

# Linear Regression

Course: Data Mining

Professor: Dr. Tahaei

Author: Sina Asghari

Subject: Linear Regression Problem and Solution

December 2024

## Question

A company wants to predict its sales (in thousands of dollars) based on the amount of advertising spent (in thousands of dollars). The following data is collected:

Advertising (\$k)	Sales (\$k)
2	3
3	6
5	8
7	11
8	14

**Task:** Using linear regression, calculate the regression equation  $y = mx + c$ , where  $m$  is the slope and  $c$  is the intercept. Predict the sales when advertising is \$10k.

## Solution

### Step 1: Find the Mean of $x$ and $y$

$$\bar{x} = \frac{\text{Sum of Advertising}}{\text{Number of Points}} = \frac{2 + 3 + 5 + 7 + 8}{5} = 5$$

$$\bar{y} = \frac{\text{Sum of Sales}}{\text{Number of Points}} = \frac{3 + 6 + 8 + 11 + 14}{5} = 8.4$$

### Step 2: Calculate the Slope ( $m$ )

The formula for the slope is:

$$m = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

- Compute  $(x_i - \bar{x})$  and  $(y_i - \bar{y})$ :

$$x_i - \bar{x} = \{-3, -2, 0, 2, 3\}$$

$$y_i - \bar{y} = \{-5.4, -2.4, -0.4, 2.6, 5.6\}$$

- Compute  $(x_i - \bar{x})(y_i - \bar{y})$ :

$$\{-3 \times -5.4, -2 \times -2.4, 0 \times -0.4, 2 \times 2.6, 3 \times 5.6\} = \{16.2, 4.8, 0, 5.2, 16.8\}$$

$$\text{Sum} = 43$$

- Compute  $(x_i - \bar{x})^2$ :

$$\{-3^2, -2^2, 0^2, 2^2, 3^2\} = \{9, 4, 0, 4, 9\}$$

$$\text{Sum} = 26$$

- Calculate  $m$ :

$$m = \frac{43}{26} \approx 1.65$$

### Step 3: Calculate the Intercept ( $c$ )

The formula for the intercept is:

$$c = \bar{y} - m\bar{x}$$

$$c = 8.4 - 1.65 \times 5 = 8.4 - 8.25 = 0.15$$

### Step 4: Form the Regression Equation

The regression equation is:

$$y = 1.65x + 0.15$$

**Step 5: Predict Sales When  $x = 10$** 

$$y = 1.65(10) + 0.15 = 16.5 + 0.15 = 16.65$$

**Final Answer:** The regression equation is:

$$y = 1.65x + 0.15$$

The predicted sales when advertising is \$10k is:

\$16.65k