Problem Set 4

Applied Stats/Quant Methods 1

Due: December 4, 2022

Question 1: Economics

In this question, use the `prestige` dataset in the `car` library. First, we 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`).

Create dummy variables:

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

Create data frame to use for regression:

```
df_new <- cbind(Prestige, professional)
```

View data frame:

```
xtable(head(df_new))
```

	education	income	women	prestige	census	type	professional
gov.administrators	13.11	12351	11.16	68.80	1113	prof	1.00
general.managers	12.26	25879	4.02	69.10	1130	prof	1.00
accountants	12.77	9271	15.70	63.40	1171	prof	1.00
purchasing.officers	11.42	8865	9.11	56.80	1175	prof	1.00
chemists	14.62	8403	11.68	73.50	2111	prof	1.00
physicists	15.64	11030	5.13	77.60	2113	prof	1.00

(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.)

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

stargazer(model1, type = "latex", out = "model1.latex", title =
              "Prestige and income-professional Regression")
```

Table 1: Prestige and income-professional Regression

	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.

prestige = 21.142 + (0.003*income) + (37.781*professional) + (-0.002*37.7812800*0.0031709)

(d) Interpret the coefficient for `income`.

The coefficient for income is 0.003, which is positive. This indicates that as the value of income increases, the mean of the dependant variable also tends to increase, and a one-unit shift in income (holding all other varables constant) causes a 0.003 unit increase in prestige.

(e) Interpret the coefficient for `professional`.

The coefficient for professional is 37.781, which is positive. This indicates that income is higher for the dummy variable 'professional' than for the reference group (white and blue collar workers), and that type 'professional' indicates a 37.781 increase in income.

(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_1000increase = 21.142 + (0.003*1000) + (37.781*1) + (-0.002*37.7812800*0.0031709)
prestige_1000increase
```

```
[1] 61.92276
```

The change in prestige associated with a $1,000 increase in income is 61.92276.

(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).

```

presitge_notprof = 21.142 + (0.003*6000) + (37.781*0) + (-0.002*37.7812800*0.0031709)
presitge_notprof

presitge_prof = 21.142 + (0.003*6000) + (37.781*1) + (-0.002*37.7812800*0.0031709)
presitge_prof

newPrestige <- presitge_prof - presitge_notprof
```

```
[1] 37.781
```

This means that changing one's occupations from non-professional to professional when income is \$6,000 leads to a 37.781 increase in prestige.

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: B1 (the slope of the regression line for the effect of living in a precinct assigned lawn signs on proportion of vote to Ken Cuccinelli) = 0

H1: B1 (the slope of the regression line for the effect of living in a precinct assigned lawn signs on proportion of vote to Ken Cuccinelli) != 0

$\alpha = .05$

[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.

Get the test statistic:

Get critical value at 0.05:

```
# get critical value at 0.05
n1 <- 131

degreesf <- n1 - 3
degreesf

p_value1 <- 2*pt(abs(t1), n1-3, lower.tail = F)
```

Since our p value of 0.0097 is less than our significance level of 0.05, we reject the null hypothesis that the slope of the regression line for the effect of living in a precinct assigned lawn signs on proportion of vote to Ken Cuccinelli = 0.

(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: B2 (slope of the regression line for the effect of living in a precinct adjacent to lawn signs on proportion of vote to Ken Cuccinelli) = 0
H1: B2 (the slope of the regression line for the effect of living in a precinct adjacent to lawn signs on proportion of vote to Ken Cuccinelli) != 0

$\alpha = .05$

Get the test statistic

```

t2 <- 0.042 / 0.013
```

Get critical value at 0.05

```
# get critical value at 0.05
n2 <- 131

degreesf <- n2 - 3
degreesf

p_value2 <- 2*pt(abs(t2), n2-3, lower.tail = F)
```

Since our p value of 0.002 is greater than our significance level of 0.05, we reject the null hypothesis that the slope of the regression line for the effect of living in a precinct adjacent to lawn signs on proportion of vote to Ken Cuccinelli = 0.

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

In this model, a constant of 0.302 indicates that the value that would be predicted for the proportion of the vote to go to Ken Cucinelli if all the independent variables (yard signs and yard sign adjacent) were simultaneously equal to zero is 0.302.

(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?

H0: all regression coefficients are equal to zero
H1: At least one regression coefficient is not equal to zero

R^2 for this model is 0.094.

```
F.test <- (R.squared / (k - 1)) / ((1 - R.squared) / ( n - k))
F.test <- (0.094 / (3 - 1)) / ((1 - 0.094) / ( 131 - 3))
F.test

df1 <- 3 - 1
df2 <- 128
F.pvalue <- pf(F.test , df1 , df2)
```

A p-value of 0.9981961 is less than our significance level of 0.05, which means that we can reject our null hypothesis that all of the co-variates are equal to zero.

Our alternate hypothesis that at least one of the co-variates has a linear relationship with our outcome is supported, but it is of note that this does not tell us which one, or anything about directionality. However, we can conclude from this test that compared to other factors not modeled in this regression, yard signs do have some impact on proportion of vote going to the opponent Ken Cuccinelli.