

Para ampliar 4

Prueba de Independencia

La Prueba de Independencia o Test de Homogeneidad investiga si existe un buen grado de asociación entre dos variables que se estudian conjuntamente, o bien son independientes. Usa la distribución chi-cuadrado, que estudiarás más adelante.

La idea en la que se basa es muy simple, pero no la desarrollaremos aquí. Se considera que si dos variables son independientes, los valores de una no influirán en los de la otra, es decir, las probabilidades condicionadas serán siempre las mismas. Eso se traduce en la práctica en una proporcionalidad en las frecuencias de la tabla. Sobre esa idea se construyen unas frecuencias teóricas y se comparan con las reales. En los textos de Estadística podrás leer todo el desarrollo completo.

Volvemos al ejemplo de las faltas graves de los alumnos, según la tabla

Meses Niveles	Enero	Febrero	Marzo	Abril	Mayo
A	4	6	7	8	8
B	3	3	6	5	9
C	9	7	7	13	14

¿Que querría decir la afirmación de que *los factores nivel y mes son independientes*? Pues que la distribución de faltas (por ejemplo, en porcentaje) debería ser similar en todas las columnas, *independientemente* del mes en el que se tomen. Con más claridad, los porcentajes que obtuvimos

Meses					
Niveles	Enero	Febrero	Marzo	Abril	Mayo
A	12,1%	18,2%	21,2%	24,2%	24,2%
B	11,5%	11,5%	23,1%	19,2%	34,6%
C	18,0%	14,0%	14,0%	26,0%	28,0%

deberían haber sido los mismos independientemente del mes que consideremos.

Si esto fuera así, ***todas las frecuencias deberían ser proporcionales*** y cada una debería poderse calcular mediante proporciones, o la clásica regla de tres. Por ejemplo, la frecuencia del nivel A en el mes de Febrero debería poder calcularse como el producto del total de A por el total de Febrero, dividido después entre el total de todos los alumnos. Comprueba que resultaría esta tabla *teórica*.

Meses						
Niveles	Enero	Febrero	Marzo	Abril	Mayo	
A	4,84	4,84	6,06	7,87	9,39	33
B	3,82	3,82	4,77	6,20	7,39	26
C	7,34	7,34	9,17	11,93	14,22	50
	16	16	20	26	31	109

Comprueba algún valor: la frecuencia 6,06 ha resultado de multiplicar 33 por 20 y dividir después entre 109, la frecuencia 11,93 es el resultado de $50 \cdot 26 / 109$.

Todos estos cálculos los efectúa el modelo [homogen.ods](#). Copia los datos primitivos de los alumnos, sin totales ni porcentajes, en la zona de datos de este modelo.

Pasa a la hoja *Cálculos* y busca las ***frecuencias teóricas***. Comprueba que son las mismas que hemos obtenido nosotros por regla de tres.

Cálculo de chi-cuadrado									
						0	Hay columna	0	0
	4	6	7	8	8	5		1	33
	3	3	6	5	9	5	Columnas	1	26
	9	7	7	13	14	5	5	1	50
						0		0	0
						0		0	0
0	3	3	3	3	3				
Filas			3						
Hay fila									
0	1	1	1	1	1				
0	16	16	20	26	31				109
Frecuencias teóricas									
	0	0	0	0	0				
0	4,84	4,84	6,06	7,87	9,39				
0	3,82	3,82	4,77	6,2	7,39				
0	7,34	7,34	9,17	11,93	14,22				
0	0	0	0	0	0				
0	0	0	0	0	0				

Por último, consulta la hoja *Entrada de datos* en los que deberá figurar este resumen

Resultado de la prueba de homogeneidad		
Los datos forman un rango de	3	filas
	5	columnas
Sus grados de libertad son:	8	
Valor correspondiente de <u>chi-cuadrado</u> :	3,03	
Probabilidad correspondiente:	0,93	
Decisión al 5%	Muestras homogéneas	
Valor crítico al 5%	15,51	
Ídem al 1%	20,09	

Si trabajamos al nivel de significación del 5%, no podemos rechazar la homogeneidad de las frecuencias, no tenemos motivos para pensar que la influencia de la primavera es distinta en cada nivel. Es muy fácil que se presenten estas diferencias: tienen un 93% de posibilidades. Así que, aunque en la primavera se han incrementado las faltas, lo han hecho por igual en los tres niveles, de forma que las observadas se pueden deber al azar.

Prueba con este otro ejemplo:

La siguiente tabla representa la distribución de unos alumnos de Bachillerato en tres niveles de rendimiento en Física y Matemáticas. ¿Se observa independencia entre los resultados de una y otra asignatura?

Niveles en Matem.	Niveles en física		
Niveles en física	Alto	Medio	Bajo
Alto	8	10	2
Medio	7	23	5
Bajo	2	6	15

Copia las frecuencias en *homogen.ods* y observa los resultados: La probabilidad de estas frecuencias si la Física y las Matemáticas fueran independientes es prácticamente nula. **Las muestras no son homogéneas**, luego, como bien sabemos, estas dos asignaturas presentan una gran influencia.

Resultado de la prueba de homogeneidad	
Los datos forman un rango de	3 filas 3 columnas
Sus grados de libertad son:	4
Valor correspondiente de <u>chi-cuadrado</u> :	25,19
Probabilidad correspondiente:	0
Decisión al 5%	Muestras no homogéneas
Valor crítico al 5%	9,49
Ídem al 1%	13,28