ESERCIZIO 1

X_1	7	13	16	27	44	61	85	100	125
X_2	3	5	5	7	9	10	12	13	14
X_3	4	8	10	15	24	37	51	62	80

a) Si determini la matrice di correlazione e, in base ai risultati ottenuti si indichi, giustificando la risposta, quale delle due rette ai minimi quadrati

$$X_1 = a + bX_2 \;\; o \;\; X_1 = c + dX_3$$

presenta il migliore adattamento.

b) Si confrontino $r_{12.3}$ *e* $r_{13.2}$ con i rispettivi coefficienti di correlazione grezzi e commentare.

c) Determinare l'equazione del piano ai minimi quadrati $X_1 = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3$, e valutarne la bontà di adattamento.

d) Si effettui l'analisi dei residui relativamente al piano di cui al punto *c*, commentando i risultati ottenuti.

$$\sum_{i=1}^{9} x_{1i} = 478, \qquad \overline{X}_1 = 53,\bar{1}, \qquad \sum_{i=1}^{9} x_{1i}^2 = 39710$$

$$\sum_{i=1}^{9} x_{2i} = 78, \qquad \overline{X}_2 = 8,\bar{6}, \qquad \sum_{i=1}^{9} x_{2i}^2 = 798$$

$$\sum_{i=1}^{9} x_{3i} = 291 \qquad \overline{X}_3 = 32,\bar{3}, \qquad \sum_{i=1}^{9} x_{3i}^2 = 15195$$

$$\sigma_{11} = \sigma^2\left(X_1\right) = 1591,4321 \quad \sigma_1 = 39,8928$$

$$\sigma_{22} = \sigma^2\left(X_2\right) = 13,\bar{5} \quad \sigma_2 = 3,6818$$

$$\sigma_{33} = \sigma^2\left(X_3\right) = 642,\bar{8} \quad \sigma_3 = 25,3553$$

$$\sum x_{1i}x_{2i} = 5431 \qquad \sigma_{12} = \sigma\left(X_1, X_2\right) = +143,\overline{148}$$

$$\sum x_{1i}x_{3i} = 24545 \qquad \sigma_{13} = \sigma\left(X_1, X_3\right) = +1009,9630$$

$$\sum x_{2i}x_{3i} = 3331 \qquad \sigma_{23} = \sigma\left(X_2, X_3\right) = +89,\bar{8}$$

a) $r_{ij} = \dfrac{\sigma_{ij}}{\sigma_i \cdot \sigma_j}$. Matrice di correlazione C:

	X_1	X_2	X_3
X_1	1	r_{12}	r_{13}
X_2	r_{21}	1	r_{23}
X_3	r_{31}	r_{32}	1

	X_1	X_2	X_3
X_1	1	0,9746	0,9985
X_2		1	0,9629
X_3			1

$$I_{1.2}^2 = r_{12}^2 = 0,9498$$

$$I_{1.3}^2 = r_{13}^2 = 0,9970$$

Poiché $I_{1.3}^2 > I_{12}^2$ è da preferirsi la retta $X_1 = c + dX_3$.

b)

$$r_{12.3} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}} = 0,8851 < r_{12} = 0,9746$$

Rispetto al coefficiente di correlazione grezzo, che è positivo e di valore elevato, la correlazione tra X_1 e X_2 al netto del contributo lineare di X_3 si mantiene positiva ed elevata, pur riducendosi leggermente.

$$r_{13.2} = \frac{r_{13} - r_{12} r_{23}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{23}^2\right)}} = 0{,}9950 < r_{13} = 0{,}9985.$$

La correlazione parziale tra X_1 e X_3 al netto del contributo lineare di X_2 è elevata e dello stesso ordine di grandezza di quella grezza.

c)

$$\alpha_2 = \frac{\sigma_{12}\sigma_{33} - \sigma_{13}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} = 1,9598$$

$$\alpha_3 = \frac{\sigma_{13}\sigma_{22} - \sigma_{12}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} = 1,2970$$

$$\alpha_1 = \bar{x}_1 - \alpha_2 \bar{x}_2 - \alpha_3 \bar{x}_3 = -5,8102$$

$$\boxed{\begin{aligned} X_1 &= \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 = \\ &= -5,8102 + 1,9598 X_2 + 1,2970 X_3 \end{aligned}}$$

$$I_{1.23}^2 = \frac{Varianza\ Spiegata}{Varianza\ Totale} = \frac{\hat{\alpha}_2\sigma_{12} + \hat{\alpha}_3\sigma_{13}}{\sigma_{11}} = 0,9994$$

adattamento molto buono del piano interpolatore.

Variazione della varianza spiegata dalla miglior retta $X_1 = c + dX_3$ al piano ai minimi quadrati: (la differenza indica la frazione della varianza totale che viene spiegata nel passare dalla retta al piano)

$$I_{12.3}^2 - I_{13}^2 = 0,0024.$$

Grado di miglioramento come riduzione relativa della varianza residua: $\dfrac{I^2_{1.23} - I^2_{1.3}}{1 - I^2_{1.3}} = 0,8$ (indica la frazione della varianza residua dalla retta che viene spiegata nel passare dalla retta al piano).

d)

X_1	$\hat{X}_1$	$X_1 - \hat{X}_1$	$\left(X_1 - \hat{X}_1\right)^2$
7	5,2572	1,7428	3,0374
13	14,3648	-1,3648	1,8627
16	16,9588	-0,9588	0,9193
27	27,3634	-0,3634	0,1321
44	42,956	1,044	1,0900
61	61,7768	-0,7768	0,6034
85	83,8544	1,1456	1,3124
100	100,0812	-0,0812	0,0066
125	125,387	-0,387	0,1498
478	**477,9996**		**9,1137**

$$\sum_{i=1}^{9}\left|x_{1i}-\hat{x}_{1i}\right| = 7,8644$$

$$A_1 = \frac{1}{N}\sum_{i=1}^{9}\left|x_{1i}-\hat{x}_{1i}\right| = \frac{7,8644}{9} = 0,8738 < 1$$

$$A_1^{'} = \frac{A_1}{\overline{\overline{x}}_1} = \frac{0,8738}{53,\bar{1}} = 0,0165.$$

L'ordine medio di grandezza dei residui in valore assoluto è l'1,65% del valore medio di $\overline{X}_1$.

$$A_2^{'} = \frac{A_2}{\overline{\overline{x}}_1} = \frac{\left[\frac{1}{N}\sum\left(x_{1i}-\hat{x}_{1i}\right)^2\right]^{\frac{1}{2}}}{\overline{\overline{x}}_1} =$$

$$= \frac{\sqrt{Var\ Residua}}{\overline{\overline{x}}_1} = 0,0189.$$

ESERCIZIO: Condurre la medesima analisi per la retta $\hat{X}_1 = c + dX_3$ e confrontare gli indici $A_1, A_1^{'}, A_2, A_2^{'}$ con quelli del piano.

ESERCIZIO 2 *Tema d'esame del 12.06.01*.

Su un collettivo di Paesi vengono rilevate le seguenti variabili:

X_1 = *speranza di vita della popolazione femminile*

X_2 = *speranza di vita della popolazione maschile*

X_3 = *calorie medie assunte giornalmente dalla popolazione*

X_4 = *percentuale della popolazione alfabetizzata.*

Si ottiene la seguente matrice di correlazione:

	X_1	X_2	X_3	X_4
X_1	1	0,982	0,775	0,865
X_2		1	0,765	0,809
X_3			1	0,682
				1

e le seguenti informazioni

	Media	σ
X_1	70,16	10,57
X_2	64,92	9,27
X_3	2753,83	567,83
X_4	78,34	22,88

a) Si determinino i parametri della retta interpolante $\hat{X}_1 = a + \alpha_{12} X_2$ e se ne fornisca la relativa interpretazione.

b) Si determinino i parametri del piano interpolante $\hat{X}_1 = b + \alpha_{12.3} X_2 + \alpha_{13.2} X_3$ e se ne fornisca la relativa interpretazione. Si confrontino α_{12} e $\alpha_{12.3}$.

c) Si calcoli la varianza spiegata del piano interpolante di cui al punto *b* e si valuti la bontà di adattamento del piano.

d) Si calcolino i coefficienti di correlazione parziale $r_{12.3}$ e $r_{24.3}$ fornendone l'interpretazione e confrontandoli con i rispettivi coefficienti grezzi.

e) Si valuti il miglioramento che si ottiene passando dal piano $\hat{X}_1 = b + \alpha_{12.3} X_2 + \alpha_{13.2} X_3$ all'interpolante che include anche X_4 tra le variabili esplicative.

SOLUZIONI:

a)

$$\alpha_{12} = \frac{\sigma_{12}}{\sigma_{22}} = \frac{\sigma\left(X_1, X_2\right)}{\sigma^2\left(X_2\right)} = \frac{(0,982)(10,57)(9,27)}{(9,27)^2} = 1,1197$$

$$a = \overline{X_1} - \alpha_{12}\overline{X_2} = 70,16 - (1,1197)(64,92) = -2,5309$$

$$\boxed{\hat{X}_1 = -2,5309 + 1,1197 X_2}$$

Il parametro α_{12} della retta indica la variazione di $\hat{X}_1$ che si ha in corrispondenza di un incremento unitario di X_2 al lordo delle variazioni di X_3.

b)

$$\sigma_{13} = r_{13} \cdot \sigma_1 \cdot \sigma_3 = \left(0,775\right)\left(10,57\right)\left(567,83\right) = 4651,5214$$
$$\sigma_{23} = r_{23} \cdot \sigma_2 \cdot \sigma_3 = \left(0,765\right)\left(9,27\right)\left(567,83\right) = 4026,7948$$
$$\sigma_{11} = \left(10,57\right)^2 = 111,7249$$
$$\sigma_{33} = \left(567,83\right)^2 = 322430,9089$$

$$\hat{X}_1 = b + \alpha_{12.3} X_2 + \alpha_{13.2} X_3$$

dove

$$\alpha_{12.3} = \frac{\sigma_{33}\sigma_{12} - \sigma_{13}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} = 1,0697$$

$$\alpha_{13.2} = \frac{\sigma_{22}\sigma_{13} - \sigma_{12}\sigma_{23}}{\sigma_{22}\sigma_{33} - \sigma_{23}^2} = 0,0011$$

$$b = \overline{X_1} - \alpha_{12.3}\overline{X_2} - \alpha_{13.2}\overline{X_3} = -2,3076$$

$$\boxed{\hat{X}_1 = -2,3076 + 1,0697 X_2 + 0,0011 X_3}$$

Il parametro $\alpha_{12.3}$ (*coefficiente di regressione netto*) del piano indica la variazione di $\hat{X}_1$ che si ha in corrispondenza di un incremento unitario di X_2 se X_3 resta costante.

Analogo significato ha il parametro $\alpha_{13.2}$.

c) *Varianza spiegata*:

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{x}_{1i}-\overline{x}\right)^{2}=Var\left(\hat{X}_{1}\right)=\alpha_{12.3}\sigma_{21}+\alpha_{13.2}\sigma_{31}=$$

$$=108,0434.$$

Bontà di adattamento del piano interpolatore:

$$I_{1.23}^{2}=\frac{Varianza\ spiegata}{Varianza\ totale}=\frac{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{x}_{1i}-\overline{x}\right)^{2}}{\frac{1}{N}\sum_{i=1}^{N}\left(x_{1i}-\overline{x}\right)^{2}}=$$

$$=\frac{108,0434}{111,7249}=0,9670.$$

Il piano interpolatore spiega circa il 96,7% della variabilità totale della speranza di vita della popolazione femminile.

d)

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{\left\{1 - r_{23}^2\right\} \cdot \left\{1 - r_{13}^2\right\}}} = 0{,}9561 < r_{12} = 0{,}982$$

Se le variabili X_1 *e* X_2 fossero incorrelate con X_3 la correlazione parziale $r_{12.3}$ sarebbe uguale a quella grezza r_{12}. Negli altri casi può esservi notevole differenza.

$$r_{24.3} = \frac{r_{24} - r_{23}r_{34}}{\sqrt{\left\{1 - r_{23}^2\right\} \cdot \left\{1 - r_{34}^2\right\}}} = 0,6099$$

$$< r_{24} = 0,809$$

La correlazione parziale tra X_2 *e* X_4 al netto del contributo lineare di X_3 è minore di quella grezza.

e) Riduzione relativa della varianza residua che si ha nel passare dal piano ai minimi quadrati

$$\hat{X}_1 = \alpha + \alpha_{12.3}X_2 + \alpha_{13.2}X_3$$

all'iperpiano ai minimi quadrati

$$\hat{X}_1 = \alpha + \alpha_{12.34} X_2 + \alpha_{13.24} X_3 + \alpha_{14.23} X_4:$$

$$r_{14.23}^2 = \frac{\left(r_{14.3} - r_{12.3} \cdot r_{24.3}\right)^2}{\left(1 - r_{12.3}^2\right)\left(1 - r_{24.3}^2\right)} =$$

$$= \frac{\left[0,7279 - \left(0,9561\right)\left(0,6099\right)\right]^2}{\left(1 - 0,9561^2\right)\left(1 - 0,6099^2\right)} = 0,3886$$

essendo

$$r_{14.3} = \frac{r_{14} - r_{13} r_{34}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{34}^2\right)}} = 0,7279.$$

La diminuzione relativa della varianza residua che si è avuta aggiungendo X_4 rappresenta circa il 38,86% della variabilità residua del modello $\hat{X}_1 = \alpha + \alpha_{12.3} X_2 + \alpha_{13.2} X_3$.

ESERCIZIO 3

Una società produttiva di beni di largo consumo ha recentemente lanciato sul mercato un nuovo prodotto. Volendo analizzare i risultati delle vendite sono state raccolte le informazioni seguenti in 20 aree geografiche:

$X_1 =$ *vendite del nuovo prodotto nell'ultimo mese (migliaia di euro).*

$X_2 =$ *spese per la campagna pubblicitaria (migliaia di euro).*

$X_3 =$ *numero di punti vendita*

$X_4 =$ *quota di mercato di una società concorrente (in percentuale).*

Sulla base di tali informazioni sono state ricavate le medie aritmetiche:

$\overline{X_1} = 145 \quad \overline{X_2} = 37,89 \quad \overline{X_3} = 5,31 \quad \overline{X_4} = 25,8$

;

la matrice di varianza e covarianza:

	X_1	X_2	X_3	X_4
X_1	$2131,78$	$488,04$	$81,18$	$213,69$
X_2		$122,06$	$21,39$	$62,87$
X_3			$4,28$	$12,54$
X_4				$45,01$

a) Si calcoli il coefficiente di correlazione parziale $r_{14.2}$ e lo si confronti con il corrispondente coefficiente di correlazione grezzo, commentando il risultato ottenuto.

b) Calcolare con il metodo dei minimi quadrati i parametri dei modelli

1. $\hat{X}_1 = a + \alpha_2 X_2$
2. $\hat{X}_1 = b + \alpha_{12.4} X_2 + \alpha_{14.2} X_4$

c) Valutare la bontà di adattamento di ciascuno dei modelli previsti al punto *b* e il miglioramento ottenuto passando dalla retta al piano.

d) Sapendo che per il modello

3. $\hat{X}_1 = c + \alpha_{12.3} X_2 + \alpha_{13.2} X_3$

risulta

$$\sum_{i=1}^{20} \left(x_{1i} - \hat{x}_{1i} \right)^2 = 2968{,}52$$

scegliere tra il modello 2. e il modello 3. quello che presenta il migliore adattamento.

SOLUZIONI:

a) *Matrice di correlazione*

	X_1	X_2	X_3	X_4
X_1	1	0,9567	0,8499	0,6899
X_2		1	0,9358	0,8482
X_3			1	0,9035
X_4				1

$$r_{14.2} = \frac{r_{14} - r_{12} r_{42}}{\sqrt{\left\{1 - r_{12}^2\right\} \cdot \left\{1 - r_{42}^2\right\}}} = -0{,}7886$$

Il coefficiente di correlazione parziale

$r_{14.2}$ informa che, al netto dell'influenza della spesa per la campagna pubblicitaria X_2, all'aumentare di X_4 tende a diminuire X_1. Osserviamo differenza di segno tra $r_{14.2}$ e r_{14}.

b)

$$\alpha_2 = \frac{\sigma_{12}}{\sigma_{22}} = \frac{Cov\left(X_1, X_2\right)}{\sigma^2\left(X_2\right)} = \frac{488,04}{122,06} = 3,9984$$

$$a = \overline{x_1} - \alpha_2 \overline{x_2} = -6,4992$$

$$\boxed{\hat{X}_1 = -6,4992 + 3,9984 X_2}$$

$$\alpha_{12.4} = \frac{\sigma_{44}\sigma_{12} - \sigma_{14}\sigma_{24}}{\sigma_{22}\sigma_{44} - \sigma_{24}^2} = 5,5356$$

$$\alpha_{14.2} = \frac{\sigma_{22}\sigma_{14} - \sigma_{12}\sigma_{24}}{\sigma_{44}\sigma_{22} - \sigma_{24}^2} = -2,9846$$

$$b = \overline{x_1} - \alpha_{12.4}\overline{x_2} - \alpha_{14.2}\overline{x_4} = 12,2588$$

$$\boxed{\hat{X}_1 = 12,2588 + 5,5356 X_2 - 2,9846 X_4}$$

c) Per il modello 1. $I_{1.2}^2 = r_{12}^2 = 0,9153.$

Per il modello 2. si ha

$$\frac{Varianza\ Spiegata}{Varianza\ Totale}=\frac{Var\left(\hat{X}_1\right)}{Var\left(X_1\right)}=\frac{\alpha_{12.4}\sigma_{21}+\alpha_{14.2}\sigma_{14}}{\sigma_{11}}=$$

$$=I^2_{1.24}=\frac{2063,8151}{2131,78}=0,9681.$$

<u>Miglioramento</u>: $I^2_{1.24}-I^2_{1.2}=0,0528$

(indica la frazione di varianza totale che viene spiegata nel passare dalla retta 1. al piano 2.).

<u>Grado di miglioramento</u>:

$$\frac{I^2_{1.24}-I^2_{1.2}}{1-I^2_{1.2}}=0,6233$$

(il 62,33% della varianza residua della retta viene spiegata nel passare dalla retta 1. al piano 2.)

d)Modello 3.

$$\frac{Varianza\ Spiegata}{Varianza\ Totale}=\frac{Var\left(\hat{X}_1\right)}{Var\left(X_1\right)}=\frac{\alpha_{12.3}\sigma_{21}+\alpha_{13.2}\sigma_{13}}{\sigma_{11}}=$$

$$=I^2_{1.23}=\frac{1876,76}{2131,78}=0,8804.$$

Si sceglie il modello 2. poiché presenta un migliore adattamento.

Esempio

Si considerino i dati in tabella:

Anno	X_1	X_2	X_3
1948	100	100	100
1949	106	104	99
1950	107	106	110
1951	120	111	126
1952	110	111	113
1953	116	115	103
1954	123	120	102
1955	133	124	103
1956	137	126	98

X_1: numero indice delle importazioni di beni e servizi nel Regno Unito, a prezzi costanti (1948)

X_2: numero indice del prodotto lordo del Regno Unito ai prezzi 1948

X_3: rapporto tra indici dei prezzi delle importazioni e del prodotto totale del Regno Unito.

$$n = 9 \quad \sum x_{1i} = 1052 \quad \sum x_{2i} = 1017 \quad \sum x_{3i} = 954$$

$$\bar{x}_1 = 116{,}9 \quad \bar{x}_2 = 113 \quad \bar{x}_3 = 106$$

$\sum x_{1i}^2 = 124228 \quad \sum x_{2i}^2 = 115571 \quad \sum x_{3i}^2 = 101772$

$\sum x_{1i}x_{2i} = 119750 \quad \sum x_{1i}x_{3i} = 111433$

$\sum x_{2i}x_{3i} = 107690.$

$\sigma_{11} = 140{,}09\overline{8} \quad \sigma_{22} = 72{,}\overline{2} \quad \sigma_{33} = 72$

$\sigma_{12} = 97{,}\overline{1} \quad \sigma_{13} = -8{,}\overline{7} \quad \sigma_{23} = -12{,}\overline{4}$

Si ottiene il seguente piano:

$\hat{X}_1 = -49{,}3297 + 1{,}3642X_2 + 0{,}1139X_3$

$$I_{1.23}^2 = \frac{\hat{\alpha}_2\sigma_{12} + \hat{\alpha}_3\sigma_{13}}{\sigma_{11}} = \frac{1{,}3642 \cdot 97{,}\overline{1} - 0{,}1139 \cdot 12{,}\overline{4}}{140{,}09\overline{8}} =$$

$= 0{,}9355$

$H_0 : \alpha_2 = \alpha_3 = 0$

$H_1 : \alpha_2 \neq 0 \cup \alpha_3 \neq 0$

$$V = \frac{I_{1.23}^2 / 2}{\left(1 - I_{1.23}^2\right) / (9-3)} = 43{,}51 > F_{0,01}(2,6) = 10{,}925$$

Rifiuto l'ipotesi nulla. I coefficienti di regressione sono congiuntamente significativamente diversi da zero.

$$\sum (x_{1i} - \hat{x}_{1i})^2 = n\sigma_{11}\left(1 - I_{1.23}^2\right) = 9 \cdot 140{,}09\overline{8}(1 - 0{,}9355) =$$

$= 81{,}3276$

$$\hat{\sigma}^2 = \frac{81{,}3276}{9-3} = 13{,}5546$$

$H_0 : \alpha_2 = 0$

$H_1 : \alpha_2 \neq 0$

$$Var(\hat{\alpha}_2) = \frac{\sigma_{33}}{n \cdot (\sigma_{22}\sigma_{33} - \sigma_{23}^2)} \cdot \sigma^2 =$$

$$= \frac{72}{9(72,\overline{2} \cdot 72 - 12,4^2)} \sigma^2 .$$

Si rifiuta H_0 se

$$\frac{|1,3642 - 0|}{\sqrt{13,5546 \cdot 0,001585}} = 9,3072 > t_{1-\frac{\alpha}{2}}(6)$$

$t_{0,9995}(6) = 5,959$. Il coefficiente di regressione di X_2 è significativamente diverso da zero.

(la determinazione degli intervalli di confidenza al 90% e al 95% per i coefficienti di regressione è lasciata per esercizio).