Statistica I modulo (*S.I.A.F.A.*, AA 2024/25)

Cograduazione, dipendenza in media, covarianza, correlazione

Soluzioni

1. Ai fini del calcolo del coefficiente ρ_s di Spearman, si individuano le due graduatorie e i valori d_i^2.

Regioni	Attrattività	Convenienza	d_i	d_i^2
Campania	2	2	0	0
Basilicata	5,5	4,5	1	1
Puglia	2	4,5	-2,5	6,25
Calabria	5,5	4,5	1	1
Sicilia	4	4,5	-0,5	0,25
Sardegna	2	1	1	1

$$\rho_s = 1 - \frac{6\sum_{i=1}^{6} d_i^2}{6(36-1)} = 1 - \frac{57}{210} = 0,732$$

Esiste un'alta concordanza tra le graduatorie dei due caratteri.

2. La media complessiva è pari a

$$\bar{y} = \frac{6 \cdot 58 + 12 \cdot 35 + 18 \cdot 14 + 24 \cdot 28}{135} = 12,533$$

La prima delle medie condizionate, ovvero la media della variabile Y condizionata alla prima modalità della X (Nord-Ovest), è data da

$$\bar{y}_{x=x_1} = \frac{6 \cdot 3 + 12 \cdot 2 + 18 \cdot 0 + 24 \cdot 0}{5} = 8,4$$

In modo analogo si calcolano le altre media condizionate,

$$\bar{y}_{x=x_2} = 12$$

$$\bar{y}_{x=x_3} = 12,686$$

$$\bar{y}_{x=x_4} = 12,857$$

Il denominatore di $\eta^2_{Y|X}$ è dato da

$$\sum_{j=1}^{4} (y_j - \bar{y})^2 n_{.j} = 6585,6$$

e dunque

$$\eta^2_{Y|X} = \frac{\sigma^2_{Media(Y|X)}}{\sigma^2_Y} = \frac{\sum_{i=1}^{4} (\bar{y}_{x=x_i} - \bar{y})^2 n_{i.}}{\sum_{j=1}^{4} (y_j - \bar{y})^2 n_{.j}} = \frac{91,2}{6585,6} = 0,014.$$

Esiste un grado di dipendenza in media della variabile Y dal carattere X quasi nullo. La variabile Y può essere considerata quasi indipendente in media dal carattere X.

3. La media complessiva è pari a

$$\bar{y} = \frac{85,5 \cdot 63 + 95,5 \cdot 60 + 103 \cdot 56 + 108 \cdot 52}{231} = 97,405$$

La prima delle medie condizionate, ovvero la media della variabile Y condizionata alla prima modalità della X (Lic. Elementare), è data da

$$\bar{y}_{x=x_1} = \frac{85,5 \cdot 3 + 95,5 \cdot 6 + 103 \cdot 8 + 108 \cdot 8}{25} = 100,7$$

In modo analogo si calcolano le altre media condizionate,

$$\bar{y}_{x=x_2} = 97,196$$

$$\bar{y}_{x=x_3} = 96,432$$

$$\bar{y}_{x=x_4} = 98,781$$

Il denominatore di $\eta^2_{Y|X}$ è dato da

$$\sum_{j=1}^{4} (y_j - \bar{y})^2 \, n_{.j} = 16736,9$$

e dunque

$$\eta^2_{Y|X} = \frac{\sigma^2_{Media(Y|X)}}{\sigma^2_Y} = \frac{\sum_{i=1}^{4} (\bar{y}_{x=x_i} - \bar{y})^2 \, n_{i.}}{\sum_{j=1}^{4} (y_j - \bar{y})^2 \, n_{.j}} = \frac{446,139}{16736,9} = 0,027.$$

Esiste un grado di dipendenza in media della variabile Y dal carattere X quasi nullo. La variabile Y può essere considerata quasi indipendente in media dal carattere X.

4. (a) Vedi Figura 1

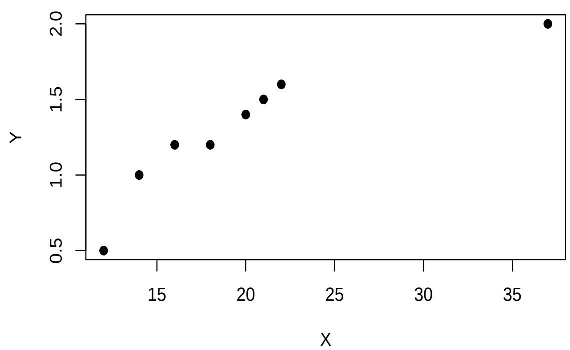

Figura 1: Diagramma di dispersione.

(b) Dalla tabella riportata sotto si calcola

$$\sigma_{XY} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{21,5}{8} = 2,6875$$

$$\sigma_X = \sqrt{\frac{414}{8}} = 7,194$$

e

$$\sigma_Y = \sqrt{\frac{1,38}{8}} = 0,415$$

da cui

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = 0,90$$

x	y	$x-\bar{x}$	$y-\bar{y}$	Prodotto	$(x-\bar{x})^2$	$(y-\bar{y})^2$
12	0,5	-8	-0,8	6,4	64	0,64
14	1	-6	-0,3	1,8	36	0,09
16	1,2	-4	-0,1	0,4	16	0,01
18	1,2	-2	-0,1	0,2	4	0,01
20	1,4	0	0,1	0	0	0,01
21	1,5	1	0,2	0,2	1	0,04
22	1,6	2	0,3	0,6	4	0,09
37	2	17	0,7	11,9	289	0,49
tot				21,5	414	1,38

Esiste un elevato grado di associazione (lineare) diretta tra le variabili X e Y.

5. (a) Vedi Figura 2.

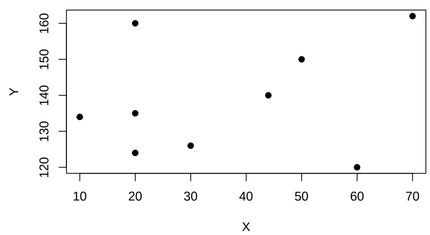

Figura 2: Diagramma di dispersione.

(b) Svolgendo i calcoli come mostrato nell'esercizio precedente si ottiene

$$\rho_{XY} = 0,259$$

Esiste un debole grado di associazione (lineare) diretta tra le variabili X e Y.