Problem Set 4

Applied Stats/Quant Methods 1 - Ariana Alves Antunes

Due: December 4, 2022

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.
- This problem set is due before 23:59 on Sunday December 4, 2022. No late assignments will be accepted.

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`).

```
#View data set to explore the variables names
View(Prestige)
head(Prestige$type)
#variable type so that professionals are coded as 1, and blue and white collar workers are coded as 0.

Prestige$professionals <- as.factor(ifelse(Prestige$type=="prof", 1, ifelse(Prestige$type=="bc" | Prestige$type=="wc", 0, NA)))
```

(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.)

```
md1 <- lm(prestige ~ as.factor(income) + as.factor(professionals),
          data = Prestige,
          na.action = na.omit)
summary(md1)

md2 <- lm(prestige ~ income*professionals, data = Prestige, na.action = na.omit)
summary(md2)

stargazer::stargazer(md2, type = "latex",
                        title = "Summary table for prestige as an outcome and income, professional, and the interaction of the two as predictors")
```

Table 1: Summary table for prestige as an outcome and income, professional, and the interaction of the two as predictors

	Dependent variable:
	prestige
income	0.003^{***}
	(0.0005)
professionals1	37.781^{***}
	(4.248)
income:professionals1	-0.002^{***}
	(0.001)
Constant	21.142^{***}
	(2.804)
Observations	98
R^2	0.787
Adjusted R^2	0.780
Residual Std. Error	8.012 (df = 94)
F Statistic	115.878^{***} (df = 3; 94)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

E(y) = 21.14 + 37.78 - 0.002

(d) Interpret the coefficient for `income`. 0.003 increase in income for every professional 37.78

(e) Interpret the coefficient for `professional`.

Increase of 37.78 for professionals every 0.003 change in their 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).

$E(y) = \alpha + \beta 1x1 + \beta 2x2$

E(y) = 21.14 + 37.78 x 1,000 - 0.002 x 1

(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).

$E(y) = \alpha + \beta 1x1 + \beta 2x2 + \beta 3x3$

E(y) = 21.14 + 0.03 x 6000 + 37.78 x 0 - 0.002 x 6000 x 0

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 $\alpha = .05$).

```
# Exercise A
#Hypothesis test for precinct assigned
#H0 : B2 = 0.042 vs. Ha : B2 =! 0.042

#Test statistics
#t = B2 / se
0.042/0.016 = 2.625

#p-value
df = n - 3
131 - 3 = 128

```

[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.

```
p_values <- 2*pt(abs(2.625) , 128, lower.tail = F)
View(p_values)

# p-value = 0.0097, H0 is not reject since p < a = 0.05
```

(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 test for precinct adjacent
#H0 :B2 = 0.042 vs. Ha :B2 =! 0.042

#Test statistics
#t = B2 / se
0.042/0.013 = 3.230

#p-value
df = n - 3
131 - 3 = 128

p_values2 <- 2*pt(abs(3.230) , 128, lower.tail = F)
View(p_values2)

# p-value = 0.0015, H0 is not reject since p < a = 0.05
```

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

The constant is 0.302 and states the proportion of the vote that went to McAuliff's opponent Ken Cuccinelli, which could be interpreted as the impact of lawn signs on vote share for exposion to the signs.

(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 model fit is R2 = 0.094, which seems to have a small value, and high significant when analyzing the importance of yard signs compered with the model and not modeled.