

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 install.packages('car')
2 library(car)
3 df<- Prestige
4 df
5 ## creating a new column
6
7 df['professional']<-ifelse(df$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 ## running a linear regression model
2 table(df$professional)
3 model_1<- lm(prestige~income+professional*income, data=df)
4 model_1
5 summary(model_1)
```

Call:

```
lm(formula = prestige ~ income + professional * income, data = df)
```

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

Residual standard error: 8.012 on 94 degrees of freedom

(4 observations deleted due to missingness)

Multiple R-squared: 0.7872, Adjusted R-squared: 0.7804

F-statistic: 115.9 on 3 and 94 DF, p-value: < 2.2e-16

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

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

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

$$\text{prestige} = 21.14 + 0.003 \times \text{income} + 37.78 \times \text{professional} - 0.0023 \times (\text{income} \times \text{professional})$$

1. β_0 - represents the intercept, expected prestige value when income and professional is zero
2. β_1 the coefficient of income, it represents the effect of income on prestige when professional is zero
3. β_2 - is the coefficient of professional it represents the effect of being professional when income is zero
4. β_3 : is the coefficient of interaction between income and professional, it represents how the effect of income on prestige changes if an individual is professional compared to non professional

[(d)] Interpret the coefficient for **income**.

the coefficient of income is 0.003, it indicates that for every 1 unit increase in income is associated with approximate increase of 0.00317 in prestige, when professional is zero(non-professional)

[(e)] Interpret the coefficient for **professional**.

the coefficient of professional is 37.78, it indicates that the effect on prestige by professional is 37.78 points higher than non professional given income is zero

[(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).

$$\text{prestige} = 21.14 + 0.00317 \times \text{income} + 37.78 \times \text{professional} - 0.0023 \times (\text{income} \times \text{professional})$$

for a \$1000 increase in income

$$\text{prestige} = 21.14 + 0.00317 \times (\text{income} + 1000) + 37.78 \times \text{professional} - 0.0023 \times ((\text{income} + 1000) \times \text{professional})$$

$$\Delta \hat{y} = \hat{y} - \hat{y}_{\text{income.inc}}$$

$$\Delta \hat{y} = (\beta_{\text{income}} + \beta_{\text{income:professional}}) \times 1000$$

$$\Delta \hat{y} = (0.003 - 0.002 * 1) * 1000 \text{(rounding off my coefficient to 3 places)}$$

```
1 Marginal_effect = (round(model_1$coefficients[2],3)+round(model_1$coefficients
  [4],3))
2 change_prestige <- Marginal_effect*1000
3 change_prestige
  > change_prestige
  1
```

therefore for a professional, an increase in income by 1000 dollars would lead to a change in prestige by 1 units

[(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).

$$\text{prestige} = 21.14 + 0.003 \times \text{income} + 37.78 \times \text{professional} - 0.0023 \times (\text{income} \times \text{professional})$$

when professional = 0

$$\text{prestige} = 21.14 + 0.003 \times \text{income} + 37.78 \times 0 - 0.0023 \times (\text{income} \times 0)$$

when professional = 1

$$\text{prestige} = 21.14 + 0.003 \times \text{income} + 37.78 \times 1 - 0.0023 \times (\text{income} \times 1)$$

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

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

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

$$\Delta \hat{y} = 37.78 - 0.002 * 1 * 6000 \text{(rounding off my coefficient to 3 places)}$$

```
1 Marginal_effect_2 <- (round(model_1$coefficients[3],3)+round(model_1$
  coefficients[4],3)*6000*1)
2 Marginal_effect_2
```

```
> Marginal_effect_2  
25.781
```

therefore at a 6000 dollars income, switching one's occupation from non professional to professional is associated with an increase in prestige by 25.78 units

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 $\alpha = .05$).

- Assumptions:
 - (a) Random sampling
 - (b) The data is continuous
 - (c) The distribution is approximately normal
- Hypothesis Testing:

H_0 : Assigning yard signs to a precinct has no effect on vote share, $\beta_{\text{assign}} = 0$.

H_1 : Assigning yard signs to a precinct has an effect on vote share, $\beta_{\text{assign}} \neq 0$.

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.

- Test statistic:

$$t = \frac{\text{Coefficient}}{\text{Standard Error}}$$

```
1 #t_value = coefficient/std.error
2 t_value = 0.042/0.016
3 t_value
```

```
> t_value
[1] 2.625
```

- Finding the p-value

```
1 df_2 = 131
2 p_value = 2*pt(abs(t_value), df_2 - 2, lower.tail = FALSE)
3 p_value
```

```
> p_value
[1] 0.00972002
```

- Conclusion:
Since the p-value is less than the significance level ($\alpha = 0.05$), we reject the null hypothesis. This indicates that there is a statistically significant effect of assigning yard signs 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 $\alpha = .05$).

- Assumption

- (a) Random sampling
- (b) The data is continuous
- (c) The distribution is approximately normal

- Hypothesis testing:

H_0 : Adjacent yard signs to a precinct have no effect on vote share, $\beta_{\text{adjacent}} = 0$

H_1 : Adjacent yard signs to a precinct have an effect on vote share, $\beta_{\text{adjacent}} \neq 0$

- Test statistic:

$$t = \frac{\text{Coefficient}}{\text{Standard Error}}$$

```
1 #t_value
2 t_value_2<- 0.042 / 0.013
3 t_value_2
```

```
> t_value_2
[1] 3.230769
```

- Finding the p-value

```
1 p_value_2<- 2*pt(t_value_2,df_2-2-1,lower.tail=FALSE)
2 p_value_2
```

```
> p_value_2
[1] 0.00156946
```

- Conclusion:

Since the p-value is less than the significance level ($\alpha = 0.05$), we reject the null hypothesis. This indicates that there is a statistically significant effect of adjacent yard signs on vote share.

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

the coefficient suggest that on average, precincts without lawn signs and not adjacent to precincts with lawn signs have a baseline voteshare of 30.2 percent, the standard error of 0.011 indicates the precision of the estimate of the constant. if Precinct are neither assigned lawn signs nor adjacent to them the expected voteshare is 30.2 percent

- (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?

since the r square value is 0.094, the model explains on 9.4 variance in the voteshare based on the lawn signs(assigned and adjacent), the value is quite low suggesting that lawn signs alone cannot be the predictor for vote share, though the lawn signs had a statistically significant effect on voteshare the overall impact of the two independent variable is not quite strong over predicting the variance in dependent variable and the voteshare is likely to be influenced by many other factors.