Problem Set 4

Applied Stats/Quant Methods 1

Due: November 26, 2021

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 in `.pdf` form.
- This problem set is due before class on Friday November 26, 2021. No late assignments will be accepted.
- Total available points for this homework is 80.

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)

Additional:

install.packages("stargazer")
library(stargazer)
install.packages("emmeans", dependencies=TRUE)
library(emmeans)
```

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 the interaction of the two as predictors (Note: this is a continuous × dummy interaction.)

```

> #(b)
>
> #Does income level mean the job is more prestigious?
>
> model1 <- lm(prestige ~ income, data = Prestige)
> summary(model1)$coef
                Estimate   Std. Error  t value                          Pr(>|
   t|)
(Intercept) 27.141176368 2.2677036186 11.96857
   0.00000000000000000000005135229
income        0.002896799 0.0002833245 10.22432
   0.00000000000000000031920039791
>
> list1 <- list(income=0)
>
> emmeans(model1, ~income, at=list1)
 income emmean   SE  df lower.CL upper.CL
      0   27.1 2.27 100     22.6     31.6

Confidence level used: 0.95
>
> #Are professional jobs more prestigious?
>
> model2 <- lm(prestige ~ professional, data = Prestige)
>
> summary(model2)$coef
             Estimate Std. Error  t value
                          Pr(>|t|)
(Intercept)  37.83284   1.199386 31.54350
   0.00000000000000000000000000000000000000000000000001831567
professional 30.01555   2.132511 14.07522
   0.00000000000000000000000045058467519774240685527439225651 28
>
> list2 <- list(professional=30)
>
> emmeans(model2, ~professional, at=list2)
 professional emmean   SE df lower.CL upper.CL
           30    938 63.3 96      813     1064

Confidence level used: 0.95
>
> modX <- lm(data = Prestige, professional ~ income)
```

```
>
> stargazer(model1, model2, type = "text", title = "Regression Results")

Regression Results
================================================================================
                                        Dependent variable:
                                    ---------------------------
                                               prestige
                                          (1)                 (2)
--------------------------------------------------------------------------------
income                                  0.003***
                                        (0.0003)

professional                                                30.016***
                                                            (2.133)

Constant                                27.141***           37.833***
                                        (2.268)             (1.199)

--------------------------------------------------------------------------------
Observations                              102                 98
R2                                        0.511               0.674
Adjusted R2                               0.506               0.670
Residual Std. Error          12.090 (df = 100)   9.817 (df = 96)
F Statistic             104.537*** (df = 1; 100) 198.112*** (df = 1; 96)
================================================================================


```

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

(d) Interpret the coefficient for `income`.

The difference made is too small to be considered statistically significant.

(e) Interpret the coefficient for `professional`.

The difference made is much more significant.

***This is too much to try and grasp in 20 hours. Sorry–I am still playing catch up.

(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).

(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).

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] 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.

(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 $\alpha = .05$).

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

(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?