Problem Set 4

Applied Stats/Quant Methods 1

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.)

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

(b) Run a linear model with prestige as an outcome and income, professional, and then interaction of the two as predictors

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

```
##
## Call:
## lm(formula = prestige ~ income + professional, data = Prestige)
##
## Coefficients:
##  (Intercept)        income  professional
##    30.618334      0.001371     22.757000
```

(c) Write the prediction equation based on the result.
Based on the above results, the prediction equation is:

$$\hat{prestige} = 30.618 + 0.00137 income + 22.757 professional$$

(d) Interpret the coefficient for income.
The coefficient of income is 0.00137. This implies that when the average income of incubants increases by 1 dollar, the Pineo-Porter prestige score for occupation increases by 0.00137

(e) Interpret the coefficient for professional.
In this case, the White collar and Blue collar occupations are the reference group. Therefore, the Professional, Managerial, and Technical occupations are being compared to to the White collar and Blue collar occupations. The coefficient of the professional is 22.757, and it is positive. Therefore, the Pineo-Porter prestige score for occupation of Professional, Managerial, and Technical occupations exceeds that of White collar and Blue collar occupations by 22.757.

(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).

```
change <- 30.618+(0.00137*1000) +(22.757*1);change
```

```
## [1] 54.745
```

When he income increases by $1,000, the prestige score for professional occupations increases by 54.745. It is the marginal effect of income when the variable professional takes the value of 1

(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)

```
change2 <- 30.618+(0.00137*6000) +(22.757*1);change2
```

```
## [1] 61.595
```

When one changes occupations from non-professional to professional when her income is $6,000, the prestige score increases to 61.595. It is the marginal effect of professional jobs when the variable income takes the value of 6,000.

Question 2: Political Science

Researchers are interested in learning the effect of all of those yard signs on voting preferences. 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).

Table 1: **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 $\alpha = .05$).
H_0 :: Having the yard signs in a precinct does not affect vote share VS H_1 : Having the yard signs in a precinct affects vote share
The p- value is 0.016, which is less than the significance level($\alpha = 0.05$). The decision rule is that we reject H_0 when $p - value \leq \alpha$. Therefore we reject the null hypothesis, and automatically fail to reject the alternative hypothesis. We conclude that Having the yard signs in a precinct affects 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 $\alpha = .05$).
H_0 :: Being next to precincts with these yard signs does not affect vote share VS H_1 : Being next to precincts with these yard signs affects vote share
The p- value is 0.013, which is less than the significance level($\alpha = 0.05$). The decision rule is that we reject H_0 when $p - value \leq \alpha$. Therefore we reject the null hypothesis, and automatically fail to reject the alternative hypothesis. We conclude that Being next to precincts with these yard signs affects vote share.

(c) Interpret the coefficient for the constant term substantively.
The contant term is actually the proportion of the vote that went to McAuliff's opponent Ken Cuccinelli when there was neither precinct assigned lawn signs nor precinct adjacent to lawn signs. The constant in linear regression indicates the value of the dependent variable if all the explanatory variables included in the model are equal to zero at a certain time period.

(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 coefficient determination (R^2) explains how well the regression model fits the observed data. In this case, $R^2 = 0.094(9.4\%)$. This shows that lawn signs have only 9.4% impact on the vote share and the other 90.6% impacts are from the factors which are not modelled. Therefore, the yard signs have some impacts in the vote share.