# Problem Set 4 - My Answer

Junjie LIU[1,*]

03 December, 2023

[1] Department of Political Science, Trinity College Dublin, 2 Clare, Street, Dublin 2, Ireland

[*] Correspondence: Junjie LIU <liuj13@tcd.ie>

## 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 3, 2023. 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`).

(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.)

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

(d) Interpret the coefficient for `income`.

(e) Interpret the coefficient for `professional`.

(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).

My Answer:

```
#install.packages("car")
## Question 1-a
library(car)
```

```
## Loading required package: carData
```

```
data(Prestige)

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

```
##
##  0  1
## 67 31
```

```
## Question 1-b
model <- lm(Prestige$prestige ~ (Prestige$income * Prestige$professional))
summary(model)
```

```
##
## Call:
## lm(formula = Prestige$prestige ~ (Prestige$income * Prestige$professional))
##
## 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
## Prestige$income                          0.0031709  0.0004993   6.351 7.55e-09
## Prestige$professional                   37.7812800  4.2482744   8.893 4.14e-14
## Prestige$income:Prestige$professional -0.0023257  0.0005675  -4.098 8.83e-05
##
## (Intercept)                             ***
## Prestige$income                         ***
## Prestige$professional                   ***
## Prestige$income:Prestige$professional ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## 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
```

For the (c), based on the regression model from (b), we can find that the prediction equation can be denoted as:

$$E[Y] = 21.142 + 0.003 \cdot \text{Income} + 37.781 \cdot \text{Processional} - 0.002 \cdot \text{Income} \cdot \text{Professional}$$

For the (d), based on the (c)'s equation, if the occupation is not professional, a \$1 increase in income leads to a 0.003 increase in prestige score.

For the (e), based on the (c)'s equation, if someone who change one's occupations from non-professional to professional, it will leads to a 37.781 increases in its prestirage.

For the (f), based on the (c)'s equation:

$$E[Y] = 21.142 + 0.003 \times 1000 + 37.781 \times 1 - 0.002 \times 1000 \times 1 = 59.923$$

Thus, the difference is 59.923.

For the (g), based on the (c)'s equation:

$$\text{Prof.} = 6000, \text{Income} = 1 : E[Y] = 21.142 + 0.003 \times 6000 + 37.781 \times 1 - 0.002 \times 6000 \times 1 = 64.923$$
$$\text{Prof.} = 6000, \text{Income} = 0 : E[Y] = 21.142 + 0.003 \times 6000 + 37.781 \times 0 - 0.002 \times 6000 \times 0 = 39.142$$

The difference when Income changes from 0 to 1 is 25.781.

## 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

[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.

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 $\alpha = .05$).

(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?

Answer of Q2:

```
estimates <- c(0.042,0.042,0.302)
std_errors <- c(0.016, 0.013, 0.011)

t_score <- (estimates - 0) / (std_errors)
p_val <- 2 * pt(abs(t_score) , df = (131 - 3), lower.tail = F)

print(t_score)
```

```
## [1]  2.625000  3.230769 27.454545
```

```
print(p_val)
```

```
## [1] 9.720020e-03 1.569460e-03 1.738775e-55
```