Problem Set 4

Stefan Keel
Applied Stats/Quant Methods 1

Due: December 3, 2023

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 3, 2023. 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`).

First I conducted a visual analysis of the dataset for a general overview. Afterwards I applied a line of code to check for missing values in the column "type" - which indicated that there are some missing values. I then dropped those lines as it is not in my knowledge to assign them to "professional" or "blue collar worker" status. To code these two levels $(0, 1)$. I then used the other line of code to code it 1 if "prof" for "professional" and 0 for all other.

```
print(any(is.na(Prestige$type)))
Prestige <- Prestige[!is.na(Prestige$type), ]

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

This created the following (head) of the dataset:

	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 $\times$ dummy interaction.)

I used this code to run the linear model:

```
model1 <- lm(prestige ~ income + professional + income*professional,
             data = Prestige)
```

Which provided me with this summary:

Table 1:

	Dependent variable:
	prestige
income	0.003^{***}
	(0.0005)
professional	37.781^{***}
	(4.248)
income:professional	-0.002^{***}
	(0.001)
Constant	21.142^{***}
	(2.804)
Observations	98
R^2	0.787
Adjusted R^2	0.780
Residual Std. Error	8.012 (df = 94)
F Statistic	115.878^{***} (df = 3; 94)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

$$\hat{Y} = 21.1422589 + 0.0031709X_1 + 37.7812800D_1 - 0.0023257X_1 * D_1$$

(d) Interpret the coefficient for `income`.

- The average prestige score for not-receiving-income (Income=0) blue collar workers is 21.142 score points.
- The average prestige score for not-receiving-income (Income=0) professionals is 58.92354 score points.
- For blue collar workers, with every additional 1 USD of average yearly income, the prestige score increases on average by 0.0031709 scale points = Income effect for blue colar workers.

- For professionals, with every additional 1 USD of average yearly income, the average prestige score increases by 0.00084528 = Income effect for professionals.
- For not-receiving-income professionals, the prestige score is 37.7812800 score points higher, in comparison to not-receiving-income blue collar workers.

The coefficient for the variable **INCOME is 0.0031709** and shows a positive association with the outcome variable (prestige).

(e) Interpret the coefficient for `professional`.

The coefficient for the variable **PROFESSIONAL is 37.7812800** and also shows a positive association with the outcome variable (PRESTIGE). In switching from "blue collar worker" to "professional" (moving from 0 to 1) (an increase of one unit) in the variable PROFESSIONAL results in an increase of 37.7812800 units in the estimated/predicted value of PRESTIGE $\hat{Y}$).

- The average prestige score for not-receiving-income (Income=0) professionals is on average 58.92354 score points.
- For professionals, with every additional 1 USD of average yearly income, the average prestige score increases by 0.00084528 = Income effect for professionals.
- For not-receiving-income professionals, the prestige score is on average 37.7812800 score points higher, in comparison to not-receiving-income blue collar workers.

(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).

I used the prediction equation and used a simple calculation:

```
prestige1000 <- (21.1422589 + (0.0031709 * 1000) + (37.7812800*1)
- (0.0023257 * 1 * 1000))
print(prestige1000)
```

A $1,000 income increase for professionals would on average result in **59.76874** for $\hat{Y}$ 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).

I used the prediction equation and used a simple calculation:

```
prestige6000_0 <- (21.1422589 + (0.0031709 * 6000) + (37.7812800*0)
- (0.0023257 * 0))

prestige6000_p <- (21.1422589 + (0.0031709 * 6000) + (37.7812800*1)
- (0.0023257 * 1 * 6000))

difference <- prestige6000_p - prestige6000_0
```

The predicted prestige score for professionals with income of \$6,000 would be 63.99474 whereas the prestige score with the same income for blue collar workers would, on average, be 40.16766. **blue collar workers have on average a 23.82708 lower** prestige score than professionals with same income of \$6,000.

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 = Yard signs in precincts do not affect outcome // B2 = 0
HA = Yard signs in precincts affect outcome // B2 not 0

I did the T Statistic by dividing the variability between groups by the variability within groups. From there I calculated the P Value. I opted for T statistic for 2a and 2b because I do not know the population standard deviation and therefore T seems to be more appropriate.

[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.

```
t1 <- (0.042 - 0) / 0.016
=2.265

p1 <- 2 * pt(t1, 128, lower.tail=FALSE)
```

The P value of 0.00972002 is smaller than 0.05 therefore there is evidence to reject HO. **There are indicators that signs on lawns may affect vote share.**

(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 $\alpha = .05$).

H0 = Yard signs in adjacent precincts do not affect outcome // B2 = 0
HA = Yard signs in adjacent precincts affect outcome // B2 not 0

```
t2 = (0.042 - 0) / 0.013

p2 <- 2 * pt(t2, 128, lower.tail=FALSE)
```

The P value of 0.00156946 is below 0.05 an therefore there is evidence to reject HO. **There are indicators that signs on lawns in adjacent precinct may affect vote share.**

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

The intercept indicates the predicted outcome for voteshare if the other predictors (treatment) are 0. Therefore, the average predicted outcome for voteshare is 0.302 without signs.

(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 of 0.094 indicates that only 9.4% of the variability in the dependent variable is explained by the independent variables, whereas 90.6% of the variability is not accounted for by the included variables. There might be other factors that are not included in the model, that could explain the variability in voteshare.