

Answer Key: Problem Set 4

QTM 200: Applied Regression Analysis

Jeffrey Ziegler

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$professional <- as.factor(ifelse(Prestige$type=="prof", 1, 0))

table(Prestige$professional)
0  1
67 31
```

- (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 \times dummy interaction.)

```
m1 <- lm(prestige ~ income*professional, data=Prestige)

summary(m1)
Call:
lm(formula = prestige ~ income * professional, data = Prestige)
Residuals:
Min      1Q  Median      3Q      Max
-14.852  -5.332  -1.272   4.658  29.932
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    21.1422589    2.8044261    7.539 2.93e-11 ***
income           0.0031709    0.0004993    6.351 7.55e-09 ***
professional    37.7812800    4.2482744    8.893 4.14e-14 ***
income:professional -0.0023257    0.0005675   -4.098 8.83e-05 ***
---
Signif. codes:  ***0.001 **0.01 *0.05
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-16
```

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

$$E[Y] = 21.142 + 0.003 * Income + 37.781 * Professional - 0.002 * Income * Professional$$

(d) *Interpret the coefficient for **income**.*

When one's occupation is NOT professional, one unit (\$1) increase in income leads to a 0.003 increase in prestige score.

(e) *Interpret the coefficient for **professional**.*

When he or she earns nothing, changing one's occupations from non-professional to professional increases prestige score by 37.781 (In reality, this does not happen).

(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).*

$$\begin{aligned}\Delta E[Y] &= 0.003 \times Income + 37.781 * Professional - 0.002 \times Income \times Professionals \\ &= 0.003 \times 1000 + 37.781 \times 1 - 0.002 \times 1000 \times 1 \\ &= 38.781\end{aligned}$$

A \$1,000 increase in income for professionals leads to a 38.78 increase in prestige score.

- (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).*

$$\begin{aligned}\Delta E[Y] &= 0.003 \times \text{Income} + 37.781 \times \text{Professional} - 0.002 \times \text{Income} \times \text{Professional} \\ &= 0.003 \times 6000 + 37.781 \times 1 - 0.002 \times 6000 \times 1 \\ &= 43.781\end{aligned}$$

A change from non-professional to professional leads to a 43.781 increase in prestige score when her income is held to \$6,000.

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 Virginia 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 McAuliffe'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).

- (a) *Use the results to determine whether having these yard signs in a precinct affects vote share (e.g., conduct a hypothesis test with $\alpha = .05$).*

For assigned lawn signs:

$$\frac{\text{Estimate}}{\text{SE}} = 0.042 / (0.016) = 2.625$$

$$2 * pt(2.625, df = 128, lower.tail = F) = 2 * 0.00486 = 0.00972$$

You can reject the null of no effect and say extra yard signs increases vote share by 0.042.

¹Donald 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.

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: $R^2=0.094$, $N=131$

- (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 $\alpha = .05$).*

For adjacent to lawn signs:

$$\frac{\text{Estimate}}{\text{SE}} = 0.042 / (0.013) = 3.230$$

$$2 * \text{pt}(3.230, \text{df} = 128, \text{lower.tail} = \text{F}) = 2 * 0.00078668 = 0.00157$$

Being next to precincts with more yard signs have the same effect on the vote share (exactly same coefficient), and the standard error is smaller (so we have a lower p value) since there are more adjacent cases than districts with assigned lawn signs.

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

The average vote share for McAuliff in a district with no assigned lawn signs and not adjacent to lawn signs is expected to be 0.302 and this constant term seems to be significant if we calculate the t-value.

- (d) *Evaluate the model fit for this regression. What does this tell us about the importance of yard signs versus other factors?*

The R^2 is quite small. The yard signs (and adjacency) only explain less than 10% of the variation in the vote share. There are other factors that affect vote share for sure.