Problem Set 4

Applied Stats/Quant Methods 1
Jianxiong Wu—23354731

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`).

```
#Question 1
#(a)
Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0)
#Prestige$professional <- as.factor(Prestige$professional)
#table(Prestige$professional)
head(Prestige)
```

(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.)

```
#(b)
Prestige_regression <- lm(prestige ~ income + professional + income:
    professional, data = Prestige)
summary(Prestige_regression)
library(stargazer)
stargazer(Prestige_regression)
```

Table 1:

	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
R^2	0.787
Adjusted R^2	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*income+37.781*professional-0.002*income*professional
```

(d) Interpret the coefficient for `income`.

```
The income coefficient for this regression model is 0.003, so when
professional = 0, each unit increase in income will increase the prestige
of blue- and white-collar workers by 0.003 units.
```

(e) Interpret the coefficient for `professional`.

```
The professional coefficient of this regression model is 37.781, so when
income is zero, the prestige scores of blue- and white-collar workers are
lower than the professionals' prestige score of 37.781 points.
```

(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+0.0032*income+37.7812*professional-0.0023*income*professional

So when profession = 1,
prestige = 21.1423+0.0032*income+37.7812-0.0023*income
= 58.9235 + 0.0009*income

Income has a coefficient of 0.009, so when income increases by $1,000,
the prestige score increases by 0.009 * 1000 = 0.9.
```

(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+37.7812*professional-0.0023*income*professional

when profession = 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

64.3235 - 40.3423 = 23.9812

At an income of $6,000, switching from non-professional to professional
increases the prestige score by about 23.98.
```

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 $\alpha = .05$).

```
coef_a <- 0.042
se_a <- 0.016
ts_a <- coef_a/se_a
#ts_q2=2.625
p_value_a <- 2 * pt(abs(ts_a), 131-2-1, lower.tail = FALSE)
#p_value=0.0097
```

```
H0: Lawn signs have no effect on vote share.
H1: Lawn signs have an effect on 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.

Since the p-value of 0.0097 is less than 0.05, the null hypothesis is rejected, so there is evidence that 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 $\alpha = .05$).

```r
coef_b <- 0.042
se_b <- 0.013
ts_b <- coef_b/se_b
#ts_q2=3.231
p_value_b <- 2 * pt(abs(ts_b), 131-2-1, lower.tail = FALSE)
#p_value=0.0015
```

H0: Being next to precincts with these yard signs does not affect vote share.
H1: Being next to precincts with these yard signs can affect vote share.

Since the p-value of 0.0015 is less than 0.05, the null hypothesis is rejected, so there is evidence that being next to precincts with these yard signs can affect vote share.

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

The constant term has a coefficient of 0.302, which represents a projected Ken Cuccinell vote of 30.2% in the absence of any yard signs and adjacent 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?

The value of R-squared in this question is 0.094, which means that only about 9.4% of Ken Cuccinell's vote share is due to precinct assigned lawn signs and precinct adjacent to lawn signs, which is a relatively low percentage, suggesting that lawn signs, although they have an impact on the vote share though This relatively low percentage indicates that although lawn signs have some effect on vote share, they are not the main factor affecting vote share and cannot fully explain the change in vote share; there are still some other variables that affect vote share.