

ESERCIZIO DI STATISTICA SU
RETTA DI REGRESSIONE, BONTÀ DI ADATTAMENTO, INDICE DI CORRELAZIONE

La direzione di una catena di fast-food ha effettuato una rilevazione dei costi in migliaia di euro (Y) in relazione alle presenze giornaliere in migliaia di unità (X), i risultati sono riportati in tabella

X	2	3	4	1
y	9	13	17	5

- Disegnare il grafico di dispersione**
- Determinare la retta di regressione**
- Disegnare la retta di regressione**
- Calcolare il coefficiente di determinazione lineare o bontà di adattamento**

Costruiamo la seguente tabella che ci permetterà di conoscere i dati che interessano

X _i	Y _i	x _i · y _i	x _i ²			
2	9	18	4			
3	13	39	9			
4	17	68	16			
1	5	5	1			
totali	130	30				

Calcoliamo le medie \bar{x} e \bar{y}

$$\bar{x} = \frac{2+3+4+1}{4} = 2,5$$

$$\bar{y} = \frac{9+13+17+5}{4} = 11$$

L'equazione della retta di regressione è $y = b_1x + b_0$

$$\text{Dove } b_1 = \frac{\text{Codev}(X,Y)}{\text{Dev}(X)} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$

Tutte le grandezze che compaiono in questa formula sono state calcolate

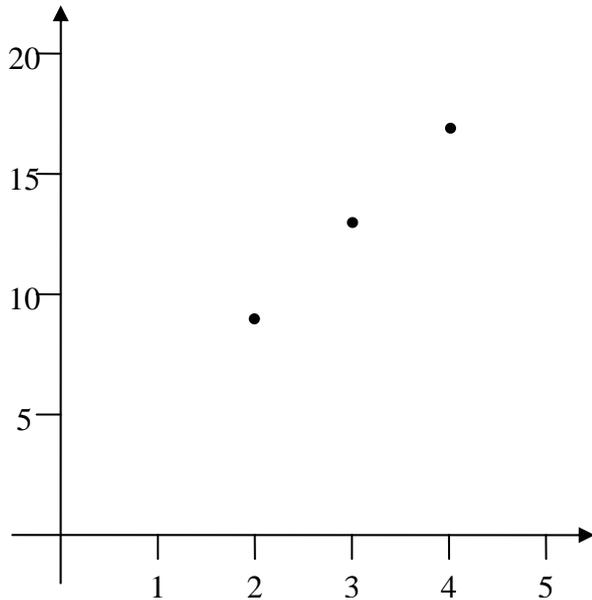
$$b_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{130 - 4 \cdot 2,5 \cdot 11}{30 - 4 \cdot 2,5^2} = \frac{130 - 110}{30 - 25} = 4$$

$$\text{Invece } b_0 = \bar{y} - b_1 \bar{x} = 11 - 4 \cdot 2,5 = 11 - 10 = 1$$

La retta di regressione $y = b_1x + b_0$ è $y = 4x + 1$

Rappresentiamo ora il grafico di dispersione e la retta di regressione.

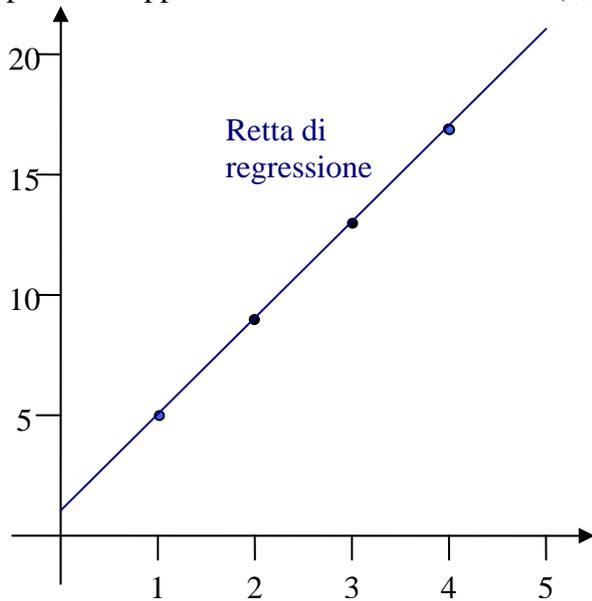
Disponiamo un opportuno diagramma cartesiano, con x che va da 0 al valore massimo 4; y va da 0 al valore massimo 17. In questo grafico rappresentiamo i punti (x_i, y_i) .



Per rappresentare la retta di regressione individuiamo almeno due punti della retta; assegniamo dei valori alla x e ricaviamo i corrispondenti valori di y. Assegniamo a x dei valori che sono tra quelli che abbiamo già rappresentato, cioè tra 0 e 4.

X	y
1	$y = 4 \cdot 1 + 1 = 5$
4	$y = 4 \cdot 4 + 1 = 17$

I punti da rappresentare hanno coordinate A(1,5), B(4,17)



Calcoliamo infine l'indice di bontà di adattamento

$$R^2 = \frac{Dev(reg)}{Dev(tot)} = \frac{\sum_{i=1}^r (\hat{y}_i - \bar{y})^2 n_i}{\sum_{j=1}^c (y_j - \bar{y})^2 n_j}$$

Per calcolare \hat{y}_i si utilizza la retta di regressione, occorre sostituire i valori di x nella equazione $y = 4x + 1$

$$x_1 = 2 \rightarrow \hat{y}_1 = 4 \cdot 2 + 1 = 9$$

$$x_2 = 3 \rightarrow \hat{y}_2 = 4 \cdot 3 + 1 = 13$$

$$x_3 = 4 \rightarrow \hat{y}_3 = 4 \cdot 4 + 1 = 17$$

$$x_4 = 1 \rightarrow \hat{y}_4 = 4 \cdot 1 + 1 = 5$$

X_i	Y_i	$x_i \cdot y_i$	x_i^2	\hat{y}_i	$(\hat{y}_i - \bar{y})^2$	$(y_j - \bar{y})^2$
2	9	18	4	9	4	4
3	13	39	9	13	4	4
4	17	68	16	17	36	36
1	5	5	1	5	36	36
totali		130	30		80	80

Ora calcoliamo

$$R^2 = \frac{\sum_{i=1}^r (\hat{y}_i - \bar{y})^2 n_i}{\sum_{j=1}^c (y_j - \bar{y})^2 n_j} = \frac{80 \cdot 4}{80 \cdot 4} = 1$$