VE401 Probabilistic Methods in Eng.
RC 7

CHEN Xiwen

UM-SJTU Joint Institute

April 23, 2020

Table of contents

Simple Linear Regression

- Simple Linear Regression Model
- Distributions of Estimators, CIs, and Hypothesis Tests
- Predictions and Model Analysis
- Simple Linear Regression using Software

Multiple Linear Regression

Simple Linear Regression Model

Model. With x to be the parameter, the mean of the ***response*** $Y|x$ is given by

$$\mu_{Y|x} = \beta_0 + \beta_1 x \qquad \text{for some } \beta_0, \beta_1 \in \mathbb{R},$$

which is equivalent to

$$Y|x = \beta_0 + \beta_1 x + E,$$

where $\mathrm{E}[E] = 0$. We want to find estimators

$$\begin{aligned} B_0 &:= \widehat{\beta_0} = \text{estimator for } \beta_0, \qquad & b_0 &= \text{estimate for } \beta_0, \\ B_1 &:= \widehat{\beta_1} = \text{estimator for } \beta_1, \qquad & b_1 &= \text{estimate for } \beta_1. \end{aligned}$$

Simple Linear Regression Model

Model. Considering X as a parameter, we have a random sample of size n of $(x, Y|x)$.

$$Y_i := Y|x_i, \qquad i = 1, \ldots, n.$$

For each measurement y_i, we have a ***residual*** given by

$$y_i = b_0 + b_1 x_i + e_i.$$

Assumptions.

- For each value of x, the random variable follows a normal distribution with variance σ^2 and mean $\mu_{Y|x} = \beta_0 + \beta_1 x$.
- The random variables $Y|x_1$ and $Y|x_2$ are independent if $x_1 \neq x_2$.

Least Squares Estimation

LSE. We have the ***error sum of squares***

$$\mathrm{SS_E} := \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2.$$

To minimize it, we take

$$\frac{\partial \mathrm{SS_E}}{\partial b_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) = 0,$$

$$\frac{\partial \mathrm{SS_E}}{\partial b_1} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) x_i = 0.$$

Least Squares Estimation

LSE. We have

$$b_1 = \frac{S_{xy}}{S_{xx}}, \qquad b_0 = \overline{y} - b_1 \overline{x},$$

where

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

$$S_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = \sum_{i=1}^{n} (y_i - \overline{y}) y_i,$$

$$\begin{aligned} S_{xy} &= \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} (x_i - \overline{x}) y_i = \sum_{i=1}^{n} (y_i - \overline{y}) x_i \\ &= \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right). \end{aligned}$$

Distribution of B_1

Theorem. The least squares estimator B_1 for β_1 follows a normal distribution with

$$\mathrm{E}[B_1] = \beta_1, \qquad \mathrm{Var}[B_1] = \frac{\sigma^2}{\sum (x_i - \overline{x})^2}.$$

Proof. Knowing $Y|x_i = \beta_0 + \beta_1 x_i + E$ and $\mathrm{E}[E_i] = 0$, the expectation is given by

$$\begin{aligned}
\mathrm{E}[B_1] &= \mathrm{E}\left[\frac{1}{S_{xx}} \sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y})\right] = \mathrm{E}\left[\frac{1}{S_{xx}} \sum (x_i - \overline{x}) Y_i\right] \\
&= \frac{1}{S_{xx}} \left(\sum (x_i - \overline{x}) \mathrm{E}[\beta_0 + \beta_1 x_i + E_i]\right) \\
&= \frac{1}{S_{xx}} \left(\beta_1 \sum (x_i - \overline{x}) x_i\right) \\
&= \beta_1.
\end{aligned}$$

Distribution of B_1

Theorem. The least squares estimator B_1 for β_1 follows a normal distribution with

$$E[B_1] = \beta_1, \qquad \mathrm{Var}[B_1] = \frac{\sigma^2}{\sum (x_i - \overline{x})^2}.$$

Proof. Similarly, given $\mathrm{Var}[E_i] = \sigma^2$, the variance is given by

$$\begin{aligned}
\mathrm{Var}[B_1] &= \frac{1}{S_{xx}^2} \mathrm{Var}\left[\sum (x_i - \overline{x}) Y_i\right] \\
&= \frac{1}{S_{xx}^2} \sum (x_i - \overline{x})^2 \mathrm{Var}[\beta_0 + \beta_1 x_i + E_i] \\
&= \frac{\sigma^2}{\sum (x_i - \overline{x})^2} \\
&= \frac{\sigma^2}{S_{xx}}.
\end{aligned}$$

Distribution of B_0

Theorem. The least squares estimator B_0 for β_0 follows a normal distribution with

$$E[B_0] = \beta_0, \qquad \mathrm{Var}[B_0] = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \overline{x})^2}.$$

Proof. Using $\sum(x_i - \overline{x}) = 0$, the expectation is given by

$$\begin{aligned} E[B_0] &= E\left[\overline{Y} - \frac{\overline{x}}{S_{xx}} \sum (x_i - \overline{x}) Y_i\right] \\ &= \beta_0 + \beta_1 \overline{x} - \frac{\overline{x}}{S_{xx}} \sum (x_i - \overline{x})(\beta_0 + \beta_1 x_i) \\ &= \beta_0 + \beta_1 \overline{x} - \frac{\overline{x}}{S_{xx}} \sum (x_i - \overline{x}) x_i \beta_1 \\ &= \beta_0. \end{aligned}$$

Distribution of B_0

Theorem. The least squares estimator B_0 for β_0 follows a normal distribution with

$$\mathrm{E}[B_0] = \beta_0, \qquad \mathrm{Var}[B_0] = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \overline{x})^2}.$$

Proof. Similarly, using $\mathrm{Var}[\overline{E}] = \sigma^2/n$, the variance is given by

$$\begin{aligned}
\mathrm{Var}[B_0] &= \mathrm{Var}\left[\overline{Y} - \frac{\overline{x}}{S_{xx}} \sum (x_i - \overline{x}) Y_i\right] \\
&= \mathrm{Var}[\beta_0 + \beta_1 \overline{x} + \overline{E}] + \frac{\overline{x}^2}{S_{xx}^2} \sum (x_i - \overline{x})^2 \mathrm{Var}[\beta_0 + \beta_1 x_i + E_i] \\
&= \frac{\sigma^2}{n} + \frac{\overline{x}^2}{S_{xx}} \sigma^2 \\
&= \frac{S_{xx} + \overline{x}^2}{n S_{xx}} \sigma^2 \\
&= \frac{\sum x_i^2}{n S_{xx}} \sigma^2.
\end{aligned}$$

Thanks for your attention!