población pueden clasificarse con dos criterios diferentes.
- Por lo tanto, es interesante saber si los dos métodos de clasificación son estadísticamente independientes.
- El interés recae en probar la hipótesis de que los dos métodos de clasificación fila-columna son independientes. Si se rechaza esta hipótesis, entonces se concluye que existe alguna interacción entre los dos criterios de clasificación.


ANALISIS DE FRECUENCIAS
Prueba de Homogeneidad y de Independencia
$E_{ij} = O_i * O_j / n$
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
$X_c^2 = \left(\sum O_{ij}^2 / E_{ij} \right) - n$
$gl = (filas - 1)(columnas - 1)$
Prueba de Bondad de ajuste
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 $E_i \geq 5$
$X_c^2 = \left(\sum O_{ij}^2 / E_{ij} \right) - n$
$gl = N - k - 1$

Ejemplo

- Un grupo de profesores universitarios quiere determinar si la satisfacción en el trabajo es independiente del cargo académico. Para ello realizó un estudio nacional entre los académicos universitarios y encontró los resultados mostrados en la tabla siguiente. Con alfa = 0.05, haga una prueba para saber si son dependientes la satisfacción en el trabajo y el cargo.

	Cargo universitario				
Satisfacción en el trabajo		Ayudante	Profesor adjunto	Profesor agregado	Profesor
	Mucha	40	60	52	63
	Regular	78	87	82	88
	Poca	57	63	66	64

	Cargo universitario					
Satisfacción en el trabajo		Ayudante	Profesor adjunto	Profesor agregado	Profesor	**Total**
	Mucha	40 (47,03)	60 (56,44)	52 (53,75)	63 (57,78)	**215**
	Regular	78 (73,28)	87 (87,94)	82 (83,75)	88 (90,03)	**335**
	Poca	57 (54,69)	63 (65,62)	66 (62,50)	64 (67,19)	**250**
	Total	**175**	**210**	**200**	**215**	**800**