

Problem Set 4

Applied Stats/Quant Methods 1

Due: December 3, 2023

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 3, 2023. 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**).

```

1 # Create a new variable
2 Prestige$professional <- ifelse(Prestige$type %in% "prof", 1,
3                                ifelse(Prestige$type %in% c("bc", "wc"),
4                                        0, NA))
5 head(Prestige)

```

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

```
1 # Run the linear model with prestige as the outcome and income,
  professional, and their interaction as predictors
2 model <- lm(prestige ~ income + professional + income*professional, data
  = Prestige)
3 summary(model)
4 stargazer(model, no.space = TRUE)
```

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

There is a positive and statistically reliable relationship between the prestige and income, such that a one unit increase in the logged difference in spending is associated with an average increase of 0.003 in the prestige.(0.3%)

There is a positive and statistically reliable relationship between the prestige and professional, such that a one unit increase in the logged difference in spending is associated with an average increase of 37.781 in the prestige.

There is a negative and statistically reliable relationship between the prestige and the interaction of the two,such that a one unit increase in the logged differencein spending is associated with an average decrease of 0.002 in the prestige.(-0.2%)

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

$$\hat{y} = \alpha + \hat{\beta}_0 \times \text{income} + \hat{\beta}_1 \times \text{professional} + \hat{\beta}_2 \times \text{income} \times \text{professional}$$
$$\text{prestige} = 21.142 + 0.003 \times \text{income} + 37.781 \times \text{professional} - 0.002 \times \text{income} \times \text{professional}$$

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

The coefficient for income is 0.003. This implies that, on average, for each one-unit increase in income, we expect the prestige score to increase by 0.3%, holding other variables constant.

Because of the coefficient for income less than 0.1, we can say that the coefficient for income is not equal to zero.

(e) Interpret the coefficient for **professional**. The coefficient for professional is 37.781.

This implies that, on average, for each one-unit increase in professional, we expect the prestige score to increase by 37.781, holding other variables constant.

Because of the coefficient for professional less than 0.1, we can say that the coefficient for professional is not equal to zero.

- (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).

$$prestige = 21.142 + 0.003 \times income + 37.781 \times professional - 0.002 \times income \times professional$$

$$\hat{y}_1 = 21.412 + 0.003 \times +37.781 \times 1 - 0.002 \times income(1)$$

$$\hat{y}_2 = 21.241 + 0.003 \times (income + 1000) + 37.781 - 0.002 \times (income + 1000)(2)$$

$$(2) - (1) = 1$$

so, the change in \hat{y} associated with a 1,000 increase in income is 1.

- (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).

$$prestige = 21.142 + 0.003 \times income + 37.781 \times professional - 0.002 \times income \times professional$$

when the value of the professional=1

$$\begin{aligned}\hat{y}_1 &= 21.412 + 0.003 \times 6000 + 37.781 \times 1 - 0.002 \times 6000 \times 1(1) \\ &= 65.193\end{aligned}$$

when the value of the professional=0

$$\begin{aligned}\hat{y}_1 &= 21.412 + 0.003 \times 6000 + 37.781 \times 0 - 0.002 \times 6000 \times 0(2) \\ &= 39.412\end{aligned}$$

$$(1)-(2)=25.871$$

When everyone's income is 6,000, the occupational prestige gap between professional and non-professional is 25.781.

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

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.

- (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 $\alpha = .05$). Hypothesis

test for the variable indicating assigned lawn signs:

Null hypothesis (H_0): $\beta_1 = 0$ (No effect of being adjacent to lawn signs on vote share).

Alternative hypothesis (H_1): $\beta_1 \neq 0$ (Being adjacent to lawn signs has an effect on vote share).

For the variable indicating assigned lawn signs (β_1):

$$t_{\beta_1} = \frac{0.042}{0.016} = 2.625$$

Calculate P_{value} in R:

```
1 pvalue_1 <- 1-pt(2.625, df=128)
2 print(pvalue_1)
```

[1] 0.00486001 We can clearly see that p-value (0.00486001) is far below the $\alpha = 0.05$, so we have sufficient evidence to reject H_0 :No effect of being adjacent to lawn signs on vote share, and we can accept H_1 .

- (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 $\alpha = .05$).
Hypothesis test for the variable indicating assigned lawn signs:

Null hypothesis (H_0): $\beta_2 = 0$ (No effect of being adjacent to lawn signs on vote share).
Alternative hypothesis (H_1): $\beta_2 \neq 0$ (Being adjacent to lawn signs has an effect on vote share).

For the variable indicating assigned lawn signs (β_2):

$$t_{\beta_1} = \frac{0.042}{0.013} = 3.231$$

Calculate P_{value} in R:

```
1 pvalue_2 <- 1-pt(3.231, df=128)
2 print(pvalue_2)
```

[1] 0.0007841451 We can clearly see that p-value (0.0007841451) is far below the $\alpha = 0.05$, so we have sufficient evidence to reject H_0 : No effect of being adjacent to lawn signs on vote share, and we can accept H_1 .

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

Represents the expected mean of the dependent variable when all explanatory variables are zero. It provides a baseline level of 0.302 when the explanatory variables all take zero.

- (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?

Based on the model output table the $R^2 = 0.094$, indicating that the independent variables included in the model can account for approximately 9.4% of the variability observed in the dependent variable, and it is likely that factors other than yard signs and proximity (as shown in the model) play an important role in shaping voting preferences. Therefore, other variables can be added later to improve the fit of the model.