QTM 200: Applied Regression Analysis- Problem Set 4

Farris Sabir

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:

```
1 install.packages("car")
2 library(car)
3 data(Prestige)
4 help(Prestige)
```

```
**Prestige of Canadian Occupations**
*Description*
The Prestige data frame has 102 rows and 6 columns. The observations are occupations.
*Usage*
Prestige
*Format*
This data frame contains the following columns:
```

education = Average education of occupational incumbents, years, in 1971.
income = Average income of incumbents, dollars, in 1971.
women = Percentage of incumbents who are women.
prestige = Pineo-Porter prestige score for occupation,
from a social survey conducted in the mid-1960s.
census = Canadian Census occupational code.
type = Type of occupation. A factor with levels (note: out of order): bc, Blue Collar;
prof, Professional, Managerial, and Technical; wc, White Collar.
Source
Canada (1971) Census of Canada. Vol. 3, Part 6. Statistics Canada [pp. 19-1{19-21].
Personal communication from B. Blishen, W. Carroll, and C. Moore, Departments of
Sociology, York University and University of Victoria.
References
Fox, J. (2016) Applied Regression Analysis and Generalized Linear Models,
Third Edition. Sage.
Fox, J. and Weisberg, S. (2019) An R Companion to Applied Regression,
Third Edition, Sage.

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 #Professionals recoded to 1, and blue and white collar workers coded as 0
2 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 prestigemodel1 <- lm(prestige ~ income + professional + income:
      professional, data = Prestige)
2 summary(prestigemodel1)
3 #Residuals:
4 #   Min       1Q    Median        3Q       Max
5 #-14.852   -5.332   -1.272    4.658    29.932
6 #Coefficients:
7 #                          Estimate  Std. Error  t value  Pr(>|t|)
8 #   (Intercept)          21.1422589   2.8044261    7.539  2.93e-11 ***
9 #   income                0.0031709   0.0004993    6.351  7.55e-09 ***
10 #   professional         37.7812800   4.2482744    8.893  4.14e-14 ***
11 #   income:professional  -0.0023257   0.0005675   -4.098  8.83e-05 ***
12 #   Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .
       0.1          1
13 #Residual standard error: 8.012 on 94 degrees of freedom
14 #(4 observations deleted due to missingness)
15 #Multiple R-squared:  0.7872, Adjusted R-squared:  0.7804
16 #F-statistic: 115.9 on 3 and 94 DF,  p-value: < 2.2e-16
```

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

$$prestige = 21.142 + 3.171 * 10^{-3} \times x_1 + 37.781 \times x_2 - 2.326 * 10^{-3} \times x_1 \times x_2$$

where x_1 = income and x_2 = professional with $x_2 = 0$ for blue or white collar and $x_2 = 1$ for professionals

(d) Interpret the coefficient for `income`.

This coefficient describes income's effect on prestige when the individual is blue or white collar.

For an individual who is blue or white collar, on average every 1 dollar increase in income leads to 0.0031709 units of increase in Pineo-Porter prestige score for occupation.

(e) Interpret the coefficient for `professional`.

This coefficient describes the effect of switching from blue or white collar to professional independent of income's effect.

For individuals who receives no income i.e. $0, on average those who are professionals have 37.7812800 units of Pineo-Porter prestige score more than those who are blue or white collar.

(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).

```
1  #Analytically deduce the answer
2  0.0031709 * 1000 - 0.0023257 * 1000 * 1
3  #Check the answer given hypothetical individual 1 vs. individual 2
4  i1 = 5000
5  i2 = 6000
6  p = 1
7  prestige1 <- 21.1422589 + 0.0031709 * i1 + 37.7812800 * p - 0.0023257 *
      i1 * p
8  prestige2 <- 21.1422589 + 0.0031709 * i2 + 37.7812800 * p - 0.0023257 *
      i2 * p
9  prestige2 - prestige1
```

The change in prestige associated with a $1000 increase in income given a professional occupation is a 0.8452 increase in units of 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).

```
1  #Analytically deduce the answer
2  37.7812800 * 1 - 0.0023257 * 6000 * 1
3  #Check the answer given hypothetical individual 1 vs. individual 2
4  i = 6000
5  p1 = 0
6  p2 = 1
7  prestige3 <- 21.1422589 + 0.0031709 * i + 37.7812800 * p1 - 0.0023257 * i
        * p1
8  prestige4 <- 21.1422589 + 0.0031709 * i + 37.7812800 * p2 - 0.0023257 * i
        * p2
9  prestige4 - prestige3
```

The change in prestige associated with changing from non-professional to professoinal given an income of $6000 is a 23.827 increase in units of Pineo-Porter 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 $\alpha = .05$).

```
1  #Hypotheses:
2  #Ho: beta_assigned = 0 & Ha: beta_assigned =/= 0
3  #Test statistic:
4  beta_assigned = 0.042
5  se_beta_assigned = 0.016
6  t = beta_assigned/se_beta_assigned
7  t #2.625
8  #P-value:
9  p = 2*pt(2.625, df = 131 - 3, lower.tail = F)
10 p #0.00972
```

Because the p-value of 0.00972 is below the alpha of 0.05, we reject the null hypothesis that having these yard signs in a precinct does not 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 $\alpha = .05$).

```
1  #Hypotheses:
2  #Ho: beta_adjacent = 0 & Ha: beta_adjacent =/= 0
3  #Test statistic:
4  beta_adjacent = 0.042
5  se_beta_adjacent = 0.013
6  t = beta_adjacent/se_beta_adjacent
7  t #3.231
8  #P-value:
9  p = 2*pt(3.231, df = 131 - 3, lower.tail = F)
10 p #0.00157
```

Because the p-value of 0.00157 is below the alpha of 0.05, we reject the null hypotheses that being next to precincts with these yard signs has no effect on vote share.

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

This means that for a precinct that is not assigned to have lawn signs and is not adjacent to a precinct with lawn signs, on average 30.2% of the vote 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?

Given an R^2 value of 0.094, the strength of fit of the linaer model is 0.094.

In other words, of the total variability in vote proportion for Cuccinelli, only 9.4% of it can be explained by yard signs.

The other 90.6% is explained by factors that are not modeled.