Section for Applied Statistics and Data Analysis

TA: Cong Mu

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

October 4, 2019

Overview

1. Some Statistics
   - Observational Data
   - Confidence Intervals
   - Diagnostics
2. Some Programming
   - Examples in Faraway
   - Exercises in Faraway

Observational Data

- Consider
$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \cdots + \widehat{\beta}_p x_p.$$
- What does $\widehat{\beta}_1$ mean?
  - A unit change in x_1 will produce a change of $\widehat{\beta}_1$ in the response?
    - Lurking variable
    - Collinearity
  - The effect of x_1 when all the other predictors are held constant?
    - Not realistict in practice
    - The specification of other variables
- Steps to make a stronger case for causality
  - Include all relevant variables
  - Use nonstatistical knowledge
  - Try a variety of models
  - Multiple studies under different conditions
  - Infer causality from an observational study

Confidence Intervals for Coefficients

- Recall

$$\mathrm{Var}\left[\widehat{\beta}\right] = \sigma^2 \left(X^\top X\right)^{-1}, \qquad \widehat{\sigma}^2 = s^2 = \frac{\mathrm{SSE}}{n-p}.$$

- General form

$$\text{Estimate} \pm \text{Critical Value} \times \text{SE of Estimate}.$$

- In particular, $100(1-\alpha)\%$ CI for $\widehat{\beta}_i$

$$\widehat{\beta}_i \pm t_{n-p}^{(\alpha/2)} \widehat{\sigma} \sqrt{(X^\top X)_{ii}^{-1}}.$$

Confidence Intervals for Predictions

- Recall

$$\widehat{y}_0 = x_0^\top \widehat{\beta}, \qquad Var\left[\widehat{y}_0\right] = \sigma^2 x_0^\top \left(X^\top X\right)^{-1} x_0.$$

- $100(1-\alpha)\%$ CI for a single future response

$$\widehat{y}_0 \pm t_{n-p}^{(\alpha/2)} \widehat{\sigma} \sqrt{1 + x_0^\top \left(X^\top X\right)^{-1} x_0}.$$

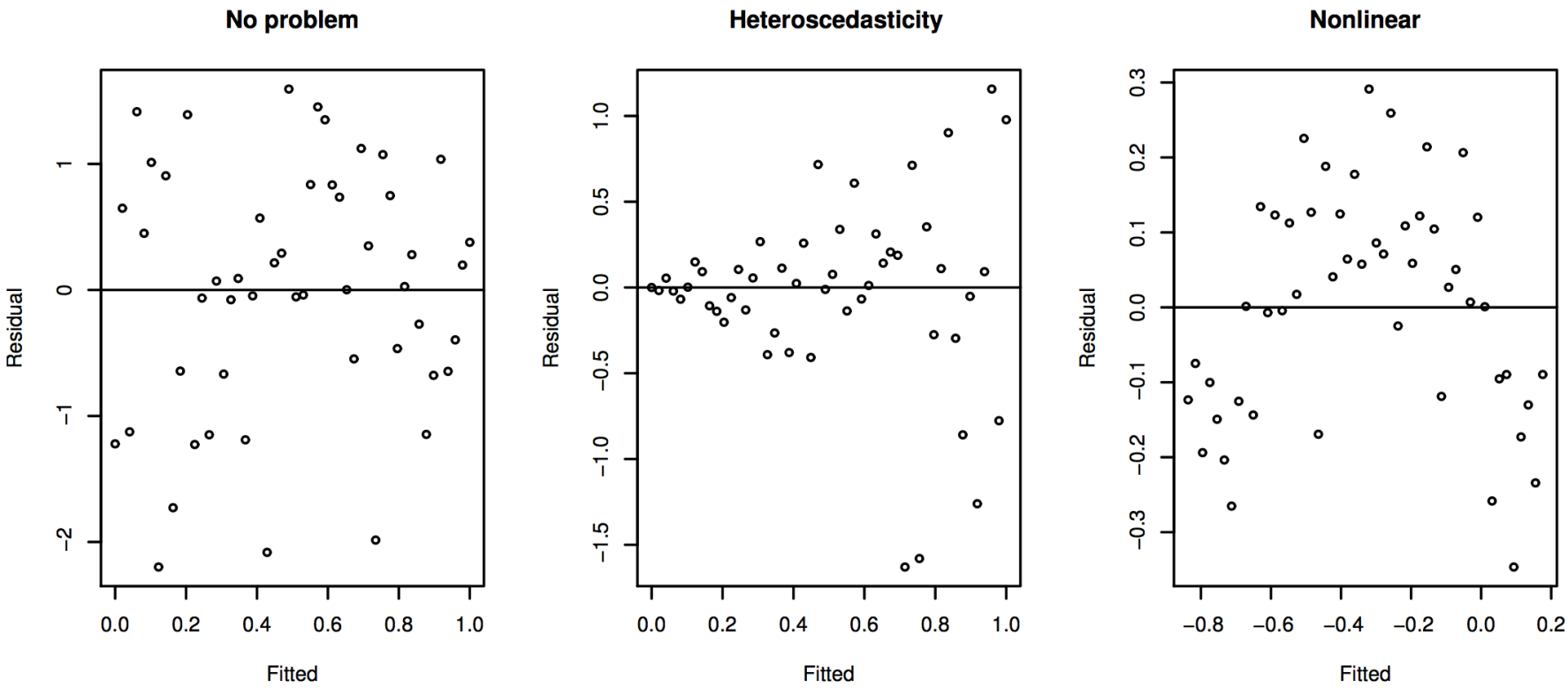
- $100(1-\alpha)\%$ CI for the mean response for given x_0

$$\widehat{y}_0 \pm t_{n-p}^{(\alpha/2)} \widehat{\sigma} \sqrt{x_0^\top \left(X^\top X\right)^{-1} x_0}.$$

Diagnostics

- Recall
$$\epsilon \sim \mathcal{N}\left(0, \sigma^2 \mathrm{I}\right).$$
- Checking Error Assumptions
  - Constant Variance
  - Normality
  - Correlated Errors
- Finding Unusual Observations
  - Leverage
  - Outliers
  - Influential Observations
- Checking the Structure of the Model

Constant Variance



(Figure from Linear Models with R)

Examples and Exercises in Faraway Chapter 3 & 4

- **Example:** `savings` dataset
- **Example:** `gala` dataset
- **Exercise 1:** `prostate` dataset

Thanks for listening!