Problem Set 4

Applied Stats/Quant Methods 1

Due: November 18, 2024

Name: Eimhin ONeill
Student Number: 20332107

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 Monday November 18, 2024. 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`).

```
Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0)
print(Prestige$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 × dummy interaction.)

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

# get the summary
summary(prestige_model)
```

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

Prestige = 21.1422589 + 0.0031709 x Income + 37.7812800 x Professional - 0.0023257 x (Income x Professiomal)

(d) Interpret the coefficient for `income`.

Income coefficient = 0.0031709

So, for each unit increase in income for non professionals (when professional = 0), it increases their prestige by a really small amount

For a blue or white collar person to increase their prestige by 3.2 points, they would have to earn a 1000 dollars more

(e) Interpret the coefficient for `professional`.

Professional coefficient = 37.7812800

For anyone in a professional job, assuming the same income as a blue or white collar worker, they would have 37.78 more prestige points.

This shows that there is a much bigger effect for those in jobs seen as professional rather than their 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).

```r
# marginal effect is income coeff + interaction coeff

income_coeff <- coef(prestige_model)["income"]
interaction_coeff <- coef(prestige_model)["income:professional"]

marginal_effect_income <- income_coeff + interaction_coeff
print(marginal_effect_income)
# 0.0008452

0.0008452 * 1000
# 0.8452
```

In this case, Prestige only increases by roughly 0.9. This furthers my previous inclination that maybe it is not all about the money for professionals.

It seems the job titles could be doing more of that prestige signalling.

(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).

```r
# get the professional coeff
prof_coeff <- coef(prestige_model)["professional"]
# 37.7812800

# the effect is going to be the prof coeff + interaction coeff x 6000

mod_interaction_coeff <- interaction_coeff * 6000
# -13.95425

# so the effect is going to be
prof_coeff + mod_interaction_coeff
#  23.82703
```

In this case, Prestige increases by 23.82703. This shows us that if someone earns 6000 dollars and switches from white/blue collar to professional, their Prestige increases but the Income is still an inhibiting factor reducing the potential full effect.

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 $\alpha = .05$).

```
# 1 - assumptions
# normality, homoscedasticity, independence

# 2 - hypoth formulation
# null
# beta (lawn signs) = 0

# alternate
# beta (lawn signs) is not = 0
```

[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.

```r
# 3 - test stat
# precinct with yard signs has 30 samples
# coefficient
lawn_coeff <- 0.042

# standard error
lawn_se <- 0.016

# tstat
lawn_t <- lawn_coeff / lawn_se
# 2.625

# 4 - p - value
# so for the testing
# compare with p-value
# where df = 131 - 3 (number of explanatory variables here)
lawn_p <- 2 * pt(-abs(lawn_t), df = 128)
# 0.00972002
```

Conclusion

Since the calculated p-value is less than a = 0.05, we reject the null and see that lawn signs in yards have a statistically significiant effect on voteshare.

(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 $\alpha = .05$).

```r
# 1 - assumptions
# normality, homoscedasticity, independence

# 2 - hypoth formulation
# null
# beta (adjacent to lawn signs) = 0

# alternate
# beta (adjacent to lawn signs) is not = 0

# 3 - test stat
# precinct with yard signs has 30 samples
# coefficient
adj_lawn_coeff <- 0.042

# standard error
adj_lawn_se <- 0.013

# tstat
adj_lawn_t <- adj_lawn_coeff / adj_lawn_se
# 3.230769

# 4 - p - value
# so for the testing
# compare with p - value
# where df = 131 - 3 (number of predictors; lawn, adjacent, ctrl)
adj_lawn_p <- 2 * pt(-abs(adj_lawn_t), df = 128)
print(adj_lawn_p)
# 0.00156946
```

Conclusion

Since the calculated p-value is less than a = 0.05, we reject the null and see that being adjacent to precincts with lawn signs in yards have a statistically significiant effect on vote share

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

30.2 percent is the vote share for McAuliffe's competitors in when precincts do not have yard signs and are not adjacent to precincts with yard signs.

This interests us because it helps better understand the isolated effects of the explanatory variables any changes above or below 30.2 per cent are pretty much caused because of the variables

so if there are no yard signs and irrelevant whether close to precincts with yard signs, McAuliffes opponent will receive less than a third of the votes

If I was the opponent and did not know about the fit of the model, it shows that I would have to make the campaign visible and do my best to influence votes.

(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?

R Squared is 0.094

The model only explains 9.4 per cent of total variability in vote share suggesting to us that yard signs and adjacency have some influence but the most of the variation comes from other factors not included.

Yard signs and adjacency are statistically significant and do affect vote share but only in a minor way, as shown by the low R Squared.

Other factors play a larger part in determining voteshare could be economic climate (like interest rates, sovereign debt rates as seen currently in the US) could be precinct demographics (i.e. younger voters might vote more for mcauliffes opponents)

It could be down to strategy and other tools in the campaign toolkit like podcast appearances (which was supposedly a major decider in the recent US election).