Università degli Studi di Cassino, Anno accademico 2004-2005
Corso di Statistica 2, Prof. M. Furno

Esercitazione del 15/03/2005
dott. Claudio Conversano

Esercizio 1
La distribuzione di un campione di appartamenti in affitto a Cassino, in base alla superficie in mq ed al canone mensile in Euro è la seguente:

	Canone mensile				
Mq	200|-300	300|-400	400|-500	500|-600	Totale
40|-60	20	16	4	0	40
60|-80	24	92	22	2	140
80|-120	0	32	50	18	100
120|-180	0	0	6	14	20
Totale	44	140	82	34	300

Determinare i parametri della retta di regressione (costi – superficie) e calcolare il coefficiente di correlazione lineare.

SVOLGIMENTO

Per una distribuzione doppia i parametri della retta di regressione sono:

$$\hat{\beta}=\frac{Cov(X,Y)}{Var(X)}=\frac{E(X,Y)-E(X)E(Y)}{E(X^2)-\left[E(X)\right]^2}=\frac{\sum_{i=1}^{k}\sum_{j=1}^{h}\hat{x}_i\hat{y}_j p(\hat{x}_i\hat{y}_j)-\left(\sum_{i=1}^{k}\hat{x}_i p(x_i)\cdot\sum_{j=1}^{h}\hat{y}_j p(\hat{y}_j)\right)}{\sum_{i=1}^{k}\hat{x}_i^2 p(\hat{x}_i)-\left(\sum_{i=1}^{k}\hat{x}_i p(\hat{x}_i)\right)^2}$$

$$\hat{\alpha}=E(Y)-\hat{\beta}E(X)=\sum_{j=1}^{h}\hat{y}_j p(\hat{y}_j)-\hat{\beta}\sum_{i=1}^{k}\hat{x}_i p(x_i)$$

Per calcolare il termine $\sum_{i=1}^{k}\sum_{j=1}^{h}\hat{x}_i\hat{y}_j p(\hat{x}_i\hat{y}_j)$ è consigliabile costruire la tabella delle $\hat{x}_i\hat{y}_j p(\hat{x}_i\hat{y}_j)$, ossia:

$\hat{x}_i$ \ $\hat{y}_i$	250	350	450	550
50	833	933	300	0
70	1400	7513	2310	257
100	0	3733	7500	3300
150	0	0	1350	3850

La somma degli elementi all'interno di tale tabella è pari a $\sum_{i=1}^{k}\sum_{j=1}^{h}\hat{x}_i\hat{y}_j p(\hat{x}_i\hat{y}_j)=33280$ che corrisponde al momento misto $E(X,Y)$, primo termine della formula della covarianza.

Per il calcolo della covarianza, dei parametri della retta di regressione e della correlazione è utile considerare la seguente tabella:

$\hat{x}_i$	$p(\hat{x}_i)$	$\hat{y}_j$	$p(\hat{y}_j)$	$\hat{x}_i p(\hat{x}_i)$	$\hat{y}_j p(\hat{y}_j)$	$\hat{x}_i^2$	$\hat{x}_i^2 p(\hat{x}_i)$	$\hat{y}_j^2$	$\hat{y}_j^2 p(\hat{y}_j)$
50	0.13	250	0.15	6.67	36.67	2500	333.33	62500	9166.67
70	0.47	350	0.47	32.67	163.33	4900	2286.67	122500	57166.67
100	0.33	450	0.27	33.33	123.00	10000	3333.33	202500	55350.00
150	0.07	550	0.11	10.00	62.33	22500	1500.00	302500	34283.33
				82.67	385.33		7453.33		155966.67

$$Cov(XY) = 33280 - 82.67 \cdot 385.33 = 1425.78$$

$$Var(X) = 7453.33 - (82.67)^2 = 619.55$$

$$\hat{\beta} = \frac{1425.78}{619.55} = 2.30$$

$$\hat{\alpha} = 385.33 - 2.30 \cdot 82.67 = 195.09$$

Per una distribuzione doppia la correlazione è data da:

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{E(X,Y) - E(X)E(Y)}{\sqrt{E(X^2) - [E(X)]^2 \cdot E(Y^2) - [E(Y)]^2}} =$$

$$= \frac{\sum_{i=1}^{k}\sum_{j=1}^{h} \hat{x}_i \hat{y}_j p(\hat{x}_i \hat{y}_j) - \left(\sum_{i=1}^{k} \hat{x}_i p(x_i) \cdot \sum_{j=1}^{h} \hat{y}_j p(\hat{y}_j)\right)}{\sqrt{\left[\sum_{i=1}^{k} \hat{x}_i^2 p(\hat{x}_i) - \left(\sum_{i=1}^{k} \hat{x}_i p(\hat{x}_i)\right)^2\right] \cdot \left[\sum_{i=1}^{h} \hat{y}_j^2 p(\hat{y}_j) - \left(\sum_{i=1}^{k} \hat{y}_j p(\hat{y}_j)\right)^2\right]}}$$

Dai dati del campione risulta:

$$Corr(X,Y) = \frac{1425.78}{\sqrt{7453.33 \cdot (155966.67 - 385.33^2)}} = 0.66$$

Esiste correlazione diretta tra la superficie ed il canone mensile di locazione.