Problem Set 4

Applied Stats/Quant Methods 1

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`).

```
Code:
    Professional = ifelse(Prestige$type == 'prof', 1, 0)
Professional

Join the new variable to the data frame:

Prestige <- cbind(Professional, Prestige)
```

(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.)

```
Remove the NA values:
Prestige <- na.omit(Prestige)

Run the regression model:

model2 <- lm(prestige ~ income + Professional + income:Professional, data=Prestige)
summary(model2)

OUTPUT:

Call:
lm(formula = prestige ~ income + Professional + income:Professional,
    data = Prestige)

Residuals:
    Min      1Q  Median      3Q     Max
-14.852  -5.332  -1.272   4.658  29.932

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         21.1422589  2.8044261   7.539 2.93e-11 ***
income               0.0031709  0.0004993   6.351 7.55e-09 ***
Professional        37.7812800  4.2482744   8.893 4.14e-14 ***
income:Professional -0.0023257  0.0005675  -4.098 8.83e-05 ***
```

```
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.012 on 94 degrees of freedom
Multiple R-squared:  0.7872,Adjusted R-squared:  0.7804
F-statistic: 115.9 on 3 and 94 DF,  p-value: < 2.2e-16
```

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

$\hat{y}$ = b0 + b1X1 + b2X2 + b3X1X2

Prestige = 21.14 + .003(income) +
37.78(professional)+ (-.002)(income)(professional)

(d) Interpret the coefficient for `income`.

y = 21.14 + .003(income) for White and blue Collar

y = (21.14 + 37.78) + (.003 - .002)(income) for professionals

For white and blue collar workers:
Income = 21.14 - (prestige) / .003

For professionals:
Income = 58.92 - (prestige) / .001

(e) Interpret the coefficient for `professional`.

Y = 21.14 + .003(income) for White and blue Collar

y = 58.92 + .001(income) for professionals

(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 = 21.14 + .003(income)
+ 37.78(professional)
+ (-.002)(income)(professional)

prestige = 24.14 + 37.78(professional) -2(professional)

White Collar and Blue Collar:

Prestige = 24.14

For professionals:
Prestige = 59.92
```

(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).

```
Income = 6000 Profession = 0 = non professional

Prestige = 21.14 + (.003)(6000) + 37.78(0)+(-.002)(6000)(0)

Prestige = 39.14

Income = 6000 Profession = 1 = professionals

Prestige = 21.14 + (.003)(6000) + 37.78 (1) + (-.002)(6000)(1)

Prestige = 64.92

The change in ^y = 25.78
```

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 $\alpha = .05$).

```
Hypotheis test:
Null H0 : b3 = 0
Alt Ha : b3 not = 0

y = .302 + .042b1 + .042b2

Test statistic:
t = b3 hat - 0 / seB3
t = .042 - 0 / .016
```

[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.

```
t = 2.625

P value:
df = n -3 (3 variables in model)
df = 128

The p value based on these parameters:
 p = 0.00972

Based on the p value threshold of .05
we reject the null hypothesis that having these yard signs
in a precinct does not 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 $\alpha = .05$).

```
Hypotheis test:
Null H0 : b3 = 0
Alt Ha : b3 not = 0

Test statistic:
t = b3 hat - 0 / seB3
t = .042 - 0 / .013
t = 3.231

P value:
df = n -3 (3 variables in model)
df = 128

The p value based on these parameters:
p =  0.001568

Based on the p value threshold of .05
we reject the null hypothesis that being in a precint area adjacent
to yard signs does not affect vote share.
```

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

y = .302 + .042b1 + .042b2

b0 is the constant

b0 = (votes to Ken Cuccinelli) -
.042(Precinct assigned lawn signs + Precinct adjacent to lawn signs)

The constant coefficient is the value at which the regression model crosses the y axis.
This is known as the y-intercept.
When the variables b1 and b2 are equal to 0 the line intercepts the y axis at .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?

This model explains 9.4% of the variance in votes that went to the opponent Ken Cuccinel

This indicates that there are other variables that need to be explored to create a stronger model that predicts variance in vote share.