Simple Linear Regression

❑ Assessing the accuracy of the coefficient estimates

Standard error

The standard error of an estimator reflects how it varies under repeated sampling

$$\left\{\begin{aligned} \mathrm{Var}(\hat{\beta}_0) &= \mathrm{SE}(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] \\ \mathrm{Var}(\hat{\beta}_1) &= \mathrm{SE}(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \end{aligned}\right.$$

Proof:

For simple linear regression: $y_i = \beta_1 x_i + \beta_0 + \varepsilon$, ε is a random variable, so y_i is also a random variable.

x_i不是随机变量

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

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

Because $\sum_{i=1}^n (x_i - \bar{x}) = 0$, $\bar{y}\frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 0$

$$Var(\hat{\beta}_1) = Var(\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i)}{\sum_{i=1}^n (x_i - \bar{x})^2}) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 Var(y_i)}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2}, \quad Var(y_i) = Var(\varepsilon) = \sigma^2$$

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

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \Rightarrow Var(\hat{\beta}_0) = Var(\bar{y} - \hat{\beta}_1 \bar{x}) = Var(\bar{y}) + \bar{x}^2 Var(\hat{\beta}_1) - 2\bar{x}Cov(\bar{y}, \hat{\beta}_1)$$

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

$$Cov(\bar{y}, \hat{\beta}_1) = Cov(\frac{1}{n}\sum y_i, \sum c_j y_j) = \frac{1}{n}\sum c_j \sum Cov(y_i, y_j) = \frac{1}{n}\sum c_j Cov(y_j, y_j) = \frac{1}{n}\sum c_i = 0$$

So $Var(\hat{\beta}_0) = Var(\bar{y}) + \bar{x}^2 Var(\hat{\beta}_1) = \sigma^2[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}]$