

Problemas de independencia

- Se seleccionó una muestra de 3000 naranjas de València. Cada naranja se clasificó según su color (claro, medio y oscuro) y se determinó su contenido de azúcar (dulce o no dulce). Los resultados fueron:

Color	Muy Dulce	No Dulce	Totales
Claro	1300	200	1500
Medio	500	500	1000
Oscuro	200	300	500
Totales	2000	1000	3000

Probar la hipótesis de que la dulzura y el color son independientes.

- Nos dan las notas de cierta asignatura de 3 grupos de alumnos A , B i C :

A	4.6	5.	5.1	5.6	4.6	5.	5.7	5.4	4.4	8.
B	4.6	3.4	5.3	4.	3.5	4.	5.	4.7	3.6	4.1
C	7.2	7.3	5.7	4.1	5.7	6.1	6.	7.8	7.	3.8

Los clasificamos según 2 criterios: por grupo y por nota teniendo en cuenta que: **Suspenso** significa una nota más pequeña que 5 (nota < 5) y **Aprobado** significa una nota entre 5 y 6 ($5 \leq 6$), **Notable** significa una nota mayor que 6 (nota > 6).

Hallar a partir del test χ^2 el p-valor para aceptar que los dos criterios son independientes

- Clasificamos N individuos según dos criterios. Cada criterio tiene dos niveles. La tabla de contingencia es la siguiente:

$C_2 \backslash C_1$	A_1	A_2
B_1	10	5
B_2	5	10

Hallar el p-valor para poder aceptar que los dos criterios son independientes usando el test χ^2 . 1. En un estudio se quiso contrastar si había relación entre el grupo sanguíneo de una persona y el hecho que sea o no portador de un cierto antígeno raro. Para hacerlo, se eligieron 150 portadores del antígeno y 500 no portadores y se les miró el grupo sanguíneo. Los resultados son los de la tabla siguiente:

Grupo	Portadores	No portadores
0	72	230
A	54	192
B	16	63
AB	8	15

Usar el test χ^2 sobre estos datos para contrastar si las frecuencias de los diferentes tipos sanguíneos son diferentes en los portadores y los no portadores.

- La tabla siguiente da la distribución de la población de los EE. UU. por edades

Edades	Menos de 18	18-24	25-34	35-44	45-64	65-79	80 o más
Porcentaje %	25	9.6	13.4	14.4	25.4	9.4	2.8

Hemos estudiado una muestra de 130 habitantes de los EE. UU. sin seguro médico, y los resultados han sido los siguientes:

Edades	Menos de 18	18-24	25-34	35-44	45-64	65-79	80 o más
Frecuencia	30	14	18	22	40	5	1

Usar el test χ^2 para contrastar si hay evidencia con nivel de significación 0.01 que la distribución por edades de los habitantes de los EE. UU. sin seguro médico es diferente de la distribución por edades de todos los habitantes de los EE. UU.