Problem Set 4 - Laura McPhillips

Applied Stats/Quant Methods 1

Due: November 26, 2021

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)
```

(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.)

```
lm(Prestige$prestige ~ Prestige$income + professional + Prestige$income:
    professional)
```

Y-intercept = 21.142259
Income co-efficient = 0.003171
professional co-efficient = 37.781280
Income:professional co-efficient = -0.002326

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

```
yi = bo + bixi + b2di + b3xidi
yi <- 21.14 + 0.00317179*Income + 37.78*di - 0.00232*Income*di
```

(d) Interpret the coefficient for `income`.

For every unit increase in income, its is predicted that there is an 0.00371 unit increase in the expected value of prestige, holding every other variable constant at their empirical means.

(e) Interpret the coefficient for `professional`.

When a person's job changes from non-professional to professional, it is predicted that there is a 37.78 point increase in the expected value of prestige, holding every other variable constant at their empirical means.

(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).

There is a 0.845 increase in prestige points when income increases by \$1000.

```
y_1 <- 21.14 + 0.00317179*0 + 37.78*1 - 0.002326*0*1
y_2 <- 21.14 + 0.00317179*1000 + 37.78*1 - 0.002326*1000*1
y_2 - y_1 = 0.845
```

(I calcualted prestige at \$0, and then at \$1000, to find the \$1000 increase.)

(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).

When income is \$6000, it is predicted that changing from a non-professional job to a professional job will increase prestige points by 23.82 from 40.17 to 63.99, holding every other variable constant at their empirical means.

```
y_non_prof <- 21.14 + 0.00317179*6000 + 37.78*0 - 0.002326*6000*0
y_prof <- 21.14 + 0.00317179*6000 + 37.78*1 - 0.002326*6000*1
y_non_prof = 40.17
y_prof = 63.99
y_prof - y_non_prof = 23.82
```

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 lawn signs have no impact on the voteshare of Cuccinelli. $\beta = 0$

Ha: Assigned lawn signs do have an impact on the voteshare of Cucinelli. $\beta \neq 0$

(First, we find t-statistic $(\frac{\beta}{SE(\beta)})$, then interpret p-value at the 0.05 level of significance. Two-tailed test.)

[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.

```
t_stat <- 0.042/0.016
t_stat = 2.625
p_value <- .005
p_value*2 = 0.01
```

P-value $= 0.01 < \alpha = 0.05$, so we reject the null hypothesis, that assigned lawn signs have no impact on voteshare.

(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: Adjacent lawn signs have no impact on the voteshare of Cuccinelli. $\beta = 0$

Ha: Adjacent lawn signs do have an impact on the voteshare of Cucinelli. $\beta \neq 0$

(First, we find t-statistic ($\frac{\beta}{SE(\beta)}$), then interpret p-value at the 0.05 level of significance. Two-tailed test.)

```
t_stat2 <- 0.042/0.013
t_stat2 = 3.23
p_value2 <- 0.000787
p_value2*2 = 0.0015
```

P-value $= 0.0015 < \alpha = 0.05$, so we reject the null hypothesis that adjacent lawn signs have no impact on voteshare.

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

When the two variables, assigned lawn signs and adjacent lawn signs are 0, i.e. there are no lawn signs, there is still a 30percent voteshare for Cucinelli. The lawn signs then increase voteshare by 4percent. (When x1 and x2 are 0, y-intercept is 0.3).

(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

R^2 is 0.094. This means that only 9.4percent of the voteshare is explained by lawn signs. This is quite low. This means that there are other variables affecting vote share that are not examined here. There is a lot variation among the data points; there are high residuals in this model.