Lezione 30

16/01/24

STATISTICA BIVARIATA

MISURIAMO DELLA STESSA UNITA' STATISTICA DUE VARIABILI CONTINUE

ESEMPIO [Alberi]

Misuro in una popolazione x_i = il diametro dell'i-esimo albero
y_i = l'età dell'i-esimo albero

Ci chiediamo se queste due variabili sono legate in maniera lineare

$$y_i \overset{?}{\simeq} \alpha x_i + \beta$$

MEDIA	VARIANZA
$\bar{x}$	σ_x^2
$\bar{y}$	σ_y^2

DEFINIZIONE

Si definisce <u>COVARIANZA</u> tra x e y la quantità

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})$$

VARIABILI CONCORDI

SE $\sigma_{xy} > 0 \Rightarrow$ SE X CRESCE Y CRESCE e VICEVERSA

VARIABILI DISCORDI

SE $\sigma_{xy} < 0 \Rightarrow$ SE X CRESCE Y DECRESCE e VICEVERSA

SE $\sigma_{xy} = 0 \Rightarrow$ VARIABILI SONO SCORRELATE

DEFINIZIONE Coefficiente di $\boxed{\text{CORRELAZIONE LINEARE}}$ (o di Bravais - Pearson)

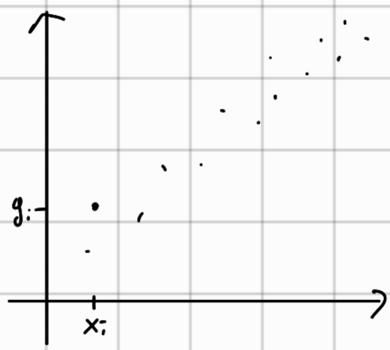
$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$\begin{cases} \sigma_x = 0 \\ \sigma_y = 0 \end{cases}$$

SI PONE $\rho_{yx} = 0$

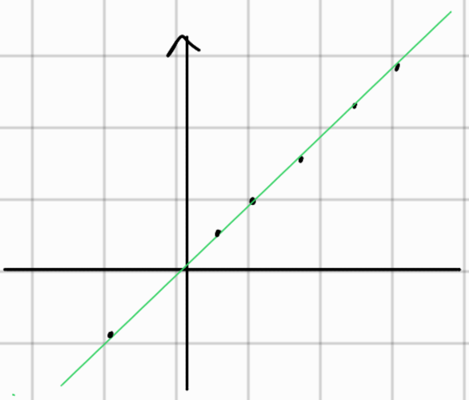
PER RAPPRESENTARE I DATI SI UTILIZZA UNO SCATTERSHOT

$$-1 \leq \rho_{xy} \leq 1$$

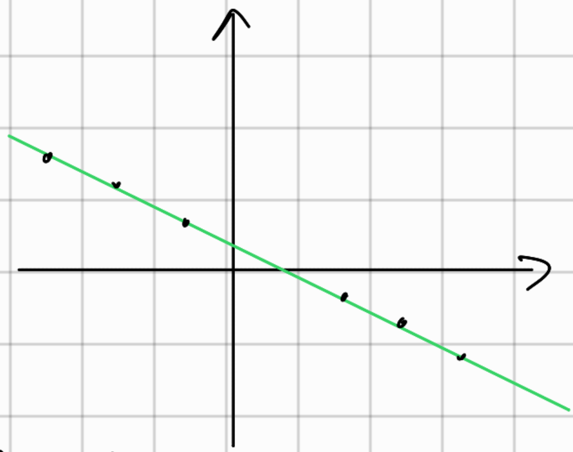


ρ_{xy} CI DICE QUANTO I DATI CHE ABBIAMO SONO VICINI A UNA RETTA

CORRELAZIONE LINEARE $\boxed{\text{ESATTA}}$ $\rho_{xy} = \pm 1$

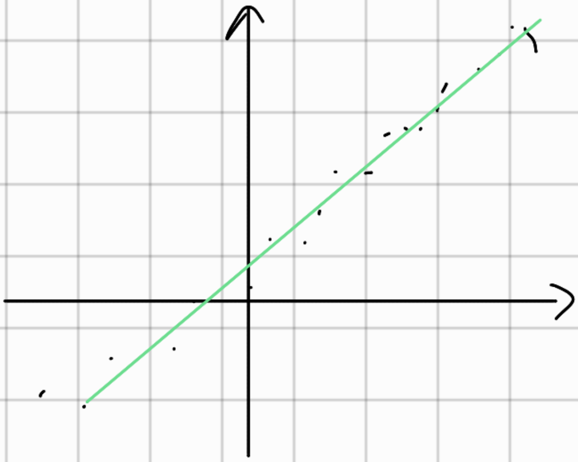


$\rho_{xy} = 1$

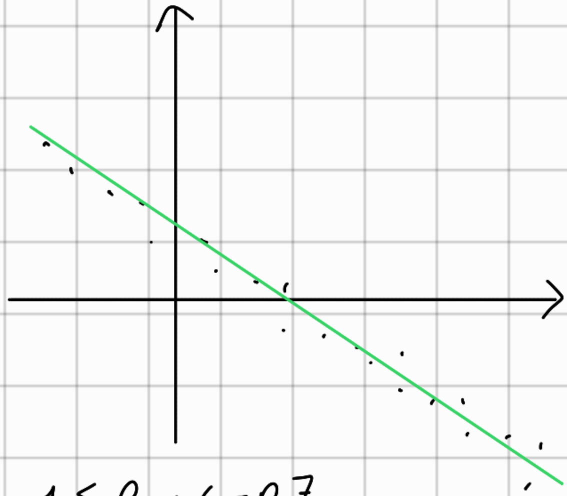


$\rho_{xy} = -1$

CORRELAZIONE LINEARE $\boxed{\text{FORTE}}$ $0.7 < |\rho_{xy}| < 1$

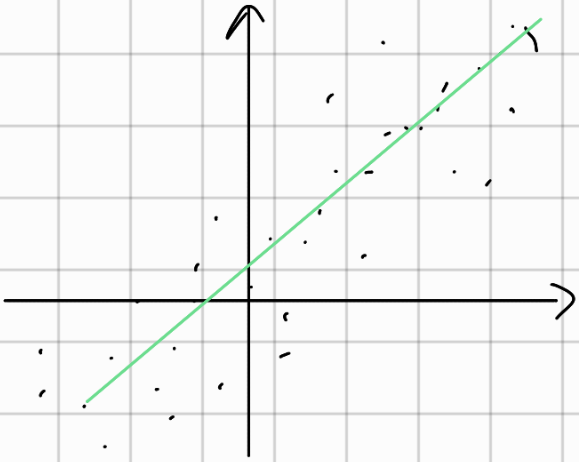


$0.7 < \rho_{xy} < 1$

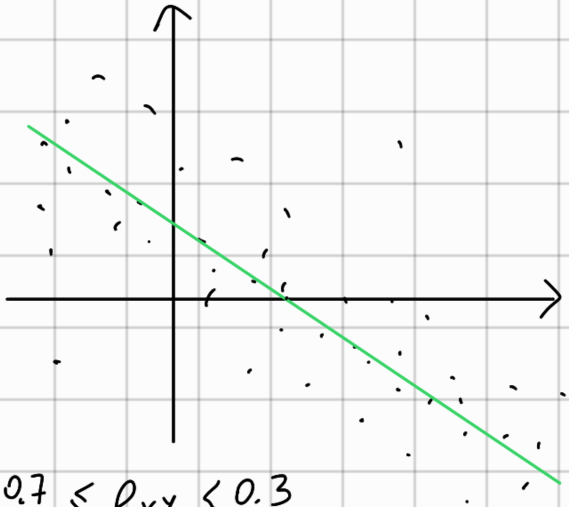


$-1 < \rho_{xy} < -0.7$

CORRELAZIONE LINEARE $\boxed{\text{MODERATA}}$ $0.3 \leq |\rho_{xy}| < 0.7$

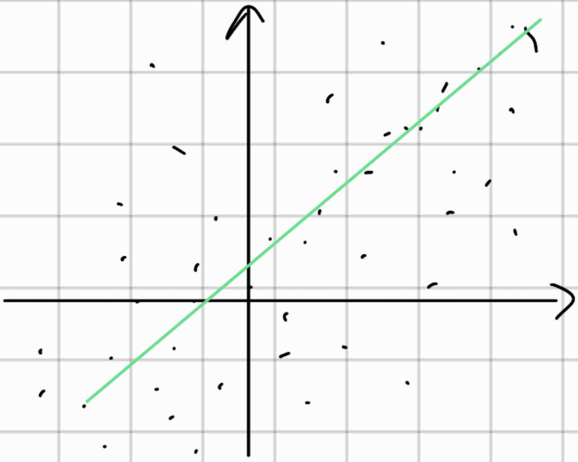


$0.3 < \rho_{xy} \leq 0.7$

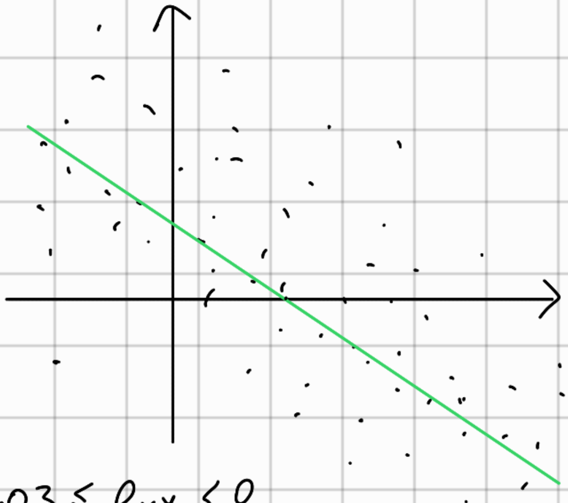


$-0.7 \leq \rho_{xy} < 0.3$

CORRELAZIONE LINEARE $\boxed{\text{DEBOLE}}$ $0.3 \leq |\rho_{xy}| < 1$

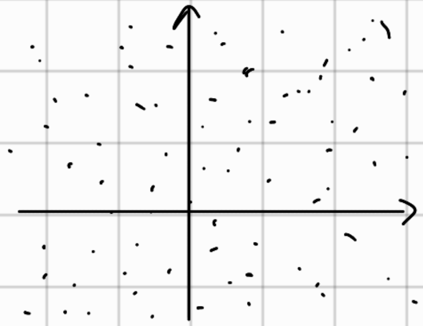


$0 < \rho_{xy} \leq 0.3$



$-0.3 \leq \rho_{xy} \leq 0$

CORRELAZIONE LINEARE $\boxed{\text{NULLA}}$ $\rho_{XY} = 0$



$\rho_{XY} = 0$

ESEMPIO: $\{(1,1),(5,9),(4,7),(10,19),(2,3)\}$ $N=5$

(x_i, y_i)

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{5}(1+5+4+10+2) = \frac{22}{5} = 4.4$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{5}(1+9+7+19+3) = \frac{39}{5} = 7.8$$

$$\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2 = \frac{1}{N}\left[(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+(x_4-\bar{x})^2+(x_5-\bar{x})^2\right]$$

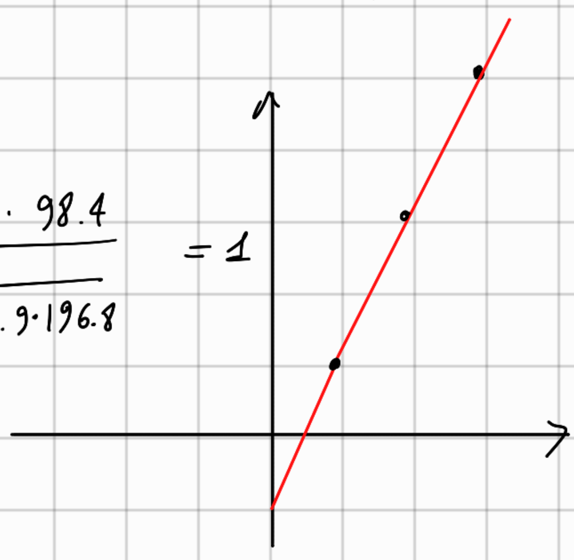
$$= \frac{1}{5}\left[(-3.4)^2+(0.6)^2+(-0.4)^2+(5.6)^2+(-2.4)^2\right] = \frac{49.2}{5} = 9.84$$

$$\sigma_y^2 = \dots = \frac{196.8}{5}$$

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}) = \frac{1}{5}\left[(-3.4)(-6.8)+(0.6)(1.2)+(-0.4)(-0.8)+(5.6)(11.2)+(-2.4)(-4.8)\right]$$

$$= 98.4$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x \sigma_y}} = \frac{98.4}{\sqrt{\frac{42.9}{5}\cdot\frac{196.8}{5}}} = \frac{5\cdot 98.4}{\sqrt{42.9\cdot 196.8}} = 1$$



CORRELAZIONE LINEARE <u>ESATTA</u>

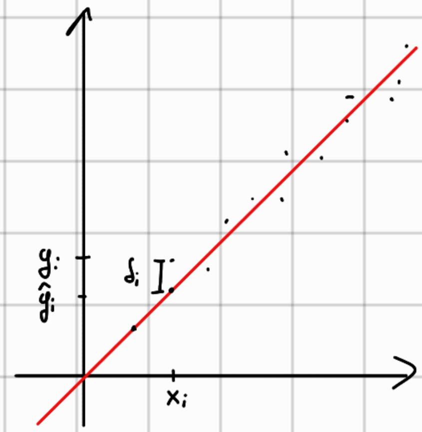
$y = 2x - 1$

SE LA CORRELAZIONE NON È ESATTA

La retta che meglio approssima i dati è detta

$\boxed{\text{RETTA DI REGRESSIONE LINEARE}}$

$y = \alpha x + \beta$



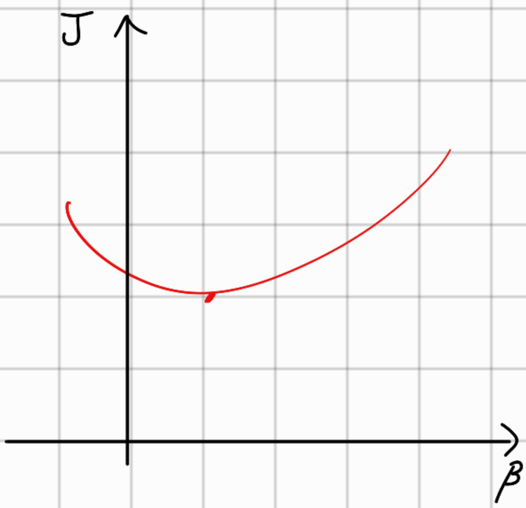
$$\hat{y}_i = \alpha x_i + \beta$$

$$\delta_i = |y - \hat{y}_i|$$

VOGLIAMO MINIMIZZARE GLI ERRORI

$\Downarrow$

METODO DEI MINIMI QUADRATI



$$J(\alpha,\beta) = \frac{1}{N}\sum_{i=1}^{N} \delta_i^2 =$$

$$= \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 =$$

$$= \frac{1}{N}\sum_{i=1}^{N} (y_i - \alpha x_i - \beta)^2$$

MINIMIZZIAMO PRIMA RISPETTO a β e POI RISPETTO a α

$$\min_{\alpha}\left[\min_{\beta} J(\alpha,\beta)\right]$$

CALCOLO LA DERIVATA PRIMA RISPETTO A β $\left(\text{SI INDICA CON } \frac{\partial J}{\partial \beta}\right)$

$$J(\alpha,\beta) = \frac{1}{N}\left(\sum_{i=1}^{N}(y_i - \alpha x_i - \beta)^2\right)$$

$$\frac{\partial J}{\partial \beta} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial \beta}\left[y_i - \alpha x_i - \beta\right]^2$$

$$= -\frac{1}{N}\sum_{i=1}^{N} 2\left[y_i - \alpha x_i - \beta\right]$$

E' COME SE FOSSE

$$(3-\beta)^2$$

$$2(3-\beta)(-1)$$

$$= -2\left[\underbrace{\frac{1}{N}\sum y_i}_{} - \alpha\frac{1}{N}\sum_{i=1}^{N} x_i - \frac{1}{\cancel{N}}\cancel{N}\beta\right]$$

$$-2\left[\bar{y} - \alpha\bar{x} - \beta\right] = 0$$

$$\Rightarrow \quad \beta = \bar{y} - \alpha\bar{x}$$

$$\min_{\alpha}\left[J(\alpha, \bar{y} - \alpha\bar{x})\right] = \min_{\alpha}\left[\frac{1}{N}\sum_{i=1}^{N}\left[y_i - (\alpha x_i + \bar{y} - \alpha\bar{x}\right)^2\right]$$

$$= \min_{\alpha}\left[\frac{1}{N}\sum_{i=1}^{N}\left[y_i - \bar{y} - \alpha(x_i - \bar{x})\right]^2\right]$$

FACENDO I CALCOLI:

$$\boxed{\alpha = \frac{\sigma_{xy}}{\sigma_x^2}} \Rightarrow \boxed{\beta = \bar{y} - \frac{\sigma_{xy}}{\sigma_x^2}\bar{x}}$$

ESEMPIO $N=5$

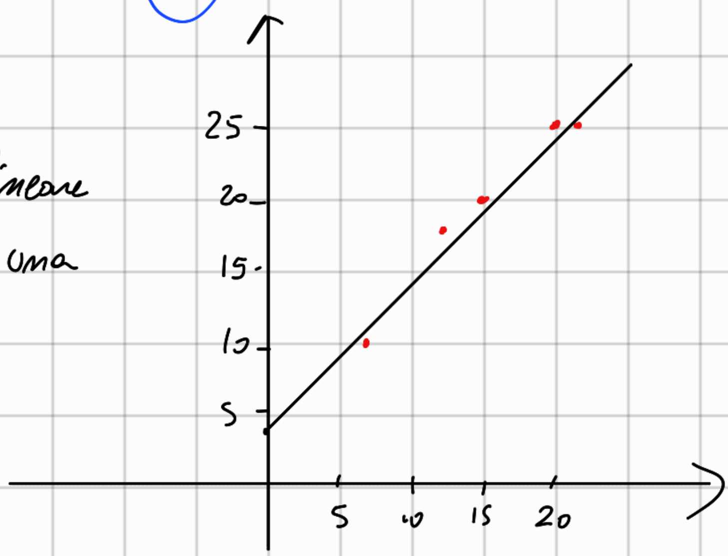
Somministro una dose x_i di farmaco e osservo una diminuzione di pressione arteriosa y_i

X (VARIABILE INDIPENDENTE) DOSE (mg)	DIMINUZIONE (mm Hg) Y (VARIABILE DIPENDENTE)	
7	10	70
12	18	196
15	20	300
20	25	500
22	25	550

a) Scrivere la retta di regressione lineare

b) Che dose dobbiamo dare se vogliamo una riduzione di circa 15 mm Hg

c) Di quanto diminuirà la pressione se do una dose di 30 mm Hg ?



$$y = \alpha x + \beta \qquad \alpha = \frac{\sigma_{xy}}{\sigma_x} \qquad \beta = \bar{y} - \frac{\sigma_{xy}}{\sigma_x}\bar{x}$$

$$\bar{x} = \frac{1}{5}\left(7+12+15+20+22\right) = 15.2$$

$$\bar{y} = \frac{1}{5}\left(10+18+20+25+25\right) = 19.6$$

$$\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y}) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \bar{x}\bar{y}$$

$$\sigma_{xy} = \Big[(7-15.2)(10-19.6) + (12-15.2)(18-19.6) + (20-15.2)(20-19.6) +$$
$$+ (22-15.2)(25-19.6) + (15-15.2)(25-19.6)\Big] \cdot \frac{1}{5}$$

$$= 29.28$$

$$\sigma_{xy} = \frac{1}{5}(1616) - (15.2)(19.6) = 29.28$$

$$\sigma_x^2 = \frac{1}{5}\Big[(7-15.2)^2 + (12-15.2)^2 + (15-15.2)^2 + (20-15.2)^2 + (22-15.2)^2$$

$$\frac{1}{5}\Big[(-8.2)^2 + (-3.2)^2 + (-0.2)^2 + (4.8)^2 + (6.8)^2\Big] = 29.36$$

$$\alpha = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{29.28}{29.36} \simeq 0.997$$

$$\beta = \bar{y} - \alpha\bar{x} = 19.6 - (0.997)(15.2) = 4.44$$

LA RETTA DI REGRESSIONE LINEARE È

$$y = 0.997\,x + 4.44$$

SE VOGLIO UNA DIMINUZIONE DI 15 mmHg che dose devo dare?

$$15 = 0.997\,x + 4.44$$

$$15 - 4.44 = 0.997\,x$$

$$\frac{15 - 4.44}{0.997} = x_{15} \qquad x_{15} \simeq 10.59$$

SE DO UNA DOSE DI 30 mmHg ?

$$y_{30} = 0.997 \cdot 30 + 4.44 = 34.36 \quad mmHg$$