Problem Set 4

Eleni Karagianni

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 df <- df %>%
2   mutate(professional = ifelse(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 × dummy interaction.)

```
1 model1 <- lm(prestige ~ income + professional + income:professional, data
       = df)
```

Table 1: Coefficient Table of Model 1

	Dependent variable:
	Prestige
Income	0.003***
	(0.0005)
Professional	37.781***
	(4.248)
Income X Professional	−0.002***
	(0.001)
Constant	21.142***
	(2.804)
Observations	98
R^2	0.787
Adjusted R^2	0.780
Residual Std. Error	8.012 (df = 94)
F Statistic	115.878*** (df = 3; 94)
Note:	*p<0.1; **p<0.05; ***p<0.01

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

$$\text{Prestige} = 21.142 + 0.003 \cdot \text{Income} + 37.781 \cdot \text{Professional} - 0.002 \cdot \text{Income*Professional} + \epsilon$$

(d) Interpret the coefficient for `income`.

According to Table 1, for every unit increase of the income, the prestige of the occupation increases by 0.003 percentage points, on average, when the 'professional' variable is equal to 0, therefore when the occupation falls under the category of Blue Collar or White Collar.

(e) Interpret the coefficient for `professional`.

If the occupation falls under the Professional category (aka. when it is equal to 1), the prestige increases, on average, by 37.781 percentage points, when the income is 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).

For the marginal effect of income we need to calculate the partial derivative of the prediction equation with respect to income:

$$\frac{\partial \hat{y}}{\partial \text{income}} = \beta_{\text{income}} + \beta_{\text{income:professional}} \cdot \text{professional}$$

For professional occupations:

$$\Delta \hat{y} = (0.00317 + (-0.00232)) \cdot 1000 = 0.85$$

Therefore, for professional occupations, a $1,000 increase in income results in a 0.85-point increase in the prestige score.

We can cofnrim the same in R as follows:

```
1  marginal_effect_income <- (coef["income"] + coef["income:professional"])
      * 1000
```

```
    income
    0.8452
```

(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).

We are interested in the difference of professional changing from the value 0 to the value 1 when income has the value $6,000.

According to the equation in (c) the predicted prestige for professionals (professional = 1) with an income of $6,000 is:

$$\hat{y}\text{professional} = \beta_0 + \beta\text{income} \cdot 6000 + \beta_{\text{professional}} + \beta_{\text{income:professional}} \cdot 6000$$

And the predicted prestige for non-professionals (professional = 0) with the same income of $6,000 is:

$$\hat{y}\text{non-professional} = \beta_0 + \beta\text{income} \cdot 6000$$

4

The effect of switching to professional is the difference:

$$\Delta \hat{y} = \hat{y}\text{professional} - \hat{y}\text{non-professional}$$

which becomes:

$$\Delta \hat{y} = (\beta_0 + \beta_{\text{income}} \cdot 6000 + \beta_{\text{professional}} + \beta_{\text{income:professional}} \cdot 6000) - (\beta_0 + \beta_{\text{income}} \cdot 6000)$$

$$\Delta \hat{y} = \beta_{\text{professional}} + \beta_{\text{income:professional}} \cdot 6000$$

By plugging in the coefficients from Table 1:

$$\Delta \hat{y} = 37.781 + (-0.00232 \cdot 6000) = 23.861$$

We can check this with R as follows:

```
# (g) 6000 income - change from professional to non professional
# Set the income value to 6000
income_value <- 6000

# Predicted values for professional = 1 and 0
y_prof <- coef["(Intercept)"] + coef["income"] * income_value + coef["
    professional"] + coef["income:professional"] * income_value
y_non_prof <- coef["(Intercept)"] + coef["income"] * income_value

# Effect of changing to professional
effect_g <- y_prof - y_non_prof
effect_g
```

```
(Intercept)
23.82703 # differences due to rounding up
```

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 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 $\alpha = .05$).

First, we state the null and alternative hypotheses:

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

It is a two-tailed test. We use the coefficient and standard error from the regression table to calculate the z-statistic (because N = 131).

$$z = \frac{\hat{\beta}_1}{SE_{\hat{\beta}_1}} = \frac{0.042}{0.016} = 2.625$$

---

[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.

The critical value for $\alpha = 0.05$ and 128 df is 1.96.

Since $2.625 > 1.96$, we have enough evidence to reject the null hypothesis. Therefore, yard signs in a precinct have a statistically significant effect on vote share at the $\alpha = 0.05$ level.

(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 $\alpha = .05$).

We state the null and alternative hypotheses as we did above:

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

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

Again, it is a two-tailed test and we calculate the z-statistic because we have more than 30 observations:

$$z = \frac{\hat{\beta}_2}{SE_{\hat{\beta}_2}} = \frac{0.042}{0.013} = 3.231$$

Since

$$3.231 > 1.96$$

, we have enough evidence to reject the null hypothesis. Therefore, adjacent yard signs have a statistically significant effect on vote share at the $\alpha = 0.05$ level.

(c) Interpret the coefficient for the constant term substantively.
The constant term $\hat{\beta}_0 = 0.302$ represents the expected vote share for Cuccinelli in precincts where:

- No yard signs are posted, and
- The precinct is not adjacent to precincts with yard signs.

This implies that in non-treated precincts, Cuccinelli's vote share is, on average, 30.2%.

(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 R^2 value of 0.094 indicates that only 9.4% of the variation in Cuccinelli's vote share is explained by the presence of yard signs and adjacency effects. This suggests that there are other factors (e.g., local issue saliency, politicians' popularity, economic factors etc.) that could affect the vote share of the challenger politicial, despite the coefficients being statistically significant