Section for Applied Statistics and Data Analysis

TA: Cong Mu

Office Hour: Wednesday 10:00AM - 12:00PM

November 8, 2019

Overview

1. Some Statistics
   - Categorical Predictors
     - Factors and Quantitative Predictors
     - Interpretation with Interaction Terms
     - Factors With More Than Two Levels
2. Some Programming
   - Examples in Faraway

Factors and Quantitative Predictors

- **Factor Predictors** (`factor` in R): A two-level factor variable could be represented by one dummy variable c where

$$c = \begin{cases} 0 & \text{reference level} \\ 1 & \text{treatment level} \end{cases}$$

- **Regression with Factors and Quantitative Predictors**

$$y = \beta_0 + \beta_1 x + \beta_2 c + \beta_3 xc.$$

  - Reference Level ($c = 0$)

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

  - Treatment Level ($c = 1$)

$$y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x.$$

Interpretation with Interaction Terms

- **Regression with Factors and Quantitative Predictors**

$$y = \beta_0 + \beta_1 x + \beta_2 c + \beta_3 xc.$$

  - Reference Level ($c = 0$)

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

  - Treatment Level ($c = 1$)

$$y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) x.$$

- Center x by replacing it with $x - \bar{x}$

Factors With More Than Two Levels

- In general, a d-level factor variable could be represented by $d-1$ dummy variables $c_2, \cdots, c_d$ where (assuming level 1 is the reference level)

$$c_i = \begin{cases} 0 & \text{not level } i \\ 1 & \text{level } i \end{cases}$$

- **Sequential Analysis of Variance** (anova in R)
- **Transformation**

Examples in Faraway Chapter 14

- **Example:** `sexab` dataset
- **Example:** `whiteside` dataset
- **Example:** `fruitfly` dataset

Thanks for listening!