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`).

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

> professional
[1]   1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
[17]  1  1  1  1  1  1  1  1  1  1  1  0  1  1  0  1
[33]  0 NA  0  0  0  0  0  0  0  0  0  0  0  0  0  0
[49]  0  0  0  0 NA  0  0  0  0  0  0  0  0  0 NA  0
[65]  0  0 NA  0  0  0  0  0  0  0  0  0  0  0  0  0
[81]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1
[97]  0  0  0  0  0  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.)

```
lm(formula = Prestige$prestige ~ Prestige$income + professional +
Prestige$income:professional)

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

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

Residual standard error: 8.012 on 94 degrees of freedom
(4 observations deleted due to missingness)
Multiple R-squared:  0.7872,  Adjusted R-squared:  0.7804
F-statistic: 115.9 on 3 and 94 DF,  p-value: < 2.2e-16t...
```

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

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_3 X_i$

$\hat{y}_i = 21.1422589 + (0.0031709)X + (-0.0023257)$

(d) Interpret the coefficient for `income`.

`A unit increase in income is associated with a rise of 0.003 in the Pineo-Porter prestige score for occupation, holding all other variables constant.`

(e) Interpret the coefficient for `professional`.

`With the value of income held constant, the Pineo-Porter prestige score for occupation is 0.002 less for group 1 than group 0.`

(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).

$\hat{y} = 21.1422589 + (0.0031709)(1000) + (-0.0023257)$
$\hat{y} = 24.3377$

(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).

$\hat{y} = 21.1422589 + (0.0031709)(6000)$
$\hat{y} = 39.14$

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 $\alpha = .05$).

Testing for partial effect of variable 1 on y (proportion of the vote that went to Cuccinelli).
$H_o : \beta_1 = 0$
$H_a : \beta_1 \neq 0$

[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.

$t = \frac{\beta_i}{se}$
$df = n - (k+1) = n - 3 = 131 - 3$
$df = 128$
$t = \frac{\beta_1}{se} = \frac{0.042}{0.016}$
$t = 2.625$

Finding the p-value in R:

```
      > p_value <- pt(q = 2.65, df= 128)
      > p_value
       [1] 0.9954678content...

```

P-value is greater than 0.05 and therefore we cannot reject null.

(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 $\alpha = .05$).
$H_o : \beta_2 = 0$
$H_a : \beta_2 \neq 0$ $t = \frac{\beta_i}{se}$
$df = n - (k+1) = n - 3 = 131 - 3$
$df = 128$
$t = \frac{\beta_2}{se} = \frac{0.042}{0.013}$
$t = 3.23$

Finding the p-value in R:

```
      > p_value2 <- pt(q = 3.23, df = 128)
      > p_value2
       [1] 0.9992133

```

P-value is greater than 0.05 and therefore we cannot reject null.

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

The constant is the mean value of the response variable when all the predictor variables equal zero. In other words the y-intercept.

(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?

```
0.094:  a relatively low r-squared value, suggests only a small amount of
variability is explained by the model.  This suggests that other variables
contributed to the variation in the response variable - vote share.
```