Problem Set 4

Doireanna Craven

Due: November 26, 2021

Question 1: Economics

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_1 <- mutate(Prestige, prof = type)
# Create a new column. I have copied the column named 'type' and called
# the new column 'prof'

Prestige_1 <- na.exclude(Prestige_1)
# Exclude observations which record NA

Prestige_1$prof <- ifelse(Prestige_1$prof == "prof", 1, 0)
# This ifelse condition reassigns elements in 'prof' as 1 or 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.)

```
# Prestige depends on income, professional status (dummy variable),
# and the interaction of the two
interaction_regression <- lm (prestige ~ income + prof +
                                        income:prof, data = Prestige_1)
summary(interaction_regression) # Returns intercept, slopes and se's
```

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

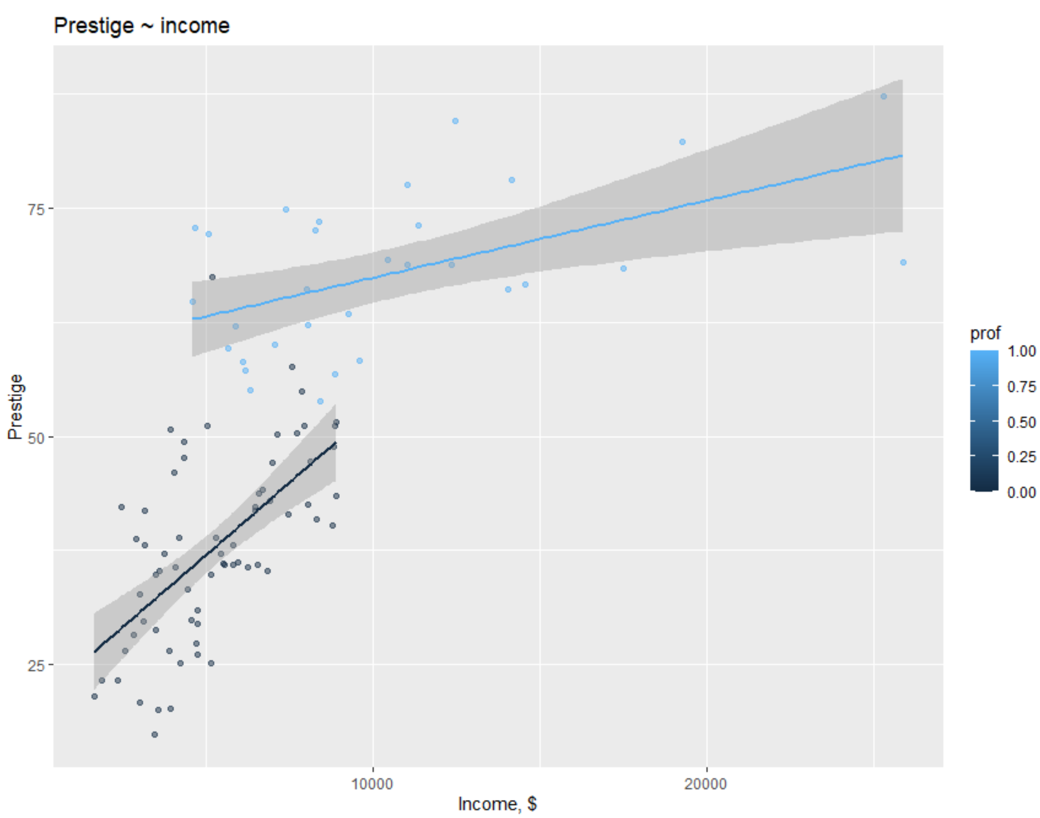
P_i = 21.1 + 0.003I_i + 37.8S_i - 0.002I_iS_i + ϵ_i
where P_i is prestige, I_i is income, and S_i is the dummy variable for professional status.

(d) Interpret the coefficient for `income`.

When professional status is white- or blue-collared workers (e.g., dummy variable = zero), on average, a $1,000 increase in income is associated with a 3 unit increase on the Pineo-Porter prestige score for occupation (a scale which runs from 10 to 90).

(e) Interpret the coefficient for `professional`.

When income is zero, the effect of becoming a professional worker is on average associated with a 37.8 unit increase on the Pineo-Porter prestige scale. This doesn't intuitively make sense since a worker would not be both earning zero and classed as a blue or white-collared worker. The graph below models the relationship between prestige and income for professional and non-professional workers'. Extrapolating the regression lines to the vertical axis demonstrates the implications of the coefficient for `professional`: the intercepts would be such that earning zero a professional class worker would have prestige approximately 40 units higher than a non-professional.



(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).

The formula is:

$P_i = 21.1 + 0.003I_i + 37.8S_i - 0.002I_iS_i + \epsilon_i$

With $S_i = 1$, increasing I_i by \$1,000 will affect the second and fourth terms in the formula, yielding an increase of $0.003 \times 1{,}000$ and a decrease of $0.002 \times 1{,}000$. The net gain is 1 unit.

For professional occupations, an increase in income of \$1,000 is on average associated with a 1 unit increase in prestige.

(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).

Changing S_i from 0 to 1 in the equation will introduce the third and fourth terms, therefore adding 37.8 units and $-0.002I_i$ units. With $I_i = 6{,}000$, the gain is 37.8 - 12 = 25.8 units of prestige.

The marginal effect on prestige of changing one's occupation from non-professional to professional when income is \$6,000 is 25.8 units, on average.

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 $\alpha = .05$).

Equation for additive model: $Y_i = \beta_0 + \beta_{assigned}X_1 + \beta_{adjacent}X_2 + \epsilon$

We need to conduct a Student's t-test on $\beta_{assigned}$:

H_0: $\beta_{assigned} = 0$, yard signs have no effect on voteshare in the precinct where they are posted.

H_a: $\beta_{assigned} \neq 0$, yard signs have an effect on voteshare in the precinct where they are posted.

[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.

Test statistic: t = $\beta_{assigned}/se_{assigned}$ = 0.042/0.016 = 2.625

DF = n - k - 1 = 131 - 2 - 1 = 128

P-value = 2 $\times$ Pr(t_{128} > 2.625) = 0.0097

```r
a_test_stat <- 0.042/0.016
a_p_value <- 2*pt(a_test_stat, df = 128, lower.tail = FALSE)
```

The p-value is less than 0.05, therefore we can reject the null hypothesis that yard signs have no effect on voteshare in the precinct where they are posted, at the $\alpha = 0.05$ level.

(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 $\alpha = .05$). Equation for additive model: $Y_i = \beta_0 + \beta_{assigned}X_1 + \beta_{adjacent}X_2 + \epsilon$

We need to conduct a Student's t-test on $\beta_{adjacent}$:

H_0: $\beta_{adjacent} = 0$, yard signs have no effect on voteshare in the precinct adjacent to where they are posted.

H_a: $\beta_{adjacent} \neq 0$, yard signs have an effect on voteshare in the precinct adjacent to where they are posted.

Test statistic: t = $\beta_{adjacent}/se_{adjacent}$ = 0.042/0.013 = 3.23

P-value = 2 $\times$ Pr(t_{128} > 3.23) = 0.0016

```r
b_test_stat = 0.042/0.013
b_p_value <- 2*pt(b_test_stat, df = 128, lower.tail = FALSE)
```

The p-value is less than 0.05, therefore we can reject the null hypothesis that yard signs have no effect on voteshare in the precintc adjacent to where they are posted, at the $\alpha = 0.05$ level.

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

The constant term indicates Ken Cuccinelli's average voteshare in the precincts which neither have placards, nor are adjacent to precincts with placards.

(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^2 value is given as 0.094, which suggests that just 9.4% of the observed variation in voteshare is explained by the placards. This does not mean that the variables are not useful factors - in fact we have seen from the p-values of $\beta_{assigned}$ and $\beta_{adjacent}$ that they are statsitically significant - but we can say that the model does not explain the vast majority of the variation from the mean. This remaining variation is bound up in the constant term.