GRADO EN INTERNATIONAL BUSINESS

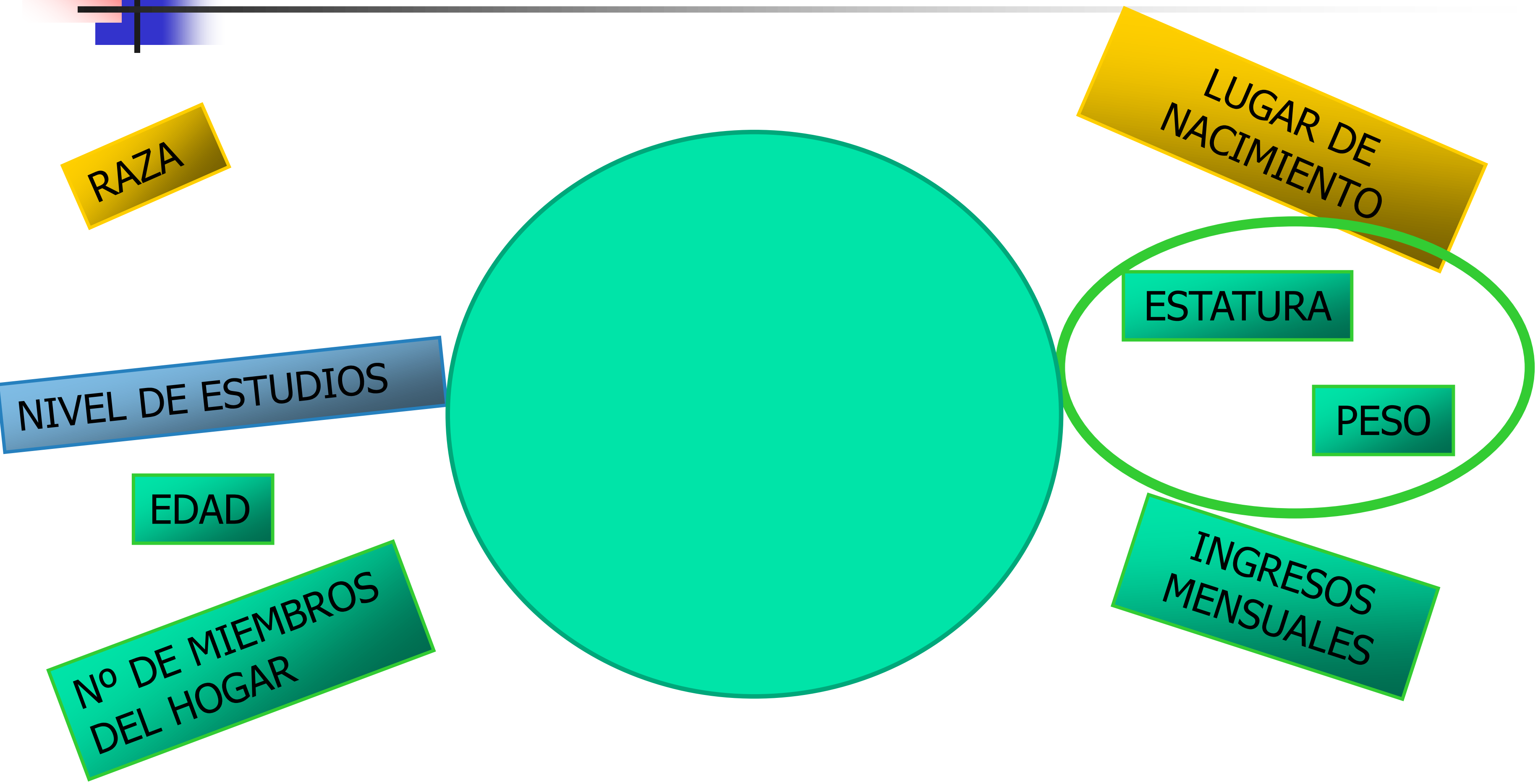
ESTADÍSTICA (35887)
CURSO ACADÉMICO 2018/19

VNIVERSITAT DE VALÈNCIA
OpenCourseWare

TEMA 2
"ANÁLISIS DE DATOS MULTIDIMENSIONALES"

- **2.1. Introducción.**
- **2.2. Datos multidimensionales: distribuciones conjuntas y marginales.**
  - **2.2.1. Distribuciones conjuntas: Representación numérica y gráfica.**
  - **2.2.2. Distribuciones Marginales.**
- **2.3. Independencia Estadística.**
- **2.4. Medidas de relación lineal:**
  - **2.4.1. Covarianza.**
  - **2.4.2. Coeficiente de correlación lineal.**
- **2.5. Independencia estadística e Incorrelación.**
- **2.6. Combinación lineal de variables estadísticas.**
- **2.7. Regresión.**

2.1.- INTRODUCCIÓN



2.2.1. Distribuciones conjuntas: Representación numérica y gráfica

REPRESENTACIÓN NUMÉRICA

TABLA DE DOBLE ENTRADA

Valores o Modalidades de la característica Y

X \ Y	y_1	y_2	…	y_j	…	y_J
x_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1J}
x_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2J}
⋮	⋮	⋮	…	⋮	…	⋮
x_i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{iJ}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_I	n_{I1}	n_{I2}	…	n_{Ij}	…	n_{IJ}

Valores o Modalidades de la característica X

f_{ij}= frecuencia relativa conjunta= $\frac{n_{ij}}{N}$

$$N = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}$$

Frecuencia absoluta conjunta (= nº de veces que se presentan conjuntamente X_i e Y_j)

Si X e Y son medibles (variables) → TABLA DE CORRELACIÓN

Si X e Y NO son medibles (cualitativas) → TABLA DE CONTINGENCIA

2.2.1. Distribuciones conjuntas: Representación numérica y gráfica

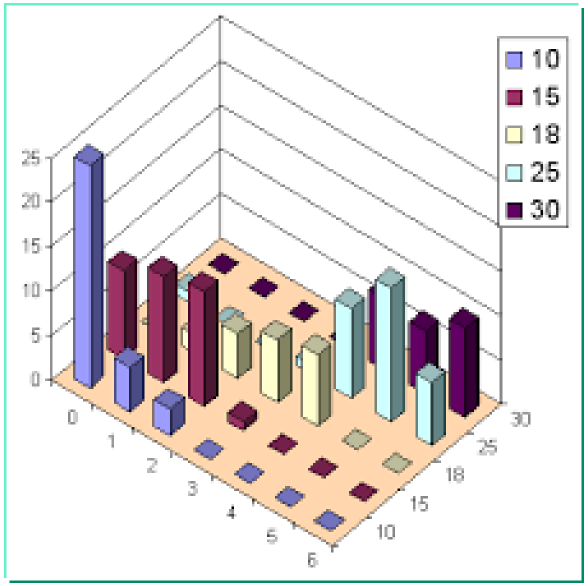
REPRESENTACIÓN NUMÉRICA

Si las frecuencias absolutas conjuntas son unitarias

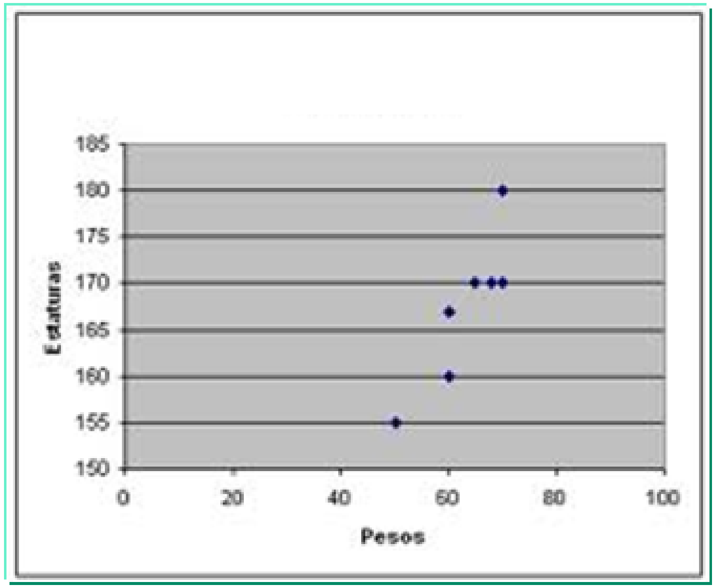
X	Y
x_1	y_1
x_1	y_2
$\vdots$	$\vdots$
x_1	y_J
$\vdots$	$\vdots$
x_I	y_1
x_I	y_2
$\vdots$	$\vdots$
x_I	y_J

2.2.1. Distribuciones conjuntas: Representación numérica y gráfica

REPRESENTACIONES GRÁFICAS



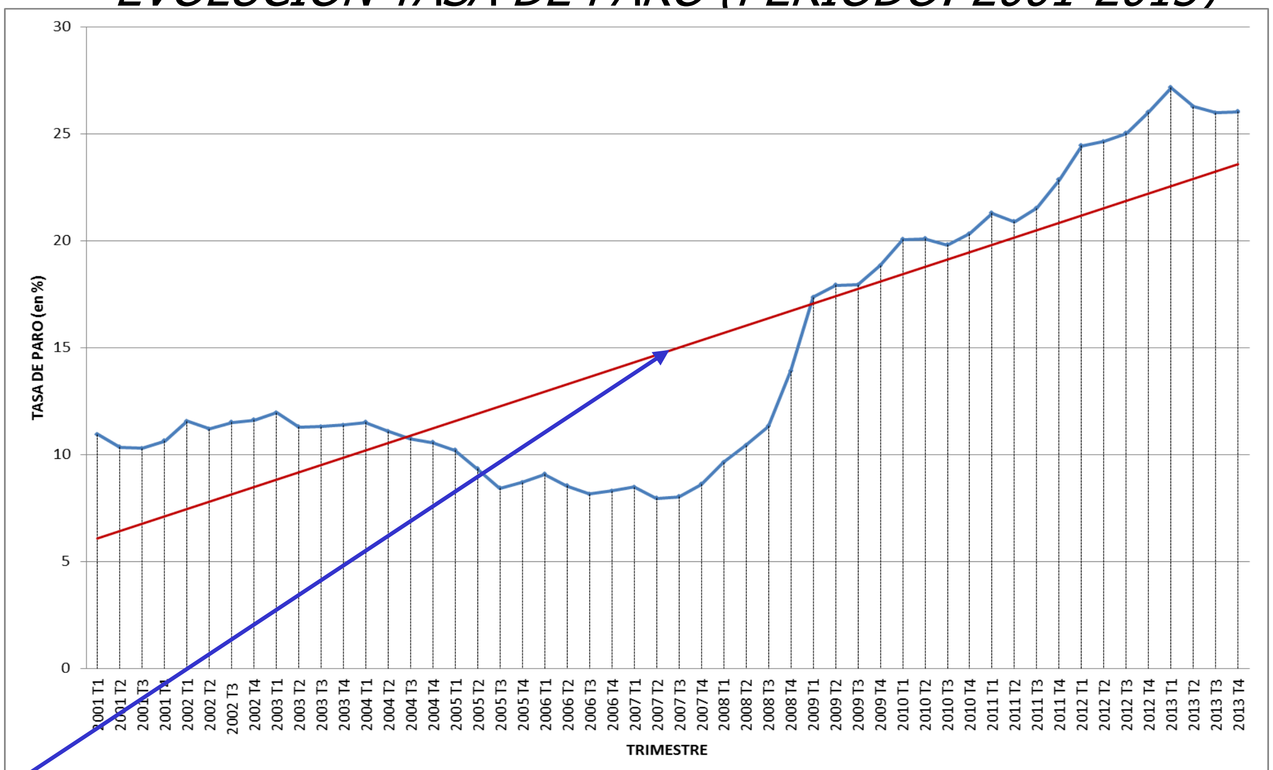
VARIABLES: DIAGRAMA DE DISPERSIÓN



2.2.1. Distribuciones conjuntas: Representación numérica y gráfica

SERIE TEMPORAL (t, Y)

EVOLUCIÓN TASA DE PARO (PERIODO: 2001-2013)

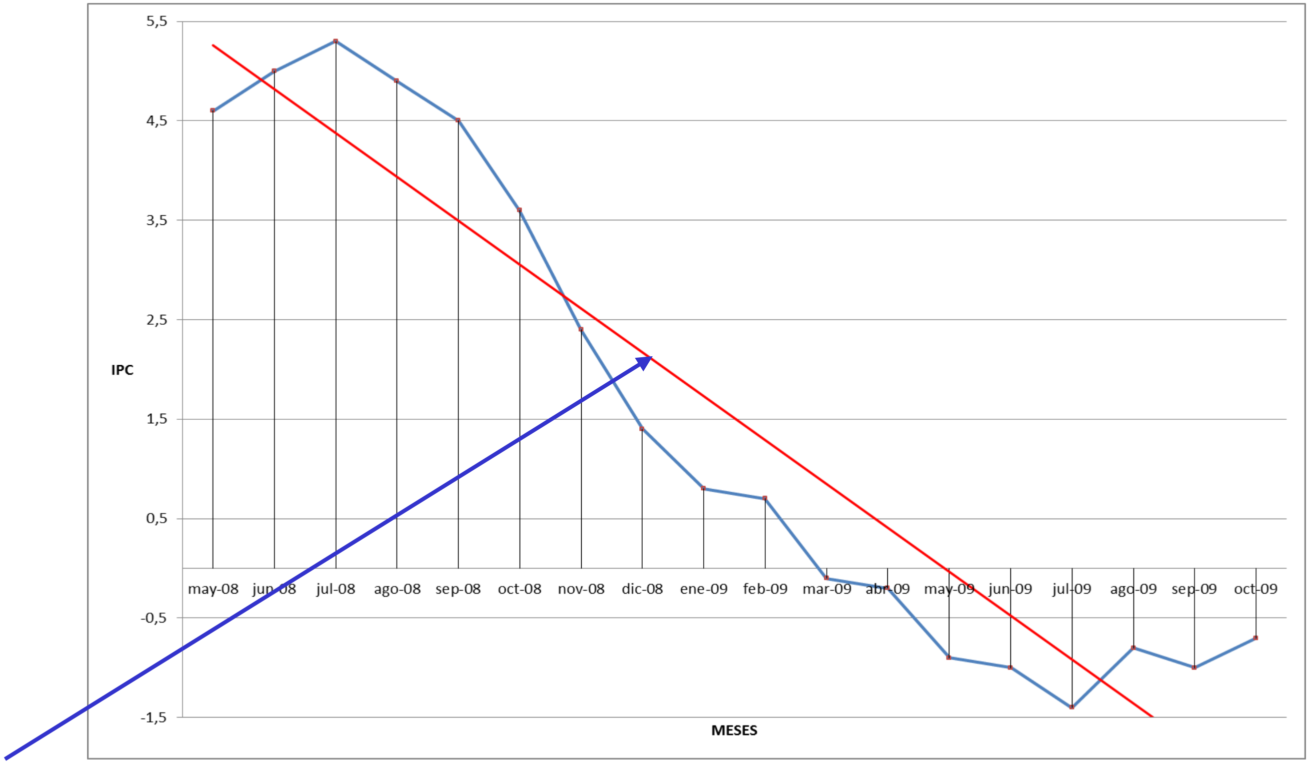


Fuente datos: INE

2.2.1. Distribuciones conjuntas: Representación numérica y gráfica

SERIE TEMPORAL (t, Y)

EVOLUCIÓN IPC (PERIODO: Mayo 2008- Octubre 2009)



Fuente datos: INE

8

2.2.2. Distribuciones marginales

$$n_{i\bullet} = \sum_{j=1}^{J} n_{ij}$$

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}	…	n_{ij}	…	n_{iJ}	$n_{i\bullet}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_I	n_{I1}	n_{I2}	…	n_{Ij}	…	n_{IJ}	$n_{I\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	…	$n_{\bullet j}$	…	$n_{\bullet J}$	N

$$n_{\bullet j} = \sum_{i=1}^{I} n_{ij}$$

$$N = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} = \sum_{i=1}^{I} n_{i\bullet} = \sum_{j=1}^{J} n_{\bullet j}$$

Marginal X

X	$n_{i\bullet}$
x_1	$n_{1\bullet}$
x_2	$n_{2\bullet}$
⋮	⋮
x_i	$n_{i\bullet}$
⋮	⋮
x_I	$n_{I\bullet}$
	$\sum = N$

Marginal Y

Y	$n_{\bullet j}$
y_1	$n_{\bullet 1}$
y_2	$n_{\bullet 2}$
⋮	⋮
y_j	$n_{\bullet j}$
⋮	⋮
y_J	$n_{\bullet J}$
	$\sum = N$

2.2.2. Distribuciones marginales

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}	…	n_{ij}	…	n_{iJ}	$n_{i\bullet}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_I	n_{I1}	n_{I2}		n_{Ij}		n_{IJ}	$n_{I\bullet}$
$n_{\bullet j}$	$n_{\bullet 1}$	$n_{\bullet 2}$	…	$n_{\bullet j}$	…	$n_{\bullet J}$	N

Marginal X

X	$n_{i\bullet}$
x_1	$n_{1\bullet}$
x_2	$n_{2\bullet}$
⋮	⋮
x_i	$n_{i\bullet}$
⋮	⋮
x_I	$n_{I\bullet}$
	$\sum = N$

Marginal Y

Y	$n_{\bullet j}$
y_1	$n_{\bullet 1}$
y_2	$n_{\bullet 2}$
⋮	⋮
y_j	$n_{\bullet j}$
⋮	⋮
y_J	$n_{\bullet J}$
	$\sum = N$

2.3. INDEPENDENCIA ESTADÍSTICA

X e Y independientes

X e Y son estadísticamente independientes $\Longleftrightarrow$ $n_{ij} = \dfrac{n_{i\bullet} \cdot n_{\bullet j}}{N} \quad \forall i, \forall j$

X e Y relacionadas funcionalmente

2.4.1. Covarianza

X	Y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
$\vdots$	$\vdots$
$\vdots$	$\vdots$
$\vdots$	$\vdots$
x_N	y_N

¿EXISTE RELACIÓN LINEAL ENTRE X e Y?

$$S_{XY} = S_{YX} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})\cdot(y_i - \overline{y})}{N} = \frac{\sum_{i=1}^{N} x_i y_i}{N} - \overline{x}\cdot\overline{y}$$

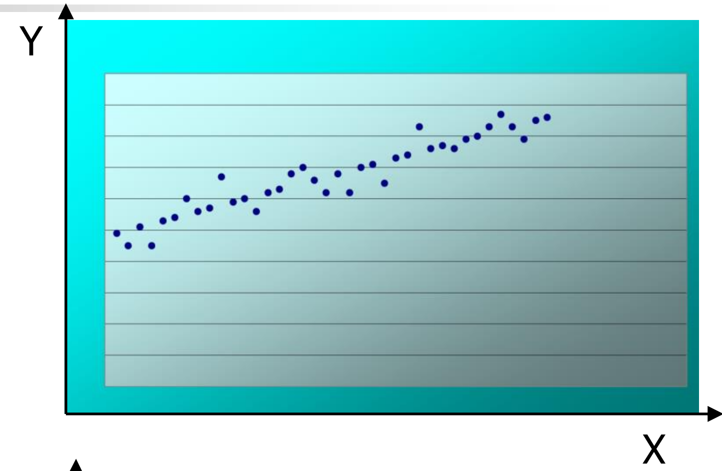
$S_{XY} \neq 0$ → Existe relación lineal entre X e Y

- $S_{XY} > 0$ (Relación lineal positiva)
- $S_{XY} < 0$ (Relación lineal negativa)

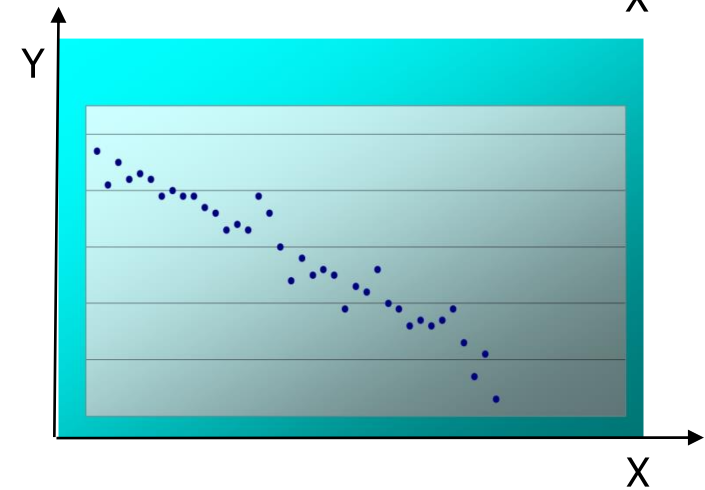
$S_{XY} = 0$ → No existe relación lineal entre X e Y

2.4.1. Covarianza

$S_{XY} > 0$ → Relación lineal positiva entre X e Y



$S_{XY} < 0$ → Relación lineal negativa entre X e Y



2.4.2. Coeficiente de correlación lineal

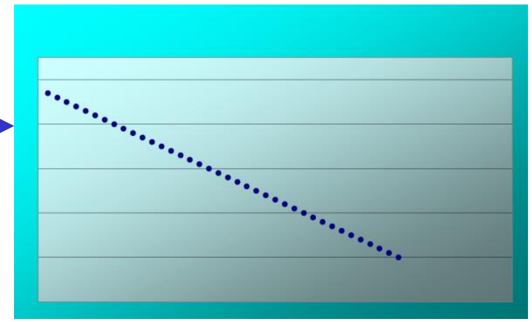
$$r_{XY} = r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

- Proporciona la misma información que la covarianza (igual signo)
- Es adimensional
- Está acotado ($-1 \leq r_{XY} \leq 1$)
- Si $r_{XY} = 0 \longrightarrow$ X e Y están incorrelacionadas

$$S_{XY} = 0$$

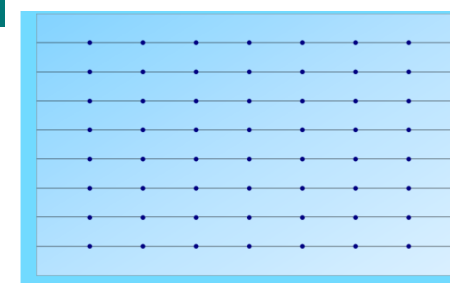
2.4.2. Coeficiente de correlación lineal

$r_{XY} = -1$



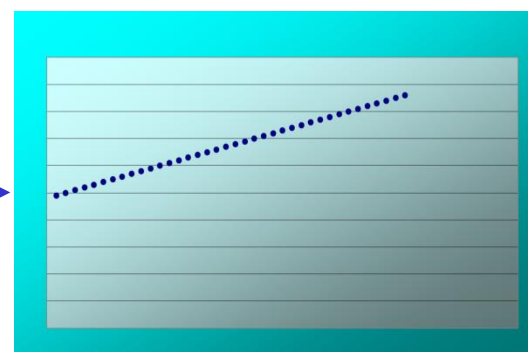
Relación lineal perfecta (negativa) entre X e Y

$r_{XY} = 0$



No existe Relación <u>LINEAL</u> entre X e Y

$r_{XY} = 1$



Relación lineal perfecta (positiva) entre X e Y

 15

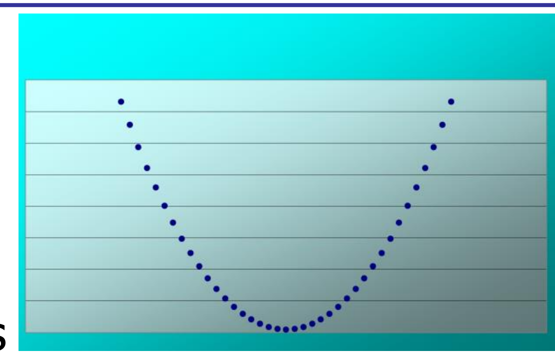
2.5. INDEPENDENCIA ESTADÍSTICA E INCORRELACIÓN

Si X e Y son estadísticamente independientes $\Longrightarrow$ X e Y están incorrelacionadas ($r_{XY} = 0$)

$\not\Longleftarrow$

$Y = X^2$

- Están incorrelacionadas
- No son estadísticamente independientes



2.6. COMBINACIÓN LINEAL DE VARIABLES ESTADÍSTICAS

X e Y= variables estadísticas y a, b y c constantes

$$\boxed{Z=aX+bY+c}$$

$$\overline{Z} = a\overline{X} + b\overline{Y} + c$$

$$S_Z^2 = a^2 S_X^2 + b^2 S_Y^2 + 2abS_{XY}$$

Si X e Y están incorrelacionadas ($S_{XY} = 0$) $\rightarrow$ $\boxed{S_Z^2 = a^2 S_X^2 + b^2 S_Y^2}$

2.7.- REGRESIÓN

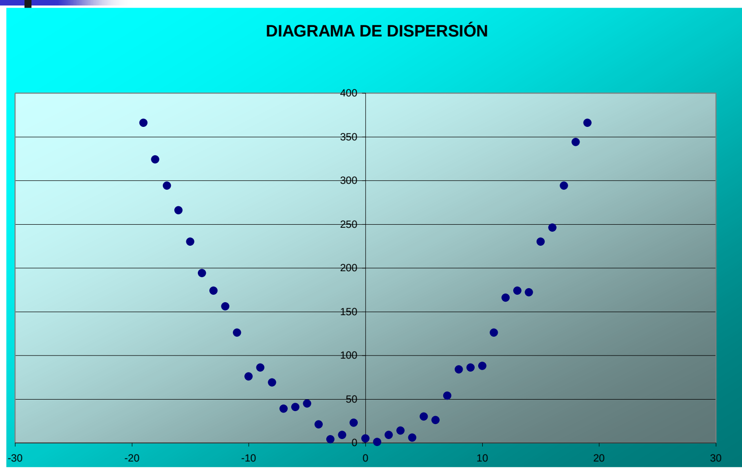
(X, Y)= v.e. bidimensional → TEORÍA DE LA CORRELACIÓN

(X, Y)= v.e. bidimensional → TEORÍA DE LA REGRESIÓN

http://pages.uv.es/piclickers/cat/MenuH_Camtasia.wiki

2.7.- REGRESIÓN

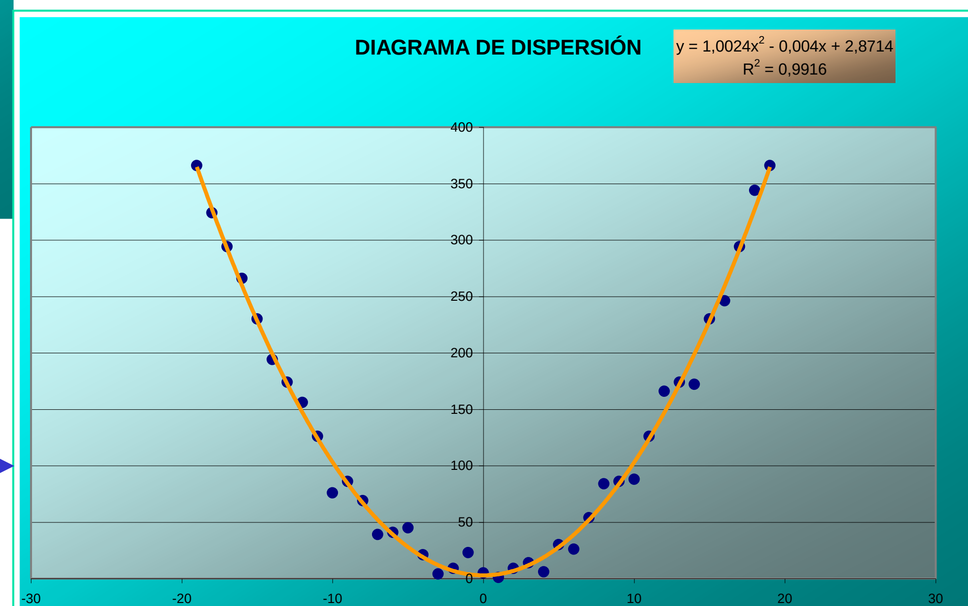
NO LINEAL



¿Existe relación no lineal?

Regresión Parabólica

$Y=a+bX+cX^2$



2.7.- REGRESIÓN

NO LINEAL

Caso Potencial: $Y=aX^b$ → Linealizar la función

$$\log Y = \log(a) + b \cdot \log(X)$$

$$V = A + bU$$

Caso Exponencial: $Y=ab^X$ → Linealizar la función

$$\log Y = \log(a) + X \cdot \log(b)$$

$$V = A + B X$$