STAT 331: Applied Linear Regression

Professor Kun Liang
LaTeXer Iris Jiang

Spring 2020

Contents

Chapter 1

Introduction to Regression

1.1 What is regression

Definition 1.1 (Regression analysis)**.** Regression analysis is a statistical methodology that models the functional relationship between a response variable y and one or more explanatory variables $x_1, x_2, \ldots, x_p$.

A typical regression model is:

$$y = f(x_1, x_2, \ldots, x_p) + \epsilon$$

- y: dependent variable or response variable
- $x_1, x_2, \ldots, x_p$: covariates, explanatory variables, independent variables, or predictors
- ϵ: random error term

Regression models can be used to:

- Identify important predictors
- Estimate regression coefficients
- Estimate the response for given values of predictors
- Predict of future values of response

In STAT 331, we focus on the simplest form of regression: linear models

$$\begin{aligned} y &= f(x_1, x_2, \ldots, x_p) + \epsilon \\ &= \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon \end{aligned}$$

where the β's are the regression parameters(coefficients).
Linear in the parameter (not predictor). Linear model is the basic building block of more complicated models

Remark. *We refer to the model as linear in the parameters β's ($\frac{\partial f}{\partial \beta_i}$ do not depende on the parameters)*

Example 1.1. Are the following models linear?

1. $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$

2. $f(x) = \beta_0 + \beta_1 e^{\beta_2 x}$

3. $f(x_1, x_2) = \beta_0 + \beta_1 x_1 x_2$

Solution.

1. This is a linear model. The predictor is x, this is not a linear model on the predictor but we define the linear model as to parameter, $\beta_0, \beta_1, \beta_2$ in this case.

2. This is not a linear model. If taking derivative to β_1, the result involves β_2.

3. This is a linear model.

$\square$

1.2 Why linear model?

- Linear model is easy to implement and interpret
- All functions can be approximated locally by a linear function
- The simplest starting model to fit

1.3 Sample vs. population

Definition 1.2 (Sample)**.** A sample is the collection of units (people, animals, cities, whatever you study) that is actually measure or surveyed.

Definition 1.3 (Population)**.** The population is the large group of unites we are interested in, from which the sample was selected.

Remark. *We assume the data we have a representative sample (random sample) from a larger population*