

6. 4. Medidas de asociación entre atributos ordinales.

A veces, se plantean problemas que exigen medir el grado de interdependencia entre dos características, **A** y **B**, que no pueden cuantificarse pero sí ordenarse de menor a mayor, en base a algún criterio.

A esta ordenación se le denomina “clasificación por rangos” y el número ordinal que corresponde a cada observación, por el lugar que ocupa en la clasificación se denomina “rango”.

6.4.1. Coeficiente de correlación por rangos de Spearman.

El coeficiente de correlación lineal de Pearson entre dos sucesiones ordinales da lugar a una nueva expresión de dicho coeficiente que se denomina **Coeficiente de correlación por rangos de Spearman**. Así, la **Correlación por Rangos de Spearman** mide la asociación entre atributos ordinales, y permitirá decir si ambos criterios son concordantes, discordantes o independientes.

Aunque las modalidades del atributo **A** sean $A_i, i=1,...,k$, y las del atributo **B** sean $B_j, j=1,...,m$; para fijar ideas supongamos que $k=m=N$. Con lo que las N observaciones bidimensionales obtenidas pueden resumirse en la *tabla con elección de modalidades* siguiente:

A	B
A_1	B_1
A_2	B_2
..	.
.	.
.	.
A_N	B_N

Si a las $k=N$ modalidades diferentes del atributo **A** les asociamos los rangos $x_i=1,2,...,k$; y a las $m=N$ modalidades diferentes del atributo **B**, les asociamos los rangos $y_j=1,2,...,m$; y hacemos $d_i=x_i-y_j$, tendremos:

A	B	$d_i = x_i - y_i$
x_1	y_1	d_1
x_2	y_2	d_2
..	.	.
.	.	.
.	.	.
x_N	y_N	d_N

Con lo que:

$$\bar{X} = \frac{\sum x_i}{N} = \frac{1 + \dots + N}{N} = \frac{\frac{1+N}{2}N}{N} = \frac{1+N}{2}$$

$$\bar{Y} = \frac{\sum y_j}{N} = \frac{1 + \dots + N}{N} = \frac{\frac{1+N}{2}N}{N} = \frac{1+N}{2}$$

y el coeficiente de correlación lineal queda:

$$\rho_{XY} = r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} = 1 - \frac{\sum_i d_i^2}{N^3 - N}$$

Cuando las dos ordenaciones coinciden, es decir, la concordancia es perfecta, el coeficiente vale 1. Cuando la concordancia es inversa o equivalentemente, el ranking asociado a cada individuo en una de las ordenaciones será exactamente el opuesto al de la otra, el valor del coeficiente es -1. La expresión del coeficiente se obtiene en el caso de que no haya empates en cada una de las ordenaciones, o lo que es lo mismo a dos sujetos nunca les corresponda el mismo rango. Sin embargo, suele aplicarse aun cuando haya empates siempre que sean pocos, pero en este caso los límites no tienen por qué ser exactamente -1 y 1. Una práctica habitual es asignar a los individuos empatados la media o mediana de los rangos correlativos que les hubiesen correspondido en caso de no haberse producido tales empates.

6.4.2. Coeficiente Gamma, γ , de Goodman y Kruskal.

Cuando hay muchas observaciones y muy pocos valores alcanzables, habrá muchos empates; es el caso de las típicas tablas de contingencia. En ese caso, en lugar de aplicar el coeficiente de correlación por rangos de Spearman se recomienda aplicar el coeficiente γ **de Goodman y Kruskal**.

Este coeficiente hace referencia a la concordancia o discordancia entre los rangos de los factores para los individuos observados.

Un par de individuos se dice par concordante si los rangos de ambos factores siguen idéntica dirección (ambos crecen o ambos decrecen) en los dos individuos; mientras que se dice par discordante en caso contrario.

Sea P = nº de pares concordantes y

Sea Q = nº de pares discordantes.

Se define la medida de asociación γ **de Goodman y Kruskal** entre los dos atributos como el coeficiente:

$$\gamma = \frac{P-Q}{P+Q}; -1 \leq \gamma \leq 1$$

Veamos como funciona con un ejemplo. Se pretende estudiar la relación existente entre el *nivel de experiencia de un colectivo* y el *nivel de prohibición* que les parece adecuado sobre un determinado aspecto de su profesión. Para ello, hemos obtenido los siguientes datos:

		Nivel de Prohibición			Total
		Nulo	Con ciertas restricciones	Radicalmente	
Nivel de Experiencia	Baja	8 (A)	16 (B)	9 (C)	33
	Mediana	11 (D)	12 (E)	5 (F)	28
	Alta	51 (G)	46 (H)	32 (I)	129
Total		70	74	46	190

Obsérvese que cada celda forma *pares concordantes* con cada una de las demás que está situada por debajo y a su derecha (crece el nivel de experiencia y crece simultáneamente el nivel de prohibición):

A-E B-F D-H E-I

A-F B-I D-I

A-H

A-I

Obsérvese que cada celda forma *pares disconcordantes* con cada una de las demás que esté situada por debajo y a su izquierda

B-D C-E E-G F-H

B-G C-D F-G

C-H

C-G

Con lo cual, se tiene que:

$$P = 8 \cdot (12 + 5 + 46 + 32) + 16 \cdot (5 + 32) + 11 \cdot (46 + 32) + 12 \cdot 32 = 2594$$

$$Q = 16 \cdot (11 + 51) + 9 \cdot (12 + 11 + 51 + 46) + 12 \cdot 51 + 5 \cdot (46 + 51) = 3169$$

$$(P - Q) / (P + Q) = -0.099774423$$

La interpretación es la siguiente:

- Si los factores son independientes, entonces existirá el mismo nº de pares concordantes que dicordantes, luego $P=Q$, y por tanto, $\gamma=0$.
- Si existe asociación positiva perfecta, entonces todos los pares son concordates, luego $Q=0$, y por tanto $\gamma=1$.

- Si existe asociación negativa perfecta, entonces todos los pares son discordantes, luego $P=0$, y por tanto $\gamma=-1$.
- Mientras más cerca de cero, menos asociación, y cuanto más cerca de 1 (de -1) más asociación positiva (negativa).