Estimation and testing in linear regression

Arjun Bhattacharya

6/16/2021

What is linear regression?

- A statistical model for the relationship between a dependent variable Y and one or more predictors
- What we'll cover:
  - Some basic mathematical theory and assumptions
  - Estimation and hypothesis testing of parameters
  - Interpretation of a linear regression model
  - Implementation of linear regression in R

The linear regression model

Simple linear regression:

$$E[Y] = \beta_0 + \beta_1 X + \epsilon$$

Multiple linear regression:

$$E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_k X_k + \epsilon$$

- If we keep $X_2, \dots, X_k$ constant, $E[Y]$ is still a linear (straight-line) function of X_1!

A few specifications and definitions

Linear regression model:

$$E[Y] = \beta_0 + \beta_1 X + \epsilon$$

- This ϵ parameter represents "random error", or the portion of Y that can't be explained by X.
- We will refer to *estimates* of β_0 and β_1 as $\hat{\beta}_0$ and $\hat{\beta}_1$
- The *fitted values* $\hat{Y}$ are our estimates of Y when we plug in $\hat{\beta}_0$ and $\hat{\beta}_1$ to the model: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$.
- The *residuals* $\hat{\epsilon}$ are the differences between the actual values of Y and the fitted values: $\hat{\epsilon} = Y - \hat{Y}$.

Ordinary least squares

For our simple linear regression model:

$$E[Y] = \beta_0 + \beta_1 X + \epsilon,$$

we estimate β_0 and β_1 by finding $\hat{\beta}_0$ and $\hat{\beta}_1$