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:

```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)
```

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

```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)

### 1.1 ### 90% confidence interval
upper <- round(qnorm(0.95),2)
lower <- round(qnorm(0.05),2)
```

```r
# mean
mean_y <- mean(y)
# sd
sd_y <- sd(y)/sqrt(length((y)))
# upper&lower bound
upper_90 = mean_y+upper*sd_y
lower_90 = mean_y+lower*sd_y
# output
cat("90% confidence interval for the average student IQ in the school is :(", round(lower_90, 2), ",", round(upper_90, 2),")")
```

Answer: 90% confidence interval for the average student IQ in the school is: (94.15 , 102.73)

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 $\alpha = 0.05$.

```r
### 1.2 ### t test
print(t.test(y, mu = 100))
print("t-test results indicate that the average student IQ in this school is not significantly different from the average IQ score of 100 among all schools in the country( p = 0.5569 > 0.05)")
```

Answer: t-test results indicate that the average student IQ in this school is not significantly different from the average IQ score of 100 among all schools in the country(p = 0.5569 >0.05)

Question 2: 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*

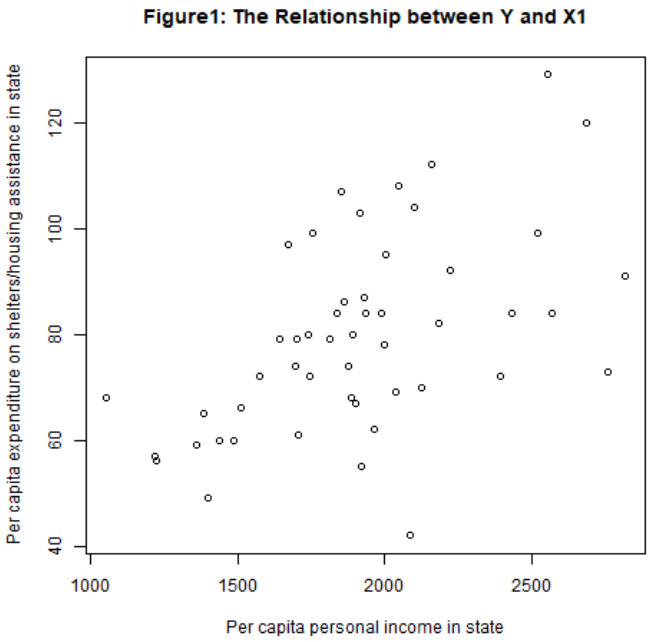
Explore the `expenditure` data set and import data into `R`.

```
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)?

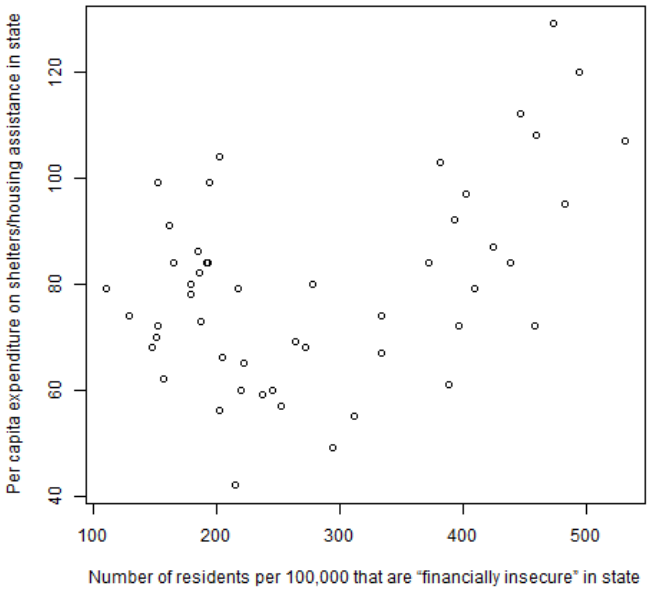
```
### 2.1 ###
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
## Plot Y~X1
png(file = "Y~X1.png")
plot( expenditure$X1,
      expenditure$Y,
      ylab = "Per capita expenditure on shelters/housing assistance in state",
      xlab = "Per capita personal income in state",
      main = "Figure1: The Relationship between Y and X1")
dev.off()
cor(expenditure$Y, expenditure$X1)
print("The correlation coefficient between Y and X1 is 0.531, Figure 1 illustrates that as X1 increases, the value of Y also increases gradually.")

## Plot Y~X2
png(file = "Y~X2.png")
plot( expenditure$X2,
      expenditure$Y,
      ylab = "Per capita expenditure on shelters/housing assistance in state",
      xlab = "Number of residents per 100,000 that are    financially insecure  in state",
      main = "Figure2: The Relationship between Y and X2")
dev.off()
cor(expenditure$Y, expenditure$X2)
```

```
print("The correlation coefficient between Y and X2 is 0.448, Figure 2 illustrates that as X2 increases, Y decreases until X2 reaches approximately 300, at which point Y begins to increase with further increases in X2.")

## Plot Y~X3
png(file = "Y~X3.png")
plot( expenditure$X3,
      expenditure$Y,
      ylab = "Per capita expenditure on shelters/housing assistance in state",
      xlab = "Number of people per thousand residing in urban areas in state",
      main = "Figure3: The Relationship between Y and X3")
dev.off()
cor(expenditure$Y, expenditure$X3)
print("The correlation coefficient between Y and X3 is 0.463, Figure 3 illustrates that as X3 increases, the value of Y also increases gradually.")
```



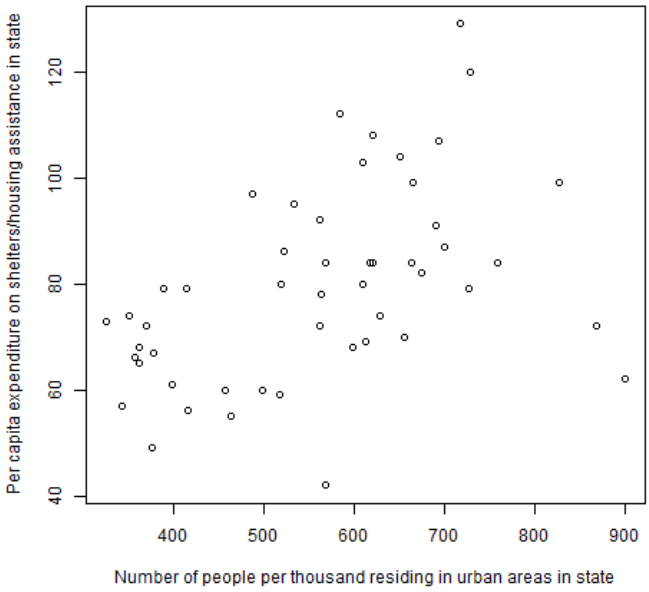
The correlation coefficient between Y and X1 is 0.531, Figure 1 illustrates that as X1 increases, the value of Y also increases gradually.

Figure2: The Relationship between Y and X2



The correlation coefficient between Y and X2 is 0.448, Figure 2 illustrates that as X2 increases, Y decreases until X2 reaches approximately 300, at which point Y begins to increase with further increases in X2.

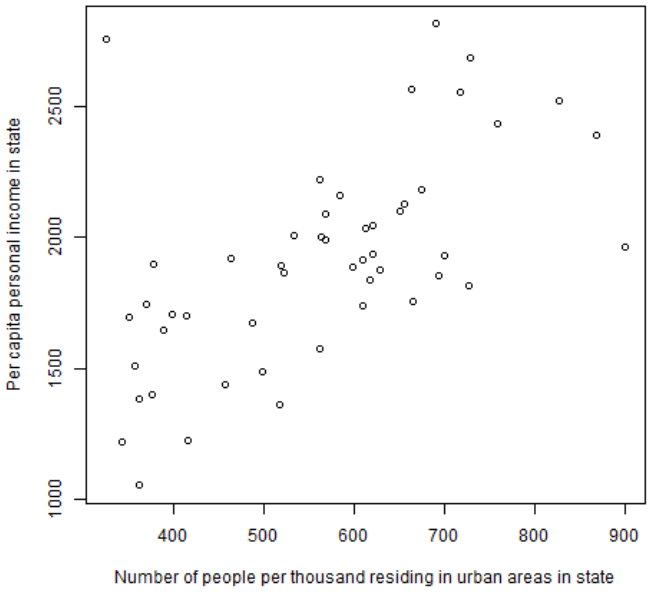
Figure3: The Relationship between Y and X3



The correlation coefficient between Y and X3 is 0.463, Figure 3 illustrates that as X3 increases, the value of Y also increases gradually.

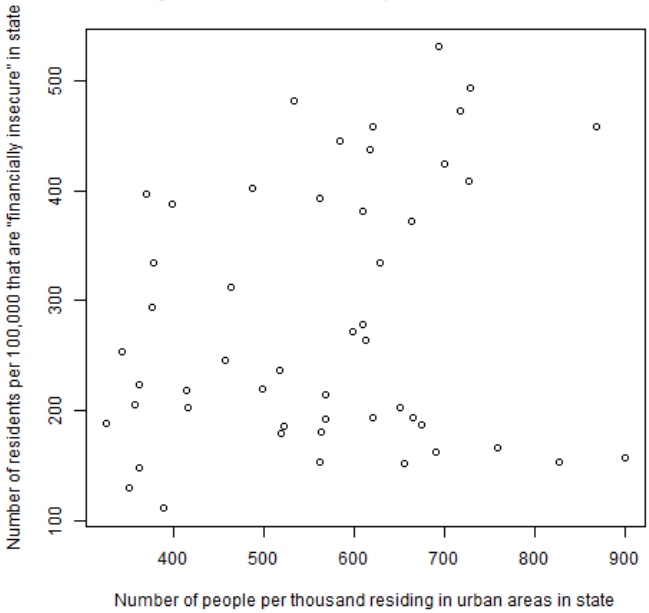
```
## Plot X1~X3
png(file = "X1~X3.png")
plot( expenditure$X3,
      expenditure$X1,
      ylab = "Per capita personal income in state",
      xlab = "Number of people per thousand residing in urban areas in state",
      main = "Figure4: The Relationship between X1 and X3")
dev.off()
cor(expenditure$X1, expenditure$X3)
print("The correlation coefficient between X1 and X3 is 0.595, Figure4 demonstrate that as X3 increase, X1 also rises correspondingly.")

## Plot X2~X3
png(file = "X2~X3.png")
plot( expenditure$X3,
      expenditure$X2,
      ylab = "Number of residents per 100,000 that are    financially insecure  in state",
      xlab = "Number of people per thousand residing in urban areas in state",
      main = "Figure5: The Relationship between X2 and X3")
dev.off()
cor(expenditure$X2, expenditure$X3)
print("The correlation coefficient between X2 and X3 is 0.221, indicating a weak positive correlation between X2 and X3. Figure 5 illustrates this relationship.")

## Plot X1~X2
png(file = "X1~X2.png")
plot( expenditure$X2,
      expenditure$X1,
      ylab = "Per capita personal income in state",
      xlab = "Number of residents per 100,000 that are    financially insecure  in state",
      main = "Figure6: The Relationship between X1 and X2")
dev.off()
cor(expenditure$X1, expenditure$X2)
```

Figure4: The Relationship between X1 and X3



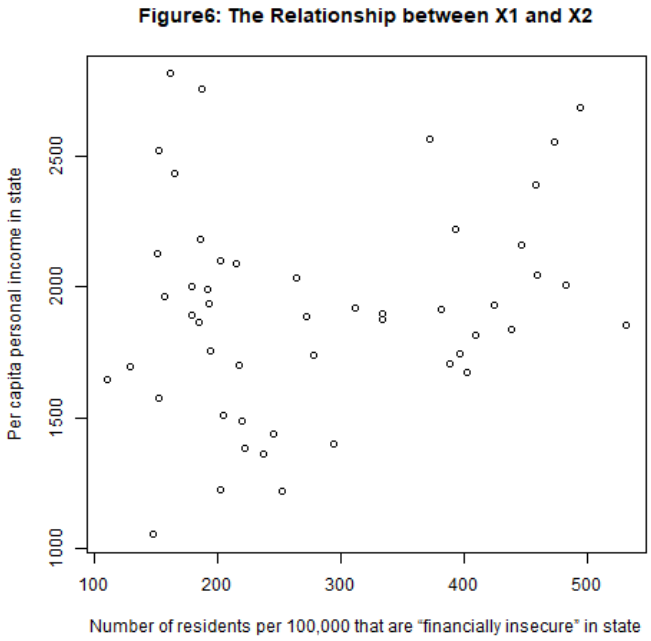
The correlation coefficient between Y and X3 is 0.595, Figure4 demonstrate that as X3 increase, X1 also rises correspondingly.

Figure5: The Relationship between X2 and X3



The correlation coefficient between Y and X3 is 0.463, Figure 3 illustrates that as X3

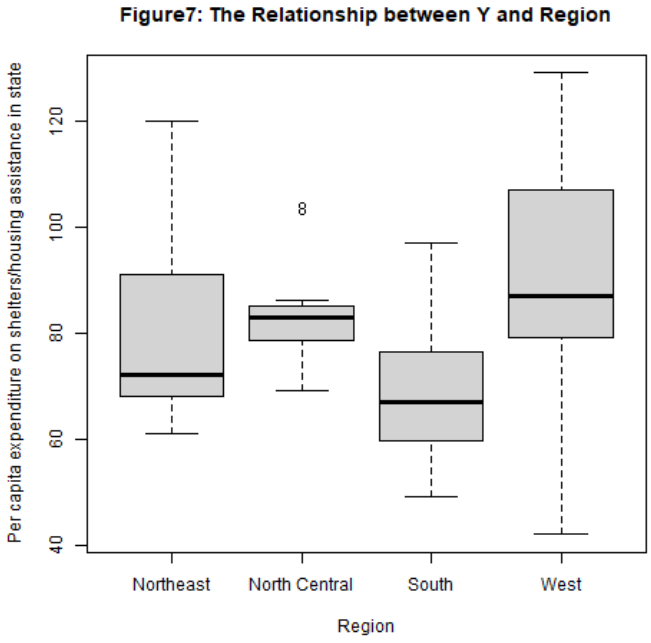
increases, the value of Y also increases gradually.



The correlation coefficient between Y and X3 is 0.463, Figure 3 illustrates that as X3 increases, the value of Y also increases gradually.

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

```
### 2.2 ###
png(file = "Y~Region.png")
boxplot( expenditure$Y ~ expenditure$Region,
         ylab = "Per capita expenditure on shelters/housing assistance in
    state",
         xlab = "Region",
         main = "Figure7: The Relationship between Y and Region",
         names=c("Northeast", "North Central","South","West"))
dev.off()
print("The box plot (Figure7) indicates that West Region (Region 4) has
   the highest per capita expenditure on housing assistance.")
```



The box plot (Figure7) indicates that West Region (Region 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.

```
### 2.3 ###
png(file = "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.off()
regression1 <- lm(Y~X1, data=expenditure)
regression1
print("Figure8 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.")
```

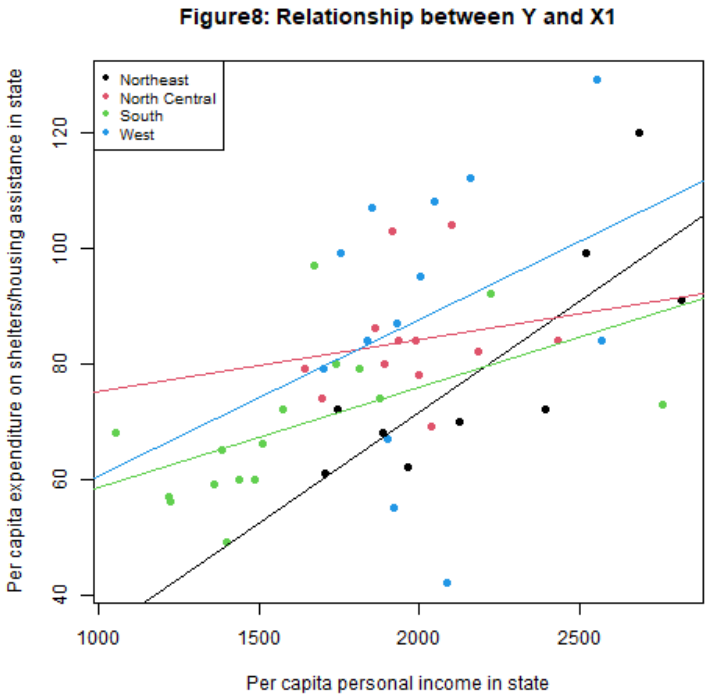


Figure8 indicates that as per capita personal income increases, the per capita expenditure on housing assistance also increases accordingly. This suggests that states with higher economic development and per capita income may be more inclined to invest more funds in housing assistance.