Regression

Regression is a supervised learning technique used to model the relationship between input(s) and continuous output.

Regression is a Machine Learning Algorithm, but more precisely, it's a type of supervised learning algorithm used for predicting continuous values. However, it’s also fair to say that “regression” is a technique — because it's not one single algorithm, but rather a family of algorithms (like linear regression, logistic regression, ridge regression, etc.).

Types of Regression in Machine Learning:

1. Linear Regression

What it does: Predicts a continuous output (like house price, temperature, weight).

Assumption: The relationship between input (x) and output (y) is linear.

Output: A real number.

Example: Predicting the price of a car based on mileage.

Equation: $y=mx+b$

2. Logistic Regression

What it does: Used for classification, not regression!

Name is misleading, but it’s called "regression" because it models the probability using a logistic (sigmoid) function.

Output: A probability (between 0 and 1), which is then mapped to a class label (e.g., 0 or 1).

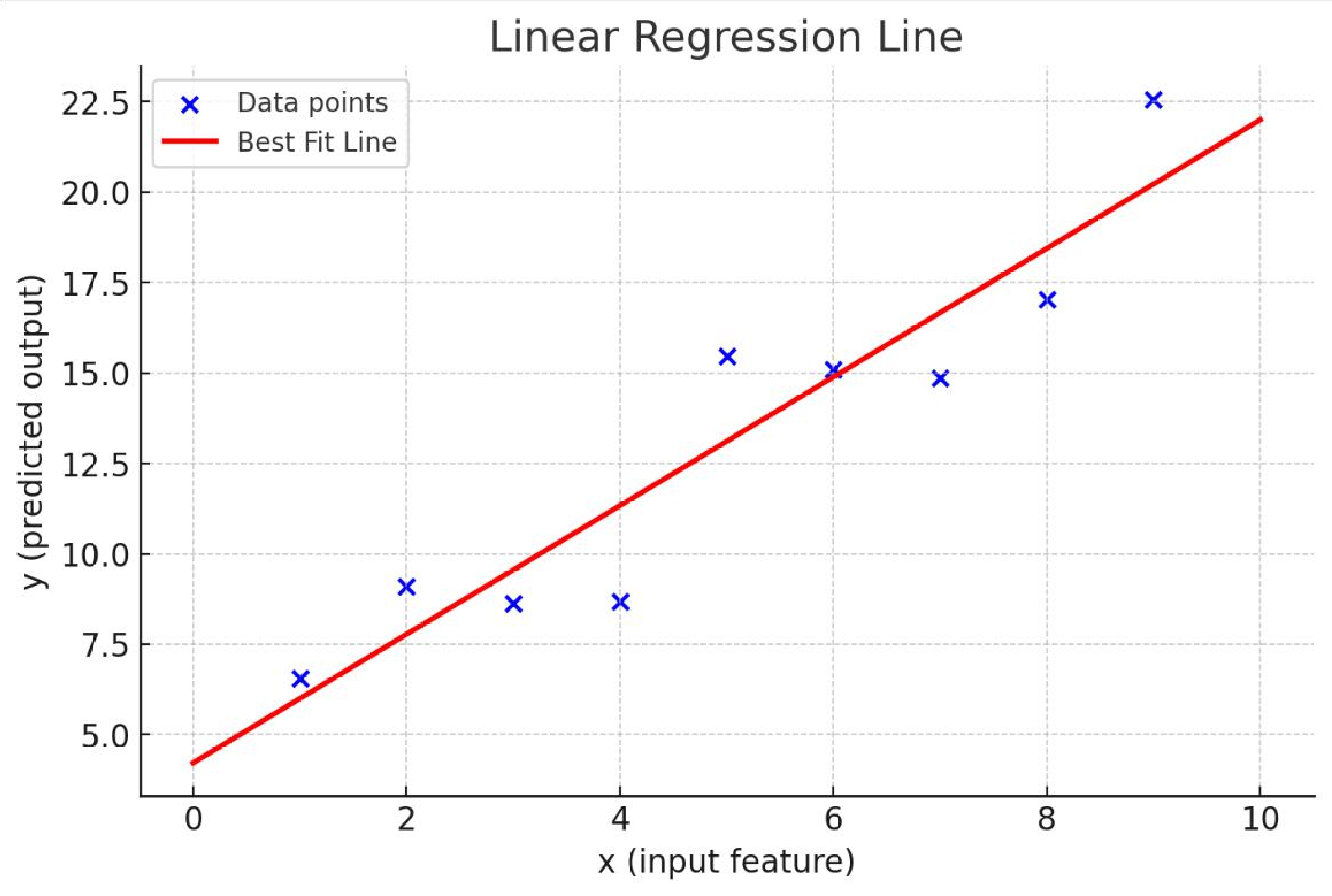
Example: Predicting if a student will pass (1) or fail (0) an exam based on study hours.

Equation: $P(y = 1|x) = \frac{1}{1 + e^{-(mx+b)}}$

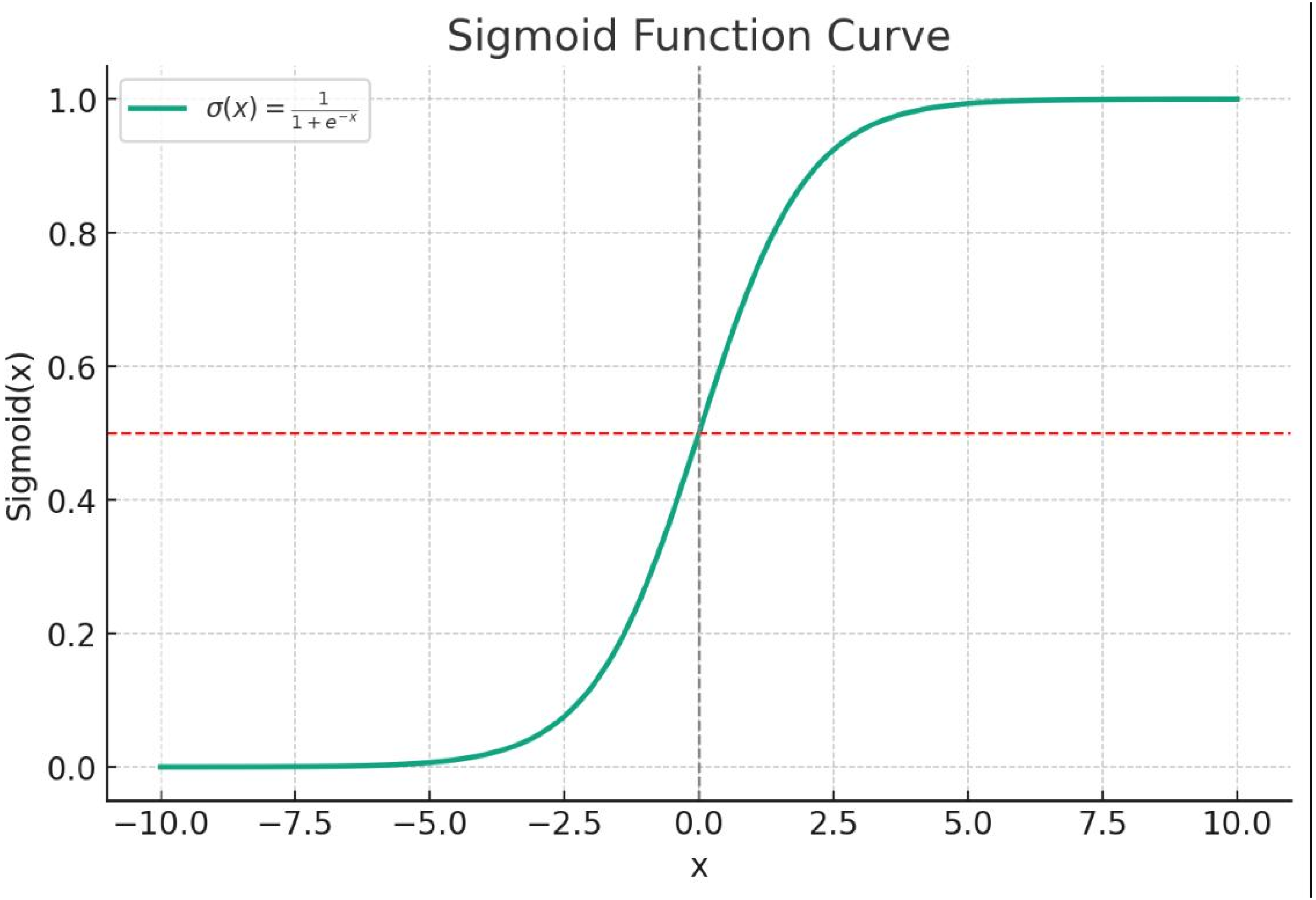
Main Difference Between Linear and Logistic Regression:

Feature	Linear Regression	Logistic Regression
Output type	Continuous number	Probability (0 to 1)
Use case	Prediction (e.g., salary)	Classification (e.g., pass/fail)
Output function	Straight line	Sigmoid curve
Algorithm type	Regression	Classification

Linear Regression Line



Sigmoid Fuction Curve



Linear Regression
(Least Square Method)

Line of Best Fit (Least Square Method)

- A **line of best fit** is a straight line that is the best approximation of the given set of data.
- It is used to study the nature of the relation between two variables. (We're only considering the two-dimensional case, here.)
- A line of best fit can be roughly determined using an eyeball method by drawing a straight line on a scatter plot so that the number of points above the line and below the line is about equal (and the line passes through as many points as possible).

- A more accurate way of finding the line of best fit is the least square method .
- Use the following steps to find the equation of line of best fit for a set of ordered pairs $(x_1,y_1),(x_2,y_2),\ldots(x_n, y_n)$
- Step 1: Calculate the mean of the x values and the mean of the y values.

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Step 2: The following formula gives the slope of the line of best fit:

$$m = \frac{\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{X})^2}$$

Step 3: Compute the y -intercept of the line by using the formula:

$$b = \overline{Y} - m\overline{X}$$

Step 4: Use the slope m and the y -intercept b to form the equation of the line.

Example:

Use the least square method to determine the equation of line of best fit for the data. Then plot the line.

x	8	2	11	6	5	4	12	9	6	1
y	3	10	3	6	8	12	1	4	9	14

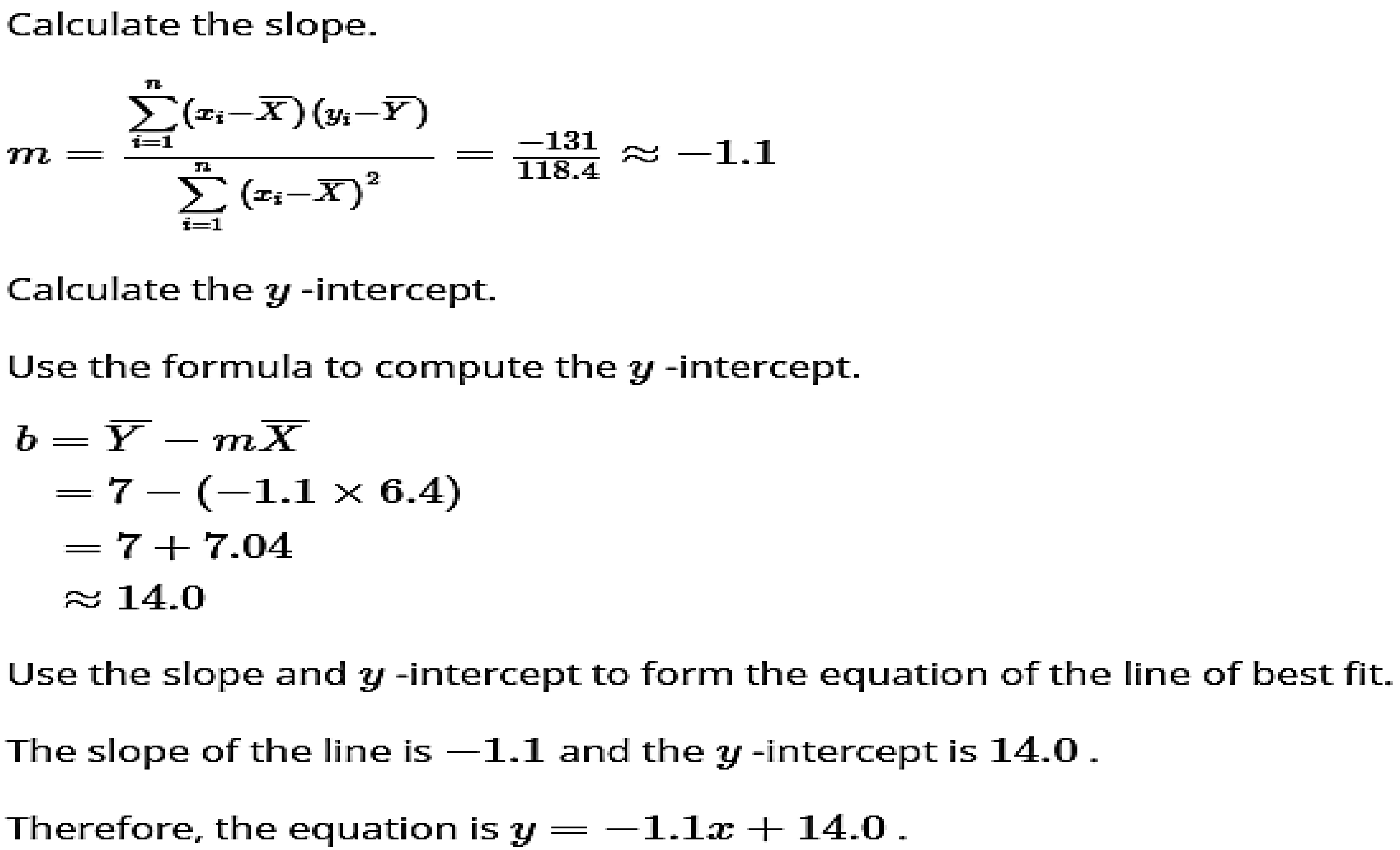
Calculate the means of the x -values and the y -values.

$$\overline{X} = \frac{8+2+11+6+5+4+12+9+6+1}{10} = 6.4$$

$$\overline{Y} = \frac{3+10+3+6+8+12+1+4+9+14}{10} = 7$$

Now calculate $x_i - \overline{X}$, $y_i - \overline{Y}$, $(x_i - \overline{X})(y_i - \overline{Y})$, and $(x_i - \overline{X})^2$ for each i .

i	x_i	y_i	$x_i - \overline{X}$	$y_i - \overline{Y}$	$(x_i - \overline{X})(y_i - \overline{Y})$	$(x_i - \overline{X})^2$
1	8	3	1.6	−4	−6.4	2.56
2	2	10	−4.4	3	−13.2	19.36
3	11	3	4.6	−4	−18.4	21.16
4	6	6	−0.4	−1	0.4	0.16
5	5	8	−1.4	1	−1.4	1.96
6	4	12	−2.4	5	−12	5.76
7	12	1	5.6	−6	−33.6	31.36
8	9	4	2.6	−3	−7.8	6.76
9	6	9	−0.4	2	−0.8	0.16
10	1	14	−5.4	7	−37.8	29.16
					$\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y}) = -131$	$\sum_{i=1}^{n}(x_i - \overline{X})^2 = 118.4$



Calculate the slope.

$$m = \frac{\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{X})^2} = \frac{-131}{118.4} \approx -1.1$$

Calculate the y -intercept.

Use the formula to compute the y -intercept.

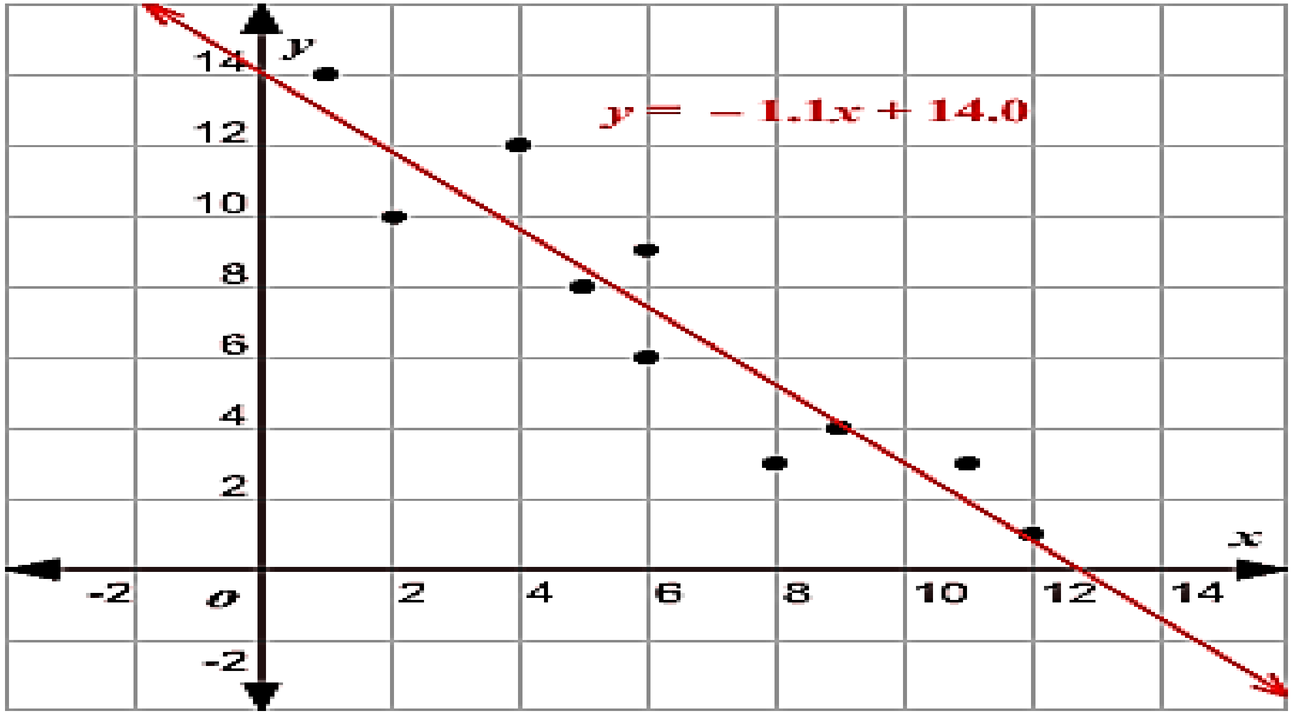
$$\begin{aligned} b &= \overline{Y} - m\overline{X} \\ &= 7 - (-1.1 \times 6.4) \\ &= 7 + 7.04 \\ &\approx 14.0 \end{aligned}$$

Use the slope and y -intercept to form the equation of the line of best fit.

The slope of the line is -1.1 and the y -intercept is 14.0 .

Therefore, the equation is $y = -1.1x + 14.0$.

Draw the line on the scatter plot.



Statistics 252 (March 11, 2005)

Linear Regression and Least Squares

Consider the linear regression model $Y = \beta_0 + \beta_1 x + \varepsilon$ where ε is a mean zero random variable. Our goal is to predict the linear trend

$$\mathbb{E}(Y) = \beta_0 + \beta_1 x$$

by estimating the intercept and the slope of this line. That is, we seek estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ such that

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Our goal is to choose $\hat{\beta}_0$ and $\hat{\beta}_1$ in such a way that they minimize SSE, the sum of the squares of the errors. That is, we want to minimize

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where y_i is the ith observation of the random variable Y, and corresponds to the input x_i. Substituting the value of $\hat{y}$ into the equation for SSE we see that SSE can be viewed as a function of $\hat{\beta}_0$ and $\hat{\beta}_1$. We can now use the methods of elementary calculus to minimize this function. Namely, we find the first derivative, set it equal to 0, and solve for the critical points. We then use the second derivative test to check that the critical points are indeed minimizers. Thus,

$$\mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

Next, we find

$$\begin{aligned} \frac{\partial}{\partial \hat{\beta}_0} \mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= -2 \sum_{i=1}^{n} y_i + 2n\hat{\beta}_0 + 2\hat{\beta}_1 \sum_{i=1}^{n} x_i \\ &= -2n\overline{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1 \overline{x} \end{aligned}$$

and

$$\frac{\partial}{\partial \hat{\beta}_1} \mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = -2 \sum_{i=1}^{n} x_i y_i + 2\hat{\beta}_0 \sum_{i=1}^{n} x_i + 2\hat{\beta}_1 \sum_{i=1}^{n} x_i^2.$$

From the first equation,

$$\frac{\partial}{\partial \hat{\beta}_0} \mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = 0$$

implies

$$-2n\overline{y} + 2n\hat{\beta}_0 + 2n\hat{\beta}_1 \overline{x} = 0$$

so that

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}.$$

From the second equation,

$$\frac{\partial}{\partial \hat{\beta}_1} \mathrm{SSE}(\hat{\beta}_0, \hat{\beta}_1) = 0$$

implies

$$-2\sum_{i=1}^{n} x_i y_i + 2\hat{\beta}_0 \sum_{i=1}^{n} x_i + 2\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

so that

$$-\sum_{i=1}^{n} x_i y_i + (\overline{y} - \hat{\beta}_1 \overline{x}) \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0.$$

Distributing, and collecting $\hat{\beta}_1$ gives

$$-\sum_{i=1}^{n} x_i y_i + \overline{y} \sum_{i=1}^{n} x_i - \hat{\beta}_1 \overline{x} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

so that

$$\hat{\beta}_1 \left(\overline{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2 \right) = \overline{y} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i.$$

Thus, we find

$$\hat{\beta}_1 = \frac{\overline{y} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i y_i}{\overline{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2}.$$

Now, we can write this in a nicer way:

$$\hat{\beta}_1 = \frac{n\overline{y}\,\overline{x} - \sum_{i=1}^{n} x_i y_i}{n\overline{x}^2 - \sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{y}\,\overline{x}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}.$$

With a bit of algebra, we can write the numerator as

$$\sum_{i=1}^{n} x_i y_i - n\overline{y}\,\overline{x} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) =: S_{XY}$$

and the denominator as

$$\sum_{i=1}^{n} x_i^2 - n\overline{x}^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2 =: S_{XX}.$$

Thus, we can write $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}.$$

This version might be useful for remembering the formula, and it often appears when you are doing numerical calculations with computer software (for example, with SAS). The software tends to return the values of S_{XY} and S_{XX} because they are useful for *residual analysis* in general (this is not Stat 252 stuff). One final note: the formulæ for $\hat{\beta}_0$ and $\hat{\beta}_1$ agree with those in the Stat 151 textbook (pages 529–532 in the first edition and pages 491–494 in the second edition).