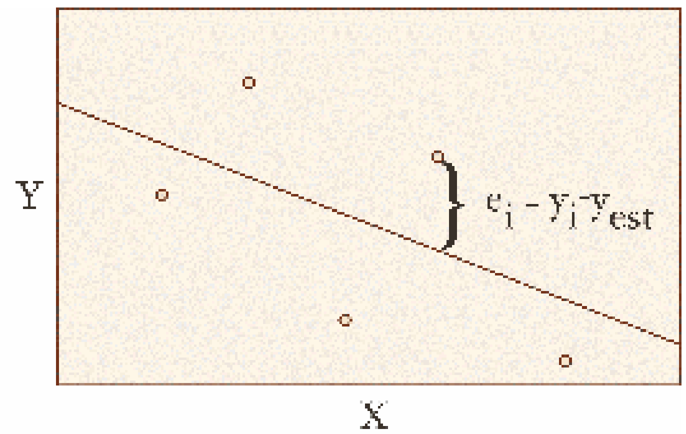
EPSEB
ESCOLA POLITÈCNICA SUPERIOR D'EDIFICACIÓ DE BARCELONA

MA1

ESTADÍSTICA APLICADA

T2 Distribucions Bidimensionals



T2 – 2.2: Distribucions Bidimensionals

MA1 EPSEB

- Distribució bidimensional: cada observació d'un element de la mostra està representada per un parell de valors (x,y)

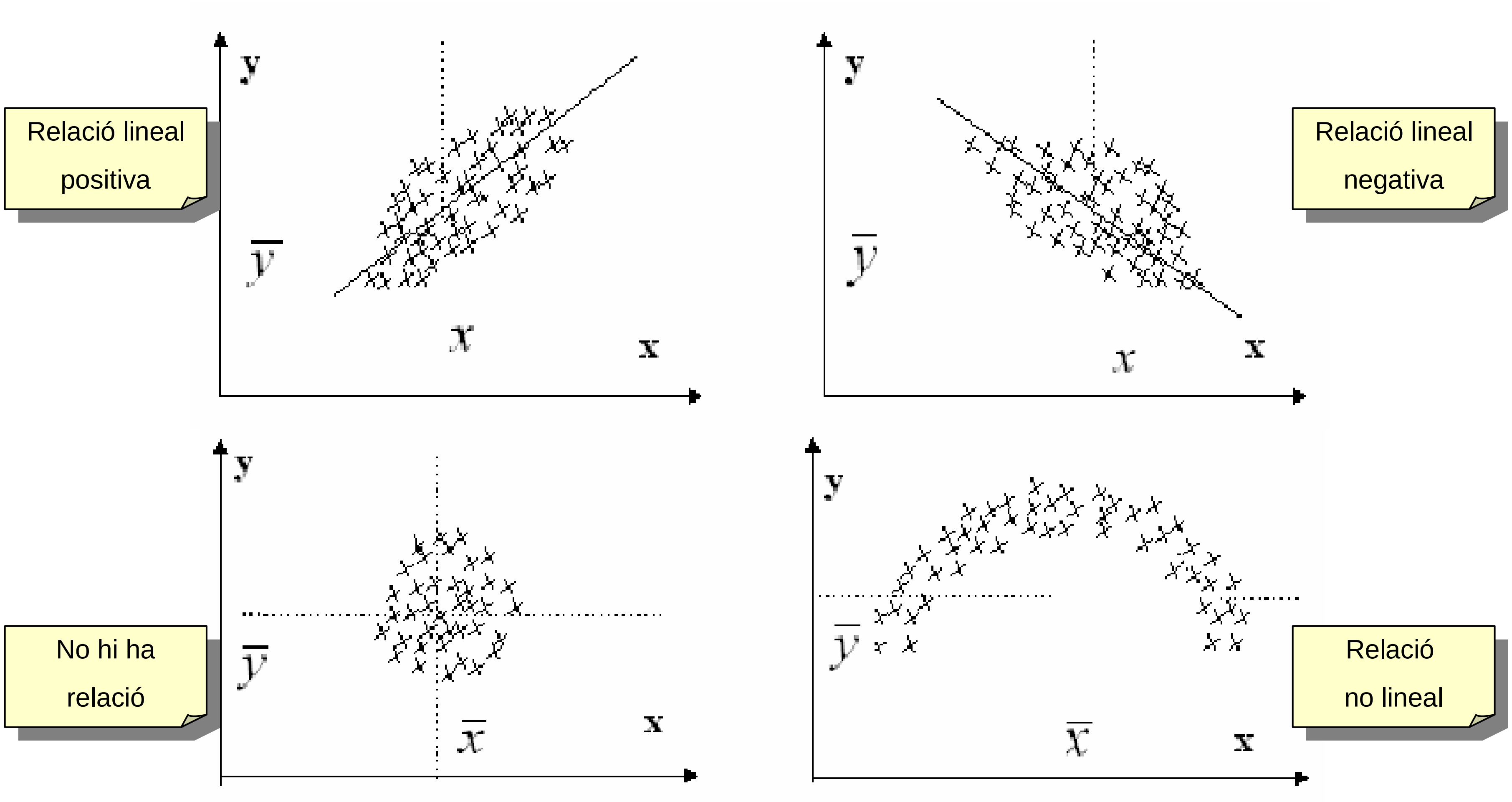
Variable X	Variable Y
x1	y1
x2	y2
x3	y3
...	...
xk	yk

Exemple 1

Velocidad del ultrasonido Km/seg.	Resistencia del hormigón N/mm2
2.98	5,1
3.25	19,1
3.26	15,4
3.42	12,8
3.55	20,9
3.59	28,0
3.62	19,9
3.71	17,4

T2 – 2.3: Diagrames Bivariants

- La representació gràfica dels parells (x,y) s'anomena núvol de punts (*scatterplot*)



T2 – 2.5: Covariància

- La covariància és un indicador del grau d'interdependència lineal entre dues variables X i Y:

$$Cov(X,Y) \equiv S_{xy} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{n} = \frac{1}{n}\sum x_i \cdot y_i - \overline{x} \cdot \overline{y}$$

- Significat de la covariància:

  Cov(X,Y) > 0 → relació lineal positiva

  Cov(X,Y) < 0 → relació lineal negativa

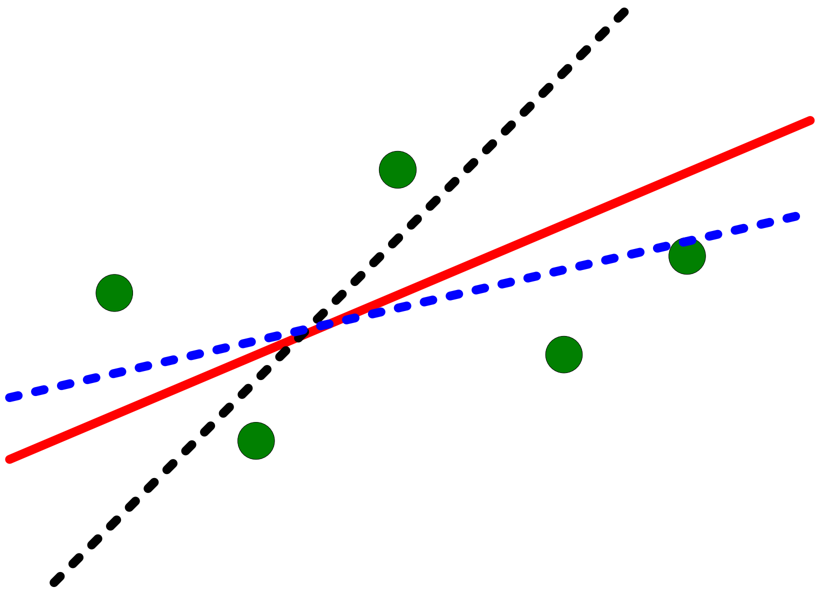
  Cov(X,Y) = 0 → no hi ha relació lineal (n'hi pot haver d'altre tipus)

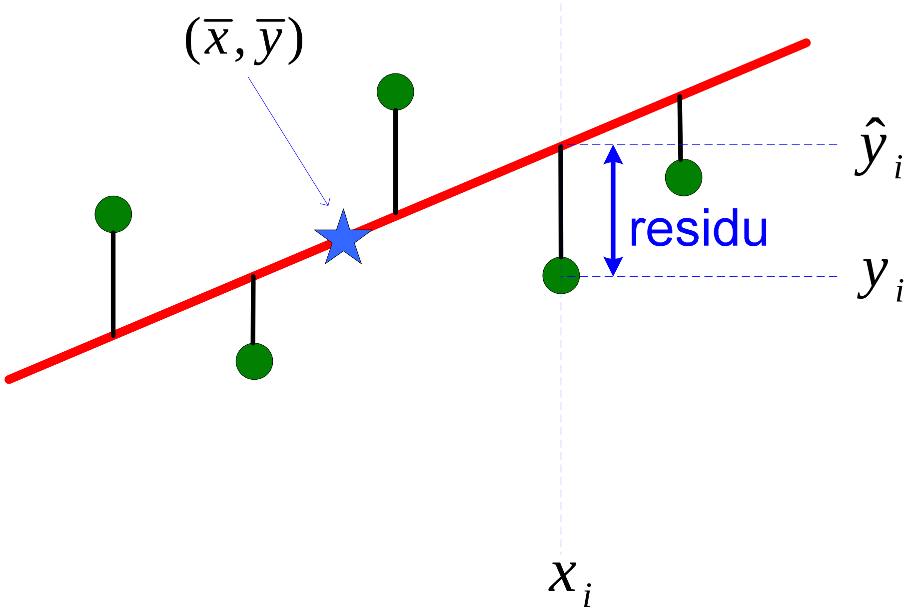
- Problema: la covariància està en funció de les unitats de mesura de les variables (no està estandarditzada).
- Solució: a partir de la covariància, definirem un altre indicador (estandarditzat), el coeficient de correlació lineal

T2 – 2.6: Recta de regressió

MA1 EPSEB

- La recta de regressió de Y sobre X és aquella recta (model) de la forma $\hat{y} = a + b \cdot x$ que "millor s'ajusta" al núvol de punts
- Per a determinar els valors òptims dels coeficients *a* i *b*, s'empra el mètode dels mínims quadrats, el qual minimitza la suma dels quadrats dels residus (diferències entre els valors observats, y_i , i els teòrics predits pel model, $\hat{y}_i$):





(Continua)

T2 – 2.6: Recta de regressió

MA1 EPSEB

- Aplicant mínims quadrats, la recta de regressió de Y (variable dependent) sobre X (variable independent) és:

$$\hat{y} = a + b \cdot x$$

$$b = \frac{Cov(X,Y)}{Var(X)} = \frac{S_{xy}}{S_x^2}$$

$$a = \overline{y} - b \cdot \overline{x}$$

- Alternativament: $y - \overline{y} = b \cdot (x - \overline{x})$
- El pendent, *b*, s'anomena coeficient de regressió de Y sobre X (no confondre amb el coef. de correlació lineal)
- Observacions:
  - recta X sobre Y ≠ recta Y sobre X
  - $(\overline{x}, \overline{y})$ és sempre un punt de la recta
  - recta Y sobre X permet predir *y* conegut *x* (i viceversa)

T2 – 2.7: Correlació

MA1 EPSEB

- Coeficient de correlació lineal, *r*, (indicador estandarditzat del grau d'interdependència lineal):

$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

- Propietats de *r* :
  - $-1 \leq r \leq +1$
  - $\text{signe}(r) = \text{signe}(S_{xy})$
  - Si $r = +1$ ó $r = -1$ → punts sobre la recta (correlació lineal total)
  - Si $r > 0$ → correlació lineal positiva (recta creixent)
  - Si $r < 0$ → correlació lineal negativa (recta decreixent)
  - Si $r = 0$ → no hi ha correlació lineal

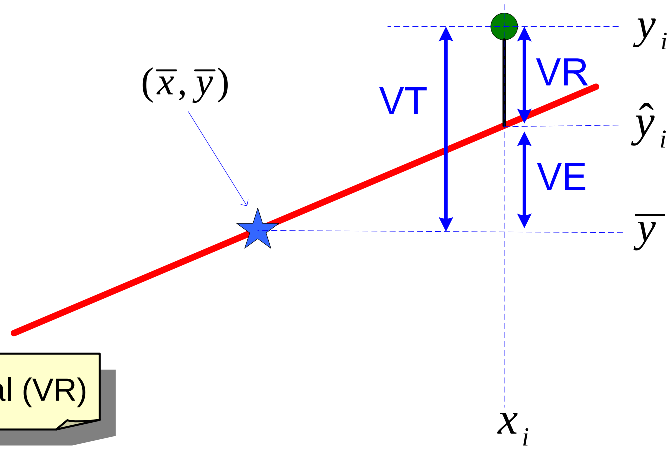
(Continua)

T2 – 2.7: Correlació

- Variància residual, S_r^2 (variància dels residus):

$$S_r^2 = \frac{1}{n}\sum\left(y_i - \hat{y}_i\right)^2 = S_y^2\left(1 - r^2\right)$$

- Variació total de la variable Y, *VT*:



$$VT = \sum\left(y_i - \overline{y}\right)^2 = \sum\left(\hat{y}_i - \overline{y}\right)^2 + \sum\left(y_i - \hat{y}_i\right)^2$$

Variació Total (VT) | Variació Explicada (VE) | Variació Residual (VR)

- El coeficient de determinació, R^2, ens diu quin percentatge de la variació en Y és explicada per la variació en X i el model:

$$R^2 = r^2 = \frac{VE}{VT}$$

- El model seria vàlid per a fer prediccions si aprox $R^2 \geq 80\%$