

Pruebas Basadas en la Distribución Cuadrado

Maria Teresa Silva Elgueta

Estadístico

Magister en Docencia Universitaria.

Master en Epidemiología y Salud Pública

Pruebas Chi Cuadrado (χ^2)

Las pruebas **chi cuadrado** o *ji-cuadrado* se debe a que (χ^2 , donde χ es la letra griega llamada “chi”), pero es un variable elevada al cuadrado.

se utiliza para determinar el comportamiento de cierta variable y también cuando se quiere saber si dos o más variables son independientes estadísticamente.

¿Cuándo se utilizan?

- Cuando tenemos datos de conteo o frecuencias.
- Cuando se desea determinar si la frecuencia de casos en una u otra condición tienen un comportamiento que nos llama la atención.
- Cuando tenemos un comportamiento lógico como patrón y vemos que en ciertas condiciones ese patrón no se repite.

Pruebas Chi Cuadrado χ^2

- **Bondad de Ajuste**

Se utiliza para la comparación de la distribución de una muestra con alguna distribución teórica que se supone describe a la población de la cual se extrajo.

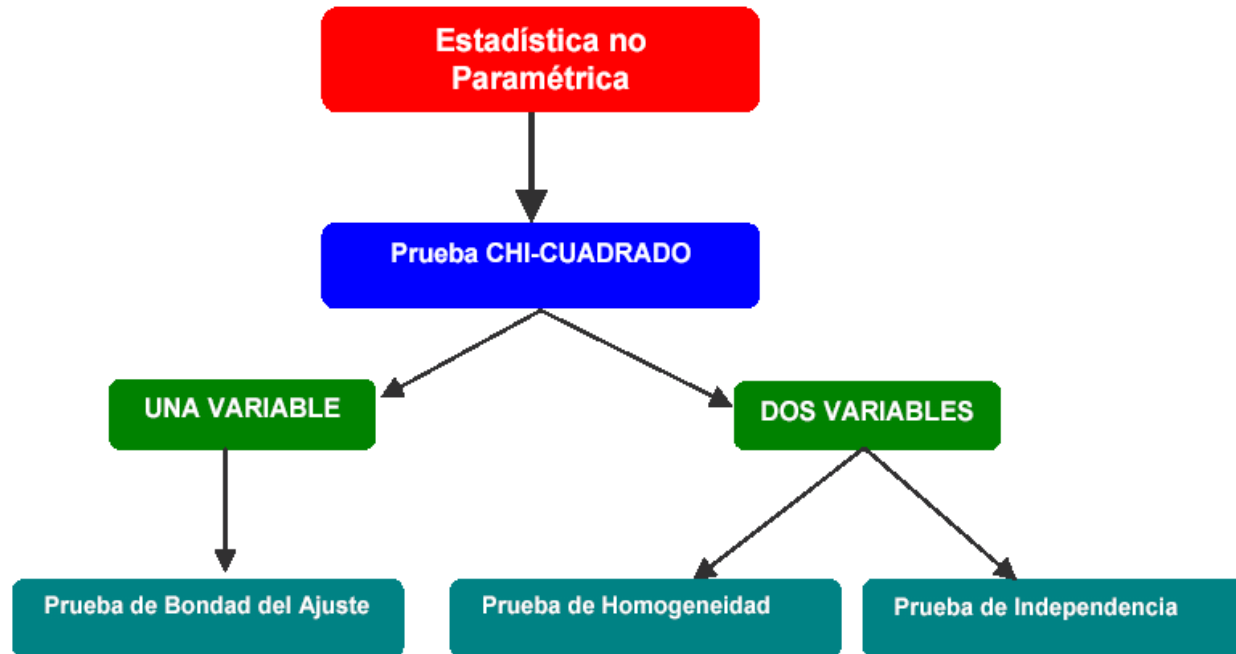
- **Independencia**

La H_0 indica que 2 variables o criterios de clasificación son independientes cuando se aplican a un conjunto de individuos (unidades de observación). Los totales Marginales son aleatorios

- **Homogeneidad**

Se extraen Muestras Independientes de varias poblaciones y se prueban para ver si son homogéneas con respecto a algún criterio de clasificación. Un total marginal es fijo mientras el otro es aleatorio.

Prueba Chi Cuadrado χ^2



Chi Cuadrado de AJUSTE (BONDAD DE AJUSTE)

Explicación.

Un pediatra señala que la enfermedad de niños menores de 5 años en Chiloé se distribuye en inviernos según lo siguiente:

Enfermedades	E1	E2	E3	E4	E5
Chiloé	30%	30%	10%	20%	10%

En Puerto Montt, una profesional del área señala que, de una muestra de 100 niños de la misma edad, los resultados fueron (**FRECUENCIAS OBSERVADAS**):

Enfermedades	E1	E2	E3	E4	E5	Niños Total
Puerto Montt	20	10	50	20	5	100

¿Se puede confirmar estadísticamente que la distribución de las enfermedades es similar?

Enfermedades	E1	E2	E3	E4	E5
Puerto Montt	20	10	50	20	5
%	20%	10%	50%	20%	5%

Si la distribución de las enfermedades en Puerto Montt fuera exactamente igual a los de Chiloé fuese similar, lo encontrado en Puerto Montt hubiese sido, (**que son las frecuencias ESPERADAS**)

Enfermedades	E1	E2	E3	E4	E5
Esperado E_i	30	30	10	20	10

Hipótesis:

- H_0 : La distribución real se ajusta a la teórica.
- H_1 : La distribución real no se ajusta a la teórica

La muestra es al azar filas y columnas

Chi cuadrado de Independencia

Se quiere estudiar si existe asociación entre el grupo sanguíneo y la severidad de una afección en una población. Investigadores reunieron los datos de 1500 personas.

Severidad de la condición	Tipo de Sangre				Total
	A	B	AB	O	
Ausente	500	200	95	471	1266
Moderada	70	22	8	31	131
Severa	45	20	7	31	103
Total	615	242	110	533	1500

Severidad de la condición	Tipo de Sangre				Total
	A	B	AB	O	
Ausente	33,3%	13,3%	6,3%	31,4%	84,4%
Moderada	4,7%	1,5%	0,5%	2,1%	8,7%
Severa	3,0%	1,3%	0,5%	2,1%	6,9%
Total	41,0%	16,1%	7,3%	35,5%	100,0%

Hipótesis:

- H_0 : No hay asociación entre el tipo de sangre y la severidad de la condición
- H_1 : No hay asociación entre el tipo de sangre y la severidad de la condición

Chi cuadrado de HOMOGENIDAD

Uno de los totales es fijo o filas o columnas y evalúa la distribución

Se quiere estudiar si la distribución de la severidad de una condición difiere según tipo de sangre. Para ellos se extrae información de 100 paciente de cada uno de los grupos sanguíneos, completando un total de 400 registros. Entonces, la distribución de niveles de severidad de la condición es similar según grupo sanguíneo

Severidad de la condición	Tipo de Sangre				Total
	A	B	AB	O	
Ausente	50	45	40	30	165
Moderada	25	30	40	30	125
Severa	25	25	20	40	110
Total	100	100	100	100	400

Severidad de la condición	Tipo de Sangre				Total
	A	B	AB	O	
Ausente	50%	45%	40%	30%	41%
Moderada	25%	30%	40%	30%	31%
Severa	25%	25%	20%	40%	28%
Total	100%	100%	100%	100%	100%

Hipótesis:

- La distribución de niveles de severidad de la condición es similar según grupo sanguíneo
- La distribución de niveles de severidad de la condición NO es similar según grupo sanguíneo

Condiciones para las pruebas Chi Cuadrado

- Muestra aleatoria, con tamaño ≥ 50 .
- Las observaciones independientes.
- Frecuencia esperadas > 0 .
- Se admite solo un 20% de las frecuencias esperadas menores que 5. Si esto no se cumple, se debe reagrupar.

Coefficiente phi, ϕ
Para tabla de doble entrada de (2x2).
Variables dicotómicas

Cuando las dos distribuciones que se desean correlacionar son realmente dicotómicas, debemos usar el coeficiente, ϕ llamado también para distribuciones puntuales, lo que supone que las dos clases tienen solamente dos valores o son atributo imposible de medir.

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

El coeficiente phi varía entre 0 y +1.

- Un valor 0 implica independencia de los atributos.
- Un valor 1 denota asociación perfecta.
- Valores cercanos a 1 indican un grado de asociación fuerte mientras que valores próximos a 0 implican un grado de asociación débil.

Coeficiente V de Cramer.

Interpretación

- Independiente de la cantidad de filas y columnas el V de Cramér's está entre 0 y 1. Puede usarse para tablas de contingencia de cualquier tamaño.
- Respecto del V de Cramér no se pueden hacer afirmaciones acerca de la dirección de la relación.
- Convenios para descifrar la magnitud de la asociación en la tabla de contingencia (Rea & Parker p. 203):

Valor de V Cramer's	Descripción
0,00 - <0,10	Asociación despreciable
0,10 - <0,20	Asociación débil
0,20 - <0,40	Asociación moderada
0,40 - <0,60	Asociación relativamente fuerte
0,60 - <0,80	Asociación fuerte
0,80 - 1,00	Asociación muy fuerte

Coeficiente de contingencia C de Pearson.

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

$$C_{\max} = \sqrt{\frac{(k-1)}{1+k}}$$

Como el máximo C, depende del número de hileras o columnas para tablas de contingencia.

El Coeficiente de contingencia o C de Pearson toma valores comprendidos entre 0 y C_{\max} siendo:

El valor K es = Min {número de filas, número de columnas}

- El valor C_{\max} denota asociación perfecta.
- Un valor 0 indica independencia

Fin



Bibliografía.

- Villa Romero AR, Moreno Altamirano L, García de la Torre GS. eds. Epidemiología y estadística en salud pública New York, NY: McGraw-Hill; 2012. <http://accessmedicina.mhmedical.com/content.aspx?bookid=1464§ionid=101049541>. Accessed febrero 05, 2019
 - García J, Alvarenga J, Ponce F, Tapia Y, Pérez L, Bernal A. eds. Metodología de la investigación, bioestadística y bioinformática en ciencias médicas y de la salud, 2e New York, NY: McGraw-Hill; . <http://accessmedicina.mhmedical.com/content.aspx?bookid=1721§ionid=115929370>. Accessed febrero 05, 2019.
 - Moreno Altamirano L. eds. Epidemiología clínica, 3e New York, NY: McGraw-Hill; 2013. <http://accessmedicina.mhmedical.com/content.aspx?bookid=1442§ionid=101158221>. Accessed febrero 05, 2019.
-