

# Problem Set 4

## Applied Stats/Quant Methods 1

Due: November 18, 2024

### 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 and load the car package
2 install.packages("car")
3 library(car)
4 data(Prestige)
5 help(Prestige)
6
7 # Create the professional variable
8 Prestige$professional <- ifelse(Prestige$type == "prof", 1, 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  $\times$  dummy interaction.)

```
1 # Fit the linear model
2 model <- lm(prestige ~ income * professional, data = Prestige)
3 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:

	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 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.012 on 94 degrees of

(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.

$$\hat{y} = 21.1422589 + 0.0031709 * \text{income} + 37.7812800 * \text{professional} + -0.0023257 * (\text{income} * \text{professional})$$

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

The coefficient for **income**: Represents the change in prestige score associated with a \$1 increase in income, for non-professional jobs (professional=0).

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

The coefficient for **professional**: Represents the difference in prestige scores between professionals (professional=1) and non-professionals (professional=0), when income = 0.

- (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).

When **professional**=1, the marginal effect of income is:

$$\frac{\partial \hat{y}}{\partial \text{income}} = 0.0031709 - 0.0023257$$

For a \$1,000 increase in income ( $\Delta \text{income} = 1000$ ) :

$$\Delta \hat{y} = 1000 * (0.0031709 - 0.0023257)$$

- (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).

Substitute **income**=6000 and calculate the change in  $\hat{y}$  for switching from non-professional (**professional**=0) to professional (**professional**=1):

$$\Delta \hat{y} = 37.7812800 + 6000 * -0.0023257$$

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

- (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 if yard signs significantly affect vote share, conduct a hypothesis test for the coefficient of  $X_1$ (assigned lawn signs):

Null and Alternative Hypotheses:

$$H_0 : \beta_1 = 0 \quad (\text{Yard signs have no effect on vote share})$$

$$H_a : \beta_1 \neq 0 \quad (\text{Yard signs have an effect on vote share})$$

---

<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.

Test Statistic:

$$t = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} = \frac{0.042}{0.016} = 2.625$$

Critical Value:

Using  $\alpha = 0.05$ , the two-tailed critical value of t for approximately df=129 (large sample size) is approximately  $\pm 1.96$ .

Decision:

Since  $|t| = 2.625 > 1.96$ , reject  $H_0$ . This means the effect of yard signs in precincts is statistically significant at  $\alpha = 0.05$ .

- (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  $\alpha = .05$ ).

Repeat the hypothesis test for  $X_2$ (adjacent precincts with yard signs):

Null and Alternative Hypotheses:

$$H_0 : \beta_2 = 0 \quad (\text{Adjacent precincts have no effect on vote share})$$

$$H_a : \beta_2 \neq 0 \quad (\text{Adjacent precincts have an effect on vote share})$$

Test Statistic:

$$t = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)} = \frac{0.042}{0.013} = 3.231$$

Critical Value:

Same as before  $\pm 1.96$ .

Decision:

Since  $|t| = 3.231 > 1.96$ , reject  $H_0$ . The effect of adjacent precincts is also statistically significant at  $\alpha = 0.05$ .

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

The constant term  $\hat{\beta}_0 = 0.302$  represents the expected proportion of votes for Cucinelli in precincts without yard signs and not adjacent to yard signs.



Substantively:

In a baseline scenario (no exposure to yard signs), Cuccinelli is expected to receive 30.2% of the vote in these precincts.

(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?

- $R^2=0.094$  indicates that only 9.4% of the variation in vote share is explained by the presence of yard signs (assigned or adjacent precincts).
- The remaining 90.6% of variation is due to other factors not captured in the model.

Implications:

- While yard signs have a statistically significant effect, the practical significance is modest because most of the variation in vote share is driven by factors outside this model.
- This suggests that other factors are likely much more important than yard signs in determining voting outcomes.