Problem Set 4

Conor Twomey
Student No. 22996168
Applied Stats/Quant Methods 1

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`).

```
#1a create new variable
Prestige <- Prestige %>%
  mutate(professional = ifelse(
    type == "prof", "1",
    ifelse(type == "bc", "0",
           ifelse(type == "wc", "0",
                  ifelse(is.na(type), NA, "Other")))))
```

(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.)

```
lm1 <- lm(prestige ~ income + professional + income:professional, data =
    Prestige)
lm1
```

Table 1:

	Dependent variable:
	prestige
income	0.003^{***}
	(0.0005)
professional1	37.781^{***}
	(4.248)
income:professional1	-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.

$$y = \beta_0 + \beta_1 x + \beta_2 d + \beta_0 3xd$$
$$y = 21.142 + .003x + 37.781d + -.002xd$$

(d) Interpret the coefficient for `income`.

For a one dollar increase, controlling for the type of work and the interaction effect between income and type of work, there is a .003171 increase in prestige.

(e) Interpret the coefficient for `professional`.

Being a white collar worker, controlling for the type of work and the interaction effect between income and type of work, is associated with a 37.781 increase in prestige.

(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).

d = 1 / yhat = (b0 + b2) + (b1 + b3)x

yhat = 21.142 + .003x + 37.781d + -.002xd

yhat = 21.142 + .003x + 37.781 (1) + -.002x(1)

yhat = 21.142 + .003x + 37.781 - .002x

yhat = (21.142 + 37.781) + (.003x - .002x)

yhat = 58.923 + .001x

yhat = 58.923 + .001(1000)

yhat = 58.923 + 1

yhat = 59.923

(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). when d = 0

yhat = b0 + b1x + b2d + b3xd

yhat = 21.142 + .003x + 37.781d + -.002xd

yhat = 21.142 + .003x + 37.781(0) +- .002x(0)

yhat = 21.142 + .003x

yhat = 21.142 + .003(6000)

yhat = 21.142 + 18

yhat = 39.142

when d = 1

yhat = 58.923 + .001(6000)

yhat = 58.923 + 6

yhat = 64.923

64.923 - 39.142

25.781

The marginal effect of changing from non-professional to professional when income is $6,000$ is 25.781 increase in prestige.

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 $\alpha = .05$).

Step 1 :

I am assuming the distribution is normal.

Step 2 :

The null hypothesis is vote share is not effected by lawn signs.

The alternative hypothesis is that vote share is effected by lawn signs.

[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.

Step 3:

Calculate t-score

```
# t.score = ybar - mu/ se
ts <- (.042-0)/.016
ts
#2.625
```

Step 4:

Calculate p-value for two-tailed test

```
# p_value <- 2 x pt(ts, df=length(y)-parameters, lower.tail=FALSE)
pvalue <- 2*(pt(ts, 131-3, lower.tail = F))
pvalue
#0.00972002
```

Step 5:

Conclusion

Because the p value is less than alpha (.05), we can reject the null hypothesis.

(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 $\alpha = .05$).

Step 1 :

I am assuming the distribution is normal.

Step 2 :

The null hypothesis is vote share is not effected by lawn signs.

The alternative hypothesis is that vote share is effected by lawn signs.

Step 3:

Calculate t-score

```
# t.score = ybar - mu/ se
ts1 <- (.042-0)/.013
ts1
#3.230769
```

Step 4:

Calculate p-value for two-tailed test

```
# p_value <- 2 x pt(ts, df=length(y)-parameters, lower.tail=FALSE)
pvalue1<- 2*(pt(ts1, 131-3, lower.tail = F))
pvalue1
#0.00156946
```

Step 5:

Conclusion

Because the p value is less than alpha (.05), we can reject the null hypothesis.

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

When the all other covariates are zero, the constant represents the predicted value for the outcome variable. So in this case, this is the predicted proportion of the vote share that went to McAuliff's opponent Ken Cuccinelli if there were no lawn signs at all.

(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 has an Rsquared value of .092 which is very low. This means that the model does not account for much of the variance which, unsurprisingly, indicates that other factors may be at play for the swing to Cuccinelli other than one's proximity to a lawn sign.