Problem Set 4

Samanta Nedzinskaite

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)
```

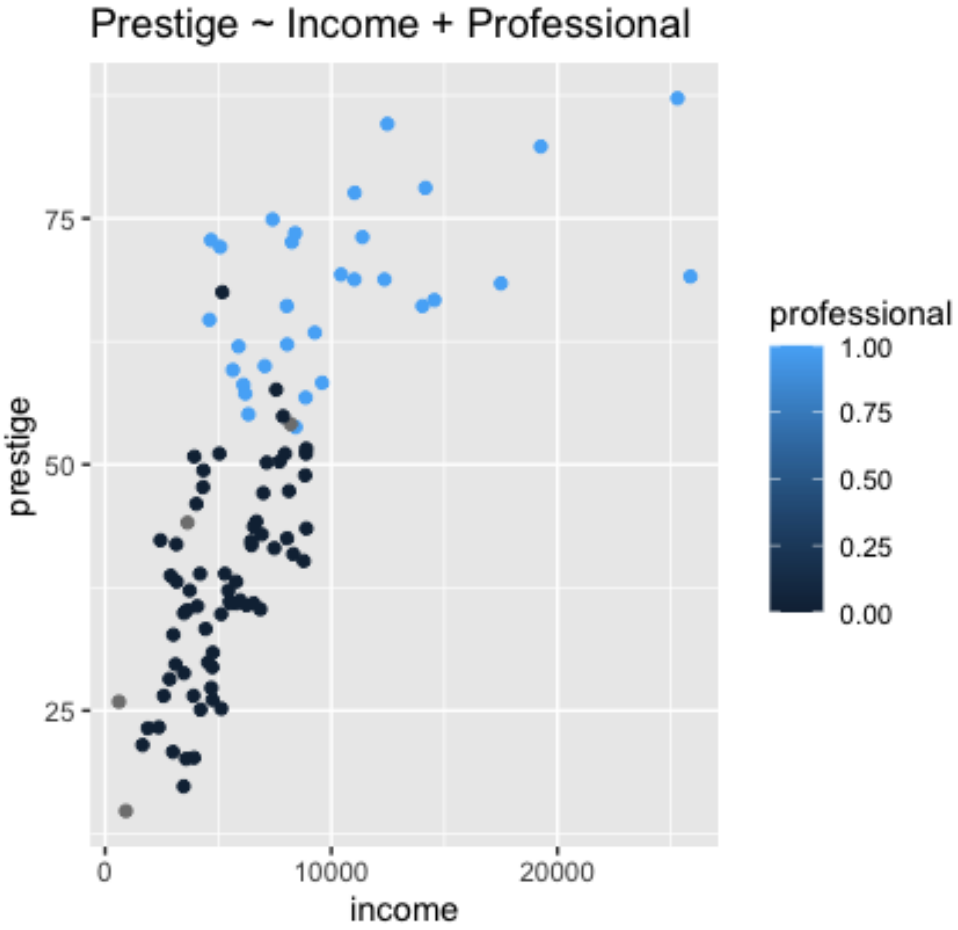
We would like to study whether individuals with higher levels of income have more prestigious jobs. Moreover, we would like to study whether professionals have more prestigious jobs than blue and white collar workers.

(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`).

```
data$professional <- ifelse(data$type == "prof", 1, 0)
head(data)
```

	education	income	women	prestige	census	type	professional
gov.administrators	13.11	12351	11.16	68.8	1113	prof	1
general.managers	12.26	25879	4.02	69.1	1130	prof	1
accountants	12.77	9271	15.70	63.4	1171	prof	1
purchasing.officers	11.42	8865	9.11	56.8	1175	prof	1
chemists	14.62	8403	11.68	73.5	2111	prof	1
physicists	15.64	11030	5.13	77.6	2113	prof	1

(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.)



```r
Income_Prestige_Prof <- data %>%
  lm(formula = prestige ~ income + professional)
```

Prestige, Income and Professional

	Dependent variable:
	prestige
income	0.001***
	(0.0003)
professional	22.757***
	(2.318)
Constant	30.618***
	(1.714)
Observations	98
R2	0.749
Adjusted R2	0.744
Residual Std. Error	8.652 (df = 95)
F Statistic	141.836*** (df = 2; 95)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

$Y = 30.618334 + (0.001371 * Income) + (22.757 * X2)$

(d) Interpret the coefficient for `income`.

```
Call:
lm(formula = prestige ~ income + professional, data = .)

Coefficients:
 (Intercept)        income  professional
   30.618334      0.001371     22.757000
```

From the regression model above, we can interpret that for non-professional workers (white collor or black collor), as their income goes up by `0.001371 units`, the prestige score goes up by `1 unit`.

(e) Interpret the coefficient for `professional`.

For professional workers, as their income goes up by `(0.001371 + 22.75700) units`, the prestige score goes up by `1 unit`.

(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).

$Y = 30.618334 + (0.001371 * 1000) + 22.757$

$Y = 54.74633$

(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).

Non-professional workers: $Y2 = 30.618334 + (0.001371 * 6000)$

$Y2 = 38.84433$

Professional workers: $Y2_2 = 30.618334 + (0.001371 * 6000) + 22.757$

$Y2_2 = 61.60133$

Change in Y: $38.84433 - 61.60133 = -22.757$

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 $\alpha = .05$).

H0: there was no effect on vote share in presincts with lawn signs ($ie : beta == 0.$)

HA: there was an effect on vote share in presincts with lawn signs $ie : beta = / = 0.$)

```
#testing individual regression coefficients
#Presincts assigned lawn signs
t = 0.042 / 0.016
p_value = 2*(1-pt(t, 131-3))
```

[1] 0.00972002

```
#Presincts adjacent to law signs
t2 = 0.042 / 0.013
p_value2 = 2*(1-pt(t2, 131-3))
```

[1] 0.00156946

[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 $\alpha = .05$).

```
#Presincts adjacent to law signs
t2 = 0.042 / 0.013
p_value2 = 2*(1-pt(t2, 131-3))
```

[1] 0.00156946

Since 0.00156946 $<$ 0.05, the effect on vote in presincts adjacent to those that had lawn signs, was significantly different from 0, with $\alpha = .05$.

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

The constant in this regression is 0.302. This value represents the vote share that would go to Ken Cuccinelli if all the predictor variables were simultaneously equal to zero. In this instance, this would mean that about 30 percent of the votes would have went to Ken Cuccinelli, if there were no lawn signs in any presinct against McAuliff. From our previous hypothesis testing, it is therefore perceivable that Cuccinelli's vote share increased by some significant value when lawn signs against McAuliff were in place.

(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?

```
#Testing overall model fit with F-test
F <- (131 - 2 -1)/2*0.094/(1-0.094)
F
p_value<- 1- pf(F, 2, 128)
```

[1] 0.001803868

Since 0.001803868 $<$ 0.05, at least one of the regression coefficients is statistically significantly different from 0. The two predictors have some signficant effect on vote share.