

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:

```
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.

```
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)
2
3 ### 1.1 ### 90% confidence interval
4 upper <- round(qnorm(0.95), 2)
5 lower <- round(qnorm(0.05), 2)
```

```

6 # mean
7 mean_y <- mean(y)
8 # sd
9 sd_y <- sd(y)/sqrt(length((y)))
10 # upper&lower bound
11 upper_90 = mean_y+upper*sd_y
12 lower_90 = mean_y+lower*sd_y
13 # output
14 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 $\alpha = 0.05$.

```

1 ### 1.2 ### t test
2 print(t.test(y, mu = 100))
3 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.

```
1 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)?

```
1 ### 2.1 ###
2 setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
3 ## Plot Y~X1
4 png(file = "Y~X1.png")
5 plot(expenditure$X1,
6       expenditure$Y,
7       ylab = "Per capita expenditure on shelters/housing assistance in
8               state",
9       xlab = "Per capita personal income in state",
10      main = "Figure1: The Relationship between Y and X1")
11 dev.off()
12 ## Plot Y~X2
13 png(file = "Y~X2.png")
14 plot(expenditure$X2,
15       expenditure$Y,
16       ylab = "Per capita expenditure on shelters/housing assistance in
17               state",
18       xlab = "Number of residents per 100,000 that are financially
19               insecure in state",
20       main = "Figure2: The Relationship between Y and X2")
21 dev.off()
22 ## Plot Y~X3
23 png(file = "Y~X3.png")
24 plot(expenditure$X3,
25       expenditure$Y,
26       ylab = "Per capita expenditure on shelters/housing assistance in
27               state",
```

```

24     xlab = "Number of people per thousand residing in urban areas in
      state",
25     main = "Figure3: The Relationship between Y and X3")
26 dev.off()
27 print("The scatter plots demonstrate that as X1 and X3 increase, Y also
      rises correspondingly (Figure1, 3). As X2 increases, Y decreases until
      X2 reaches approximately 300, at which point Y begins to increase
      with further increases in X2 (Figure 2).")

```

Figure1: The Relationship between Y and X1

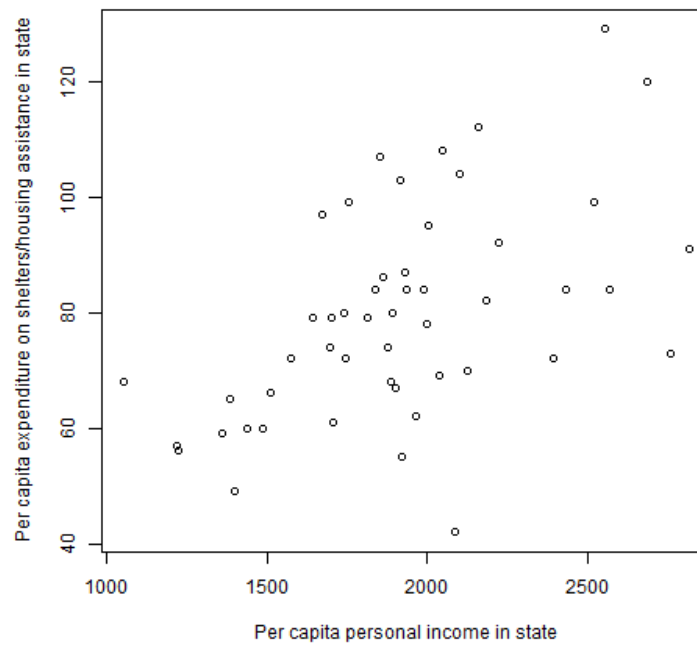


Figure2: The Relationship between Y and X2

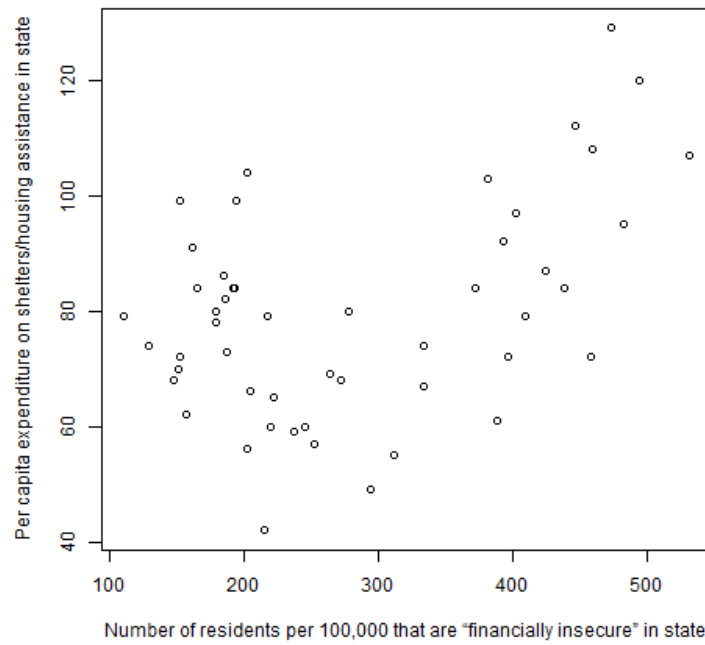
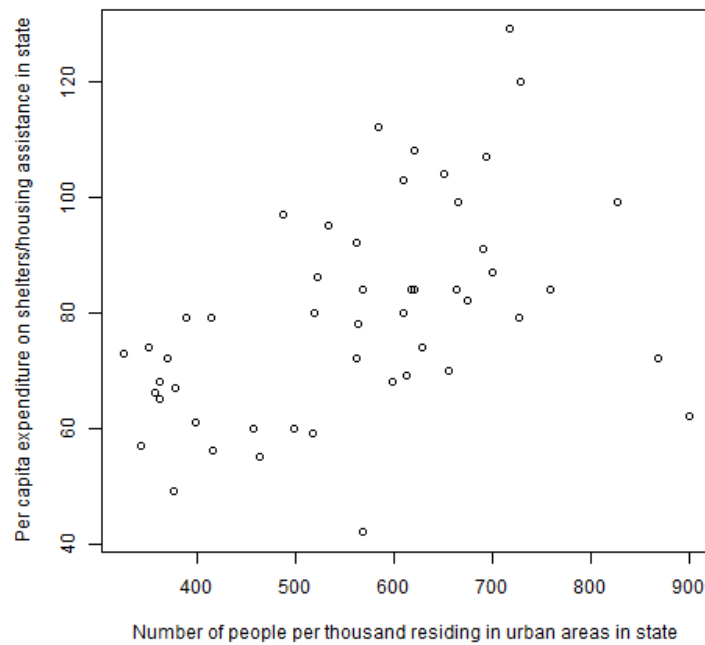


Figure3: The Relationship between Y and X3



The scatter plots demonstrate that as X1 and X3 increase, Y also rises correspond-

ingly (Figure1, 3). As X2 increases, Y decreases until X2 reaches approximately 300, at which point Y begins to increase with further increases in X2 (Figure 2).

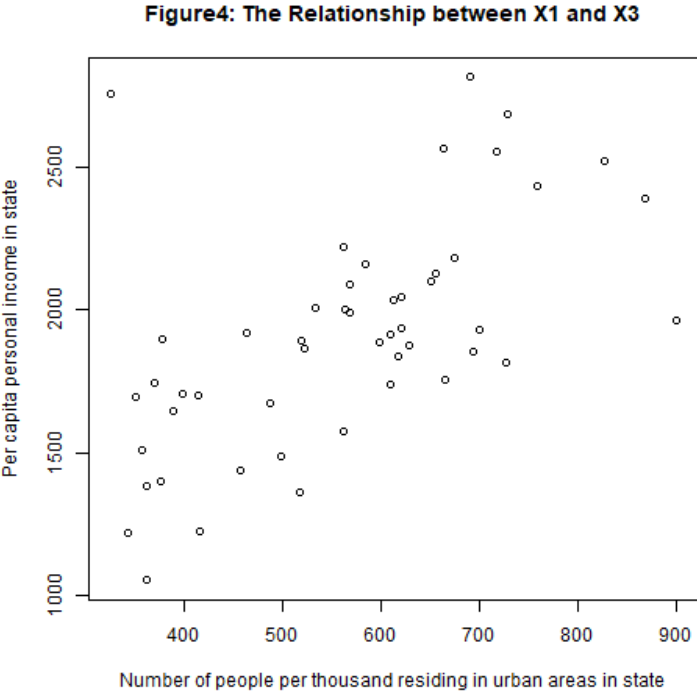


Figure5: The Relationship between X2 and X3

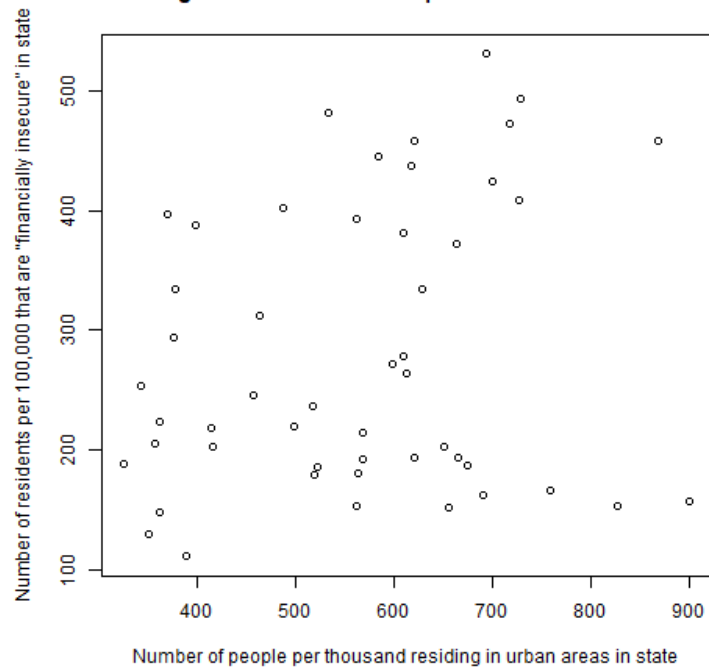
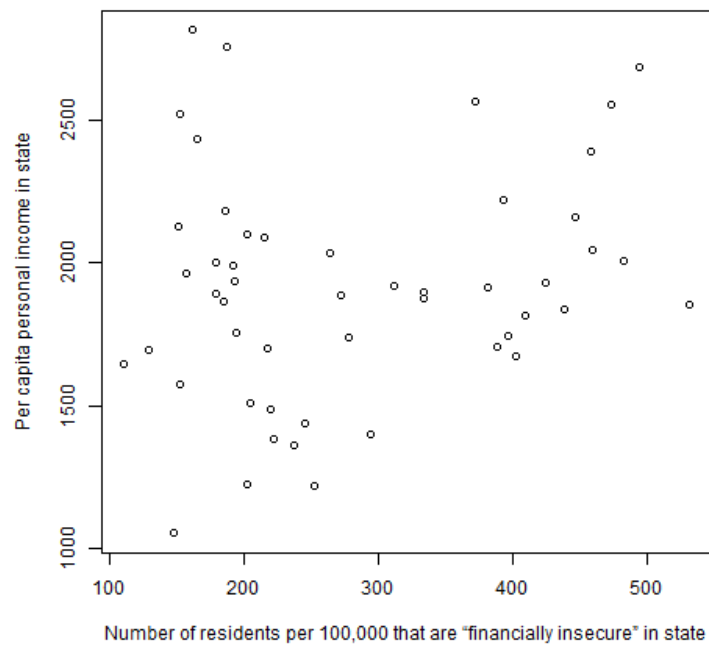


Figure6: The Relationship between X1 and X2



- Please plot the relationship between Y and $Region$? On average, which 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 different regions with different types of symbols and colors.