

Prueba de Homogeneidad de Proporciones

Introducción

- ▶ El objetivo de la prueba es conocer si la distribución de la variable estudiada difiere en las “I” poblaciones de donde fueron tomadas las muestras.
- ▶ Las muestras independientes están clasificadas según una variable cualitativa.

Estructura de los datos

Se obtienen muestras aleatorias de dos o más poblaciones, y cada individuo se clasifica por valores de una variable categórica.

		Categoría						Total
		1	2	...	j	...	J	
Población	1	n_{11}	n_{12}	...	n_{1j}	...	n_{1J}	$n_{1.}$
	2	n_{21}	n_{22}	...	n_{2j}	...	n_{2J}	$n_{2.}$

	i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{iJ}	$n_{i.}$

	I	n_{I1}	n_{I2}	...	n_{Ij}	...	n_{IJ}	$n_{I.}$
Total		$n_{.1}$	$n_{.2}$...	$n_{.j}$...	$n_{.J}$	n

Donde:

n_{ij} : Número de observaciones de la celda (ij).

$n_{i.}$: Número de observaciones de la fila i.

$n_{.j}$: Número de observaciones de la columna j.

Procedimiento general

1) Se **formulan las hipótesis**:

H_0 : Las proporciones son iguales para todas las poblaciones

H_1 : Las proporciones no son iguales para todas las poblaciones

O equivalentemente:

$H_0 : p_1 = p_2 = \dots = p_I = p$

H_1 : Al menos dos son diferentes

2) Fijar el **nivel de significación** α .

3) Calcular el **valor del estadístico de prueba**:

$$\chi_0^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_{(I-1)(J-1)}^2$$

Donde:

$$E_{ij} = \frac{n_{i.} \times n_{.j}}{n} = \frac{(\text{Total fila } i)(\text{Total columna } j)}{\text{Total general}}$$

Procedimiento general

4) Determinar la **región crítica** y **regla de decisión**:

$$RC = \left\langle \chi^2_{((I-1)(J-1); 1-\alpha)}; \infty \right\rangle \Rightarrow \text{Se } \textbf{rechaza } H_0 \text{ si } \chi_0^2 > \chi^2_{((I-1)(J-1); 1-\alpha)}$$

5) Decisión y **conclusión**: **Rechazar** H_0 si el valor del estadístico de prueba pertenece a la región crítica. Se concluye que no todas las proporciones son iguales.

Ejemplo:

Se desea determinar si existen diferencias significativas entre las proporciones de niños que prefieren el tipo de cereal Z de las ciudades de Lima, Cusco, Arequipa y Trujillo. Para esto, se seleccionaron 200 niños al azar en cada una de estas ciudades, obteniéndose lo siguiente:

		Preferencia		Total
		Prefieren el cereal Z	No prefieren el cereal Z	
Ciudad	Lima	130	70	200
	Cusco	125	75	200
	Arequipa	135	65	200
	Trujillo	140	60	200
Total		530	270	800

¿Las proporciones de niños que prefieren el tipo de cereal Z son diferentes en las 4 ciudades? Usar $\alpha = 0.05$

Solución:

1) Se formulan las hipótesis:

H_0 : La proporción de niños que prefieren el cereal Z en las 4 ciudades son iguales

H_1 : La proporción de niños que prefieren el cereal Z en las 4 ciudades no son iguales

2) $\alpha = 0.05$

3) Valor del estadístico de prueba:

$$\chi_0^2 = \sum_{i=1}^4 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = 2.795 \sim \chi_3^2$$

► El estadístico de prueba tiene una distribución Chi-Cuadrado con $(I - 1)(J - 1) = (4 - 1)(2 - 1) = 3$ grados de libertad.

Tabla de frecuencias esperadas

		Preferencia	
		Prefieren el cereal Z	No prefieren el cereal Z
Ciudad	Lima	132.5	67.5
	Cusco	132.5	67.5
	Arequipa	132.5	67.5
	Trujillo	132.5	67.5

$$\bullet E_{11} = \frac{n_{1.} \times n_{.1}}{n} = \frac{530 \times 200}{800} = 132.5 \quad \bullet E_{12} = \frac{n_{1.} \times n_{.2}}{n} = \frac{270 \times 200}{800} = 67.5$$

$$\dots \quad \dots \quad \dots \quad \dots \quad \bullet E_{42} = \frac{n_{4.} \times n_{.2}}{n} = \frac{270 \times 200}{800} = 67.5$$

$$\chi_0^2 = \frac{(130 - 132.5)^2}{132.5} + \frac{(70 - 67.5)^2}{67.5} + \dots + \frac{(60 - 67.5)^2}{67.5} = 2.795$$



4) Región crítica y regla de decisión:

$$RC = \left\langle \chi^2_{((4-1)(2-1); 1-\alpha)}; \infty \right\rangle = \left\langle \chi^2_{(3; 0.95)}; \infty \right\rangle = \langle 7.815; \infty \rangle$$

Además, se **rechaza** H_0 si $\chi_0^2 > 7.815$.

5) **Decisión y conclusión:** Como $\chi_0^2 = 2.795 \notin RC$; por lo tanto, no se rechaza la hipótesis nula.

Se concluye que no existe suficiente evidencia para decir que la proporción de niños a favor del cereal Z es diferente en las 4 ciudades, al 5% de significación.

Ingreso de datos en R

```
prefieren <- c(130,125,135,140)
noprefieren <- c(70,75,65,60)

resultados <- matrix(c(prefieren,noprefieren), nrow = 4, ncol = 2,
                      dimnames = list(c("Lima","Cusco","Arequipa","Trujillo",
                                         c("Prefieren Z","No prefieren Z"))))

resultados
```

##	Prefieren Z	No prefieren Z
## Lima	130	70
## Cusco	125	75
## Arequipa	135	65
## Trujillo	140	60

Prueba de Homogeneidad de Proporciones en R

```
chisq.test(resultados)

##
##  Pearson's Chi-squared test
##
## data:  resultados
## X-squared = 2.7952, df = 3, p-value = 0.4243
```

Decisión y conclusión: Como $p\text{-value} = 0.4243 > \alpha = 0.05$, entonces no se rechaza la hipótesis nula. Por lo tanto, no existe suficiente evidencia para concluir que la proporción de niños a favor del cereal Z es diferente en las 4 ciudades, al 5% de significación.

Recursos Adicionales |

- Devore, J. (2019). *Introducción a la probabilidad y estadística para ingeniería y ciencias*. Cengage, 1 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_CENGAGE.
- Johnson, R. A. (2012). *Probabilidad y estadística para ingenieros*. Pearson Educación, 8 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_PEARSON.
- Kokoska, S. (2015). *Introductory Statistics*. W. H. Freeman and Company, 2 edition.
- Mendenhall, W., Beaver, R. J., and Beaver, B. M. (2015). *Introducción a la probabilidad y estadística*. Cengage, 14 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_CENGAGE.

Recursos Adicionales II

Millones, R., Barreno, E., Vásquez, F., and Castillo, C. (2017). *Estadística Descriptiva y Probabilidades: Aplicaciones en la ingeniería y los negocios*. Lima: Fondo Editorial de la Universidad de Lima, 1 edition. Código Biblioteca U.Lima: 519.53 E.

Triola, M. (2018). *Estadística*. Pearson Educación, 12 edition. Tomado de http://webaloe.ulima.edu.pe/portalUL/bi/baseDatosEtech/index.jsp?BD=BI_RUTA_PEARSON.