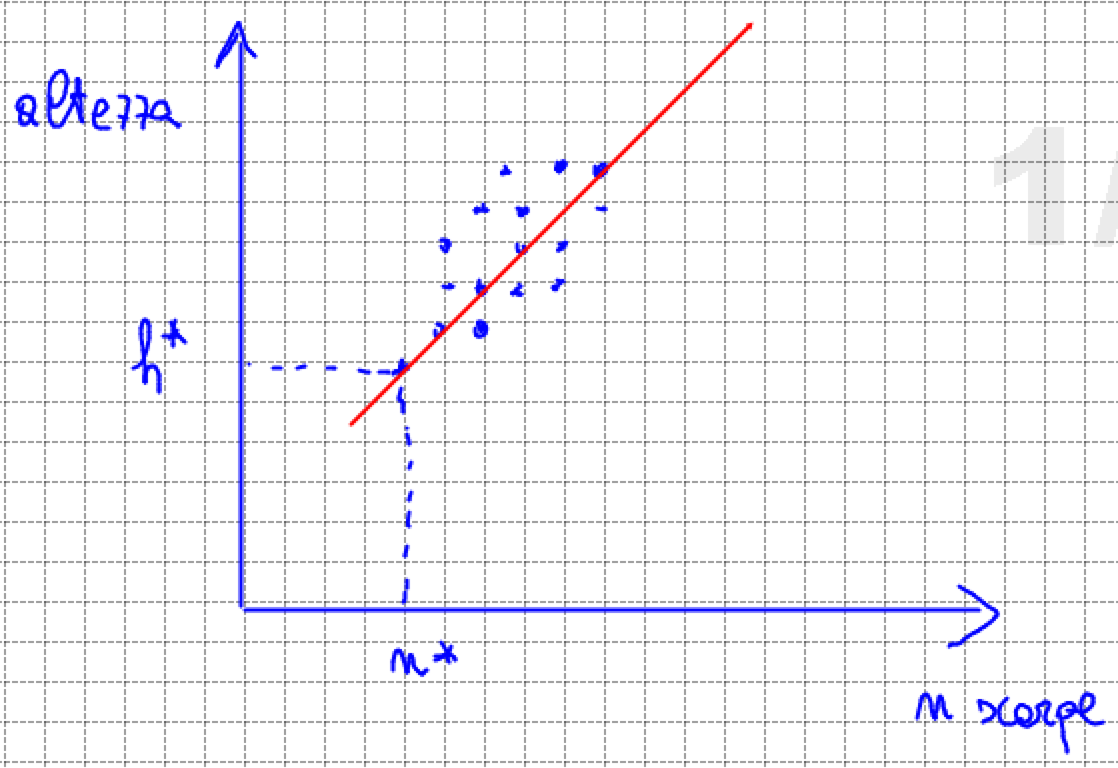
CORRELAZIONI:

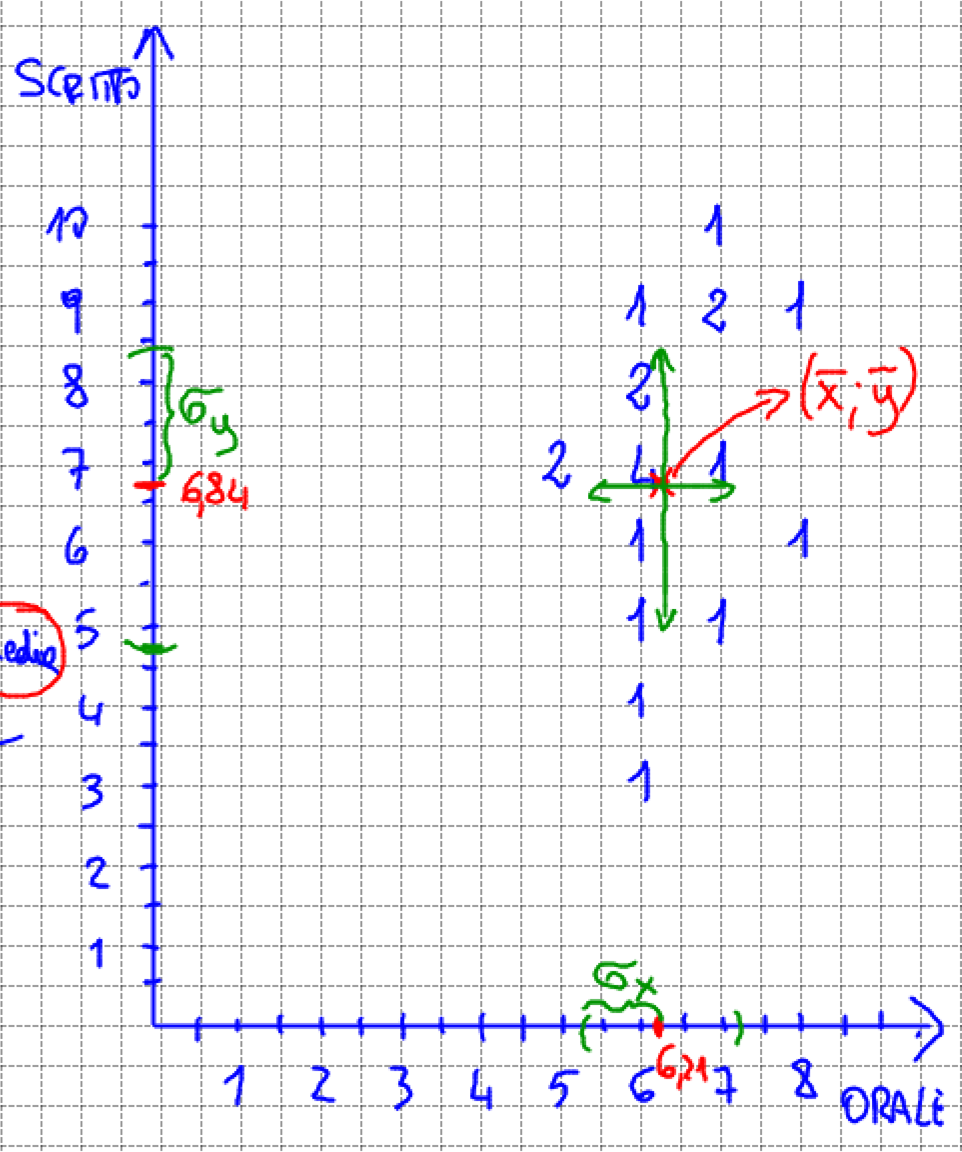


ESEMPIO

Voti di matematica del primo trimestre della classe 5P

	A	B	C	
1	O	S	O	S
2	7	10	6	6
3	6	9	8	6
4	7	9	6	5
	7	9	7	5
	8	9	6	4
	6	8	6	3
	6	8	6,21	6,84 media
	5	7	0,69	1,76 σ
	5	7		
	6	7		
	6	7		
	6	7		
	6	7		
	7	7		

$$\sigma = \sqrt{\frac{\sum \Delta x^2}{n}}$$



FOGLIO ELETTRONICO

Per la media: = MEDIA(A2:A22)

: = MEDIA(B2:B22)

Per il sigma: = dev.st.pop(A2:A22)

: = dev.st.pop(B2:B22)

σ = deviazione standard della popolazione.

Troviamo ora la retta di REGRESSIONE (retta che meglio interpreta i dati)

$$y = mx + q$$

$$y - \bar{y} = m(x - \bar{x}) \qquad (\bar{x};\bar{y}) \text{ sono i valori medi}$$

$$r = m = \text{coefficiente di correlazione} = \frac{cov}{\sigma_x \sigma_y}$$

cov = covarianza (variabile che viene fuori quando siamo in presenza di due variabili)

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} \qquad \sigma^2 = \text{varianza} = \frac{\sum \Delta x^2}{n}$$

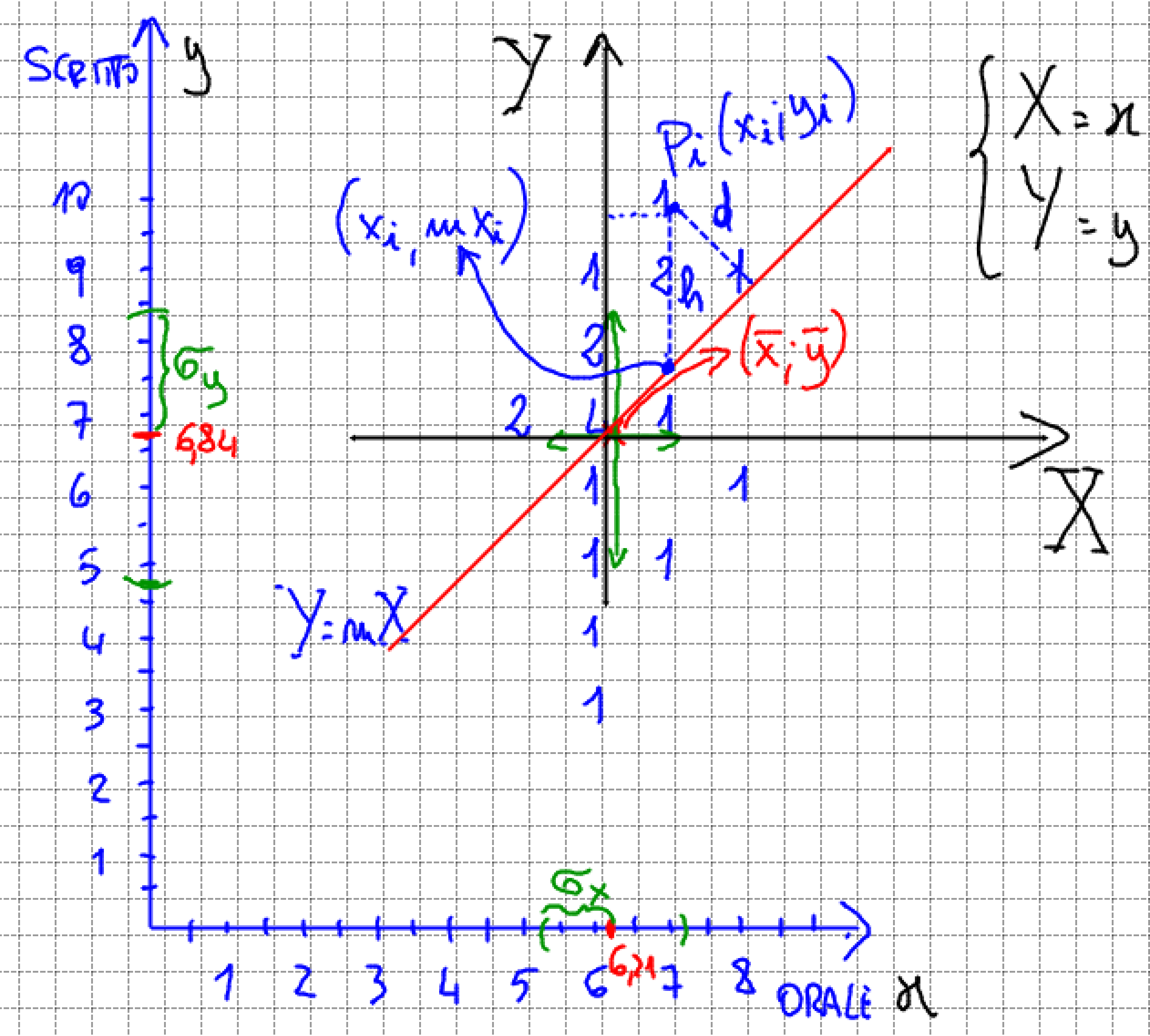
$$cov = \frac{\sum \Delta x \Delta y}{n} = \frac{\sum (x-\bar{x})(y-\bar{y})}{n}$$

cov = covarianza = 0,19

La funzione corrispondente della covarianza è:

= covarianza (A2 : A22 ; B2 : B22)

$m = 0{,}16$



$$\begin{cases} X = x - \bar{x} \\ Y = y - \bar{y} \end{cases}$$

Analizziamo i dati: ogni punto $P_i(x_i;y_i)$ $i = 1, \dots, 20$ è tale che la distanza $d(P_i; Y = mX)$ e la somma dei quadrati di queste distanze deve essere minimo.
È la stessa cosa studiare le altezze h fra i punti P_i e l'intersezione della perpendicolare per P_i con la retta $Y = mX$

$$\sum h_i^2 =$$
$$= \sum (Y_i - mX_i)^2 = \sum (Y_i^2 - 2mX_iY_i + m^2X_i^2) =$$
$$= \sum Y_i^2 - 2m\sum X_iY_i + m^2\sum X_i^2 =$$
$$= Y_1^2 - 2mX_1Y_1 + m^2X_1^2 +$$
$$+ Y_2^2 - 2mX_2Y_2 + m^2X_2^2 +$$
$$+ \dots +$$
$$+ Y_{20}^2 - 2mX_{20}Y_{20} + m^2X_{20}^2$$

$$\sum h_i^2 = \underbrace{\sum X_i^2}_{a}\, m^2 \underbrace{-2\sum X_iY_i}_{b}\, m \underbrace{+\sum Y_i^2}_{c} \quad \text{eq. parabola}$$

Il Vertice di questa parabola ci minimizza

$$V(x_v; y_v) \qquad x_v = -\frac{b}{2a} \qquad x_v = \cancel{+}\frac{\cancel{+2}\sum X_iY_i}{\cancel{2}\sum X_i^2}$$

$$\boxed{x_v = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}} = m$$

$\sum (x_i - \bar{x})(y_i - \bar{y})$ → covarianza

$\sum (x_i - \bar{x})^2$ → varianza