Problem Set 1

Applied Stats/Quant Methods 1

Due: October 1, 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 October 1, 2023. No late assignments will be accepted.

- Total available points for this homework is 80.

Question 1 (40 points): Education

A school counselor was curious about the average of IQ of the students in her school and took a random sample of 25 students' IQ scores. The following is the data set:

```
1 y <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113, 112, 98,
    80, 97, 95, 111, 114, 89, 95, 126, 98)
```

1. Find a 90% confidence interval for the average student IQ in the school.

```
1 y <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113,
    112, 98, 80, 97, 95, 111, 114, 89, 95, 126, 98)
2 #sample mean and standard deviation
3 mean<-mean(y)
```

```r
sd<-sd(y)
n<-length(y)
#standard error
se<-sd/sqrt(n)
#degrees of freedom
df<-n-1
#calculate t_value
t_value<-qt(0.95,df)
#margin of error
me<-t_value*se
#calculate confidence interval
a<-mean-me
b<-mean+me
#result
cat("The 90% confidence interval is (",a,""","","",b,")")
```

2. Next, the school counselor was curious whether the average student IQ in her school is higher than the average IQ score (100) among all the schools in the country.

```r
y <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113,
    112, 98, 80, 97, 95, 111, 114, 89, 95, 126, 98)
all_schools_average<-100
#one-sample t_test
t_test<-t.test(y,mu=all_schools_average, alternative="greater")
#p-value
p_value<-t_test$p.value
   <-0.05
#result
cat("p-value:", p_value)
if (p_value <    ) {
  cat("reject H0")
} else {
  cat("accept Ha")
}
```

Using the same sample, conduct the appropriate hypothesis test with $\alpha = 0.05$.

Question 2 (40 points): Political Economy

Researchers are curious about what affects the amount of money communities spend on addressing homelessness. The following variables constitute our data set about social welfare expenditures in the USA.

State	*50 states in US*
Y	*per capita expenditure on shelters/housing assistance in state*
X1	*per capita personal income in state*
X2	*Number of residents per 100,000 that are "financially insecure" in state*
X3	*Number of people per thousand residing in urban areas in state*
Region	*1=Northeast, 2= North Central, 3= South, 4=West*

Explore the `expenditure` data set and import data into `R`.

- Please plot the relationships among *Y*, *X1*, *X2*, and *X3*? What are the correlations among them (you just need to describe the graph and the relationships among them)?

```
1  expenditure <- read.table("https://raw.githubusercontent.com/ASDS-TCD/
     StatsI_Fall2023/main/datasets/expenditure.txt", header=T)
2  par(mfrow=c(2,3))
3  #X1_Y
4  X1_Y<-plot(expenditure$X1,expenditure$Y,main="X1&Y",xlab="X1",ylab="Y")
5  X1_Y
6  #X2_Y
7  X2_Y<-plot(expenditure$X2,expenditure$Y,main="X2&Y",xlab="X2",ylab="Y")
8  X2_Y
9  #X3_Y
10 X3_Y<-plot(expenditure$X3,expenditure$Y,main="X3&Y",xlab="X3",ylab="Y")
11 X3_Y
12 #X1_X2
13 X1_X2<-plot(expenditure$X1,expenditure$X2,main="X1&X2",xlab="X1",ylab="X2
     ")
14 X1_X2
15 #X1_X3
16 X1_X3<-plot(expenditure$X1,expenditure$X3,main="X1&X3",xlab="X1",ylab="X3
     ")
17 X1_X3
18 #X2_X3
19 X2_X3<-plot(expenditure$X2,expenditure$X3,main="X2&X3",xlab="X2",ylab="X3
     ")
20 X2_X3
21 par(mfrow=c(1,1))
```

```
22  #check correlations
23  all_correlation_coes<-c(cor(expenditure$X1,expenditure$Y),cor(expenditure
        $X2,expenditure$Y),cor(expenditure$X3,expenditure$Y),cor(expenditure$
        X1,expenditure$X2),cor(expenditure$X1,expenditure$X3),cor(expenditure$
        X2,expenditure$X3))
24  all_correlation_coes
25  for(j in all_correlation_coes){
26      if(j>=0){
27          if(j>0){
28              print("positive correlation")
29          }else{
30              print("irrelevant")
31          }
32      }else{
33          print("negative correlation")
34      }
35  }
```

- Please plot the relationship between *Y* and *Region*? On average, which region has the highest per capita expenditure on housing assistance?

```
1   Region_1<-mean(expenditure[expenditure$Region == 1,]
2       $Y)
3   Region_2<-mean(expenditure[expenditure$Region == 2,]
4               $Y)
5   Region_3<-mean(expenditure[expenditure$Region == 3,]
6               $Y)
7   Region_4<-mean(expenditure[expenditure$Region == 4,]
8               $Y)
9   region_cat = c("1", "2", "3", "4")
10  region_mean = c(79, 84, 69, 88)
11  Region_Y_bar<-barplot(region_mean, names.arg = region_cat, main = "Region
        _Y_bar",
12          xlab = "region_cat", ylab = "region_mean")
13  result<-cat("the region of NO.4 has the highest per capita expenditure on
          housing assistance")
```

- Please plot the relationship between *Y* and *X1*? Describe this graph and the relationship. Reproduce the above graph including one more variable *Region* and display different regions with different types of symbols and colors.

```
1   region_car_var<-as.character(expenditure$Region)
2   region_car_var
3   colors <- c("red", "blue", "green", "yellow")
4   for (i in 1:length(region_car_var)) {
5       points(expenditure[[i]]$X1, expenditure[[i]]$Y, col = colors[i], pch =
          19)
6   }
7   legend("topright", legend = unique(expenditure$Region), col = colors, pch
          = 19)
```