Problem Set 4

Applied Stats/Quant Methods 1

Due: December 3, 2023

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 3, 2023. 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`).

   [

```r
professional <- 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: Linear Model of Prestige with Income and Professional

	Dependent variable:
	prestige
income	0.003***
	(0.0005)
professional	37.781***
	(4.248)
income:professional	−0.002***
	(0.001)
Constant	21.142***
	(2.804)
Observations	98
R²	0.787
Adjusted R²	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.

Prestige = 21.142 + 0.003 x Income + 37.781 x Professional - 0.002 (Income x Professional)

(d) Interpret the coefficient for `income`.

A one unit increase in income is associated with a 0.003 increase in prestige score points, with the condition that all other variables remain constant

(e) Interpret the coefficient for `professional`.

A one unit increase in profressional is associated with a 37.781 in prestige score points, with the condition that all other variables remain constant

(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  Fnon_yHat1000 <-21.142+(0.003*1000)+(37.781*0)+(-0.002*(1000*0))
3  24.142
4
5  Fprof_yHat1000 <- 21.142+(0.003*1000)+(37.781*1)+(-0.002*(1000*1))
6  59.923
7
8  ME_1000 <- (0.003+(-0.002))*1000
9  1
```

We see the marginal effect of income as 1 so we would see that yHat increase by 1 score point of presidge with a 1000dollar income increase as the professional variable is catagorical which makes the interaction linear.

(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  Fnon_yHat6000 <- 21.142+(0.003*6000)+(37.781*0)+(-0.002*(1000*0))
2  39.142
3
4  Fprof_yHat6000 <- 21.142+(0.003*6000)+(37.781*1)+(-0.002*(1000*1))
5  74.923
6
7  ME_6000 <-(0.003+(-0.002))*6000
8  6
```

] We see the marginal effect of income as 6 so we would see that yHat increase by 6 score points of presidge with a 6000dollar income increase.

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  t_inYard <- 0.042/0.016
2  2.625
3
4  p_inYard <- pt(2.625, 29, lower.tail = FALSE)
5  0.007
```

] Our H0 is that having signs assigned onto precinct lawns had no affect on the votes given to Ken Cuccinelli. As we calculate the t-statistic and p-value we see that the p-value is smaller than our confidence interval, thus rejecting the H0. Accepting our hypothesis that having signs in precinct yards did have an affect on the vote share.

---

[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.

8

(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 $\alpha = .05$).

[

```
1  t_AdjYard <- 0.042/0.013
2  3.230
3  p_AdjYard <- pt(3.230, 75, lower.tail = FALSE)
4  0.001
```

] Our H0 is that having signs next to precincts lawns had no affect on the votes given to Ken Cuccinelli As we calculate the t-statistic and p-value we see that the p-value is smaller than our confidence interval, thus rejecting the H0. Accepting our hypothesis that having signs adjacent to precinct yards did have an affect on the vote share.

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

As assigned lawn signs increases by one unit (adjacent or otherwise), votes for the opponent increases by aproximately 0.042, as long as all other variables are held constant. The intercept/constant of 0.302 gives us the estimated number of votes for the opponent when the number of precincts assigned lawn signs is zero.

(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?

We can assume with the model fit that yard signs can be taken in as an important factor in the outcomes of these elections. While other variables that we are not taking into account (omitted variable bias)ould hold a differnt impact of said elections that could change the estimated relationship.