Problem Set 4

Applied Stats/Quant Methods 1

Qin Guo 24338859

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 Monday November 18, 2024. 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
# Install and load the car package
install.packages("car")
library(car)
# Load the Prestige dataset
data(Prestige)

#a
# Create a new variable professional
# use the ifelse function to create a new variable professional based on the type variable in the Prestige dataset.
Prestige$professional <- ifelse(Prestige$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.)

```r
#b
# Fit the linear model
model <- lm(prestige ~ income + professional + income:professional, data = Prestige)
#fit a linear model with prestige as the outcome and income, professional, and their interaction as predictors.
summary(model)
```

```
Call:
lm(formula = prestige ~ income + professional + income:professional,
    data = Prestige)

Residuals:
    Min      1Q  Median      3Q     Max
-14.852  -5.332  -1.272   4.658  29.932

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
  (因为不存在，4个观察量被删除了)
Multiple R-squared:  0.7872,    Adjusted R-squared:  0.7804
F-statistic: 115.9 on 3 and 94 DF,  p-value: < 2.2e-16
```

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

$$Prestige = 21.142 + 0.003 * Income + 37.781 * Professional - 0.002 * Income * Professional \times \text{difflog}$$

(d) Interpret the coefficient for `income`.
When other variables are controlled, especially when the professional variable is 0, the average change in prestige (prestige score) for each unit increase in income is 0.For blue-collar and white-collar workers, the average change in prestige (prestige score) for each unit increase in income is 0.003 units.

(e) Interpret the coefficient for `professional`.
The coefficient of professional indicates the average change in prestige when changing from a non-professional(blue-collar or white-collar workers) to a professional, while controlling for other variables (such as income).The average prestige score of professional worker is about 37.781 units higher than an equally income 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).

```r
#f
#Given coefficients
income_coefficient <- 0.003
professional_coefficient <- 37.781
interaction_coefficient <- -0.002

# Calculate the effect of a $1,000 increase in income on prestige score
    for professionals
income_coefficient <- 0.003
interaction_coefficient <- -0.002
income_effect_professional <- (income_coefficient + interaction_
    coefficient) * 1000
income_effect_professional
```

```
> income_effect_professional <- (income_coefficient + interaction_coeffi
00
> income_effect_professional
[1] 1
>
```

According to the result, there is only a 1 prestige point difference for a 1000dollars increase in income for professional occupations.

(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).

```
# Calculate the effect of changing from non-professional to professional
    at an income of $6,000
professional_coefficient <- 37.781
income_level <- 6000
professional_effect <- professional_coefficient + (interaction_
    coefficient * income_level)
professional_effect
```

```
> professional_effect
[1] 25.781
```

According to the result, it shows that there would be a predicted 25.781 increase in prestige score from non-professional to professional when her income is 6000dollars.

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).

[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.

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 $\alpha = .05$).

```
#Question2
#a
#calculate t-statistic
coef_signs <- 0.042
se_signs <- 0.016
statistic_signs <- coef_signs / se_signs
#calculate P-value
#df = N - K - 1
df <- 131 - 2 - 1
p_value <- 2* pt(statistic_signs , df , lower.tail = FALSE )
#t=2.625  p=0.0097
```

Null Hypothesis: The presence of yard signs has no effect on the vote share.
H0:$\beta = 0$
Alternative Hypothesis: The presence of yard signs has an effect on the vote share.
Ha:$\beta \neq 0$

Conclusion: P-value of 0.0097 is less than 0.05, therefore, we reject the null hypothesis. So there is a significant effect of yard signs on the 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 $\alpha = .05$).

```
#b
#calculate t-statistic
coef_adjacent <- 0.042
se_adjacent <- 0.016
statistic_adjacent <- coef_adjacent / se_adjacent
#calculate P-value
#df = N - K - 1
df <- 131 - 2 - 1
p_value_adjacent <- 2* pt(statistic_adjacent , df , lower.tail = FALSE )
#t=3.231  p=0.00157
```

Null Hypothesis: being next to precincts with these yard signs does not effect on the vote share.
H0:$\beta = 0$
Alternative Hypothesis: being next to precincts with these yard signs effects on the vote share.
Ha:$\beta \neq 0$

Conclusion: P-value of 0.00157 is less than 0.05, therefore, we reject the null hypothesis. So there is a significant effect of being next to precincts with yard signs on the vote share.

(c) Interpret the coefficient for the constant term substantively.
The constant term coefficient of 0.302 signifies the average proportion of votes that Ken Cuccinelli, McAuliffe's opponent, would receive in precincts where no lawn signs were installed and which are not in proximity to any precincts with such signs.

(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?
An R-squared value of 0.094 indicates that the influence of yard signs accounted for only 9.4 % of the vote share. This suggests that while yard signs might have some impact, they are not strong predictors of voting behavior.