Università di Cassino

Corso di Statistica

Esercitazione del 26 febbraio 2010

Simona Balzano

ESERCIZIO 1

Considerando le le variabili EFFETTO SERRA e VALORE AGGIUNTO per le sole 10 regioni caratterizzate dai più alti valori di effetto serra, calcolare:

a) la covarianza
b) il coefficiente di correlazione

Soluzione

a)

1. formule classiche

$$\sigma_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)(y_i - \mu_Y)$$

$$\rho_{XY} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2\right) \times \left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_Y)^2\right)}}$$

Nella tabella che segue, accanto ai dati, contenuti nelle prime due colonne, sono raccolte le quantità necessarie per il calcolo sia della covarianza sia del coefficiente di correlazione, considerando che:

$$\mu_X = \frac{1}{10}373,39 = 37,34$$

$$\mu_Y = \frac{1}{10}1079,32 = 107,93$$

Effetto serra X	**Valore aggiunto Y**	$(x_i - \mu_X)$	$(y_i - \mu_Y)$	$(x_i - \mu_X)(y_i - \mu_Y)$	$(x_i - \mu_X)^2$	$(y_i - \mu_Y)^2$
31.71	104.59	-5.63	-3.35	18.84	31.72	11.19
68.71	270.87	31.38	162.94	5112.40	984.45	26549.56
38.85	121.90	1.51	13.96	21.07	2.28	194.97
17.83	35.69	-19.51	-72.24	1409.60	380.76	5218.40
38.36	111.21	1.02	3.28	3.34	1.04	10.75
26.86	86.01	-10.47	-21.92	229.65	109.71	480.70
33.04	139.87	-4.30	31.94	-137.31	18.49	1019.94
14.99	81.05	-22.34	-26.88	600.67	499.25	722.69
59.69	57.44	22.36	-50.49	-1128.80	499.78	2549.51
43.34	70.70	6.00	-37.23	-223.51	36.04	1386.32
373.39	**1079.32**	**0**	**0**	**5905.96**	**2563.51**	**38144.04**
					256.35	**3814.40**
				0.6		

Sostituendo i valori ottenuti, la covarianza è, dunque, pari a:

$$\sigma_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)(y_i - \mu_Y) = \frac{1}{10} \times 5905,96 = 590,6$$

b)

Sostituendo i valori calcolati, il coefficiente di correlazione, nella formula classica, risulta pari a:

$$\rho_{XY} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)(y_i - \mu_Y)}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_X)^2\right) \times \left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_Y)^2\right)}}$$

$$= \frac{590,6}{\sqrt{\frac{1}{10}2563,51 \times \frac{1}{10}38144,04}} = \frac{590,6}{\sqrt{256,35 \times 3814,4}} = 0,6$$

<u>2. Formule alternative</u>

$$\sigma_{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \mu_X \mu_Y$$

$$\rho_{XY} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \mu_X \mu_Y}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu_X^2\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \mu_Y^2\right)}}$$

Nella tabella le quantità necessarie al calcolo della covarianza e del coefficiente di correlazione, secondo le formule alternative:

Effetto serra X	**Valore aggiunto Y**	$x_i y_i$	x_i^2	y_i^2
31.71	104.59	3316.11	1005.32	10938.37
68.71	270.87	18612.95	4721.71	73372.09
38.85	121.90	4735.40	1509.16	14858.53
17.83	35.69	636.27	317.75	1274.06
38.36	111.21	4265.83	1471.35	12367.80
26.86	86.01	2310.55	721.70	7397.32
33.04	139.87	4621.16	1091.59	19563.32
14.99	81.05	1215.33	224.85	6569.02
59.69	57.44	3428.84	3563.43	3299.33
43.34	70.70	3064.23	1878.51	4998.36
		46206.65	**16505.38**	**154638.20**

Sostituendo:

a)

La covarianza è pari a:

$$\sigma_{XY} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \mu_X \mu_Y = \frac{1}{10} 46206,65 - 37,34 \times 107,93 = 4620,67 - 4030,07 = 590,6$$

b)

Il coefficiente di correlazione è:

$$\rho_{XY} = \frac{\operatorname{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \mu_X \mu_Y}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu_X{}^2\right) \times \left(\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \mu_Y{}^2\right)}} =$$

$$= \frac{590,6}{\sqrt{\left(\frac{1}{10} 16505,38 - 37,34^2\right) \times \left(\frac{1}{10} 154638,2 - 107,93^2\right)}} =$$

$$= \frac{590,6}{\sqrt{(1650,54 - 1394,19) \times (15463,82 - 11649,42)}} =$$

$$= \frac{590,6}{\sqrt{256,35 \times 3814,4}} = \frac{590,6}{\sqrt{977825,63}} = 0,6$$

ESERCIZIO 2

Si calcoli il coefficiente di correlazione per la seguente distribuzione doppia dei caratteri EFFETTO SERRA (X) e VALORE AGGIUNTO (Y):

X \ Y	0-25	25-50	50-100	100-300	**Totale**
0-10	5	2	0	0	**7**
10-30	0	4	2	0	**6**
30-70	0	0	2	5	**7**
Totale	**5**	**6**	**4**	**5**	**20**

Soluzione

Per le distribuzioni doppie il coefficiente di correlazione è ottenuto come segue:

$$\rho_{XY} = \frac{\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{h}\hat{x}_i\hat{y}_j n_{ij} - \mu_X\mu_Y}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{k}\hat{x}_i^2 n_{i\bullet} - \mu_X^{\ 2}\right)\left(\frac{1}{n}\sum_{i=1}^{k}\hat{y}_j^2 n_{\bullet j} - \mu_Y^{\ 2}\right)}}$$

Nella seguente tabella sono riportati sulla prima riga e sulla prima colonna i valori centrali $\hat{x}_i$ e $\hat{y}_j$ delle classi di X ed Y ed all'interno delle celle i prodotti $\hat{x}_i\hat{y}_j n_{ij}$:

$\hat{x}_i\hat{y}_j n_{ij}$	**12.5**	**37.5**	**75**	**200**	
5	312.5	375	0	0	
20	0	3000	3000	0	
50	0	0	7500	50000	
					64187.5

Ad esempio il valore 312,5 della prima cella è il risultato del calcolo:

$$\underset{\hat{x}_i}{5} \times \underset{\hat{y}_j}{12{,}5} \times \underset{n_{ij}}{5}$$

La somma delle celle della tabella è pari a:

$$\sum_i\sum_j \hat{x}_i\hat{y}_j n_{ij} = 64187{,}5$$

La tabella che segue contiene tutte le quantità necessarie al calcolo del coefficiente di correlazione:

$\hat{x}_i$	$n_{i.}$	$\hat{y}_j$	$n_{.j}$	$\hat{x}_i n_{i.}$	$\hat{y}_j n_{.j}$	$\hat{x}_i^2$	$\hat{y}_j^2$	$\hat{x}_i^2 n_{i.}$	$\hat{y}_j^2 n_{.j}$
5	7	12.5	5	35	62.5	25	156.25	175	781.25
20	6	37.5	6	120	225	400	1406.25	2400	8437.5
50	7	75	4	350	300	2500	5625	17500	22500
		200	5	**505**	1000		40000	**20075**	200000
					1587.5				**231718.75**

$$\mu_X = \frac{1}{n}\sum_i \hat{x}_i n_{i.} = \frac{505}{20} = 25,25$$

$$\mu_Y = \frac{1}{n}\sum_j \hat{y}_j n_{.j} = \frac{1587,5}{20} = 79,38$$

$$\sigma_{XY} = \frac{1}{n}\sum_i \sum_j \hat{x}_i \hat{y}_j n_{ij} - \mu_x \mu_y = \frac{64187,5}{20} - 25,25 \times 79,38 = 1205,16$$

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{k} c_i^2 n_i - \mu_x^2} = \sqrt{\frac{1}{20}20075 - 25,25^2} = \sqrt{1003,75 - 637,56} = 19,14$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum_{j=1}^{h} c_j^2 n_j - \mu_y^2} = \sqrt{\frac{1}{20}231718,75 - 79,38^2} = \sqrt{11585,94 - 6300,39} = 72,7$$

Il coefficiente di correlazione è, infine:

$$\rho_{XY} = \frac{1205,16}{19,14 \times 72,7} = 0,86$$