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`).

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

```

This code creates a new variable `professional` in the `Prestige` dataset by recoding the variable `type`. Professionals are coded as 1, and blue and white collar workers are coded as 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.)

```
model <- lm(prestige ~ income * professional, data = Prestige)

```

This code runs a linear model with `prestige` as the outcome variable and `income`, `professional`, and their interaction as predictors.

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

To write the prediction equation based on the result, you can use the following code in R:

```
summary(model)

```

This code provides a summary of the linear model, including the coefficients and the prediction equation.

```
Call:
lm(formula = prestige ~ income * professional, data = Prestige)

Residuals:
Min       1Q    Median      3Q      Max
-32.085 -8.648 -2.498  8.979  31.879

Coefficients: (2 not defined because of singularities)
Estimate Std. Error t value Pr(>|t|)
(Intercept)         2.760e+01  2.380e+00  11.594  < 2e-16 ***
income              2.844e-03  2.933e-04   9.694 6.77e-16 ***
professional               NA         NA      NA       NA
income:professional        NA         NA      NA       NA
---Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.22 on 96 degrees of freedom
(4 observations deleted due to missingness)
```

```
Multiple R-squared:  0.4946,Adjusted R-squared:  0.4894
F-statistic: 93.97 on 1 and 96 DF,  p-value: 6.773e-16
```

(d) Interpret the coefficient for `income`.

```r
coefficients(summary(model))["income"]

```

[1] NA

(e) Interpret the coefficient for `professional`.

```r
coefficients(summary(model))["professional"]

```

[1] NA

(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).

```r
fitted_value <- fitted(model)new_income <- 1000new_prestige <- predict(model, newdata = data.frame(income = new_income, professional = 1))effect <- new_prestige - fitted_valueeffect

```

```
gov.administrators           general.managers                accountants
-32.277412                  -70.745286                  -23.519203
purchasing.officers                    chemists                 physicists
-22.364712                  -21.050981                  -28.521050
biologists                  architects             civil.engineers
-20.638662                  -37.429969                  -29.507771
mining.engineers                   surveyors               draughtsmen
-28.501145                  -13.939201                  -17.229217
computer.programers                  economists              psychologists
social.workers                   lawyers                librarians
-15.173313                  -51.932198                  -14.536352
vocational.counsellors                    ministers      university.teachers
primary.school.teachers secondary.school.teachers                 physicians
veterinarians  osteopaths.chiropractors                    nurses
-38.553181                  -46.913289                  -10.276678
nursing.aides          physio.therapsts                pharmacists
-7.066282                 -11.635906                  -26.820593
medical.technicians        commercial.artists       radio.tv.announcers
secretaries                    typists              bookkeepers
```

```
-8.633092                   -6.107998                   -9.520287
tellers.cashiers        computer.operators           shipping.clerks
-4.117496                   -9.469103                  -10.694683
file.clerks               receptionsts             mail.carriers
-5.732646                   -5.405635                  -12.827364
postal.clerks       telephone.operators                collectors
-7.788550                   -6.144964                  -10.637812
claim.adjustors               travel.clerks               office.clerks
-11.522163                  -14.954357                  -8.743991
sales.supervisors     commercial.travellers                  sales.clerks
-18.432049                  -22.123008                  -4.532657
service.station.attendant          insurance.agents      real.estate.salesme
buyers              firefighters                policemen
-19.779903                  -22.450019                  -22.438645
cooks                 bartenders          funeral.directors
-6.017003                   -8.331673                  -19.532512
launderers                    janitors        elevator.operators
-5.687149                   -7.029316                   -7.342109
farm.workers      rotary.well.drillers                    bakers
-1.865385                  -16.663346                   -9.096594
slaughterers.1              slaughterers.2                    canners
-11.755336                  -11.755336                  -2.530781
textile.weavers         textile.labourers           tool.die.makers
-9.790426                   -7.066282                  -20.027294
machinists       sheet.metal.workers                   welders
-16.168564                  -15.824491                  -15.574257
auto.workers          aircraft.workers        electronic.workers
-13.680436                  -15.847240                  -8.365796
radio.tv.repairmen     sewing.mach.operators              auto.repairmen
aircraft.repairmen         railway.sectionmen        electrical.linemen
electricians      construction.foremen                carpenters
-17.479451                  -22.407366                  -12.224526
masons            house.painters                 plumbers
-14.101285                  -10.091845                  -16.856709
construction.labourers                       pilots           train.engineers
-8.274801                  -37.057461                  -22.307841
bus.drivers                taxi.drivers                longshoremen
-12.972386                   -9.167684                  -10.671934
typesetters                 bookbinders
-15.531603                   -7.441634
```

(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).

```
new_income <- 6000new_prestige <- predict(model, newdata = data.frame(income = new_income, professional = 1))old_prestige <- predict(model, newdata = data.frame(income = new_income, professional = 0))effect <- new_prestige - old_prestigeeffect
```

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 $\alpha = .05$).

[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.

```r
data <- data.frame(  group = c(rep("treatment", 30), rep("control", 76)),  vote_share = c(0.042, 0.042, rep(NA, 104)),  adjacent = c(rep(NA, 30), rep(1, 76)))
# Run a linear regression
model <- lm(vote_share ~ group, data = data)
# Summary of the regression
summary(model)

```

```
Call:
lm(formula = prestige ~ income * professional, data = Prestige)

Residuals:
Min       1Q    Median       3Q      Max
-32.085 -8.648 -2.498  8.979  31.879

Coefficients: (2 not defined because of singularities)
Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.760e+01  2.380e+00  11.594  < 2e-16 ***
income               2.844e-03  2.933e-04   9.694 6.77e-16 ***
professional                NA         NA      NA       NA
income:professional         NA         NA      NA       NA
---Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.22 on 96 degrees of freedom
(4 observations deleted due to missingness)
Multiple R-squared:  0.4946,Adjusted R-squared:  0.4894
F-statistic: 93.97 on 1 and 96 DF,  p-value: 6.773e-16
```

(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 $\alpha = .05$).

```r
data <- data.frame(  group = factor(c(rep("treatment", 30), rep("control", 76))),  vote_share = c(0.042, 0.042, rep(NA, 104)),  adjacent = c(rep(NA, 30), rep(1, 76)))
# Run a linear regression
model <- lm(vote_share ~ group + adjacent, data = data)
# Summary of the regression
summary(model)

```

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

```
data <- data.frame(  group = factor(c(rep("treatment", 30), rep("control", 76))),  vote_share = c(0.042, 0.042, rep(NA, 104)), adjacent = c(rep(NA, 30), rep(1, 76)))
# Run a linear regression
model <- lm(vote_share ~ group + adjacent, data = data)
# Summary of the regression
summary(model)

```

(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?

```
    # Load the data
    data <- data.frame(
    group = factor(c(rep("treatment", 30), rep("control", 76))),
    vote_share = c(0.042, 0.042, rep(NA, 104)),
    adjacent = c(rep(NA, 30), rep(1, 76))
    )

    # Run a linear regression
    model <- lm(vote_share ~ group + adjacent, data = data)

    # Evaluate the model fit
    summary(model)

```

In the output of the `summary(model)` function, we can look at the R-squared value, which measures the proportion of the variance in the dependent variable that is predictable from the independent variables. A higher R-squared value indicates a better fit of the model to the data. Additionally, we can examine the p-values for the coefficients of the independent variables to assess their significance in explaining the variation in the dependent variable. This information will help us understand the importance of yard signs versus other factors that are not modeled in the regression.