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`.)

```
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 × dummy interaction.)

```
presitge_regression <- lm(prestige ~ income+professional + income:
    professional, data = Prestige)
summary(presitge_regression)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-14.852  -5.332  -1.272   4.658  29.932

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         21.1422589  2.8044261   7.539 2.93e-11
income               0.0031709  0.0004993   6.351 7.55e-09
professional        37.7812800  4.2482744   8.893 4.14e-14
income:professional -0.0023257  0.0005675  -4.098 8.83e-05

(Intercept)         ***
income              ***
professional        ***
income:professional ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.012 on 94 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.7872,Adjusted R-squared:  0.7804
F-statistic: 115.9 on 3 and 94 DF,  p-value: < 2.2e-16
```

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

```
Y = 21.142259 + 0.003171(incomex1) + 37.781280(professionalx2)
-0.002326(incomex1)(professionalx2)
```

(d) Interpret the coefficient for `income`.

`For every 1 dollar increase in income, there is a 0.003171 increase in the Pineo-Porter prestige score for an occupation, when type of profession is held constant.`

(e) Interpret the coefficient for `professional`.

`Being a professional, instead of a white or blue collar worker, leads to a 37.781280 increase in the Pineo-Porter prestige score for an occuptaion when income is held 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).

```
Y_hat_income <- 21.142259 + 0.003171*(1000) + 37.781280*(1) -0.002326*
    (1000)*(1) # 59.76854
Y_hat_income_increase <- 21.142259 + 0.003171*(2000) + 37.781280*(1)
    -0.002326*(2000)*(1) #60.61354
Delta_y_hat_income <- Y_hat_income_increase-Y_hat_income #0.845
```

A \$1,000 increase in income leads to, on average, a 0.845 increase in the prestige score for professional occupations.

(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).

```
Y_hat_nonprof <- 21.142259 + 0.003171*(6000) + 37.781280*(0) -0.002326*
    (6000)*(0) #40.16826
Y_hat_prof <- 21.142259 + 0.003171*(6000) + 37.781280*(1) -0.002326*
    (6000)*(1) #63.99354
Delta_y_hat_prof <- Y_hat_prof-Y_hat_nonprof
```

A professional occupation with a \$6,000 income has, on average, a 23.82528 increse in the prestige score than a non-professional occuption with a \$6,000 income.

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 $\alpha = .05$).

```
Hypothesis: Ho: B1 = 0, Ha: B1 =/ 0
```

```r
test_stat1 <- (0.042 - 0)/0.016
n1 <- 30
p_val1 <- 2*pt(test_stat1, n1-1, lower.tail = FALSE)
p_val1 #0.01368397
```

```
With a significance level of 0.05 and a p-value of 0.01368397, I reject
the null hypothesis and conclude that having yard signs affects 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 $\alpha = .05$).

Hypothesis: Ho: B2 = 0, Ha: B2 =/ 0

```
test_stat2 <- (0.042- 0)/0.013
n2 <- 76
p_val2 <- 2*pt(test_stat2, n2-1, lower.tail = FALSE)
p_val2 #0.001834302
```

With a significance level of 0.05 and a p-value of 0.001834302, I reject the null hypothesis and conclude that living in a precint next to yard signs affects vote share.

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

In a precinct that does not have yard signs and is not adjacent to a precint with yard signs, 30.2% of the vote share, on average, goes to 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 R^2 value for this model is 0.094. This indicates that only 9.4% of the variability is explained by the model. This R^2 value demonstrates that there are other, more important, confounding variables which infuence vote share.