

Esercitazione 8 del corso di Statistica (parte 1)

Dott.ssa Paola Costantini

4 Dicembre 2008

Esercizio 1

Considerando il DATASET *INFORMATION TECHNOLOGY*:

Calcolare media aritmetica, moda, mediana, quartili, varianza e sarto quadratico medio per il carattere % FATTURATO SOFTWARE a partire:

1.1) dalla distribuzione semplice del carattere;

%Fatturato Software
15
35
34
40
40
40
55
30
40
20
50
85
40
40
90
50
50
30
60
65
50
40
50
30
5

%Fatturato Software	n_i	f_i	F_i
5	1	0,04	0,04
15	1	0,04	0,08
20	1	0,04	0,12
30	3	0,12	0,24
$Q_1=34$	1	0,04	0,28
35	1	0,04	0,32
Me 40	7	0,28	0,6
$Q_3=50$	5	0,2	0,8
55	1	0,04	0,84
60	1	0,04	0,88
65	1	0,04	0,92
85	1	0,04	0,96
90	1	0,04	1

Media aritmetica

$$\mu = \frac{\sum_{i=1}^n x_i \times n_i}{n} = \frac{5 \times 1 + 15 \times 1 + 20 \times 1 + 30 \times 3 + \dots + 90 \times 1}{25} = 43,36$$

Moda = 40 Me = 40 $Q_1 = 34$ $Q_3 = 50$

$$\text{Varianza} = \sigma_x^2 = \frac{1}{n} \sum_{i=1}^k (x_i - \mu)^2 n_i = \frac{1}{25} \sum (5 - 43,36)^2 \cdot 1 + (15 - 43,36)^2 \cdot 1 + \dots + (90 - 43,36)^2 = 344,15$$

$$\text{Scarto Quadratico Medio} = \sqrt{344,15} = 18,55$$

1.2) dalla distribuzione nelle classi: [5 – 33];]33 – 61];]61 – 90];

x_i	c_i	n_i	f_i	N_i	F_i
[5,33]	19	6	0,24	6	0,24
]33,61]	47	16	0,64	22	0,88
]61,90]	75,5	3	0,12	25	1
Totali		25	1,00		

Media aritmetica

$$\mu = \frac{\sum_{i=1}^n c_i \times n_i}{n} = \frac{19 \times 6 + 47 \times 16 + 75,5 \times 3}{25} = 43,7$$

Moda = classe]33,61]

$$Me \cong x_{Me-1} + (x_{Me} - x_{Me-1}) \frac{0,5 - F_{Me-1}}{F_{Me} - F_{Me-1}}$$

$$Me = 33 + (61 - 33) \cdot \frac{0,5 - 0,24}{0,88 - 0,24} = 44,375$$

$$\text{Primo quartile} \quad Q_1 \cong x_{Q_1-1} + (x_{Q_1} - x_{Q_1-1}) \frac{0,25 - F_{Q_1-1}}{F_{Q_1} - F_{Q_1-1}}$$

$$Q_1 = 33 + (33 - 5) \cdot \frac{0,25 - 0,24}{0,88 - 0,24} = 33,4375$$

$$\text{Terzo quartile} \quad Q_3 \cong x_{Q_3-1} + (x_{Q_3} - x_{Q_3-1}) \frac{0,75 - F_{Q_3-1}}{F_{Q_3} - F_{Q_3-1}}$$

$$Q_3 = 33 + (33 - 5) \cdot \frac{0,75 - 0,24}{0,88 - 0,24} = 55,31$$

$$\text{Varianza} = \sigma_x^2 = \frac{1}{n} \sum_{i=1}^k (c_i - \mu)^2 n_i = \frac{1}{25} \sum (19 - 43,7)^2 \cdot 6 + (47 - 43,7)^2 \cdot 16 + (75,5 - 43,7)^2 = 274,4$$

$$\text{Scarto Quadratico Medio} = \sqrt{274,4} = 16,56$$

Esercizio 2

Considerando il datate Information technology (informazioni relative a 25 aziende operanti nel settore dell'*information technology*):

Id. Aziende	Appartenenza gruppo	Fatturato (milioni di lire)	%Fatturato Hardware	%Fatturato Software	Numero clienti	Numero dipendenti	Localizzazione clienti	Area attività clienti	Importanz mercato
1	1	62000	15	15	1500	350	Nazionale	Pubblica Amministrazione	1
2	0	8000	30	35	40	55	Nazionale	Società servizi	4
3	1	2300	33	34	30	7	Locale	Industria/Commercio	2
4	1	32500	0	40	350	210	Nazionale	Banche/Assicurazioni	3
5	1	12000	0	40	30	95	Nazionale	Banche/Assicurazioni	2
6	1	2500	30	40	300	12	Nazionale	Industria/Commercio	1
7	1	4500	20	55	1000	30	Nazionale	Pubblica Amministrazione	2
8	0	24000	5	30	100	100	Locale	Pubblica Amministrazione	2
9	1	49500	0	40	10	260	Locale	Pubblica Amministrazione	2
10	1	10000	20	20	200	55	Nazionale	Pubblica Amministrazione	2
11	1	5000	0	50	150	7	Nazionale	Industria/Commercio	3
12	1	44000	10	85	3000	200	Nazionale	Industria/Commercio	3
13	0	10000	5	40	150	100	Nazionale	Società servizi	3
14	1	4000	20	40	200	14	Nazionale	Industria/Commercio	3
15	0	5000	0	90	40	70	Nazionale	Pubblica Amministrazione	1
16	0	50000	0	50	100	270	Nazionale	Banche/Assicurazioni	2
17	1	4000	0	50	170	25	Nazionale	Pubblica Amministrazione	2
18	0	3500	70	30	20	20	Nazionale	Industria/Commercio	4
19	1	8000	10	60	130	40	Locale	Industria/Commercio	1
20	1	1000	5	65	100	10	Nazionale	Pubblica Amministrazione	3
21	1	3000	0	50	100	10	Locale	Pubblica Amministrazione	3
22	0	5000	0	40	250	25	Nazionale	Pubblica Amministrazione	2
23	0	14000	0	50	500	45	Locale	Industria/Commercio	3
24	1	5000	0	30	15	50	Locale	Banche/Assicurazioni	2
25	0	10500	0	5	40	60	Locale	Industria/Commercio	2

si calcoli la correlazione tra i caratteri % FATTURATO SOFTWARE E NUMERO DI CLIENTI. A tale riguardo si suddividano entrambi i caratteri in tre classi equi-ampie.

Soluzione

L'ampiezza costante delle tre classi per il carattere "%fatturato software" (espresso in migliaia di euro) si ottiene come:

$$\text{range} (\% \text{FATTURATO SOFTWARE}) = 90 - 5 = 85$$

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

Ampiezza %FATTURATO SOFTWARE = range/3 = 85/3 = 28,33 (per cui avremo due classi di ampiezza 28 e una di ampiezza 29).

L'ampiezza costante delle tre classi per il carattere "NUMERO CLIENTI" si ottiene come:

$$\text{range}(\text{NUMERO CLIENTI}) = 3000 - 10 = 2990$$

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

Ampiezza NUMERO CLIENTI = range/3 = 2990/3 = 996,67 (per cui avremo due classi di ampiezza 1000 e una di ampiezza 990)

% Fatturato Soft.	[5,33]]33,61]]61,90]	Totale
Numero Clienti				
[10,1010]	5	16	2	23
]1010,2010]	1	0	0	1
]2010,3000]	0	0	1	1
Totale	6	16	3	25

Valori Centrali di $x = 510; 1510; 2505$

Valori Centrali di $y = 19; 47; 75,5$

\hat{x}_i	n_i	\hat{y}_i	n_j	$\hat{y}_i n_j$	$\hat{x}_i n_i$	\hat{y}_j^2	$\hat{y}_j^2 n_i$	\hat{x}_i^2	$\hat{x}_i^2 n_i$
510	23	19	6	114	11730	361	2166	260100	5982300
1510	1	47	16	752	1510	2209	35344	2280100	2280100
2505	1	75,5	3	226,5	2505	5700,25	17100,75	6275025	6275025
				1092,5	15745		54610,75		14537425

Media delle $x = \mu_x = 15745/20 = 629,8$

Media delle $y = \mu_y = 1092,5/25 = 43,7$

Per calcolare il termine $\sum_i \sum_j \hat{x}_i \hat{y}_j n_{ij}$ è necessario costruire la tabella delle $\hat{x}_i \hat{y}_j n_{ij}$ come segue:

\hat{y}_i	19	47	75,5
\hat{x}_i			
510	48450	23970	0
1510	459040	0	0
2505	95190	0	189127,5

La somma degli elementi all'interno della tabella è pari a 815777,5 da cui:

$$\frac{\sum_i \sum_j \hat{x}_i \hat{y}_j n_{ij}}{n} = \frac{815777,5}{25} = 32631,1$$

$$\text{Cov} = \sigma_{x,y} = \mu_{x,y} - (\mu_x * \mu_y) = 32631,1 - (629,8 * 43,7) = 32631,1 - 27522,26 = 5108,4$$

$$\sigma^2_x = VAR(Y) = [\hat{Y}^2 \cdot n_i] - \mu^2 = \frac{54610,75}{25} - 43,7^2 = 2184,42 - 1909,69 = 274,73 \rightarrow \sigma = \sqrt{274,73} = 16,57$$

$$\sigma^2_y = VAR(X) = [\hat{X}^2 \cdot n_j] - \mu^2 = \frac{14537425}{25} - 629,8^2 = 581497 - 396648 = 184849 \rightarrow \sigma = \sqrt{184849} = 429,9$$

$$\text{Corr} = \rho_{x,y} = \frac{\text{Cov}_{x,y}}{\sigma_x \cdot \sigma_y} = \frac{5108,4}{16,57 \cdot 429,9} = \frac{5108,4}{7123,443} = 0,72$$

Tale valore va confrontato con l'intervallo [-1, 1], quindi indica una correlazione lineare positiva piuttosto alta.