Problem Set 4

Applied Stats/Quant Methods 1

Due: December 3, 2023

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 3, 2023. 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  Prestige$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 × dummy interaction.)

```
1  lm1 <- lm(Prestige$prestige ~ Prestige$income +
2                                Prestige$professional +
3                                Prestige$income*Prestige$professional)
4  summary(lm1)
```

Table 1:

	Dependent variable:
	prestige
income	0.003***
	(0.0005)
professional	37.781***
	(4.248)
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.

$$\hat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \widehat{\beta}_2 x_2 + \beta_3 \times (x_1 \times x_2)$$
$$= 21.142 + 0.003 \times \text{income} + 37.781 \times \text{professional} - 0.002 \times (\text{income} \times \text{professional})$$

(d) Interpret the coefficient for `income`.

Our model includes an interaction term. Because of this, there are different coefficients for professionals and blue and white collared workers. For blue and white collared workers (professional $= 0$), every dollar increase in income is associated with an average 0.003 increase in prestige of 0.003. Meanwhile, to interpret the association between income and prestige for professionals, we simply have to add the interaction effect of -0.002 to the previous coefficient, giving us 0.001. This means that for professionals, a one dollar increase in income is associated with an on average 0.001 increase in prestige.

(e) Interpret the coefficient for `professional`.

In comparison to blue and white collar workers, the prestige of professionals is on average higher by 37.781, when income is 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).

$$
\begin{aligned}
\text{Marginal Effect} &= \text{Coefficient of income} + \text{Coefficient of(income} \times \text{professional)} \\
&\quad \times \text{Value of professional} \\
&= 0.003 - 0.002 \times 1 \\
&= 0.001
\end{aligned}
$$

$$
\begin{aligned}
\text{Effect} &= \text{Marginal Effect} \times \text{Change in income} \\
&= 0.001 \times 1000 \\
&= 1
\end{aligned}
$$

This shows that for professionals, a 1000$ increase in income is associated with a 1 point 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).

To calculate this we need to know the marginal effects of `professional` being equal to 0 and 1. As we already know the marginal effect of professional $= 1$, we just need to calculate the marginal effect of professional $= 0$.

3

$$\text{Marginal Effect} = \text{Coefficient of income} + \text{Coefficient of(income} \times \text{professional)}$$
$$\times \text{Value of professional}$$
$$= 0.003 - 0.002 \times 0$$
$$= 0.003$$

Then, we calculate the effects when `professional` $= 0$ and $= 1$.

$$\text{Effect} = \text{Marginal Effect} \times \text{Change in income}$$
$$= 0.001 \times 6000$$
$$= 6$$

$$\text{Effect} = \text{Marginal Effect} \times \text{Change in income}$$
$$= 0.003 \times 6000$$
$$= 18$$

Finally we subtract the effect of `professional` $= 0$ from `professional` $= 1$.

$$6 - 18 = -12$$

This shows us that, when income is \$6000, $\hat{y}$ decreases by 12 points, when changing from non-professional to professional.

Question 2: Political Science

Researchers are interested in learning the effect of all of those yard signs on voting prefer-ences.[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 $\alpha = .05$).

1. Hypotheses

H_0: Having yard signs in the precinct does not affect vote share.

H_1: Having yard signs in the precinct affects vote share.

2. Test statistic

$$t = \frac{\beta_1}{se_{\beta_1}} = \frac{0.042}{0.016} = 2.625$$

3. P-Value df = n - 3 = 30 - 3 = 27 $p - value = 2 \times Pr(t_{27} > |2.625|)$

---

[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 experi-ments." Electoral Studies 41: 143-150.

```
1  # Given values
2  t_statistic <- 2.625
3  degrees_of_freedom <- 27
4
5  # Calculate p-value for a two-sided test
6  p_value <- 2 * (1 - pt(abs(t_statistic), df = degrees_of_freedom))
```

p-value: 0.01409096

4. Conclusion

As the p-value is lower than α, we can conclude that there is enough evidence to reject the the null hypothesis. This means that having lawn signs in the precinct is associated with a 0.042 point increase in 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. Hypotheses

H_0: Having yard signs in the precinct does not affect vote share.

H_1: Having yard signs in the precinct affects vote share.

2. Test statistic

$$t = \frac{\beta_2}{se_{\beta_2}} = \frac{0.042}{0.013} = 3.231$$

3. P-Value df = n - 3 = 76 - 3 = 73 $p-value = 2 \times Pr(t_{73} > |3.231|)$

```
1  # Given values
2  t_statistic2 <- 3.230769
3  degrees_of_freedom2 <- 73
4
5  # Calculate p-value for a two-sided test
6  p_value2 <- 2 * (1 - pt(abs(t_statistic2), df = degrees_of_freedom2))
```

p-value: 0.001852656

4. Conclusion As the p-value is lower than α, we can conclude that there is enough evidence to reject the the null hypothesis. This means that having lawn signs in the adjacent precinct is associated with a 0.042 point increase in vote share.

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

The coefficient for the constant term is essentially the intercept. This means that when there are no lawn signs in the precinct or adjacent precincts, the estimated mean vote share is 0.302.

(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 coefficients were found to have a statistically significant effect, meaning that lawn signs do affect vote share positively. However, this effect is small (0.042) for both signs in the precinct and adjacent to it.

The R^2 value is low, which suggests that the model only explains approximately 9.4% of the variance in vote share. This implies that while yard signs do have some effect on vote share, there are other factors that have an effect on this that have not been included in the model.