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`.)

```
1
2  ######### Question 1: Economics #########
3 # We would like to study whether individuals with higher levels of income
      have more
4 # prestigious jobs. Moreover, we would like to study whether
      professionals have more
5 # prestige jobs than blue and white collar workers.
6
7 # (a) Create a new variable 'professional' by recoding the variable 'type
      '
8 # so that 'professionals' are coded as 1, and blue and white color
      workers
9 # are coded as 0 (Hint: ifelse)
10
11 # This calls for a dummy variable
12 # A dummy variable is a type of variable that we create in
13 # regression analysis so that we can represent a categorical variable
14 # as a numerical variable that takes on one of two values: zero or one.
15 # Create the dummy variable with ifelse() function
16
17
18
19 # create new variable 'professional'
20 # recode the variable 'type'
21 # prof to "1"
22 # bc to "0" or blue collar
23 # wc to "0" or white collar
24
25 ## referenced https://www.statology.org/dummy-variables-in-r/
26 # ifelse() function to define dummy varaiables
27
28 #Prestige$professional <- mutate(Prestige, prof = type)
29
30 Prestige$professional <- ifelse(Prestige$type == 'prof', 1, 0)
31 str(Prestige)
32
33
```

(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
2   # (b) Run a linear (lm) model with 'prestige' as an outcome and 'income
      ', 'professional'
3 # and the interaction of the two as predictors (Note: this is a continous
      X dummy
4 # interaction)
5
6 # prestige depends on income
7 prof_rg <- lm(prestige ~ income + professional + income * professional,
      data = Prestige)
8 summary(prof_rg)
9
10 ##
11 #Coefficients:
12 #  Estimate Std. Error t value Pr(>|t|)
13 # (Intercept)         21.1422589  2.8044261   7.539 2.93e-11 ***
14 # income               0.0031709  0.0004993   6.351 7.55e-09 ***
15 # professional        37.7812800  4.2482744   8.893 4.14e-14 ***
16 # income:professional -0.0023257  0.0005675  -4.098 8.83e-05 ***
17
18
19
```

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

```
1
2   # (c) Write the prediction equation based in the result
3
4 # intercept
5 # y (hat)      = 21.1422589 + 37.7812800D +  0.0031709x - 0.0023257xD + e
6
7
8
```

(d) Interpret the coefficient for `income`.

```
1
2
3   # (d) Interpret the coefficient for income variable
4
5 # income coeInfficient = 0.0031709
6 # and additional dollar for income will mean there is a 0.0031709
      increase
7 # in the level of prestige
8
9
10
```

(e) Interpret the coefficient for `professional`.

```
1
2   # (e) Interpret the coefficient for professional
3 # similiarly the professional coefficient is 37.7812800 relates to the
      level of
4 # prestige associated with a professional worker or a non blue and white
      collar
5 # professional
6
7
```

(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
2   # (f) What is the effect of a $1,000 increase in income on prestige
      score for
3 # professional occupation? In other words, we are interested in the
      marginal
4 # effect of income when the variable 'professional' takes the value of 1.
5 # Calculate the change of    associated with a $1,000 increase in income
6 # based on your answer for (c).
7
8 # for additional/increase of 1000 in income results in
9 # 0.0031709 x 1000 -0.0023257 x 1000 = 0.8452
10 # professional increase by 1000
11 # profession:income increase by 1000
12 # prestige points now show an increase of 0.8452 for an additional
      increase of $1,000
13
```

(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
2   # (g) What is the effect of changing one's occupation from non-
      professional
3 # to professional when her income if $6,000? We are interested in the
      marginal effect
4 # of professional jobs when the variable 'income' takes the value of $
      6,000.
5 # Calulate the change of y (hat) based on your answer for (c).
6
7 # more or less the same for (f) .. but non-professional
8 # income for professional + income professional
9 # income takes the value of 6000
10 # 37.7812800 -0.0023257 x 6000 = 23.82708
11
12
```

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 (e.g., conduct a hypothesis test with $\alpha = .05$).

```
1
2   # (a) Use the results from a linear regression to determine whether
      having
3 # these yard signs in a precinct affects vote share (e.g., conduct a
      hypothesis
4 # test with    = .05 which is the significance level in hypothesis test)
5
6 # hypothesis test 0.05
7 # Null hypothesis H0
8 # Alternative hypothesis Ha is   1
9 #        (hat) is the estimator of
10
```

[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.

```
11
12 # t-test is testing for a non-linear relationship
13 # t =    j    / se    j     where is the estimated beta and se is
       standard error
14 #student t-distribution
15
16 # gives 0.042 / (0.016) = 2.625
17
18 # N = 131
19 pt(2.625, df = 131)
20
21
```

(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 $\alpha = .05$).

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

(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?