

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 # New factor variable (dummy variable)
2 # Recoding professional workerse as 1, blue/white collar workers as 0
3 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 \times dummy interaction.)

```
1 # Linear model with prestige as outcome and income, professional, and the
  interaction of the two as predictors
2 lm1 <- lm(prestige ~ income + professional + income:professional, data =
  Prestige)
3 summary(lm1)
```

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

$$\text{prestige} = 21.1423 + 0.0032 \times \text{income} + 37.7813 \times \text{professional} - 0.0023 \times \text{income} \times \text{professional}$$

- (d) Interpret the coefficient for **income**.

0.0032 is the coefficient associated with income for blue collar/white collar group. Since they take the value of 0 for the professional variable, the professional variable and the interaction variable will all cancel out. A 1 dollar increase income is associated with a 0.0032 unit increase in prestige for blue/white collar workers.

0.0009 is the coefficient associated with income for the professional group. Since they take the value of 1 for the professional variable, the coefficient attached to the interaction variable (-0.0023) is combined with the coefficient on the income variable to result in 0.0009 (0.0032 - 0.0023). A 1 dollar increase in income is associated with a 0.0009 unit increase in prestige for professionals.

- (e) Interpret the coefficient for **professional**.

37.7813, the coefficient for professional variable, is the effect associated with the professional group when controlling for income. The professional group exhibits, on average, 37.7813 units more prestige than blue/white collar workers.

- (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).

$$\text{prestige} = 21.1423 + 0.0032 \times \text{income} + 37.7813 \times \text{professional} - 0.0023 \times \text{income} \times \text{professional}$$

```
1 prestige_0 <- 21.1423 + 37.7813 # 58.9236 units of prestige with $0
  income
```

```

2 prestige_1000 <- 21.1423 + 0.0032 * 1000 + 37.7813 - 0.0023 * 1000 #
  59.8236 units of prestige with $1000 income
3 prestige_1000 - prestige_0 # difference between the two -> 0.9 unit
  change in prestige with $1000 increase

```

The marginal effect of a \$1000 increase in income is 0.9 units of prestige when we use the prediction equation and substitute in the respective values.

- (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 prestige_prof <- 21.1423 + 0.0032 * 6000 + 37.7813 - 0.0023 * 6000 #
  64.3236 units of prestige for professional with $6000 income
2 prestige_bc <- 21.14 + 0.0032 * 6000 # 40.34 units of prestige for blue/
  white collar workers with $6000 income
3 prestige_prof - prestige_bc # 23.9836 units change in prestige between
  professionals and blue/white collar workers

```

The marginal effect of professional jobs when income is \$6000 is 23.9836 units of prestige when we use the prediction equation and substitute in the respective values.

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 Virginia 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 McAuliffe’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 $\alpha = .05$).

First, I set the null hypothesis as $H_0 : \beta_1 = 0$ and the $H_a : \beta_1 \neq 0$. Then I calculated the t-test statistic using the outputs given above using the formula $t = \frac{\hat{\beta} - 0}{se}$. The degrees of freedom was $n - 3$ ($df = n - 1 - k$) since there were two variables for k . I calculated the p-value by finding the upper tail area and multiplying it by two.

```
1 n <- 30 # sample size
2 ts <- (0.042 - 0) / 0.016 # calculating the test statistic given the R
  outputs
3 p_value <- 2 * pt(abs(ts), df = n - 3, lower.tail = F) # calculating the
  p-value
4 p_value # 0.014
```

¹Donald 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.

Since the p-value of 0.014 is less than the alpha value of 0.05, we reject the null hypothesis that the β_1 is equal to 0. Therefore, we have sufficient evidence to conclude that having yard signs in the precincts affect 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 $\alpha = .05$).

First, I set the null hypothesis as $H_0 : \beta_2 = 0$ and the $H_a : \beta_2 \neq 0$. Then the test statistic, degrees of freedom, and the p-value were calculated using the same formulas in part a.

```
1 n2 <- 76 # sample size
2 ts2 <- (0.042 - 0) / 0.013 # calculating the test statistic
3 p_value2 <- 2 * pt(abs(ts2), df = n2 - 3, lower.tail = F) # calculating
  the p-value
4 p_value2 # 0.0019
```

Since the p-value of 0.0019 is less than the alpha value of 0.05, we reject the null hypothesis that the β_2 is equal to 0. Therefore, we have sufficient evidence to conclude that having yard signs in neighboring precincts affect vote share.

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

The coefficient for the constant term which equals 0.302 represents, on average, the vote share for Cuccinelli when there are no yard signs in precincts or the neighboring precincts. 30.2% of the vote share would go to Cucinelli in this scenario.

- (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?

Based on the R-squared value which equals 0.094, we can interpret that 9.4% of the variation in vote share can be explained by the current model. Therefore, yard signs are not a significant contributor to vote share, compared to other factors that were not considered in this study, for example like political party identification or demographics.