

Problem Set 4

Applied Stats/Quant Methods 1

Due: December 4, 2022

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 4, 2022. 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 unique(Prestige$type)
2 Prestige$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 mod <- lm(prestige ~ income + professional + income:professional,
2           data = Prestige)
3 stargazer(mod)
```

Table 1: The effect on `prestige` of `income`, `professional`, and their interaction.

	<i>Dependent variable:</i>
	<code>prestige</code>
<code>income</code>	0.003*** (0.0005)
<code>professional</code>	37.781*** (4.248)
<code>income:professional</code>	-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)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

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

$$y = \beta_0 + \beta_1 x + \beta_2 d + \beta_3 x d$$

$$y = 21.142 + 0.003x + 37.781d - 0.002xd$$

If $d = 0$:

$$y = \beta_0 + \beta_1 x + \cancel{\beta_2 d} + \cancel{\beta_3 x d}$$

$$y = 21.142 + 0.003x$$

If $d = 1$:

$$y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x$$

$$y = (21.142 + 37.781) + (0.003 - 0.002)x$$

$$y = 58.923 + 0.001x$$

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

On average, for each additional dollar a blue or white collar worker ($d = 0$) earns, his prestige score increases by 0.003; while for each additional dollar earned by a professional worker ($d = 1$), his prestige score increases by 0.001.

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

Holding income constant, a professional worker ($d = 1$) has an increase of 37.781 in his prestige score compared to blue and white collar workers ($d = 0$).

- (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).**

For $d = 1$, the prediction equation assumes the following form:

$$y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x$$

$$y = (21.142 + 37.781) + (0.003 - 0.002)x$$

$$y = 58.923 + 0.001x$$

That means that an addition of \$1,000 in a professional worker's income is associated to an increase of 1 in his prestige score. In the prediction equation:

$$y = 58.923 + 0.001 \cdot 1,000$$

$$y = 58.923 + 1 = 59.923$$

- (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).

On average, a *non-professional worker* earning \$6,000 has a prestige score of $y_{non-prof} = 39.142$, given by the following equation:

For $d = 0$:

$$y_{non-prof} = \beta_0 + \beta_1 x$$

$$y_{non-prof} = 21.142 + 0.003x$$

$$y_{non-prof} = 21.142 + 0.003 \cdot 6,000 = 21.142 + 18 = 39.142$$

A *professional worker*, earning the same \$6,000, has on average a prestige score of $y_{prof} = 64.523$, given by the following equation:

For $d = 1$:

$$y_{prof} = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x$$

$$y_{prof} = (21.142 + 37.381) + (0.003 - 0.002)x$$

$$y_{prof} = 58.523 + 0.001x$$

$$y_{prof} = 58.523 + 0.001 \cdot 6,000 = 58.523 + 6 = 64.523$$

Therefore, holding income constant (in this case, at \$6,000), a professional worker's prestige score will be on average 25.381 higher than a non-professional worker's score ($y_{prof} - y_{non-prof} = 64.523 - 39.142$). That variation is explained by **the individual effect of professional occupations** ($\beta_2 = 37.381$) plus the **interaction effect** that changes the slope of x and the impact of income on prestige ($\beta_3 = -0.002$).

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 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 $\alpha = .05$).

Model:

$$y = \beta_0 + \beta_{prec}x_{prec} + \beta_{adj}x_{adj}$$
$$y = 0.302 + 0.042x_{prec} + 0.042x_{adj}$$

¹Donald 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.

1. **Assumptions:** Randomization, linearity, normality, and constant variance.

2. **Hypotheses:**

$$H_0 : \beta_{prec} = 0$$

$$H_a : \beta_{prec} \neq 0$$

3. **T-test:**

$$t = \frac{\hat{\beta}_{prec} - 0}{se_{\hat{\beta}_{prec}}} = \frac{0.042}{0.016} = 2.625$$

4. **P-value (two-tail):**

```
1 2*pt(2.625, df = 131 - 3, lower.tail = F)
```

$$p = 0.009$$

5. **Conclusion:** Considering $\alpha = 0.05$, there is enough statistical evidence to *reject the null hypothesis* according to which yard signs in the precinct have no effect on 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 $\alpha = .05$).**

1. **Assumptions:** Randomization, linearity, normality, and constant variance.

2. **Hypotheses:**

$$H_0 : \beta_{adj} = 0$$

$$H_a : \beta_{adj} \neq 0$$

3. **T-test:**

$$t = \frac{\hat{\beta}_{adj} - 0}{se_{\hat{\beta}_{adj}}} = \frac{0.042}{0.013} = 3.231$$

4. **P-value (two-tail):**

```
1 2*pt(3.231, df = 131 - 3, lower.tail = F)
```

$$p = 0.001$$

5. **Conclusion:** Considering $\alpha = 0.05$, there is enough statistical evidence to *reject the null hypothesis* according to which lawn signs in adjacent precincts have no effect on vote share.

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

The constant term ($\beta_0 = 0.302$) is the average proportion of votes that went to McAuliff's opponent (Ken Cuccinelli) in precincts that did not have lawn signs and were not adjacent to precinct with lawn 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?**

Even though the associations in the model are statistically relevant ($p_{prec} = .009$ and $p_{adj} = 0.001$), $R^2 = 0.094$ is very low. Such low R^2 means that the model does not explain much of data variance and, probably, other factors which were not considered may have more weight in explaining the difference of voters' preferences among precincts.