TEMA III Descripción Estatística Conxunta de Várias Variábeis

"Táboa de correlación ou continxéncia"

x/y	y1	y2	...	yh	marxinal y
x1	f11	f12	...	f1h	f1. = Σ fil
x2	f21	f22	...	f2h	f2. = Σ fil
...	...	...		...	...
xk	fk1	fk2	...	fkh	fk. = Σ fil
marxinal x	f.1 = Σ col	f.2 = Σ col	...	f.h = Σ col	1

(x,y) variábel bidimensional

distribución condicionada
- escoller unha fil. (y/x=xj)
- escoller unha col. (x/y=yj)

os cuadros da táboa tamén se usan asi:

nij	
	fij

x e y estatisticamente **independentes** $\Leftrightarrow$ fij = fi. x f.j $\forall$ i, j

- MOMENTOS

$$a_{r,s} = \sum_{(x)} \sum_{(y)} x_i^r \, y_j^s \, f_{ij} \qquad a_{1,0} = \overline{x} \qquad a_{0,1} = \overline{y}$$

$$m_{r,s} = \sum_{(x)} \sum_{(y)} (x_i - \overline{x})^r \, (y_j - \overline{y})^s \, f_{ij} \qquad m_{2,0} = \sigma_x^2 \qquad a_{0,2} = \sigma_y^2$$

- COVARIANZA

$$cov(x,y) = m_{1,1} = S_{xy} = \sigma_{xy} = \sum \sum (x_i - \overline{x})\,(y_j - \overline{y})\, f_{ij}$$

para calculá-la usa-se $a_{11} - \overline{x}\,\overline{y} = \overline{xy} - \overline{x}\,\overline{y}$

- $S_{xy} > 0 \Rightarrow$ x e y varian no mesmo sentido, senón en contrário.
- x, y son independentes $\Rightarrow \sigma_{x,y} = 0$
(a inversa non se cumpre!)

- VECTOR DE MÉDIAS

$$\vec{M} = \begin{pmatrix} \overline{x} \\ \\ \overline{y} \end{pmatrix} = \begin{pmatrix} \Sigma\, x_i f_{i.} \\ \\ \Sigma\, y_j f_{.j} \end{pmatrix}$$

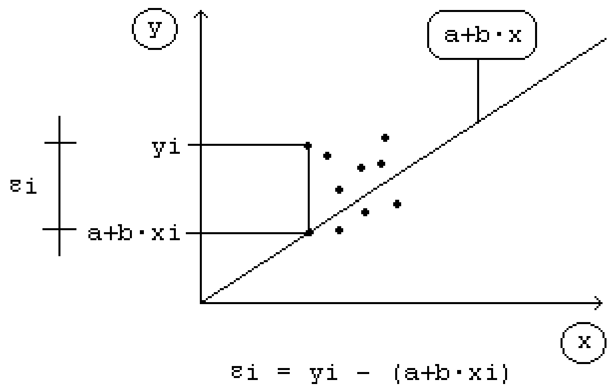
- MATRIZ DE VARIANZAS E COVARIANZAS

dá unha aproximación da dispersión

$$M = \begin{pmatrix} S_x^2 & S_{xy} \\ S_{xy} & S_y^2 \end{pmatrix}$$

$|M| = \sigma_x^2\sigma_y^2-\sigma_{xy}^2 \approx$ área ocupada polo conxunto de dados
(varianza xeralizada)

--variábel bidimensional (x,y)
--observacións = nube de pontos
--queremos saber se existe relación lineal entre x e y: y = a+bx
--para calcular a, b de recta que mellor se axusta usa-se o método dos mínimos cuadrados que minimiza o erro

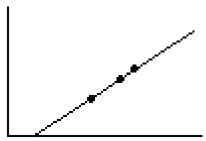
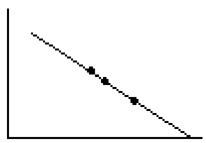
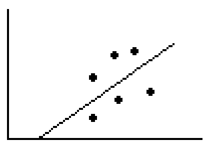
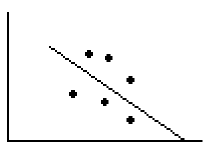


coeficiente de regresión lineal $b = S_{xy} / S_x^2$

$a = \bar{y} - b\,\bar{x}$

} para calcular y = a+b·x

ρ coeficiente de correlación lineal $r_{xy} = S_{xy} / S_xS_y$
para saber se y = a+b·x se axusta ben
$(-1 \leq x \leq 1)$

b>0, r=1  b<0, r=-1 	relación lineal exacta y = a+b·x
b>0, 1>r>0  b<0, 0>r>-1 	dependéncia aleatória lineal
r=0	non existe relación lineal variábeis incorreladas ⇓ ⇑ non si ⇓ ⇑ variábeis independentes

(o modelo de regresión lineal só é válido dentro do renago mostral)