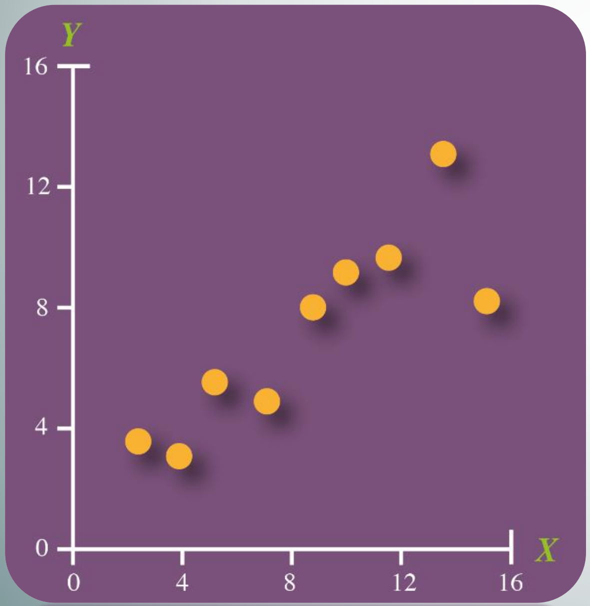
Estadística
Grado en Relaciones Laborales

Resumen de los contenidos básicos

Tema 4. Distribuciones de frecuencias bidimensionales



Lorena Remuzgo Pérez
Carmen Trueba Salas
DEPARTAMENTO DE ECONOMÍA

Contenidos

Sean X e Y dos variables cuyos valores son $x_1, \ldots, x_i, \ldots, x_k$ e $y_1, \ldots, y_j, \ldots, y_h$, respectivamente. La observación conjunta de ambas variables proporciona pares de observaciones (x_i, y_j).

4.1 Tabla de correlación

Permite presentar la información relativa a las variables X e Y de manera ordenada.

$x_i \setminus y_j$	y_1	...	y_j	...	y_h	$n_{i.}$
x_1	n_{11}	...	n_{1j}	...	n_{1h}	$n_{1.} = n_{11} + \ldots + n_{1h}$
x_2	n_{21}	...	n_{2j}	...	n_{2h}	$n_{2.} = n_{21} + \ldots + n_{2h}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_i	n_{i1}	...	n_{ij}	...	n_{ih}	$n_{i.} = n_{i1} + \ldots + n_{ih}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
x_k	n_{k1}	...	n_{kj}	...	n_{kh}	$n_{k.} = n_{k1} + \ldots + n_{kh}$
$n_{.j}$	$n_{.1} = n_{11} + \ldots + n_{k1}$	...	$n_{.j} = n_{1j} + \ldots + n_{kj}$	...	$n_{.h} = n_{1h} + \ldots + n_{kh}$	$\sum_{i=1}^{k} \sum_{j=1}^{h} n_{ij} = N$

La distribución de frecuencias bidimensional de (X, Y) viene dada por el conjunto de pares de valores junto con sus correspondientes frecuencias absolutas conjuntas $(x_i, y_j; n_{ij})$ o frecuencias relativas conjuntas $(x_i, y_j; f_{ij})$.

4.2 Distribuciones marginales

Permiten estudiar el comportamiento de cada variable de forma aislada.

Distribución marginal de *X*

x_i	$n_{i.}$
x_1	$n_{1.}$
x_2	$n_{2.}$
$\vdots$	$\vdots$
x_i	$n_{i.}$
$\vdots$	$\vdots$
x_k	$n_{k.}$
	N

Distribución marginal de *Y*

y_j	$n_{.j}$
y_1	$n_{.1}$
y_2	$n_{.2}$
$\vdots$	$\vdots$
y_j	$n_{.j}$
$\vdots$	$\vdots$
y_h	$n_{.k}$
	N

4.3 Distribuciones condicionadas

Permiten estudiar el comportamiento de cada variable cuando la otra permanece constante.

Distribución de X condicionada por $Y = y_j$

$x_i \mid Y = y_j$	$n_{i\mid j}$
x_1	n_{1j}
x_2	n_{2j}
$\vdots$	$\vdots$
x_i	n_{ij}
$\vdots$	$\vdots$
x_k	n_{kj}
	$n_{.j}$

Distribución de Y condicionada por $X = x_i$

$y_j \mid X = x_i$	$n_{j\mid i}$
y_1	n_{i1}
y_2	n_{i2}
$\vdots$	$\vdots$
y_j	n_{ij}
$\vdots$	$\vdots$
y_h	n_{ik}
	$n_{i.}$

Sea la distribución de frecuencias bidimensional (x_i, y_j; n_{ij}).

4.4 Independencia estadística

Dos variables son estadísticamente independientes cuando la variación de una de ellas no influye sobre la variación de la otra.

Condición de independencia estadística

$$n_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{N} \qquad \forall\, i, j$$

4.5 Relación lineal o correlación

Covarianza

Proporciona información acerca de la existencia de relación lineal entre las variables *X* e *Y.*

$$S_{XY} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{h}(x_i - \overline{x})(y_j - \overline{y})n_{ij} = \frac{1}{N}\sum_{i=1}^{k}x_i\sum_{j=1}^{h}y_j n_{ij} - \overline{xy}$$

Interpretación

Si $S_{XY} = 0$, no existe relación lineal entre *X* e *Y*.

Si $S_{XY} > 0$, existe relación lineal positiva entre *X* e *Y*.

Si $S_{XY} < 0$, existe relación lineal negativa entre *X* e *Y*.

4.5 Relación lineal o correlación

Coeficiente de correlación lineal

Proporciona información acerca del grado de relación lineal existente entre las variables *X* e *Y*.

$$r_{XY} = \frac{S_{XY}}{S_X \cdot S_Y} \qquad -1 \leq r_{XY} \leq 1$$

Interpretación

Si $r_{XY} = 0$, no existe relación lineal entre *X* e *Y*.

Si $r_{XY} = 1$, existe relación lineal perfecta positiva entre *X* e *Y*.

Si $r_{XY} = -1$, existe relación lineal perfecta negativa entre *X* e *Y*.

Si $0 < r_{XY} < 1$, existe relación lineal positiva entre *X* e *Y*.

Si $-1 < r_{XY} < 0$, existe relación lineal negativa entre *X* e *Y*.

4.6 Diagrama de dispersión o nube de puntos

Representación de los pares de observaciones (x_i, y_j) mediante puntos en el espacio bidimensional.