ASOCIACIÓN (Tablas de Contingencia)

A través de este contraste pretendemos probar si existe independencia/asoción entre dos variables o atributos (en el conjunto de la población) a partir de las observaciones de las dos característica (en una muestra).

Las frecuencias observadas las podemos disponer en una tabla de contingencia:

X\Y	y_1	y_2	y_j	y_m	
x_1	$n_{1,1}$	$n_{1,2}$	·	·	$\mathbf{n_{1,*}}$
x_2	$n_{2,1}$	$n_{2,2}$	·	·	$\mathbf{n_{2,*}}$
x_i	·	·	$n_{i,j}$	·	$\mathbf{n_{i,*}}$
·	·	·	·	·	
x_n	·	·	·	$n_{n,m}$	$\mathbf{n_{n,*}}$
	$\mathbf{n_{*,1}}$	$\mathbf{n_{*,2}}$	$\mathbf{n_{*,j}}$	$\mathbf{n_{*,m}}$	**N**

Donde : $n_{i,j}$ es la frecuencia conjunta

$n_{i,*}$ es la frecuencia marginal de x

$n_{*,j}$ es la frecuencia marginal de y

Si la hipótesis de independencia se cumple, y por el teorema de caracterización , se deberá cumplir que todas las frecuencias relativas conjuntas sean iguales al producto de las respectivas frecuencias relativas marginales:

$$\frac{n_{i,j}}{N} = \frac{n_{i,\bullet}}{N} \bullet \frac{n_{\bullet,j}}{N} \quad \forall i, \forall j$$

luego en el caso de independencia cada una de las ij frecuencias conjuntas teóricas serán :

$$n_{i,j}^{T} = \frac{n_{i,\bullet} \bullet n_{\bullet,j}}{N}$$

si establecemos el mismo método del test de la chi-2 crearemos el estadístico

$$\chi^2 = \sum_{\forall i} \sum_{\forall j} \frac{\left(n_{i,j} - n_{i,j}^{T}\right)^2}{n_{i,j}^{T}}$$

Se pueden establecer tres coeficientes:

PERSON $$C = \sqrt{\frac{\chi^2}{\chi^2+N}}$$

si tabla es de 2x2 entre 0 y 1 cero independencia 1 máxima asociación

CRAMER

$$V = \sqrt{\frac{\chi^2}{N(A-1)}} \qquad siendo\ A\ el\ mínimo\ de\ i\ o\ j$$

i número de valores distintos de x, j número de valores distintos de y

su valor está entre 0 y 1 cero independencia 1 máxima asociación

TSCHUPROW

$$T = \sqrt{\frac{\chi^2}{N\sqrt{(i-1)(j-1)}}}$$

i número de valores distintos de x, j número de valores distintos de y

su valor está entre 0 y 1 cero independencia 1 máxima asociación

ejemplo 1

Se dispone de las observaciones del color del pelo y de los ojos de 400 individuos según la siguiente tabla:

	ojos azules	ojos negros	ojos pardos	
rubios	120	20	20	160
castaños	50	30	60	140
morenos	50	10	40	100
	220	60	120	400

Construyamos primero la tabla de frecuencias teóricas: aplicando para cada valor la expresión

$$n_{i,j}^{T} = \frac{n_{i,\cdot} \cdot n_{\cdot,j}}{N} \qquad \text{ejemplo : } 88 = \frac{160 \cdot 220}{400}$$

construimos la tabla de contingencia de frecuencias teóricas

	ojos azules	ojos negros	ojos pardos	
rubios	88	24	48	160
castaños	77	21	42	140
morenos	55	15	30	100
	220	60	120	400

Construimos el estadístico

$$\chi^2 = \sum_{\forall i} \sum_{\forall j} \frac{\left(n_{i,j} - n_{i,j}^{T}\right)^2}{n_{i,j}^{T}}$$

que tomará el valor 55,13

$\frac{(120-88)^2}{88} = 11{,}6363$	$\frac{(20-24)^2}{24} = 0{,}666$	$\frac{(20-48)^2}{48} = 16{,}33$
$\frac{(50-77)^2}{77} = 9{,}4675$	$\frac{(30-21)^2}{21} = 3{,}8571$	$\frac{(60-42)^2}{42} = 7{,}7142$
$\frac{(50-55)^2}{55} = 0{,}4545$	$\frac{(10-15)^2}{15} = 1{,}666$	$\frac{(40-30)^2}{30} = 3{,}333$

Pearson

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{55{,}13}{55{,}13 + 400}} = 0{,}348$$

Cramer

$$V = \sqrt{\frac{\chi^2}{N(\ A - 1)}} = \sqrt{\frac{55{,}13}{400(\ 3 - 1)}} = 0{,}262511$$

nivel bajo menor que 0,3

Tschuprow

$$T = \sqrt{\frac{\chi^2}{N\sqrt{(i-1)(j-1)}}} = \sqrt{\frac{55{,}13}{400\sqrt{(3-1)(3-1)}}} = 0{,}262511$$

ejemplo 2

Para intentar mejorar la posición en el mercado de cierto producto se llevaron a cabo tres campañas de promoción entre los minoristas distribuidores en otras tantas localidades : A,B,C. Se desea comprobar si las tres campañas son indistintas o influyen en el incremento de las ventas en las tiendas .Para ello se han recogido los siguientes datos:

	localidad A	localidad B	localidad C	$n_{i,\bullet}$
tiendas aumentan ventas	165	141	152	458
tiendas no aumentan ventas	256	142	98	496
$n_{\bullet,j}$	421	283	250	954

si calculamos las frecuencias teóricas mediante $n_{i,j}^{T} = \frac{n_{i,\bullet} \cdot n_{\bullet,j}}{N}$ tendremos :

	localidad A	localidad B	localidad C	$n_{i,\bullet}$
tiendas aumentan ventas	202,1153	135,86373	120,02096	458
tiendas no aumentan ventas	218,8847	147,13626	129,97903	496
$n_{\bullet,j}$	421	283	250	954

aplicando la expresión del estadístico

$$\chi^2 = \sum_{\forall i} \sum_{\forall j} \frac{\left(n_{i,j} - n_{i,j}^{T}\right)^2}{n_{i,j}^{T}}$$

$$\chi^2 = 29{,}8711$$

Pearson

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} = \sqrt{\frac{29{,}8711}{29{,}8711 + 954}} = 0{,}1742$$

Cramer

$$V = \sqrt{\frac{\chi^2}{N(\, A - 1)}} = \sqrt{\frac{29{,}8711}{954(\, 2 - 1)}} = 0{,}1769$$

nivel bajo menor que 0,3

Tschuprow

$$T = \sqrt{\frac{\chi^2}{N\sqrt{(i-1)(j-1)}}}$$

$$= \sqrt{\frac{29{,}8711}{954\sqrt{(2-1)(3-1)}}} = 0{,}1487$$

ejemplo 3

En nuestra empresa tenemos organizados tres turnos de producción .Hemos recogido información aleatoria sobre las cantidades de piezas que se producen por turno a lo largo de 100 días , resultando la siguiente información recogida en la tabla.

Turno/producción	mañana	tarde	noche
100-200 piezas	12	12	13
200-300 piezas	7	12	12
300-400 piezas	10	8	14

comprobar mediante el coeficiente de asociación si existe relación entre la producción y el turno

Estableceremos la siguiente tabla donde aparecen las frecuencias conjuntas observadas en la muestra y las que teóricamente corresponderían al caso de homogeneidad. Así

Turno/producción	mañana		tardes		noche		
	obs	teo	obs	teo	obs	teo	
100-200 piezas	12	10,73	12	11,84	13	14,43	37
		0,15		**0,0021**		**0,14**	
200-300 piezas	7	8,99	12	9,92	12	12,09	31
		0,44		**0,436**		**0,0006**	
300-400 piezas	10	9,28	8	10,24	14	12,48	32
		0,055		**0,49**		**0,185**	
	29		32		39		100

$$n^{T}_{i,j} = \frac{n_{\bullet j} \bullet n_{i\bullet}}{N} \rightarrow n^{T}_{1,1} = \frac{n_{\bullet 1} \bullet n_{1\bullet}}{N} = \frac{29 \bullet 37}{100} = 10,73$$

y de la misma manera para todas las frecuencias conjuntas teóricas cada uno de estas celdas sería

$$\frac{\left(n_{o,i}-n_{t,i}\right)^2}{n_{t,i}}$$

en este caso

$$\frac{\left(7-8{,}99\right)^2}{8{,}99} = 0{,}44$$

y así para todas las diferencias entre observada y teóricas y siendo el estadístico :

$$\chi^2 = \sum_{\forall i}\sum_{\forall j} \frac{\left(n_{o,i}-n_{t,i}\right)^2}{n_{t,i}}$$

Cuyo valor sería =0,15+0,0021+0,14+0,44+0,43+0+0,05+0,49+0,18=**1,9**

Pearson

$$C = \sqrt{\frac{\chi^2}{\chi^2+N}} = \sqrt{\frac{1{,}9}{1{,}9+100}} = 0{,}1365$$

Cramer

$$V = \sqrt{\frac{\chi^2}{N(\,A-1)}} = \sqrt{\frac{1{,}9}{100(\,3-1)}} = 0{,}0974$$

nivel bajo menor que 0,3

Tschuprow

$$T = \sqrt{\frac{\chi^2}{N\sqrt{(i-1)(j-1)}}}$$

$$= \sqrt{\frac{1{,}9}{100\sqrt{(3-1)(3-1)}}} = 0{,}0974$$

Coeficiente correlación de SPEARMAN , para variables ordinales

mide la relación entre dos variables númericas ordenables

Su expresión es $\rho = 1 - \frac{6\sum D^2}{N(N^2-1)}$

donde D es la diferencia de orden entre X e Y siendo N es total de pares de valores

su rango como el coeficiente de coróelacion está entre -1 y 1

ejemplo 1

X	Y
22	12
12	11
15	9
7	14

orden que ocupan los valores

X	Y	D	D2
4	3	1	1
2	2	0	0
3	1	2	4
1	4	-3	9

por tanto $\rho = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6\cdot 14}{4(4^2-1)} =$

$$= 1 - 1{,}4 = -0{,}4$$

ejemplo 2

X es el coeficiente intelectual e Y el número "WhatsApp" a la hora

X (C.I)	**Y (Wapp)**
77	**40**
80	**35**
80	**32**
80	**32**
110	**12**
111	**15**
123	**10**

órdenes

x	**y**	**Ordenx**	**Orden y**	**D**	**D2**
77	**40**	**1**	**7**	**-6**	**36**
80	**35**	**3**	**6**	**-3**	**9**
80	**32**	**3**	**4,5**	**-1,5**	**2,25**
80	**32**	**3**	**4,5**	**-1,5**	**2,25**
110	**12**	**5**	**2**	**3**	**9**
111	**15**	**6**	**3**	**3**	**9**
123	**10**	**7**	**1**	**6**	**36**
					103,5

por tanto $\rho = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6\cdot 103{,}5}{7(7^2-1)} =$

$$= 1 - 1{,}8482 = -0{,}8482$$

por tanto muy alta y negativa