Problem Set 4

Tianxin Zhang/21352232/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`).

```
nstall.packages(car)
library(car)
data(Prestige)
help(Prestige)
library(stargazer)

#Q1:

# a)

Prestige$professional_dummy <- ifelse(Prestige$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 × dummy interaction.)

```
# b)

# multiple regression with interaction
interact_reg_pres_inc_pro <- lm(prestige ~ income + professional_dummy +
    income:professional_dummy, data = Prestige)
summary((interact_reg_pres_inc_pro))
stargazer(interact_reg_pres_inc_pro, title="Regression Results with
    interaction: prestige ~ income + professional_dummy")
```

p-value of the coefficient for income is smaller than 0.01, we can reject the null hypothesis that there is no association statistically significant between prestige and income at the 99% level.
p-value of the coefficient for professional_dummy is smaller than 0.01, we can reject the null hypothesis that there is no association statistically significant between prestige and professional_dummy at the 99% level.
p-value of the coefficient for interaction of income and professionaldummy is smaller than 0.01, we can reject the null hypothesis that there is no association statistically

Table 1: Regression Results with interaction: prestige income + professional_dummy

	Dependent variable:
	prestige
income	0.003^{***}
	(0.0005)
professional_dummy	37.781^{***}
	(4.248)
income:professional_dummy	-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:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

significant between prestige and interaction of income and professional_dummy at the 99% level.

(c) Write the prediction equation based on the result.
$\hat{y} = 21.142 + 0.003x_1 + 37.781x_2 - 0.002x_1x_2$

(d) Interpret the coefficient for `income`.
When holding the effects of other variables in the model constant, with an increase in income by 1 dollar, we expect to see an increase in prestige by 0.003.

(e) Interpret the coefficient for `professional`.
When holding the effect of other variables in the model constant, with a difference in professional, we expect to see an average difference of prestige by 37.781.

(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).
$0.003 * 1000 - 0.002 * 1000 * 1 = 3$

```
# f)

y1 = 0.003 * 1000 - 0.002*1000*1
# y1 = 1
```

$\hat{y} = 1$ The effect is an increase in prestige by 1

(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).
$\hat{y} = 37.781 * 1 - 0.002 * 6000 * 1$

```
# g)

y2 = 37.781 * 1 - 0.002 * 6000 *1
# y2 = 25.781
```

$\hat{y} = 25.781$ The effect is an increase in prestige by 25.781

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

(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: $H_0 : B_1 = 0 vs. H_a : B_1 \neq 0$
Test-Statistics:
$standarderror : 0.016, t = (0.042 - 0)/0.016 = 2.625, degreeoffreedom : n - 3 = 131 - 3 = 128$

```
sd1 = 0.016
ts1 <- (0.042-0) /0.016
ts1
# test_statistics = 2.625
```

P-value:

[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.

```
p_value1 <- 2*pt(abs(ts1), 128, lower.tail = FALSE)
p_value1
# p_value1 =  0.00972002
```

$p - value = 0.0097 < 0.05$
Conclusion: we can reject the null hypothesis that the slope of having these yard signs in a precinct is 0.

(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: $H_0 : B_2 = 0 vs. H_a : B_2 \neq 0$

Test-Statistics:

$standarderror : 0.013, t = (0.042 - 0)/0.013 = 2.625, degreeoffreedom : n - 3 = 131 - 3 = 128$

```
sd2 = 0.013
ts2 = (0.042-0) /0.013
ts2
# test_statistics = 3.231
```

P-value:

```
p_value2 <- 2*pt(abs(ts2), 128, lower.tail = FALSE)
p_value2
# p_value2 =  0.00156946
```

$p - value = 0.0016 < 0.05$

Conclusion: we can reject the null hypothesis that the slope of being next to precincts with yard signs is 0.

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

When the precinct is not assigned lawn signs or the precinct is not adjacent to lawn signs, the vote share of this precinct went to McAuliff's opponent Ken Cuccinelli is 30.2%.

(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 R^2 of the model is 9.4%, which means 9.4% of the variations are explained by the yard signs, and the rest 90.6% of variations are explained by other unmodeled factors.