

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 professional<-ifelse(Prestige$type=="prof",1,0)
2 prtab<-table(professional)
3 prtab<-as.data.frame(prtab)
```

Table below describes the variable 'Professional'

Table 1: Professional Variable Table

Statistic	N	Mean	St. Dev.	Min	Max
Freq	2	49.000	25.456	31	67

- (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 presreg<-lm(Prestige$prestige ~ Prestige$income*professional)
```

Table 2: Linear Model with Outcome Variable: Prestige and Explanatory Variables: Income, Professional, and interaction of Income and Professional

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

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

The prediction equation from this regression model is:

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

where prestige is the outcome variable, and income, professional, and the interaction between income and professional are the explanatory variables.

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

The coefficient for "income" is 0.003, indicating that for each additional unit of income the prestige will increase by 0.003 units, while professional is held equal to 0. The three asterixes on the coefficient indicate that this relationship is statistically significant at the level p less than 0.01.

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

The coefficient on the variable "professional" is 37.781. This indicates that when income is held at 0, for each unit increase in professional the outcome variable prestige will increase by 37.781 units. Again, the three asterixes on this coefficient infer that this relationship is statistically significant at the level of p being less than 0.01.

- (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).

```

1 cf<-coef(presreg)
2 meinc<-(cf["Prestige$income"] + cf["Prestige$income:professional"])
3 meinc
4 stargazer(meinc, type="latex", title="Marginal Effect of income when the
   variable takes the value of $1", out = file.path(fp, "ps04q1one$.tex"
   ))
5 meinc2<-meinc*1000
6 meinc2

```

The marginal effect of income when the variable professional takes the value of one dollar is exactly 0.0008452.

Table 3: Marginal Effect of income when the variable takes the value of one dollar (rounded)

Prestige\$income
0.001

The change in the outcome variable prestige associated with a one thousand dollar increase in income is exactly 0.8452.

Table 4: Change in Prestige associated with a one thousand dollar increase in income (rounded)

Prestige\$income
0.845

- (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).

```
1 chngprof<-cf["professional"]+cf["Prestige$income:professional"]*6000
2 chngprof
```

The marginal effect of increasing professional jobs by one unit on outcome prestige when income is six thousand dollars is an increase in 23.82703 units.

Table 5: Marginal Effect of professional jobs when income is six thousand dollars

professional
23.827

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 $\alpha = .05$).

```
1 lawn_cf <- 0.042
2 lawn_se <- 0.016
3 adjlawn_cf <- 0.042
4 adjlawn_se <- 0.013
5 n_lawn <- 30
6 n_adjlawn <- 76
7 int_cf <- 0.302
8 int_se <- 0.011
9 rsqr <- 0.094
10 n <- 131
11 lawn_tstat <- lawn_cf / lawn_se
12 adjlawn_tstat <- adjlawn_cf / adjlawn_se
```

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.

```

13 int_tstat<-int_cf/int_se
14 df_lawn=n_lawn-1
15 df_adjlawn=n_adjlawn-1
16 df_int<-n-1
17 p_lawn<-2*(1-pt(lawn_tstat, df=df_lawn))
18 p_adjlawn<-2*(1-pt(adjlawn_tstat, df=df_adjlawn))
19 p_int<-2*(1-pt(int_tstat, df=df_int))
20 p_value<-function(pval) {
21   if(pval>0.05){
22     cat("P Value is:", pval, "The p value is greater than 0.05,
23       therefore we fail to reject the null hypothesis")
24   }
25   else {
26     cat("P value is:", pval, "The p value is less than 0.05, thus
27       there is sufficient statistical evidence to reject the null
28       hypothesis")
29   }
30 }
31 p_value(p_lawn)
32 p_value(p_adjlawn)
33 p_value(p_int)

```

The coefficient for precincts with yard signs, has a P value of 0.01368397. This is less than 0.05, thus there is sufficient evidence to reject the null hypothesis. In this model there is a statistically significant relationship between the presence of yard signs in a precinct and 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 $\alpha = .05$).

The coefficient for regions adjacent to precincts with yard signs has a P value of 0.001834302. This is less than 0.05, thus there is sufficient evidence to reject the null hypothesis. There is a statistically significant relationship between the being next to precincts with of yard signs and vote share in this regression model.

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

The coefficient for the constant term is 0.302. The constant term is the intercept - that is, when controlling for the variables being in a precinct with yard signs and being next to a precinct with yard signs the proportion of the vote that went to Cuccinelli was 0.302. It is the predicted value of vote shares that went to Cuccinelli when the explanatory variables (precincts with yard signs, and adjacent precincts with yard signs) are equal to zero. The constant in this model indicates that the predicted proportion of the vote which would go to Cuccinelli would be 30.2 percent if the explanatory variables were equal to zero.

- (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 p values for all variables in this regression were below 0.05. That means, that for an alpha level of 0.05, the variables selected all have statistically relevant impacts on the outcome variable - that the presence of yard signs in a precinct, or adjacent to a precinct had a statistically important relationship with proportion of votes which went to Cuccinelli. However, the R squared value is very low - it is only 0.094. This indicates that despite the variables selected having strong statistical relationships with the outcome variable, that including other variables in the model could help improve it. Only 9.4 percent of variance in the proportion of vote shares which went to Cuccinelli was explained by this model, thus looking only at yard signs in a precinct, or adjacent to a precinct is insufficient in explaining the variation of the proportion of votes which went to Cuccinelli. Clearly other variables need to be considered and included into his model in order to improve the predictability of Cuccinelli's vote proportion.