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`).

Using the ifelse() function To create a new variable: "professional" in the "Prestige" dataset by the existing variable : "type." It assigns a value of 1 if the corresponding value in the "type" variable is "prof" and 0 to others.

```
Prestige$professional <- ifelse(Prestige$type == "prof", 1, 0)
table(Prestige$professional)

```

(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.) using lm() function, income: professional represents a continuous and dummy interaction for income and professional

```
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

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.

The intercept is 21.1423, which means the estimated prestige when all 3 predictors remain zero.

Prestige = 21.1423+0.0031709×Income+37.7813×Professional−0.0023257×(Income× Professional)

(d) Interpret the coefficient for `income`.

0.0031709 is the coefficient of the "Income" variable, representing on average, for each one-unit increase in the "Income" variable, the the predicted value of prestige is expected to increase by 0.0031709, while the rest predictor variables remain the same.

(e) Interpret the coefficient for `professional`.

The coefficient 37.7813 represents: On average, the estimated change in the prestige for people who are classified as professional compared to those who are not, while other predictor variables hold the same.

When Professional is binary 1 (means professional), and the coefficient 37.7813 represents the estimated change in prestige for professionals compared to non-professionals.

Professional is binary 0 (means non-professional), the coefficient 37.7813 does not influence the prestige.

(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).

Using the coefficient for the "Income" variable in the linear regression model, to calculate the effect ($\hat{y}$) of a \$1,000 increase in income on the outcome prestige for professional type, the result as follow:

($\hat{y}$) = Coefficient for Income $\times$ Change in Income = 0.0031709$\times$1000= 3.17

(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).

Using the coefficient for 1 unit change in the "Professional" variable and 1 unit change in the interaction : "Income:Professional" in the linear regression model to calculate the effect of changing one's occupation from non-professional to professional when the income is\$6,000

($\hat{y}$) = Coefficient for Professional+ Coefficient for Income:Professional $\times$Income = 37.7813+(0.0023257) $\times 6,000 = 28.8271$

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 $\alpha = .05$).

The null hypothesis (H0) : yard signs in a precinct has no effect (coefficient is zero) on vote share

The alternative hypothesis (H1) : yard signs in a precinct has effect (coefficient is not zero) on vote share

There are two variables related to yard signs: Precinct assigned lawn signs / Precinct adjacent to lawn signs

[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.

DF = 131-2= 129 Two-Tailed Test here: $\alpha/2 = 0.025$ in each tail.

t-statistic (assigned) = Coefficient / Standard Error of the Coefficient = 0.042/ 0.016= 2.625

Critical value (assigned) = 1.979

The absolute value of the t-statistic (2.625) is greater than the critical value (1.979). It suggests there is evidence at the 0.05 significance level that the presence of assigned lawn signs has a statistically significant on the proportion of the vote, so reject the null hypothesis

In conclusion: yard signs in a precinct,it has a statistically significant impact on the proportion of the vote , we can reject the null hypothesis.

(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 $\alpha = .05$).

t-statistic (adjacent) = Coefficient / Standard Error of the Coefficient = 0.042/ 0.013= 3.23

Critical value (adjacent) = 1.979

The absolute value of the t-statistic (3.23) is greater than the critical value (1.979). It suggests there is evidence at the 0.05 significance level that the Precinct adjacent to lawn signs has a statistically significant on the proportion of the vote, so reject the null hypothesis, in conclusion : There is evidence that being next to precincts with assigned yard signs affects the vote

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

The constant term is 0.302 is the estimated intercept of the predictive equation. It represents the estimated portion of the vote when all predictor variables are 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?

R^2 is a measure of the proportion of the variance in the outcome variable , which explained by predictors.

$R^2 = 0.094$, representing that approximately 9.4% of the variability in the proportion of the vote is explained by the variables in the model.

So the overall model fit is quite low, other factors play a significant role in explaining variations in the vote share

The coefficients for predictor variables: "Precinct assigned lawn signs" and "Precinct adjacent to lawn signs" are both 0.042 associated different standard errors. 0.042 represent : on average the estimated change in the proportion of the vote associated with a one-unit change in the corresponding predictor, whiel holding other variables the same. In summary, yard signs may have a specific importance on the vote share from current statistics.