

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.

```
1 install.packages(car)
2 library(car)
3 data(Prestige)
4 help(Prestige)
5
6 summary(Prestige)
```

- (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 #a)
2 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 #b)
2 LinearModel1=lm(prestige ~ income + professional+ income*professional,
3                 data = Prestige)
3 LinearModel1
4 summary(LinearModel1)
```

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

$$\text{Prestige} = 21.1422589 + 0.0031709 * (\text{income}) + 37.7812800 * (\text{professional}) - 0.0023257 * (\text{income} * \text{professional})$$

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

The coefficient for income is a positive number, which suggests that income is positively correlated with prestige. Moreover, the coefficient suggests that when there is a dollar increase in income, there will also be 0.0031709 unit of increase in the value of prestige, given all other factors in the model are constant.

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

The coefficient for professional is a positive number, which suggests that professional is positively correlated with prestige. Moreover, the coefficient suggests that when this person is identified as professional, his prestige value will immediately increase by 37.78 units, given all other factors in the model are constant.

- (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).

$$\text{Prestige} = 21.1422589 + 0.0031709 * (\text{income}) + 37.7812800 * (\text{professional}) - 0.0023257 * (\text{income} * \text{professional})$$

Because we are talking about professional only

$$\text{professional} = 1$$

$$\text{Prestige} = (21.1422589 + 37.7812800) + (0.0031709 - 0.0023257) * (\text{income}) = 58.92 + 0.0008 * \text{income}$$

When income increases by 1000 dollars

$$\text{difference} = 0.0008 * 1000 = 0.8$$

The effect of 1000 dollar increase in the income given professional is 1 will be a 0.8 unit of value increases for prestige.

- (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).

$$\text{Prestige} = 21.1422589 + 0.0031709 * (\text{income}) + 37.7812800 * (\text{professional}) - 0.0023257 * (\text{income} * \text{professional})$$

When income = 6000 dollars, and professional = 0

$$\text{Prestige} = 21.1422589 + 0.0031709 * (6000) + 37.7812800 * (0) - 0.0023257 * (6000 * 0)$$

$$\text{Prestige} = 40.168$$

When income = 6000 dollar, but professional = 1

$$\text{Prestige} = 21.1422589 + 0.0031709 * (6000) + 37.7812800 * (1) - 0.0023257 * (6000 * 1)$$

$$\text{Prestige} = 63.994$$

$$\text{Difference} = 63.994 - 40.168 = 23.826$$

Even though the person's income is not changed, when he switches from being a non-professional to be a professional, the value of prestige increases by 23.826 units

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: $R^2=0.094$, $N=131$

Set the coefficient in front of “Precinct assigned lawn signs” as beta1 and the coefficient in front of “Precinct adjacent to lawn signs” as beta2

$$\text{vote share} = 0.302 + 0.042 * (\text{assigned}) + 0.042 * (\text{adjacent})$$

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

$$\text{beta1} = 0.042 ; \text{se}(\text{beta1}) = 0.016$$

Hypothesis:

$$H_0: \text{beta1} = 0$$

$$H_a: \text{beta1} \text{ not equal to } 0$$

$$\text{Test Statistics} = (0.042 - 0) / 0.016 = 2.625$$

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.

$df = 131 - 3 = 128$

$p \text{ value} = 2 * pt(2.626, 128, \text{lower.tail} = F)$

$P \text{ value} = 0.0096$

Because P value is smaller than 0.05, there is enough evidence to reject the H_0 , so having these yard signs in a precinct affects 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 $\alpha = .05$).

$$\text{beta2} = 0.042; \text{se}(\text{beta2}) = 0.013$$

Hypothesis:

$$H_0: \text{beta2} = 0$$

$$H_a: \text{beta2} \text{ not equal to } 0$$

$$\text{Test Statistics} = (0.042 - 0) / 0.013 = 3.23$$

$$\text{df} = 131 - 3 = 128$$

$$p \text{ value} = 2 * \text{pt}(3.23, 128, \text{lower.tail} = F)$$

$$P \text{ value} = 0.0015$$

Because P value is smaller than 0.05, there is enough evidence to reject the H_0 , so being next to precincts with these yard signs affects vote share

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

$$\text{constant term} = 0.302$$

Interpretation: Regardless of whether having these yard signs in a precinct or being next to precincts with these yard signs, there is always going to be 30.2 percent of the vote that went to McAuliff's opponent 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?

The strength of the fit of a linear model is mostly evaluated using R Square, in this model, the value of R Square is only 0.094, which suggest that only 9.4 percent of the variability in the proportion of the vote share that went to McAuliff's opponent Ken Cuccinelli is explained by factors related to the yard signs. Thus, other vectors that are not included in the model must have more influence on the proportion of vote share than the yard signs. In another word, comparing with other vectors, yard sign's effect on the outcome variable is rather trivial.