ES. REGRESSIONE SEMPLICE - INVENTATO

X	Y
3	5
5	4
7	7
8	6
11	7
12	8

$\bar{x} = 7.67$

$\bar{y} = 6.17$

$n = 6$

1) Qual è l'equazione di previsione?
2) Quali sono i valori di R^2 e r?
3) Qual è la statistica test associata al coefficiente angolare? Con $\alpha = 0.05$ rifiuteresti l'ipotesi nulla?

SVOLGIMENTO

1) $\boxed{\hat{y} = \alpha + \beta x} \rightarrow \boxed{\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum y_i \\ \sum (x_i \cdot y_i) \end{bmatrix}} =$

$$= \begin{bmatrix} 6 & 46 \\ 46 & 412 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 37 \\ 305 \end{bmatrix} \overset{*}{=} \begin{bmatrix} 1.16 & -0.13 \\ -0.13 & 0.017 \end{bmatrix} \cdot \begin{bmatrix} 37 \\ 305 \end{bmatrix} = \begin{bmatrix} 3.41 \\ 0.36 \end{bmatrix} \rightarrow \hat{y} = 3.41 + 0.36x$$

α (l'intercetta) mi dice che: il valore atteso della var. dipendente, quando la var. indipendente è uguale a 0, è 3.41.

β (il coefficiente angolare) mi dice che: all' ~~variare~~ aumentare (o diminuire) di un' unità della var. indipendente, il valore atteso di y aumenta (o diminuisce) di 0.36.

$*$ $\boxed{Det = a \cdot d - b \cdot c} = (6 \cdot 412) - (46 \cdot 46) = 356 \rightarrow$ la matrice è invertibile

$$\Rightarrow \begin{bmatrix} 6 & 46 \\ 46 & 412 \end{bmatrix}^{-1} = \begin{bmatrix} 412/356 & -46/356 \\ -46/356 & 6/356 \end{bmatrix} \simeq \begin{bmatrix} 1.16 & -0.13 \\ -0.13 & 0.017 \end{bmatrix}$$

2) $\boxed{R^2 = \frac{SQreg}{SQtot}} \rightarrow \boxed{SQreg = \sum(\hat{y}_i - \bar{y})^2 = (\hat{y} - \bar{y})' \cdot (\hat{y} - \bar{y})}$ $\boxed{SQtot = \sum(y_i - \bar{y})^2 = (y_i - \bar{y})' \cdot (y_i - \bar{y})}$

$$\hat{y}_i = \begin{bmatrix} 4.49 \\ 5.21 \\ 5.93 \\ 6.29 \\ 7.37 \\ 7.73 \end{bmatrix} \rightarrow SQreg = \left(\begin{bmatrix} 4.49 \\ 5.21 \\ 5.93 \\ 6.29 \\ 7.37 \\ 7.73 \end{bmatrix} - \begin{bmatrix} 6.17 \\ 6.17 \\ 6.17 \\ 6.17 \\ 6.17 \\ 6.17 \end{bmatrix} \right)' \cdot \left(\begin{bmatrix} 4.49 \\ 5.21 \\ 5.93 \\ 6.29 \\ 7.37 \\ 7.73 \end{bmatrix} - \begin{bmatrix} 6.17 \\ 6.17 \\ // \\ // \\ // \\ // \end{bmatrix} \right) =$$

ottenuta calcolando $\hat{y} = 3.41 + 0.36x$ per ognuno dei valori di X.

$$= \begin{pmatrix} -1.68 \\ -0.96 \\ -0.24 \\ 0.12 \\ 1.20 \\ 1.56 \end{pmatrix}' \cdot \begin{pmatrix} -1.68 \\ -0.96 \\ -0.24 \\ 0.12 \\ 1.20 \\ 1.56 \end{pmatrix} =$$

$$= \begin{bmatrix} -1.68 & -0.96 & -0.24 & 0.12 & 1.20 & 1.56 \end{bmatrix} \cdot \begin{bmatrix} -1.68 \\ -0.96 \\ -0.24 \\ 0.12 \\ 1.20 \\ 1.56 \end{bmatrix} = 7.67$$

$$SQ_{tot} = \left(\begin{bmatrix}5\\4\\7\\6\\7\\8\end{bmatrix} - \begin{bmatrix}6.17\\6.17\\ \text{"}\\ \text{"}\\ \text{"}\\ \text{"}\end{bmatrix} \right)' \cdot \left(\begin{bmatrix}5\\4\\7\\6\\7\\8\end{bmatrix} - \begin{bmatrix}6.17\\6.17\\ \text{"}\\ \text{"}\\ \text{"}\\ \text{"}\end{bmatrix} \right) = \left(\begin{bmatrix}-1.17\\-2.17\\0.83\\-0.17\\0.83\\1.83\end{bmatrix} \right)' \cdot \left(\begin{bmatrix}-1.17\\-2.17\\0.83\\-0.17\\0.83\\1.83\end{bmatrix} \right) =$$

$$= \begin{bmatrix}-1.17 & -2.17 & 0.83 & -0.17 & 0.83 & 1.83\end{bmatrix} \cdot \begin{bmatrix}-1.17\\-2.17\\0.83\\-0.17\\0.83\\1.83\end{bmatrix} = \underline{10.83}$$

$$R^2 = \frac{7.67}{10.83} = \underset{(=71\%)}{0.71} \rightarrow \boxed{r = \sqrt{R^2}} \Rightarrow \underline{r = \sqrt{0.71} = 0.84}$$

R^2 mi indica che utilizzando $\hat{y}$, per prevedere i valori di y, invece di $\bar{y}$ commetto il 71% di errori in meno.

Un altro modo di interpretarlo è: $\hat{y}$ spiega il 71% della varianza di y.

r mi indica che: all'aumentare (o diminuire) di una deviazione standard della var. indipendente, y aumenta (o diminuisce) di 0.84 deviazioni standard.

3) $$\boxed{t = \frac{\beta - 0}{\sqrt{\frac{\sigma^2}{\sum(x-\bar{x})^2}}}} \quad \text{dove} \quad \boxed{\sigma^2 = \frac{\sum(y-\hat{y})^2}{n-2}} \quad \boxed{\left(\sum(y-\hat{y})^2 = SQ_{err}\right) \rightarrow \sum(y-\hat{y})^2 = (y-\hat{y})' \cdot (y-\hat{y})}$$

$$SQ_{err} = \left(\begin{bmatrix}5\\4\\7\\6\\7\\8\end{bmatrix} - \begin{bmatrix}4.49\\5.21\\5.93\\6.29\\7.37\\7.73\end{bmatrix} \right)' \cdot \left(\begin{bmatrix}5\\4\\7\\6\\7\\8\end{bmatrix} - \begin{bmatrix}4.49\\5.21\\5.93\\6.29\\7.37\\7.73\end{bmatrix} \right) = \left(\begin{bmatrix}0.51\\-1.21\\1.073\\-0.29\\-0.37\\0.28\end{bmatrix} \right)' \cdot \left(\begin{bmatrix}0.51\\-1.21\\1.073\\-0.29\\-0.37\\0.28\end{bmatrix} \right) =$$

$$= \begin{bmatrix}0.51 & -1.21 & 1.073 & -0.29 & -0.37 & 0.28\end{bmatrix} \cdot \begin{bmatrix}0.51\\-1.21\\1.073\\-0.29\\-0.37\\0.28\end{bmatrix} = \underline{3.16}$$

$$\Rightarrow \sigma^2 = \frac{3.16}{4} = 0.791$$

$$\boxed{\sum(x-\bar{x})^2 = (x-\bar{x})' \cdot (x-\bar{x})} \Rightarrow \left(\begin{bmatrix}3\\5\\7\\8\\11\\12\end{bmatrix} - \begin{bmatrix}7.67\\7.67\\ \text{"}\\ \text{"}\\ \text{"}\\ \text{"}\end{bmatrix} \right)' \cdot \left(\begin{bmatrix}3\\5\\7\\8\\11\\12\end{bmatrix} - \begin{bmatrix}7.67\\7.67\\ \text{"}\\ \text{"}\\ \text{"}\\ \text{"}\end{bmatrix} \right) = \left(\begin{bmatrix}-4.67\\-2.67\\-0.67\\0.33\\3.33\\4.33\end{bmatrix} \right)' \cdot \left(\begin{bmatrix}-4.67\\-2.67\\-0.67\\0.33\\3.33\\4.33\end{bmatrix} \right) =$$

$$= \begin{bmatrix}-4.67 & -2.67 & -0.67 & 0.33 & 3.33 & 4.33\end{bmatrix} \cdot \begin{bmatrix}-4.67\\-2.67\\-0.67\\0.33\\3.33\\4.33\end{bmatrix} = \underline{59.33}$$

$\Rightarrow t = \dfrac{3.16}{\sqrt{\dfrac{0.79}{59.33}}} = \underline{3.12}$ $\rightarrow$ dalla tavola B a pag. 527 del libro vedo che il ~~p-valore associato~~ a t-score associato a un $\alpha = 0.05$, per 4 gdl, è pari a 2.776.

$\Rightarrow$ Essendo $3.12 > 2.776$ posso concludere che il mio p-valore è minore di 0.05. Quindi rifiuto l'ipotesi nulla. $\rightarrow H_0: \beta = 0$; $H_a: \beta \neq 0$ $\Rightarrow$ {β prevede in modo significativo ~~[illegible]~~ la y}

$\rightarrow$ Il punto 3 è stato svolto nell'esercitazione del 28/05, per mancanza di tempo. Lo svolgimento l'ho comunque riportato qui per facilitarne la comprensione.

RISOLUZIONE CON IL METODO DEL LIBRO DI AGRESTI e FINLAY.

$\boxed{\beta = \dfrac{\sum[(x-\bar{x})\cdot(y-\bar{y})]}{\sum(x-\bar{x})^2}}$ $\boxed{\alpha = \bar{y} - b\cdot\bar{x}}$

X	Y
3	5
5	4
7	7
8	6
11	7
12	8

$\bar{x} = 7.67$
$\bar{y} = 6.17$
$n = 6$

1) $\beta = \dfrac{[(3-7.67)\cdot(5-6.17)] + [(5-7.67)\cdot(4-6.17)] + \dots}{(3-7.67)^2 + (5-7.67)^2 + \dots} =$

$= \dfrac{21.33}{59.33} = \underline{0.36}$

$\alpha = 6.17 - 0.36 \cdot 7.67 = \underline{3.41}$ $\Rightarrow \underline{\hat{y} = 3.41 + 0.36x}$

$\rightarrow$ Le varie interpretazioni non cambiano dal metodo con le matrici.

2) $\boxed{r = \left(\dfrac{s_x}{s_y}\right)\cdot\beta}$ dove $\boxed{s_x = \sqrt{\dfrac{\sum(x-\bar{x})^2}{n-1}}}$ e $\boxed{s_y = \sqrt{\dfrac{\sum(y-\bar{y})^2}{n-1}}}$

$s_x = \sqrt{\dfrac{59.33}{5}} = \underline{3.45}$ $s_y = \sqrt{\dfrac{(5-6.17)^2 + (4-6.17)^2 + \dots}{5}} = \sqrt{\dfrac{10.83}{5}} = \underline{1.47}$

$r = \dfrac{3.45}{1.47}\cdot 0.36 = \underline{0.84}$ $\Longrightarrow \boxed{R^2 = r^2} = 0.71$

3) $\boxed{t = \dfrac{\beta - 0}{se}}$ dove $\boxed{se = \dfrac{s}{\sqrt{\sum(x-\bar{x})^2}}}$ dove $\boxed{s = \sqrt{\dfrac{SSE}{n-2}}}$ dove $\boxed{SSE = \sum(y-\hat{y})^2}$

$\Rightarrow SSE = (5-4.49)^2 + (4-5.21)^2 + \dots = \underline{3.16}$ $\Rightarrow s = \sqrt{\dfrac{3.16}{4}} = \underline{0.89}$ $\Rightarrow se = \dfrac{0.89}{\sqrt{59.33}} = \underline{0.12}$

$t = \dfrac{0.36}{0.12} = \underline{3.11}$