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`).

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.)

model ¡- lm(prestige income + professional, data = Prestige)

(c) Write the prediction equation based on the result. prestige = 30.62 + 0.001371 *

income + 22.76 * professional

(d) Interpret the coefficient for `income`.

There is a positive correlation between income and prestige of the job. However, the impact of the income to the prestige of the job compared to being a professional is very minor.

(e) Interpret the coefficient for `professional`.

Professional's coefficient is 2.276e+01. This means there is a positive correlation between being a professional and prestige of the job.

(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).

marginal impact of 1000 income increase on the prestige score to be calculated by multiplying 1000 with income coefficient

marginalimpactincome ¡- model$coefficients$["$income$"]$*$1000$print$($marginal$impactincome)

(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).

Since we are focusing on the marginal effect of changing occupations from professional to non-professional the value of income doesn't have any impact. There is no marginal effect when income is the same when occupation type changes. The marginal effect to be calculated by using the professional coefficient.

marginalimpactprof ¡- model$coefficients$["$professional$"]$*$(0−1)$print$($marginal$impactprof)
changing occupation type has negative marginal effect (-22.757) on 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 $\alpha = .05$).

Null Hypothesis (H0): The presence of yard signs in a precinct does not affect vote share. The coefficient for "Precinct assigned lawn signs" is equal to zero.

Alternative Hypothesis (HA): The presence of yard signs in a precinct affects vote share. The coefficient for "Precinct assigned lawn signs" is not equal to zero.

t statistic: t = (coefficient - hypothesized value) / standard error t = 0.042 / 0.016 = 2.625 degree of freedom = 128 significance level of 0.05 is approximately ±1.980

[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 $\alpha = .05$).

Since the calculated t-statistic (2.625) is greater than the critical t-value (1.980), we can reject the null hypothesis. This suggests that the presence of yard signs in a precinct does have a statistically significant effect on vote share.

Therefore, based on the calculations, we can conclude that having these yard signs in a precinct does affect vote share.

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

The constant term represents the baseline value or the starting point of the dependent variable (proportion of the vote that went to McAuliffe's opponent Ken Cuccinelli) when all other predictors have zero effect. We can say that , without any influence from the presence of lawn signs, the baseline proportion of the vote that went to McAuliffe's opponent is estimated to be 0.302.(Since the constant term is 0.302)

(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?