

# 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 Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0)
2
3 Prestige$professional[is.na(Prestige$type)] <- 0
4
5 head(Prestige[, c("type", "professional")])
```

- (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 model <- lm(prestige ~ income * professional, data = Prestige)
2
3 summary(model)
4
5
6 install.packages("stargazer")
```

Table 1: Regression Results

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	20.8035	2.5387	8.194	$9.74 \times 10^{-13}$ ***
Income	0.0033	0.0005	7.161	$1.49 \times 10^{-10}$ ***
Professional	38.1200	4.0798	9.344	$3.22 \times 10^{-15}$ ***
Income x Professional	-0.0024	0.0005	-4.570	$1.42 \times 10^{-5}$ ***

*Significance codes:* 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.018 on 98 degrees of freedom

Multiple R-squared: 0.7893, Adjusted R-squared: 0.7828

F-statistic: 122.3 on 3 and 98 DF, p-value:  $< 2.2 \times 10^{-16}$

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

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

$$\text{prestige} = 20.8035 + 0.0033 \cdot \text{income} + 38.1200 \cdot \text{professional} - 0.0024 \cdot (\text{income} \cdot \text{professional}) + \epsilon$$

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

For non-professionals, meaning blue-and white collar workers (where "Professional" is equal to 0), a one-unit increase in "Income" is associated with a 0.0033 unit increase of the "Prestige" scale. As "Professional" is a dummy variable, this interpretation applies specifically when "Professional" is held constant at 0.

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

The coefficient of "Professional" represents an increase of 38.1200 units of "Prestige" for professionals ("Professional" = 1) compared to white-and blue collar workers ("Professional" = 0), while "Income" is held constant.

- (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 interaction coefficient of "IncomeXProfessional" is -0.0024, which represents the change in the effect of "Income" on "Prestige" when switching from 0 (non-professionals) to 1 (professionals). To now calculate the effect of "Income" on "Prestige" for professionals, one must summarise the interaction coefficient, with the coefficient for "Income":  $\beta_1 + \beta_3 = 0.0033 - 0.0024 = 0.0009$ . This value represents, that for professionals (Professional = 1), a one-unit increase in "Income" results in a 0.0009 unit increase in "Prestige". To now calculate the change in "Prestige" for a 1000USD increase in "Income" for professionals, this value of 0.0009 must be multiplied by 1000:  $0.0009 \times 1000 = 0.9$ . This value can be interpreted as follows: A 1000USD increase in "Income" for professionals, is associated with a 0.9 unit increase in "Prestige".

- (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 regression output above shows a regression coefficient of  $\beta_2 = 38.1200$ , meaning that holding "Income" constant, a switch from non-professional (Professional=0) to professional (Professional=1), is associated with a 38.1200 increase in "Prestige". And as described above, the interaction coefficient  $\beta_3 = -0.0024$ , represents the change in

the effect of "Income" on "Prestige", when switching from non professionals (Professional=0) to professional (Professional = 1), meaning the effect of "Income" on "Prestige" is slightly lower for professionals, compared to non-professionals. Now, to calculate the effect of switching from a non-professional to a professional job when income=6000USD, one must summarise the coefficients of both "Professional" and the interaction term, and multiply this by the value for income:  $\beta_2 + \beta_3 X_{\text{income}}$ . Resulting in the following formula:  $38.1200 + (-0.0024) \times 6000 = 23.7200$ . This value can be interpreted as follows: Switching from a non-professional job to a professional job, when "Income" is 6000USD, is associated with an increase of 23.72 units in "Prestige".

## 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  $\alpha = .05$ ).

H0: The anti-McAuliffe signs in a precinct do not impact the voteshare for Cuccinelli ( $\beta_1=0$ ).

HA: The anti-McAuliffe signs in a precinct do have an impact on the voteshare for Cuccinelli ( $\beta_1$  not equal to 0).

$$t\text{-statistic}=\beta_1/SE\beta_1=0.042/0.016=2.625$$

$$df=N-k-1=131-2-1=128$$

For a two-tailed test, the critical value of  $\alpha=0.05$ , is approximately 1.96. Since  $t=2.625$  is larger than 1.96, we have evidence to reject the Null hypothesis. Hence, the presence of yard signs in the precinct, has a statistically significant effect on the vote share at the 5% significance level.

---

<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.

- (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  $\alpha = .05$ ).

N0: Living adjacent to precinct with anti-McAuliffe yard signs does not impact the voteshare for Cuccinelli ( $\beta_2=0$ ).

HA: Living adjacent to precinct with anti-McAuliffe yard signs does impact the vote-share for Cuccinelli ( $\beta_2$  not equal to 0).

t-statistic:  $\beta_2/SE\beta_2=0.042/0.013=3.231$

Since the t-statistic of 3.231 is larger than the critical value of 1.96, we have evidence to reject the null hypothesis for the adjacent precincts as well. Thus, being adjacent to precincts with yard signs also has a statistically significant effect on the voteshare of Cuccinelli at the 5% significance level.

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

The coefficient of the constant term is 0.302 with a standard error of 0.011. The constant term represents the value of the dependent variable (vote share for Cuccinelli) when all other variables are equal to zero. In the context of this study, this means, that precincts that were neither assigned anti-McAuliffe signs, nor adjacent to precincts with signs, the predicted vote share for Cuccinelli is 30.2%: This explanation holds due to the fact that the two explanatory variables are categorical variables, which must be coded as a dummy variable. E.g. a precinct can either have yard signs or have no yard signs.



- (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?

The fit of the model is shown by the value of Rsquared, which is 0.094 in this scenario. This value is relatively low, indicating that only about 9.4% of the variation in vote share for Cuccinelli is explained by the two explanatory variables. This low Rsquared suggests that, while the presence of yard signs and adjacency to signs has some effect on vote share, this effect is relatively small in explaining the overall voting outcome. In other words, about 90.6% of the variation in Cuccinelli's vote share is caused by factors not explained by this model.

Likely other impactful factors might include demographic characteristics of voters, political affiliations, overall campaign strategy, or socioeconomic conditions of individual precincts.

To conclude, while yard signs have a statistically significant effect on vote share, their overall impact is relatively small when considering all the possible factors that drive voter behavior. Thus, to identify substantive drivers of voter behavior, a more complex and comprehensive model is required.