Simple Linear Regression

Prof. Lauren Perry

Simple Linear Regression

Model:

$$Y \approx \beta_0 + \beta_1 X$$

where X consists of a single predictor variable.

- The *intercept*, β_0, and the *slope*, β_1, make up the models *parameters* or *coefficients*.

When we use the estimated model to make predictions, we write

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

- Conceptually, this is a 2D extension of using a sample mean $\bar{x}$ to estimate a population mean μ.

Estimating the Coefficients

- We can think of our data as n points of the form (x_i, y_i).
- Our goal is to estimate β_0 and β_1 so that the model fits the data well.
  - That is, so that
    $$y_i \approx \hat{\beta}_0 + \hat{\beta}_1 x_i$$
    for each $i \in \{1, \ldots, n\}$.
  - Idea: the line is as close as possible to all n data points.

Error (Residuals)

$$e_i = y_i - f(x_i)$$

where e_i is the ith *residual*.

Our goal is to *minimize* overall error.

Why can't we jump right in with minimizing this quantity?

Goal: Minimize Error

One possibility: absolute error $|y_i - f(x_i)|$

Another possibility: squared error $(y_i - f(x_i))^2$

Squared error is used far more often than absolute error.

Why do you think that is?

Least Squares

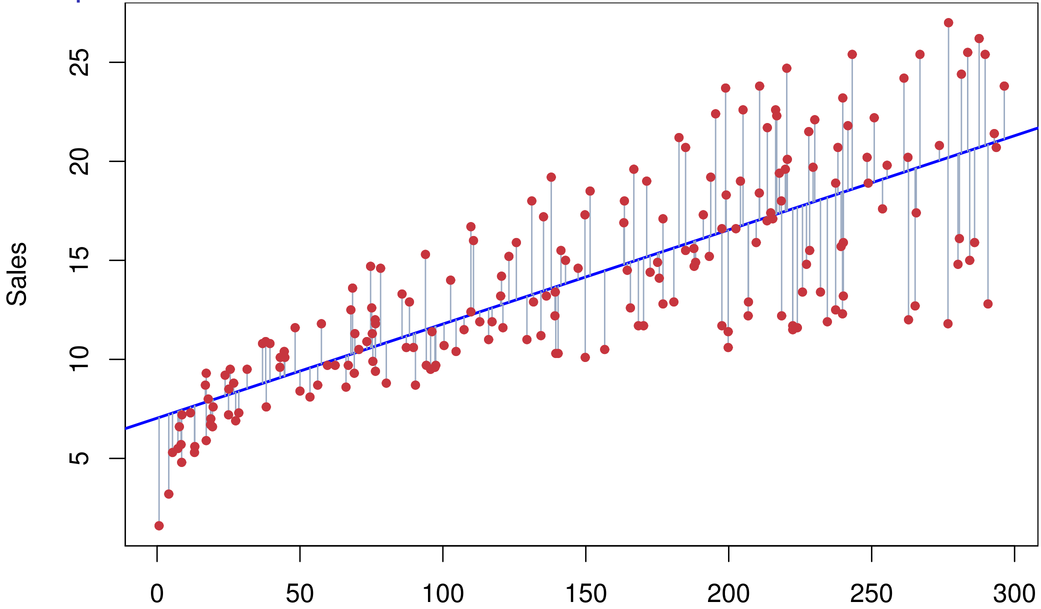
The *least squares criterion* focuses on "closeness" as a measure of how close each response value y is to the predicted value $\hat{y}$:

$$e_i = y_i - \hat{y}_i$$

Then the *residual sum of squares* is

$$\text{RSS} = e_1^2 + e_2^2 + \cdots + e_n^2$$

Least Squares



Least Squares

The least squares approach chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize the RSS.

$$\begin{aligned}\text{RSS} &= \sum_{i=1}^{n} e_i^2 \\ &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2\end{aligned}$$

which we minimize by taking the derivatives

$$\frac{\delta \text{RSS}}{\delta \hat{\beta}_0} \quad \text{and} \quad \frac{\delta \text{RSS}}{\delta \hat{\beta}_1}$$

Least Squares

Note: least squares is a convex optimization problem.

- That is, every local minimum is a global minimum.
  - (We don't need to do any kind of second derivative check.)

Least Squares

This minimization problem yields

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\ &= \frac{s_{xy}}{s_x^2} \\ &= r_{xy}\frac{s_y}{s_x}\end{aligned}$$

Notice how the sample correlation, covariance, variances, and coefficients are all related.

Here, $\hat{\beta}_0 = 7.03$ and $\hat{\beta}_1 = 0.0475$.