UNIMORE
UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

Dipartimento di Comunicazione
ed Economia

Analisi della relazione tra variabili quantitative

Relazioni tra variabili quantitative

Uno dei casi di maggiore interesse riguarda lo studio della relazione esistente tra due caratteri quantitativi.

Un primo passo per mettere in evidenza la relazione tra i due caratteri è rappresentato dallo strumento grafico del *diagramma di dispersione*:

Relazioni tra variabili quantitative

Uno dei casi di maggiore interesse riguarda lo studio della relazione esistente tra due caratteri quantitativi.

Un primo passo per mettere in evidenza la relazione tra i due caratteri è rappresentato dallo strumento grafico del *diagramma di dispersione*:

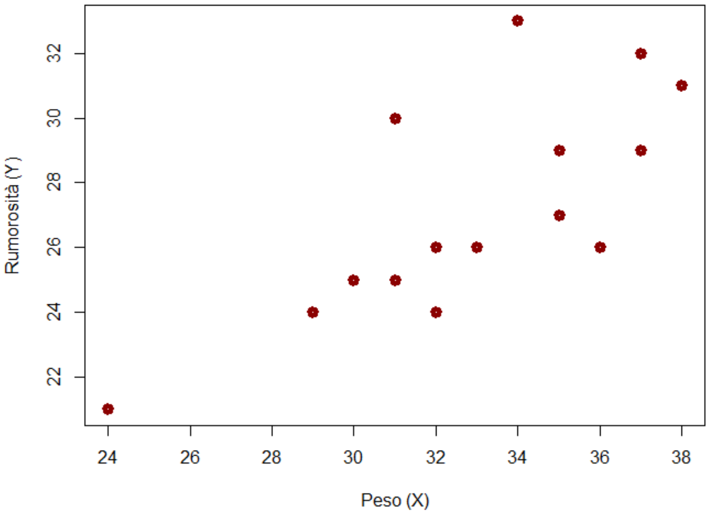
> Si tratta di rappresentare su un piano cartesiano tutte le coppie di osservazioni delle due variabili in esame

Esempio: Le lavatrici

Su 15 lavatrici sono stati misurati il peso del carico di lavaggio (X) e la rumorosità (Y), ottenendo i seguenti risultati:

Peso (X)	Rumorosità (Y)
31	25
33	26
37	29
38	31
29	24
35	29
32	26
35	27
30	25
24	21
36	26
31	30
34	33
37	32
32	24

Esempio: (cont)



Come leggere il diagramma di dispersione

I punti sul grafico tendono a seguire un andamento lineare crescente $\Rightarrow$ se uno dei due caratteri aumenta, anche l'altro carattere tende ad assumere valori crescenti.

Come leggere il diagramma di dispersione

I punti sul grafico tendono a seguire un andamento lineare crescente $\Rightarrow$ se uno dei due caratteri aumenta, anche l'altro carattere tende ad assumere valori crescenti.

In questo caso, diremo che i caratteri sono *concordanti*.

Come leggere il diagramma di dispersione

I punti sul grafico tendono a seguire un andamento lineare crescente $\Rightarrow$ se uno dei due caratteri aumenta, anche l'altro carattere tende ad assumere valori crescenti.

In questo caso, diremo che i caratteri sono *concordanti*.

Viceversa, se i punti sul grafico tendono a seguire un andamento lineare decrescente (vale a dire che se uno dei due caratteri aumenta, l'altro carattere tende ad assumere valori decrescenti)$\Rightarrow$ diremo che i caratteri sono *discordanti*.

Misurare la concordanza/discordanza

Osservazioni:

- Il diagramma di dispersione non consente di valutare "oggettivamente" la correlazione tra due variabili

Misurare la concordanza/discordanza

Osservazioni:

- Il diagramma di dispersione non consente di valutare "oggettivamente" la correlazione tra due variabili
- Occorre definire un indice in grado di misurare la concordanza (o la discordanza) tra due caratteri quantitativi che sia
  - *positivo*, se i caratteri sono concordi
  - *negativo*, se i caratteri sono discordi

Misurare la concordanza/discordanza

Osservazioni:

- Il diagramma di dispersione non consente di valutare "oggettivamente" la correlazione tra due variabili

- Occorre definire un indice in grado di misurare la concordanza (o la discordanza) tra due caratteri quantitativi che sia

  - *positivo*, se i caratteri sono concordi

  - *negativo*, se i caratteri sono discordi

Nota Bene: nel proseguo è necessario distinguere il caso in cui i dati si presentano sotto forma di distribuzioni unitarie dal caso in cui sono organizzati in una tabella a doppia entrata.

Misurare la concordanza/discordanza

La covarianza - distribuzioni unitarie

Consideriamo le variabili scarto

$$X' = X - M(X) \qquad \text{e} \qquad Y' = Y - M(Y)$$

ottenute come differenze dei valori di ciascuna variabile dalla rispettiva media.

Ritornando all'esempio

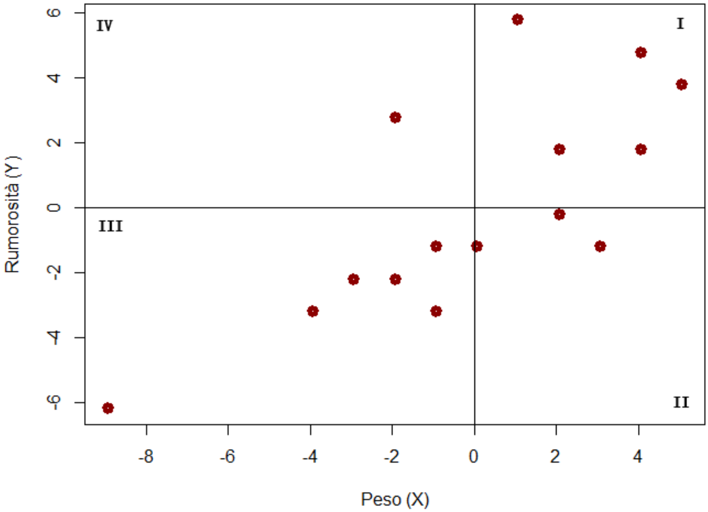
Sappiamo che $M(X) = 32.933$ e $M(Y) = 27.2$

Ritornando all'esempio

Sappiamo che $M(X) = 32.933$ e $M(Y) = 27.2$

Peso (X)	Rumorosità (Y)	X-M(X)	Y-M(Y)
31	25	-1.933	-2.2
33	26	0.067	-1.2
37	29	4.067	1.8
38	31	5.067	3.8
29	24	-3.933	-3.2
35	29	2.067	1.8
32	26	-0.933	-1.2
35	27	2.067	-0.2
30	25	-2.933	-2.2
24	21	-8.933	-6.2
36	26	3.067	-1.2
31	30	-1.933	2.8
34	33	1.067	5.8
37	32	4.067	4.8
32	24	-0.933	-3.2

Misurare la concordanza/discordanza



Misurare la concordanza/discordanza

- I QUADRANTE:
  Possiede sia X' che Y' positivi
- III QUADRANTE:
  Possiede sia X' che Y' negativi
- II QUADRANTE:
  Possiede X' positivi e Y' negativi
- IV QUADRANTE:
  Possiede X' negativi e Y' positivi

Misurare la concordanza/discordanza

- I QUADRANTE:
  Possiede sia X' che Y' positivi
- III QUADRANTE:
  Possiede sia X' che Y' negativi

$\searrow$ $\nearrow$ scostamenti concordi

- II QUADRANTE:
  Possiede X' positivi e Y' negativi
- IV QUADRANTE:
  Possiede X' negativi e Y' positivi

$\searrow$ $\nearrow$ scostamenti discordi

Misurare la concordanza/discordanza

La covarianza - distribuzioni unitarie

Un indice simmetrico che misura la concordanza (o discordanza) tra due caratteri quantitativi è la *covarianza*

Si definisce **covarianza** la media dei prodotti degli scostamenti delle variabili X ed Y dalle rispettive medie

$$\sigma_{XY} = \frac{1}{n} \sum_{i=1}^{n} [x_i - M(X)][y_i - M(Y)]$$

Esempio: Le lavatrici

Calcoliamo la covarianza

Esempio: Le lavatrici

Calcoliamo la covarianza

Peso (X)	Rumorosità (Y)	X-M(X)	Y-M(Y)	$[X - M(X)][Y - M(Y)]$
31	25	-1.933	-2.2	4.2526
33	26	0.067	-1.2	-0.0804
37	29	4.067	1.8	7.3206
38	31	5.067	3.8	19.2546
29	24	-3.933	-3.2	12.5856
35	29	2.067	1.8	3.7206
32	26	-0.933	-1.2	1.1196
35	27	2.067	-0.2	-0.4134
30	25	-2.933	-2.2	6.4526
24	21	-8.933	-6.2	55.3846
36	26	3.067	-1.2	-3.6804
31	30	-1.933	2.8	-5.4124
34	33	1.067	5.8	6.1886
37	32	4.067	4.8	19.5216
32	24	-0.933	-3.2	2.9856
				129.200

Esempio: Le lavatrici

Calcoliamo la covarianza

Peso (X)	Rumorosità (Y)	X-M(X)	Y-M(Y)	$[X - M(X)][Y - M(Y)]$
31	25	-1.933	-2.2	4.2526
33	26	0.067	-1.2	-0.0804
37	29	4.067	1.8	7.3206
38	31	5.067	3.8	19.2546
29	24	-3.933	-3.2	12.5856
35	29	2.067	1.8	3.7206
32	26	-0.933	-1.2	1.1196
35	27	2.067	-0.2	-0.4134
30	25	-2.933	-2.2	6.4526
24	21	-8.933	-6.2	55.3846
36	26	3.067	-1.2	-3.6804
31	30	-1.933	2.8	-5.4124
34	33	1.067	5.8	6.1886
37	32	4.067	4.8	19.5216
32	24	-0.933	-3.2	2.9856
				129.200

$$\sigma_{XY} = \frac{129.2}{15} = 8.613$$

La covarianza - distribuzioni unitarie

Osservazioni:

- Misura la variazione congiunta delle due variabili rispetto alle corrispondenti medie
- È un indice *simmetrico* di dipendenza lineare $\Rightarrow \sigma_{XY} = \sigma_{YX}$
- Può assumere valori sull'insieme dei numeri reali:
  - se assume valori > 0 tra i due caratteri c'è *concordanza*
  - se assume valori < 0 tra i due caratteri c'è *discordanza*
- Se X e Y sono *indipendenti*, allora $\sigma_{XY} = 0$. Tuttavia, se $\sigma_{XY} = 0$. <u>non è detto</u> che X e Y siano indipendenti
- Dipende dall'unità di misura delle osservazioni ($\Rightarrow$ non può essere usata per fare confronti!)

Misurare la concordanza/discordanza

Il coefficiente di correlazione

La covarianza può assumere valori nell'intervallo

$$-\sigma_X\sigma_Y \leq \sigma_{XY} \leq \sigma_X\sigma_Y$$

dove σ_X e σ_Y sono le deviazioni standard di X ed Y

Misurare la concordanza/discordanza

Il coefficiente di correlazione

Quindi, è possibile ottenere l'indice relativo noto come *coefficiente di correlazione di Bravais e Pearson*

Il **coefficiente di correlazione di Bravais e Pearson** è dato da:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Misurare la concordanza/discordanza

Il coefficiente di correlazione

È possibile verificare che

$$-1 \leq \rho_{XY} \leq 1$$

In particolare:

- $\rho_{XY} = 1$ nel caso di perfetto legame lineare (X e Y concordi)
- $\rho_{XY} = -1$ nel caso di perfetto legame lineare (X e Y discordi)
- Se X e Y sono *indipendenti*, allora $\rho_{XY} = 0$. Tuttavia, se $\rho_{XY} = 0$. <u>non è detto</u> che X e Y siano indipendenti (la relazione potrebbe essere non lineare)

Esempio: Le lavatrici

Calcoliamo l'indice di correlazione di Bravais Pearson

Esempio: Le lavatrici

Peso (X)	Rumorosità (Y)	X-M(X)	Y-M(Y)	$[X - M(X)]^2$	$[Y - M(Y)]^2$
31	25	-1.933	-2.2	3.7365	4.84
33	26	0.067	-1.2	0.0045	1.44
37	29	4.067	1.8	16.5405	3.24
38	31	5.067	3.8	25.6745	14.44
29	24	-3.933	-3.2	15.4685	10.24
35	29	2.067	1.8	4.2725	3.24
32	26	-0.933	-1.2	0.8705	1.44
35	27	2.067	-0.2	4.2725	0.04
30	25	-2.933	-2.2	8.6025	4.84
24	21	-8.933	-6.2	79.7985	38.44
36	26	3.067	-1.2	9.4065	1.44
31	30	-1.933	2.8	3.7365	7.84
34	33	1.067	5.8	1.1385	33.64
37	32	4.067	4.8	16.5405	23.04
32	24	-0.933	-3.2	0.8705	10.24
				190.9333	158.4

Esempio: Le lavatrici

Peso (X)	Rumorosità (Y)	X-M(X)	Y-M(Y)	$[X - M(X)]^2$	$[Y - M(Y)]^2$
31	25	-1.933	-2.2	3.7365	4.84
33	26	0.067	-1.2	0.0045	1.44
37	29	4.067	1.8	16.5405	3.24
38	31	5.067	3.8	25.6745	14.44
29	24	-3.933	-3.2	15.4685	10.24
35	29	2.067	1.8	4.2725	3.24
32	26	-0.933	-1.2	0.8705	1.44
35	27	2.067	-0.2	4.2725	0.04
30	25	-2.933	-2.2	8.6025	4.84
24	21	-8.933	-6.2	79.7985	38.44
36	26	3.067	-1.2	9.4065	1.44
31	30	-1.933	2.8	3.7365	7.84
34	33	1.067	5.8	1.1385	33.64
37	32	4.067	4.8	16.5405	23.04
32	24	-0.933	-3.2	0.8705	10.24
				190.9333	158.4

$$\sigma_X^2 = \frac{190.93}{15} = 12.73 \quad \Rightarrow \sigma_X = \sqrt{12.73} = 3.57$$

$$\sigma_Y^2 = \frac{158.40}{15} = 10.56 \quad \Rightarrow \sigma_Y = \sqrt{10.56} = 3.25$$

Esempio: Le lavatrici

Peso (X)	Rumorosità (Y)	X-M(X)	Y-M(Y)	$[X - M(X)]^2$	$[Y - M(Y)]^2$
31	25	-1.933	-2.2	3.7365	4.84
33	26	0.067	-1.2	0.0045	1.44
37	29	4.067	1.8	16.5405	3.24
38	31	5.067	3.8	25.6745	14.44
29	24	-3.933	-3.2	15.4685	10.24
35	29	2.067	1.8	4.2725	3.24
32	26	-0.933	-1.2	0.8705	1.44
35	27	2.067	-0.2	4.2725	0.04
30	25	-2.933	-2.2	8.6025	4.84
24	21	-8.933	-6.2	79.7985	38.44
36	26	3.067	-1.2	9.4065	1.44
31	30	-1.933	2.8	3.7365	7.84
34	33	1.067	5.8	1.1385	33.64
37	32	4.067	4.8	16.5405	23.04
32	24	-0.933	-3.2	0.8705	10.24
				190.9333	158.4

$$\sigma_X^2 = \frac{190.93}{15} = 12.73 \quad \Rightarrow \sigma_X = \sqrt{12.73} = 3.57$$

$$\sigma_Y^2 = \frac{158.40}{15} = 10.56 \quad \Rightarrow \sigma_Y = \sqrt{10.56} = 3.25$$

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{8.613}{3.57 \cdot 3.25} = 0.74$$

Formula alternativa per il calcolo della covarianza

È possibile dimostrare che la covarianza può essere anche scritta come

$$\sigma_{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - M(X)M(Y)$$

Esempio: Le lavatrici

Peso (X)	Rumorosità (Y)	$x_i y_i$
31	25	775
33	26	858
37	29	1073
38	31	1178
29	24	696
35	29	1015
32	26	832
35	27	945
30	25	750
24	21	504
36	26	936
31	30	930
34	33	1122
37	32	1184
32	24	768
		13566

$$\sigma_{XY} = \frac{13566}{15} - 32.933 \cdot 27.2 = 904.4 - 894.88 = 8.62$$