The Method of Least Squares

Lecture 50
Section 7.4

Robb T. Koether

Hampden-Sydney College

Fri, Apr 28, 2017

Objectives

Objectives

- Derive the equation for the least-squares regression line.
- Apply the equation to examples.
- Derive other least-squares equations.

The Regression Line

The Regression Line

- The data points are

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n).$$

- Given a line $y = a + bx$, the point (x_i, y_i) **deviates** from the line by the amount

$$(a + bx_i) - y_i.$$

- The **regression line** is the (unique) line that minimizes the sum of the squared deviations

$$\sum_{i=1}^{n} ((a + bx_i) - y_i)^2.$$

Finding the Regression Line

Finding the Regression Line

- The sum of squared deviations is a function of the parameters a and b.
- Thus, we want to find values of a and b that minimize the function

$$f(a,b) = \sum_{i=1}^{n} ((a + bx_i) - y_i)^2.$$

Finding the Regression Line

Finding the Regression Line

- Find the first partials.

$$\begin{aligned}\frac{\partial f}{\partial a} &= \sum 2((a+bx_i)-y_i) \\ &= 2\sum (a+bx_i-y_i), \\ \frac{\partial f}{\partial b} &= \sum 2((a+bx_i)-y_i)(x_i) \\ &= 2\sum (ax_i+bx_i^2-x_iy_i).\end{aligned}$$

- Solve the system of equations

$$\begin{aligned}2\sum (a+bx_i-y_i) &= 0, \\ 2\sum (ax_i+bx_i^2-x_iy_i) &= 0.\end{aligned}$$

Finding the Regression Line

Finding the Regression Line

- We can rewrite these as

$$na + b\sum x_i = \sum y_i,$$
$$a\sum x_i + b\sum x_i^2 = \sum x_i y_i.$$

Finding the Regression Line

Finding the Regression Line

- Now solve for a and b.

$$na + b\sum x_i = \sum y_i,$$
$$a\sum x_i + b\sum x_i^2 = \sum x_i y_i.$$

- The solution is

$$a = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2},$$
$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}.$$

Finding the Regression Line

Finding the Regression Line

We usually write these equations as

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

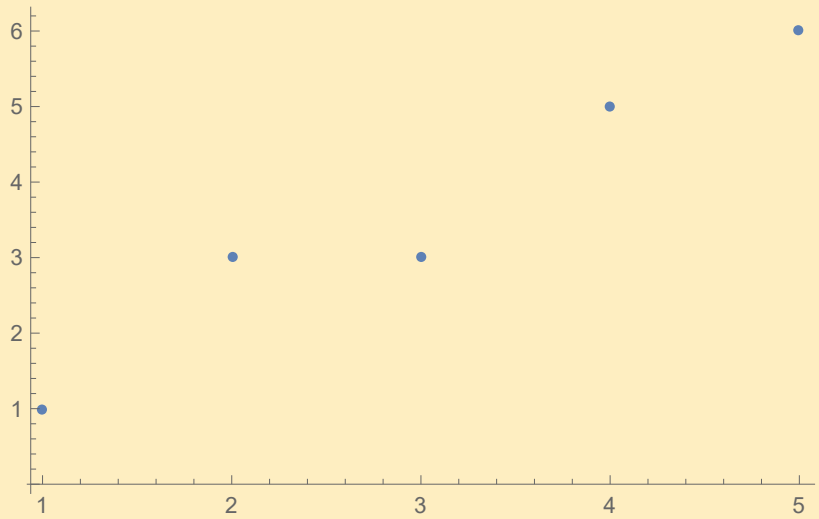
$$a = \frac{\sum y - b\sum x}{n}.$$

Linear Regression

Example

Find the least-squares regression line for the data

$$(1, 1), (2, 3), (3, 3), (4, 5), (5, 6).$$

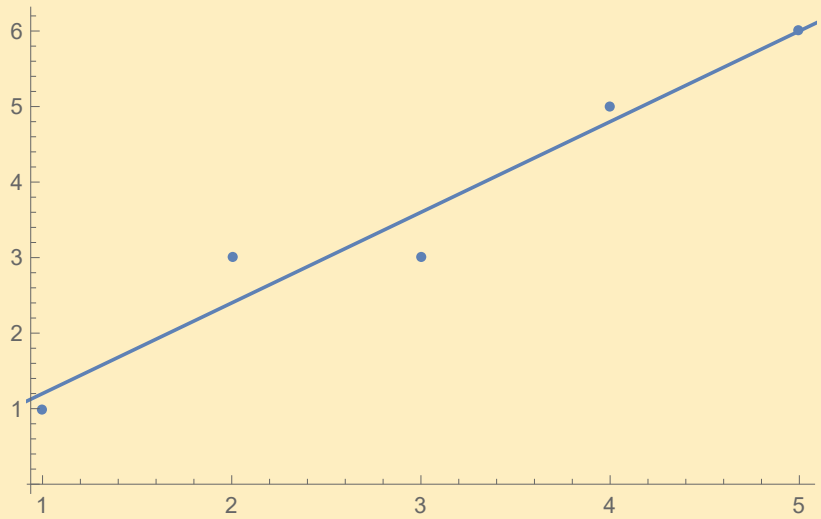


Linear Regression

Example

Find the least-squares regression line for the data

$$(1,1),(2,3),(3,3),(4,5),(5,6).$$



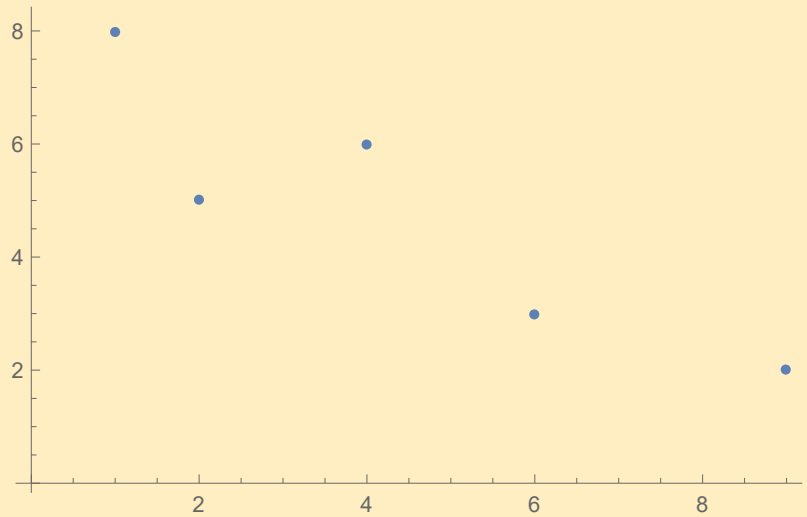
$$y = 0.0 + 1.2x$$

Linear Regression

Example

Find the least-squares regression line for the data

$$(1, 8), (2, 5), (4, 6), (6, 3), (9, 2).$$

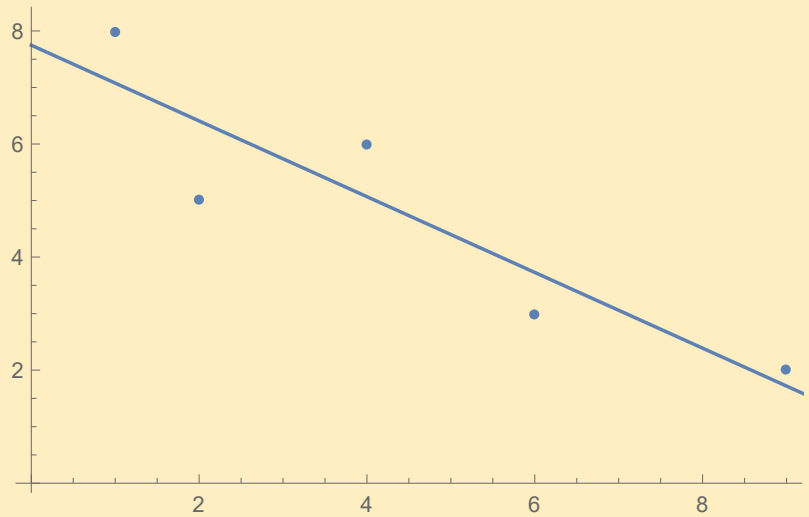


Linear Regression

Example

Find the least-squares regression line for the data

$$(1, 8), (2, 5), (4, 6), (6, 3), (9, 2).$$



$$y = 7.7476 - 0.6699x$$