Problem Set 4

QTM 200: Applied Regression Analysis

Due: February 24, 2020

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in `R`, please include the code you used to get your answers. Please also include the `.R` file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.

- Your homework should be submitted electronically on the course GitHub page in `.pdf` form.

- This problem set is due at the beginning of class on Monday, February 24, 2020. No late assignments will be accepted.

- Total available points for this homework is 100.

Question 1 (50 points): Economics

In this question, use the `prestige` dataset in the `car` library. First, run the following commands:

```
install.packages(car)
library(car)
data(Prestige)
help(Prestige)
```

We would like to study whether individuals with higher levels of income have more prestigious jobs. Moreover, we would like to study whether professionals have more prestigious jobs than blue and white collar workers.

(a) Create a new variable **professional** by recoding the variable **type** so that professionals are coded as 1, and blue and white collar workers are coded as 0 (Hint: **ifelse**.)

```
1  library(dplyr)
2  Prestige1<-Prestige %>% mutate(type=ifelse(Prestige$type=="prof",1,0))
```

(b) Run a linear model with prestige as an outcome and **income**, **professional**, and the interaction of the two as predictors (Note: this is a continuous $\times$ dummy interaction.)

```
1  Model1<-lm(prestige~income + type + income:type,data=Prestige1)
2  summary(Model1)
```

(c) Write the prediction equation based on the result.

```
1  #prestige = 21.142 + .003(income) + 37.781(type) -.002(income)(type)
```

(d) Interpret the coefficient for `income`.

```
1 #For every increase in average income of $1, the prestige score of one's
      occupation increases by .003 on average
```

(e) Interpret the coefficient for `professional`.

```
1 #Being a professional increases the prestige score of one's occupation by
      37.781 on average in comparison to being white collar or blue collar
```

(f) What is the effect of a \$1,000 increase in income on prestige score for professional occupations? In other words, we are interested in the marginal effect of income when the variable `professional` takes the value of 1. Calculate the change in $\hat{y}$ associated with a \$1,000 increase in income based on your answer for (c).

```
1  prestige1f<-21.142 + .003*(1000) + 37.781 - .002*(1000)
2  prestige1f
3  #A $1000 increase in income for professional occupations results in an
       average increase in prestige score of 59.923
```

(g) What is the effect of changing one's occupations from non-professional to professional when her income is \$6,000? We are interested in the marginal effect of professional jobs when the variable `income` takes the value of $6,000$. Calculate the change in $\hat{y}$ based on your answer for (c).

```
1  prestige1g<-(21.142 + .003*(6000) + 37.781 -.002*(6000)) - (21.142 + .003
       *(6000))
2  prestige1g
3  #Changing one's occupation from non-professional to professional at an
       income of $6000 results in an average increase in prestige score of
       25.781
```

Question 2 (50 points): Political Science

Researchers are interested in learning the effect of all of those yard signs on voting preferences.[1] Working with a campaign in Fairfax County, Virginia, 131 precincts were randomly divided into a treatment and control group. In 30 precincts, signs were posted around the precinct that read, "For Sale: Terry McAuliffe. Don't Sellout Virgina on November 5."

Below is the result of a regression with two variables and a constant. The dependent variable is the proportion of the vote that went to McAuliff's opponent Ken Cuccinelli. The first variable indicates whether a precinct was randomly assigned to have the sign against McAuliffe posted. The second variable indicates a precinct that was adjacent to a precinct in the treatment group (since people in those precincts might be exposed to the signs).

Impact of lawn signs on vote share

Precinct assigned lawn signs (n=30)	0.042
	(0.016)
Precinct adjacent to lawn signs (n=76)	0.042
	(0.013)
Constant	0.302
	(0.011)

Notes: R^2=0.094, N=131

(a) Use the results to determine whether having these yard signs in a precinct affects vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

```
1  test.stat2a<-.042/.016
2  p2a<-2*pt(test.stat2a,127,lower.tail=F)
3  p2a # = .010
```

[1]Donald P. Green, Jonathan S. Krasno, Alexander Coppock, Benjamin D. Farrer, Brandon Lenoir, Joshua N. Zingher. 2016. "The effects of lawn signs on vote outcomes: Results from four randomized field experiments." Electoral Studies 41: 143-150.

(b) Use the results to determine whether being next to precincts with these yard signs affects vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

```
1  test.stat2b<-.042/.013
2  p2b<-2*pt(test.stat2b,127,lower.tail=F)
3  p2b #.002
```

(c) Interpret the coefficient for the constant term substantively.

```
1  #In precincts that neither had yard signs themselves nor were adjacent to
       precincts with yard signs, the average vote share for Cuccinelli was
       .302.
```

(d) Evaluate the model fit for this regression. What does this tell us about the importance of yard signs versus other factors that are not modeled?

```
1  #Since the r^2 value is only .094, this model only explains 9.4% of the
       variation in vote share. This means that there are a lot of other
```

variables that influence vote share, which likely have a much larger effect.