

# Distribución Ji-Cuadrada

Tal como en el caso de otras distribuciones de muestreo, es posible usar métodos matemáticos para llegar a la distribución de frecuencia teórica deseada. Puesto que existe solamente un número limitado de valores posibles para las frecuencias de las celdas en la Tabla de datos, existe sólo un número limitado de valores de  $x^2$  posibles. Así pues, la distribución teórica de  $x^2$  debe ser una distribución discreta. Puesto que una distribución discreta con muchos valores posibles requiere la aplicación de cálculos largos y tediosos, consideraciones de índole práctica demanda una aproximación sencilla y continua de la distribución discreta de  $x^2$ , en forma similar a la aproximación normal de la distribución binomial. Esta distribución continua la tenemos en lo que se conoce como la distribución de ji-cuadrada. Es una circunstancia poco afortunada el que la distribución continua se aproxima a la distribución discreta de ji-cuadrada se llame también distribución ji-cuadrada; sin embargo, no habrá lugar a confusión debido a que la distribución continua será la única que se une.

**Ejemplo 1.** Considérese el siguiente problema de genética: en experimentos de cruce de flores de cierta especie, un experimentador ha obtenido 120 flores color violeta con estigma verde, 48 flores violeta con estigma rojo, 36 flores rojas con estigma verde y 13 flores rojas con estigma rojo. La teoría mendeliana predice que las flores de estos tipos deberían obtenerse de las relaciones 9:3:3:1. ¿Son compatibles los resultados anteriores con la teoría?

En términos de la notación de prueba de hipótesis, este es un problema de prueba de hipótesis:

$$H_0: p_1 = \frac{9}{16}, p_2 = \frac{3}{16}, p_3 = \frac{3}{16}, p_4 = \frac{1}{16}$$

Para este problema,  $n = 217$  y  $k = 4$ . Las frecuencias esperadas para las 4 celdas se obtienen multiplicando 217 por cada una de las probabilidades  $H_0$ . Las frecuencias observadas junto con las esperadas, correctas al entero más próximo, aparecen en la Tabla 1.

$o_i$	120	48	36	13
$e_i$	122	41	41	14

Tabla 1

Los cálculos dan:

$$x^2 = \frac{(120 - 122)^2}{122} + \frac{(48 - 41)^2}{41} + \frac{(36 - 41)^2}{41} + \frac{(13 - 14)^2}{14} = 1.9$$

Una característica notable en la distribución  $x^2$  es que su forma depende solamente del número de celdas. La figura 2 da las gráficas de 6 curvas  $x^2$  correspondientes al número de celdas variable de 2 a 7. Es costumbre identificar como una distribución  $x^2$  por medio del parámetro  $v = k - 1$ , llamado número de grados de libertad, en lugar del número de celdas. Los grados de libertad de la frase se refieren al número de frecuencias independientes de las celdas. Puesto que la suma de las 4 frecuencias observadas en la Tabla 1 debe igualar a 217, la cuarta frecuencia se determina tan pronto como se han especificado las 3 primeras frecuencias de las celdas. Así pues, existen  $v = 3$  grados de libertad para este problema. El valor de  $x^2$  que corta en la cola derecha el 5% de la distribución para  $v = 3$  resulta ser 7.8. Puesto que el valor de  $x^2$  para la Tabla 2, es decir 1.9, no se encuentra en la región crítica, el resultado no es significativo. No existe razón, como consecuencia de esta prueba, para dudar de que la teórica mendeliana se aplica a los datos de la Tabla 1.

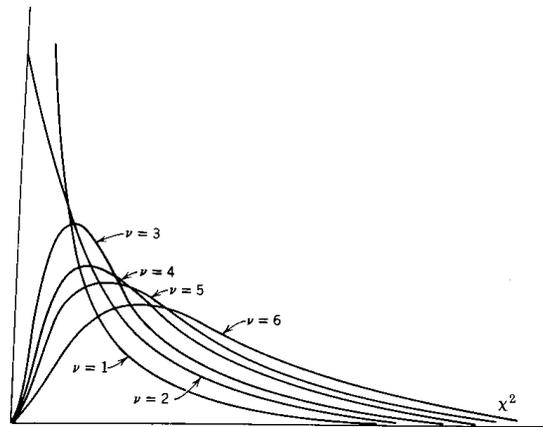


Figura 1 Distribución de  $x^2$  para varios grados de libertad.

Los valores de  $x^2$  que han determinado las regiones críticas del 5% se han obtenido de la Tabla VII. El valor crítico del 5% ha sido encontrado en la columna con el encabezado 0.05 y la línea correspondiente al número apropiado de grado de libertad  $\nu = k - 1$ .

La prueba ji-cuadrada es aplicable a problemas en que las probabilidades de las celdas no han sido especificadas, pero que pueden calcularse mediante una hipótesis de "homogeneidad". El siguiente es un ejemplo de esto.

**Ejemplo 2.** En el Journal of Industrial Economics de noviembre de 1972 se publicó un artículo que catalogaba a las 100 más grandes compañías de manufactura del Reino Unido para el año 1948. Se dividió esta clasificación en cuatro grupos: las 12 más grandes, las 13 siguientes, las 25 "terceras" y los 50 restantes. Se determinó cuántas de estas compañías estaban aun en operación (vivas) en 1968. Los resultados del estudio son los que se muestran en la tabla siguiente:

Rango en 1948	Sobrevivientes hasta 1968
1-12	10
13-25	8
26-50	14
51-100	32

El problema consiste en determinar si es verdad que cuanto mayor es la compañía, mayores son sus oportunidades de supervivencia. Puede usarse la prueba de ji-cuadrada para ayudar en esta determinación probando la hipótesis de que la tasa de supervivencia es la misma para todas las categorías. Esta hipótesis de "homogeneidad" permite calcular las probabilidades de las celdas. Para hacerlo es necesario, primero, enlistar el número de compañías en las celdas. La lista de la tabla siguiente.

Número en celda en 1948	Número de sobrevivientes en 1968
12	10
13	8
25	14
50	32

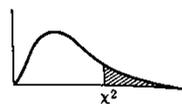
Como el número de compañías originales es 100, las proporciones de esas compañías en las cuatro celdas para los datos de 1948 son 0.12, 0.13, 0.25 y 0.50. Estas son por tanto las posibilidades que hay que asignar a las cuatro celdas para los datos de 1968, con el supuesto de tasa uniforme de supervivencia. Las frecuencias esperadas se obtienen multiplicando estas posibilidades por 64 que es el número total de compañías sobrevivientes. Esto nos da la siguiente tabla de frecuencias obtenidas y esperadas.

$o_i$	10	8	14	32
$e_i$	8	8	16	32

El valor de  $\chi^2$  es 0.75. Como hay 3 grados de libertad para este problema y el valor crítico es 7.8, se acepta la hipótesis de homogeneidad. Las sobrevivencias reales son sorprendentemente cercanas a las esperadas con el supuesto de que el tamaño de la compañía nada tenía que ver con sus posibilidades de supervivencia.

**TABLA VII. La Distribución D  $\chi^2$**

En la primera columna aparece el número de grados de libertad ( $\nu$ ). Los encabezados de las otras columnas dan las probabilidades ( $P$ ) de que  $\chi^2$  exceda el valor de la casilla. Para  $\nu > 100$ , tratar  $\sqrt{2\chi^2} - \sqrt{2\nu} - 1$  como una variable normal estándar.



$\nu$	0.995	0.975	0.050	0.025	0.010	0.005
1	0.004327	0.009821	3.84146	5.02389	6.63490	7.87944
2	0.010025	0.050636	5.99147	7.37776	9.21034	10.5966
3	0.071721	0.215795	7.81473	9.34840	11.3449	12.8381
4	0.206990	0.484419	9.48773	11.1433	13.2767	14.8602
5	0.411740	0.831211	11.0705	12.8325	15.0863	16.7496
6	0.675727	1.237347	12.5916	14.4494	16.8119	18.5476
7	0.989265	1.68987	14.0671	16.0128	18.4753	20.2777
8	1.344419	2.17973	15.5073	17.5346	20.0902	21.9550
9	1.734926	2.70039	16.9190	19.0228	21.6660	23.5893
10	2.15585	3.24697	18.3070	20.4831	23.2093	25.1882
11	2.60321	3.81575	19.6751	21.9200	24.7250	26.7569
12	3.07382	4.40379	21.0261	23.3367	26.2170	28.2995
13	3.56503	5.00874	22.3621	24.7356	27.6883	29.8194
14	4.07468	5.62872	23.6848	26.1190	29.1413	31.3193
15	4.60094	6.26214	24.9958	27.4884	30.5779	32.8013
16	5.14224	6.90766	26.2962	28.8454	31.9999	34.2672
17	5.69724	7.56418	27.5871	30.1910	33.4087	35.7185
18	6.26481	8.23075	28.8693	31.5264	34.8053	37.1564
19	6.84398	8.90655	30.1435	32.8523	36.1908	38.5822
20	7.43386	9.59083	31.4104	34.1696	37.5662	39.9968
21	8.03366	10.28293	32.6705	35.4789	38.9321	41.4010
22	8.64272	10.9823	33.9244	36.7807	40.2894	42.7956
23	9.26042	11.6885	35.1725	38.0757	41.6384	44.1813
24	9.88623	12.4001	36.4151	39.3641	42.9798	45.5585
25	10.5197	13.1197	37.6525	40.6465	44.3141	46.9278
26	11.1603	13.8439	38.8852	41.9232	45.6417	48.2899
27	11.8076	14.5733	40.1133	43.1944	46.9630	49.6449
28	12.4613	15.3079	41.3372	44.4607	48.2782	50.9933
29	13.1211	16.0471	42.5569	45.7222	49.5879	52.3356
30	13.7867	16.7908	43.7729	46.9792	50.8922	53.6720
40	20.7065	24.4331	55.7585	59.3417	63.6907	66.7659
50	27.9907	32.3574	67.5048	71.4202	76.1539	79.4900
60	35.5346	40.4817	79.0819	83.2976	88.3794	91.9517
70	43.2752	48.7576	90.5312	95.0231	100.425	104.215
80	51.1720	57.1532	101.879	106.629	112.329	116.321
90	59.1963	65.6466	113.145	118.136	124.116	128.299
100	67.3276	74.2219	124.342	129.561	135.807	140.169

Tabla 7 La distribución  $D\chi^2$

**Referencias:**

Mendenhall, W. (2010). *Introducción a la Probabilidad y Estadística*. Cengage Learning Editores.

Hoel, P. G. (1984). *Elementary Statistics*. John Wiley & Sons.

Kelmansky, D. (2009). *Estadística para todos, Estrategias de pensamiento y herramientas para la solución de problemas*. Recuperado a partir de:

<http://www.bnm.me.gov.ar/giga1/documentos/EL001858.pdf>