Introduzione

Domanda 1

Le vendite dei jeans sono legate all'età dei consumatori?

Domanda 2

La spesa per consumi di una famiglia è legata al numero di componenti della famiglia?

Domanda 3

Le vendite di costumi da bagno sono legate alle condizioni metereologiche delle località di acquisto?

Domanda 4

Le vendite nel mercato automobilistico sono legate alle vendite nel settore dell'alta moda?

Domanda 5

I costi del settore della produzione sono legati alla redditività del settore vendite?

CONCORDANZA/DISCORDANZA

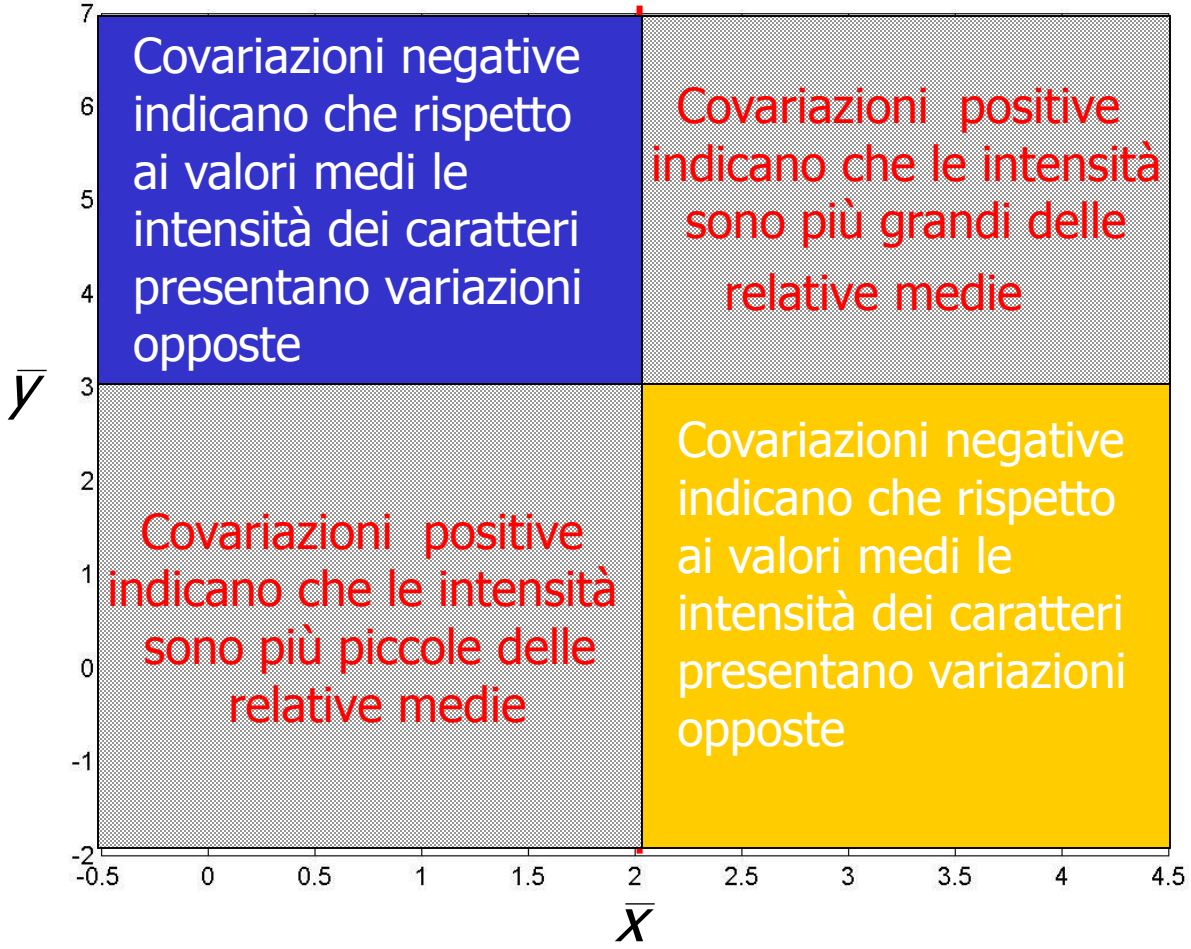
L'analisi della correlazione tra due variabili X e Y implica il calcolo delle seguenti quantità:

- Valor medio di ciascuna variabile → $\bar{x} \quad \bar{y}$
- Varianza di ciascuna variabile → $\sigma_x^2 \quad \sigma_y^2$
- Covarianza →

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

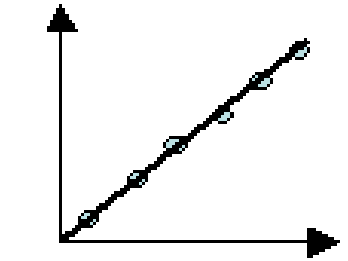
- *algebricamente si esprime come la media (Σ/n) dei prodotti degli scarti delle variabili dalle rispettive medie*
- *È una misura della variabilità congiunta di due caratteri (X e Y) rispetto alle proprie medie (MISURA DELLA CONCORDANZA)*

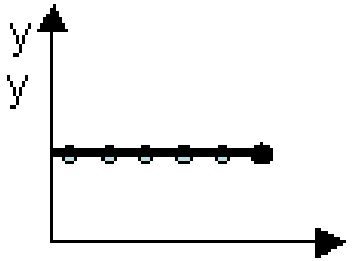
Graficamente ....

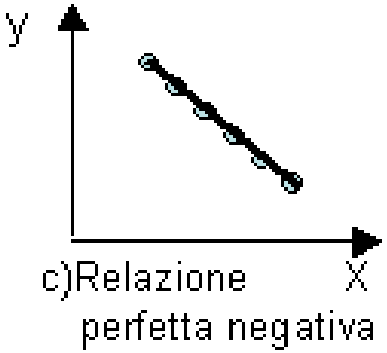


Gli scarti presentano segno positivo quando le intensità (modalità) presentano, rispetto ai valori medi, variazioni analoghe (entrambe maggiori o minori delle medie!)

Relazioni tra variabili







Codevianza e covarianza

- La somma dei prodotti degli scarti prende il nome di *codevianza*

$$Cod(X,Y)=\sum_{i=1}^{N}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)$$

- Dividendo per N si avrà la *covarianza*

$$Cov(X,Y)=\frac{1}{N}\sum_{i=1}^{N}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)$$

scarti fornisce la variabilità congiunta: …ricordo è necessario calcolare i prodotti degli scarti…??

Positivo (“+” x “+” o “-” x “-”) **indica che le variabili presentano modalità che si comportano nello stesso modo rispetto alla media (entrambe più alte o più basse)**

Negativo (“+” x “-”) **indica che le variabili presentano modalità che si comportano in modo diverso rispetto alla media (se una è più grande, l’altra è più bassa della media)**

	n. esami sostenuti X	ore/giorno in palestra Y	scarti x	scarti y	prodotti tra scarti	
	6	3	6-6 =0	3-4=-1	-	0
	4	3	4-6=-2	3-4=-1	+	2
	3	8	3-6=-3	8-4=4	- →	-12
	8	4	2	0	+	0
	10	5	4	1	+	4
	5	1	-1	-3	+	3
media	6	4				

...e poi...perché è necessario calcolare la media dei prodotti degli scarti...???

...è necessario calcolare la media perché da un certo numero di scarti si vuole ricavare un'unica misura che sintetizzi tutti gli scarti...con i rispettivi segni (positivi e negativi)!!!

	n. esami sostenuti X	ore/giorno in palestra Y	scarti x	scarti y	prodotti tra scarti
	6	3	6-6 =0	3-4=-1	0
	4	3	4-6=-2	3-4=-1	2
	3	8	3-6=-3	8-4=4	-12
	8	4	2	0	0
	10	5	4	1	4
	5	1	-1	-3	3
media	6	4			-0.5

COVARIANZA

Codevianza e covarianza

La covarianza è positiva (>0), se nella somma prevalgono prodotti di scarti con lo stesso segno (i cui risultati sono +)

I caratteri *tendono ad assumere valori concordanti* all'aumentare (diminuire) di un carattere aumenta (diminuisce) anche l'altro; cioè, entrambi assumono valori maggiori (minori) delle proprie medie

Esempio:

all'aumentare (diminuire) della cilindrata aumenta (diminuisce) il prezzo delle automobili

Codevianza e covarianza

La covarianza è negativa (<0), se nella somma prevalgono prodotti di scarti di segno opposto (i cui risultati sono -)

i caratteri *tendono ad assumere valori discordanti*:

all'aumentare (diminuire) di un carattere, l'altro diminuisce (aumenta); cioè, mentre uno assume valori maggiori (minori) della propria media, l'altro assume valori minori (maggiori)

Esempio: all'aumentare della cilindrata diminuiscono i km/litro

- con una cilindrata 1100 corrispondono 8 km/litro
- con una cilindrata 1500 corrispondono 5 km/litro

Codevianza e covarianza

La covarianza è nulla (=0), quando non si verifica alcuna relazione tra i caratteri, ossia alle variazioni di un carattere non corrisponde nessuna variazione dell'altro carattere

Esempio :

- al variare del tempo impiegato per raggiungere una determinata località il numero dei km percorsi non varia.

STEP per stimare la covarianza:

a) **(SCATTER-PLOT)**
b) **Calcolo media della variabile X**
c) **Calcolo media della variabile Y**
d) **Calcolo scarti dalla media per la variabile X**
e) **Calcolo scarti dalla media per la variabile Y**
f) **Prodotto tra coppie di scarti per ogni modalità di X e di Y**

   ricordando che: ***(-) *(-)=(+)*(+)="+"*** ***e*** ***(-) *(+)="-"***

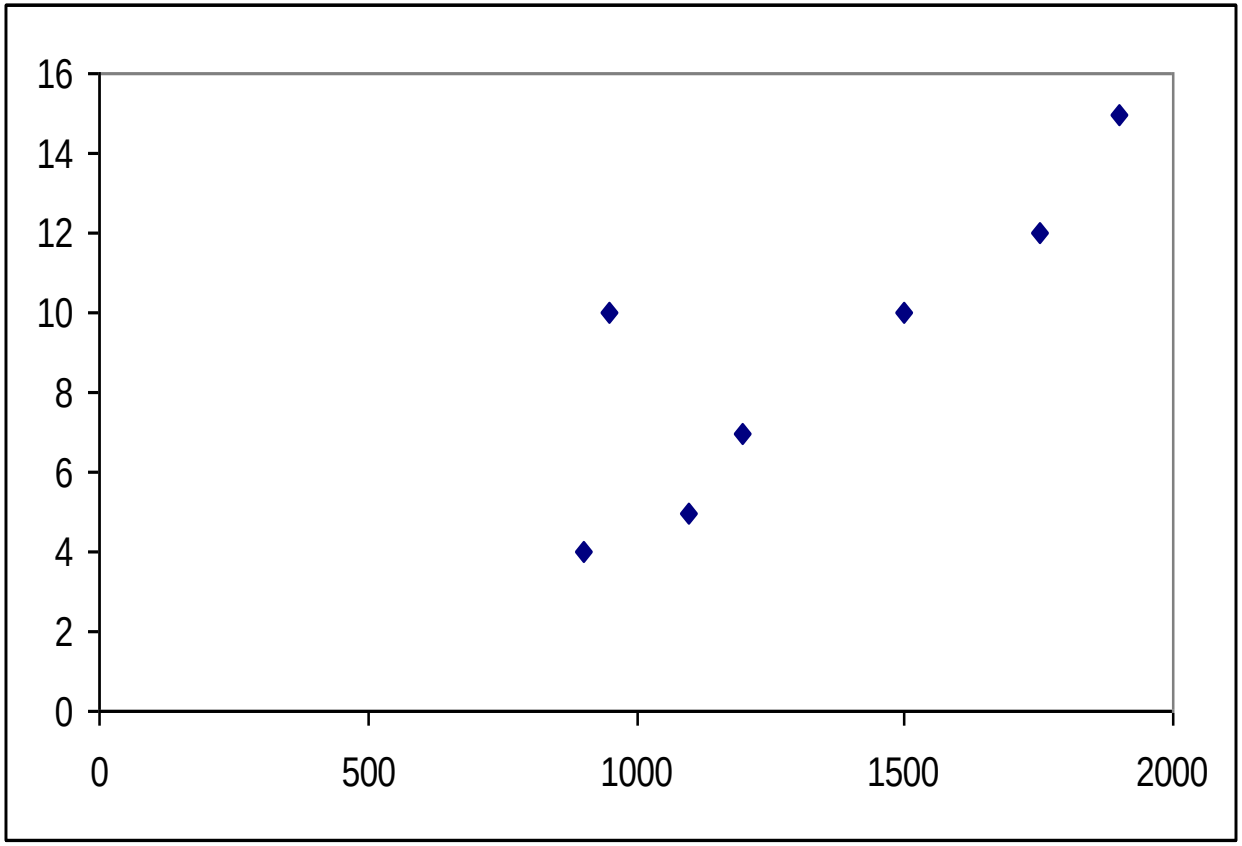
g) **Somma dei prodotti ottenuti (CODEVIANZA)**
h) **Rapporto tra codevianza e N (COVARIANZA)**
i) **Interpretazione risultati con commenti**

Esercizio 1

Si richiede di verificare, in un collettivo di 7 auto, se esiste covariazione tra la cilindrata ed i consumi (litri/100km)

Id	Cc	L/100km
1	900	4
2	950	10
3	1,100	5
4	1,200	7
5	1,500	10
6	1,750	12
7	1,900	15
totale	9,300	63
medie	1,329	9

Scatterplot



Esercizio 1

		x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$
	Id	Cc	L/100km		
	1	900	4	-428.6	-5
	2	950	10	-378.6	1
	3	1,100	5	-228.6	-4
	4	1,200	7	-128.6	-2
	5	1,500	10	171.4	1
	6	1,750	12	421.4	3
	7	1,900	15	571.4	6
totale		9,300	63		
Medie		1,329	9		

$$\bar{x} = \frac{9{,}300}{7} = 1{,}328.6 \qquad \bar{y} = \frac{63}{7} = 9$$

$$x_1 - \bar{x} = 900 - 1{,}328.6 = -428.6 \qquad y_1 - \bar{y} = 4 - 9 = -5$$

Esercizio 1

		x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
	Id	Cc	L/100km			
	1	900	4	-428.57	-5	2,142.86
	2	950	10	-378.57	1	-378.57
	3	1,100	5	-228.57	-4	914.29
	4	1,200	7	-128.57	-2	257.14
	5	1,500	10	171.43	1	171.43
	6	1,750	12	421.43	3	1,264.29
	7	1,900	15	571.43	6	3,428.57
Totale		**9,300**	**63**			**7,800**
Medie		**1,328.6**	**9.00**			**1,114.29**

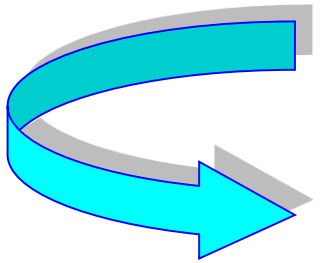
$$(x_1 - 1{,}328.57)(y_1 - 9) = 2{,}142.86$$

$$(x_2 - 1{,}328.57)(y_2 - 9) = -378.57$$

$$\mathrm{Cov}(X, Y) = \frac{1}{7} \sum_{i=1}^{7} (x_i - 1{,}328.6)(y_i - 9) = \frac{7{,}800}{7} = 1{,}114.29$$

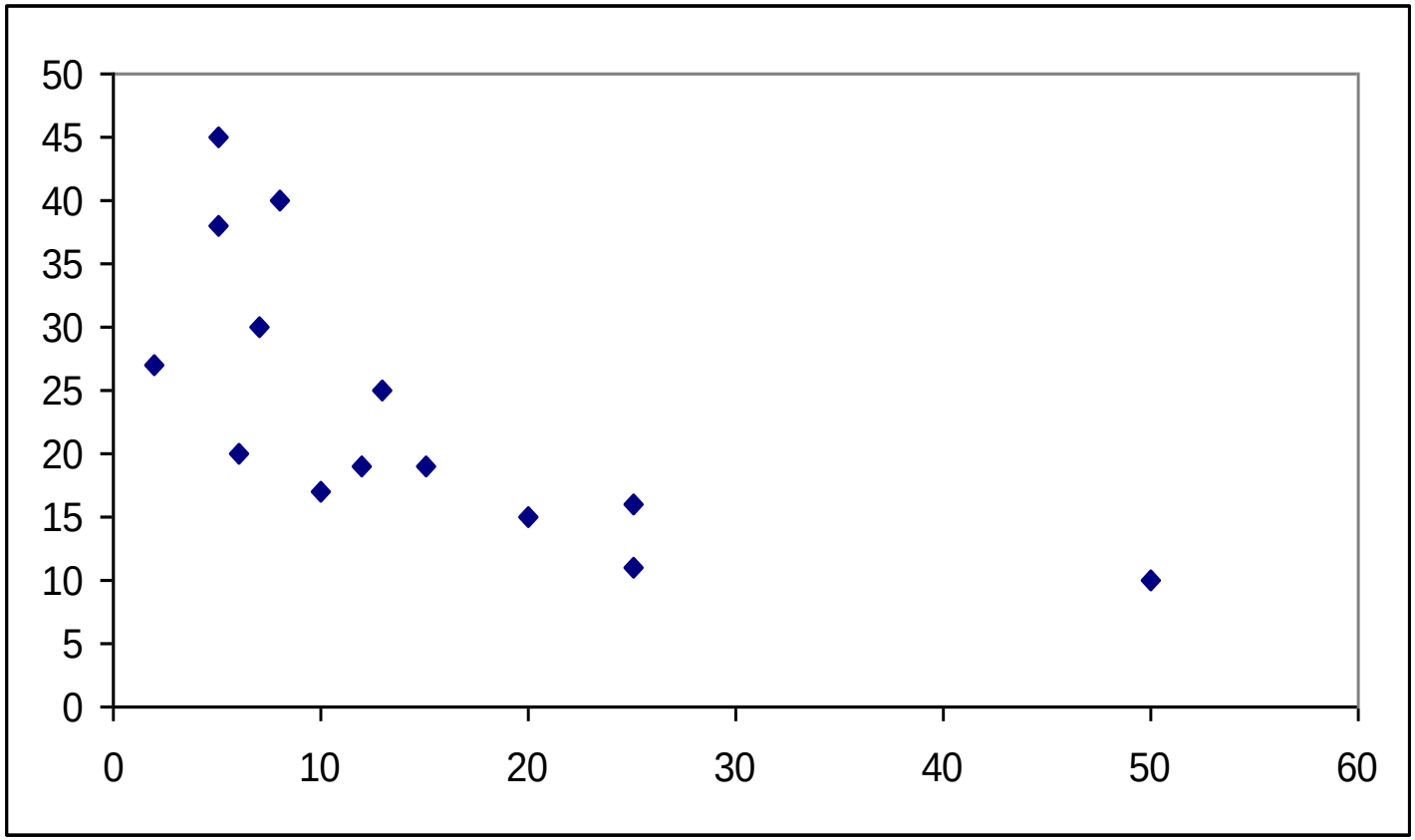
Esercizio 2

Si richiede di verificare se esiste covarianza tra il numero di film e gli spettacoli teatrali visti da un collettivo di 14 studenti in un semestre



	xi	yi
Id	film	spettacoli
1	50	10
2	25	11
3	20	15
4	25	16
5	10	17
6	15	19
7	12	19
8	6	20
9	13	25
10	2	27
11	7	30
12	5	38
13	8	40
14	5	45

Scatter - plot



Esercizio 2

	Id	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
	1	50	10	35.5	-13.71	-486.86
	2	25	11	10.5	-12.71	-133.50
	3	20	15	5.5	-8.71	-47.93
	4	25	16	10.5	-7.71	-81.00
	5	10	17	-4.5	-6.71	30.21
	6	15	19	0.5	-4.71	-2.36
	7	12	19	-2.5	-4.71	11.79
	8	6	20	-8.5	-3.71	31.57
	9	13	25	-1.5	1.29	-1.93
	10	2	27	-12.5	3.29	-41.07
	11	7	30	-7.5	6.29	-47.14
	12	5	38	-9.5	14.29	-135.71
	13	8	40	-6.5	16.29	-105.86
	14	5	45	-9.5	21.29	-202.21
Somma		203	332		COD(X,Y)	-1,212.00
Medie		14.50	23.71		COV(X,Y)	-86.57

Coefficiente di correlazione

Un indice relativo che misura il legame di interdipendenza tra 2 variabili X e Y e che utilizza la covarianza è il coefficiente di correlazione di Bravais-Pearson:

$$\rho(X,Y) = \frac{Cov(X,Y)}{sqm(x)\cdot sqm(y)} = \frac{\sigma_{xy}}{\sigma_X \sigma_Y}$$

$$\rho(X,Y) = \frac{\dfrac{\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})}{N}}{\sqrt{\dfrac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N}} \cdot \sqrt{\dfrac{\sum_{i=1}^{N}(yi - \bar{Y})^2}{N}}} = \frac{\sum_{i=1}^{N}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{X})^2} \cdot \sqrt{\sum_{i=1}^{N}(yi - \bar{Y})^2}} = \frac{Cod(X,Y)}{\sqrt{Dev(X)} \cdot \sqrt{Dev(Y)}}$$

Coefficiente di correlazione

Il coefficiente di correlazione quantifica, in sintesi, il grado di concordanza e di discordanza tra variabili:

$$-1 \leq \rho(X, Y) \leq +1$$

$\rho = +1 \rightarrow$ perfetta correlazione positiva (concordanza)

$\rho = -1 \rightarrow$ perfetta correlazione negativa (discordanza)

$\rho = 0 \rightarrow$ assenza di correlazione

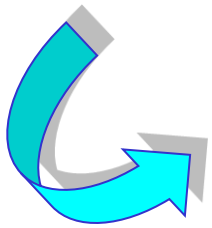
Esercizio 3

I dati riportati nella tabella seguente fanno riferimento ad un'intervista effettuata all'uscita da un cinema su 5 spettatori di un particolare film.

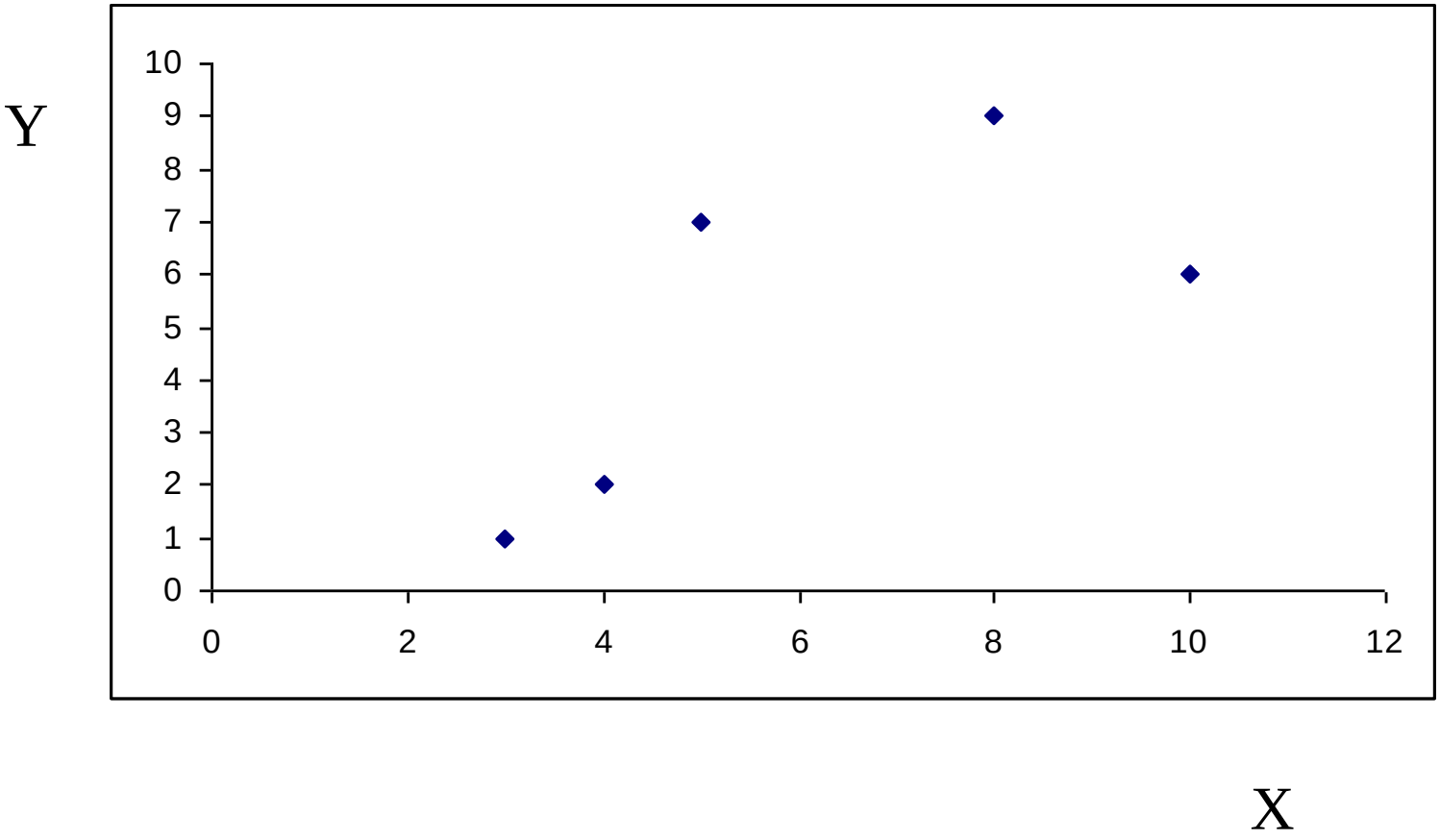
Si richiede di verificare se esiste correlazione tra i pareri espressi sul film e sulla sua colonna sonora

N.B. i pareri sono espressi in scala crescente da 1 a 10

Film (x)	Colonna sonora (y)
5	7
8	9
4	2
3	1
10	6



Scatter - plot



STEP:

1) CALCOLARE LE MEDIE

2) CALCOLARE GLI SCARTI DALLA MEDIA

3) EFFETTUARE IL PRODOTTO TRA COPPIE DI SCARTI E SOMMARE, OTTENENDO COD(X,Y) (se si divide COD(X,Y) per N si ottiene COV(X,Y))

4) CALCOLARE LE DEVIANZE DI X E Y, EFFETTUARE IL PRODOTTO E CALCOLARNE LA RADICE QUADRATA (se al numeratore è presente COV(X,Y) calcolare gli s.q.m. di X e di Y)

5) DIVIDERE COD (X,Y) PER IL RISULTATO DEL PUNTO 4) OTTENENDO $\rho(X,Y)$ (dividere COV (x,y) per il prodotto degli s.q.m.)

Esercizio 3

	Film(X)	Colonna sonora (Y)	$(x_i - \overline{X})$	$(y_i - \overline{Y})$	$(x_i - \overline{X})(y_i - \overline{Y})$	$(x_i - \overline{X})^2$	$(y_i - \overline{Y})^2$
	5	**7**	**-1**	**2**	**-2**	**1**	**4**
	8	**9**	**2**	**4**	**8**	**4**	**16**
	4	**2**	**-2**	**-3**	**6**	**4**	**9**
	3	**1**	**-3**	**-4**	**12**	**9**	**16**
	10	**6**	**4**	**1**	**4**	**16**	**1**
Totale	30	25			28	34	46
MEDIE	**6**	**5**					
CODEVIANZA					**28**		
DEVIANZA						34	46
S.Q.M.						2.608	3.03

Esercizio 3

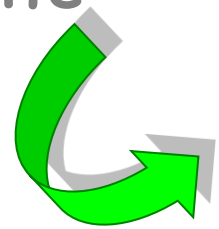
il coefficiente di correlazione di Bravais- Pearson è:

$$\rho(X,Y) = \frac{\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})}{\frac{1}{N}\sqrt{\sum_{i=1}^{N}(x_i - \overline{X})^2 \sum_{i=1}^{N}(yi - \overline{Y})^2}} = \frac{Cod(X,Y)}{\sqrt{Dev(X)Dev(Y)}} = \frac{28}{\sqrt{34 \times 46}} = 0.708$$

Tra le variabili è presente una forte correlazione positiva molto vicina al valore +1!

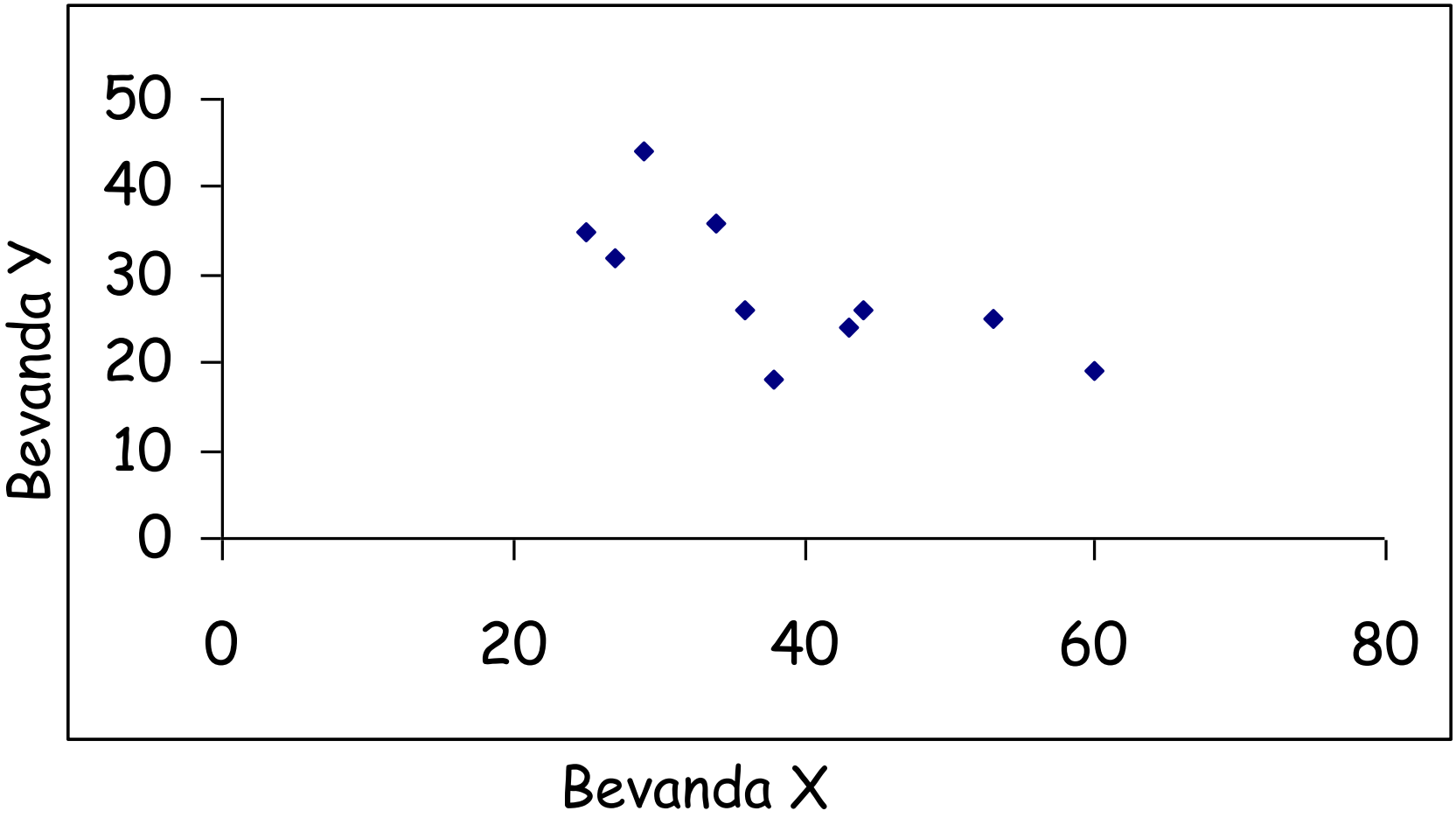
Esercizio 4

In tabella sono riportati i dati relativi alle quote di mercato detenute da due aziende operanti nel settore delle bibite analcoliche



bevanda X	bevanda Y
36	26
25	35
27	32
34	36
44	26
29	44
53	25
38	18
60	19
43	24

Scatter - plot



Esercizio 4

	bevanda X	bevanda Y	$(x_i - \bar{X})$	$(y_i - \bar{y})$	$(x_i - \bar{X})(y_i - \bar{y})$	$(x_i - \bar{X})^2$	$(y_i - \bar{y})^2$
	36	26	-2.90	-2.50	7.25	8.41	6.25
	25	35	-13.90	6.50	-90.35	193.21	42.25
	27	32	-11.90	3.50	-41.65	141.61	12.25
	34	36	-4.90	7.50	-36.75	24.01	56.25
	44	26	5.10	-2.50	-12.75	26.01	6.25
	29	44	-9.90	15.50	-153.45	98.01	240.25
	53	25	14.10	-3.50	-49.35	198.81	12.25
	38	18	-0.90	-10.50	9.45	0.81	110.25
	60	19	21.10	-9.50	-200.45	445.21	90.25
	43	24	4.10	-4.50	-18.45	16.81	20.25
totale	389	285			-586.50	1,152.90	596.50
medie	38.9	28.5					
codevianza					-586.50		
devianza						1,152.90	596.50

Esercizio 4

il coefficiente di correlazione di Bravais- Pearson è:

$$\rho(X,Y) = \frac{\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})}{\frac{1}{N}\sqrt{\sum_{i=1}^{N}(x_i - \overline{X})^2 \sum_{i=1}^{N}(yi - \overline{Y})^2}} = \frac{Cod(X,Y)}{\sqrt{Dev(X)Dev(Y)}} = \frac{-586.50}{\sqrt{1{,}152.90 \times 596.50}} = -0.71$$

Tra le variabili è presente una forte correlazione negativa molto vicina al valore - 1!