Problem Set 4

Paula Montano/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`.)

```
Prestige$professional <- ifelse(Prestige$type == "prof", 1,0)
str(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 × dummy interaction.)

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

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

1. Yi (prestige) = B0 (Intercept) + B1 (income)*Xi + B2 (professional)*Di + Ei
2. Yi = prestige outcome variable, the predicted value
3. B0 = (intercept: 30.6183338)
4. B1 = income (coefficient: 0.0013706) *Xi
5. B2 = professional (coefficient: 22.7569999) *Di
6. Ei = standard error

(d) Interpret the coefficient for `income`.

The income coefficient (0.0013706) shows a positive relationship with the outcome variable prestige. The variable prestige is predicted to increase (0.0013706) when the variable income goes up one.

(e) Interpret the coefficient for `professional`.

The professional coefficient (22.7569999) shows a positive effect on the outcome variable prestige. The variable prestige is predicted to increase (22.7569999) when the variable professional increases by one.

(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).

```
y_incomeVar1 <- 30.6183338 + 0.0013706*1000 + 22.7569999*1 + (1000*
    0.0002)
y_incomeVar1 = 54.9459337
```

When the effect of a $1,000 increase in income, prestige increases as well. The change in Y is 54.9459337

(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).

```
y_incomeVar2 <- 30.6183338 + 0.0013706*6000 + 22.7569999*1 + (6000*
    0.0002)
y_incomeVar2 = 62.7989337
```

The effect of professional when her income increases $6,000, prestige increases significantly. The change in Y is 62.7989337

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 $\alpha = .05$).

1. Set null and alternative hypothesis:

   Ho: Yard signs in precincts do not affect vote share. B2 = 0

   Ha: Yard signs in precincts affect vote share. B2 different from 0.

[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.

2. Test Statistic:

```
## t = (0.042-0) / (0.016)
## t =  2.625
```

3. Calculate p value:

```
## df = N - k = 131 - 3 = 128
p <- 2*pt(2.625, 128, lower.tail = F)
## p = 0.0097200197
```

(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 $\alpha = .05$).

After conducting hypothesis testing, our p value is smaller than alpha (0.05). Therefore, there is enough evidence to reject the null hypothesis (Ho). We can consider that our alternative hypothesis could be true and yard signs in precincts could affect vote share.

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

The intercept or constant value on vote share (outcome) is expected to be 0.302 when law signs (predictor) is zero. The intercept value is important in this regression analysis because it helps us to understand if there is an effect on the outcome variable when the predictor variable is 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? The R square coefficient

(0.094) tells us that the distance of the data from the mean all squared is closer to one. Therefore, the R square in the regression model suggests a good fit for the data and the points fall closer to the real slope.