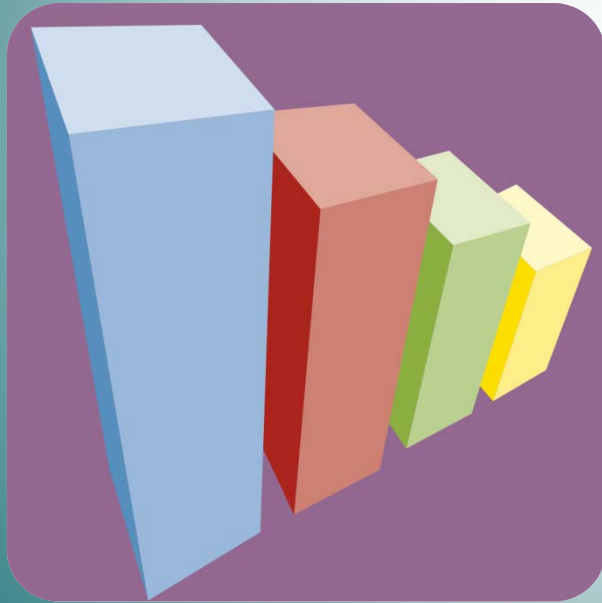


# Estadística I

## Tema 6. Variables cualitativas o atributos



**Carmen Trueba Salas**  
**Lorena Remuzgo Pérez**  
**Vanesa Jordá Gil**  
**José María Sarabia Alegría**

DPTO. DE ECONOMÍA

Este tema se publica bajo Licencia:

[Creative Commons BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)



# Introducción

- Las variables cualitativas o atributos se caracterizan por no tomar valores numéricos.
- Estas variables describen características que se denominan como modalidades o categorías.

## Clasificación

- Atributos en escala nominal: no es posible establecer una jerarquía entre sus modalidades
- Atributos en escala ordinal: Es posible establecer alguna jerarquía entre sus modalidades.

# Tabla de doble entrada

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

→ Frecuencia marginal de A

↓  
Frecuencia marginal de B

$n_{ij}$  : Frecuencia absoluta conjunta

# Coeficiente básico de dependencia

$$D = \frac{n_{11}n_{22} - n_{12}n_{21}}{N},$$

## Interpretación

- Si  $D = 0$  no existe asociación entre los atributos.
- Si  $D > 0$  existe asociación positiva entre los atributos.
- Si  $D < 0$  existe asociación negativa entre los atributos.

# Coeficiente de asociación Q de Yule

$$Q = \frac{n_{11}n_{22} - n_{12}n_{21}}{n_{11}n_{22} + n_{12}n_{21}},$$

## Interpretación

- Si  $Q=0$  no existe asociación entre los atributos.
- Si  $Q=1$  existe asociación positiva perfecta entre los atributos.
- Si  $Q=-1$  existe asociación negativa perfecta entre los atributos.
- Si  $0 < Q < 1$  existe asociación positiva (mayor grado de asociación cuanto mayor es el coeficiente).
- Si  $-1 < Q < 0$  existe asociación negativa entre las modalidades (mayor grado de asociación cuanto menor es el coeficiente).

# Estadístico chi-cuadrado

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

donde  $n_{ij}$  es la frecuencia absoluta conjunta y  $e_{ij}$  es la frecuencia esperada:

$$e_{ij} = \frac{n_{i.} \times n_{.j}}{N},$$

## Interpretación

- Si  $\chi^2 = 0$ , no existe asociación entre los atributos.
- Si  $\chi^2 > 0$ , existe asociación entre los atributos.

# Coeficiente de contingencia

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

## Interpretación

- Si  $C=0$  no existe asociación entre los atributos.
- Si  $C=1$  indica asociación completa entre los atributos.
- Si  $0 < C < 1$  cuanto mayor es el coeficiente mayor es la asociación entre los atributos.

# Coeficiente V de Cramer

$$V = \sqrt{\frac{\chi^2}{N[\text{mín}(k, h) - 1]}}$$

## Interpretación

- Si  $V=0$  no existe asociación entre los atributos.
- Si  $V=1$  asociación completa entre los atributos.
- Si  $0 < V < 1$  cuanto mayor es el coeficiente mayor es el grado de asociación entre los atributos.



# Coeficiente T de Tschuprow

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

## Interpretación

- Si  $T=0$  no existe asociación entre los atributos.
- Si  $T=1$  asociación completa entre los atributos.
- Si  $0 < T < 1$  cuanto mayor es el coeficiente mayor es la asociación entre los atributos.

# Coeficiente de correlación por rangos de Spearman

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

## Interpretación

- Si  $\rho=0$  los criterios de clasificación son independientes.
- Si  $\rho =1$  existe concordancia perfecta entre los rangos.
- Si  $\rho =-1$  existe discordancia perfecta entre los rangos.
- Si  $0 < \rho < 1$  cuanto mayor es el coeficiente mayor es la concordancia entre los rangos.
- Si  $-1 < \rho < 0$  cuanto menor es el coeficiente mayor es la discordancia entre los rangos.

# Coeficiente de correlación $\tau$ de Kendall

$$\tau = \frac{CC - DC}{N(N-1)/2}.$$

## Interpretación

- Si  $\tau=0$  los criterios de clasificación son independientes.
- Si  $\tau=1$  existe concordancia perfecta entre los rangos.
- Si  $\tau=-1$  existe discordancia perfecta entre los rangos.
- Si  $0 < \tau < 1$  cuanto mayor es el coeficiente mayor es la concordancia entre los rangos.
- Si  $-1 < \tau < 0$  cuanto menor es el coeficiente mayor es la discordancia entre los rangos.