La prueba de ji-cuadrado

Elaborada por Luis Temis y Dana Gómez

En teoría de la probabilidad y en estadística, la **distribución ji al cuadrado** (también llamada **distribución de Pearson** o **distribución** χ^2) con $k \in \mathbb{N}$ grados de libertad es la distribución de la suma del cuadrado de k variables aleatorias independientes con distribución normal estándar. La distribución chi cuadrada es un caso especial de la distribución gamma y es una de las distribuciones de probabilidad más usadas en Inferencia Estadística, principalmente en pruebas de hipótesis y en la construcción de intervalos de confianza.

El estadístico ji-cuadrado (o chi cuadrado), que tiene distribución de probabilidad, sirve para someter a prueba hipótesis referidas a distribuciones de frecuencias, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula.

Distribución χ^2= asociación entre dos variables utilizando una situación hipotética y datos simulados.

Evalúa cuán buena puede resultar una distribución teórica.

- ➔ Pretende representar la distribución real de los datos de una muestra determinada.
- ➔ Se le llama evaluar la bondad de un ajuste. Probar la bondad de un ajuste es ver en qué medida se ajustan los datos observados a una distribución teórica o esperada.
- ➔ Se utiliza una segunda situación hipotética y datos simulados.

Supongamos que un investigador está interesado en evaluar la asociación entre uso de cinturón de seguridad en vehículos particulares y el nivel socioeconómico del conductor del vehículo. Con este objeto se toma una muestra de conductores a quienes se clasifica en una tabla de asociación, encontrando los siguientes resultados:

Uso de cinturón	Nivel socioeconómico bajo	Nivel socioeconómico medio	Nivel socioeconómico alto	TOTAL
SI	8	15	28	51
NO	13	16	14	43
TOTAL	21	31	42	94

Tabla I. Tabla de asociación, valores observados.

H0: “El uso de cinturón de seguridad es independiente del nivel socioeconómico”.
H1: “El uso de cinturón de seguridad depende del nivel socioeconómico”.

Las frecuencias esperadas se obtendrán de la distribución de frecuencias del total de los casos, 51 personas de un total de 94 usan el cinturón y 43 de 94 no lo usan. Esa misma proporción se debería dar al interior de los tres grupos de nivel socioeconómico, de manera que el cálculo responde al siguiente razonamiento: si de 94 personas 51 usan cinturón; de 21 personas, ¿cuántas debieran usarlo?

La respuesta a esta pregunta se obtiene aplicando la “regla de tres” y es 11,4. Este procedimiento debe repetirse con todas las frecuencias del interior de la tabla.
El detalle de los cálculos es el siguiente:
Nivel bajo: (21x51/94)=11,4-(21x43/94)=9,6
Nivel medio: (31x51/94)=16,8-(31x43/94)=14,2
Nivel alto: (42x51/94)=22,8-(42x43/94)=19,2

En este caso, el estadístico de prueba es Ji-cuadrado que, como dijimos al comienzo, compara las frecuencias que entregan los datos de la muestra (frecuencias observadas) con las frecuencias esperadas, y tiene la siguiente fórmula cálculo:

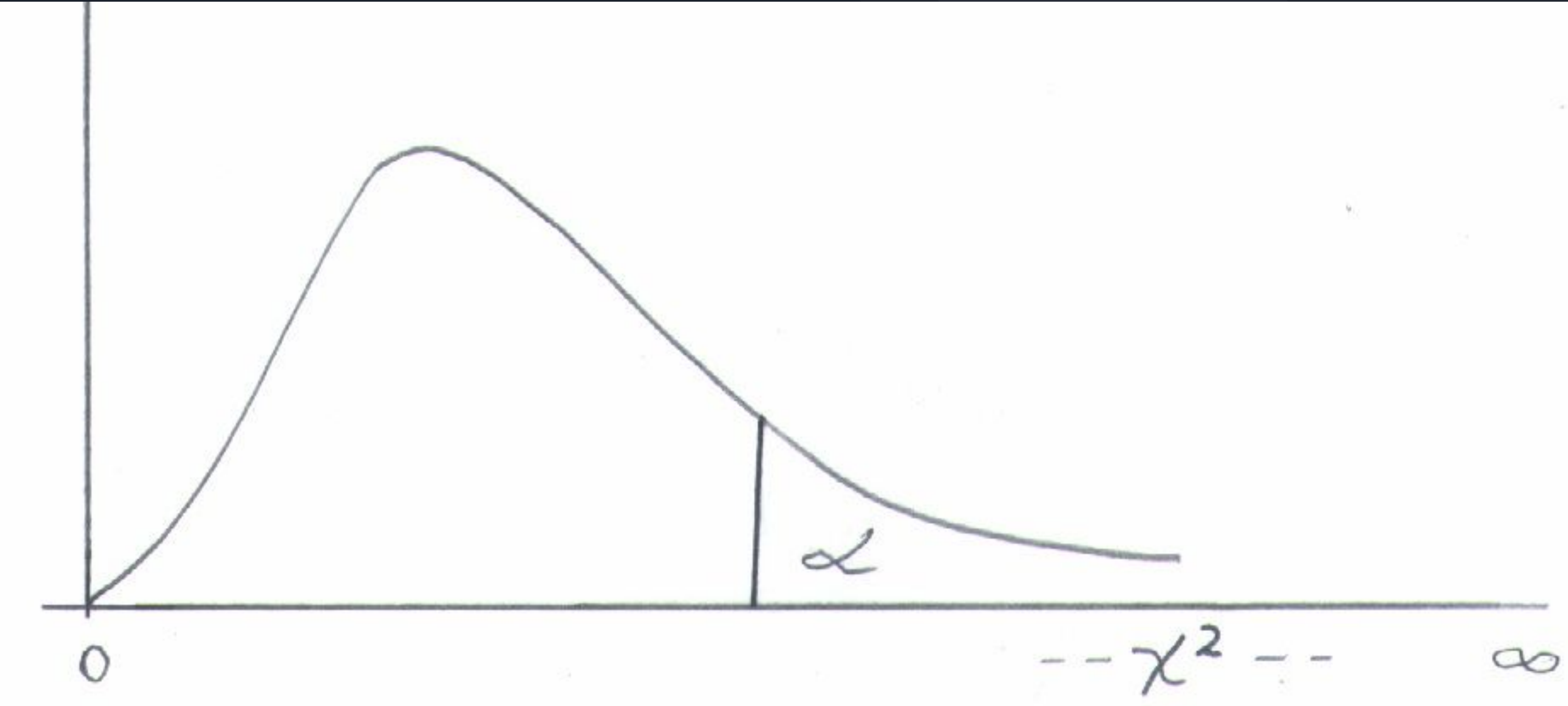
$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$$

Donde oi representa a cada frecuencia observada y ei representa a cada frecuencia esperada. De este modo el valor del estadístico de prueba para este problema será:

$$\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i} =$$

$$\frac{(8-11{,}4)^2}{11{,}4} + \frac{(13-9{,}6)^2}{9{,}6} + \frac{(15-16{,}8)^2}{16{,}8} + \frac{(16-14{,}2)^2}{14{,}2} + \frac{(28-22{,}8)^2}{22{,}8} + \frac{(14-19{,}2)^2}{19{,}2} = 5{,}23$$

Entonces X2= 5,23. Este es el valor de nuestro estadístico de prueba que ahora, siguiendo el procedimiento de problemas anteriores (paso 4), debemos comparar con un valor de la tabla de probabilidades para ji-cuadrado (x2). Esta tabla es muy parecida a la tabla t de student, pero tiene sólo valores positivos porque ji-cuadrado sólo da resultados positivos. Véase gráfico 1, que muestra la forma de la curva, con valores desde 0 hasta infinito.



α
0
- - χ² - - -
∞

1. Uso de tabla ji-cuadrado

La tabla de ji-cuadrado tiene en la primera columna los grados de libertad y en la primera fila la probabilidad asociada a valores mayores a un determinado valor del estadístico.
Los grados de libertad dependen del número de celdas que tiene la tabla de asociación donde están los datos del problema y su fórmula de cálculo es muy sencilla:

Grados de libertad (gl)=(n° de filas–1)x(n° de columnas–1)

Así, en nuestro ejemplo, en que hay 2 filas y 3 columnas, los grados de libertad serán:

gl=(2-1)x(3-1)=2

Nótese que no se consideran la fila ni la columna de los totales.

Al comienzo elegimos un nivel de significación alfa=0,05. Entonces un valor de tabla para x2 asociado a 2 grados de libertad y alfa 0,05 es 5,99.

Por lo tanto, como en el gráfico 2 vemos que 5,23 se encuentra a la izquierda de 5,99, la probabilidad asociada a valores superiores a 5,23 es mayor que alfa (0,05).

Grados de libertad

alfa

π/ϕ	0.995	0.99	0.975	0.95	0.9	0.75	0.5	0.25	0.1	0.05	0.025	0.01	0.005	π/ϕ
1	3.93E-05	1.57E-04	9.82E-04	3.93E-03	1.58E-02	0.102	0.455	1.323	2.71	3.84	5.02	6.63	7.88	1
2	1.00E-02	2.01E-02	5.06E-02	0.103	0.211	0.575	1.386	2.77	4.61	5.99	7.38	9.21	10.60	2
3	7.17E-02	0.115	0.216	0.352	0.584	1.213	2.37	4.11	6.25	7.81	9.35	11.34	12.84	3
4	0.207	0.297	0.484	0.711	1.064	1.923	3.36	5.39	7.78	9.49	11.14	13.28	14.86	4
5	0.412	0.554	0.831	1.145	1.610	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	5
6	0.676	0.872	1.237	1.635	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	6
7	0.989	1.239	1.690	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.3	7
8	1.344	1.647	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.1	22.0	8
9	1.735	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.7	23.6	9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.5	23.2	25.2	10
11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.9	24.7	26.8	11
12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.0	23.3	26.2	28.3	12
13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.4	24.7	27.7	29.8	13
14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.1	23.7	26.1	29.1	31.3	14
15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.3	25.0	27.5	30.6	32.8	15
16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.5	26.3	28.8	32.0	34.3	16
17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.5	24.8	27.6	30.2	33.4	35.7	17
18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.6	26.0	28.9	31.5	34.8	37.2	18
19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.7	27.2	30.1	32.9	36.2	38.6	19
20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.8	28.4	31.4	34.2	37.6	40.0	20
21	8.03	8.90	10.28	11.59	13.24	16.34	20.3	24.9	29.6	32.7	35.5	38.9	41.4	21
22	8.64	9.54	10.98	12.34	14.04	17.24	21.3	26.0	30.8	33.9	36.8	40.3	42.8	22
23	9.26	10.20	11.69	13.09	14.85	18.14	22.3	27.1	32.0	35.2	38.1	41.6	44.2	23
24	9.89	10.86	12.40	13.85	15.66	19.04	23.3	28.2	33.2	36.4	39.4	43.0	45.6	24
25	10.52	11.52	13.12	14.61	16.47	19.94	24.3	29.3	34.4	37.7	40.6	44.3	46.9	25
26	11.16	12.20	13.84	15.38	17.29	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	26
27	11.81	12.88	14.57	16.15	18.11	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	27
28	12.46	13.56	15.31	16.93	18.94	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	28
29	13.12	14.26	16.05	17.71	19.77	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	29
30	13.79	14.95	16.79	18.49	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	30
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	40
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5	50
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0	60
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	104.2	70
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.3	116.3	80
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3	90
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	140.2	100
Z_α	-2.58	-2.33	-1.96	-1.64	-1.28	-0.674	0.000	0.674	1.282	1.645	1.96	2.33	2.58	Z_α