Problem Set 4

Applied Stats/Quant Methods 1 Jia Lyu-2337006

Due: November 18, 2024

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
# Create a new variable
Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0

# Check the first few rows to confirm
head(Prestige)
```

```
                    education income women prestige census
gov.administrators      13.11  12351 11.16     68.8   1113
general.managers        12.26  25879  4.02     69.1   1130
accountants             12.77   9271 15.70     63.4   1171
purchasing.officers     11.42   8865  9.11     56.8   1175
chemists                14.62   8403 11.68     73.5   2111
physicists              15.64  11030  5.13     77.6   2113
                    type professional
gov.administrators  prof            1
general.managers    prof            1
accountants             prof            1
purchasing.officers prof            1
chemists                    prof            1
physicists                  prof            1
```

(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
# Fit the linear model
prestige_model <- lm(prestige ~ income * professional, data = Prestige)

# View the summary of the model
summary(prestige_model)
```

```
Call:
lm(formula = prestige ~ 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
(4 observations deleted due to missingness)
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+0.003income+37.781professional-0.02incomeprofessional

(d) Interpret the coefficient for `income`.

The income coefficient in this regression model is 0.003. For non-professional workers (blue and white collar), when professional=0, each $1,000 increase in income is associated with an expected increase of 0.003 points in prestige.

(e) Interpret the coefficient for `professional`.

The coefficient for professional in this regression model is 37.781. This means that when professional=1, and income is 0, professionals are expected to have a job prestige score 37.781 points higher than non-professional workers (blue- and white-collar workers).

(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).

```
Prestige=21.1423+0.0032Income+37.7813Professional0.0023Incomepofessional

When Professional=1:
Prestige=(0.00320.0023)Income+58.9236
Prestige=0.0009Income+58.9236

For professionals, a $1,000 increase in revenue is expected to result in a
reputation score of 0.9 units.
```

(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).

```
Prestige=21.1423+0.0032*Income+37.7813*Professional0.0023*Income*Professional

When Professional=1:
prestige = 21.1423+0.0032*6000+37.7812-0.0023*6000 = 64.3235

when profession = 0,
prestige = 21.1423+0.0032*6000 = 40.3423

Difference Between the Two Prestige Values:
64.3235 - 40.3423 = 23.9812

Switching from non-professional to professional prestige score increases
the prestige score by about 23.98 points when earning $6,000
```

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 $\alpha = .05$).

```
Null Hypothesis ( H0 ): The presence of lawn signs does not affect the
vote share for Ken Cuccinelli (coefficient = 0).
Alternative Hypothesis ( H1): The presence of lawn signs does affect the
vote share for Ken Cuccinelli (coefficient  0).

Coefficient:  1=0.042
Standard error: 0.016

Test Statistic Calculation:
```

[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.

t=Coefficient/Standard Error =0.042/0.016 =2.625

With 131-2=129 degrees of freedom (since we have two predictors and a constant in the model), the critical value for a two-tailed test with =0.05 is approximately ±1.96.

Since 2.625 > 1.96,we reject the null hypothesis and conclude that having yard signs in a precinct does affect 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 $\alpha = .05$).

Null Hypothesis (H0): A candidate's age does not affect the voting share (coefficient = 0).
Alternative Hypothesis (H1): A candidate's age affects the voting share(coefficient 0).

Coefficient: 1=0.042
Standard error: 0.013

Test Statistic Calculation:
t=Coefficient/Standard Error =0.042/0.013 =3.231

With 131-2=129 degrees of freedom (since we have two predictors and a constant in the model), the critical value for a two-tailed test with =0.05 is approximately ±1.96.

Since 3.231 > 1.96,we reject the null hypothesis and conclude that being adjacent to a precinct with yard signs does affect vote share.

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

The coefficient for the constant term is 0.302, which represents the average proportion of the vote that went to McAuliff's opponent Ken Cuccinelli in precincts where neither the treatment nor adjacency to the treatment was present. This suggests that, on average, Cuccinelli received about 30.2% of the vote in these precincts.

(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-squared value for the regression model is 0.094, which indicates that the model explains only about 9.4% of the variability in vote share. This suggests that yard signs and adjacency to precincts with yard signs are relatively unimportant factors in determining vote share compared to other factors that are not included in the model. It is possible that other factors, such as candidate characteristics, campaign spending, or voter demographics, may have a larger impact on vote share.