

Linear Regression Problem

Linear Regression equation $y = a + bx$

$$\text{where } a \text{ (intercept)} = \frac{\sum y \sum x^2 - \sum x \sum xy}{(\sum x^2) - (\sum x)^2}$$

$$b \text{ (slope)} = \frac{n \sum xy - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Q: Find the linear regression equation for following data:

x	2	4	6	8
y	3	7	5	10

Solution:

Construct this table

x	y	x^2	xy
2	3	4	6
4	7	16	28
6	5	36	30
8	10	64	80
$\sum x$ = 20	$\sum y$ = 25	$\sum x^2$ = 120	$\sum xy$ = 144

$$b = \frac{n \sum xy - (\sum x)(\sum y)}{n \sum x^2 - (\sum x)^2} = \frac{4 \times 144 - 20 \times 25}{4 \times 120 - 400} = 0.95$$

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n(\sum x^2) - (\sum x)^2} = \frac{25 \times 120 - 20 \times 144}{4 \times 120 - 400} = 1.5$$

$$\text{So } y = 1.5 + 0.95x \quad (y = a + bx)$$

If you are asked to find y for $x = 10$ then

$$y = 1.5 + 0.95 \times 10 = 1.5 + 9.5 = 11$$