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`.)

```
#changing the variable to professional from type
summary(Prestige)
prestige_data<- Prestige %>% rename(professional = type)
view(prestige_data)
```

(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.)

```
#plotting prestige as outcome, income, professional, and interaction of
    the two as predictors
lm(prestige ~ income + professional + professional:income, data =
    prestige_data)
```

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

$$y = 13.90 + .00402x_i + 45.02D1_i + 18.99D2_i - 0.003178x_iD1_i - .0022X_iD2_i$$

(d) Interpret the coefficient for `income`.

Given the individual in the dataset, on average, as income increases by 1 dollar, the individual's prestige will go up by 0.004023 points regardless of one's profession.

(e) Interpret the coefficient for `professional`.

If the individual's profession is a professional, on average, the individual's prestige will go up by 45.02 points whereas if the individuals profession is white collar, on average, the individual's prestige will go up by 18.98 points

(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).

```
13.904517 + (.004023 * 1000) + (45.019022 *1) + (18.980739 * 0) -
    (.003178 *1000 *1)
#59.769 when professional with income of 1000

13.904517 + (.004023 * 0) + (45.019022 *1) + (18.980739 * 0) - (.003178 *
    1000 * 0)
#55.754 when income is 0
59.769 - 58.92354
```

The change in $\hat{y}$ associated with a \$1000 increase in income is 0.84546

(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).

```
13.904517 + (.004023 * 6000) + (45.019022 *1) + (18.980739 * 0) -
    (.003178 *1000 *1) - (.002171 * 1000 *0)
#79.88354
13.904517 + (.004023 * 6000) + (45.019022 *0) + (18.980739 * 1) -
    (.003178 *1000 *0) - (.002171 * 1000 *1)
#54.85226
79.88354-54.85226
```

The change in $\hat{y}$ associated with an income of \$6000 when going to a professional status is 25.03 in prestige

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 $\alpha = .05$).

```
xbar = .042 # sample mean
mu0 = .302 # hypothesized value
#std = se * sqrt(n)
sigma = .016*(sqrt(30)) #standard error converted to standard deviation =
    .088
n = 131 # sample size
z = (xbar-mu0)/(sigma/sqrt(n)) #-16.25 is the z score

alpha = .05
z.half.alpha = qnorm(1-alpha/2)
c(-z.half.alpha, z.half.alpha) #from -1.96 to 1.96 for 95% CI
```

[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.

With our z-score of -16.25, which does not lie within the 95% CI range of -1.96 to 1.96, we can reject the null, that there is an affect on vote share based on having yard signs in a precinct.

(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 $\alpha = .05$).

```
### Part B ###
xbar2 = .042 # sample mean
mu02 = .302 # hypothesized value
sigma2 = .013*(sqrt(76)) #standard error converted to standard deviation
    = 113
n2 = 131 # sample size
z2 = (xbar2-mu02)/(sigma2/sqrt(n2)) #-20 is the z score

alpha2 = .05
z.half.alpha2 = qnorm(1-alpha2/2)
c(-z.half.alpha2, z.half.alpha2) #from -1.96 to 1.96 for 95% CI
```

With our z-score of -20, which does not lie within the 95% CI range of -1.96 to 1.96, we can reject the null, that there is an affect on vote share based on being next to precincts with these yard signs.

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

The constant (.302, with a standard error of .011) has been identified as the proportion of the vote that went to McAuliff's opponent Ken Cuccinelli. As a result, if there were no yard signs, the vote share proportion would be .302 in favor of Cuccinelli. From our data analysis, there can be a difference having the signs up for degrading McAuliff, to boost Cuccinelli's proportion of the vote share whether the method is to add the signs within certain precincts, or even be neighboring to precincts with these signs.

(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 model fit for this regression is not good, with an $R^2 = .094$, meaning, the proportion of the variance in the dependent variable that is explained by the independent variable, is quite low. As a result, the yard signs alone are not a good predictor of whether or not there is an impact on vote share.

One thing that I could suggest as a possible variable, is the precinct location as a whole, depending on their demographic make-up as well as their previous political voting history.