

Homework 3

Due on Thursday, Oct. 26th by 11:00 am

*You are expected to solve all the problems by yourselves at first and then discuss with your group members. **Please turn in one report per group with names of group members. If you don't join in group discussion, please turn in your own homework report individually.** Make sure you show all of your work and attach your R-script for full credit. The datasets for the problems can be downloaded from Canvas. Please turn in your homework report right before the lecture time.*

1. (20 points) In a study, we have observed response variable Y and predictor variables X_1 and X_2 , which shows some linear pattern in the scatter plot. So we propose a multiple linear model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon, \quad (1)$$

- (a) We consider the test

$$H_0 : \beta_1 = 0 \text{ vs. } H_a : \beta_1 \neq 0,$$

and find the p-value is 0.9 and hence fail to reject the null hypothesis. What does your conclusion mean? No association between Y and X_1 , or something else?

- (b) If we are only interested in whether there is linear association between Y and X_1 , could we use model (1) and the hypothesis test in part (a) to make the same decision? If no, please write down the alternative model and the corresponding hypothesis test on the parameter in the new model.
2. (30 points) Sec. 5.4, Problem 1 in textbook. You are expected to answer the following three parts.
 - (a) Get the scatter plot of “LATE” vs “BILL” and label the residential accounts with “R” and commercial accounts with “C” in the plot. (*hint: the dataset does not include the variable “TYPE”, but you could define it in R as “TYPE = c(rep(1,48),rep(0,48))”, where 1 denotes “residential” and 0 denotes “commercial”*).
 - (b) Based on the scatterplot, propose a linear model and implement it in R.
 - (c) We want to know whether there is any difference between residential accounts and commercial accounts in terms of overdue dates. Write hypothesis test for this question and implement it in R to make a conclusion.
 3. (30 points) Sec. 5.4, Problem 2 in textbook. In each sub-question, you are expected to write down the model you use and the corresponding hypothesis test and then implement it in R to get the answer. (*hint: The original data used the actual value of “Year”, but we could relabel them with 1/0. For example, you could use the R-code: “Year[which(Year == 2004)] <- 1 Year[which(Year == 1994)] <- 0”*)
 4. (20 points) Sec. 5.4, Problem 3 in textbook. (*hint: figure out a hypothesis test first and then get the p-value based on the output posted to answer the question in part (a)*)