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`).

```r
# Create a new variable 'professional' where 'prof' is coded as 1, others as 0
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.)

```r
# Create an interaction term between 'income' and 'professional'
Prestige$income_professional <- Prestige$income * Prestige$professional

# Run the linear regression model with interaction term
model1 <- lm(prestige ~ income + professional + income_professional, data=Prestige)
summary(model1)
```

Call:

lm(formula = prestige ~ income + professional + 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

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

Prediction equation:

$$\hat{y} = \beta_0 + \beta_1 \cdot \text{income} + \beta_2 \cdot \text{professional} + \beta_3 \cdot (\text{income} \times \text{professional})$$

The prediction equation based on the result will be:

$$\hat{prestige} = 21.1422589 + 0.0031709 \cdot \text{income} + 37.7812800 \cdot \text{professional} - 0.0023257 \cdot (\text{income} \times \text{professional}$$

(d) Interpret the coefficient for `income`.

The coefficient for income ($\beta_1 = 0.0031709$) indicates that, for non-professionals ($professional = 0$), each additional unit of income is associated with an average increase of 0.0031709 points in the prestige score.

(e) Interpret the coefficient for `professional`.

The coefficient for professional ($\beta_2 = 37.7812800$) indicates that, when income is 0, professionals have an average prestige score that is 37.7812800 points higher than non-professionals.

(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:

$$\hat{y} = \beta_0 + \beta_1 \cdot \text{income} + \beta_2 \cdot 1 + \beta_3 \cdot (\text{income} \times 1)$$

$$= \beta_0 + \beta_1 \cdot \text{income} + \beta_2 + \beta_3 \cdot \text{income} = \beta_0 + \beta_2 + (\beta_1 + \beta_3) \cdot \text{income}$$

Now, if income increases by 1000, we need to calculate:

$$\Delta\hat{y} = \hat{y}(\text{income} + 1000) - \hat{y}(\text{income})$$

Substitute:

$$\hat{y}(\text{income} + 1000) = \beta_0 + \beta_2 + (\beta_1 + \beta_3) \cdot (\text{income} + 1000)$$

$$\hat{y}(\text{income}) = \beta_0 + \beta_2 + (\beta_1 + \beta_3) \cdot \text{income}$$
$$\Delta\hat{y} = [\beta_0 + \beta_2 + (\beta_1 + \beta_3) \cdot (\text{income} + 1000)] - [\beta_0 + \beta_2 + (\beta_1 + \beta_3) \cdot \text{income}]$$
$$= (\beta_1 + \beta_3) \cdot 1000$$

Calculation:

$$\Delta\hat{y} = \beta_1 + \beta_3 \cdot \text{professional} = 0.0031709 - 0.0023257 \times 1 = 0.0008452$$

Since the income unit is $1, an increase of 1$,000 corresponds to 1000 units. Therefore:

$$\Delta\hat{y}_{\$1000} = 0.0008452 \times 1000 = 0.8452$$

For professionals, an increase of $\$1,000$ in income is associated with an average increase of approximately 0.8452 points in the prestige score.

(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).

When we want to calculate the effect of changing from non-professional to professional status, we are actually calculating:

$$\Delta\hat{y} = \hat{y}(\text{professional} = 1) - \hat{y}(\text{professional} = 0)$$

When professional=1:

$$\hat{y}_1 = \beta_0 + \beta_1 \cdot \text{income} + \beta_2 \cdot 1 + \beta_3 \cdot (\text{income} \times 1) = \beta_0 + \beta_1 \cdot \text{income} + \beta_2 + \beta_3 \cdot \text{income}$$

When professional=0:

$$\hat{y}_0 = \beta_0 + \beta_1 \cdot \text{income} + \beta_2 \cdot 0 + \beta_3 \cdot (\text{income} \times 0) = \beta_0 + \beta_1 \cdot \text{income}$$

Therefore, the change effect $\Delta\hat{y}$ is:

$$\Delta\hat{y} = \hat{y}_1 - \hat{y}_0 = (\beta_0 + \beta_1 \cdot \text{income} + \beta_2 + \beta_3 \cdot \text{income}) - (\beta_0 + \beta_1 \cdot \text{income}) = \beta_2 + \beta_3 \cdot \text{income}$$

Calculation:

$$\Delta\hat{y} = \beta_2 + \beta_3 \cdot \text{income} = 37.7812800 - 0.0023257 \times 6000 \approx 23.827$$

When income is $\$6,000$, changing from a non-professional to a professional is associated with an average increase of approximately 23.827 points in the prestige score.

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

[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.

(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 $\alpha = .05$).

Hypothesis:

$$H_0 : \beta(\text{yard signs}) = 0$$

$$H_1 : \beta(\text{yard signs}) \neq 0$$

Test statistic:

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

Degrees of freedom $= 131 - 3 = 128$ (sample size minus number of parameters)

```
# Results
coef_assigned <- 0.042
se_assigned <- 0.016

# Calculate t-statistic
t_value_assigned <- coef_assigned / se_assigned

# Calculate two-sided p-value
p_value_assigned <- 2 * pt(-abs(t_value_assigned), df=128)

# Output results
t_value_assigned
p_value_assigned
```

At $\alpha = 0.05$ significance level:

$$|t| = 2.625 > t(0.025, 128) \approx 1.96$$

$$p = 9.72\text{e}^{-3} < 0.05$$

Therefore, we reject the null hypothesis, indicating that yard signs have a significant 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 $\alpha = .05$).

Hypothesis:

$$H_0 : \beta(\text{adjacent}) = 0$$

$$H_1 : \beta(\text{adjacent}) \neq 0$$

Test statistic:

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

```r
# Results
coef_assigned <- 0.042
se_assigned <- 0.013

# Calculate t-statistic
t_value_assigned <- coef_assigned / se_assigned

# Calculate two-sided p-value
p_value_assigned <- 2 * pt(-abs(t_value_assigned), df=128)

# Output results
t_value_assigned
p_value_assigned
```

At $\alpha = 0.05$ significance level:

$$|t| = 3.231 > t(0.025, 128) \approx 1.96$$

$$p = 1.57\mathrm{e}^{-3} < 0.05$$

Therefore, we reject the null hypothesis, indicating that being adjacent to precincts with yard signs also has a significant effect on vote share.

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

The constant term (0.302) indicates that Cuccinelli's expected vote share is 30.2% when there are no yard signs and no yard signs in adjacent 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?

$R^2 = 0.094$ means that only 9.4% of the variation in vote share can be explained by yard signs. This suggests that while yard signs have a statistically significant effect, their actual explanatory power is small. Other factors not included in the model likely have a larger influence on voting behavior.