

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.

```
y <- c(105, 69, 86, 100, 82, 111, 104, 110, 87, 108, 87, 90, 94, 113, 112, 98, 80,
```

```
Short version: n<30: t-test for CI  
t.test(y, conf.level = 0.90)
```

```
Long version:
```

```

n <- length(y)
mean_y <- mean(y)
sd_y <- sd(y)
se_y <- sd_y / sqrt(n)

Using critical value from t-distribution
t90 <- qt(1 - (0.10/2), df = n - 1)
ci_90_lower_t <- mean_y - t90 * se_y
ci_90_upper_t <- mean_y + t90 * se_y

```

Answer: The results show a mean of 98.44, with a confidence interval of (93.9 ; 102.9).

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 $\alpha = 0.05$.

Explanation: For this task, we will conduct a one sided hypothesis test. Our null hypothesis (H_0) states that the school mean is 100. Our alternative hypothesis (H_1) states that the school mean is greater than 100. To support the alternative hypothesis, we need sufficient evidence to reject the H_0 .

```

.test(y, mu = 100, alternative = "greater", conf.level = 0.95)

# test statistics (TS)
TS <- (mean_y-100)/se_y
pt(q=TS, df = 24)

# variant 1
1-pt(q=TS, df = 24)
# variant 2
pt(q=TS, df = 24, lower.tail = FALSE)

```

Interpretation: The results give us a p-value of 0.72, which is the probability of seeing a value as extreme as this one (or higher) if the H_0 was true. As such, we do not reject the H_0 , and the school's average IQ is not greater than the national average of 100.

Question 2: Political Economy

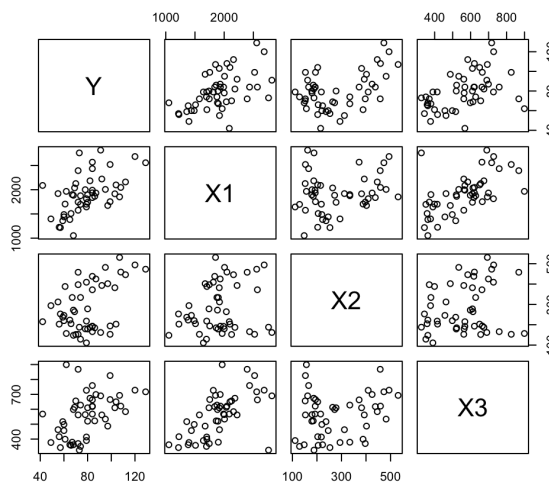
Researchers are curious about what affects the amount of money communities spend on addressing homelessness. The following variables constitute our data set about social welfare expenditures in the USA.

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)?

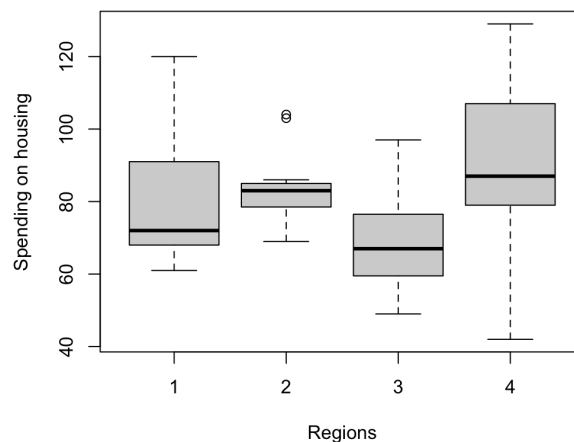
```
expenditure <- read.table("https://raw.githubusercontent.com/ASDS-TCD/StatsI_2025/")
head(expenditure)
str(expenditure)
pairs(expenditure[,c(2, 3, 4, 5)])
```



Answer: In this plot matrix, we observe positive correlations for the plots (X1, Y), (X1, X3). For (X3, Y), the correlation is not as clear. For (X2, Y), (X2, X1), (X2, X3) there is no clear linear relation, but rather some form of relation (one can see a curvature in the plot).

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

```
plot(x = expenditure$X1, y = expenditure$Y)
```



Answer: It appears that the Region "West" has the highest spending, while Region "South" has the lowest spending.

- 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.

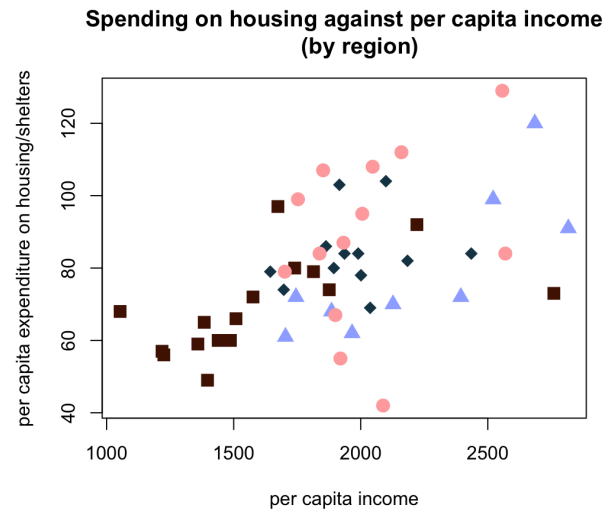
```
colors <- paletteer_c("scico::berlin", n=4)
shapes <- c(17, 18, 15, 19)

plot(x = expenditure$X1,
     y = expenditure$Y,
```

```

col = colors[expenditure$Region],
pch = shapes[expenditure$Region],
cex = 1.5,
main = "Spending on housing against per capita income
(by region)",
xlab = "per capita income",
ylab = "per capita expenditure on housing/shelters")

```



Answer: There appears to be a positive relationship/correlation between X_1 and Y .