Problem Set 4

Applied Stats/Quant Methods 1

Due: November 26, 2021

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 GitHub in `.pdf` form.
- This problem set is due before class on Friday November 26, 2021. No late assignments will be accepted.
- Total available points for this homework is 80.

Question 1: 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`.)

```
install.packages("car")
library(car)
data(Prestige)
help(Prestige)

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.)

```
reg1 <- lm (data=Prestige, prestige ~ income + professional + income:
    professional )
summary(reg1)
```

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

Prestige $\hat{y} = 21.142$(intercept) + 0.003*income(x1) + 37.781*professional(x2) - 0.002(income(x1)*professional(x2))

(d) Interpret the coefficient for `income`.

One dollar increase on the income of incumbents increases on average 0.003 the prestige score while holding constant on the average the profession type and the interaction of professional and income.

(e) Interpret the coefficient for `professional`.

An individual with an occupation type classified as professional(=1) increases on average 37.781 of the prestige score compared to other non-professional types, holding income constant and the interaction of professional and income.

(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_1000income <- 21.142 + 0.003*1000 + 37.781 + (-0.002*1000)

y_2000income <- 21.142 + 0.003*2000 + 37.781 + (-0.002*2000)

y_3000income <- 21.142 + 0.003*3000 + 37.781 + (-0.002*3000)

y_3000income-y_2000income
y_2000income-y_1000income
```

According to the model in (b), an increase of $1000 in the income of an individual that has an occupation type professional has an marginal effect of 1 unit on the prestige score. Simulating the predicted $\hat{y}$ for incomes of 1000, 2000 and 3000 as the code above shows, and keeping constant the occupation as professional (=1), there is an increase of 1 associated in the prestige scale.

(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_nonprof <- 21.142 + 0.003*6000 + (37.781*0) + (-0.002*(6000*0))

y_prof <- 21.142 + 0.003*6000 + (37.781*1) + (-0.002*(6000*1))

y_prof-y_nonprof
```

According to the model in (b), an individual with an income of $6,000 that changes her occupation to a professional type, has an marginal effect on the prestige score of 25.781. As the code abose shows, the $\hat{y}$ with an individual on 6,000 income but in a non-professional occupation has on average the prestige score of 39.142, while if the same individual changes to a professional occupation, the prestige will likely to increase to 64.923.

Question 2: 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 from a linear regression to determine whether having these yard signs in a precinct affects vote share

```
# Hypotheses: H0 : B1 = 0 vs. Ha : B1 != 0

t_stat1 <- (0.042-0)/0.016 #Calculate t-statistic
df <- 131 -3 # calculate degrees of freedom = # of obs. - # of variables

p_value1 <- 2*pt(abs(t_stat1), df, lower.tail = F)

#0.009728528
```

We can reject the null hypothesis that B1 = 0 (at significance level alpha = .05). Precincts randomly assigned to have the sign against McAuliffe has a postive and significant effect on the proportion of the vote that went to McAuliff's opponent Ken Cuccinelli (p-value 0.0097).

[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.

```
# Hypotheses: H0 : B2 = 0 vs. Ha : B2 != 0

t_stat2 <- (0.042-0)/0.013 #Calculate t-statistic
p_value2 <- 2*pt(abs(t_stat2), df, lower.tail = F)

#0.00156946
```

We can reject the null hypothesis that B2 = 0 (at significance level alpha = .05). Precinct that was adjacent to a precinct in the treatment group has a postive and significant effect on the proportion of the vote that went to McAuliff's opponent Ken Cuccinelli (p-value 0.0016)

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

The expected proportion of vote that went to McAuliff's opponent is 0.302 when the Precinct assigned lawn signs and the Precinct adjacent to lawn signs are zero.

(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?

Approximately 9.4% of the variation in the Precinct assigned lawn signs and in the Precinct adjacent to lawn signs explains the variation of the proportion of the vote that went to McAuliff's opponent Ken Cuccinelli. This shows us that there are other factors that influence the vote share not included as covariants in the model.