TEMA 2

ANÁLISIS DE DATOS MULTIDIMENSIONALES

EJEMPLOS

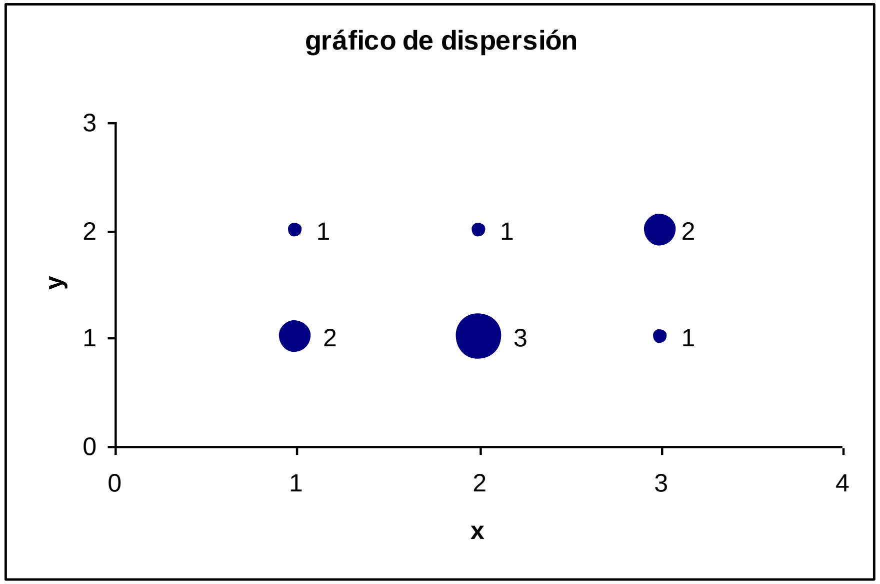
VNIVERSITAT DE VALÈNCIA
OpenCourseWare

Ejemplo 2.1. **CONJUNTA.** Ordenar la siguiente serie de datos bidimensionales en una distribución conjunta o distribución de frecuencias bidimensional (tabla de correlación):

X	1	1	2	2	3	3	1	2	3	2
Y	1	2	1	2	1	2	1	1	2	1

Solución:

X \ Y	1	2
1	*2*	*1*
2	*3*	*1*
3	*1*	*2*



Ejemplo 2.2. **MARGINALES**

Obtener las distribuciones de frecuencias marginales de X e Y a partir de la distribución conjunta del ejemplo 2.1.

X \ Y	1	2	$n_{i\cdot}$
1	*2*	*1*	**3**
2	*3*	*1*	**4**
3	*1*	*2*	**3**
$n_{\cdot j}$	**6**	**4**	N=10

El objetivo fundamental de estudiar conjuntamente dos variables (X, Y) sobre una misma población es cuantificar la posible relación entre ambas. Para ello se ha de definir el concepto de independencia estadística (no relación).

INDEPENDENCIA ESTADÍSTICA.

CARÁCTERIZACIÓN DE LA INDEPENDENCIA

X \ ***Y***	y_1	y_2		y_j		y_J	$n_{i\bullet}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1J}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2J}	$n_{2\bullet}$
x_i	n_{i1}	n_{i2}	...	$\boldsymbol{n_{ij}}$	...	n_{iJ}	$\boldsymbol{n_{i\bullet}}$
x_I	n_{I1}	n_{I2}	...	n_{Ij}	...	n_{IJ}	$n_{I\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	...	$\boldsymbol{n_{\bullet j}}$	...	$n_{\bullet J}$	$\boldsymbol{N}$

X e Y son INDEPENDIENTES si:

$$f_{ij} = f_{i\bullet} \times f_{\bullet j} \quad \forall i, j$$

es decir:

$$\frac{n_{ij}}{N} = \frac{n_{i\bullet}}{N} \times \frac{n_{\bullet j}}{N} \quad \forall i, j$$

Equivalente a:

$$n_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{N}$$

CONSIDERACIONES SOBRE LA INDEPENDENCIA.

- En general, una variable bidimensional (X, Y) (su distribución conjunta) se obtiene al medir simultáneamente las dos variables sobre los mismos elementos de una población. Si se miden (estudian) por separado las variables X e Y y se obtienen las distribuciones unidimensionales de X e Y (marginales), a partir de ellas no se puede construir la distribución conjunta de (X, Y).
- Pero si las variables X e Y son independientes, la caracterización de la independencia indica que lo conjunto se descompone en producto de lo marginal $f_{ij} = f_{i\bullet} \times f_{\bullet j} \quad \forall i, j$, por tanto:
- Si las variables X e Y son independientes, se puede obtener fácilmente la distribución conjunta a partir de las marginales: $f_{i\bullet} \times f_{\bullet j} = f_{ij} \quad \forall i, j$

Ejemplo 2.3. **EJEMPLO DE VARIABLES INDEPENDIENTES.**

Sea la siguiente distribución conjunta de dos variables (X, Y):

X \ Y	1	2
1	2	1
2	2	1
3	4	2

Compruébese que son independientes a partir de la caracterización de la independencia.

Solución:

Obsérvese que las columnas de frecuencias conjuntas son claramente proporcionales, al igual que las filas de frecuencias conjuntas.

POR LA CARACTERIZACIÓN DE LA INDEPENDENCIA:

Y / X	1	2	$n_{i\bullet}$
1	2	1	3
2	2	1	3
3	4	2	6
$n_{\bullet j}$	8	4	12

$$n_{ij} = \frac{n_{i\bullet} \times n_{\bullet j}}{N}$$

Por ejemplo:

$$n_{21} = \frac{n_{2\bullet} \times n_{\bullet 1}}{N}$$

$$2 = \frac{3 \times 8}{12}$$

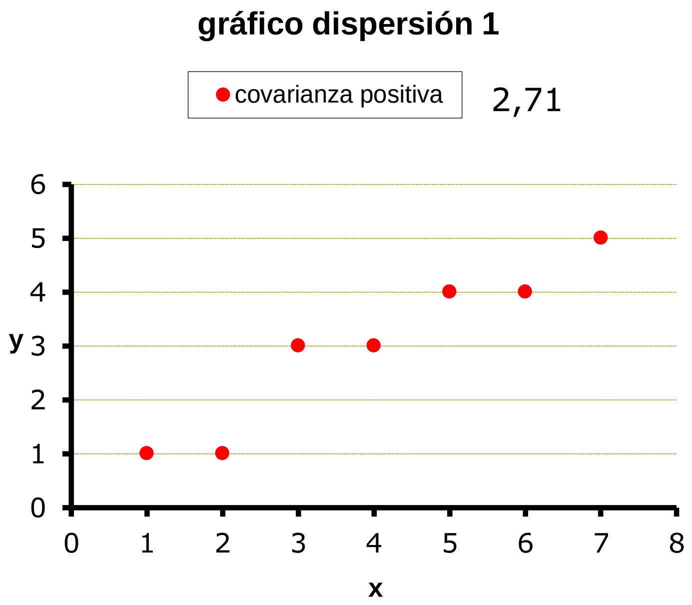
$4 = \dfrac{6 \times 8}{12}$ y así sucesivamente, lo cumplen todas las frecuencias conjuntas.

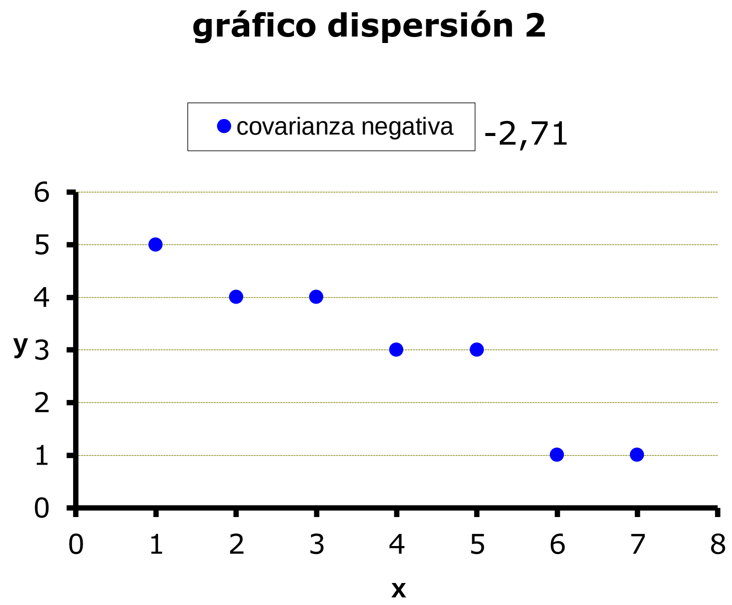
(Basta comprobarlo para *(I-1)x(J-1)* celdas n_{ij}). En este caso: 2x1.

Por tanto las variables son independientes.

Ejemplo 2.4. **COVARIANZA CORRELACIÓN LINEAL.**

Calculad la covarianza y el coeficiente de correlación lineal en los siguientes gráficos de dispersión:





Solución:

A partir del gráfico 1, las medias y varianzas de X e Y serían:

Se observan n = 7 parejas de valores,

x	y	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$x \cdot y$
1	1	9	4	1
2	2	4	4	2
3	3	1	0	9
4	3	0	0	12
5	4	1	1	20
6	4	4	1	24
7	5	9	4	35
28	*21*	*28*	*14*	*103*

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{28}{7} = 4$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{21}{7} = 3$$

$$s_X^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} = \frac{28}{7} = 4$$

$$s_Y^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n} = \frac{14}{7} = 2$$

Correlación lineal positiva

$$s_{XY} = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{n} - \bar{x}\cdot\bar{y} = \frac{103}{7} - 4\cdot 3 = 14{,}7143 - 12 = 2{,}7143$$

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{2{,}7143}{2\cdot\sqrt{2}} = 0{,}96$$

Análogamente, a partir del gráfico 2:
Se observan n = 7 parejas de valores. Las medias y varianzas son las mismas:

x	y	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$x \cdot y$
1	5	9	4	5
2	4	4	1	8
3	4	1	1	12
4	3	0	0	12
5	3	1	0	15
6	1	4	4	6
7	1	9	4	7
28	*21*	*28*	*14*	*65*

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{28}{7} = 4$$
$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{21}{7} = 3$$

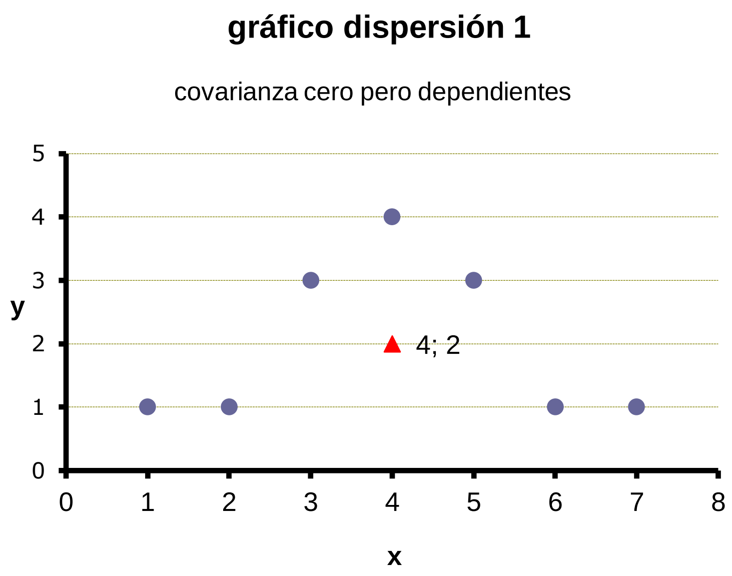
$$s_X^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n} = \frac{28}{7} = 4$$
$$s_Y^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n} = \frac{14}{7} = 2$$

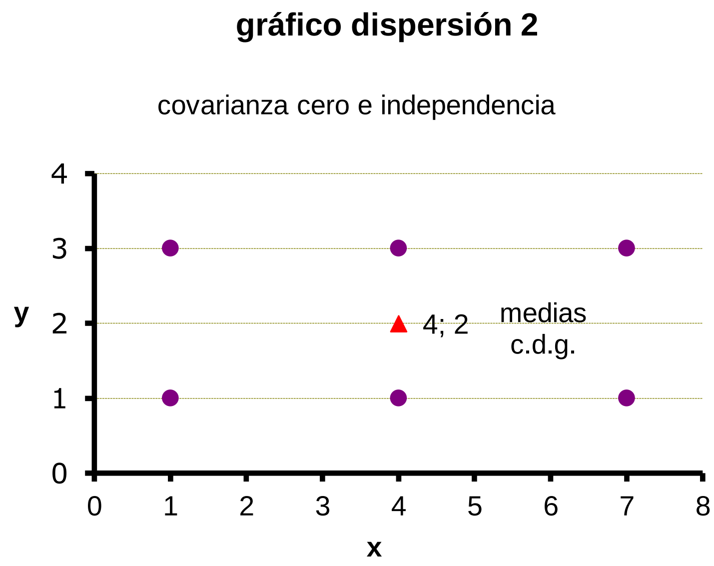
Cambia el signo de la covarianza
Correlación lineal negativa

$$s_{XY} = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{n} - \bar{x}\cdot\bar{y} = \frac{65}{7} - 4\cdot 3 = 9{,}2857 - 12 = -2{,}7143$$
$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{-2{,}7143}{2\cdot\sqrt{2}} = -0{,}96$$

Ejemplo 2.5. EJEMPLOS COVARIANZA NULA $s_{XY} = 0$

Covariación nula: no hay una variabilidad conjunta lineal dominante (positiva o negativa) entre X e Y. Las variables están incorreladas (linealmente)





- **PROPIEDAD**: si las variables X e Y son independientes estadísticamente, la covarianza es cero. El recíproco no es necesariamente cierto.

$$X \ \ e \ \ Y \ \ independientes \Rightarrow s_{XY} = 0$$

En el primer gráfico la covarianza es:

Se observan n = 7 parejas de valores.

x	y	$x \cdot y$
1	1	1
2	1	2
3	3	9
4	4	16
5	3	15
6	1	6
7	1	7
28	*14*	*56*

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{28}{7} = 4$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{14}{7} = 2$$

$$s_{XY} = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{n} - \bar{x} \cdot \bar{y} = \frac{56}{7} - 4 \cdot 2 = 8 - 8 = 0$$

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} = \frac{0}{s_X s_Y} = 0$$

No hay correlación lineal pero se observa otro tipo de relación funcional. Las variables no son independientes.

Se puede comprobar que en el segundo gráfico las variables son independientes y la covarianza también es cero.