Homework Assignment 04

Numerical Statistics Fall, 2022

Jake Underland - 1A193008

2022-01-24

1.

Consider the set of data $\{(x_i, y_i)|i = 1, 2, \ldots, 6\} = \{(-2, -3), (-1, 0), (0, 1), (2, 3), (3, 4), (4, 7)\}$.

(1-1) Find the Pearson correlation coefficient r_{xy} based on the data.
Solution.

$$\bar{x} = \frac{(-2 + -1 + 0 + 2 + 3 + 4)}{6} = 1$$
$$\bar{y} = \frac{(-3 + 0 + 1 + 3 + 4 + 7)}{6} = 2$$

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
-2	-3	-3	-5	15	9	25
-1	0	-2	-2	4	4	4
0	1	-1	-1	1	1	1
2	3	1	1	1	1	1
3	4	2	2	4	4	4
4	7	3	5	15	9	25
Total		0	0	40	28	60

$$\Longrightarrow r_{xy} = \frac{40}{\sqrt{28 \times 60}} = \frac{2\sqrt{105}}{21} \ldots \square$$

(1-2) Find the linear regression line, $y = \alpha + \beta x$, of y on x based on the data, and then express α and β as irreducible fractions.
Solution.
From the table above, $\bar{x} = 1,\ \bar{y} = 2,\ \sum_{i=1}^{6}(x_i - \bar{x})(y_i - \bar{y}) = 40,\ \sum_{i=1}^{6}(x_i - \bar{x})^2 = 28.$ Thus,

$$\beta = \frac{40}{28} = \frac{10}{7}$$
$$\alpha = 2 - \frac{10}{7} \times 1 = \frac{4}{7} \ldots \square$$

(1-3) Find the coefficient of determination R^2 and then express it as an irreducible fraction.
Solution.

$$R^2 = r_{xy}^2 = \left(\frac{2\sqrt{105}}{21}\right)^2 = \frac{20}{21} \ldots \square$$

(1-4) Find the explained variation $\sum_{i=1}^{6}(y_{est,i} - \bar{y})^2$ and then express it as an irreducible fraction.
Solution.
Since $r_{xy}^2 = \frac{\sum_{i=1}^{6}(y_{est,i}-\bar{y})^2}{\sum_{i=1}^{6}(y_i-\bar{y})^2}$, we have

$$\sum_{i=1}^{6}(y_{est,i} - \bar{y})^2 = \frac{20}{21} \times 60 = \frac{400}{7} \ldots \square$$

(1-5) Find the unexplained variation $\sum_{i=1}^{6}(y_i - y_{est,i})^2$ and then express it as an irreducible fraction.
Solution.
Since $\sum_{i=1}^{6}\left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{6}\left(y_{est,i} - \bar{y}\right)^2 + \sum_{i=1}^{6}\left(y_i - y_{est,i}\right)^2$, we have

$$\begin{aligned}\sum_{i=1}^{6}\left(y_i - y_{est,i}\right)^2 &= \sum_{i=1}^{6}\left(y_i - \bar{y}\right)^2 - \sum_{i=1}^{6}\left(y_{est,i} - \bar{y}\right)^2 \\ &= 60 - \frac{400}{7} = \frac{20}{7} \ldots \square\end{aligned}$$