Problem Set 1

Applied Stats/Quant Methods 1

Due: September 30, 2024

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 Monday September 30, 2024. No late assignments will be accepted.

Question 1: 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:

```
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.

```
#Defining
iq_scores <- 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)
# Calculate the sample mean
mean_iq <- mean(iq_scores)
#Calculate sample standard deviation
sd_iq <- sd(iq_scores)
# Calculate n-1 (degrees of freedom)
n <- length(iq_scores) df <- n - 1
```

```
# Find the t-value for a 90% confidence level (0.1 for two-sided and 0.05 for one-sided)
t_ci <- qt(0.95, df)
# Because it is a two-sided 0.1, the t value of 0.05 on one side is checked and then used for the two-sided calculation
# Calculating the margin of error
Margin_of_error <- t_ci * (sd_iq / sqrt(n))
# Calculating confidence intervals
ci_lower <- mean_iq - margin_of_error ci_upper <- mean_iq + margin_of_error
# Output
cat("90%Confidence interval: [", ci_lower, ", ", ci_upper, "]\n")
```

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.

```
using the same sample, conduct the appropriate hypothesis test with Q = 0.05.
# National average IQ
mu_0 <- 100
# t-test
t_test_result <- t.test(iq_scores, mu = mu_0, alternative = "greater", conf.level = 0.95)
# Output
print(tt检验_test_result)
```

Question 2: Political Economy

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

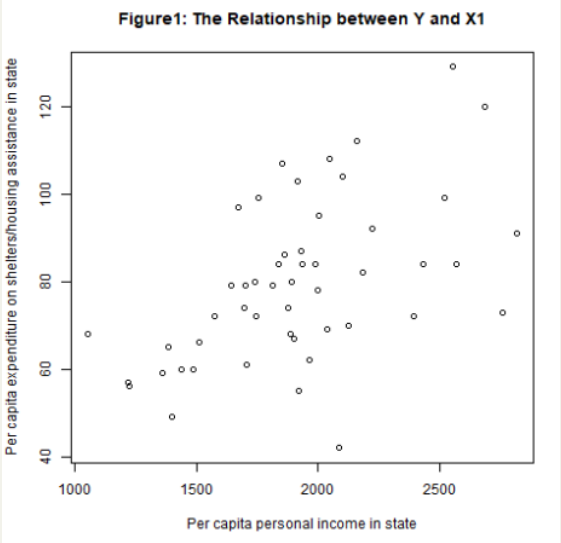
State	50 states in US
Y	per capita expenditure on shelters/housing assistance in state
X1	per capita personal income in state
X2	Number of residents per 100,000 that are "financially insecure" in state
X3	Number of people per thousand residing in urban areas in state
Region	1=Northeast, 2= North Central,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

Explore the expenditure data set and import data into R.

```
expenditure <- read.table("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2024/main/datasets/expenditure.txt", header=T)
```

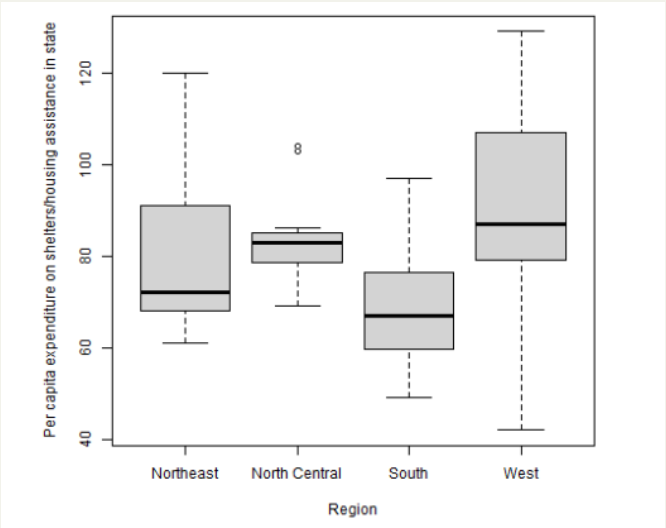
- Please plot the relationships among Y, X1, X2, and X3? What are the correlations among them (you just need to describe the graph and the relationships among them)?

```
Setwd(dirname(rstudioapi : : getActiveDocumentContext()$path))
Png(file = "Y~X1.png"
Plot ( expenditure $X1,
      expenditure $Y,
       Ylab = "Per  capita expenditure on shelters / housing assistance i n
state " ,
      Xlab = "Per capita personal income in state"
      Main = "Figure1: The Relationship between Y and X1")
Dev.off()
Cor (expenditure$Y, expenditure $X1)
```



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

```
Png(file = “Y~Region.png”)
Boxplot( expenditure $Y~ expenditure$ Region,
          Ylab = “per capita expenditure on shelters/housing assistance in state”,
          Xlab = “Region”,
         Main =” Figure2:The Relationship between Y and Region”,
         Names= c(“Northeast”, “North Central”,”South”, “West”))
Dev.off()
Print(“The bos plot indicates that West Region 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 diferent regions with diferent types of symbols and colors.

```
png ( f i l e = ”Y˜X1 Region . png” )
 plot ( expenditure $X1,
         expenditure$Y,
         col = expenditure$Region ,
         pch = 19 ,
         ylab = ”Per capita expenditure on shelters / housing assistance in
        state ” ,
        xlab = ”Per capita personal income in state ” ,
        main = ”Figure8 : Relationship between Y and X1” )
 legend ( ” topleft ” ,
            legend = c ( "Northeast ” , ”North Central ” , ”South” ,"West” ) ,
```

```
            col = c ( ”1” , ”2” , ”3” , ”4” ) ,
            pch = 19 ,
            cex = 0.8 )
for ( region in unique ( expenditure $Region ) ) {
   region_data <- subset ( expenditure , Region == region )
    fit <- lm(Y˜X1, data = region_data )
   abline ( fit , col = region ) }
dev . o f f ( )
regression1 <- lm(Y˜X1, data=expenditure )
regression1
print ( ”Figure3 indicates that as per capita personal income increases ,
the per capita expenditure  on housing assistance also increases accordingly.This suggests that
 states with higher economic development and per capita income may be more inclined to invest more
 funds in housing assistance . ” )
```