

Problem Set 4

Applied Stats/Quant Methods 1

Due: November 26, 2021

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 in **.pdf** form.
- This problem set is due before class on Friday November 26, 2021. No late assignments will be accepted.
- Total available points for this homework is 80.

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**.)

- `dataprestige` | Prestige

Rename the column 'type' as 'professionals'

- `names(dataprestige)[names(dataprestige) == 'type']` | 'professionals'

Write if/else loop to code professionals under this column as 1, and blue and white collar workers as 0.

- `dataprestigeprofessionals <- ifelse(dataprestigeprofessionals`

- `dataprestige`

- (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.)

- `prestigeregression` | `lm(prestige ~ professionals+income, data=dataprestige)`

- `summary(prestigeregression)`

- `plot(prestigeregression)`

- `abline(prestigeregression)`

Intercept = 27.863, prof slope = 17.761, income slope = 0.002

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

$$\text{Equation} = y = 27.863 + 17.761 \cdot x_1 + 0.002 \cdot x_2$$

X1 and X2 are randomly chosen explanatory (or x) variables that can predict y (or outcome values) when they occur.

E.g. Take the first value for professionals which equals 1 as the x1 value and the first value for income which equals 12351 as the x2 value to get prestige (or y value predicted when these x1 and x2 values occur)

$$- y = 27.863 + 17.761 \cdot 1 + 0.002 \cdot 12351$$

- y

- Predicted y is 70.326

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

Coefficient = 0.02. Close to zero coefficient signals a null relationship between income and prestige.

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

Coefficient = 17.761. Indicates strongly positive relationship between professionals and prestige.

- (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).

Add 1000 income to income from (c) equation which equals 12351.

- 12351 + 1000

ANS = 13351

Place in new equation instead of 12351 income in question (c)

- y1 j- 27.863+17.761*1+0.002*13351

- y1

Pineo-Porter prestige score = 72.326 when income is increased by 1000.

This is compared with 70.326 from answer to (c).

Therefore 1000 income increase associated with 2 point increase in prestige points.

Prestige therefore increases by a little bit if income is increased.

- (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).

Create object y2 with professionals as x1

- y2 :- 27.863+17.761*1+0.002*6000

- y2

ANS = 57.624 prestige points

Create object y3 with non-professionals value (or 0) as x1

- y3 :- 27.863+17.761*0+0.002*6000

ANS = 39.863 prestige points

Subtract both prestige points to get effect of profession on prestige at same income rate of 6,000

- 57.624 - 39.863 or y2 - y1

ANS = 17.761.

Therefore, even if they earn the same income of 6,000, a professional is likely to receive 17.761 more prestige points than blue and white collar workers.

Therefore, this indicates that the impact of profession is very large on prestige relative to income.

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 $\alpha = .05$).

Yard signs is explanatory variable, vote share is outcome variable

- 1) Assumptions: Sample size $n = 30$, slope = 0.042, standard error = 0.016
- 2) Null and Alt Hypotheses:
 - H_0 = the effect of yard signs on vote share is null (i.e. the slope = 0),
 - H_A = the effect of yard signs on vote share is present (i.e. the slope does not = 0)
- 3) Calculate test stat by dividing slope and null hyp for slope (or 0) by standard error (i.e. 0.016)
 - $0.042/0.016$

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.

$T \text{ stat} = 2.625$

4) P-value: Use T table

- $N - 1df$ (i.e. $30-1$) $ANS = 19$. T stat on row 19 of t table falls between 0.01 and 0.005. Therefore, 0.0075 for one tail. Multiply by 2 for two-tail test = 0.05

5) Conclusions: p-value is = to level of significance 0.05.

Therefore, the null hypothesis can just about be rejected as p value equals 0.05.

We can determine that it is possible that the prescence of these yard signs within a precinct may impact 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 $\alpha = .05$).

1) Assumptions: Sample size $n = 76$, intercept = 0.042, standard error = 0.013

2) Null and Alt Hypotheses:

H_0 = no effect on vote share (slope = 0)

H_A = effect on vote share (i.e. slope does not = 0)

3) Test Stat: $-0.042/0.013$

T stat = 3.230769

4) P-value = 0.9991

5) Conclusions: P value greater than 0.05. Therefore, we cannot reject the null hypothesis. It cannot be determined that being adjacent to precincts using yard signs impacts vote share

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

The constant 0.302 here is the Y-intercept (where the line crosses the y axis). Standard error of intercept is 0.011.

It tells us that if lawn signs present within and adjacent to precincts are zero (i.e. there are no yard signs), vote share = 0.302.

Therefore, with no yard signs, 0.302 of vote share will go to Ken Cuccinelli.

- (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?

Interpret Rsquared 0.094 to evaluate model fit. Rsquared always between 0-100 percent - Rsquared for model here = 0.094 or 9.4 percent

Closer to 100 percent, the more the model explains outcome variable variance

Therefore, model explains only 9.4 percent of vote share variance

Indicates that other factors that are not modeled are likely have a greater effect on vote share than yard signs. This is because 90.6 percent of vote share variance is not explained by yard signs.