Sulla retta di regressione

applicata alla vendita di gelati

In una località turistica, nel corso di una settimana del mese di luglio, un gestore di un chiosco-bar ha registrato il numero di gelati venduti nell'orario di apertura. Per gli stessi giorni sono state registrate le temperature massime raggiunte nelle giornate.

I dati sono riportati nella tabella Tab.1

Tab.1	Alte temperature e gelati	
Giorno	Temperatura max	N. Gelati venduti
1	25	30
2	27	35
3	24	30
4	27	40
5	28	40
6	30	50
7	32	60

Quesiti

1) Indicando con x_i la temperatura massima raggiunta nella giornata i-sima ed y_i il numero di gelati venduti nella stessa giornata**, determinare l'equazione della retta di regressione di y su x**.
2) Calcolare il grado di accostamento tra i valori teorici calcolati $\hat{y}_i$ ed i valori effettivi y_i assumendo come indice il valore del **coefficiente di correlazione.**
3) Rappresentare in uno stesso riferimento cartesiano il grafico a dispersione delle coppie (x_i; y_i) e quello della retta di regressione.

Soluzione

1) Per analizzare i dati, e dovendo determinare la retta di regressione di Y su X, è necessario disporre in ordine crescente i valori x_i e conseguentemente disporre i corrispondenti valori y_i. Il lavoro necessario è eseguito con un foglio di lavoro con Excel; i dati elaborati sono riportati in due distinte tabelle.

	Elaborazioni per la retta di regressione			
	x_i	y_i	x_iy_i	x_i^2
	24	30	720	576
	25	30	750	625
	27	35	945	729
	27	40	1080	729
	28	40	1120	784
	30	50	1500	900
	32	60	1920	1024
Σ	193	285	8035	5367

Equazione della retta di regressione

Scrivendo la retta di regressione nella forma $y = bx + a$, i valori dei coefficienti *a* e *b* si determinano risolvendo il seguente sistema di equazioni

$$\begin{cases} \left(\sum_{i=1}^{n} x_i^2\right) b + \left(\sum_{i=1}^{n} x_i\right) a = \sum_{i=1}^{n} x_i y_i \\ \left(\sum_{i=1}^{n} x_i\right) b + na = \sum_{i=1}^{n} y_i \end{cases}$$

e si ottiene

$$b = \frac{n \cdot \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \quad a = \frac{\left(\sum_{i=1}^{n} x_i^2\right) \cdot \sum_{i=1}^{n} y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

L'equazione della retta di regressione lineare è: y=3,875x-66,125

2) Il **coefficiente di correlazione** tra le serie di dati X, Y è

$$r = \frac{Cov(X,Y)}{\sigma_X \sigma_X} = \frac{\sum_{i=1}^{n}(x_i - M_X)(y_i - M_Y)}{n\sqrt{\frac{\sum_{i=1}^{n}(x_i - M_X)^2}{n} \cdot \frac{\sum_{i=1}^{n}(y_i - M_Y)^2}{n}}}$$

Nella formula, M_X indica il valore medio dei dati x_i, M_Y il valore medio dei dati y_i, σ_X e σ_X sono gli scarti quadratici medi rispettivamente delle serie di dati X ed Y; $Cov(X,Y)$ rappresenta la covarianza tra le due serie di dati. Risulta:

$M_X = 27,57$; $M_Y = 40,71$;

$\sigma_X = 2,5555$; $\sigma_Y = 10,1519$;

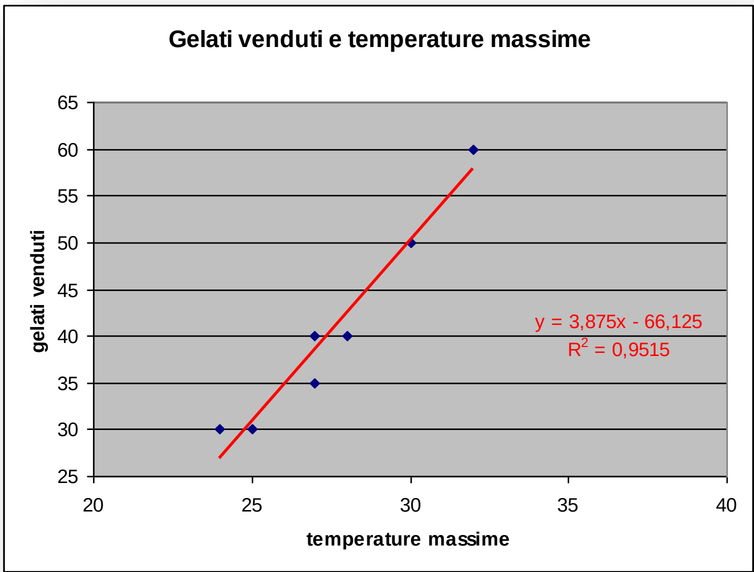
$$Cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i - M_X)(y_i - M_Y)}{n} = 25,3061$$

Elaborazioni per la Covarianza e gli scarti σ_X, σ_Y				
$x'_i=x_i-M_X$	$y'_i=y_i-M_Y$	$x'_iy'_i$	$(x_i-M_X)^2$	$(y_i-M_Y)^2$
-3,57143	-10,7143	38,26531	12,7551	114,7959
-2,57143	-10,7143	27,55102	6,612245	114,7959
-0,57143	-5,71429	3,265306	0,326531	32,65306
-0,57143	-0,71429	0,408163	0,326531	0,510204
0,428571	-0,71429	-0,30612	0,183673	0,510204
2,428571	9,285714	22,55102	5,897959	86,22449
4,428571	19,28571	85,40816	19,61224	371,9388
		Σ=177,1429	Σ=45,71429	Σ=721,4286

A margine sono riportate le elaborazioni per il calcolo del coefficiente di correlazione. Il valore del coefficiente di correlazione è

$$r = \frac{Cov(X,Y)}{\sigma_X \sigma_X} = 0,9754$$

Il valore ottenuto è molto prossimo ad uno. Ricordiamo che si ha **perfetta correlazione lineare positiva tra le serie X, Y se il coefficiente di correlazione vale 1**, mentre si ha perfetta correlazione lineare negativa se il coefficiente di correlazione vale -1. Nel caso studiato le due serie di dati sono molto ben correlate positivamente.



3) Le rappresentazioni grafiche richieste sono riportate a margine.

Applicazione

Il gestore del chiosco-bar, se è a conoscenza che per un determinato giorno del periodo estivo al quale si riferiscono i dati registrati per le temperature massime il valore della temperatura massima prevista sarà 31 °C, può ragionevolmente ritenere che venderà un numero di gelati pari al valore teorico che si ottiene applicando l'equazione della retta di regressione lineare, cioè:

$$y(31°C) = 3{,}875 \cdot 31 - 66{,}125 = 54$$