



Distribución de Pearson

Teoría y aplicación en las pruebas de asociación

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:

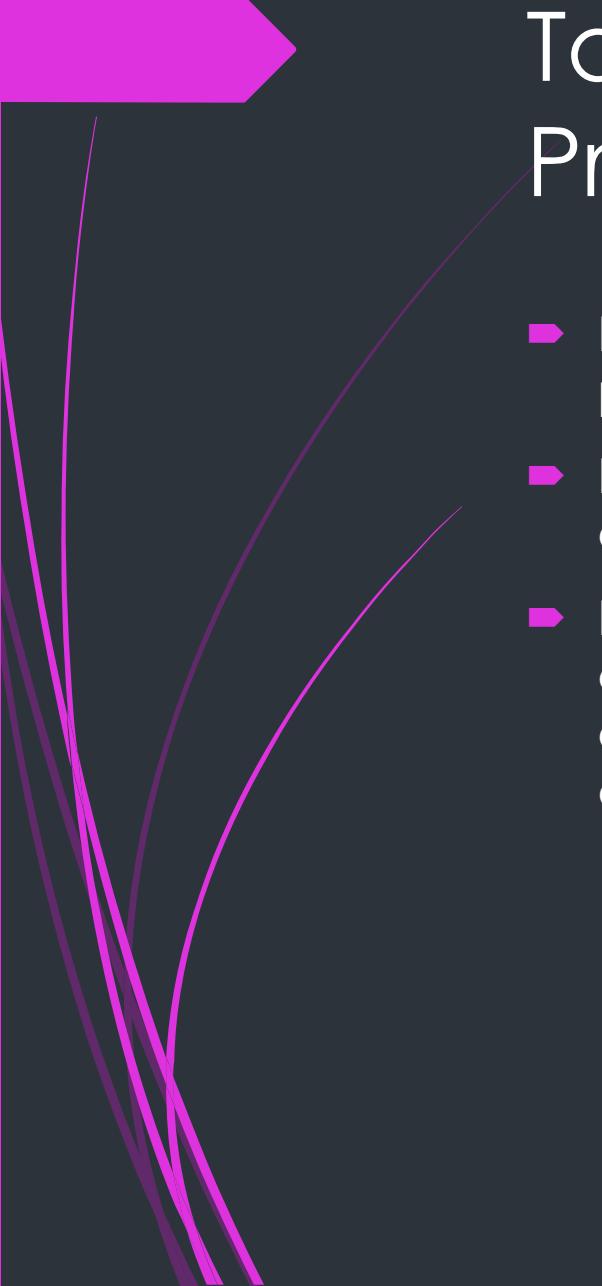
$$\chi = z_1^2 + \cdots + z_k^2$$

Características

- ▶ El resultado de la función es siempre positivo por ser 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

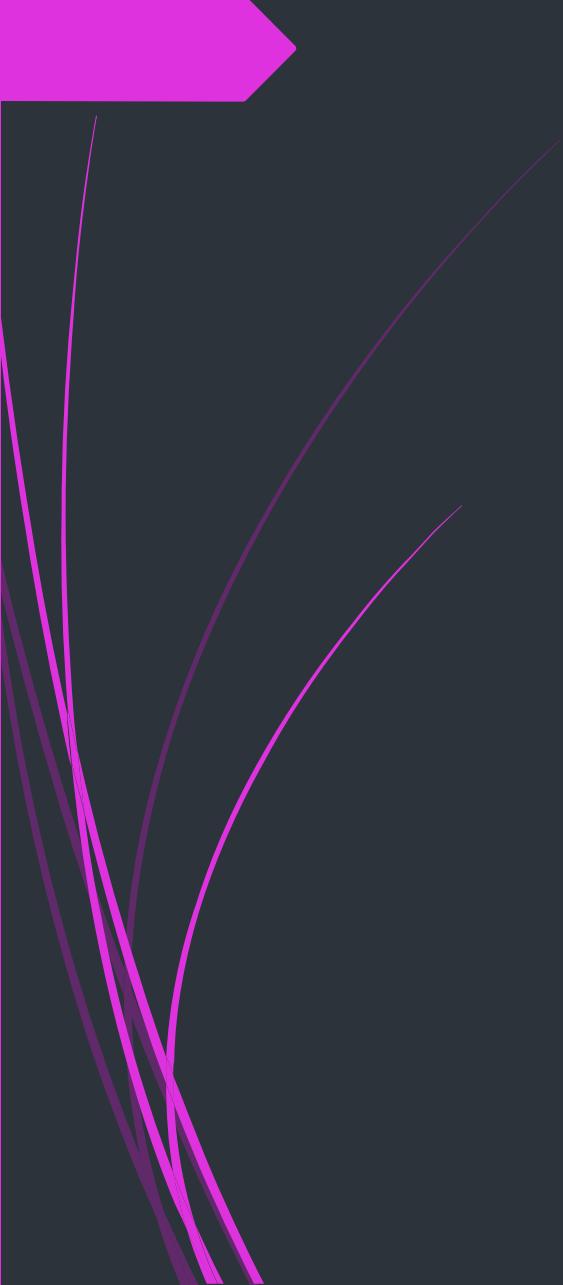
- ▶ El uso mas extendido es la prueba de Chi cuadrado, utilizada para prueba de independencia, como bondad de ajuste y en estimación de varianzas.
- ▶ También está relacionada con el problema de estimar la media de una población normalmente distribuida.
- ▶ Se usa en la estimación de la pendiente de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



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.


$$\chi^2_{calc} = \sum \frac{(f_0 - f_e)^2}{f_e}$$

f_0 : Frecuencia del valor observado.

f_e : Frecuencia del valor esperado.

$$gl = (\text{filas} - 1)(\text{columnas} - 1)$$

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			
		Ayudante	Profesor adjunto	Profesor agregado	Profesor
Satisfacción en el trabajo	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	

Pregunta 1

- En el trabajo titulado Brote por Salmonella enteritidis en trabajadores de un hospital, ¿que prueba realizaría para investigar si hay asociación entre el sexo de los trabajadores y la presencia de enfermedad?
 - a) Test de comparación de medias para muestras pareadas
 - b) Test de comparación de medias para muestras independientes
 - c) Test de chi cuadrado
 - d) Estudio de correlación

Pregunta 2

- ¿Cuál es el valor del estadístico calculado en este caso?
a) 2,46
b) 3,84
c) 6,69
d) 31,4

Cuadro I
CARACTERÍSTICAS DEMOGRÁFICAS
DE CASOS Y CONTROLES. INTOXICACIÓN ALIMENTARIA.
MÉXICO D.F., 1998

Variable	Casos (129)	Controles (150)
Género		
Masculino	39 (30.2 %)	68 (45.3%)
Femenino	90 (69.8 %)	82 (54.7%)
Edad		
Media	32 (\pm 9)	34 (\pm 10)
Coprocultivo		
Positivo	59 (46%)	6 (4%)

La probabilidad de ser caso sintomático y positivo a *Salmoella* fue 14 veces mayor que en los controles ($p < 0.001$, RM 14.2)

Pregunta 3

- Respecto a la respuesta anterior, ¿qué conclusión saca?
 - a) Encontramos evidencia estadística, a un nivel de significación del 5% de que la mayoría de las mujeres enfermaron
 - b) Encontramos evidencia estadística, a un nivel de significación del 5% de que hay asociación entre el sexo y la presencia de enfermedad
 - c) No encontramos evidencia estadística, a un nivel de significación del 5% de que haya asociación entre el sexo y la presencia de enfermedad
 - d) No encontramos evidencia estadística, a un nivel de significación del 5% de que los hombres enfermen más frecuentemente que las mujeres

Pregunta 4

- Si se quiere investigar si existen diferencias significativas en la edad entre casos y controles, ¿cuál sería la hipótesis alternativa con la que trabajarías?
- a) $H_0: \mu_{casos} = \mu_{controles}$
 - b) $H_1: p_{casos} \neq p_{controles}$
 - c) $H_1: \mu_{casos} \neq \mu_{controles}$
 - d) $H_1: \mu_{casos} < \mu_{controles}$
 - e) $H_1: \bar{x}_{casos} \neq \bar{x}_{controles}$

Pregunta 5

- Para aplicar el test planteado en la pregunta anterior, ¿Qué estadístico utilizaría?
 - a) T con 128 grados de libertad
 - b) T con 1 grado de libertad
 - c) Chi cuadrado con 1 grado de libertad
 - d) z

Pregunta 6

- El valor del estadístico calculado fue 1,75, ¿qué concluyes a un nivel de significación del 5%?
 - a) Hay diferencias significativas en la edad promedio entre los sanos y los enfermos
 - b) No se pusieron en evidencia diferencias significativas en la edad promedio entre sanos y enfermos
 - c) Hay evidencia estadística de que la enfermedad fue más frecuente entre los jóvenes

Pregunta 7

- En la sección resultados del estudio “Valoración funcional en pacientes con osteoartrosis de rodilla tratados con glucosamina y un programa de rehabilitación” dice: ‘En el grupo A (glucosamina) se incluyeron 12 pacientes de sexo femenino y 3 masculino, en tanto que el grupo B (placebo) incluyó 13 femeninos y 2 masculinos ($p=0,50$)’ De acuerdo con el apartado estadística, esta afirmación refiere a un test de chi cuadrado. La hipótesis alternativa del test referido puede escribirse:
 - a) La pertenencia a uno de los grupos de estudio (glucosamina/placebo) no depende del sexo
 - b) La proporción de mujeres es similar en los dos grupos de estudio (glucosamina/placebo)
 - c) La proporción de mujeres entre los pacientes tratados con placebo es mayor que entre los tratados
 - d) Hay asociación entre el sexo y la pertenencia a uno de los grupos de estudio

Pregunta 8

- La frecuencia esperada de hombres en el grupo que recibió glucosamina es:
 - a) 2.0
 - b) 2.5
 - c) 3.0
 - d) 3.5
 - e) 5.0

Pregunta 9

- El valor del estadístico calculado en este contraste es:
 - a) 0,24
 - b) 0,45
 - c) 0,48
 - d) 0,50

Pregunta 10

- Según el cuadro IV, para comparar la eficacia de los tres tratamientos, ¿cuál de las hipótesis nulas es la mas adecuada?
- a) Existe asociación entre la erradicación de la parasitosis y el esquema de tratamiento
 - b) No existe asociación entre la erradicación de la parasitosis y el esquema de tratamiento
 - c) El tratamiento con Albendazol es mas efectivo que el tratamiento con Nitazoxanida
 - d) El tratamiento con Albendazol es menos efectivo que el tratamiento con Nitazoxanida

Cuadro IV
EFICACIA COMPARATIVA DE LOS TRES ESQUEMAS DE TRATAMIENTO EN PARASITOSIS INTESTINAL. MÉXICO, 2002

Esquema de tratamiento	Albendazol 400 mg, dosis única n (%)	Nitazoxanida 15 mg/kg/día, por tres días n (%)	Nitazoxanida, dosis única n (%)	p*
Erradicación				
Sí	22 (81.5)	23 (67.6)	22 (71.0)	
No	5 (18.5)	11 (32.4)	9 (29.0)	0.46

*p: se obtuvo con la prueba exacta de Fisher

Pregunta 11

El test a efectuar sería

- a) Test de chi cuadrado
- b) Test de comparación de medias dependientes
- c) Test de comparación de medias independientes
- d) Test de comparación de 2 proporciones

Pregunta 12

La región crítica para esta situación correspondería a los valores

- a) Mayores a 3,84
- b) Menores a 3,84
- c) Mayores a 5,99
- d) Menores a -5,99