

Tema 8: Relación entre variables

Estadística I

Universidad de Salamanca

Curso 2010/2011

Outline

- 1 Independencia
- 2 Medidas de asociación
- 3 La covarianza y el coeficiente de correlación
 - La covarianza
 - El coeficiente de correlación lineal
 - Matriz de varianzas-covarianzas
 - Matriz de correlaciones

Independencia

Definición

Diremos que dos variables son estadísticamente independientes si

$$f_{ij} = f_{i.} = f_{.j} \Leftrightarrow n_{ij} = \frac{n_{i.} \cdot n_{.j}}{N}$$

Propiedades

- Si dos variables son independientes $\Rightarrow S_{XY} = 0$

Medidas de asociación

Coeficiente de χ^2 de Pearson

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^p \left[\frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{N} \right)^2}{\frac{n_{i.} n_{.j}}{N}} \right]$$

Interpretación

- Si $\chi^2 = 0 \Rightarrow$ Independencia
- A mayor $\chi^2 \Rightarrow$ Mayor asociación

Outline

- 1 Independencia
- 2 Medidas de asociación
- 3 **La covarianza y el coeficiente de correlación**
 - **La covarianza**
 - El coeficiente de correlación lineal
 - Matriz de varianzas-covarianzas
 - Matriz de correlaciones

Covarianza entre dos variables

Definición

$$S_{XY} = \text{Cov}(X, Y) = \frac{\sum_{i=1}^k \sum_{j=1}^p (x_i - \bar{X})(y_j - \bar{Y})n_{ij}}{N} =$$
$$\sum_{i=1}^k \sum_{j=1}^p (x_i - \bar{X})(y_j - \bar{Y})f_{ij} = \frac{\sum_{i=1}^N (x_i - \bar{X})(y_i - \bar{Y})}{N}$$

Propiedades

- Mide la relación lineal entre X e Y
 - Si $S_{XY} > 0$ Existe relación lineal positiva
 - Si $S_{XY} = 0$ No existe relación
 - Si $S_{XY} < 0$ Existe relación lineal inversa

Covarianza entre dos variables

Propiedades

- $S_{XY} = S_{YX}$
- $S_{XX} = S_X^2$
- $Cov(aX, cY) = ac Cov(X, Y)$
- $Cov(b + X, Y) = Cov(X, Y)$
- $Cov(b + aX, d + cY) = ac Cov(X, Y)$
- Fórmula práctica

$$S_{XY} = \frac{\sum_{i=1}^k \sum_{j=1}^p x_i y_j n_{ij}}{N} - \overline{XY}$$

Outline

- 1 Independencia
- 2 Medidas de asociación
- 3 La covarianza y el coeficiente de correlación**
 - La covarianza
 - El coeficiente de correlación lineal**
 - Matriz de varianzas-covarianzas
 - Matriz de correlaciones

El coeficiente de correlación lineal

Definición

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}}$$

Propiedades

- Mide el grado de relación lineal entre X e Y sin eliminar la influencia del resto de las variables
 - $r_{XY} > 0 \Rightarrow S_{XY} > 0$ Relación lineal positiva. Existe correlación positiva
 - $r_{XY} = 0 \Leftrightarrow S_{XY} = 0$ No existe relación lineal. Las variables están incorreladas
 - $r_{XY} < 0 \Rightarrow S_{XY} < 0$ Relación lineal negativa. Existe correlación negativa
 - $r_{XY} = 1$ Correlación positiva máxima
 - $r_{XY} = -1$ Correlación negativa máxima

El coeficiente de correlación lineal

Propiedades

- $-1 \leq r_{XY} \leq 1$
- Si

$$u = aX + b$$

$$v = cY + d$$

entonces

$$r_{uv} = r_{XY}$$

Outline

- 1 Independencia
- 2 Medidas de asociación
- 3 La covarianza y el coeficiente de correlación**
 - La covarianza
 - El coeficiente de correlación lineal
 - Matriz de varianzas-covarianzas**
 - Matriz de correlaciones

Matriz de varianzas-covarianzas

Definición

$$S = \begin{pmatrix} S_X^2 & S_{XY} \\ S_{YX} & S_Y^2 \end{pmatrix}$$

Propiedades

- En la diagonal principal aparecen las varianzas
- En las diagonales secundarias aparecen las covarianzas
- Es cuadrada 2x2
- Es simétrica
- Es definida positiva

Outline

- 1 Independencia
- 2 Medidas de asociación
- 3 La covarianza y el coeficiente de correlación**
 - La covarianza
 - El coeficiente de correlación lineal
 - Matriz de varianzas-covarianzas
 - Matriz de correlaciones**

Matriz de correlaciones

Definición

$$S = \begin{pmatrix} 1 & r_{XY} \\ r_{YX} & 1 \end{pmatrix}$$

Propiedades

- Es cuadrada 2x2
- Es simétrica
- Es definida positiva