Simple Linear Regression

Dr. Kiah Wah Ong

ILLINOIS TECH

Previously on Regression

We assume that the x and the y are related by a straight line equation

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

with β_0 and β_1 unknown numbers.

Simple Linear Regression (SLR)

For data points (x_i, y_i), $i = 1, 2, \cdots, n$, we write

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

(i) y_i is the value of the response variable in the i-th trial

(ii) β_0 and β_1 are parameters

(iii) x_i is a **known constant**, namely, the value of the predictor variable in the i-th trial

(iv) ϵ_i is a random error with

$$E(\epsilon_i) = 0 \quad \text{and} \quad \sigma^2(\epsilon_i) = \sigma^2$$

and ϵ_i, ϵ_j are uncorrelated so that their covariance is zero, i.e. $\sigma(\epsilon_i, \epsilon_j) = 0$ for all i, j, $i \neq j$, $i = 1, 2, \cdots, n$.

Simple Linear Regression (SLR)

By minimizing

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

the residual sum of squares. we obtained

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{S_{xx}} = \sum_{i=1}^{n} c_i y_i$$

where

$$c_i = \frac{x_i - \overline{x}}{S_{xx}}, \quad i = 1, \cdots, n,$$

and

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

Properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

Now let us look at

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{S_{xx}} = \sum_{i=1}^{n} c_i y_i$$

where

$$c_i = \frac{x_i - \overline{x}}{S_{xx}}, \quad i = 1, \cdots, n, \quad \text{so}$$

$$\begin{aligned}
E(\hat{\beta}_1) &= E(\sum_{i=1}^{n} c_i y_i) = \sum_{i=1}^{n} c_i E(y_i) \\
&= \sum_{i=1}^{n} c_i E(\beta_0 + \beta_1 x_i + \epsilon_i) \\
&= \sum_{i=1}^{n} c_i \left(E(\beta_0 + \beta_1 x_i) + E(\epsilon_i)\right) \\
&= \sum_{i=1}^{n} c_i(\beta_0 + \beta_1 x_i) = \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i x_i
\end{aligned}$$

ILLINOIS TECH

Properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

Continue from the previously slide, we have

$$E(\hat{\beta}_1) = E(\sum_{i=1}^{n} c_i y_i) = \sum_{i=1}^{n} c_i(\beta_0 + \beta_1 x_i) = \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i x_i$$

We leave it as an exercise for you to show that

$$\sum_{i=1}^{n} c_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} c_i x_i = 1$$

This gives

$$E(\hat{\beta}_1) = \beta_1$$

We say that $\hat{\beta}_1$ is an **unbiased estimator** of β_1.

Properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

From

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

we have

$$\begin{aligned}
E(\hat{\beta}_0) &= E(\overline{y} - \hat{\beta}_1 \overline{x}) \\
&= E\left(\frac{\sum_{i=1}^n y_i}{n}\right) - \overline{x} E(\hat{\beta}_1)) \\
&= \frac{1}{n} \sum_{i=1}^n E(y_i) - \overline{x}\beta_1 \\
&= \frac{1}{n} \left(\sum_{i=1}^n (\beta_0 + \beta_1 x_i) \right) - \overline{x}\beta_1 \\
&= \frac{1}{n} (n\beta_0 + \beta_1 n \overline{x}) - \overline{x}\beta_1 \\
&= \beta_0 + \beta_1 \overline{x} - \overline{x}\beta_1 = \beta_0
\end{aligned}$$

We say that $\hat{\beta}_0$ is an **unbiased estimator** of β_0.

ILLINOIS TECH

Properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

Recall that

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{S_{xx}} = \sum_{i=1}^{n} c_i y_i$$

where

$$c_i = \frac{x_i - \overline{x}}{S_{xx}}, \quad i = 1, \cdots, n$$

hence

$$\mathrm{Var}(\hat{\beta}_1) = \mathrm{Var}\left(\sum_{i=1}^{n} c_i y_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \mathrm{Cov}(y_i, y_j)$$

Because the observation y_i is uncorrelated, namely $\mathrm{Cov}(y_i, y_j) = 0$ for $i \neq j$ the above equation gives

$$\mathrm{Var}(\hat{\beta}_1) = \mathrm{Var}\left(\sum_{i=1}^{n} c_i y_i\right) = \sum_{i=1}^{n} c_i^2 \mathrm{Var}(y_i)$$

Properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

With

$$c_i = \frac{x_i - \overline{x}}{S_{xx}} \quad \text{and} \quad S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

we have

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_1) &= \sum_{i=1}^{n} c_i^2 \mathrm{Var}(y_i) = \sigma^2 \sum_{i=1}^{n} c_i^2 \\
&= \sigma^2 \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{S_{xx}} \right)^2 \\
&= \frac{\sigma^2}{S_{xx}^2} \sum_{i=1}^{n} (x_i - \overline{x})^2 \\
&= \frac{\sigma^2}{S_{xx}^2} S_{xx} \\
&= \frac{\sigma^2}{S_{xx}}
\end{aligned}$$

Properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

Using

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

and

$$\mathrm{Var}(aX + bY) = a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y) + 2ab\mathrm{Cov}(X, Y)$$

the variance of $\hat{\beta}_0$ is computed as

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_0) &= \mathrm{Var}(\overline{y} - \hat{\beta}_1 \overline{x}) \\
&= \mathrm{Var}(\overline{y}) + \overline{x}^2 \mathrm{Var}(\hat{\beta}_1) - 2\overline{x}\mathrm{Cov}(\overline{y}, \hat{\beta}_1)
\end{aligned}$$

Properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

From the previous slide, we see that

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_0) &= \mathrm{Var}(\overline{y} - \hat{\beta}_1 \overline{x}) \\ &= \mathrm{Var}(\overline{y}) + \overline{x}^2 \mathrm{Var}(\hat{\beta}_1) - 2\overline{x}\mathrm{Cov}(\overline{y}, \hat{\beta}_1)\end{aligned}$$

Now

$$\begin{aligned}\mathrm{Var}(\overline{y}) &= \mathrm{Var}\left(\frac{\sum_{i=1}^n y_i}{n}\right) \\ &= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^n y_i\right) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(y_i) \\ &= \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}\end{aligned}$$

We will leave it as an exercise for you to show that

$$\mathrm{Cov}(\overline{y}, \hat{\beta}_1) = 0$$

Properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

Using

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}, \mathrm{Var}(\overline{y}) = \frac{\sigma^2}{n} \text{ and } \mathrm{Cov}(\overline{y}, \hat{\beta}_1) = 0$$

we conclude that

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_0) &= \mathrm{Var}(\overline{y} - \hat{\beta}_1 \overline{x}) \\
&= \mathrm{Var}(\overline{y}) + \overline{x}^2 \mathrm{Var}(\hat{\beta}_1) - 2\overline{x}\mathrm{Cov}(\overline{y}, \hat{\beta}_1) \\
&= \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}} \right)
\end{aligned}$$

Recap

The least squares estimators of β_0 and β_1, denoted as $\hat{\beta}_0$ and $\hat{\beta}_1$, have the following means and variances:

$$E(\hat{\beta}_0) = \beta_0 \quad \text{and} \quad E(\hat{\beta}_1) = \beta_1$$

That is, $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimator of β_0 and β_1, respectively.

As for the variance, we have

$$\mathrm{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{S_{xx}} \right)$$

and

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

Properties of the least squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$

The **Gauss-Markov Theorem** states that for the regression model

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

with the assumptions

$$E(\epsilon) = 0, \mathrm{Var}(\epsilon) = \sigma^2$$

and uncorrelated errors, the least squares estimators are unbiased and have minimum variance when compared with all other unbiased estimators that are linear combinations of the y_i.

We say that the least squares estimators are **best linear unbiased estimators**, where “best” implies minimum variance.