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`).

```
lapply(c("stargazer","car"), pkgTest)
# Load the Prestige dataset
data("Prestige")
#alternatively as instructed in PS i had run the following:
#install.packages(car)
#library(car)
#data(Prestige)
#help(Prestige)
# Recode the 'type' variable to create 'professional'
Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0)
Prestige$professional <- as.factor(Prestige$professional)
```

Note that I further recoded the variable to be a factor to reflect it's categorical data type.

(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.)

```
#odel <- lm(prestige ~ income + professional + income:professional, data
    = Prestige_clean)
# Linear model with interaction term
model <- lm(prestige ~ income + professional + income:professional, data
    = Prestige)
summary(model)
```

Table 1:

	Dependent variable:
	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)
Note:	*p<0.1; **p<0.05; ***p<0.01

(c) Write the prediction equation based on the result.
Prestige = 21.142 + 0.003*Income + 37.781*Professional - 0.002*Income*Professional

(d) Interpret the coefficient for `income`.
The coefficient for income indicates how much the prestige score is expected to change on average for each unit increase in income when the professional term is zero(when the professional is white collar or blue colloar). In this case, for white collar/ blue collar worker, for each additional dollar the worker makes his/her presitige score will increase on average 0.003.

(e) Interpret the coefficient for `professional`.
The coefficient for professional indicates how much the prestige score is expected to change on average from being white/blue collar to professional, when the work has no incomes. In this case, holding income constant(0), the professional work is expected to have 37.781 prestige score more than a equally 0 income white/blue colalr worker.

(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).

```
stargazer(model)
#Prestige = Intercept + Coef_income*Income + Coef_professional*
    Professional + Coef_interaction*Income*Professional
coef_model <- coef(model)

# Calculate effect of $1,000 increase for professionals
effect_increase <- coef_model["income"] + coef_model["income:
    professional1"]
```

the marginal increase in prestige score is 0.8452

(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).

```

# Calculate the effect of changing to professional at $6,000 income
change_effect <- coef_model["professional1"] + coef_model["income:
    professional1"] * 6000
```

If we change one's occupation from non-professional to professional worker when her income is 6000 dolalr, the mariginal effect on the prestige score is 23.82703 increase.

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 $\alpha = .05$).

```

#q2
# Assuming a standard significance level of 0.05
coef_value <- 0.042
std_error <- 0.016
t_value <- coef_value / std_error
p_value <- 2 * (1 - pt(abs(t_value), df = 131 - 2))  # two-tailed test
```

Null Hypothesis (H0): Lawn yard signs do not affect vote share. (i.e., the effect of lawn signs is zero) Alternative Hypothesis (HA): yards signs do affect vote share. (i.e., the effect of lawn signs is not zero) Since the p-value is 0.0097 is less than the significance

[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.

level of 0.05, we have found evidence to reject the null hypothesis that the yard signs in a presicnt do not affect vote share. There is a statistically significant evidence to suggest that lawn signs in a precinct do have an 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 $\alpha = .05$).

```

# Repeat the hypothesis test procedure with the coefficient and standard
   error for adjacent precincts
coef_value_adjacent <- 0.042
std_error_adjacent <- 0.013
t_value_adjacent <- coef_value_adjacent / std_error_adjacent
p_value_adjacent <- 2 * (1 - pt(abs(t_value_adjacent), df = 131 - 2))
```

Null Hypothesis (H0): adjacent yard signs do not affect vote share. (i.e., the effect of adjacent signs is zero) Alternative Hypothesis (HA): next to yards signs do affect vote share. (i.e., the effect of adjacent signs is not zero) Since the p-value 0.0015 is less than the significance level of 0.05, we have found evidence to reject the null hypothesis that next to precincts with the yard signs affect vote share. There is a statistically significant evidence to suggest that next to lawn signs in a precinct have an effect on vote share.

(c) Interpret the coefficient for the constant term substantively.
The constant coefficient of 0.302 represents the baseline vote share in precincts without lawn signs and not adjacent to any precincts with lawn signs. It indicates the expected vote share for Ken Cucinelli is 0.302 when there are no lawn signs (control group), assuming other factors are held constant.

(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-squared value of 0.094 suggests that only about 9.4percent of the variance in the vote share is explained by the presence of lawn signs in and adjacent to precincts. This indicates that while lawn signs might have some effect, a significant portion of the vote share variance is likely due to other unmodeled factors. Variables may have been omitted in this model. Lawn signs and adjencent lawn signs may not be sufficient to explain the variation in the propotion of vote that wen to Mcauliff's openent Ken Cucinelli.