

Resumen de los contenidos básicos

Tema 6. Dependencia entre atributos



Carmen Trueba Salas
Lorena Remuzgo Pérez

DEPARTAMENTO DE ECONOMÍA

Este tema se publica bajo Licencia:
[Creative Commons BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)



Contenidos

- 6.1 Tabla de contingencia
- 6.2 Atributos en escala nominal
- 6.3 Atributos en escala ordinal

Sean A y B dos atributos cuyas modalidades son $a_1, \dots, a_i, \dots, a_k$ y $b_1, \dots, b_j, \dots, b_h$, respectivamente. La observación conjunta de ambas atributos proporciona pares de modalidades (a_i, b_j) .

6.1 Tabla de contingencia

Permite presentar la información relativa a los atributos A y B de manera ordenada.

$a_i \setminus b_j$	b_1	...	b_j	...	b_h	$n_{i.}$
a_1	n_{11}	...	n_{1j}	...	n_{1h}	$n_{1.} = n_{11} + \dots + n_{1h}$
a_2	n_{21}	...	n_{2j}	...	n_{2h}	$n_{2.} = n_{21} + \dots + n_{2h}$
\vdots	\vdots		\vdots		\vdots	\vdots
a_i	n_{i1}	...	n_{ij}	...	n_{ih}	$n_{i.} = n_{i1} + \dots + n_{ih}$
\vdots	\vdots		\vdots		\vdots	\vdots
a_k	n_{k1}	...	n_{kj}	...	n_{kh}	$n_{k.} = n_{k1} + \dots + n_{kh}$
$n_{.j}$	$n_{.1} = n_{11} + \dots + n_{k1}$...	$n_{.j} = n_{1j} + \dots + n_{kj}$...	$n_{.h} = n_{1h} + \dots + n_{kh}$	$\sum_{i=1}^k \sum_{j=1}^h n_{ij} = N$

La distribución de frecuencias bidimensional de (A, B) viene dada por el conjunto de pares de modalidades junto con sus correspondientes frecuencias absolutas conjuntas $(a_i, b_j; n_{ij})$ o frecuencias relativas conjuntas $(a_i, b_j; f_{ij})$.

6.2 Atributos en escala nominal

Sea la distribución de frecuencias $(a_i, b_j; n_{ij})$.

La dimensión de la tabla de contingencia determina las medidas que se pueden utilizar para estudiar la dependencia entre atributos.

Atributos con dos modalidades cada uno (tablas 2 x 2)

- Coeficiente básico de dependencia
- Coeficiente de asociación Q de Yule

Atributos con k y h modalidades (tablas $k \times h$)

- Estadístico chi-cuadrado
- Coeficiente de contingencia
- Coeficiente V de Cramer
- Coeficiente T de Tschuprow

6.2 Atributos en escala nominal

Coeficiente básico de dependencia

$$D = \frac{n_{11}n_{22} - n_{12}n_{21}}{N} = n_{11} - \frac{n_{1.} \times n_{.1}}{N}$$

Interpretación

Si $D = 0$, los atributos A y B son independientes.

Si $D > 0$, existe asociación positiva entre las modalidades a_1 y b_1 .

Si $D < 0$, existe asociación negativa entre las modalidades a_1 y b_1 .

6.2 Atributos en escala nominal

Coeficiente de asociación Q de Yule

$$Q = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}} = \frac{N \cdot D}{n_{11}n_{22} + n_{12}n_{21}} \quad -1 \leq Q \leq 1$$

Interpretación

Si $Q = 0$, los atributos A y B son independientes.

Si $Q = 1$, existe asociación positiva completa entre las modalidades a_1 y b_1 .

Si $Q = -1$, existe asociación negativa completa entre las modalidades a_1 y b_1 .

Si $0 < Q < 1$, existe asociación positiva entre las modalidades a_1 y b_1 .

Si $0 < Q < -1$, existe asociación negativa entre las modalidades a_1 y b_1 .

6.2 Atributos en escala nominal

Estadístico chi-cuadrado

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^h \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad \chi^2 \geq 0$$

donde $e_{ij} = \frac{n_{i.} \cdot n_{.j}}{N}$ es la frecuencia absoluta esperada.

Interpretación

Si $\chi^2 = 0$, los atributos A y B son independientes.

Si $\chi^2 > 0$, los atributos A y B no son independientes.

6.2 Atributos en escala nominal

Coefficiente de contingencia

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} \quad 0 \leq C \leq 1$$

Interpretación:

Si $C = 0$, los atributos A y B son independientes.

Si $C = 1$, existe asociación completa entre los atributos A y B .

Si $0 < C < 1$, existe asociación entre los atributos A y B .

6.2 Atributos en escala nominal

Coeficiente V de Cramer

$$V = \sqrt{\frac{\chi^2}{N[\min(k, h) - 1]}} \quad 0 \leq V \leq 1$$

Interpretación

Si $V = 0$, los atributos A y B son independientes.

Si $V = 1$, existe asociación completa entre los atributos A y B .

Si $0 < V < 1$, existe asociación entre los atributos A y B .

6.2 Atributos en escala nominal

Coeficiente T de Tschuprow

$$T = \frac{\chi^2}{N\sqrt{(k-1)(h-1)}} \quad 0 \leq T \leq 1$$

Interpretación

Si $T = 0$, los atributos A y B son independientes.

Si $T = 1$, existe asociación completa entre los atributos A y B .

Si $0 < T < 1$, existe asociación entre los atributos A y B .

6.3 Atributos en escala ordinal

Sea la distribución de frecuencias $(a_i, b_j; n_{ij})$.

Es posible estudiar la concordancia o discordancia entre las ordenaciones de los individuos de la población atendiendo a los atributos A y B .

Coefficientes de correlación por rangos

- Coeficiente de correlación de Spearman
- Coeficiente de correlación τ de Kendall

6.3 Atributos en escala ordinal

Coefficiente de correlación de Spearman

$$\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{N(N^2 - 1)} \quad -1 \leq \rho \leq 1$$

donde $d_i = r_i^1 - r_i^2$.

Interpretación

Si $\rho = 0$, los criterios de ordenación son independientes.

Si $\rho = 1$, existe concordancia perfecta entre las ordenaciones.

Si $\rho = -1$, existe discordancia perfecta entre las ordenaciones.

Si $0 < \rho < 1$, existe concordancia entre las ordenaciones.

Si $-1 < \rho < 0$, existe discordancia entre las ordenaciones.

6.3 Atributos en escala ordinal

Coeficiente de correlación τ de Kendall

$$\tau = \frac{C - D}{N(N-1)/2} \quad -1 \leq \tau \leq 1$$

donde C es el número de concordancias y D es el número de discordancias.

Interpretación

Si $\tau = 0$, los criterios de ordenación son independientes.

Si $\tau = 1$, existe concordancia perfecta entre las ordenaciones.

Si $\tau = -1$, existe discordancia perfecta entre las ordenaciones.

Si $0 < \tau < 1$, existe concordancia entre las ordenaciones.

Si $-1 < \tau < 0$, existe discordancia entre las ordenaciones.