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)
```

```
1 #read in data
2 install.packages("car")
3 library(car)
4 data(Prestige)
5 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`).

```
1 #a. Delete NA of type
2 Prestige <- na.omit(Prestige)
3 #Creat a new dummy variable by coding the variable type
4 Prestige$professional <- ifelse(Prestige$type == "prof",1,0)
```

Di=0 if the collar are blue and write, it's non-professional
Di=1 if the collar is prof, it's professional

(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.)

```
1 #b. Creat a dummy for progression
2 #run a linear model with prestige as an outcome
3 model_b <- lm(prestige~income*professional,data = Prestige)
4 summary(model_b)
```

Table 1: Outcome variable is `prestige` and the explanatory variable are `income`,`professional`and `income*professional`.

	Model 1
(Intercept)	21.1423***
	(2.8044)
income	0.0032***
	(0.0005)
professional	37.7813***
	(4.2483)
income*professional	-0.0023***
	(0.0006)
R^2	0.7872
Adj. R^2	0.7804
Num. obs.	98

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

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

The prediction quations are:

$$prestige = 21.1423 + 0.0032 \times income \quad +37.7813 \times professional - 0.0023 \times (income \times prof$$
$$Di = 0 : prestige = 21.1423 + 0.0032 \times income$$
$$Di = 1 : prestige = 58.9236 + 0.0009 \times income$$

(d) Interpret the coefficient for `income`. The value of coefficient for income is 0.0032 ,it's the slope associated with prestige when controlling for professional. Controlling for professional, a unit increase in prestige is associated with 0.0032 increase in prestige.

(e) Interpret the coefficient for `professional`. The value of coefficient for professional is 37.7843 ,it's the effect associated with prestige when controlling for income.
Controlling for income, the professional group, on average,37.7843 greater prestige than non-professional.
Controlling for income, the non-professional group, on average,37.7843 less prestige than professional.

(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). The interaction term

measures how the effect of income on prestige differ for professional compared non-professional. In this case, the coefficient for the interaction term is -0.0023. Therefor, for professional, the effect of income on prestige is reduced by 0.0023 compared to non-professional.
In the equation of Di, it shows us that for a $1000 increase in income, the prestige score for professional is expected to increase by 0.0009.

(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).

$$\text{prestige} = 21.1423 + 0.0032 \times 6000 + 37.7813 \times 1 - 0.0023 \times (6000 \times 1) = 64.2236$$

When we the variable income takes the value of 6000, the marginal effect of changing one's occupations from non-professional to professional is 66.2236.

Question 2: Political Science

Researchers are interested in learning the effect of all of those yard signs on voting prefer-
ences.[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  #a.Ho:There is no effect of yard signs in a precinct on vote share
2  #  H1:There is effect of yard signs in a precinct on vote share
3  #Calculate t-statistics and p-value
4  TS_1 <- 0.042/0.016
5
6  #df_1=n-k-1=131-2-1=128
7  df_1 <-128
8  p_value1 <- 2*pt(abs(TS_1),df_1,lower.tail = F)
9  cat("The p-value of hypothsis is:",round(p_value1,4))
```

The p-value of hypothesis is: 0.0097
In this case, we can reject H0 because 0.0097<0.05, the p-value suggests that the

---

[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 experi-
ments." Electoral Studies 41: 143-150.

observed effect of yard sign in a precinct on vote share is statistical significant at the 0.05 level. Therefore, we have evident to suggest that having these yard signs in a precinct affects vote share.

(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  #Ho:There is no effect of being next to precints with these yard signs on
       vote share
2  #H1:There is effect of being next to precints with these yard signs on
       vote share
3  TS_2 <- 0.042/0.013
4
5  #df_1=n-k-1=131-2-1=128
6  df_2 <-128
7  p_value2 <- 2*pt(abs(TS_2),df_2,lower.tail = F)
8  cat("The p-value of hypothsis is:",round(p_value2,4))
```

The p-value of hypothesis is: 0.0016
In this case, we can reject H0 because 0.0016<0.05, the p-value suggests that the observed effect of being next to precincts with these yard signs on vote share is statistical significant at the 0.05 level. Therefore, we have evident to suggest that being next to precincts with these yard signs affects vote share.

(c) Interpret the coefficient for the constant term substantively.
The constant term is 0.302, it is the intercept of the regression line, representing the estimated vote shore when the value of others variables is 0.In this case, the estimated value of proportion of votes going to McAuliffe's opponent is 0.032 without the any influence from lawn sign.

(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?
This indicates that the model explains about 9.4% of the variability in the proportion of votes. While statistically significant, it suggests that there are other factors not included in the model that also influence vote share.

The value of (R^2) tells us that the model, including yard signs and adjacent, explains a small portion of the variation in vote share. Other factors not considered in the model might also impact voting preferences.