

## Prueba de Independencia. Test Chi Cuadrado

La prueba Chi-cuadrado de Pearson una de las técnicas estadísticas más usadas en la evaluación de datos de conteo o frecuencias, principalmente en los análisis de tablas de contingencia donde se resumen datos categóricos.

Uno de los usos de la Prueba Chi- Cuadrado es para probar la independencia de dos variables entre sí, mediante la presentación de los datos en tablas de contingencia. Se tiene una muestra de  $n$  individuos que se clasifican respecto a dos variables,  $X$  e  $Y$ , preferentemente cualitativas (nominales dicotómicas o politómicas) y se desea conocer a partir de datos muestrales, si existe asociación de estas a nivel poblacional.

Supongamos tenemos los datos resumidos en una *tabla de contingencia*, una de las formas más comunes de resumir datos categóricos, están compuestas por  $m$  filas (horizontales), para la información de una variable y  $k$  columnas (verticales) para la información de la otra variable. Estas filas y columnas delimitan *celdas* donde se vuelcan las frecuencias de cada combinación de las variables analizadas.

Se consideran  $X$  e  $Y$  dos variables con  $m$  y  $k$  categorías respectivamente, y la siguiente tabla de contingencia de  $2 \times 2$

Variable X Categorías	Variable Y Categorías		
	$k_1$	$k_2$	Total
$m_1$	$n_{m_1 k_1}$	$n_{m_1 k_2}$	Fila $m_1$
$m_2$	$n_{m_2 k_1}$	$n_{m_2 k_2}$	Fila $m_2$
Total	Columna $k_1$	Columna $k_2$	$n$

Se plantea la hipótesis:

$$H_0: \text{Existe independencia entre las variables estudiadas } X \text{ e } Y$$

$$H_1: \text{No existe independencia entre las variables estudiadas } X \text{ e } Y$$

Se considera el estadístico de prueba:

$$\chi^2_{\text{calculado}} = \sum_{i=1}^k \frac{(n_i - e_i)^2}{e_i} \sim \chi^2_{(1-\alpha, (m-1)(k-1))}$$

Donde:

- $m$  = número de categorías para la variable  $X$
- $k$  = número de categorías para la variable  $Y$
- $n_i$  = frecuencia observada de una muestra, resumida en una tabla de contingencia
- $e_i$  = frecuencia esperada

Se tiene que en cada frecuencia observada en una tabla de contingencia de  $m \times k$ , hay una frecuencia esperada de cada frecuencia observada, se calcula con la siguientes formula

$$e_i = \frac{\text{Total columna } i \cdot \text{Total fila } i}{\text{Total general}}$$

Toda la información se puede resumir en una Tabla de contingencia

	<b>Categorías</b>			
$n_i$ : frecuencia observada	$n_{m_1 k_1}$	$n_{m_2 k_1}$	$n_{m_1 k_2}$	$n_{m_2 k_2}$
$e_i$ : frecuencia esperada	$e_1$	$e_2$	$e_3$	$e_4$

Se calcula el valor critico

$$\chi^2_{(1-\alpha, (m-1)(k-1))}$$

obtenido en una tabla de distribución de probabilidad

luego, la regla de decisión es:

Se rechaza  $H_0$  al nivel de significación  $\alpha$ , si  $\chi^2_{calculado} > \chi^2_{(1-\alpha, (m-1)(k-1))}$

Se concluye que se rechaza  $H_0$ , es decir, no existe independencia entre las variables estudiadas,  $X$  e  $Y$

**Observación:** La muestra debe ser lo suficientemente grande. Si menos del 20% de las celdas de la tabla de contingencia, presentan valores esperados menores o iguales a 5, no se recomienda aplicar la prueba.

### Ejemplo

A un grupo de 350 adultos que participaron en una escuela de salud, se les preguntó si llevaban o no una dieta. Las respuestas por sexo, son las siguientes:

<b>Sexo</b>	<b>Dieta</b>		
	<b>Con Dieta</b>	<b>Sin Dieta</b>	<b>Total</b>
<b>Masculino</b>	14	159	173
<b>Femenino</b>	25	152	177
<b>Total</b>	39	311	350

¿Sugieren estos datos que el estar o no la dieta, depende del tipo de sexo? Para un nivel de significancia del 5%.

### Desarrollo

**Datos:**

X = Tipo de sexo, con  $m = 2$

Y = Dieta, con  $k = 2$

**Hipótesis:**

$H_0$  : No influye que el estar o no a dieta, dependa del tipo de sexo (las dos variables en estudio son independientes)

$H_1$  : Sí influye que el estar o no a dieta, dependa del tipo de sexo (las dos variables en estudio son dependientes)

**Calculo de frecuencias esperadas y resumen en tabla de contingencia:**

Calculamos las frecuencias esperadas

$$e_1 = \frac{39 \cdot 173}{350} = 19,277, \quad e_2 = \frac{39 \cdot 177}{350} = 19,723, \quad e_3 = \frac{311 \cdot 173}{350} = 153,723, \quad e_4 = \frac{311 \cdot 177}{350} = 157,277$$

resumimos en la siguiente tabla

$n_i$ frecuencia observada	14	25	159	152
$e_i$ frecuencia esperada	19,277	19,723	153,723	157,277

**Estadístico de prueba (calculado):**

$$\begin{aligned} \chi^2_{\text{calculado}} &= \sum_{i=1}^k \frac{(n_i - e_i)^2}{e_i} = \frac{(14 - 19.277)^2}{19.277} + \frac{(25 - 19.723)^2}{19.723} + \frac{(139 - 153.723)^2}{153.723} \\ &\quad + \frac{(152 - 157.277)^2}{157.277} \\ &= 1.445 + 1.412 + 0.181 + 0.177 = 3.215 \end{aligned}$$

**Valor critico (por tabla):**

$$\chi^2_{(1-\alpha, (m-1)(k-1))} = \chi^2_{(1-0.05, (2-1)(2-1))} = \chi^2_{(0.95, 1)} = 3,841$$

**Decisión:**

$$\text{Como } \chi^2_{\text{calculado}} = 3.215 < \chi^2_{(1-\alpha, (m-1)(k-1))} = 3.841$$

No se puede rechazar  $H_0$ , es decir, que no existe relación entre las dos variables en estudio.

Por lo que con un 95% de confianza no hay dependencia en el estar a dieta o no, y el género de la persona

## Ejercicios

1. La asociación de lesiones papilomatosas, según el sexo en delfines *Tursiops truncatus*, estudiados en la costa norte central de Cuba, presenta a continuación una tabla de contingencia, con datos de acuerdo a sexo y síntoma de la enfermedad:

Sexo	Síntomas de la enfermedad		
	Sintomáticos	Asintomáticos	Total
Hembras	27	122	149
Machos	52	62	114
Total	79	184	263

¿Influye el tipo de sexo, en los síntomas de la enfermedad? Considere un nivel de significancia de 0,05

### Desarrollo

2. Se desea comparar la efectividad de dos análisis de laboratorio para detectar la presencia de bacterias en equinodermos (erizo rojo), para ello se selecciona dos muestras independientes de *Loxechinus albus* (erizo rojo comestible) en un sector de caleta coloso y los resultados de los análisis fueron los siguientes:

Análisis	Detección de la Bacteria		
	Si	No	Total
1	11	39	50
2	8	42	50
Total	19	81	100

Pruebe la hipótesis de que la detección de la bacteria es independiente del tipo de análisis usado, con un nivel de significación de 0.05.

### Desarrollo

3. Se desea conocer si existe asociación entre el hábito de fumar y el bajo peso al nacer en una población, para lo cual se selecciona una muestra aleatoria de 100 recién nacidos, obteniéndose los resultados siguientes:

Hábito de fumar	Bajo peso al nacer		
	Si	No	Total
Presente	30	10	40
Ausente	20	40	60
Total	50	50	100

¿Permiten estos datos afirmar que el Bajo peso del recién nacido depende del hábito de fumar?  
Usar un nivel de significación de 0,05

### Desarrollo

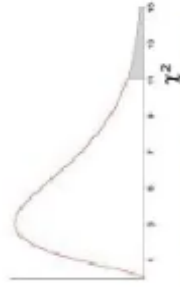
**Ejercicio propuesto**

4. Un experimento para investigar el efecto de vacunación de animales de laboratorio refleja la siguiente tabla:

Vacuna	Animal laboratorio		
	Enfermo	Sano	Total
Vacunado	9	42	51
No Vacunado	18	28	46
Total	27	70	97

Con un nivel de significación de 0,05, ¿Es conveniente vacunar?

Tabla D.7: VALORES CRÍTICOS DE LA DISTRIBUCIÓN JI CUADRADA



$\alpha$	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	$\alpha$
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,302	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
10	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15
16	39,252	34,267	32,000	29,633	28,845	28,191	27,136	26,296	23,542	21,793	20,465	19,369	18,418	17,565	16,780	16
17	40,790	35,718	33,409	30,995	30,191	29,523	28,445	27,587	24,769	22,977	21,615	20,489	19,511	18,633	17,824	17
18	42,312	37,156	34,805	32,346	31,526	30,845	29,745	28,869	25,989	24,155	22,760	21,605	20,601	19,699	18,868	18
19	43,820	38,582	36,191	33,687	32,852	32,158	31,037	30,144	27,204	25,329	23,900	22,718	21,689	20,764	19,910	19
20	45,315	39,997	37,566	35,020	34,170	33,462	32,321	31,410	28,412	26,498	25,038	23,828	22,775	21,826	20,951	20
21	46,797	41,401	38,932	36,343	35,479	34,759	33,597	32,671	29,615	27,662	26,171	24,935	23,858	22,888	21,991	21
22	48,268	42,796	40,289	37,659	36,781	36,049	34,867	33,924	30,813	28,822	27,301	26,039	24,939	23,947	23,031	22
23	49,728	44,181	41,638	38,968	38,076	37,332	36,131	35,172	32,007	29,979	28,429	27,141	26,018	25,006	24,069	23
24	51,179	45,559	42,980	40,270	39,364	38,609	37,389	36,415	33,196	31,132	29,553	28,241	27,096	26,063	25,106	24
25	52,620	46,928	44,314	41,566	40,646	39,880	38,642	37,652	34,382	32,282	30,675	29,339	28,172	27,118	26,143	25
26	54,052	48,290	45,642	42,856	41,923	41,146	39,889	38,885	35,563	33,429	31,795	30,435	29,246	28,173	27,179	26
27	55,476	49,645	46,963	44,140	43,195	42,407	41,132	40,113	36,741	34,574	32,912	31,528	30,319	29,227	28,214	27
28	56,892	50,993	48,278	45,419	44,461	43,662	42,370	41,337	37,916	35,715	34,027	32,620	31,391	30,279	29,249	28
29	58,301	52,336	49,588	46,693	45,722	44,913	43,604	42,557	39,087	36,854	35,139	33,711	32,461	31,331	30,283	29
30	59,703	53,672	50,892	47,962	46,979	46,160	44,834	43,773	40,256	37,990	36,250	34,800	33,530	32,382	31,316	30
31	61,098	55,003	52,191	49,226	48,232	47,402	46,059	44,985	41,422	39,124	37,359	35,887	34,598	33,431	32,349	31
32	62,487	56,328	53,486	50,487	49,480	48,641	47,282	46,194	42,585	40,256	38,466	36,973	35,665	34,480	33,381	32
33	63,870	57,648	54,776	51,743	50,725	49,876	48,500	47,400	43,745	41,386	39,572	38,058	36,731	35,529	34,413	33
34	65,247	58,964	56,061	52,995	51,966	51,107	49,716	48,602	44,903	42,514	40,676	39,141	37,795	36,576	35,444	34
35	66,619	60,275	57,342	54,244	53,203	52,335	50,928	49,802	46,059	43,640	41,778	40,223	38,859	37,623	36,475	35
40	73,402	66,766	63,691	60,436	59,342	58,428	56,946	55,758	51,805	49,244	47,259	45,616	44,165	42,848	41,622	40
60	99,607	91,952	88,379	84,580	83,298	82,125	80,482	79,082	74,397	71,341	68,972	66,981	65,127	63,628	62,135	60
80	124,839	116,321	112,329	108,069	106,629	105,422	103,459	101,879	96,578	93,106	90,405	88,130	86,120	84,284	82,566	80
90	137,208	128,299	124,116	119,648	118,136	116,869	114,806	113,145	107,565	103,904	101,054	98,650	96,524	94,581	92,761	90
100	149,449	140,169	135,807	131,142	129,561	128,037	126,079	124,342	118,437	114,657	111,667	109,141	106,906	104,862	102,946	100
120	173,617	163,648	158,950	153,918	152,211	150,780	148,447	146,567	140,233	136,062	132,806	130,055	127,616	125,383	123,289	120
140	197,451	186,847	181,840	176,471	174,648	173,118	170,624	168,613	161,827	157,352	153,854	150,894	148,269	145,863	143,604	140