Comparación bivariada de variables categóricas, paramétrico

Martínez-Bello, D

Universidad de Santander,Maestria en Biotecnología Bucaramanga, Colombia

5 de diciembre de 2025

Introducción

- La prueba de chi cuadrado se utiliza para analizar tablas de contingencia.
- Permite evaluar si existe asociación entre dos variables categóricas.

Objetivo

- Determinar si existe relación significativa entre dos variables.
- Comparar frecuencias observadas con frecuencias esperadas.

Supuestos

- Las observaciones son independientes.
- Las frecuencias esperadas son mayores a 5.
- Las variables son categóricas.

Hipótesis

- Hipótesis nula: H_0: No hay asociación entre las variables.
- Hipótesis alternativa: H_1: Existe asociación entre las variables.

Tabla de Contingencia

- Se organiza la información en una tabla de doble entrada.
- Cada celda representa la frecuencia de una combinación de categorías.

Frecuencias Esperadas

$$E_{ij} = \frac{(F_{i\cdot})(F_{\cdot j})}{N}$$

Donde:

- $F_{i\cdot}$: Total fila i
- $F_{\cdot j}$: Total columna j
- N: Total general

Estadístico de Prueba

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Donde:

- O_{ij}: Frecuencia observada
- E_{ij}: Frecuencia esperada
- El estadístico sigue una distribución χ^2 con $(r-1)(c-1)$ grados de libertad.
- Se compara el valor calculado con el valor crítico.

Ejemplo en Salud Animal

- Evaluar si existe asociación entre tipo de alimentación e inflamación de la piel en gatos.
- Variables: Alimentación (Natural, Comercial), Parásitos (Sí, No)

Datos del Ejemplo

	Sí	No
Natural	30	20
Comercial	10	40

Cálculo de Frecuencias Esperadas

- Total Natural: 50, Total Comercial: 50
- Total Sí: 40, Total No: 60
- Total general: 100
- Ejemplo: $E_{11} = \frac{50 \times 40}{100} = 20$

Cálculo del Estadístico

$$\chi^2 = \frac{(30-20)^2}{20} + \frac{(20-30)^2}{30} + \frac{(10-20)^2}{20} + \frac{(40-30)^2}{30}$$

$$\chi^2 = 5 + 3{,}33 + 5 + 3{,}33 = 16{,}66$$

Decisión

- Grados de libertad: $(2-1)(2-1) = 1$
- Valor crítico para $\alpha = 0{,}05$: 3.84
- Como $\chi^2 = 16{,}66 > 3{,}84$, se rechaza H_0

Conclusión

- Existe asociación entre tipo de alimentación e inflamación de la piel.

Proporciones

Comparación de dos Proporciones de dos poblaciones para muestras grandes.

- Las hipotesis
- Hipótesis a dos colas:

$$H_0 : \pi_1 = \pi_2; H_A : \pi_1 \neq \pi_2$$

- Hipótesis a una cola:

$$H_0 : \pi_1 \leq \pi_2; H_A : \pi_1 > \pi_2;$$
$$H_0 : \pi_1 \geq \pi_2; H_A : \pi_1 < \pi_2;$$

- El nivel de significancia (0.1, 0.05, 0.01), recordar esto lo fija el investigador.

Proporciones

Comparación de dos Proporciones de dos poblaciones para muestras grandes.

- La regla de decisión a dos colas:

$$H_0 : \pi_1 = \pi_2; \text{ si } |z^*| \leq z(1-\alpha/2), \text{concluir } H_0$$
$$H_A : \pi_1 \neq \pi_2; \text{ si } |z^*| > z(1-\alpha/2), \text{concluir } H_A$$

Proporciones

Comparación de dos proporciones de dos poblaciones para muestras grandes.

- Las proporciones de las muestras de cada población

$$P_1 = \frac{\sum_{i=1}^{n} X_i}{n_1}$$
$$P_2 = \frac{\sum_{i=1}^{n} X_i}{n_2}$$

- El estadístico

$$\frac{P_1 - P_2 - (\pi_1 - \pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}} = z^* \sim \text{Normal}(0, 1)$$

Proporciones

Comparación de dos Proporciones de dos poblaciones para muestras grandes.

- Bajo el supuesto de la hipótesis nula $\pi_1 = \pi_2 = \pi$ se calcula una proporción conjunta π
  Pero el parámetro π normalmente es desconocido y lo estimamos de la muestra y lo denotamos P

$$P = \frac{X_1 + X_1}{n_1 + n_2}$$

- Y el estadístico queda finalmente

$$\frac{P_1 - P_2 - \pi_1 - \pi_2}{\sqrt{P(1-P)(\frac{1}{n_1} + \frac{1}{n_2})}} \sim N(0,1)$$

Proporciones

Ejemplo de Comparación de dos Proporciones de dos poblaciones para muestras grandes.

Tenemos dos grupos de equinos con similares condiciones de manejo, alimentación, raza, genética. Queremos conocer el efecto de un medicamento sobre la presentación o no de diarrea post administración intravenosa del fármaco, comparada con la administración oral del fármaco. El grupo uno (intravenosa) presenta 45 animales y el grupo dos(oral) presenta 54 animales. En el grupo 1 se presentan 14 animales con diarrea post administración y el grupo 2 presenta 23 animales con diarrea post administración.

Proporciones

Comparación de dos Proporciones de dos poblaciones para muestras grandes.

- Pregunta científica, la proporción de caballos con diarrea post administración es igual entre poblaciones tratadas por vía intravenosa y oral.
- Las hipótesis (a dos colas)

$$H_0 : \pi_{intravenoso} = \pi_{oral}$$
$$H_A : \pi_{intravenoso} \neq \pi_{oral}$$

- El nivel de significancia se fija para este análisis en 0.05

Proporciones

Comparación de dos Proporciones de dos poblaciones para muestras grandes.

- La regla de decisión a dos colas

$$H_0 : \pi_{intravenoso} = \pi_{oral};\ \text{si } |z^*| \leq z(1-\alpha/2), \text{concluir } H_0$$
$$H_A : \pi_{intravenoso} \neq \pi_{oral};\ \text{si } |z^*| > z(1-\alpha/2), \text{concluir } H_A$$

- o reexpresado

$$H_0 : \pi_{intravenoso} = \pi_{oral};\ \text{si } |z^*| \leq z(0{,}975), \text{concluir } H_0$$
$$H_A : \pi_{intravenoso} \neq \pi_{oral};\ \text{si } |z^*| > z(0{,}975), \text{concluir } H_A$$

Proporciones

Comparación de dos Proporciones de dos poblaciones para muestras grandes.

- Las proporciones de cada muestra

$$P_1 = \frac{\sum_{i=1}^{n} X_i}{n_1} = \frac{14}{45} = 0{,}31$$
$$P_2 = \frac{\sum_{i=1}^{n} X_i}{n_2} = \frac{23}{54} = 0{,}43$$
$$P = \frac{X_1 + X_2}{n_1 + n_2} = \frac{14 + 23}{45 + 54} = 0{,}37$$

- El estadístico

$$z^* = \frac{P_1 - P_2 - \pi_1 - \pi_2}{\sqrt{P(1-P)(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0{,}31 - 0{,}43}{\sqrt{0{,}37(0{,}63)(0{,}041)}} = \frac{-0{,}12}{0{,}098} = -1{,}2$$

Proporciones

Comparación de dos Proporciones de dos poblaciones para muestras grandes.

- La regla de decisión dice que $|z^*| \leq z(0{,}975)$ aceptamos la hipótesis nula, encontramos que 1.22$\leq$1.96 luego aceptamos la nula, y concluimos, que hay evidencia estadística al 5 % que no hay diferencias en la proporción de enfermos entre las dos poblaciones.

¿Qué son los Odds?

Los odds son la razón entre el número de eventos y el número de no eventos.

$$Odds = \frac{\text{Número de casos}}{\text{Número de no casos}}$$

Ejemplo: 10 vacas enfermas, 90 sanas
Odds $= \frac{10}{90} = 0{,}111$

¿Qué es la Razón de Odds (OR)?

La razón de odds compara los odds entre dos grupos.

$$OR = \frac{\text{Odds}_{expuestos}}{\text{Odds}_{no\_expuestos}}$$

Ejemplo: Odds en expuestos = 0.111, Odds en no expuestos = 0.022
$\text{OR} = \frac{0{,}111}{0{,}022} = 5{,}05$

Interpretación de la Razón de Odds

- OR $= 1$: No hay asociación.
- OR > 1: Asociación positiva entre exposición y enfermedad.
- OR < 1: Asociación negativa.

Una OR de 5.05 sugiere fuerte asociación entre exposición y enfermedad.

Tabla de Contingencia 2x2

	Enfermos	Sanos
Expuestos	a	b
No Expuestos	c	d

- Riesgo expuestos $= \frac{a}{a+b}$
- Riesgo no expuestos $= \frac{c}{c+d}$
- Odds expuestos $= \frac{a}{b}$
- Odds no expuestos $= \frac{c}{d}$

Ejemplo en Epidemiología Animal

En una granja:

	Enfermos	Sanos
Expuestos	20	80
No Expuestos	5	95

- Riesgo expuestos $= \frac{20}{100} = 0{,}20$
- Riesgo no expuestos $= \frac{5}{100} = 0{,}05$
- $RR = \frac{0{,}20}{0{,}05} = 4$

Cálculo de Odds y OR

- Odds expuestos $= \frac{20}{80} = 0{,}25$
- Odds no expuestos $= \frac{5}{95} \approx 0{,}0526$
- $\mathrm{OR} = \frac{0{,}25}{0{,}0526} \approx 4{,}75$

Interpretación: Los animales expuestos tienen 4.75 veces más odds de enfermar.

Prueba de hipótesis e intervalo de confianza de 95 % para la razón de odds.

Hipótesis estadística para el logaritmo de la razón de odds, a dos colas:

$$H_0 : \log\left[\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}\right] = 0$$

$$H_A : \log\left[\frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}\right] \neq 0$$

Prueba de hipótesis e intervalo de confianza de 95 % para la razón de odds.

Para probar hipótesis o construir intervalos de confianza, se puede usar el siguiente estadístico (para muestras grandes):

$$Z^* = \frac{\log\left[\frac{p_1/(1-p_1)}{p_2/(1-p_2)}\right]}{\sqrt{\frac{1}{n_1(1-p_1)} + \frac{1}{n_1 p_1} + \frac{1}{n_2 p_2} + \frac{1}{n_2(1-p_2)}}} \sim \text{Normal}(0,1)$$

El intervalo de confianza del (1-α) %

$$\log\left[\frac{p_1/(1-p_1)}{p_2/(1-p_2)}\right] \pm Z_{\alpha/2} * \sqrt{\text{Var}\left(\log\left[\frac{p_1/(1-p_1)}{p_2/(1-p_2)}\right]\right)}$$

$$\text{Var}\left(\log\left[\frac{p_1/(1-p_1)}{p_2/(1-p_2)}\right]\right) = \frac{1}{n_1(1-p_1)} + \frac{1}{n_1 p_1} + \frac{1}{n_2 p_2} + \frac{1}{n_2(1-p_2)}$$

Tabla de contingencia dos x dos

Comparando proporciones en tablas de contingencia.

- Descripción y distribución de variables predictoras de un cuestionario acerca de factores de manejo de hato posiblemente asociados con virus de diarrea viral bovina VDVB en 55 hatos en Arequipa, Peru 2003-2004

Variable	Categoría	Número de Hatos VDVB negativo	VDVB positivo
Tamaño de hato	<20	21	14
	20 a 100	10	5
	>100	4	1
compra de animales	No	17	15
	Si	18	5
Vacunación	No	30	18
	Si	5	2

Tabla de contingencia dos x dos

Comparando proporciones en tablas de contingencia.

- Pruebe la hipótesis de igualdad de proporciones de la seropositividad a VDVB de animales vacunados y no vacunados
- Cual es la diferencia de proporciones entre animales seropositivos a VDVB con vacunación y sin vacunación.
- Cual es el riesgo relativo entre animales seropositivos a VDVB con vacunación y sin vacunación.
- Cual es la razón de Odds entre animales seropositivos a VDVB con vacunación y sin vacunación.
- Pruebe la hipótesis de igualdad de proporciones de la seropositividad a VDVB entre animales provenientes de fincas con menos de 20 animales comparado con fincas de más de 20 animales.
- Cual es la diferencia de proporciones entre animales seropositivos a VDVB animales provenientes de fincas con menos de 20 animales comparado con fincas de más de 20 animales.
- Cual es el riesgo relativo entre animales seropositivos a VDVB animales provenientes de fincas con menos de 20 animales comparado con fincas de más de 20 animales.

Tabla de contingencia dos x dos

Comparación de proporciones dependientes en tablas de contingencia. Prueba de Mc Nemar's.

Hipótesis estadística a dos colas:

$$H_0 : \pi_{1+} = \pi_{+1}$$
$$H_A : \pi_{1+} \neq \pi_{+1}$$

Hipótesis estadística a dos colas:

$$H_0 : \pi_{1+} > \pi_{+1}$$
$$H_A : \pi_{1+} \leq \pi_{+1}$$

Tabla de contingencia dos x dos

Comparación de proporciones dependientes en tablas de contingencia. Prueba de Mc Nemar's.

Estadístico: Bajo la hipótesis nula esperamos la misma frecuencia para los conteos n_{12} y n_{21}. Bajo la hipótesis nula, n_{12} y n_{21} son éxito y fracaso de una probabilidad binomial con media $n^* = n_{12} + n_{21}$ y probabilidad de éxito 1/2 ($p = 1/2$). El valor p es igual a

$$\binom{n^*}{n_{12}} p^{n_{12}} (1-p)^{n^* - n_{12}}$$

Cuando $n^* = 10$ se puede utilizar la aproximación normal.

$$Z^* = \frac{n_{12} - n_{21}}{\sqrt{n_{12} + n_{21}}} \sim \text{Normal}(0, 1)$$