Simple Linear Regression

Dr. Kiah Wah Ong

ILLINOIS TECH

Least Squares Method

In our simple linear regression model

$$y = \beta_0 + \beta_1 x + \epsilon$$

the parameters β_0 and β_1 are unknown and must be estimated using sample data

$$(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n).$$

The **least squares method** is used to estimate β_0 and β_1. To do this, from

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

we write

$$\epsilon_i = y_i - (\beta_0 + \beta_1 x_i)$$

and call the following as

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

the residual sum of squares.

Least Squares Method

We now estimate β_0 and β_1 by minimizing the residual sum of squares. Using calculus, we see that the **least squares estimators** of β_0 and β_1, written as $\hat{\beta}_0$ and $\hat{\beta}_1$, must satisfy

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$\left.\frac{\partial S}{\partial \beta_1}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) x_i = 0$$

A bit more detail

From

$$\left.\frac{\partial S}{\partial \beta_0}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

we obtain

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

hence

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

A bit more detail

From

$$\left.\frac{\partial S}{\partial \beta_1}\right|_{\hat{\beta}_0,\hat{\beta}_1} = -2\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)x_i = 0$$

we obtain

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

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

and

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

Least Squares Method

That is, we have the following

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

These are known as the *normal equations*. With

$$\overline{y} = \sum_{i=1}^{n} \frac{y_i}{n} \quad \text{and} \quad \overline{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

We have

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

A bit more detail

From

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

and

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

we have

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

Solve for $\hat{\beta}_1$, we get

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

Least Squares Method

From

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

because

$$S_{xx} = \sum_{i=1}^n (x_i - \overline{x})^2 = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}$$

and

$$S_{xy} = \sum_{i=1}^n y_i(x_i - \overline{x}) = \sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right)}{n}$$

we can write $\hat{\beta}_1$ in a more compact form as

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}.$$

Least Squares Method Recap

Now given a set of n data

$$(x_1, y_1), \cdots, (x_n, y_n)$$

we are able to compute the least square estimator of β_0 and β_1 as

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

and

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{S_{xy}}{S_{xx}}.$$