md example

Michael Arango

11/6/2018

Problem 3

A simple linear regression model is given by $Y = \beta_0 + \beta_1 X + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$.

a. Use the method of least squares, as discussed in class to derive the least squares estimators for β_0 and β_1. The multiple regression model was derived in class.

Solution

Suppose we have $i = 1, ..., n$ data points where $x_1, ..., x_n$ is the independent variable and $y_1, ..., y_n$ is the dependent variable. We want $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$, but we obtain $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \epsilon_i$. We need to try to minimize the sum of the squared residuals defined as follows:

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

We want to minimize this function L with respect to β_0 and β_1. So, our problem becomes

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2}$$

which simplifies to

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{3}$$

We get our key equations by taking the partial derivative with respect to each of the parameters we want to minimize and setting them equal to zero:

$$\frac{\partial L}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} -2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{4}$$

$$\frac{\partial L}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} -2x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{5}$$

Solving (4), we find that:

$$\sum_{i=1}^{n} -2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{6}$$

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{7}$$

$$\text{Recall that } \sum_{i=1}^{n} y_i = n\bar{y} \text{ since } \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y} \tag{8}$$

$$n\bar{y} - n\hat{\beta}_0 - n\hat{\beta}_1 \bar{x} = 0 \tag{9}$$

$$n\hat{\beta}_0 = n\bar{y} - n\hat{\beta}_1 \bar{x} \tag{10}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{11}$$

Solving (5), we see that:

$$\sum_{i=1}^{n} -2x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{12}$$

$$\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{13}$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \hat{\beta}_0 x_i - \sum_{i=1}^{n} \hat{\beta}_1 x_i^2 = 0 \tag{14}$$

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} (\bar{y} - \hat{\beta}_1 \bar{x}) x_i - \sum_{i=1}^{n} \hat{\beta}_1 x_i^2 = 0 \tag{15}$$

$$\sum_{i=1}^{n} x_i y_i - \bar{y} \sum_{i=1}^{n} x_i + \hat{\beta}_1 \bar{x} \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \hat{\beta}_1 x_i^2 = 0 \tag{16}$$

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} + n\hat{\beta}_1 \bar{x}^2 - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0 \tag{17}$$

$$\sum_{i=1}^{n} \hat{\beta}_1 x_i^2 - n\hat{\beta}_1 \bar{x}^2 = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \tag{18}$$

$$\hat{\beta}_1 \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \tag{19}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{20}$$

Using the fact that

$$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

and

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

we can rewrite this in the difference of means form:

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

b. Now, forcing the intercept coefficient to be zero, re-derive the least squares estimator for the slope. Under what circumstances do you think it would be useful?

Solution

By forcing the intercept to zero, we want to model $y_i = \beta x_i + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$, but we obtain $\hat{y}_i = \hat{\beta} x_i + \epsilon_i$. This time, we are minimizing the function

$$L = \sum_{i=1}^{n} (y_i - \hat{\beta} x_i)^2 \tag{21}$$

We want to minimize this function L with respect to β. So, our problem becomes

$$\min_{\hat{\beta}} \sum_{i=1}^{n} (y_i - \hat{\beta} x_i)^2 \tag{22}$$

Taking the derivative and setting the equation equal to zero, we have our first order condition:

$$\frac{dL}{d\hat{\beta}_0} = \sum_{i=1}^{n} -2x_i(y_i - \hat{\beta} x_i) = 0 \tag{23}$$

Solving (23), we find

$$\sum_{i=1}^{n} x_i(y_i - \hat{\beta} x_i) = 0 \tag{24}$$

$$\sum_{i=1}^{n} x_i y_i - \hat{\beta} \sum_{i=1}^{n} x_i^2 = 0 \tag{25}$$

$$\hat{\beta} \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \tag{26}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \tag{27}$$

$$\tag{28}$$

Most of the time it does not make sense to use a model without an intercept in practice, but it can be helpful if there are physical constraints on the variables. For example, if we want to use the speed a car is traveling to predict the stopping distance of the car, then it might make sense to force the intercept coefficient to zero.