Problem Set 4

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`).

```
professional <- ifelse(Prestige$type == "prof",
                       c(1), c(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.)

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

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

```
Y = 21.1422589 + 0.0031709(X1) + 37.7812800(X2) -0.0023257(X1*X2)
Where Y = prestige, X1 = Income, X2 = Professional
```

(d) Interpret the coefficient for `income`.

For a non-professional, every one unit
increase in income is associated with an
increase in prestige by 0.0031709 units.

For a professional, the association between income
and prestige is less significant, every additional unit of income
increases prestige by 0.0023257 less units than for other workers.

(e) Interpret the coefficient for `professional`.

Controlling for income, being a professional is
associated with a value for prestige 37.7812800 units
higher than for a blue-collar or white-collar worker.

(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).

```
Increase in ŷ = +0.0031709(1000) -0.0023257(1000)(1)
Increase in ŷ = +3.1709 -2.3257
Increase in ŷ = +0.8452
```

(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).

```
Change in ŷ = + 37.7812800  - 0.0023257(6000)
Change in ŷ = + 37.7812800 - 13.9542
Change in ŷ = +23.82708
```

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 $\alpha = .05$).

```
First I state my null and alternative hypotheses.

H0: there is no association between precincts having lawn signs and vote share.
i.e. the average vote share in precints with lawn signs = average vote share
in precints without lawn signs.

H1: there is some positive association between precints having lawn signs and
vote share. i.e. average vote share in precints with lawn signs > average vote
in precints without 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.

```
Next I conduct a t-test.
Difference in means = 0.04200
95\% confidence interval: 0.00146 to 0.08254
t = 2.0777
df = 53
P-score = 0.04260
Because P < 0.05, I can reject the null hypothesis, that lawn signs and vote share
are not associated with one another.
```

(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 $\alpha = .05$).

First I state my null and alternative hypotheses.

H0: there is no association between precincts being next to one with lawn signs and vote share. i.e. the average vote share in precints next to those with lawn signs = average vote share in precints not next to those with lawn signs.

H1: there is some positive association between precints being next to those with lawn signs and vote share. i.e. average vote share in precints next to those with lawn signs > average vote in precints not next to those with lawn signs.

Next I conduct a t-test.
Difference in means = 0.04200
95\% confidence interval: -0.00480 to 0.08880
t = 1.7808
df = 99
P-score = 0.07801
Because P > 0.05, I cannot reject the null hypothesis, that being next to a precint with lawn signs and vote share are not associated with one another.

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

```
In precints which neither had lawn signs, nor were adjacent to precints
with lawn signs, Ken Cuccinelli would be expected to get 30.2\% of the vote.
```

(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 suggests that 0.94\% of variance in the dependent
variable (vote share for Ken Cuccinelli) is explained by lawn signs
in the precint or in a neighboring precint. This suggests that the
 vast majority of variance is explained by other factors which are
not included in the model.
```