Problem Set 1

Applied Stats/Quant Methods 1

Due: October 9, 2025

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 Thursday October 9, 2025. 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. Find a 90% confidence interval for the average student IQ in the school.

```r
# part 1: 90% CI
n <- length(na.omit(student_iqs))
sample_mean <- mean (student_iqs, na.rm = TRUE)
sample_sd <- sd(student_iqs, na.rm = TRUE)
t90 <- qt((1 - .90)/2, df = 24, lower.tail = FALSE)
lower_90 <- sample_mean- (t90 * (sample_sd/sqrt(n)))
upper_90 <- sample_mean + (t90 * (sample_sd/sqrt(n)))
confint_90 <- c(lower_90, upper_90)

#We are 90% confident that the true population mean of student IQ in the
    school lie around 93.96 and 102.92.
```

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
#Ho = 100
#Ha > 100
t.test(student_iqs, mu = 100, alternative = "greater")
print(critical_t <- qt(1 - 0.05, 24))

#For this hypothesis test, the null hypothesis is that the population average of student IQ is 100. However, the alternative hypothesis states that the average is higher than 100. Therefore, this question is looking for a one sided t test, testing if there is sufficient evidence to reject the null. Using the information about the sample that was found from the part above (mean and standard deviation), the t statistic can be calculated with this equation: t = (sample mean − hypothesized population mean)/(standard deviation/square root of the sample size). In order to discover if the null hypothesis can be rejected the calculated t value must be compared to a critical t value. In this case, the code above shows how to find the the critical t value in r. In order to find the critical t value three things are needed: type of test, alpha value, and the degrees of freedom. Since this test is a one sided test, with the alternative hypothesis stating the true population mean is larger than the hypothesized mean, the problem is looking at the upper tail, which is why the code shows 1 − 0.05 (alpha value). In this problem the critical value (1.712) is greater than the calculated t value (−0.596) therefore we fail to reject the null hypothesis and there is insufficient evidence to support the alternative hypothesis.
```

Question 2: Political Economy

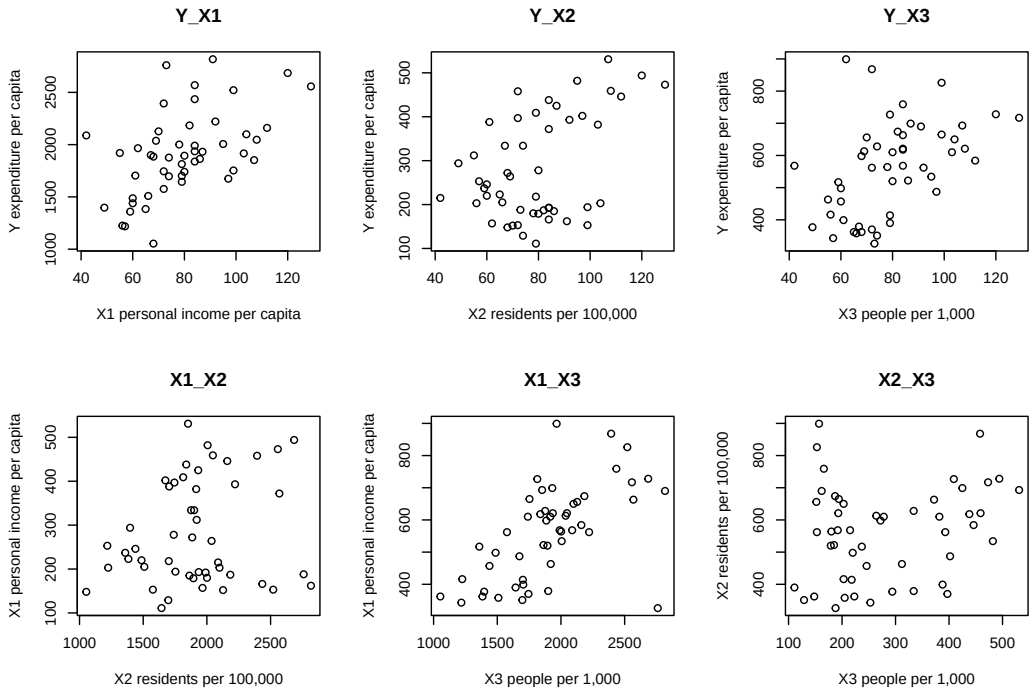
Researchers are curious about what affects the amount of money communities spend on addressing homelessness. The following variables constitute our data set about social welfare expenditures in the USA.

`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)?

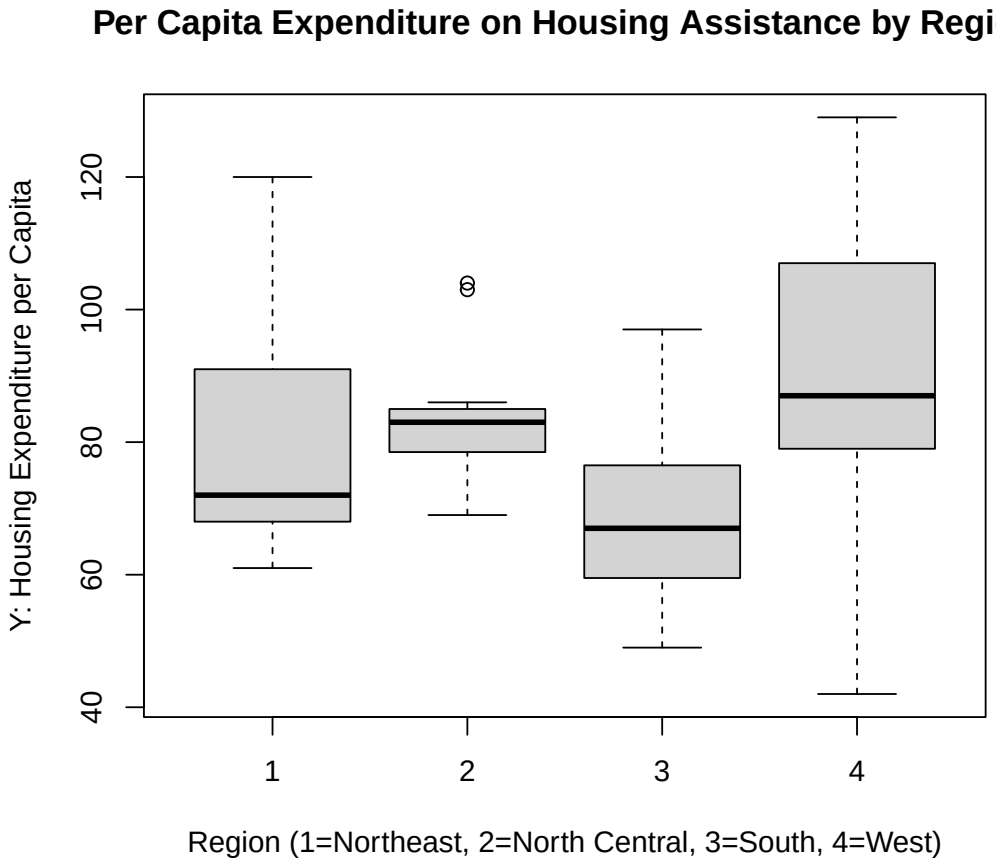
Figure 1: Relationships between Y, X1, X2, and X3.




```
#For the Y_X1 graph, the data is looking at the relationship between per
    capita expenditure on shelter/housing assistance and per capita
    personal income in state. The graph shows that as the personal income
    increases, so does the expenditure on shelter and housing assistance,
    therefore, we can say there is a correlation. However, the points aren
    't perfectly lined up so, the correlation is weak to moderate.
#For the Y_X2 graph, the data is looking at the relationship between per
    capita expenditure on shelter/housing assistance and number of
    residents that are "financially insecure" per 100,000 residents. The
    graph shows that as "finacially insecure" rsidents rise by 100,000,
    expenditure doesn't always increase as well. In certain areas were the
     expenditure is 300 and above there is more of a correlation, but
    lower than that the correlation is more scattered.
```

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

Figure 2: Relationships Y and Region.



- 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: Relationships Y and Region.

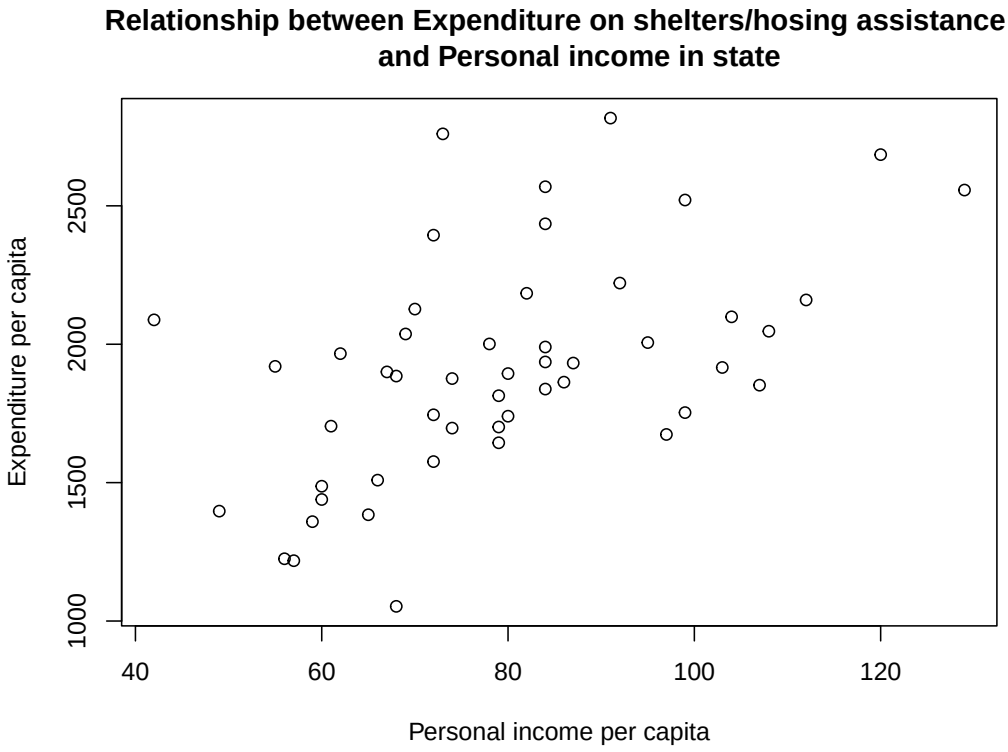


Figure 4: Relationships Y and Region.