STAT 408
Applied Regression Analysis

Nan Miles Xi

Department of Mathematics and Statistics

Loyola University Chicago

Fall 2022

Midterm Review

Time and Format

- Oct 20, Thursday, 13:00 – 14:15

- Open book, take home

- Posted on Sakai

- Five questions, 50 points in total

- There are three categories of questions
    1. Conceptual question
        Definition, terminology, model assumption
    2. Coding question
        Understand code output, connect code to methodology and concepts
    3. Proof question

Question One

- Question one is a conceptual question, you need to know
  - The commonly used terminology in linear model
  - Difference between the known part and unknown part in the model
  - Understand the random part and fixed part in the model

- You need to understand the assumption we made to conduct inference in linear model

- State all assumptions in both English and mathematical language

$$\varepsilon_i \sim N\left(0, \sigma^2\right), i = 1, 2, \ldots, n$$

$$\varepsilon_i \perp \varepsilon_j \text{ for } i \neq j$$

Question Two

- You will be given one real dataset, the code, and output to analyze the dataset

- You need to know
  - The output of lm function, what is the meaning of each term
  - Different tests conducted in lm function, how to interpret the test result
  - The anova function, what does it do and what is the output
  - Be able to write down the null and alternative hypothesis based on previous output
  - Understand the model degree of freedom, what is the relation among sample size, number of parameters, and number of predictors

Question Three

- You will be given another real dataset, the code, and output to analyze the dataset

- Give appropriate interpretations for model parameters, especially for categorical predictors

- You will be given some code which you may or may not see before, try to understand them either by guessing or search

- Understand how we can transform the predictors into other form, compare the difference of parameter interpretation after the transformation

Question Four

- Question one is a conceptual question

- Given a plot of Y vs. X, identify outlier or influential point based on the criteria we discussed in lectures

- Note the there may be overlap among those categories, list all possible identities

Question Five

- Question five is a proof question

- In this question, you will use least square estimation to derive a specific $\hat{\beta}$

- The goal of this question is to show why we use the average of response to estimate the null model

- I will show you this question now, you can try to find your solution and write them in the test

Question Five

Consider the simple linear regression model with only intercept (null model):

$$Y_i = \beta_0 + \varepsilon_i$$

where $i = 1, 2, \ldots, n$. Prove that the least square estimation of intercept $\widehat{\beta_0}$ is the average of response $\bar{y}$. Use two methods: plain derivate and matrix form.