ES. REGRESSIONE MULTIPLA - INVENTATO

α	x_1	x_2	x_3	y
1	23	3	4	5
1	31	1	8	4
1	42	5	10	8
1	33	5	16	8
1	25	2	12	6
1	29	7	14	7

$\bar{x}_1 = 30.5$

$\bar{x}_2 = 3.8\bar{3}$

$\bar{x}_3 = 10.67$

$\bar{y} = 6.33$

$n = 6$

$k = 3$

1) Qual è l'equazione di previsione?
2) Applicare un t-test per l'intercetta e per ogni coefficiente angolare; con $\alpha = 0.05$ ~~rifiuto~~ rifiuteresti l'ip. nulla?
3) Applicare il test F per verificare la significatività del modello completo con $\alpha = 0.05$
4) Applicare il test F per ~~ul~~ confronto fra il modello ridotto con solo x_1 e il modello completo ($\alpha = 0.05$) $(q = 2)$

SVOLGIMENTO

1) $\hat{y} = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 = BX + \varepsilon$

$$\boxed{B = (X' \cdot X)^{-1} \cdot X'Y}$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 23 & 31 & 42 & 33 & 25 & 29 \\ 3 & 1 & 5 & 5 & 2 & 7 \\ 4 & 8 & 10 & 16 & 12 & 14 \end{bmatrix} \cdot X = \begin{bmatrix} 1 & 23 & 3 & 4 \\ 1 & 31 & 1 & 8 \\ 1 & 42 & 5 & 10 \\ 1 & 33 & 5 & 16 \\ 1 & 25 & 2 & 12 \\ 1 & 29 & 7 & 14 \end{bmatrix} \Rightarrow X' \cdot X = \begin{bmatrix} 6 & 183 & 23 & 64 \\ 183 & 809 & 728 & 1994 \\ 23 & 728 & 113 & 272 \\ 64 & 1994 & 272 & 776 \end{bmatrix}$$

$$\Rightarrow (X' \cdot X)^{-1} = \begin{bmatrix} 4.614 & -0.127 & 0.0617 & -0.075 \\ -0.127 & 0.0051 & -0.0043 & -0.0011 \\ 0.0617 & 0.0043 & 0.062 & -0.0157 \\ -0.075 & -0.0011 & -0.0157 & 0.0157 \end{bmatrix} \rightarrow (X' \cdot X)^{-1} \cdot X' \cdot Y = \begin{bmatrix} 1.17 \\ 0.08 \\ 0.36 \\ 0.13 \end{bmatrix} (= B)$$

$$X' \cdot Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 23 & 31 & 42 & 33 & 25 & 29 \\ 3 & 1 & 5 & 5 & 2 & 7 \\ 4 & 8 & 10 & 16 & 12 & 14 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 4 \\ 8 \\ 8 \\ 6 \\ 7 \end{bmatrix} = \begin{bmatrix} 38 \\ 1192 \\ 160 \\ 430 \end{bmatrix}$$

$\rightarrow$ La retta di previsione è $\underline{\hat{y} = 1.17 + 0.08 x_1 + 0.36 x_2 + 0.13 x_3}$

α (l'intercetta) mi dice che: il valore atteso della var. dipendente, quando <u>tutte</u> le var. indipendenti sono uguali a 0, è 1.17.

β_1 (coefficiente angolare di x_1) mi dice che: all'~~au~~ aumentare (o diminuire) di un'unità della var. indipendente x_1, <u>controllando le altre var. indipendenti</u>, il valore atteso della var. dipendente aumenta (o ~~dimin~~ diminuisce) di 0.08.

$\rightarrow$ L'interpretazione degli altri coefficienti è analoga, variando semplicemente il valore considerato [$x_2 \rightarrow 0.36$; $x_3 \rightarrow 0.13$]

2) $\boxed{t_i = \dfrac{\beta_i - 0}{\sqrt{Se^2 \cdot (X' \cdot X)^{-1}_{ii}}}}$ dove $\boxed{Se^2 = \dfrac{e' \cdot e}{n-k-1}}$ dove $\boxed{e' \cdot e = SQerr_c = \sum (y - \hat{y})^2 = (y - \hat{y})' \cdot (y - \hat{y})}$

$$\hat{Y} = X \cdot B \Rightarrow \hat{Y} = \begin{bmatrix} 1 & 23 & 3 & 4 \\ 1 & 31 & 1 & 8 \\ 1 & 42 & 5 & 10 \\ 1 & 33 & 5 & 16 \\ 1 & 25 & 2 & 12 \\ 1 & 29 & 7 & 14 \end{bmatrix} \cdot \begin{bmatrix} 1.17 \\ 0.08 \\ 0.36 \\ 0.13 \end{bmatrix} = \begin{bmatrix} 4.59 \\ 5.03 \\ 7.59 \\ 7.62 \\ 5.41 \\ 7.76 \end{bmatrix}$$

$$e' \cdot e = Sqerr_c \Rightarrow \left(\begin{bmatrix} 5 \\ 4 \\ 8 \\ 8 \\ 6 \\ 7 \end{bmatrix} - \begin{bmatrix} 4.59 \\ 5.03 \\ 7.59 \\ 7.62 \\ 5.41 \\ 7.76 \end{bmatrix} \right)' \cdot \left(\begin{bmatrix} 5 \\ 4 \\ 8 \\ 8 \\ 6 \\ 7 \end{bmatrix} - \begin{bmatrix} 4.59 \\ 5.03 \\ 7.59 \\ 7.62 \\ 5.41 \\ 7.76 \end{bmatrix} \right) = \left(\begin{bmatrix} 0.41 \\ -1.03 \\ 0.41 \\ 0.38 \\ 0.59 \\ -0.76 \end{bmatrix} \right)' \cdot \left(\begin{bmatrix} 0.41 \\ -1.03 \\ 0.41 \\ 0.38 \\ 0.59 \\ -0.76 \end{bmatrix} \right) =$$

$$= \begin{bmatrix} 0.41 & -1.03 & 0.41 & 0.38 & 0.59 & -0.76 \end{bmatrix} \cdot \begin{bmatrix} 0.41 \\ -1.03 \\ 0.41 \\ 0.38 \\ 0.59 \\ -0.76 \end{bmatrix} = \underline{2.465} \Rightarrow S_e^2 = \frac{2.465}{6-3-1} = 1.233$$

-) $\boxed{(X' \cdot X)^{-1}_{ii}}$ è la cella della diagonale della matrice $(X' \cdot X)^{-1}$ corrispondente a ciò che sto testando:

$$\begin{array}{cccc} \text{intercetta} = \alpha & \beta_1 & \beta_2 & \beta_3 \end{array}$$

$$\begin{bmatrix} 4.614 & & & \\ & 0.0051 & & \\ & & 0.062 & \\ & & & 0.0157 \end{bmatrix}$$

-) $H_0 = \beta_i = 0$; $H_a = \beta_i \neq 0$

$H_0 = \alpha = 0$; $H_a = \alpha \neq 0$

$$\Rightarrow t_1 (\text{intercetta}) = \frac{1.172}{\sqrt{1{,}233 \cdot 4.614}} = \underline{0.49} \; ; \; t_2(\beta_1) = \frac{0.08}{\sqrt{1.233 + 0.0051}} = \underline{1.013}$$

$$t_3(\beta_2) = \frac{0.36}{\sqrt{1.233 + 0.062}} = \underline{1.29} \; ; \; t_4(\beta_3) = \frac{0.13}{\sqrt{1.233 + 0.0157}} = \underline{0.91}$$

-) Dalla tavola B a pag. 527 del libro vedo che il t-score associato a $\alpha = 0.05$, per 4 gdl, è pari a 2.776. Essendo nessuna delle statistiche test trovate maggiori di questo numero, non rifiuto l'ipotesi nulla per nessuno dei parametri testati. $\Rightarrow$ {nessuno dei coefficienti angolari prevede in modo significativo la Y, controllando gli altri}

3) $$\boxed{F_{(K; n-K-1)} = \frac{SQreg_c}{SQ\,err_c} \cdot \frac{n-k-1}{K}}$$ dove $\boxed{SQreg_c = \sum(\hat{y} - \bar{y})^2 = (\hat{y} - \bar{y})' \cdot (\hat{y} - \bar{y})}$

dove $\boxed{SQ\,tot = \sum(y - \bar{y})^2 = (y - \bar{y})' \cdot (y - \bar{y})}$

$$SQreg_c = \left(\begin{bmatrix} 4.59 \\ 5.03 \\ 7.59 \\ 7.62 \\ 5.41 \\ 7.76 \end{bmatrix} - \begin{bmatrix} 6.33 \\ 6.33 \\ // \\ // \\ // \\ // \end{bmatrix} \right)' \cdot \left(\begin{bmatrix} 4.59 \\ 5.03 \\ 7.59 \\ 7.62 \\ 5.41 \\ 7.76 \end{bmatrix} - \begin{bmatrix} 6.33 \\ 6.33 \\ // \\ // \\ // \\ // \end{bmatrix} \right) = \left(\begin{bmatrix} -1.74 \\ 1.31 \\ 1.25 \\ 1.29 \\ -0.93 \\ 1.43 \end{bmatrix} \right)' \cdot \left(\begin{bmatrix} -1.74 \\ 1.31 \\ 1.25 \\ 1.29 \\ -0.93 \\ 1.43 \end{bmatrix} \right) =$$

$$= \begin{bmatrix} -1.74 & 1.31 & 1.25 & 1.29 & -0.93 & 1.43 \end{bmatrix} \cdot \begin{bmatrix} -1.74 \\ 1.31 \\ 1.25 \\ 1.29 \\ -0.93 \\ 1.43 \end{bmatrix} = \underline{10.87}$$

$$SQtot = \left(\begin{bmatrix} 5\\4\\8\\8\\6\\7 \end{bmatrix} - \begin{bmatrix} 6.33\\ //\\ //\\ //\\ //\\ // \end{bmatrix} \right)' \cdot \left(\begin{bmatrix} 5\\4\\8\\8\\6\\7 \end{bmatrix} - \begin{bmatrix} 6.33\\ //\\ //\\ //\\ //\\ // \end{bmatrix} \right) = \left(\begin{bmatrix} -1.33\\2.33\\1.67\\1.67\\-0.33\\0.67 \end{bmatrix} \right)' \cdot \left(\begin{bmatrix} -1.33\\2.33\\1.67\\1.67\\-0.33\\0.67 \end{bmatrix} \right) =$$

→ Il calcolo di SQtot serve per il punto 4 ma l'ho fatto qui per comodità

$$= \begin{bmatrix} -1.33 & 2.33 & 1.67 & 1.67 & -0.33 & 0.67 \end{bmatrix} \cdot \begin{bmatrix} -1.33\\2.33\\1.67\\1.67\\-0.33\\0.67 \end{bmatrix} = \underline{13.33}$$

$$\Rightarrow F_{(3;2)} = \frac{10.87}{\cancel{4.93}\,2.465} \cdot \frac{2}{3} = \underline{2.94}$$

→ Dalla tavola D a pag. 529 del libro, trovo che: per (3;2) gradi di libertà la statistica F "critica" ha un valore di 19.16. Essendo 2.94 minore di questo valore non ho sufficiente evidenza per rifiutare l'ipotesi nulla. ⇒ {Il modello non spiega una quota di varianza statistic. significativa}

4) $\boxed{F_{(q;n-k-1)} = \frac{R^2_c - R^2_r}{1 - R^2_c} \cdot \frac{n-k-1}{q}}$ dove $\boxed{R^2_c = \frac{SQreg\,c}{SQtot}}$ dove $\boxed{R^2_r = \frac{SQreg\,r}{SQtot}}$

⇒ $SQreg\,r = 4.79$ (questo è un dato ricavato da calcoli a parte)

$$\Rightarrow R^2_c = \frac{10.87}{13.33} = \underline{0.82} = 82\% \quad ; \quad R^2_r = \frac{4.79}{13.33} = \underline{0.36 = 36\%}$$

$$\Rightarrow F_{(q;n-k-1)} = \frac{0.82 - 0.36}{1 - 0.82} \cdot \frac{2}{2} = \underline{2.466}$$

→ Dalla tavola D a pag. 529 del libro trovo che: per (2;2) gdl la statistica F "critica" ha un valore di 19.00. Essendo 2.466 minore di questo valore, non ho sufficiente evidenza per rifiutare l'ipotesi nulla.

⇒ {L'utilizzo del ~~un~~ modello completo rispetto quello ridotto non spiega una quota di varianza significativamente ~~maggiore~~, più alta}