Problem Set 4

Vanessa Wong

QTM 200: Applied Regression Analysis

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 the course GitHub page in `.pdf` form.
- This problem set is due at the beginning of class on Monday, February 24, 2020. No late assignments will be accepted.
- Total available points for this homework is 100.

Question 1 (50 points): 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$type
Prestige$profession <- ifelse(profession == "prof", 1, 0)
Prestige$profession
```

(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.)

```
# y = prestige (outcome variable)
# x = income, professional, interaction
reg1b <- lm(prestige ~ income + profession + income:profession, data=
    Prestige)
summary(reg1b)
```

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

1. $\hat{y} = 21.1422 + 0.00317x_{income} + 37.7813x_{profession} - 0.00233x_{income}x_{profession}$

(d) Interpret the coefficient for `income`.

- For blue and white collar workers, a one unit increase in income (one dollar increase) is associated with a 0.00317 unit increase in prestige.

(e) Interpret the coefficient for `professional`.

- On average, holding all other variables constant at their means, professionals have 37.813 more units of prestige than blue and white collar workers.

(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).

- The expected change in prestige score for professional occupations as a result of a \$1,000 increase in income is 0.84.

```
# Professional coded as 1
# expected prestige score = (21.1422 + 37.7813) + (0.00317 + (-0.00233))x
upper <- (21.1422 + 37.7813) + ((0.00317 + (-0.00233))*1000)
lower <- (21.1422 + 37.7813)
upper - lower
```

(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 expected change in prestige score for an individual whose income is \$6,000 and is changing from a non-professional to a professional occupation is 23.8013.

```
# Professional coded as 1
# Non-professional coded as 0
# prediction equation: y = 21.1422 + 0.00317xincome + 37.7813xprofession
   - 0.00233xincomexprofessionx
ynonprof <- 21.1422 + (0.00317*6000)
yprof <- 21.1422 + (0.00317*6000) + (37.7813*1) - (0.00233*6000*1)
yprof - ynonprof
```

Question 2 (50 points): 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 to determine whether having these yard signs in a precinct affects vote share (e.g., conduct a hypothesis test with $\alpha = .05$).

- p = 0.0137 < 0.05. Therefore, we reject the null hypothesis that having yard signs in a precinct does not affect vote share.

```
ts <- (0.042)/(0.016)
2*pt(ts, df=29, lower.tail=FALSE)
```

[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 $\alpha = .05$).

- p = 0.0105 < 0.05. Therefore, we reject the null hypothesis that being next to precincts with yard signs does not affect vote share.

```
ts2 <- (0.042)/(0.013)
2*pt(ts, df=75, lower.tail=FALSE)
```

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

- In precincts that are not assigned to have yard signs posted and are not adjacent to precincts with yard signs, Ken Cuccinelli's predicted vote share is 0.302.

(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?

- 9.4% of the variation in vote share is explained by yard signs in a precint. This is a relatively low percentage, indicating that the yard signs have little explanatory power relative to other factors that are not modeled – 90.6% of the variation in Cucinelli's voteshare is explained by other factors.

```
#######################
# load libraries
# set wd
# clear global .envir
#######################

# remove objects
rm(list=ls())
# detach all libraries
detachAllPackages <- function() {
  basic.packages <- c("package:stats", "package:graphics", "package:grDevices
    ", "package:utils", "package:datasets", "package:methods", "package:base")
  package.list <- search()[ifelse(unlist(gregexpr("package:", search()))==1,
    TRUE, FALSE)]
  package.list <- setdiff(package.list, basic.packages)
  if (length(package.list)>0)  for (package in package.list) detach(package,
    character.only=TRUE)
}
detachAllPackages()

# load libraries
pkgTest <- function(pkg){
  new.pkg <- pkg[!(pkg %in% installed.packages()[,  "Package"])]
  if (length(new.pkg))
    install.packages(new.pkg,  dependencies = TRUE)
  sapply(pkg,  require,  character.only = TRUE)
}
install.packages(car)
library(car)
data(Prestige)
help(Prestige)


# here is where you load any necessary packages
# ex: stringr
# lapply(c("stringr"),  pkgTest)

lapply(c("stringr"),  pkgTest)

# set working directory, import datasets
setwd("~/GitHub/QTM200Spring2020/problem_sets/PS4")


#######################
# Problem 1:
#######################
Prestige$type
Prestige$profession <- ifelse(profession == "prof", 1, 0)
Prestige$profession
#######################
# y = prestige (outcome variable)
```

```
# x = income, professional, interaction
reg1b <- lm(prestige ~ income + profession + income:profession, data=Prestige)
summary(reg1b)
########################
# prediction equation: y = 21.1422 + 0.00317xincome + 37.7813xprofession -
    0.00233xincomexprofessionx
########################
# For blue and white collar workers, a one unit increase in income (one dollar
     increase) is associated with a 0.00317 unit increase in prestige.
########################
# On average, holding all other variables constant at their means,
    professionals have 37.813 more units of prestige than blue and white
    collar workers.
########################
# Professional coded as 1
# expected prestige score = (21.1422 + 37.7813) + (0.00317 + (-0.00233))x
upper <- (21.1422 + 37.7813) + ((0.00317 + (-0.00233))*1000)
lower <- (21.1422 + 37.7813)
upper - lower
# expected change in prestige score for professional occupations as a result
    of a $1,000 increase in income is 0.84.
########################
# Professional coded as 1
# Non-professional coded as 0
# prediction equation: y = 21.1422 + 0.00317xincome + 37.7813xprofession -
    0.00233xincomexprofessionx
ynonprof <- 21.1422 + (0.00317*6000)
yprof <- 21.1422 + (0.00317*6000) + (37.7813*1) - (0.00233*6000*1)
yprof - ynonprof
# the expected change in prestige score for an individual whose income is $6
    ,000 and is changing from a non-professional to a professional occupation
    is 23.8013.
########################

########################
# Problem 2:
########################
ts <- (0.042)/(0.016)
2*pt(ts, df=29, lower.tail=FALSE)
# p = 0.0137 < 0.05. Therefore, we reject the null hypothesis that having yard
     signs in a precinct does not affect vote share.
########################
ts2 <- (0.042)/(0.013)
2*pt(ts, df=75, lower.tail=FALSE)
# p = 0.0105 < 0.05. Therefore, we reject the null hypothesis that being next
    to precincts with yard signs does not affect vote share.
########################
#In precincts that are not assigned to have yard signs posted and are not
    adjacent to precincts with yard signs, Ken Cuccinelli's predicted vote
    share is 0.302.
########################
```

```
88 # 9.4% of the variation in vote share is explained by yard signs in a precint.
       This is a relatively low percentage, indicating that the yard signs have
      little explanatory power relative to other factors that are not modeled --
       90.6% of the variation in Cucinelli's voteshare is explained by other
      factors.
```