Problem Set 4

Applied Stats/Quant Methods 1

Charis Rogers (19333698)

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)
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 $\times$ dummy interaction.)

```
interaction_model <- lm(prestige ~ income + professional + income:
    professional,
                        data = Prestige)
```

	Dependent variable:
	Prestige
Constant	21.142^{***}
	(2.804)
Income	0.003^{***}
	(0.0005)
Professional	37.781^{***}
	(4.248)
Income:Professional	-0.002^{***}
	(0.001)
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.

Prestige = 21.142 + 0.003*Income + 37.781*Professional - 0.002*Income*Professional

(d) Interpret the coefficient for `income`.

As seen in the above equation the coefficient of `income` is 0.003. This is the effect of income on prestige, when the interaction term is zero. This means that for blue and white collar workers, when income increases by one unit, prestige increases by 3%.

(e) Interpret the coefficient for `professional`.

The coefficient of `professional` refers to the binary variable of professionals compared to blue and white collar workers. For professionals, they are 38 units more prestigious than blue and white collar workers, when income is zero.

(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).

Prestige = 21.142 + 0.003*Income + 37.781*Professional - 0.002*Income*Professional

As seen in the equation above, the change in prestige when `professional` is zero is 0.003. When `professional` is one (when we are looking at the group of professionals) the interaction term must also be taken into account. This means that the a $1,000 increase in income will result in a 0.8452 change in the prestige score.

```
Y <- 0.0031709*(1000) - 0.0023257*(1000)*(1)
Y #0.8452
```

(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).

Calculating the effect of changing from non-professional to professional when income is zero is 37.781. However, when income is $6,000$, the interaction effect must be taken into account so that the change in $\hat{y}$ is 23.827.

```
Y <- 37.7812800*1 - 0.0023257*(6000)*(1)
Y #23.82708
```

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 $\alpha = .05$).

Vote Share = 0.302 + 0.042*Assigned + 0.042*Adjacent

As seen in the equation above we are interested in the coefficient of assigned lawn signs, 0.042. In order to test if there is a linear relationship between assigned lawn signs and the vote share, controlling for the adjacent precinct, we can construct hypotheses to test.

Null hypothesis: There is no linear relationship between these yard signs and vote share, controlling for the adjacent precinct.

Alternative hypothesis: There is a relationship between these yard signs and the vote share, controlling for the adjacent precinct.

[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.

Calculating the test statistic (2.625), we can then calculate p value, 0.0097. This p value is smaller than 0.05 which means that we can reject the null hypothesis that there is no relationship. This means that, controlling for the lawn signs in adjacent precincts, there seems to be a significant relationship between the lawn signs and vote share.

```
TSa <- 0.042/0.016
p_value_a <- 2*pt(abs(TSa), 131-2-1, lower.tail = FALSE)
p_value_a #0.00972002
```

(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 $\alpha = .05$).

To test if being next to precincts with these yard signs affect vote share we construct hypothesis with the second variable.

Null hypothesis: There is no linear relationship between being next to precincts with these yard signs and vote share, controlling for the yard signs being posted.

Alternative hypothesis: There is a relationship between being next to precincts with these yard signs and the vote share, controlling for the yard signs being posted.

Using the test statistic (3.231), we can calculate the p value which is 0.0016. This p value is smaller than 0.05 which means that we can reject the null hypothesis that there is no linear relationship. This indicates that being in a precinct next to precincts with yard signs posted is related to the vote share, controlling for the yard signs being posted.

```
TSb <- 0.042/0.013
p_value_b <- 2*pt(abs(TSb), 131-2-1, lower.tail = FALSE)
p_value_b #0.00156946
```

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

The constant term in this model is 0.302. This means that Ken Cuccinelli has 30.2% of the vote share in precincts in which (1) the precinct was not assigned to have signs posted or (2) not beside precincts in which the signs were posted.

(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 modelled?

We can see from the table above that R^2=0.094, which means that the current model explains 9.4% of the variation in vote share. This means that there is about 91% of variation in vote share that is not related to the two variables modelled, (1) if the precinct had a sign against McAuliff posted or (2) if the precinct was beside one of these precincts. Essentially, these two variables explain about 9% of the variability of Cuccinelli's vote share.