TEMA 5: REGRESIÓN Y CORRELACIÓN LINEAL

¿Qué es la correlación?

X	Y	P(X,Y)
1	1	0.30
1	2	0.05
2	1	0.15
3	2	0.08
2	5	0.12
4	1	0.30
		1.00

X	P(X)
1	0.35
2	0.27
3	0.08
4	0.30
	1.00

Y	P(Y)
1	0.75
2	0.13
5	0.12
	1.00

$\mathbb{P}(X = x, Y = y)$ d. conjunta

$$\left.\begin{array}{l}\mathbb{P}(X = x) \\ \mathbb{P}(Y = y)\end{array}\right\}\ \text{marginales}$$

¿Qué nivel de relación existe entre X e Y?

¿Cómo se puede medir?

$$E[X] = 1\cdot 0.35 + 2\cdot 0.27 + 3\cdot 0.08 + 4\cdot 0.30 := \mu_x$$

$$E[Y] = 1\ 0.75 + 2\cdot 0.13 + 5\cdot 0.12 = \mu_y$$

$$E[X-\mu_x] = 0 \qquad Var(X) = \sigma_x^2$$

$$E[Y-\mu_y] = 0 \qquad Var(Y) = \sigma_y^2$$

Se cumple que

$$E\left[\frac{X-\mu_x}{\sigma_x}\right] = 0 \qquad Var\left(\frac{X-\mu_x}{\sigma_x}\right) = 1$$

$$E\left[\frac{Y-\mu_y}{\sigma_y}\right] = 0 \qquad Var\left(\frac{Y-\mu_y}{\sigma_x}\right) = 1$$

$$Z_{\cancel{x}} = \left(\frac{X-\mu_x}{\sigma_x}\right) \quad Z = Z_y = \left(\frac{Y-\mu_y}{\sigma_y}\right)$$

Las dos tienen media 0 y varianza 1

Z_x	$P(Z_x)$
$\frac{1-\mu_x}{\sigma_x}$	0.35
$\frac{2-\mu_x}{\sigma_x}$	0.27
$\frac{3-\mu_x}{\sigma_x}$	0.08
$\frac{4-\mu_x}{\sigma_x}$	0.30

Z_y	$P(Z_y)$
$\frac{1-\mu_y}{\sigma_y}$	0.75
$\frac{2-\mu_y}{\sigma_y}$	0.13
$\frac{5-\mu_y}{\sigma_y}$	0.12

Construimos ahora

Z_xZ_x	Z_y	$P(Z_y, Z_j)$
$\frac{1-\mu_x}{\sigma_x}$	$\frac{1-\mu_y}{\sigma_y}$	0.30
$\frac{1-\mu_x}{\sigma_x}$	$\frac{2-\mu_y}{\sigma_y}$	0.05
$\frac{2-\mu_x}{\sigma_x}$	$\frac{1-\mu_y}{\sigma_y}$	0.15
$\frac{3-\mu_x}{\sigma_x}$	$\frac{2-\mu_x}{\sigma_x}$	0.08 0
$\frac{2-\mu_x}{\sigma_x}$	$\frac{5-\mu_x}{\sigma_x}$	0.12
$\frac{4-\mu_x}{\sigma_x}$	$\frac{1-\mu_y}{\sigma_y}$	0.30

Coeficiente de correlación

$$\rho = E[Z_x Z_y]$$

$$E\left[\left(\frac{X-\mu_x}{\sigma_x}\right)\left(\frac{Y-\mu_y}{\sigma_y}\right)\right] =$$

$$E\left[\frac{1}{\sigma_x\sigma_y}(X-\mu_x)(Y-\mu_y)\right] =$$

$$\frac{1}{\sigma_x\sigma_y}E\left[(X-\mu_x)(Y-\mu_y)\right] =$$

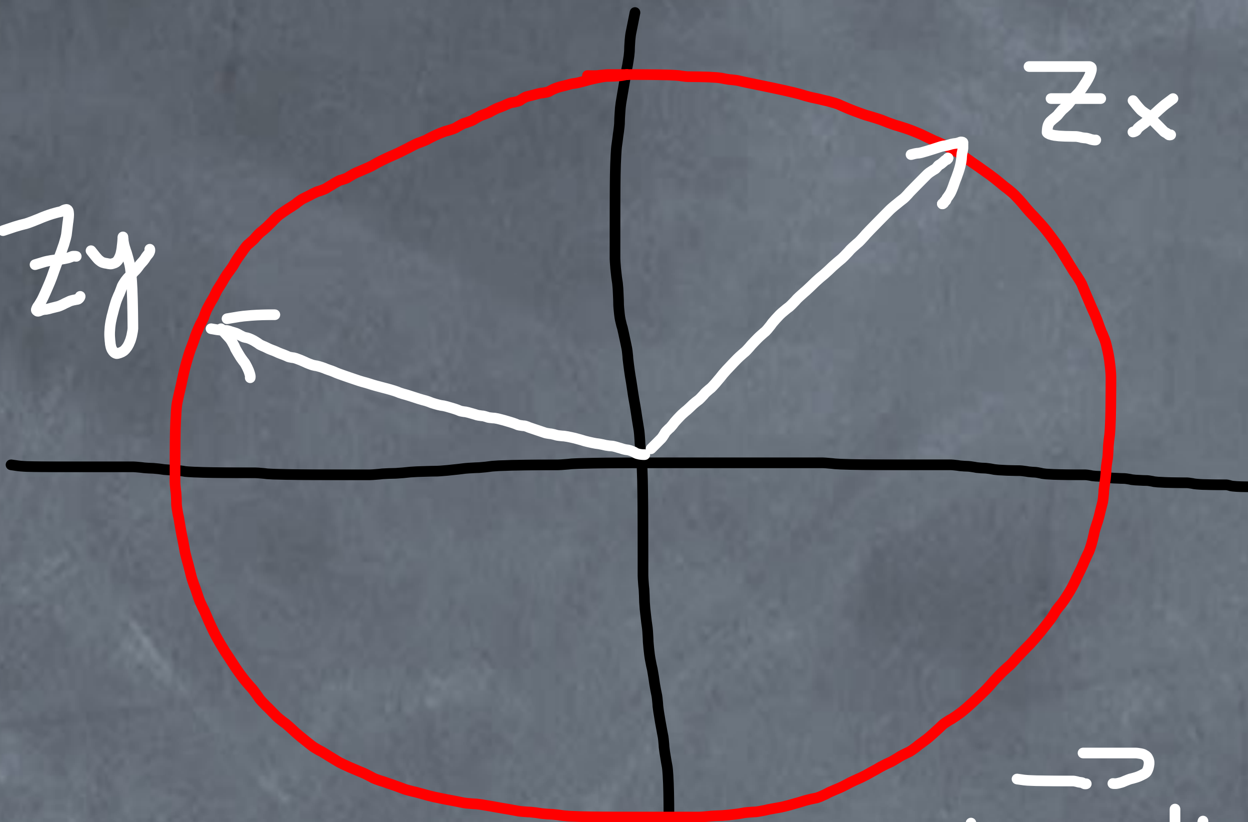
$$\boxed{\rho = \frac{E\left[(X-\mu_x)(Y-\mu_y)\right]}{\sigma_x \quad \sigma_y}}$$

A $E\left[(X-\mu_x)(Y-\mu_y)\right]$ se le llama <u>covarianza</u> de X e Y

¿Qué es el coeficiente de correlación realmente?

$$\begin{array}{ccc} Z_x & Z_y & \\ \left[\begin{matrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{matrix}\right] & & Z_y \end{array}$$

$$\begin{bmatrix} P_{11} \\ P_{22} \\ P_{33} \end{bmatrix}$$



$$\|\vec{x}\|^2 = 1$$

$$x^2 + y^2 + z^2 = 1$$

$$Z_x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad Z_y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$E[Z_x Z_y] = x_1 y_1 P_{11} + x_2 y_2 P_{22} + x_3 y_3 P_{33}$$ siendo

$$P_{11} + P_{22} + P_{33} = 1$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \underline{x_1 y_1 P_{11} + x_2 y_2 P_{22} + x_3 y_3 P_{33}}$$

es un producto escalar en $\mathbb{R}^3$

$$\left\| \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right\| = \sqrt{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}}$$

$$= \sqrt{P_{11} x_1^2 + P_{22} x_2^2 + P_{33} x_3^2}$$

que define un módulo o norma

$$E[Z_x Z_y] = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

$$= \left\| \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right\| \cdot \left\| \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \right\| \cos \theta_{xy}$$

$$= 1 \cdot 1 \cdot \cos \theta_{xy}$$

$$\left\| \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right\| = \sqrt{p_{11} x_1^2 + p_{22} x_2^2 + p_{33} x_3^2}$$

$$= \sqrt{E\left[Z_x^2\right]}$$

$$= \sqrt{Var(Z_x)} = 1$$

$$\boxed{\rho = E\left[Z_x Z_y\right] = \cos\theta_{xy}}$$

$$\boxed{-1 \leq \rho \leq 1}$$

Si $\rho = 0$ $(Z_x \perp Z_y)$ se dice que $\underline{X}$ e $\underline{Y}$ estan incorrelacionados (son linealmente independientes)

Si $\rho = \pm 1$ $\theta_{xy} = 0, \pi$

$\longrightarrow$ z_x, z_y

$\longleftarrow z_x \quad z_y \longrightarrow$

entonces estan ~~totalmente~~ ~~correlacio~~ correlacionados (son linealmente dependientes)

$$Z_y = \pm Z_x$$

$$\left(\frac{Y - \mu_y}{\sigma_y}\right) = \pm \left(\frac{X - \mu_x}{\sigma_x}\right)$$

$$\boxed{Y = \pm \frac{\sigma_y}{\sigma_x} \cdot X - \pm \frac{\sigma_y}{\sigma_x} \mu_x + \mu_y}$$

¿Dados (X, Y) con $p(x,y)$ existen α y β de forma que

$$Y = \beta X + \alpha \, ?$$

$$Y - \beta X - \alpha = 0$$

en el peor de los casos

$$(Y - \beta X - \alpha)^2 \geq 0$$

$$\boxed{\min_{(\beta,\alpha)} E\left[(Y - \beta X - \alpha)^2\right]}$$

$$\shortparallel$$

$$f(\beta, \alpha)$$

$$\frac{\partial f}{\partial \beta} = 2E\left[(Y-\beta X-\alpha)(-X)\right] = 0$$

$$\frac{\partial f}{\partial \alpha} = 2E\left[(Y-\beta X-\alpha)(-1)\right] = 0$$

$$\begin{cases} E[XY] - \beta E[X^2] - \alpha E[X] = 0 \\ E[Y] - \beta E[X] - \alpha = 0 \end{cases}$$

$$\alpha = E[Y] - \beta E[X]$$

$$E[XY] - \beta E[X^2] - \left(E[Y] - \beta E[X]\right)E[X] = 0$$

$$\hat{\beta} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{Var(X)}$$

$$\frac{\sigma_y}{\sigma_x} \frac{\sigma_Y}{\sigma_x} \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x \, \sigma_y} \rho$$

$$\hat{\beta} = \frac{\sigma_y}{\sigma_x} \rho$$

$$\hat{\alpha} = \mu_y - \mu_x \frac{\sigma_y}{\sigma_x} \rho$$

RECTA DE MÍNIMOS CUADRADOS

$$Y = \hat{\beta} X + \hat{\alpha}$$