Problem Set 4

Applied Stats/Quant Methods 1

Due: December 3, 2023

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 3, 2023. 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`).

I also drop any rows with NA although it doesn't really make a difference.

```
Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0)
Prestige_no_na <- na.omit(Prestige)View(Prestige_no_na)

```

(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.)

```
model_B <- lm(prestige ~ income + professional + income:professional ,
  data = Prestige_no_na)

```

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

Y = 21.1422589 + 0.0031709*x1 + 37.7812800*x2 + * -0.0023257(x1 * x2)

Y = Predicted value of prestige, 21.1422589 is the expected prestige score for a blue or white collar worker with an income of $0, x1 represents income in dollars,
for x2 a value of 1 signifies a professional and a value of 0 signifies a blue or white collared worker (reference catagory).

(d) Interpret the coefficient for `income`.

For blue or white collar workers the expected estimated effect of a one dollar increase in "income" is associate with an increase of 0.0031709 in prestige score.
There is a positive relationship between income and prestige score.

(e) Interpret the coefficient for `professional`.

When income is held constant, the estimated expected effect being "professional" compared to being "blue or white collar worker" is an increase of 37.7812800 in prestige score.

(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).

For a professional with an income income of \$1,000 the estimated prestige score is 59.76874. When income is \$2,000 the estimated prestige score is 60.61394
21.1422589 + 0.0031709 * 1000 + 37.7812800 * 1 + -0.0023257 * (1000*1) = 59.76874
21.1422589 + 0.0031709 * 2000 + 37.7812800 * 1 + -0.0023257 * (2000*1) = 60.61394

By subtracting the two prestige scores I can find the estimated differnce of what a \$1,000 increae in income makes.
60.61394-59.76874 = 0.8452
For a professional, an increase in income of \$1,000 is associated with an estimated increase in in prestige score of 0.8452. The above equation displays an increase from \$1,000 to \$2,000 but the effect on prestige would be the same for any increase of $1000 ex.0$ to $1000 or 1{,}001{,}000$ to $1,002,000 ect$

(g) What is the effect of changing one's occupations from being a blue or white collar worker 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).

This equation calculates the estimated prestige score for a professional who earns \$6,000
21.1422589 + 0.0031709 * 6000 + 37.7812800 * 0 + -0.0023257 * (6000*0) = 40.16766
This equation calculates the estimated prestige score for a blue or white collar worker who earns \$6,000
21.1422589 + 0.0031709 * 6000 + 37.7812800 * 1 + -0.0023257 * (6000*1) = 63.99474

I calculate the difference in prestige scores
63.99474 - 40.16766 = 23.82708
The estimated change of prestige score between blue or white collar workers and professionals when income held constant at \$6,000 is an increase of 23.82708 in prestige score.

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 effects vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

Ho: beta1 = 0 (slope equals zero)
The estimated effect of having yard signs in the precinct on vote share = 0
Ha: beta1 != 0 (slope does not equals zero)
The estimated effect of having yard signs in the precinct on vote share does not = 0

We want to know if yard signs effect vote share in any way (positive or negative) hence, lower.tail = FALSE and *2

```
t_stat_a <- 0.042/0.016 ## Test Statistic: coefficient/sd
df <- 128 # n - k - 1 (number of observations - number of variables - 1)
```

[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.

```
    two_a <- 2 * pt(t_stat_a, 128, lower.tail = FALSE)

```

the p-value is 0.00972002

Since the p value of 0.00972002 is smaller than the significance level of alpha = .05 we can reject the null hypothesis that the estimated effect of having yard signs in the precinct on vote share = 0. We find evidence to support the alternate hypothesis that the estimated effect of having yard signs in the precinct on vote share does not = 0

(b) Use the results to determine whether being next to precincts with these yard signs effects vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

Ho: beta1 = 0 (slope equals zero)
The estimated effect of vote share for a precinct adjacent to a precinct which displayed yard signs is 0.
Ha: beta1 != 0 (slope does not equals zero)
The estimated effect of vote share for a precinct adjacent to a precinct which displayed yard signs does not equal 0.

```
t_stat_b <- 0.042/0.013 # Test Statistic
two_b <- 2 * pt(t_stat_b, df, lower.tail = FALSE )

```

Since the p-value of 0.00156946 is lower than the significance level, alpha = .05, we can reject the Null Hypothesis that the estimated effect of vote share in a precinct adjacent to a precinct which displayed yard signs is 0. We find evidence to support the alternate hypothesis that the the estimated effect on vote share of a precinct adjacent to a precinct which displayed yard signs does not equal 0.

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

The coefficient for the constant term 0.302 is the estimated vote share for a precinct where no lawn signs were displayed and no adjacent precincts displayed lawn signs.

(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 covariate of R squared = 0.094 means that 9.4 percent of the variability of vote-share can be explained by the model. There is much unexplained variability indicating that there may be covariates which were not included. The model would fit better if more relevant variables were included.