TOPIC
Engineering Probability and Statistics – Section II – Question 3

QUESTION
Four data points have been observed as follows:

i	x_i	y_i
1	2.0	5.1
2	1.5	4.2
3	3.6	7.5
4	5.7	10.4

Using linear least-square regression, the equation that best fits this data is

(A) $y = 2.3 + 1.5x$
(B) $y = 2.3 + 2.1x$
(C) $y = 1.5 + 2.1x$
(D) $y = 1.5 + 1.5x$

HINT
For linear regression equation $y = a + bx + \varepsilon$, the parameters a and b can be estimated by the least-square method as

$$\hat{b} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad \text{and} \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

SOLUTION
The data averages are

$$\bar{x} = \frac{2+1.5+3.6+5.7}{4}$$
$$= 3.2$$
$$\bar{y} = \frac{5.1 + 4.2 + 7.5 + 10.4}{4}$$
$$= 6.8$$

The numerator

$$\sum_{i=1}^{4}(x_i - \bar{x})(y_i - \bar{y}) = 15.74$$

and the denominator

$$\sum_{i=1}^{4}(x_i - \bar{x})^2 = 10.74.$$

Therefore,

$$\hat{b} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$
$$= \frac{15.74}{10.74}$$
$$= 1.47$$

and

$$\hat{a} = \bar{y} - \hat{b}\bar{x}$$
$$= 6.8 - 1.47 \times 3.2$$
$$= 2.1$$

Therefore, the equation that best fits this data is

$y = 1.5 + 2.1x$

ANSWER
(C)

CONTRIBUTOR
Michael Weng