Some Theory Underlying Simple Linear Regression

2018-09-02

To show that the slope estimator is unbiased, we need to show that $\mathbb{E}(B_1) = \beta$. We can express the slope estimator B_1 as a linear function of the observations, namely,

$$B_1 = \sum m_i Y_i \qquad \text{where } m_i = \frac{x_i - \bar{x}}{\sum (x_i - \bar{x})^2}$$

We also make use of the fact that,

$$\begin{aligned}
\sum (x_i - \bar{x})^2 &= \sum (x_i - \bar{x})(x_i - \bar{x}) \\
&= \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \\
&= \sum x_i^2 - 2\bar{x}(n\bar{x}) + n\bar{x}^2 \\
&= \sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 \\
&= \sum x_i^2 - n\bar{x}^2
\end{aligned}$$

and the assumption of linearity in the population, namely that,

$$\mathbb{E}(Y_i) = \beta_0 + \beta_1(x_i)$$

Since $B_1 = \sum \frac{(x_i - \bar{x})Y_i}{\sum x_i^2 - n\bar{x}^2}$, then

Proof.

$$
\begin{aligned}
\mathbb{E}(B_1) &= \mathbb{E}\Bigg(\sum \frac{(x_i - \bar{x})Y_i}{\sum x_i^2 - n\bar{x}^2} \Bigg) \\
&= \mathbb{E}\Bigg(\frac{1}{\sum x_i^2 - n\bar{x}^2} \times \sum \Big[(x_i - \bar{x})Y_i \Big] \Bigg) \\
&= \frac{1}{\sum x_i^2 - n\bar{x}^2} \times \mathbb{E}\Bigg(\sum \Big[(x_i - \bar{x})Y_i \Big] \Bigg) \\
&= \frac{1}{\sum x_i^2 - n\bar{x}^2} \times \sum \Bigg(\mathbb{E}\Big[(x_i - \bar{x})Y_i \Big] \Bigg) && \text{Expected value of a sum is the sum of an expected value.} \\
&= \frac{1}{\sum x_i^2 - n\bar{x}^2} \times \sum \Bigg((x_i - \bar{x})\mathbb{E}\big[Y_i\big] \Bigg) \\
&= \frac{1}{\sum x_i^2 - n\bar{x}^2} \times \sum \Bigg((x_i - \bar{x})\big[\beta_0 + \beta_1(x_i)\big] \Bigg) && \text{Use definition of linearity assumption.} \\
&= \frac{1}{\sum x_i^2 - n\bar{x}^2} \times \sum \Bigg(x_i\beta_0 - \bar{x}\beta_0 + \beta_1 x_i^2 - \beta_1 x_i \bar{x} \Bigg) \\
&= \frac{1}{\sum x_i^2 - n\bar{x}^2} \times \beta_0 \sum x_i - n\bar{x}\beta_0 + \beta_1 \sum x_i^2 - \beta_1 \bar{x} \sum x_i \Bigg) && \text{Distribute sum and pull out constants} \\
&= \frac{1}{\sum x_i^2 - n\bar{x}^2} \times \beta_0 n\bar{x} - n\bar{x}\beta_0 + \beta_1 \Bigg(\sum x_i^2 - \bar{x}n\bar{x} \Bigg) \\
&= \frac{1}{\sum x_i^2 - n\bar{x}^2} \times \beta_1 \Bigg(\sum x_i^2 - n\bar{x}^2 \Bigg) \\
&= \beta_1
\end{aligned}
$$