

Problem Set 1

Applied Stats/Quant Methods 1
Clare Zureich

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:

```
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.

90% Confidence Interval = [93.96, 102.92]

```

1 mean <- mean(scores)
2
3 sd <- sd(scores)
4 se <- sd/sqrt(length(scores))
5
6 # Find the confidence interval, using t-score, for 90%
7
8 t <- qt((.9+(1-.9)/2), df=length(scores)-1)
9 t
10 lower_90 <- mean - t*se
11 upper_90 <- mean + t*se

```

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 $\alpha = 0.05$.

Based on the p-value of .72 and a 95% confidence level, we fail to reject the null hypothesis that the average IQ scores are less than or equal to 100.

```

1 t.test(scores, mu = 100, alternative = "greater", conf.level = 0.95)

```

Question 2: Political Economy

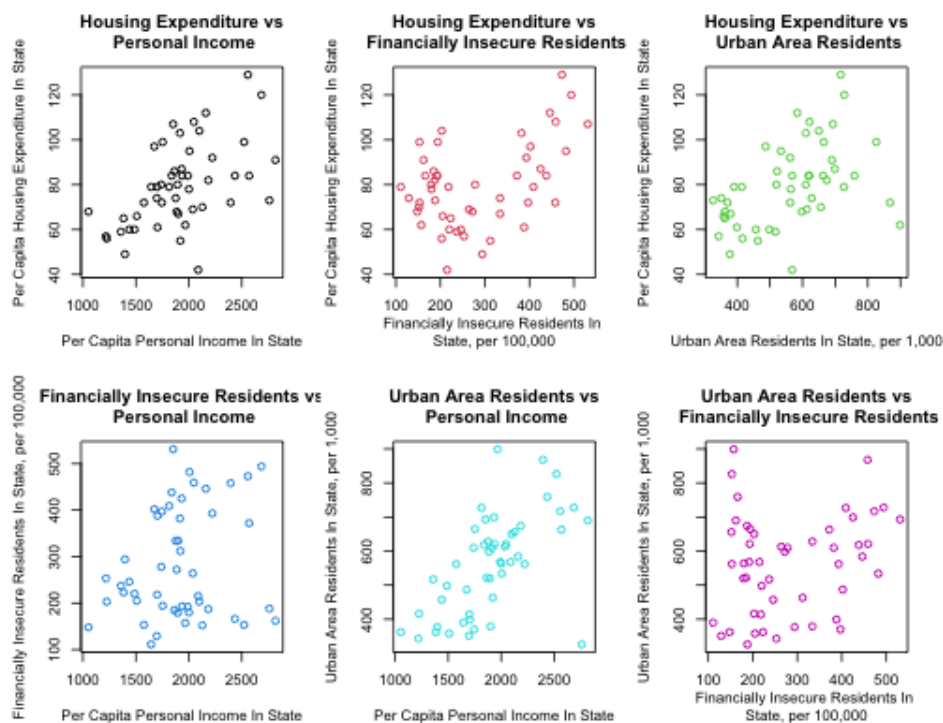
Researchers are curious about what affects the amount of money communities spend on addressing homelessness. The following variables constitute our data set about social welfare expenditures in the USA.

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.

- 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)?

Figure 1: Figure 1.



```

1 png( file=" political_economy_scatter_plot.png" ,
2     width = 500,
3     height = 400)
4 par(mfrow = c(2,3))
5 plot(expenditure$X1, expenditure$Y,col=1, ylab = "Per Capita Housing
    Expenditure In State", xlab = "Per Capita Personal Income In State",
    main = "Housing Expenditure vs\n Personal Income")
6 plot(expenditure$X2, expenditure$Y,col=2, ylab = "Per Capita Housing
    Expenditure In State", xlab = "Financially Insecure Residents In \
    nState, per 100,000", main = "Housing Expenditure vs\n Financially
    Insecure Residents")
7 plot(expenditure$X3, expenditure$Y,col=3, ylab = "Per Capita Housing
    Expenditure In State", xlab = "Urban Area Residents In State, per
    1,000", main = "Housing Expenditure vs\n Urban Area Residents")
8 plot(expenditure$X1, expenditure$X2,col=4, xlab = "Per Capita Personal
    Income In State", ylab = "Financially Insecure Residents In State, per
    100,000", main = "Financially Insecure Residents vs\n Personal Income
    ")
9 plot(expenditure$X1, expenditure$X3,col=5, xlab = "Per Capita Personal
    Income In State", ylab = "Urban Area Residents In State, per 1,000",
    main = "Urban Area Residents vs\n Personal Income")
10 plot(expenditure$X2, expenditure$X3,col=6, xlab = "Financially Insecure
    Residents In \nState, per 100,000", ylab = "Urban Area Residents In
    State, per 1,000", main = "Urban Area Residents vs\n Financially
    Insecure Residents")
11 dev.off()

```

Graph 1: Loose linear positive correlation

Graph 2: Non-linear correlation

Graph 3: Loose positive correlation

Graph 4: Nonlinear correlation

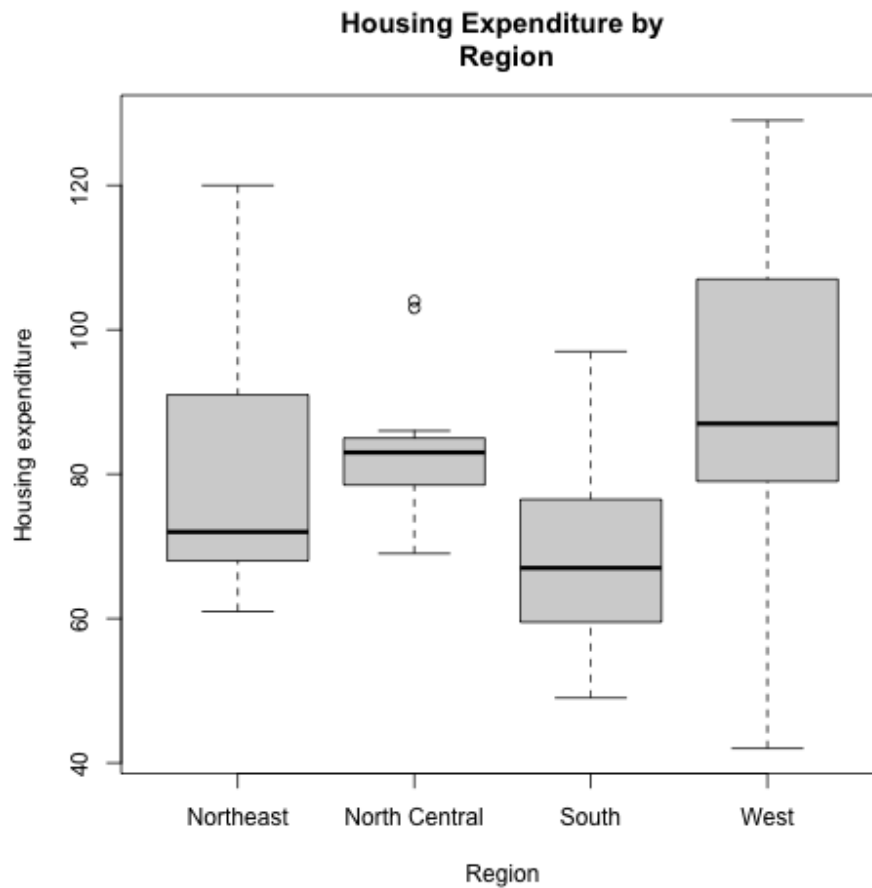
Graph 5: Loose linear positive

Graph 6: Nonlinear correlation

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

Region 4, the West, has the highest average of per capita expenditure on housing assistance

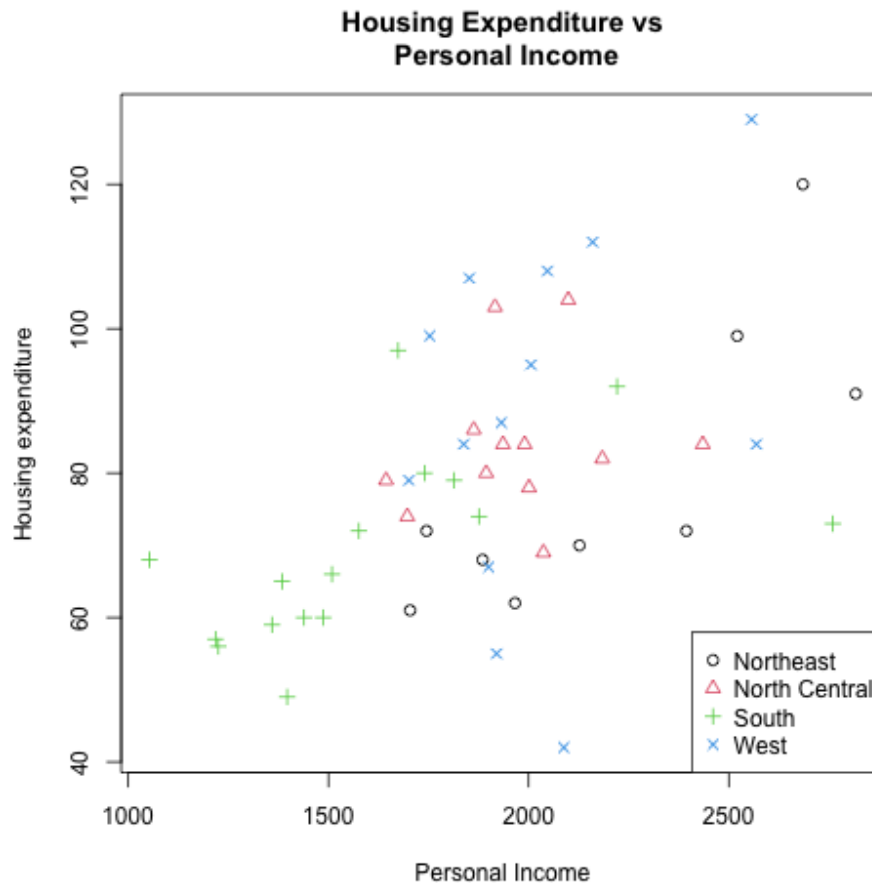
Figure 2: Figure 2.



```
1 png(file="housing_expenditure_region__scatter_plot.png")
2 par(mfrow = c(1,1))
3 region <- factor(expenditure$Region, levels = c(1, 2, 3, 4), labels = c("
  Northeast", "North Central", "South", "West"))
4 expenditure$name <- region
5 plot(region, expenditure$Y, ylab = "Housing expenditure", xlab = "Region"
  , main = "Housing Expenditure by\n Region")
6 #The west, Region 4, has the highest per capita spending on housing
7 dev.off()
```

- 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.

Figure 3: Figure 3.



```
1 png(file="housing_expenditure_personal_income__scatter_plot.png")
2 plot(expenditure$X1, expenditure$Y, col=expenditure$Region,
3       ylab = "Housing expenditure",
4       pch = expenditure$Region,
5       xlab = "Personal Income",
6       main = "Housing Expenditure vs\n Personal Income")
7 legend("bottomright", legend=unique(expenditure$name),
8       col=unique(expenditure$Region),
9       pch=unique(expenditure$Region))
10 dev.off()
```