

Problem Set 4

Applied Stats/Quant Methods 1

Due: December 4, 2022

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 4, 2022. 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**).

The following code was used to run the regression:

```
1 # Question 1 (a), creating new variable
2
3 professional <- c()
4
5
6 for (i in Prestige$type) {
7   if (is.na(i) == TRUE) {
8     professional <- append(professional, NA)
9   }
10  else if (i == "prof") {
11    professional <- append(professional, 1)
12  }
13  else {
14    professional <- append(professional, 0)
15  }
16 }
17
18
19 Prestige <- cbind(Prestige, professional)
```

- (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 \times dummy interaction.)

The following code was used to run the regression:

```
1 reg1 <- lm(prestige ~ income + professional + income*professional, data =
  Prestige)
```

From this, using Stargazer an output table was created (see Table 1)

Table 1: Regression

	<i>Dependent variable:</i>
	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)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

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

The full equation is shown below:

$$Y_i = 21.142 + 0.003X_i + 37.781D_i - 0.002X_iD_i$$

For the average prestige when the participant is a blue/white collar worker:

$$\mu_{Y|x} = 21.142 + 0.003X$$

For the average prestige when the participant is a professional:

$$\mu_{Y|x} = 58.923 + 0.001X$$

(d) Interpret the coefficient for **income**.

The income coefficient means that for blue and white collar workers, a 1 unit increase in prestige will lead to a \$0.003 increase in income

(e) Interpret the coefficient for **professional**.

The professional coefficient means that for professional workers, on average they score 37.781 higher on prestige compared to blue and white collar workers when income is held constant.

- (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).

For \$0 income:

$$Y = 21.142 + 0.003(0) + 37.781(1) - 0.002(0)(1) \\ Y = 58.923$$

For \$1,000 income:

$$Y = 21.142 + 0.003(1000) + 37.781(1) - 0.002(1000)(1) \\ Y = 59.923$$

Therefore, an increase in \$1,000 of income leads to an increase of score by 1 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).

For a professional worker at \$6,000 income:

$$Y = 21.142 + 0.003(6000) + 37.781(1) - 0.002(6000)(1) \\ Y = 64.923$$

For a blue collar/white collar worker at \$6,000 income

$$Y = 21.142 + 0.003(6000) + 37.781(0) - 0.002(6000)(0) \\ Y = 39.142$$

Changing from a non-professional to professional job both with \$6,000 income would lead to an increase in prestige score of 25.871

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 Virginia 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 McAuliffe’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 $\alpha = .05$).

First, we construct the hypotheses:

$$\begin{aligned}H_0: \beta_2 &= 0 \\H_1: \beta_2 &\neq 0\end{aligned}$$

We then calculate the test statistic:

$$T = \frac{\beta - 0}{se_{\beta}} = \frac{0.042 - 0}{0.013} = 2.625$$

¹Donald 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.

After which, we calculate the p -value, which was done in RStudio with the following code:

```
1 2*(pt(2.625, df=128, lower.tail = FALSE))
```

```
[1] 0.00972002
```

As $p < .05$, we can reject the null hypothesis that $\beta_1 = 0$, meaning that yard signs have a significant 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 $\alpha = .05$).

First, we construct the hypotheses:

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

We then calculate the test statistic:

$$T = \frac{\beta - 0}{se_{\beta}} = \frac{0.042 - 0}{0.013} = 3.231$$

After which, we calculate the p -value, which was done in RStudio with the following code:

```
1 2*(pt(3.231, df=128, lower.tail = FALSE))
```

```
[1] 0.00156829
```

As $p < .05$, we can reject the null hypothesis that $\beta_1 = 0$, meaning that being next to precincts with these yard signs affects vote share.

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

The constant term means that with no signs put up in any yards, meaning there would be no adjacent precincts to precincts with signs, the vote share for Ken Cuccinelli is at 30.2%

- (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?

While there is a non-zero effect of the two predictor variables on Ken Cuccinelli's vote share, the variance in the vote share is only explained by 9.4% of the variance in the predictor variables, $R^2 = 0.094$. Therefore, there may be other important factors not included in the model that can explain more of the variance in the dependent variable, which would lead to a better-fitting model. For example, perhaps voters' agreement with policies, candidate media presence etc. could be variables that might explain more of the variance in vote share.