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`).

Variable = "pros"

```
Prestige$pros <- ifelse(Prestige$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.)

Table 1:

	Dependent variable:
	prestige
income	0.003^{***}
	(0.0005)
pro	37.781^{***}
	(4.248)
pro	-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 = b0 + b1 + b2 + b3
Prestige = 21.14 + 0.003*income + 37.78*pro + 0.002*income:pro

(d) Interpret the coefficient for `income`.

When controlling for the other variables in the model, workers experience a 0.003 increase in prestige, as measured by the researchers, per dollar increase in their income.

(e) Interpret the coefficient for `professional`.

Moving to the professional category would see a 37.78 increase in prestige points, when controlling for income and it's interaction with professional status.

(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).

A \$1,000 increase in salary increases a professionals prestige points by 1

```
21.14 + 37.78
58.98 + 0.003*1000
```

(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).

At an income of \$6,000, a non-professional moving to a professional occupation would gain 37.84 prestige points

```
58.98 + 0.003*6000
21.14 + 0.003*6000
76.98 - 39.14
```

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 $\alpha = .05$).

P-value .009 is less than alpha .05 so we can reject the null hypothesis that there is no relationship between lawn signs in the precinct and Cuccinelli's vote share

```
T.stat <- .042 / .016
df <- 131 - 2
p.val <- 2*pt(q = T.stat, df = df, lower.tail = F)
```

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

P-value is .001 is less than alpha .05 so we can reject the null hypothesis that there is no relationship between lawns signs in the adjacent precinct and Cuccinelli's vote share

```
T.stat.deux <- 0.042/  0.013
df.deux <- 76 - 2
p.val.deux <- 2*pt(q = T.stat.deux, df = df, lower.tail = F)
```

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

It is the value of y within the equation if the explanatory variables where equal to 0. In this case, it suggests Cuccinelli's vote share would be .302 without lawns signs either the precinct or those adjacent. This should probably be interpreted as a statistical artefact however.

(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?

$R^2 = 0.094$ suggests the model explains 9.4% of the variation in Cuccinelli's vote share. This suggests the model accounts for a non-trivial of variation in vote share but the overwhelming majority of variation is unexplained the EVs in question.