Problem Set 4

Applied Stats/Quant Methods 1

Zexi Wang

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 ###
data(Prestige)
Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0)
Prestige$professional <- as.factor(Prestige$professional)
# It can be seen that there are 4 missing values in
# the "type", which makes their "professional" NA.
# If we wanna dropping them, here is the code
# Prestige_clean <- na.omit(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 ###
model_pres <- lm(prestige ~ income + professional + income:professional,
    data = Prestige)
summary(model_pres)
stargazer(model_pres, type = "text", out = "model_pres.txt")
```

```
================================================
Dependent variable:
---------------------------
prestige
------------------------------------------------
income                          0.003***
(0.0005)

professional1                   37.781***
(4.248)

income:professional1            -0.002***
(0.001)

Constant                        21.142***
(2.804)

------------------------------------------------
Observations                      98
R2                               0.787
Adjusted R2                      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 coefficient for income in the regression model indicates the expected change in the prestige score for each additional dollar earned.For blue-collar and white-collar workers, for each additional dollar of income, the prestige score is expected to increase by 0.003 on average.`

(e) Interpret the coefficient for `professional`.

`The coefficient for professional shows the expected change in prestige score when moving from a white-collar or blue-collar position to a professional role, with income at zero. A professional worker is expected to have a prestige score that is 37.781 points higher than an equally low-income white-collar or blue-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).

```
### f ###
increase_effect <- (coef(model_pres)["income"] + coef(model_pres)["income
    :professional"]) * 1000
```

`The marginal increase in prestige score is 0.8452`

(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).

```
### g ###
change_effect <- coef(model_pres)["income"] + coef(model_pres)["income:
    professional"] * 6000
```

`Changing one's occupation from a non-professional to a professional worker at an income of $6,000 results in an increase of 23.83 in the prestige score.`

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 $\alpha = .05$).

```
############ Question 2 ############

### a ###
coef_sign <- 0.042
std_error_sign <- 0.016
t_value_sign <- coef_sign / std_error_sign
p_value_sign <- 2 * pt(-abs(t_value_sign), df = 128)
p_value_sign < 0.05
# t = 2.625, p = 0.0097
```

```
H0:lawn yard signs do not affect vote share (effect = 0)
Ha:lawn yard signs do affect 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.

Given that the p-value of 0.0097 is less than the significance level of 0.05, we reject the null hypothesis, providing statistically significant evidence that lawn signs in a precinct do have an effect on 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
### b ###
coef_next <- 0.042
std_error_next <- 0.013
t_value_next <- coef_next / std_error_next
p_value_next <- 2 * pt(-abs(t_value_next), df = 128)
p_value_next < 0.05
# t =3.23, p = 0.0015
```

H0:being next to precincts with these yard signs does not affect vote share
Ha:being next to precincts with these yard signs affects vote share

Since the p-value of 0.0015 is less than the significance level of 0.05, we reject the null hypothesis, providing statistically significant evidence that precincts next to yard signs do have an effect on vote share.

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

The constant coefficient of 0.302 represents the baseline vote share in precincts without lawn signs and not adjacent to any precincts with lawn signs, indicating that the expected vote share for Ken Cuccinelli is 0.302 in the control group, assuming other factors remain constant.

(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 of 0.094 indicates that approximately 9.4% of the variance in vote share is accounted for by the presence of lawn signs in and adjacent to precincts. This implies that although lawn signs may have some influence, a considerable amount of the variance in vote share is likely due to other unaccounted factors.The model might be missing key variables, suggesting that lawn signs and nearby lawn signs alone may not adequately explain the variation in the proportion of votes.