Problem Set 4

Makenzie Forster

Due: December 4, 2022

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 4, 2022. 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`).

```
professional <- ifelse(Prestige$type == "prof", 1, 0)
Data <- cbind(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.)

```
mod <- lm(Data$prestige ~ Data$income + Data$professional +
              Data$income:Data$professional)
print(summary(mod))
```

```
Call:
lm(formula = Data$prestige ~ Data$income + Data$professional +
    Data$income:Data$professional)

Residuals:
    Min      1Q  Median      3Q     Max
-14.852  -5.332  -1.272   4.658  29.932

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   21.1422589  2.8044261   7.539 2.93e-11 ***
Data$income                    0.0031709  0.0004993   6.351 7.55e-09 ***
Data$professional             37.7812800  4.2482744   8.893 4.14e-14 ***
Data$income:Data$professional -0.0023257  0.0005675  -4.098 8.83e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.012 on 94 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.7872,    Adjusted R-squared:  0.7804
F-statistic: 115.9 on 3 and 94 DF,  p-value: < 2.2e-16
```

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

`Y = 21.1422589 + 0.0031709(income) + 37.7812800(professional) + -0.0023257 (income * professional)`

(d) Interpret the coefficient for `income`.

There is a small, positive, and statistically reliable relationship between income and occupational prestige, such that a one dollar increase in income is associated with an average increase of 0.003 points in occupational prestige score.

(e) Interpret the coefficient for `professional`.

There is a positive and statistically reliable relationship between having a Professional job type and occupational prestige, such that having a Professional job type is associated with an average increase of 37.78 points in occupational prestige score.

(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).

```
base <- 21.1422589 + 0.0031709*(0) + 37.7812800 * (1) + (-0.0023257 * (0)
    )
raise <- 21.1422589 + 0.0031709*(1000) + 37.7812800 * (1) + (-0.0023257 *
    (1000 * 1))
raise - base
```

We would expect that a \$1000 raise would lead to an 0.8452 point increase in prestige score.

(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).

```
prof <- 21.1422589 + 0.0031709*(6000) + 37.7812800 * (1) + (-0.0023257 *
    (6000 * 1))
nonprof <- 21.1422589 + 0.0031709*(6000) + 37.7812800 * (0) + (-0.0023257
    * (6000 * 0))
prof - nonprof
```

For an individual making \$6000, we would expect that moving from nonprofessional to professional would lead to a 23.82 point 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 $\alpha = .05$).

Ho: assigned = constant Ha: assigned $\neq$ constant (two tailed test)

SE = 0.016

```
# Step 2: Calculate the test statistic
Z <- (0.042 - 0.302)/0.016 # Z = -16.5
# Calculate the p-value
pvalue <- 2*pnorm(Z, lower.tail = F)
```

pvalue = 2 , thus we fail to reject the Null hypothesis. Lawn signs do not have a statistically significant impact on 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 $\alpha = .05$).

Ho: assigned = constant Ha: assigned $\neq$ constant (two tailed test)

SE = 0.013

```r
# Step 2: Calculate the test statistic
z <- (0.042 - 0.302)/0.013 # z = -20
# Step 3: Caluclate p value
pvalue <- 2*pnorm(z, lower.tail = F)
```

pvalue = 2 , thus we fail to reject the Null hypothesis. Being adjacent to lawn signs does not have a statistically significant impact on vote share

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

In precincts without the incluence of yard signs, Cuccinelli would receive on average 30.2% of the vote.

(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?

r2 = 0.094. Yard Signs account for only 9.4% of variation in percent vote share.