Problem Set 4

Applied Stats/Quant Methods 1: Jack O'Neill

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)
```

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`.)

```
#Create new variables 'professsional'
library(dplyr)
Prestige$professional <- recode(Prestige$type, prof = "1", bc = "0", wc =
    "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 with prestige as outcome and income, professional and interaction as
    predictors
interact_reg <- lm(data = Prestige, prestige ~ income + professional +
    income*professional)
summary(interact_reg)
```

```
(Intercept)               income        professional1 income:professional1
21.142258854          0.003170909         37.781279955         -0.002325709
```

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

$$\hat{Y_i} = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{}_2 D_i + \beta_3 X_i D_i + \epsilon_i \tag{1}$$

Using the coefficients, the equation is:

$$\hat{Y_i} = 21.1422589 + 0.0031709Xi + 37.7812800 - 0.0023257 * D + \epsilon_i \tag{2}$$

or:

$$\hat{(}Y) = 58.9235389 + (0.0008452)X_i \tag{3}$$

(d) Interpret the coefficient for `income`.

```r
coef(interact_reg)
```

The coefficient for "income" $\hat{\beta}1$ is 0.0031709.

This means that keeping the level of profession constant, with a one unit increase in income, the expected value of prestige increases by 0.0031709 unit.

(e) Interpret the coefficient for `professional`.

The coefficient for "professional" $\hat{\beta}2$ is 37.781279955.

This means that when income equals 0 (at the intercept), the average value of prestige is 37.78 units higher for the group with professional = 1 (professionals) than the group with professional = 0 (blue collar and white collar)

(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).

Allowing (Xi) to equal \$1,000 our equation is:

$$\hat{(Y)} = 58.9235389 + (0.0008452) * 1000\hat{(Y)} = 59.7687389 \tag{4}$$

Therefore the score of $\hat{(Y)}$ in for professionals is 59.7687389. To get a marginal effect of \$1000 increase we let income = \$0 which is the effect of the intercept and belonging to the category of professionals, which is 58.9235389.

Therefore the marginal effect of additional \$1000 on prestige is equals to:

$MEF1 = 59.7687389 - 58.9235389$ $MEF1 = 0.8452$

Therefore an additional \$1000 in income gives 0.8452 points on in prestige.

(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).

The marginal effect of changing the value of D from 0 to 1 will be 37.781-0.002x. If x is equal to 6,000 , th emarginal change in $\hat{y}$ =

37.781 - 0.002*6000 = 23.827.

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 $\alpha = .05$).

If there is no impact of yard signs on vote share then the coefficient of $\hat{\beta}2$ would have the value zero.

Our null hypothesis is - H0: $\hat{\beta}2 = 0$ Our alternative hypothesis is - Ha:$\hat{\beta}2 \neq 0$

We then use a t test to calculate our t statistic.

[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.

$t = (\hat{\beta}2 - 0)/\hat{\sigma}(\hat{\beta}2)$

$t = (0.042 - 0)/0.016$

$t = 2.625$

This value allows us to calculate the p-value and, using the n-p df, where p = to the number of parameters of the regression model (3).

```
p1 <- 2*pt(2.625, 128, lower.tail = F)
p1
```

Our p-value = 0.00972002

If $\alpha = 0.05$, there is evidence to support our alternative hypothesis, that having these yard signs in a precinct affects 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 $\alpha = .05$).

Our null hypothesis is - H0: $\hat{\beta}3 = 0$ Our alternative hypothesis is - Ha:$\hat{\beta}3 \neq 0$

$t = (\hat{\beta}2 - 0)/\hat{\sigma}(\hat{\beta}2)$

$t = (0.042 - 0)/0.013$

$t = 3.23076923077$

```
p1 <- 2*pt(2.625, 128, lower.tail = F)
p1
```

Our p-value = 0.00156946

If $\alpha = 0.05$, there is evidence to support our alternative hypothesis, being next to precints with yard signs affects vote share.

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

When we consider the regression equation

$\hat{Y} = \hat{\beta}0 + \hat{\beta}1Xi + \hat{\beta}2Xs$

where the response variable is the proportion of the vote that went to Ken Cuccinelli, and the response variables are Xi = a precinct was randomly assigned to have the sign against McAuliffe posted and Xs = precinct that was beside a precinct in the treatment group. Therefore, the constant term in the equation represents the coefficient of the intercept.

This is equal to 0.302. This means that on average, in a precinct that is neither part of the treatment nor adjacent to one, 30.2% voted for Ken Cuccinelli.

This would point to a small effect of the lawn sign on vote share, supported by the small R^2.

(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^2 gives us a measure of model fit. The R^2 value in the table is 0.094. This indicates that a relatively small amount (9.4%) of variance of the response variable is explained by the explanatory variables.

The value of coefficients provided in the table, and their corresponding p-values that they significantly impact Cuccinelli's votes, but the value of R^2 shows us that even though the effect of lawn sign is 'statistically' significant it represents a small portion of variation. Therefore, roughly 90% of residual variance remains unexplained.