Econ 184b, Econometrics, Assignment 4

The raw .rmd file for this assignment are available at: https://raw.githubusercontent.com/f1kidd/Econ184/main/Assignment/Assignment_4.Rmd

Note: For the math problems, you can either (1) type your solution and compile it with RMarkdown; (2) type your solution use another word-processing software and convert it into a PDF file; (3) write your solution on paper, scan it and make it into a legible PDF file. TAs can decide, at their discretion, that the submitted file is illegible and thus give zero credit to a question or the entire problem set. If you type your solutions (options 1 and 2 above), you will get **a bonus equal to 5% of your performance** on the problem set. No credit will be given if you only report the final answers without showing intermediate steps whenever appropriate.

For the programming problems, you need to submit both the compiled pdf/html file and the RMarkdown (.rmd) file on LATTE. You will get **an additional bonus equal to 5% of your performance** if your RMarkdown file is completely reproducible with minimal altercation (install packages, etc.). That is, our team (or anyone else) should be able to recompile your Rmarkdown file and reach *the same* result. You need to explicitly write out your answers - just showing programming outputs will receive zero credit.

You will be receiving a total of 15% bonus if you submit one single pdf/html compiled directly from Rmarkdown, along with the .rmd file.

The first three problems are based on the following setting.

Suppose that you are studying the mortgage market. You have collected data on 78 households. For each household, you observe the following variables:

- "interest": the mortgage interest rate
- "income": annual income of the borrower
- "credit": the credit rating of the borrower. This is a categorical variable taking three possible values: high, medium and low.
- "self": whether or not the borrower is self-employed. This is a Bernoulli variable: 1 means "self-employed" and 0 means "not self-employed".

The goal is to explain "interest".

Question 1:

Ignore "income" and "self". Run a regression of "interest" on "credit", i.e., predict "interest" using "credit". Consider the regression equation:

$$interest = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

- X_1 is binary (1 means "medium" and 0 means "not medium")
- X_2 is binary (1 means "low" and 0 means "not low")

Answer the following questions:

a. We are interested in whether or not "credit" is useful in predicting "interest". Use the above regression equation and write down the null and alternative hypotheses.
b. We are interested in whether or not the prediction for someone with "credit"=medium is the same as the prediction for someone with "credit"=low. Use the above regression equation and write down the null and alternative hypotheses.

c. We are interested in whether or not the prediction for someone with "credit"=medium is the same as the prediction for someone with "credit"=high. Use the above regression equation and write down the null and alternative hypotheses.
d. In part (b), is the null hypothesis on individual regression coefficients? (Does it involve only one of β_1 and β_2?) If not, transform the regression equation such that we can study the same comparison in part (b) by testing only one coefficient in the transformed regression. In this case, write down the transformed regression equation and the null hypothesis based on the transformed regression.

Question 2

Ignore "income". Consider the following regression:

$$interest = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_3 + \beta_5 X_2 X_3 + u$$

- X_1 is binary (1 means "medium" and 0 means "not medium")
- X_2 is binary (1 means "low" and 0 means "not low")
- X_3="self" (X_3=1 means "self-employed" and X_3=0 means "not self-employed")

Answer the following questions:

a. What is the prediction for someone who has a medium credit rating and is self-employed? Write your answer in terms of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$.
b. What is the prediction for someone who has a low credit rating and is self-employed? Write your answer in terms of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$.
c. What is the prediction for someone who has a high credit rating and is self-employed? Write your answer in terms of $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$.
d. You have a theory: for those that are self-employed, the credit rating does not predict the interest rate of their mortgage. Write down the null and the alternative hypotheses for studying this theory.

Question 3

Ignore "self". Consider the following regression:

$$interest = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_3 + \beta_5 X_2 X_3 + u$$

- X_1 is binary (1 means "medium" and 0 means "not medium")
- X_2 is binary (1 means "low" and 0 means "not low")
- X_3 = income

You use the above regression to study the effect of income on the mortgage interest rate. Answer the following questions:

a. You have a theory: for those with a low credit rating, they always get the maximum interest rate, regardless of their income. (In other words, for those with a low credit rating, "income" does not predict "interest".) Use the above regression equation and write down the null and alternative hypotheses for studying this theory.
b. You are interested in finding out whether or not the effect of income on the interest rate depends on the credit rating. Use the above regression equation and write down the null and alternative hypotheses.

Question 4 :

Does a student's GPA rise or fall as the student moves from freshman year to senior year? In this question you will explore this issue – and other factors that are correlated with GPA - using data on college students collected by researchers at the Harvard School of Public Health. The researchers surveyed 9,890 undergraduate students at 119 four-year colleges in 2001. Use the data set college.csv ("https://raw.githubusercontent.com/f1kidd/Econ184/main/Data/college.csv"). Make sure to use robust standard errors in your regressions.

a. Estimate the regression of GPA on *male* and *work*. Interpret the regression coefficients (including the intercept).

b. Estimate the regression of GPA on *freshman*, *sophomore*, *junior* and *senior* for men only. Are all regression coefficients reported in the results? Explain what happened. What is the solution?
c. Estimate the regression of GPA on *sophomore*, *junior* and *senior* for men only. What is the interpretation of all coefficients in this regression (including the intercept)?
d. For the regression in (c), test the null hypothesis that there is no difference in the GPA of male sophomores and juniors. What is the number of restrictions q for this test?
e. For the regression in (c), test the null hypothesis that the coefficients on *sophomore*, *junior* and *senior* are all zero, against the alternative that at least one coefficient is nonzero. State clearly the significance level of the test you are using and the critical value of the F statistic for this test.
f. Estimate the regression of GPA on *age*, *sophomore*, *junior* and *senior* for men only. What happens to the statistical significance of the coefficients in this regression? Explain why this is the case.
g. Estimate the same regression as in part (c), but now do this for all respondents (male and female). Include controls for *male*, *work*, *marijuana*, *lightdrinker*, *moddrinker* and *heavydrinker*. Calculate the predicted GPA for a male senior who works, is a moderate drinker and has not smoked marijuana in the past 30 days.
h. Using the regression in part (g), test the hypothesis (at the 5% significance level) that freshmen and sophomores have the same GPA on average, holding constant gender, work, type of drinking and marijuana use.
i. What is the adjusted R2 for the regression in part (h)? What does this adjusted R2 tell you about the fit of this regression? Does it indicate that omitted variable bias is likely to be a problem?

Your code could look like the following:

```
library(tidyverse)
data = read_csv("https://raw.githubusercontent.com/f1kidd/Econ184/main/Data/college.csv")
# ......
```