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

#Question 1: We would like to study whether individuals with higher levels of income have
# more pretigious jobs.

#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.

Prestige$professional <- ifelse(Prestige$type=="prof", 1, 0)
#check recoding:
Prestige$type
Prestige$professional
```

(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.)

```
#b) Run a linear model with prestige as an outcome and income, professional, and the
#   interaction of the two as predictors. (Continuous x dummy interaction)
lm(prestige ~ income + professional + income:professional, data = Prestige)
```

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

```
#c) Write the predition equation based on the result.

#Yi = 30.618 + 0.0014(incomeXi) + 22.757(professionalXi) + −0.0023
```

(d) Interpret the coefficient for `income`.

```
#d) Interpret the coefficient for income.

# Controlling for type of profession, a 1 unit increase in income is associated
# with a 0.0014 increase in the prestige of an occupation.
```

(e) Interpret the coefficient for `professional`.

```
#e) Interpret the coefficient for professional.

# Controlling for income, being a "professional" as a opposed to a white collar or
# blue collar worker is associated with a 22.757 increase in the Pineo-Porter
# prestige score of an occupation.
```

(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).

```
#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 yhat associated with an $1,000 increase in income
#   based on your answer for c.

30.618 + 0.0014*(1000) + 22.757*(1) + -0.0023
# The prestige score of a professional occupation with a $1,000 income is about 54.7727.
# Calculate the prestige score of the same with an extra $1,000 in income:
30.618 + 0.0014*(2000) + 22.757*(1) + -0.0023
# The prestige score is 56.173.
# Subtract the prestige score of the base income from the boosted income:
56.173 - 54.773
# An increase of $1,000 in income produces a marginal effect of a 1.4 increase in the prestige
# score of a professional occupation.
```

(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).

```
#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 yhat
#   based on your answer for c.

30.618 + 0.0014*(6000) + 22.757*(1) + -0.0023
# The prestige score of a professional occupation with an income of $6000
    is 61.773.
# Now calculate with same income and non-professional occupation:
30.618 + 0.0014*(6000) + 22.757*(0)
# The prestige score of a non-professional occupation with an income of $
    6000 is 39.018.
# Subtract the non-professional prestige score from the professional
    prestige score to find
# the marginal effect:
61.773 - 39.018
# The marginal effect of a professional occupation on a $6,000-income job
# is a 22.755 increase in the 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 $\alpha = .05$).

```
#Question 2: Researchers are interested in learning the effect of yard signs on voting
# preferences.
#a) Use the results of the regression table to determine whether having these yard signs in
#   a precinct affects vote share. Conduct a hypothesis test with alpha = 0.05.

#H0: B2 = 0, Ha: B2 =/= 0
n <- 30
#Find the test statistic:
(0.042 - 0)/0.016
```

[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.

```
#The test statistic is 2.625.
#Find the p-value:
2*pt(-abs(2.625), df=n-1)
#The p-value is 0.0137. This is lower than our alpha at 0.05, meaning we can reject our null
# hypothesis that having yard signs in a precinct has no effect on vote share.
```

(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 $\alpha = .05$).

```
#b) Use the results to determine whether being next to precincts with these yard signs
#   affects vote share. Conduct a hypothesis test with alpha = 0.05.

#H0: B3 = 0; Ha: B3 =/= 0
n2 <- 76
#Find the test statistic:
(0.042 - 0)/0.013
#The test statistic is 3.23.
#Find the p-value:
2*pt(-abs(3.23), df=n2-1)
#The p-value is 0.00183. This is lower than our alpha at 0.05, meaning we can reject our
# null hypothesis that being in a precinct adjacent to the yard signs does not have an effect
# on vote share.
```

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

```
#c) Interpret the coefficient for the constant term substantively.

#On average, a precinct that was neither adjacent to a precinct with the yard signs nor assigned
# the yard signs itself was expected to have 0.302, or 30.2%, of its voteshare go to McAuliff's
# opponent.
```

(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?

```
#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?

#The r-squared value is 0.094, meaning about 9.4% of the variation in voteshare can be explained by
# our model. This is not a very large percentage, indicating that there are probably other variables
# that have a much larger effect on vote share, such as income of the precinct, or polarization, for
# example.
```