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`.)

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

#1a.
Prestige$professional <- ifelse(Prestige$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 × dummy interaction.)

```
#1b.
lm1 <- lm(prestige~income+professional+income:professional, data=Prestige
    )
summary(lm1)
```

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

prestige hat = 21.142 + 0.003*(income) + 37.781(professional) -0.002(income)(professional)

(d) Interpret the coefficient for `income`.

For every one dollar increase in average income of incumbents in 1971, there is about a 0.003 increase in Pineo-Porter prestige score for occupation (from a social survey conducted in the mid-1960s).

(e) Interpret the coefficient for `professional`.

Switching from being a blue or white collar worker to being a professional results in a 37.781 increase in Pineo-Porter prestige score.

(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).

```
#1f.
prestige1 <- 21.1422589 + 0.0031709*(1000) + 37.7812800*(1) - 0.0023257*
    (1000)*(1)
prestige1
```

The effect of a 1,000 dollar increase in income on prestige score for professional occupation is an increase of 59.76874 on the Pineo-Porter prestige score

(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).

```
#1g.
prestige2 <- (21.1422589 +  0.0031709*(6000) + 37.7812800*(1) -0.0023257*
    (6000)*(1)) -(21.1422589 +  0.0031709*(6000) + 37.7812800*(0)
    -0.0023257*(6000)*(0))
prestige2
```

Someone who changes from non-professional to professional when her income is 6,000 dollars results in a 23.82708 increase in prestige score.

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 $\alpha = .05$).

```
TS1 <- 0.042/0.016
p1 <-2*pt(TS1,df = 131-2-1,lower.tail=F)
p1
```

Because the p-value of 0.00972002 is less than the alpha value of 0.05, we have sufficient evidence to reject the null hypothesis that the yard signs have no effect on vote share.

[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 $\alpha = .05$).

```
TS2 <- 0.042/0.013
p2 <-2*pt(TS2, df = 131-2-1,lower.tail=F)
p2
```

Because the p-value of 0.00156946 is less than the alpha value of 0.05, we have sufficient evidence to reject the null hypothesis that being next to precincts with these yard signs has no effect on the vote share.

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

In precincts that were not assigned lawn signs and did not have lawn signs adjacent, on average, 30.2 percent of the vote went to Ken Cuccinelli.

(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?

According to the table's provided R-squared value of 0.094, the model is only accounting for 9.4 percent of the variance in votes going to Ken Cuccinelli. Because 9.4 percent is quite small, this means that the yard signs are likely not a very important factor affecting the vote share as opposed to other factors that are not included in the model that are accounting for about 90 percent of the variance in the vote share.