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
#clean data, deleting imcomplete observation
df_na <- df[complete.cases(df[,c("income", "prestige", "type")]),]
# Create a new variable professional by recoding the variable type
# professionals are coded as 1, and blue
# and white collar workers are coded as 0
df_na$professional <- ifelse(df_na$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.)

```r
pre_inc_pro<-lm(prestige~income+professional+income:professional,data=df_
    na)
summary(pre_inc_pro)
```

Figure 1: summary of regression

```
Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)        21.1422589  2.8044261   7.539 2.93e-11 ***
income              0.0031709  0.0004993   6.351 7.55e-09 ***
professional       37.7812800  4.2482744   8.893 4.14e-14 ***
income:professional -0.0023257  0.0005675  -4.098 8.83e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.012 on 94 degrees of freedom
Multiple R-squared:  0.7872,    Adjusted R-squared:  0.7804
F-statistic: 115.9 on 3 and 94 DF,  p-value: < 2.2e-16
```

There is a positive and statistically reliable relationship between the prestige and income in different professional, such that a one unit increase in the income is associated with an average increase of 0.032 in the prestige of prefessionals and that a one unit increase in the income is associated with an average increase of 0.0009 in the prestige of non-professionals.

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

- Based on regression analysis, we can obtain the following regression equation:

  prestige=21.1423+0.0032*income+37.7812*prefessional-0.0023*income*prefessional

  When professional is 1(professionals), the prediction equation would be:
  prestige=58.9244+0.0009*income

  When professional is 0(blue and white collar workers), the prediction equation would be:
  prestige=21.1423+0.0032*income

(d) Interpret the coefficient for `income`.

The coefficient for income is 0.0032. It means when occupation is professionals each 1 dollar increase in income would lead to an average increase of 0.0032 in the prestige.

(e) Interpret the coefficient for `professional`.

The coefficient for professional is 37.7812. It means, compared to blue and white collar workers, the prestige of professionals would be 37.7812 higher on average, when income remains unchanged.

(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).

```
#When professional=1, in general the effect of a $1,000
#increase in income on prestige score would be:
x1<-runif(1, min = 0)
Y<-(58.9244+0.0009*(1000+x1))-(58.9244+0.0009*x1)#Y=0.9
```

According to the above result, a 1,000 dollar increase in incomewould lead to an average increase of 0.9 on prestige score for professional. In other words, when the prefessional equals to 1, the each unit increase in income would cause an average increase of 0.0009 on prestige score.

(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).

```r
#when her income is $6,000, in general the effect of
#changing occupations of somebody from non-professional to
#professional would be:
Y<-(21.1423+0.0032*6000)-(58.9244+0.0009*6000)
#Y=-23.9821
```

According to above result, when changing one's occupations from non-professional to professional with her income is 6,000 dollar, her prestige score would decrease 23.9821 on average

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] 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.

```r
# H0 : B_assigned= 0 vs. H1: B_assigned != 0
# get t_statistic and p_values
df_assigned<-131-2-1
TS_assigned <- ( 0.042 - 0 ) /0.016
p_value_assigned <- 2 * pt ( abs ( TS_assigned ) , df_assigned, lower.
    tail= F )
p_value_assigned
```

According to calculation, the p-value is 0.00972, which is lower than 0.05 so we make a conclusion that we can reject the Null hypothesis and accept the alternative hypothesis that the coefficient of the dummy variable in the question- Precinct assigned lawn signs - is is statistically differentiable from zero.

(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 $\alpha = .05$).

```
#H0 : B_adjacent = 0 vs. H1_adjacent : B!= 0
# get t_statistic and p_values
df_adjacent<-131-2-1
TS_adjacent <- ( 0.042 - 0 ) /0.013
p_value_adjacent <- 2 * pt ( abs ( TS_adjacent ) , df_adjacent , lower.
    tail= F )
p_value_adjacent
```

According to above calculation, the p-value is 0.001569, which is lower than 0.05 so we make a conclusion that we can reject the Null hypothesis and accept the alternative hypothesis that the coefficient of the dummy variable in the question– Precinct adjecent to lawn signs – is statistically differentiable from zero.

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

At precinct that is neither assigned lawns assign nor adjacent to lawn signs, the estimated value of the dependent variable–the mean of proportion of the vote that went to McAuliffs opponent Ken Cuccinelli–would equals to the coefficient for the constant,which is 0.302

(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 $R^2$of this regression is 0.094, which mean the regression only explain 0.094 variance of the dependent variable–the mean of proportion of the vote that went to McAuliffs opponent Ken Cuccinelli. We can make a conclusion from this R^2 that the yard sign only can explain a very smart portion of the variance of the dependent variable, so there must be others factors play a more important role on the variance of dependent variable.