Problem Set 4: Applied Stats/Quant Methods 1

Luna Goldstein

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`.)

```
Prestige$professional <- ifelse(Prestige$type=='prof', 1,0)
```

(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.)

```
model1 <- lm(prestige ~ income + professional + income:professional, data
    = Prestige)
```

The linear model :

Y-intercept = 21.142259,
Coefficient estimate (income) = 0.003171,
Coefficient estimate (professional) = 37.781280,
Coefficient estimate (interaction) = -0.002326,

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

prestige = 21.142259 + 0.003171 * income + 37.781280 * professional + (-0.002326 * income * professional)

(d) Interpret the coefficient for `income`.

The coefficient for income indicates that for every unit increase in income there is a 0.003 increase in the prestige (expected value of y).

(e) Interpret the coefficient for `professional`.

The coefficient for professional indicates that a professional has 37 prestige unit/points more than the working class (blue -and white collar jobs).

(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).

21.142259 + 0.003171*0 + 37.781280 * 1 + (-0.002326 * 0 * 1) = 58.924

21.142259 + 0.003171 * 1000 + 37.781280 * 1 + (-0.002326 * 1000 * 1) = 59.769

59.769 - 58.924 = 0.845

The effect of a 1000 increase in come on prestige score for professional occupations, results in a 0.845 increase in prestige score for professional occupation.

(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).

21.142259 + 0.003171 * 6000 + 37.781280 * 1 + (-0.002326 * 6000 * 1) = 63.994

21.142259 + 0.003171 * 6000 + 37.781280 * 0 + (-0.002326 * 6000 * 0) = 40.164

63.994 - 40.168 = 23.825

The effect of changing one's occupations from non-professional to professional, results in a 23 increase in 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 $\alpha = .05$).

0.042 / 0.016 = 2.625

$\alpha = .005$

We are able to reject the null hypothesis stating that there is no linear relationship between having these yard signs in a precinct and 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.

(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 $\alpha = .05$).

$0.042 / 0.013 = 3.23$

$\alpha = .01$

We are able to reject the null hypothesis stating that there is no linear relationship between being next to precincts with these yard signs and vote share.

(c) Interpret the coefficient for the constant term substantively

The coefficient for the constant term indicates that without lawn signs 30 percent of votes 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?

The model fit for this regression tell us that the use of lawn signs only explain four percent of the vote share. The remaining 96 are due to different factors.