Simple Linear Regression Model – Part I

Johns Hopkins Engineering

625.461 Statistical Models and Regression

Module 1 – Lecture 1D

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Model Basics

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

$$\begin{aligned} E(y \mid x) &= E(\beta_0 + \beta_1 x + \varepsilon) \\ &= \beta_0 + \beta_1 x \end{aligned}$$

$$\mathrm{Var}(y \mid x) = \mathrm{Var}(\beta_0 + \beta_1 x + \varepsilon) = \sigma^2$$

Data and Model Structures for Simple Linear Regression

Data structure:

n *independent* paired data: (y_1, x_1), …., (y_n, x_n)

Model structure:

$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, \ldots, n$

$\varepsilon_1, \ldots, \varepsilon_n$ are *statistically independent*

Least Squares Criterion and Minimization

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

Minimize S(β_0, β_1) and label the solution as $(\hat{\beta}_0, \hat{\beta}_1)$

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

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

Ordinary Least Squares (OLS) Estimators of Intercept and Slope

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

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

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

OLS Estimator of Slope in an alternative form

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

where

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

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

Residuals

$$e_i = y_i - \hat{y}_i = y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right), \quad i = 1, 2, \ldots, n$$

Properties of OLS Estimators

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

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

$$= \sum_{i=1}^{n} c_i(\beta_0 + \beta_1 x_i) = \beta_0 \sum_{i=1}^{n} c_i + \beta_1 \sum_{i=1}^{n} c_i x_i$$

$E\left(\hat{\beta}_1\right) = \beta_1$ because $\sum_{i=1}^{n} c_i = 0$, $\sum_{i=1}^{n} c_i x_i = 1$

$$E\left(\hat{\beta}_0\right) = \beta_0$$

Estimation of σ^2

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$SS_{\text{Res}} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$SS_{\text{Res}} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 - \hat{\beta}_1 S_{xy}$$

$$\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 \equiv SS_{\text{T}}$$

Estimation of σ^2

$$SS_{\text{Res}} = SS_{\text{T}} - \hat{\beta}_1 S_{xy}$$

The residual sum of squares has $(n - 2)$ degrees of freedom for estimating β_0 and β_1 and $E(SS_{\text{Res}}) = (n-2)\sigma^2$

An unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{SS_{\text{Res}}}{n-2} = MS_{\text{Res}}$$

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING