

Esercizio 1

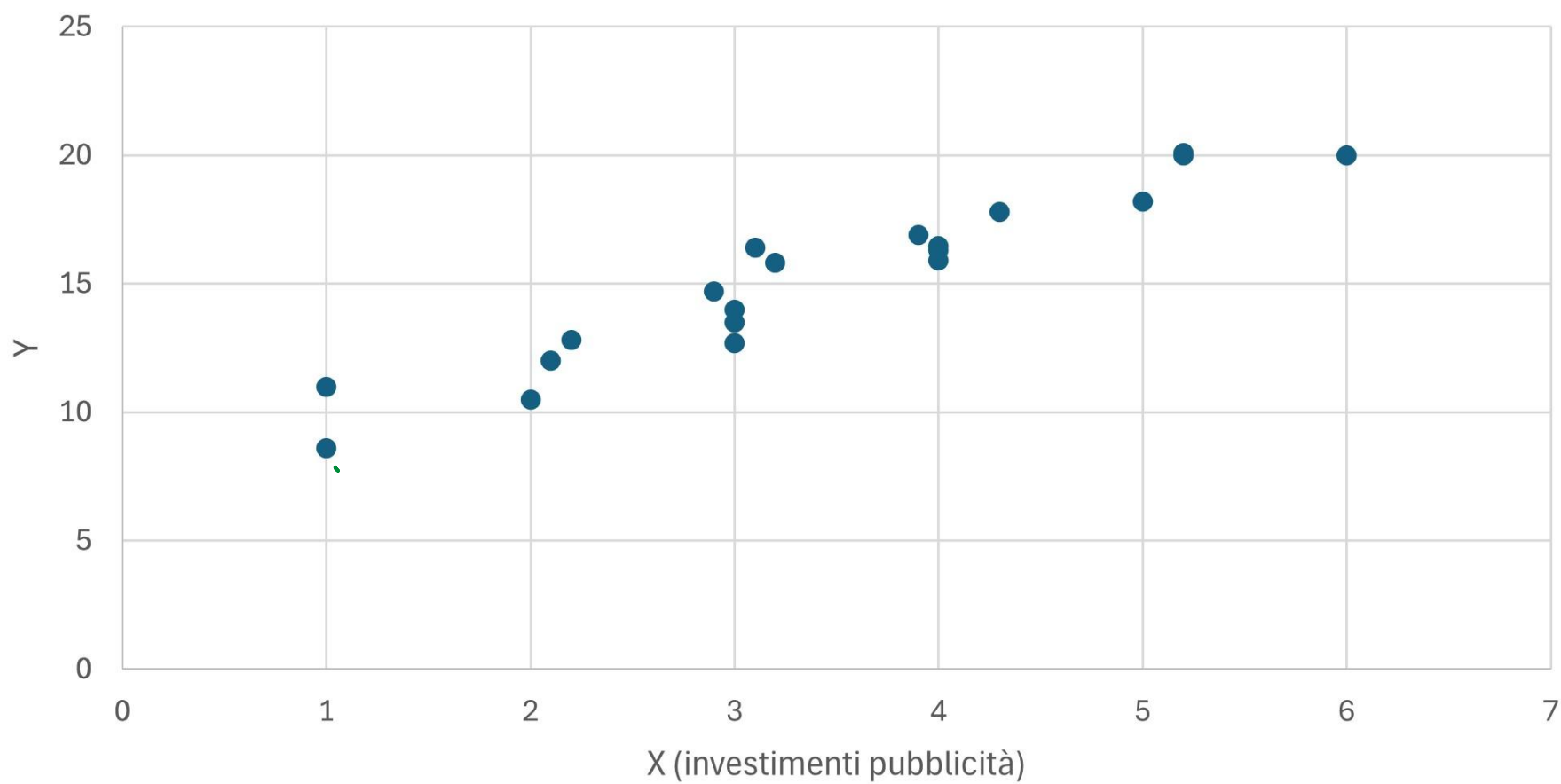
Una grande società che possiede una catena di negozi situati in diverse città italiane vuole valutare l'impatto avuto da un investimento in pubblicità su quotidiani locali e attraverso volantini fatto nel mese di settembre. A tal fine seleziona un campione casuale di 20 negozi, situati in città simili per dimensione demografica e per altre caratteristiche che possono influenzare il comportamento d'acquisto. Per ciascun negozio vengono rilevati: investimento in pubblicità (X) e volume delle vendite (Y) fra il 20 settembre e il 30 novembre

Città (codice)	X (investimenti in pubblicità)	Y (volume vendite)
1	1	11
2	1	8,6
3	2	10,5
4	2,1	12
5	2,2	12,8
6	2,9	14,7
7	3	13,5
8	3	14
9	3	12,7
10	3,1	16,4
11	3,2	15,8
12	3,9	16,9
13	4	16,3
14	4	15,9
15	4	16,45
16	4,3	17,8
17	5	18,2
18	5,2	20
19	5,2	20,1
20	6	20

		X	Y	X ²	y ²	xy
	1	1	11	1	121	11
	2	1	8,6	1	73,96	8,6
	3	2	10,5	4	110,25	21
	4	2,1	12	4,41	144	25,2
	5	2,2	12,8	4,84	163,84	28,16
	6	2,9	14,7	8,41	216,09	42,63
	7	3	13,5	9	182,25	40,5
	8	3	14	9	196	42
	9	3	12,7	9	161,29	38,1
	10	3,1	16,4	9,61	268,96	50,84
	11	3,2	15,8	10,24	249,64	50,56
	12	3,9	16,9	15,21	285,61	65,91
	13	4	16,3	16	265,69	65,2
	14	4	15,9	16	252,81	63,6
	15	4	16,45	16	270,6025	65,8
	16	4,3	17,8	18,49	316,84	76,54
	17	5	18,2	25	331,24	91
	18	5,2	20	27,04	400	104
	19	5,2	20,1	27,04	404,01	104,52
	20	6	20	36	400	120
totale		68,1	303,65	267,29	4814,083	1115,16
media		3,405	15,1825	13,365	240,7041	55,758

- a) Valutare graficamente il legame tra X e Y
- b) Quantificare la dipendenza lineare attraverso il coefficiente di correlazione lineare.
- c) Stimare i parametri β_0 e β_1 con gli stimatori dei Minimi Quadrati.
- d) Sulla base dei risultati ottenuti, formulare una valutazione riguardo il volume delle vendite che ci si aspetta di ottenere in corrispondenza di un investimento in pubblicità pari a 4500€, e dell'investimento in pubblicità che sarebbe stato necessario per ottenere un volume delle vendite pari a circa 10000€
- e) Rappresentare graficamente la retta stimata
- f) Calcolare la somma dei residui
- g) Stimare la varianza degli errori
- h) Ipotizzando che gli errori abbiano distribuzione Gaussiana, calcolare gli intervalli di confidenza al 95% per β_0 e β_1

Volume vendita



a) Valutare graficamente il legame tra X e Y

Fra i dati prevale concordanza: a valori più alti della media per una variabile si accompagnano prevalentemente valori più alti della media per l'altra, e a valori più bassi della media per una variabile si accompagnano prevalentemente valori più bassi della media per l'altra.

b) Quantificare la dipendenza lineare attraverso il coefficiente di correlazione lineare.

b)

ρ

$$\begin{aligned}\text{Cov}(X, Y) &= 55.578 - 3.405 \cdot 15.1825 \\ &= 4.068\end{aligned}$$

$$\text{Var}(X) = 13.365 - 3.405^2 = 1.77$$

$$\begin{aligned}\text{Var}(Y) &= 240.7041 - 15.1825^2 \\ &= 10.19\end{aligned}$$

$$\rho = \frac{4.069}{\sqrt{1.77 \cdot 10.12}} = 0.958$$

$$\sum_{i=1}^n x_i y_i = 1115.16 \quad \sum_{i=1}^n x_i^2 = 267.29 \quad \sum_{i=1}^n y_i^2 = 4814.08$$

$$\bar{X} = 3.41 \quad \bar{Y} = 15.18$$

$$Dev(X) = \sum_{i=1}^n x_i^2 - n\bar{X}^2 = 267.29 - 20 \times 3.41^2 = 35.41$$

$$Dev(Y) = \sum_{i=1}^n y_i^2 - n\bar{Y}^2 = 4814.08 - 20 \times 15.18^2 = 203.92$$

$$Cod(X, Y) = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} = 1115.16 - 20 \times 3.41 \times 15.18 = 81.23$$

$$\rho_{XY} = \frac{Cod(X, Y)}{\sqrt{Dev(X)Dev(Y)}} = \frac{81.23}{\sqrt{35.41 \times 203.92}} = 0.96$$

Si riscontra un'elevata correlazione lineare positiva (il valore del coefficiente di correlazione è positivo e molto vicino a 1).

c) Stimare i parametri β_0 e β_1 con gli stimatori dei Minimi Quadrati.

Le espressioni da usare sono:

$$\hat{\beta}_1 = \frac{Cod(X, Y)}{Dev(X)} = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \quad \text{e} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

e le stime, utilizzando i calcoli fatti al punto b), risultano essere

$$\hat{\beta}_1 = \frac{Cod(X, Y)}{Dev(X)} = \frac{81.23}{35.41} = 2.29 \quad \text{e} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 15.18 - 2.29 \times 3.41 = 7.37$$

Il coefficiente di regressione è positivo; all'aumentare dell'investimento in pubblicità il volume delle vendite aumenta. In particolare, ad un aumento unitario (1000€) dell'investimento in pubblicità il volume delle vendite aumenta di 2290€.

$$\hat{\beta}_1 = \frac{\text{Cov}(x, y)}{\text{Var } x}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{4.069}{1.77} = 2.29$$

$$\hat{\beta}_0 = 15.1825 - 2.29 \cdot 3.405 = 7.34$$

zette

$$y = 7.37 + 2.29x$$

$$\hat{y}_i = 7.37 + 2.29x_i$$

$$y_i = 7.37 + 2.29x_i + \epsilon_i$$

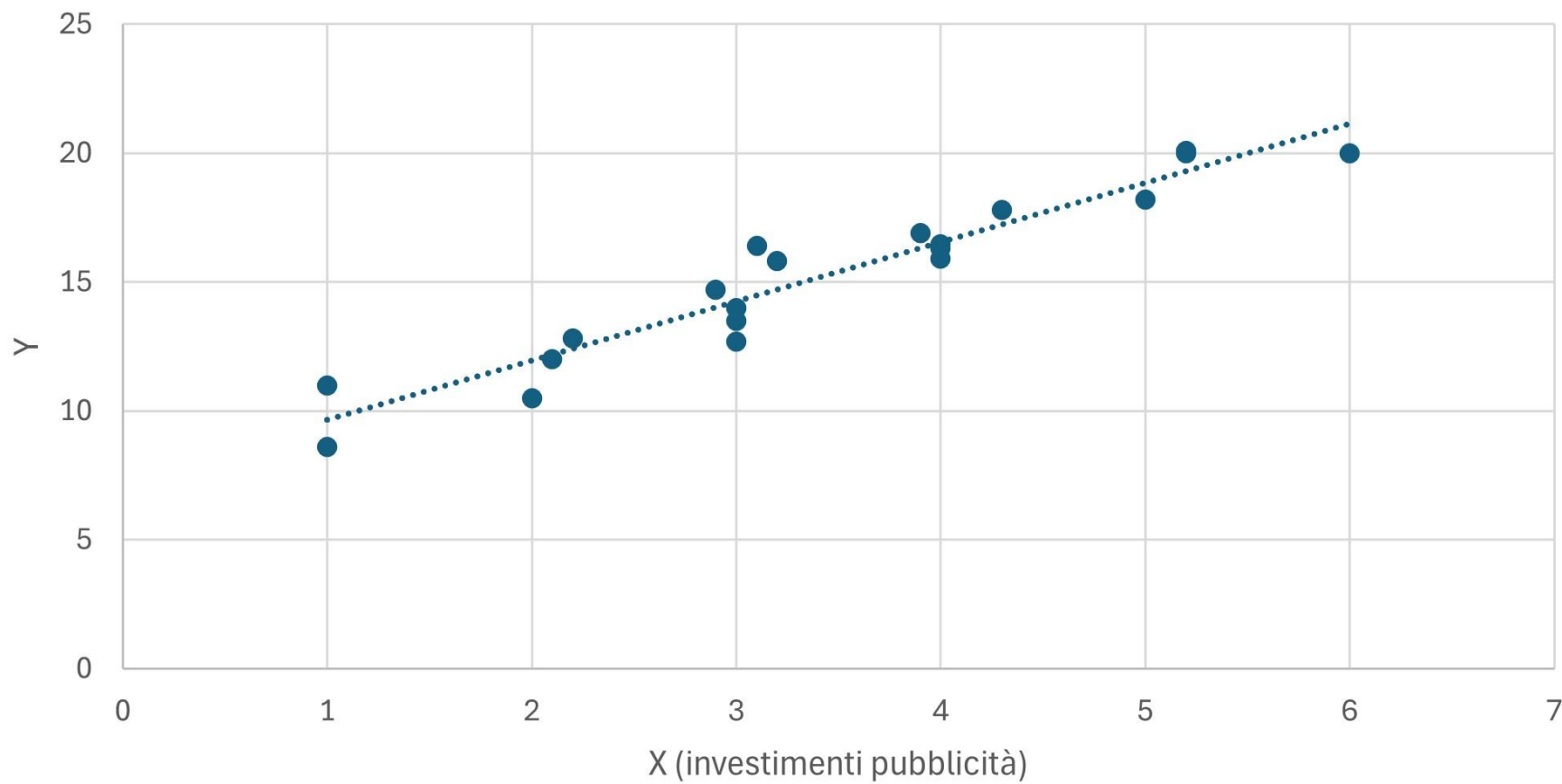
d) Sulla base dei risultati ottenuti, formulare una valutazione riguardo il volume delle vendite che ci si aspetta di ottenere in corrispondenza di un investimento in pubblicità pari a 4500€, e dell'investimento in pubblicità che sarebbe stato necessario per ottenere un volume delle vendite pari a circa 10000€

$$\hat{y} = 7.37 + 2.29 \times 4.5 = 17.675$$

$$10 = 7.37 + 2.29 \times x$$

$$x = \frac{10 - 7.37}{2.29} = 1,15$$

Volume vendita



- e) Rappresentare graficamente la retta stimata
- f) Calcolare la somma dei residui

Somma dei residui è nulla!

$$\varepsilon_i = y_i - \hat{y}_i$$

$$\sum \varepsilon_i = 0$$

$$(1 - R^2) n \text{Var} y = \sum_{i=1}^n \varepsilon_i^2$$

$$\text{Var}(\varepsilon_i) = \frac{\sum \varepsilon_i^2}{n-2}$$

Stimare la varianza degli errori

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n Y_i^2 - n\bar{Y}^2 = 4814,083 - 20 \times 15,1825^2 = 203.9164$$

$$R^2 = 0.91$$

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = (1 - 0.91) \times \sum_{i=1}^n (Y_i - \bar{Y})^2 = (1 - 0.91) \times 203.9164 = 18.3525$$

$$\hat{\sigma}^2 = \frac{18.3525}{18} = 1,02$$

Ipotizzando che gli errori abbiano distribuzione Gaussiana, calcolare gli intervalli di confidenza al 95% per β_0 e β_1

$$\begin{aligned} Var(\hat{\beta}_1) &= \sigma^2 \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} & \hat{Var}(\hat{\beta}_1) &= \hat{\sigma}^2 \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ Var(\hat{\beta}_0) &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) & \hat{Var}(\hat{\beta}_0) &= \\ &= \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \end{aligned}$$

$$\hat{Var}(\hat{\beta}_1) = 1.092 \frac{1}{35.41} = 0.0288$$

$$\hat{Var}(\hat{\beta}_0) = 1.092 \left(\frac{1}{20} + \frac{3.41^2}{35.41} \right) = 0.4132$$

Ipotizzando che gli errori abbiano distribuzione Gaussiana, calcolare gli intervalli di confidenza al 95% per β_0 e β_1

$$\hat{Var}(\hat{\beta}_1) = 1.092 \frac{1}{35.41} = 0.0288 \quad \hat{Var}(\hat{\beta}_0) = 1.092 \left(\frac{1}{20} + \frac{3.41^2}{35.41} \right) = 0.4132$$

$$t_{0.025}^{18} = 2.1009$$

$$\begin{aligned} 95\% &= \left[\hat{\beta}_1 - t_{0.025}^{18} \sqrt{\hat{Var}(\hat{\beta}_1)}; \hat{\beta}_1 + t_{0.025}^{18} \sqrt{\hat{Var}(\hat{\beta}_1)} \right] = [2.29 \pm 2.1009 \times 0.1687] \\ &= [1.933; 2.646] \end{aligned}$$

$$\begin{aligned} 95\% &= \left[\hat{\beta}_0 - t_{0.025}^{18} \sqrt{\hat{Var}(\hat{\beta}_0)}; \hat{\beta}_0 + t_{0.025}^{18} \sqrt{\hat{Var}(\hat{\beta}_0)} \right] = [7.37 \pm 2.1009 \times 0.6428] \\ &= [6.0195; 8.72] \end{aligned}$$

Esercizio 2

Con riferimento ad un collettivo di 30 famiglie sulle quali sono state rilevate le variabili: spesa per consumi (Y in migliaia di €) e reddito (X in migliaia di €) ed è stato trovato:

$$\sum_{i=1}^n x_i = 75 \quad \sum_{i=1}^n y_i = 60 \quad \sum_{i=1}^n x_i^2 = 247.5$$

$$\sum_{i=1}^n y_i^2 = 390 \quad \sum_{i=1}^n x_i y_i = 189$$

- calcolare i parametri della retta di regressione della spesa (Y) sul reddito (X) e spiegarne il significato;

- calcolare i parametri della retta di regressione della spesa (Y) sul reddito (X) e spiegarne il significato;

$$\hat{\beta}_1 = \frac{Cov(X, Y)}{Var(X)} = \frac{\frac{\sum_{i=1}^n x_i y_i}{n} - \frac{\sum_{i=1}^n x_i}{n} \frac{\sum_{i=1}^n y_i}{n}}{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2} = \frac{6.3 - 2.5 \times 2}{8.25 - 6.25} = \frac{1.3}{2} = 0.65$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2 - 0.65 \times 2.5 = 0.375$$

- calcolare un opportuno indice di bontà di adattamento del modello ai dati e commentare il risultato;

- calcolare un opportuno indice di bontà di adattamento del modello ai dati e commentare il risultato;

$$R^2 = \rho^2 = \frac{Cov^2(X, Y)}{Var(Y)Var(X)} = \frac{1.3^2}{9 \times 2} = 0.093 = 0.306^2$$

$$Var(y) = \frac{\sum_{i=1}^n y_i^2}{n} - \bar{y}^2 = 13 - 4 = 9$$

- secondo il modello, qual è la spesa di una famiglia con reddito pari a 2?

- secondo il modello, qual è la spesa di una famiglia con reddito pari a 2?

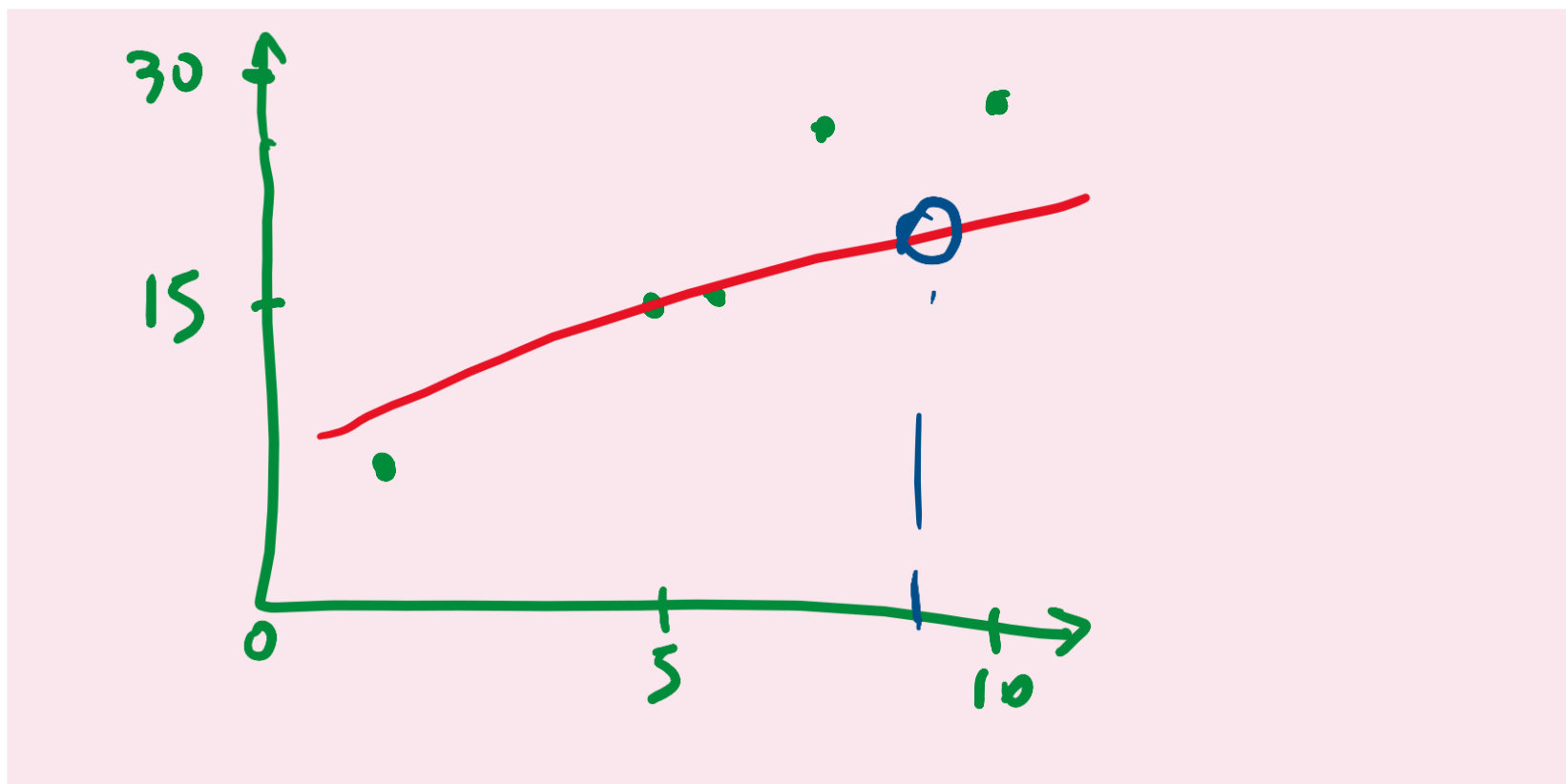
$$\hat{y} = 0.65 \times 2 + 0.375 = 1.675$$

Esercizio 3

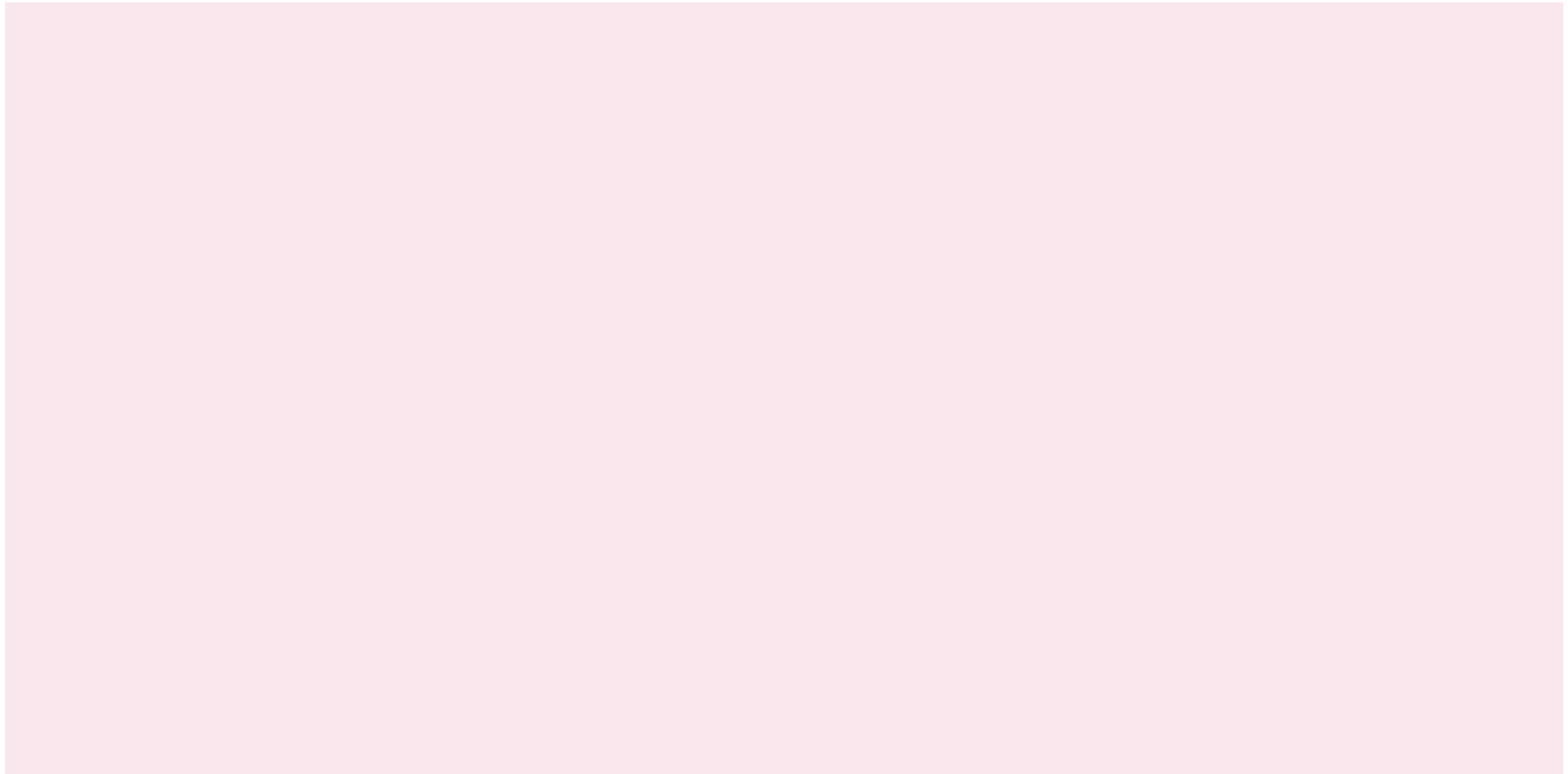
Per cinque laureati alla Facoltà di Economia si sono osservati gli anni trascorsi dalla Laurea (X) e il reddito annuale lordo (Y) misurato in migliaia euro. I risultati sono riassunti dalla seguente tabella:

Ex- Studente	A	B	C	D	E
X	2	10	5	7	6
Y	5	30	15	25	15

a) disegnare il grafico della nuvola dei punti e la retta dei minimi quadrati;



a) Calcolare i parametri della retta di regressione del reddito annuale lordo Y sugli anni dalla laurea (X) e commentare i risultati.



	x	y	x²	y²	xy
	2	5	4	25	10
	10	30	100	900	300
	5	15	25	225	75
	7	25	49	625	175
	6	15	36	225	90
Somma	30	90	214	2000	650
media	6	18	42,8	400	130

cov	22
var x	6,8
var y	76
β_1	3,24
β_0	-1,41
rho	0,97

	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	
	2	5	-4	-13	16	169	52
	10	30	4	12	16	144	48
	5	15	-1	-3	1	9	3
	7	25	1	7	1	49	7
	6	15	0	-3	0	9	0
Somma	30	90			34	380	110
media	6	18			6.8	76	22

cov	22
var x	6,8
var y	76
β_1	3,24
β_0	-1,41
rho	0,97

b) Calcolare i parametri della retta di regressione del reddito annuale lordo Y sugli anni dalla laurea (X) e commentare i risultati.

$$Y_i = -1.91 + 3.24X_i + \varepsilon_i$$

c) Si calcoli il coefficiente di determinazione R^2

$$R^2 = \frac{22^2}{68.76} = 0.936$$

d) Quale vi aspettate che sia il reddito annuale lordo di una studente a 8 anni dalla laurea

$$-1.41 + 8 \cdot 3.24 = 24.91$$

e) Valutare la somma dei residui e la somma dei residui al quadrato

somme residui: nullo

$$\sum_i e_i^2 = (1 - R^2) n \text{ Var } Y$$

$$(1 - 0.936) \cdot 5.76 = 0.32$$

Esercizio 3 In uno studio sul consumo, sono stati rilevati la spesa mensile in generi alimentari e il reddito mensile di 24 famiglie. Indicando con X il reddito e con Y la spesa, sono stati ottenuti i valori seguenti per le variabili espresse in migliaia di euro:

$$\sum_{i=1}^{24} x_i = 42.8; \quad \sum_{i=1}^{24} y_i = 12.4; \quad \sum_{i=1}^{24} x_i^2 = 81.5; \quad \sum_{i=1}^{24} y_i^2 = 6.72, \quad \sum_{i=1}^{24} x_i y_i = 23.27$$

- a) Stimare la retta di regressione $Y_i = \alpha + \beta x_i + \varepsilon_i$
- b) Qual è la variazione stimata del valore atteso di Y se X aumenta di un'unità?
Se una famiglia non avesse reddito, quale sarebbe il valore atteso di y_i stimato?
- c) Stimare la varianza degli errori
- d) Calcolare l'indice di determinazione
- e) Assumendo che $\varepsilon \sim N(0, \sigma^2)$, sottoporre a test l'ipotesi nulla $\beta = 0$ al livello di significatività del 5%

3. Sono stati intervistati 250 studenti ammessi al Corso di Laurea in Medicina ed è stato chiesto il punteggio ottenuto al test di entrata (Y) e il voto all'eame di Maturità (X) e sono stati ottenuti i seguenti risultati:

$$\sum_{i=1}^{250} x_i = 20850 \quad \sum_{i=1}^{250} y_i = 15250 \quad \sum_{i=1}^{250} x_i^2 = 1918750 \quad \sum_{i=1}^{250} y_i^2 = 979950 \quad \sum_{i=1}^{250} x_i y_i = 1320000$$

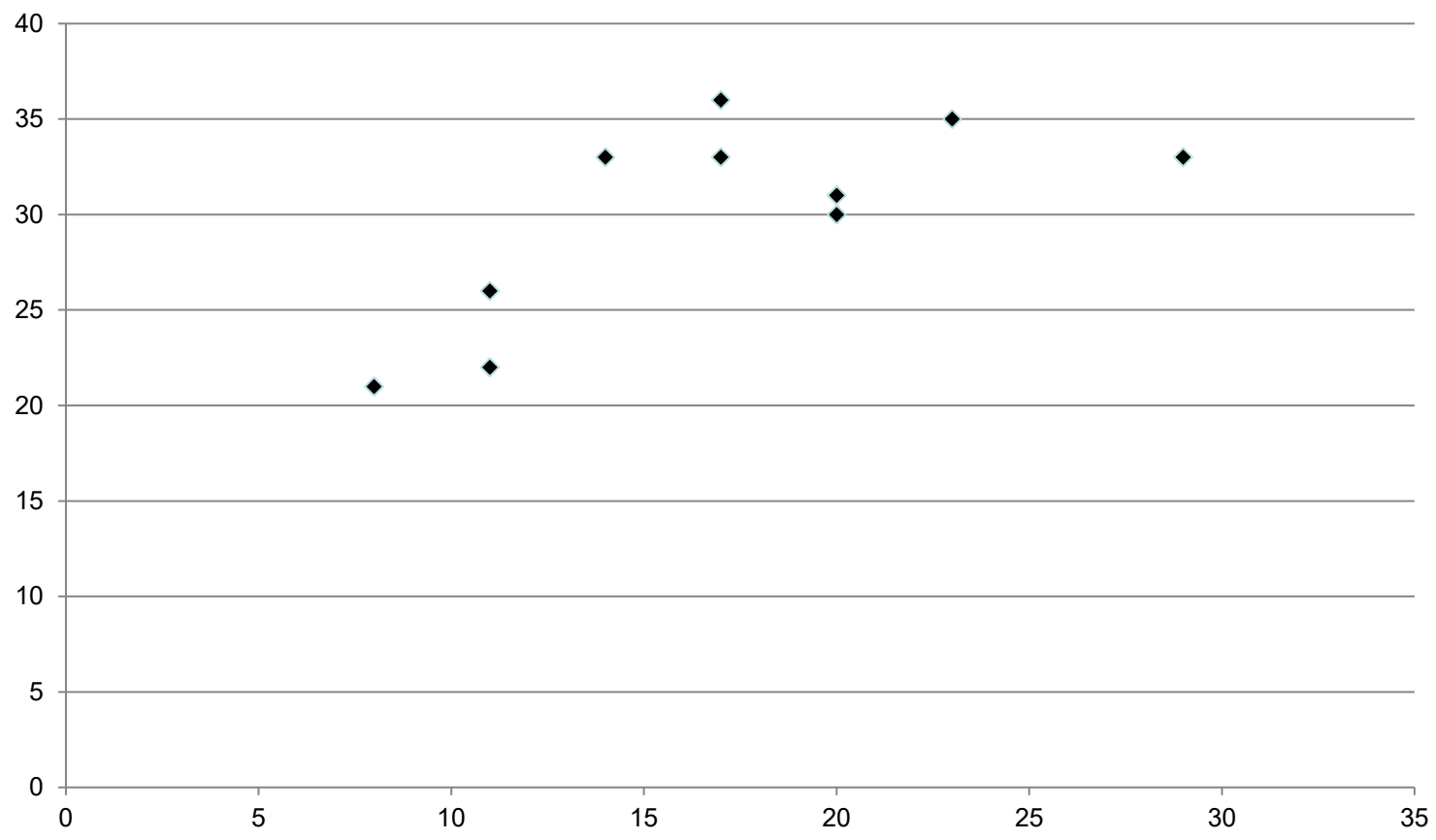
- (a) Calcolare i parametri della retta di regressione del punteggio ottenuto al test di ingresso (Y) sul voto di maturità (X) e spiegarne il significato (**2 punti**).
- (b) Si calcoli il coefficiente di determinazione (**2 punti**).
- (c) Valutare la somma dei residui (**1 punto**).
- (d) Valutare la somma dei residui al quadrato (**1 punto**).
- (e) Secondo il modello, qual è il punteggio atteso per uno studente che ha concluso il Liceo con un voto uguale a 70 alla Maturità? (**1 punto**)
- (f) E' possibile indicare quale dei due caratteri presenta maggiore variabilità? (**1 punto**)

Esempio

I dati sottostanti mostrano i voti ottenuti da 10 studenti nei test di Matematica e Fisica

Pupil	A	B	C	D	E	F	G	H	L	M
Maths	20	23	8	29	14	11	11	20	17	17
Physics	30	35	21	33	33	26	22	31	33	36

	x	y	x²	y²	xy
	20	30	400	900	600
	23	35	529	1225	805
	8	21	64	441	168
	29	33	841	1089	957
	14	33	196	1089	462
	11	26	121	676	286
	11	22	121	484	242
	20	31	400	961	620
	17	33	289	1089	561
	17	36	289	1296	612
sum	170	300	3250	9250	5313
mean	17	30	325	925	531,3



$$Var(X) = 325 - 17^2 = 36$$

$$Var(Y) = 925 - 30^2 = 25$$

$$Cov(X, Y) = 531.3 - 17 \times 30 = 21.3$$

$$Corr(X, Y) = \frac{21.3}{\sqrt{36 \times 25}} = 0.71$$

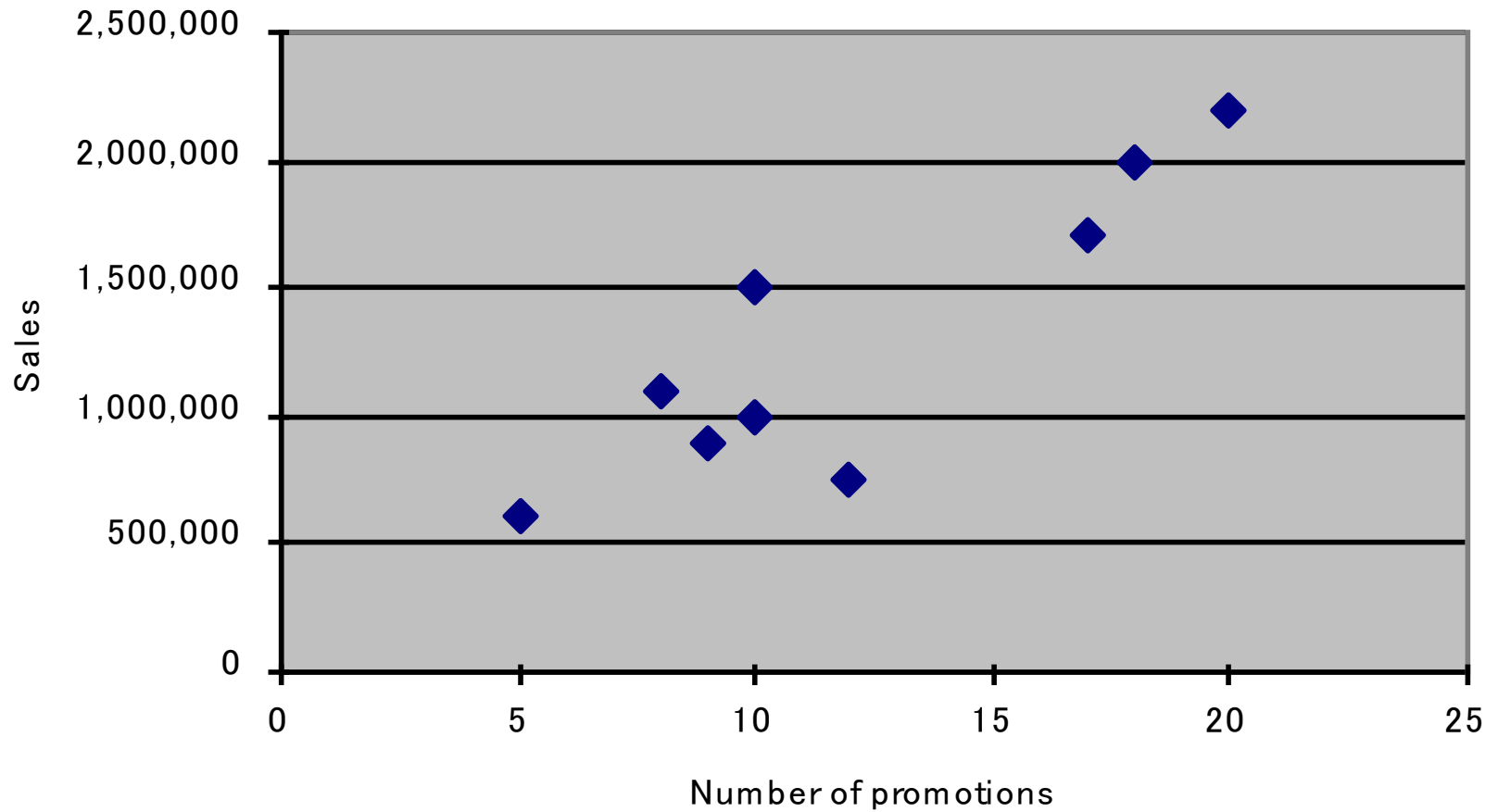
$$\beta_1 = \frac{21.3}{36} = 0.59$$

$$\beta_0 = 30 - 17 \times 0.59 = 19.94$$

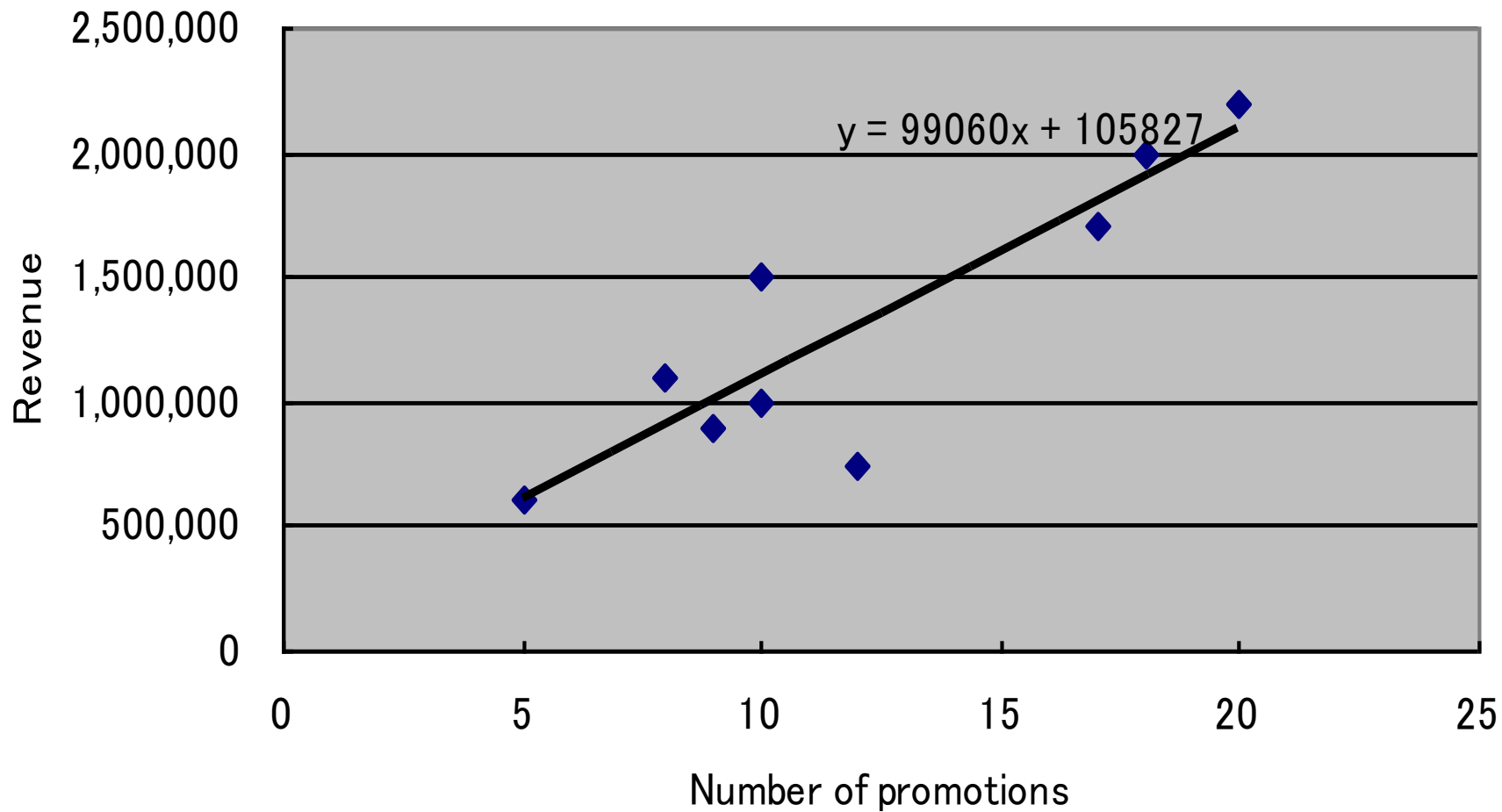
Come misurare la direzione della relazione: Esempio

Numero di promozioni	Ricavi dal prodotto A (in 1000 yen)
5	600
10	1000
8	1100
9	900
10	1500
12	750
20	2200
18	2000
17	1700

Product A: Promotion and sales



Product A: Promotion and sales



ESERCIZIO

Con riferimento ad un campione di 40 aziende, si sono state osservate le seguenti statistiche campionarie relative alle vendite in migliaia di euro (Y) ed alle spese in marketing in migliaia di euro (X):

$$\sum_{i=1}^n y_i = 520 \quad \sum_{i=1}^n y_i^2 = 8654 \quad \sum_{i=1}^n x_i = 100.5 \quad \sum_{i=1}^n x_i^2 = 544.5 \quad \sum_{i=1}^n x_i y_i = 2020.8$$

$$\beta_0 \quad \beta_1$$

$$\Rightarrow \quad \bar{x} = \frac{\sum x_i}{n} = \frac{100.5}{40} = 2.5125 \quad \bar{y} = \frac{\sum y_i}{n} = \frac{520}{40} = 13$$

$$\beta_1 = \frac{Cov(x, y)}{Var(x)} = \frac{\frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x}\bar{y}}{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2} = \frac{\frac{2020.8}{40} - 2.5125 \times 13}{\frac{544.5}{40} - 2.5125^2}$$

$$\beta_1 = \frac{17.8575}{7.2998} = 2.4463$$

$$\beta_0 \quad \beta_1$$



$$\bar{x} = \frac{\sum x_i}{n} = \frac{100.5}{40} = 2.5125$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{520}{40} = 13$$

$$\beta_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 13 - 2.4463 \times 2.5125 = 6.8537$$

$$\beta_1 = \frac{17.8575}{7.2998} = 2.4463$$

$$R^2 = \frac{Cov(x, y)^2}{Var(x)Var(y)} = \frac{17.8575^2}{7.2998 \times \left(\frac{\sum_{i=1}^n y_i^2}{n} - \bar{y}^2 \right)}$$

$$R^2 = \frac{17.8575^2}{7.2998 \times \left(\frac{8654}{40} - 13^2 \right)} = \frac{17.8575}{7.2998 \times 46.35} = 0.9226$$