

# Problem Set 4

Applied Stats/Quant Methods 1

Yu Fan/24344707

## 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 Monday November 18, 2024. 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**).

```
1 install.packages("car")
2 library(car)
3
4 # Loading the Prestige dataset
5 data(Prestige)
6 help(Prestige)
7
8 # Create professional
9 Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0)
10
11 # View new variables in the dataset
12 head(Prestige)
```

	education	income	women	prestige	census	type	professional
gov.administrators	13.11	12351	11.16	68.8	1113	prof	1
general.managers	12.26	25879	4.02	69.1	1130	prof	1
accountants	12.77	9271	15.70	63.4	1171	prof	1
purchasing.officers	11.42	8865	9.11	56.8	1175	prof	1
chemists	14.62	8403	11.68	73.5	2111	prof	1
physicists	15.64	11030	5.13	77.6	2113	prof	1

- (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.)

```
1 # Running a linear model
2 model <- lm(prestige ~ income * professional, data = Prestige)
3
4 # Output results
5 summary(model)
```

```

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:
(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 '***' 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

```

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

Based on the output of the linear model, we can write the prediction equation. The form of the equation is as follows:

$$\text{prestige} = \beta_0 + \beta_1 \times \text{income} + \beta_2 \times \text{professional} + \beta_3 \times (\text{income} \times \text{professional})$$

Substituting the estimated coefficient values into the equation:

$$\text{prestige} = 21.1423 + 0.0032 \times \text{income} + 37.7813 \times \text{professional} - 0.0023 \times (\text{income} \times \text{professional})$$

So: When professional = 0 (blue-collar or white-collar workers):

$$\text{prestige} = 21.1423 + 0.0032 \times \text{income}$$

When professional = 1 (professionals):

$$\text{prestige} = 21.1423 + 37.7813 + (0.0032 - 0.0023) \times \text{income} = 58.9236 + 0.0009 \times \text{income}$$

Thus, these two equations show how income and occupation type affect occupational prestige.

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

Coefficient of income (income = 0.0032): This coefficient indicates that, controlling for other variables (including occupation type and interactions), for every unit increase in income, the mean value of prestige increases by about 0.0032. This suggests that there is a positive correlation between income and prestige, i.e., as income increases, prestige increases slightly.

If the occupation type is Blue Collar or White Collar Worker (professional = 0), occupational prestige increases by 0.0032 units for each additional unit of income. If the occupation type is Professional (professional = 1), the effect of income on occupational prestige is adjusted by the interaction term, but in general, income still has a positive effect on occupational prestige. This coefficient is positive and statistically significant (p-value  $< 0.001$ ), indicating that income is a significant predictor variable of occupational prestige.

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

Coefficient of Occupational Type (professional = 37.7813): This coefficient indicates that the mean value of occupational prestige is 37.7813 units higher for professionals (professional = 1) than for blue-collar or white-collar workers (professional = 0), controlling for other variables, including income and interaction terms. This coefficient shows a significant effect of occupation type on occupational prestige, i.e. professionals have significantly higher occupational prestige than blue- or white-collar workers, even at the same income level.

- (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).

Based on the prediction equation from part (c), when  $\text{professional} = 1$ , the marginal effect of income on the prestige score is 0.0009. This means that for each additional unit of income, the prestige score increases by 0.0009 units.

For a \$1,000 increase in income, the change in the prestige score can be calculated as:  $0.0009 \times 1000 = 0.9$

Therefore, for professional occupations, a \$1,000 increase in income is associated with a 0.9-unit increase in the 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).

When an individual's income is \$6,000, changing their occupation from non-professional to professional results in an impact on the prestige score that can be calculated using the coefficients from part (c). The direct effect of switching to a professional job is given by the coefficient of **professional**, which is 37.7813. However, because of the interaction term **income:professional** (-0.0023), we need to account for the effect of income.

At an income level of \$6,000, the interaction effect is:

$$-0.0023 \times 6,000 = -13.8$$

Therefore, the total change in the prestige score is:

$$37.7813 - 13.8 = 23.9813$$

Thus, at an income of \$6,000, switching from a non-professional to a professional occupation increases the prestige score by approximately 23.98 units.

## Question 2: Political Science

Researchers are interested in learning the effect of all of those yard signs on voting preferences.<sup>1</sup> 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).

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*

---

<sup>1</sup>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.

- (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  $\alpha = .05$ ).

To determine whether having the yard signs in a precinct significantly affects vote share, we can conduct a hypothesis test for each coefficient using a significance level of  $\alpha = 0.05$ .

#### Hypothesis Tests

1. For Precinct Assigned Lawn Signs: - Null Hypothesis ( $H_0$ ): The coefficient for precincts assigned lawn signs is equal to zero, meaning the signs have no effect on vote share. - Alternative Hypothesis ( $H_a$ ): The coefficient for precincts assigned lawn signs is not equal to zero, meaning the signs do affect vote share. - Coefficient Estimate: 0.042 - Standard Error: 0.016 - t-value Calculation:  $0.042 / 0.016 = 2.625$  - Using a t-distribution table or a statistical software, we can find the p-value associated with  $t = 2.625$  for a two-tailed test. The p-value is less than 0.05, so we reject the null hypothesis. Conclusion: The yard signs significantly affect the vote share in precincts where they are posted.

2. For Precinct Adjacent to Lawn Signs: - Null Hypothesis ( $H_0$ ): The coefficient for precincts adjacent to lawn signs is equal to zero, meaning there is no effect on vote share. - Alternative Hypothesis ( $H_a$ ): The coefficient for precincts adjacent to lawn signs is not equal to zero, meaning there is an effect on vote share. - Coefficient Estimate: 0.042 - Standard Error: 0.013 - t-value Calculation:  $0.042 / 0.013 = 3.231$  - The p-value associated with  $t = 3.231$  is also less than 0.05, so we reject the null hypothesis. Conclusion: The yard signs significantly affect the vote share even in adjacent precincts.

Final Conclusion Both coefficients are statistically significant at the  $\alpha = 0.05$  level, suggesting that having these yard signs in a precinct, or even in an adjacent precinct, significantly affects the 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  $\alpha = .05$ ).

To determine whether being next to precincts with these yard signs affects vote share, we will conduct a hypothesis test for the coefficient of the "Precinct Adjacent to Lawn Signs" variable using a significance level of  $\alpha = 0.05$ .

#### Hypothesis Test for Precinct Adjacent to Lawn Signs

1. Null Hypothesis ( $H_0$ ): The coefficient for precincts adjacent to lawn signs is equal to zero, meaning there is no effect on vote share. 2. Alternative Hypothesis ( $H_a$ ): The coefficient for precincts adjacent to lawn signs is not equal to zero, meaning there is an effect on vote share.

Given Values - Coefficient Estimate: 0.042 - Standard Error: 0.013

Calculation - t-value:  $0.042 / 0.013 = 3.231$

Conclusion - Using a t-distribution table or statistical software, the p-value associated with  $t = 3.231$  for a two-tailed test is less than 0.05. - Since the p-value is less than  $\alpha = 0.05$ , we reject the null hypothesis.

Final Conclusion: Being next to precincts with these yard signs significantly affects vote share at the  $\alpha = 0.05$  level.

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

The coefficient for the constant term (0.302) represents the baseline proportion of the vote share that went to McAuliffe's opponent, Ken Cuccinelli, when both of the independent variables (precincts assigned lawn signs and precincts adjacent to lawn signs) are equal to zero.

Substantive Interpretation In practical terms, the constant term indicates that in precincts where there were no lawn signs posted and which were not adjacent to precincts with lawn signs, the baseline vote share for Cuccinelli is 30.2 percent. This provides a reference point for comparing how the presence of lawn signs or proximity to precincts with lawn signs affects the vote share.

- (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?

To evaluate the model fit for this regression, we can look at the  $R^2$  value provided in the regression results.

Model Fit:  $R^2 = 0.094$ : This indicates that only 9.4

Interpretation The low R-squared value suggests that the model does not explain much of the variability in vote share. This implies that yard signs, while statistically significant, account for only a small portion of the overall variation in voting preferences. Other factors not included in the model, such as demographic characteristics, political affiliation, campaign strategies, or broader socio-economic influences, likely play a more substantial role in determining vote share.