

Esercitazione 6 del corso di Statistica (parte 1)

Dott.ssa Paola Costantini

19 Febbraio 2009

Nella tabella seguente sono riportati gli anni di iscrizione universitaria ed il punteggio ad un quiz rilevati su di un collettivo di otto studenti.

Anni di Università (X)	Punteggio Quiz (Y)
1	30
3	40
5	40
4	50
2	35
5	50
3	35
2	25

Verificare l'esistenza di una relazione statistica tra le due variabili.

Considerando che entrambe le variabili sono di natura quantitativa, l'indice più opportuno per verificare l'esistenza di una relazione tra le variabili è il coefficiente di correlazione lineare.

Il coefficiente di correlazione è una misura dell'**INTERDIPENDENZA** lineare tra due fenomeni e varia tra -1 ed +1 (perfetta correlazione inversa, perfetta correlazione diretta).

Il coefficiente di correlazione di Bravais-Pearson è dato da:

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\mu_{X^2} - \mu_X^2} \sqrt{\mu_{Y^2} - \mu_Y^2}} = \frac{\frac{1}{N} \sum_{i=1}^N x_i y_i - \mu_X \mu_Y}{\sqrt{\mu_{X^2} - \mu_X^2} \sqrt{\mu_{Y^2} - \mu_Y^2}}$$

Il numeratore si chiama **Covarianza** perché è una misura della contemporanea variazione di X e Y in rapporto alle rispettive medie.

Mesi (N)	X	Y	XY	X ²	Y ²
1	1	30	30	1	900
2	3	40	120	9	1600
3	5	40	200	25	1600
4	4	50	200	16	2500
5	2	35	70	4	1225
6	5	50	250	25	2500
7	3	35	105	9	1225
8	2	25	50	4	625
Totale	25	305	1025	93	12175

Si inizia con il calcolare la media per le due variabili

$$\mu_x = 25/8 = 3,125$$

$$\mu_y = 305/8 = 38,125$$

$$\text{Cov}(X,Y) = \frac{1}{N} \sum_{i=1}^N x_i y_i - \mu_x \mu_y = \frac{1025}{8} - 3,125 * 38,125 = 128,125 - 119,1406 = 8,9844$$

$$\sigma_x = \sqrt{\mu_{x^2} - \mu_x^2} = \sqrt{11,625 - 9,765} = 1,3635$$

$$\sigma_y = \sqrt{\mu_{y^2} - \mu_y^2} = \sqrt{1521,875 - 1453,516} = 8,2679$$

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{8,9844}{1,3635 * 8,2679} = 0,7969$$

Siamo in presenza di un'alta correlazione positiva tra la variabile anni di iscrizione all'università e la variabile punteggio quiz.

Esercizio n 2

La seguente tabella riporta le informazioni relative a 25 laureati nell'anno 2005 in Economia, ad un anno dal conseguimento del titolo.

Genere	Eta	Voto	Durata	Tipo Contratto	Utilizzo della Laurea	Efficacia della Laurea
Uomini	47,8	83	4,4	Stabile	In misura elevata	Efficace
Uomini	26,6	113	7,4	Stabile	In misura ridotta	Poco efficace
Uomini	31,5	91	12,4	Atipico	In misura ridotta	Abb. efficace
Uomini	23,6	102	4,4	Inserimento/Form	In misura elevata	Efficace
Uomini	25,9	94	6,4	Stabile	Per niente	Per nulla efficace
Donne	23,6	108	4,7	Inserimento/Form	In misura ridotta	Abb. efficace
Donne	28,6	108	5,7	Atipico	In misura ridotta	Abb. efficace
Uomini	42,1	100	3	Stabile	Per niente	Efficace
Donne	24,3	113	3,4	Atipico	Per niente	Per nulla efficace
Donne	26,3	113	3,4	Atipico	Per niente	Per nulla efficace
Uomini	24,9	106	3,7	Inserimento/Form	In misura elevata	Molto efficace
Donne	24	95	4,5	Stabile	In misura elevata	Efficace
Uomini	34,7	92	4,4	Stabile	In misura elevata	Efficace
Donne	24,7	106	5,14	Atipico	In misura ridotta	Abb. efficace
Uomini	25,9	100	5,94	Atipico	In misura ridotta	Abb. efficace
Uomini	25,4	92	6,14	Atipico	In misura elevata	Molto efficace
Donne	27,6	113	4,14	Senza contratto	Per niente	Poco efficace
Donne	23,4	113	44	Inserimento/Form	In misura elevata	Molto efficace
Uomini	31,3	105	3,4	Stabile	In misura elevata	Molto efficace
Uomini	29,7	110	4,14	Atipico	In misura ridotta	Abb. efficace
Uomini	27	93	7,3	Stabile	In misura elevata	Efficace
Uomini	35,6	97	15,4	Atipico	Per niente	Poco efficace
Uomini	23,1	101	3,7	Senza contratto	Per niente	Poco efficace
Uomini	25,3	91	6,5	Atipico	Per niente	Per nulla efficace
Uomini	32,6	92	13,5	Stabile	In misura ridotta	Abb. efficace

Costruire la distribuzione di intensità per la variabile Voto di laurea

Suddividere la distribuzione in 3 classi equiampie riportando le frequenze (assolute e relative) e la funzione di ripartizione empirica.

L'ampiezza costante delle tre classi si ottiene come:

$$\text{range (VOTO DI LAUREA)} = 113 - 83 = 30$$

L'ampiezza delle classi della prima distribuzione di frequenza è pari a:

$$\text{Ampiezza classi} = \text{range}/3 = 30/3 = 10$$

Le 3 classi sono, dunque, delimitate dai seguenti estremi:

$$C1 = [83; 93]$$

$$C2 =] 93; 103]$$

$$C3 =] 103; 113]$$

Costruire la distribuzione di intensità per la variabile Durata del corso di studio

Suddividere la distribuzione in 3 classi equiampie riportando le frequenze (assolute e relative) e la funzione di ripartizione empirica.

L'ampiezza costante delle tre classi si ottiene come:

$$\text{range (durata)} = 15,4 - 3 = 12,4$$

L'ampiezza delle classi della prima distribuzione di frequenza è pari a:

$$\text{Ampiezza classi} = \text{range}/3 = 12,4/3 = 4,13$$

Le 3 classi sono, dunque, delimitate dai seguenti estremi:

$$C1 = [3; 7,13]$$

$$C2 =] 7,13; 11,27]$$

$$C3 =] 11,27; 15,4]$$

Voto \ Durata	C₁ = [83; 93]	C₂ =] 93; 103]	C₃ =] 103; 113]	tot
C1 = [3; 7,13]	4	6	10	20
C2 =] 7,13; 11,27]	1	0	1	2
C3 =] 11,27; 15,4]	2	1	0	3
Totali	7	7	11	25

Valori centrali di y: 88; 98; 108

Valori centrali di x: 5,065; 9,2; 13,33

$$\text{Media delle } y = (88 \times 7) + (98 \times 7) + (108 \times 11) / 25 = 99,6$$

$$\text{Media delle } x = (5,065 \times 20) + (9,2 \times 2) + (13,33 \times 3) / 25 = 6,4$$

$$\begin{aligned} \mu_{x,y} &= (88 \times 5,065 \times 4) + (88 \times 9,2 \times 1) + (88 \times 13,33 \times 2) + (98 \times 5,065 \times 6) + (98 \times 9,2 \times 0) + (98 \times 13,33 \times 1) + \\ &+ (108 \times 5,065 \times 10) + (108 \times 9,2 \times 1) + (108 \times 9,2 \times 0) = \end{aligned}$$

$$\mathbf{1782,88+809,6+2346+2978+1306,34+5470,2+993,615686,62 = 627,46}$$

$$Cov = \sigma_{x,y} = \mu_{x,y} - (\mu_x \cdot \mu_y) = 627,46 - (6,4 \cdot 99,6) = -9,98$$

$$\frac{1}{N} \sum (\hat{y} - \mu_y)^2 n_i = \frac{1}{25} \sum (88 - 99,6)^2 \cdot 7 + (98 - 99,6)^2 \cdot 7 + (108 - 99,6)^2 \cdot 11 =$$

Var(y)=

$$= 941,92 + 17,92 + 776,16 / 25 = 69,44 \rightarrow \sigma = \sqrt{69,44} = 8,33$$

$$\frac{1}{N} \sum (\hat{x} - \mu_x)^2 n_i = \frac{1}{25} \sum (5,065 - 6,4)^2 \cdot 20 + (9,2 - 6,4)^2 \cdot 2 + (13,33 - 6,4)^2 \cdot 3 =$$

Var(x)=

$$= 35,64 + 15,68 + 144 / 25 = 7,8 \rightarrow \sigma = \sqrt{7,8} = 2,79$$

$$\text{Corr} = \rho_{x,y} = \frac{Cov_{x,y}}{\sigma_x \cdot \sigma_y} = \frac{-9,98}{2,79 \cdot 8,33} = \frac{-9,98}{23,24} = -0,429$$

Correlazione negativa media