

## 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



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.	Alto	Medio	Bajo
Niveles en física			
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%	<b>Muestras no homogéneas</b>
Valor crítico al 5%	9,49
Ídem al 1%	13,28