

# Assignment 1

Ella Karagulyan

Applied Stats/Quant Methods 1

## Question 1

*A school counselor was curious about the average IQ of the students in her school and took a random sample of 25 students' IQ scores.*

First, we load the dataset.

```
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)
```

Next, we find the 90% confidence interval for the mean student IQ in the school.

```
1 t.test(y, conf.level = 0.90, alternative = "two.sided")
```

The mean student IQ in the school is 98.44 with 90% CI [93.96, 102.92]. Below are the results of the One-sample t-test.

One Sample t-test

```
data: y
t = 37.593, df = 24, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
90 percent confidence interval:
93.95993 102.92007
sample estimates:
mean of x
98.44
```

*Next, the school counselor was curious whether the mean student IQ in her school is higher than the mean IQ score (100) among all the schools in the country.*

To test this we need to conduct a one-tailed t-test.

```
1 t.test(y, mu = 100, alternative = "greater")
```

The One-sample t-test results provide no statistically significant evidence to conclude that the mean students' IQ of 98.4 (95% CI[93.96,  $\infty$ ]) is greater than 100,  $t(24)=-0.6$ ,  $p=0.72$ .

#### One Sample t-test

```
data: y
t = -0.59574, df = 24, p-value = 0.7215
alternative hypothesis: true mean is greater than 100
95 percent confidence interval:
93.95993      Inf
sample estimates:
mean of x
98.44
```

We also checked whether mean students' IQ is equal to the mean IQ score (100) among all the schools in the country.

```
1 t.test(y, mu = 100)
```

The One-sample t-test results provide no statistically significant evidence to conclude that the mean students' IQ of 98.4 (95% CI[93.03, 103.84]) is equal to 100,  $t(24)=-0.6$ ,  $p=0.56$ . Below are the results of the One-sample t-test.

#### One Sample t-test

```
data: y
t = -0.59574, df = 24, p-value = 0.5569
alternative hypothesis: true mean is not equal to 100
95 percent confidence interval:
93.03553 103.84447
sample estimates:
mean of x
98.44
```

## Question 2

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

First, we load the dataset, assign labels to variables and its categories for ease of use and explore the data.

```
1 expenditure <- read.table("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2024/main/datasets/expenditure.txt", header=T)
2 attributes(expenditure$Y)$label <- "Per capita expenditure on shelters/
   housing assistance in state"
3 attributes(expenditure$X1)$label <- "Per capita personal income in state"
4 attributes(expenditure$X2)$label <- "Number of residents per 100,000 that
   are financially insecure in state"
5 attributes(expenditure$X3)$label <- "Number of people per thousand residing
   in urban areas in state"
6 expenditure$Region <- factor(expenditure$Region,
7                               levels = c(1, 2, 3, 4),
8                               labels = c("Northeast", "North Central", "Southeast", "West"))
9 str(expenditure)
10 summary_exp <- summary(expenditure)
11 summary_exp
```

Y	X1	X2	X3
Min. : 42.00	Min. :1053	Min. :111.0	Min. :326.0
1st Qu.: 67.25	1st Qu.:1698	1st Qu.:187.2	1st Qu.:426.2
Median : 79.00	Median :1897	Median :241.5	Median :568.0
Mean : 79.54	Mean :1912	Mean :281.8	Mean :561.7
3rd Qu.: 90.00	3rd Qu.:2096	3rd Qu.:391.8	3rd Qu.:661.2
Max. :129.00	Max. :2817	Max. :531.0	Max. :899.0

STATE	REGION
Length:50	Northeast : 9
Class :character	North Central:12
Mode :character	Southeast :16
	West :13

Next, we explore the relationships between integer variables by plotting them against each other using the following code:

```
1 # Plotting Y and X1
2 png( file="/Users/ellakaragulyan/Documents/StatsI_Fall2024/problemSets/PS01_EK/
  scatterplot_y_x1.png")
3 plot(expenditure$X1, expenditure$Y,
4       ylab="Per capita expenditure on shelters/housing assistance",
5       xlab="Per capita personal income",
6       main="Relationship between expenditure on shelters/housing\
  nassistance and personal income in state")
7 dev.off()
8
9 # Plotting Y and X2
10 png( file="/Users/ellakaragulyan/Documents/StatsI_Fall2024/problemSets/PS01_EK/
  scatterplot_y_x2.png")
11 plot(expenditure$X2, expenditure$Y,
12       ylab="Per capita expenditure on shelters/housing assistance",
13       xlab="Number of financially insecure residents per 100,000",
14       main="Relationship between per capita expenditure on shelters/housing\
  nassistance and number of financially insecure residents \nper 100,000 in
  state")
15 dev.off()
16
17 # Plotting Y and X3
18 png( file="/Users/ellakaragulyan/Documents/StatsI_Fall2024/problemSets/PS01_EK/
  scatterplot_y_x3.png")
19 plot(expenditure$X3, expenditure$Y,
20       ylab="Per capita expenditure on shelters/housing assistance",
21       xlab="Number of people per thousand residing in urban areas",
22       main="Relationship between per capita expenditure on shelters/housing\
  nassistance and number of people per thousand residing \nin urban areas in
  state")
23 dev.off()
24
25 # Plotting X1 and X2
26 png( file="/Users/ellakaragulyan/Documents/StatsI_Fall2024/problemSets/PS01_EK/
  scatterplot_x1_x2.png")
27 plot(expenditure$X1, expenditure$X2,
28       ylab="Per capita personal income",
29       xlab="Number of financially insecure residents per 100,000",
30       main="Relationship between per capita personal income and \nnumber of
  financially insecure residents per 100,000 in state")
31 dev.off()
32
33 # Plotting X1 and X3
34 png( file="/Users/ellakaragulyan/Documents/StatsI_Fall2024/problemSets/PS01_EK/
  scatterplot_x1_x3.png")
35 plot(expenditure$X1, expenditure$X3,
36       ylab="Per capita personal income",
```

```

37     xlab="Number of people per thousand residing in urban areas" ,
38     main="Relationship between per capita personal income and \number of
    people per thousand residing in urban areas in state")
39 dev.off()
40
41 # Plotting X2 and X3
42 png( file="/Users/ellakaragulyan/Documents/StatsI_Fall2024/problemSets/PS01_EK/
    scatterplot_x2_x3.png")
43 plot(expenditure$X2, expenditure$X3,
44       ylab="Number of financially insecure residents per 100,000" ,
45       xlab="Number of people per thousand residing in urban areas" ,
46       main="Relationship between number of financially insecure \nresidents per
    100,000 and number of people \nper thousand residing in urban areas in
    state")
47 dev.off()

```

Figure 1 below shows there is a positive association between the variables with an increase in per capita income being associated with an increase in per capita state expenditure on shelters/housing in state.

Figure 1: Expenditure and personal income

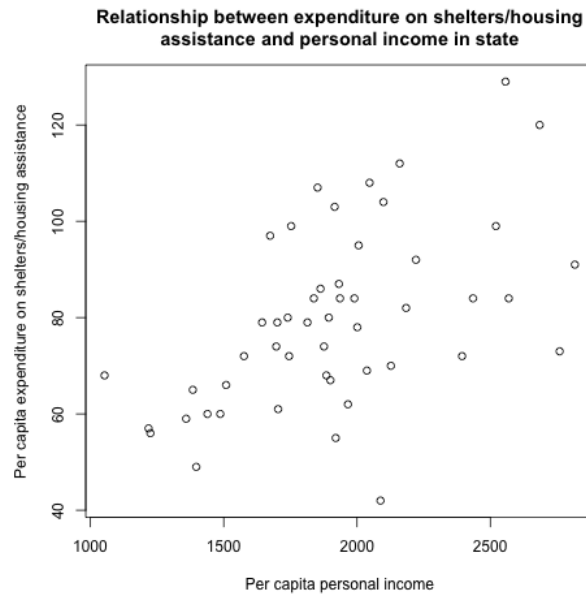


Figure 2 demonstrates a U-shaped distribution of the relationship between per capita state expenditure on shelters/housing and number of financially insecure residents per 100.000 population. The visual analysis of the plot shows that below a certain margin (around 300 financially insecure residents per 100.000) the state expenditure decreases with an increase in the number of finacially insecure residents, however, above the margin, with higher numbers of financially insecure residents in state, this association turns positive.

Figure 2: Expenditure and number of financially insecure residents

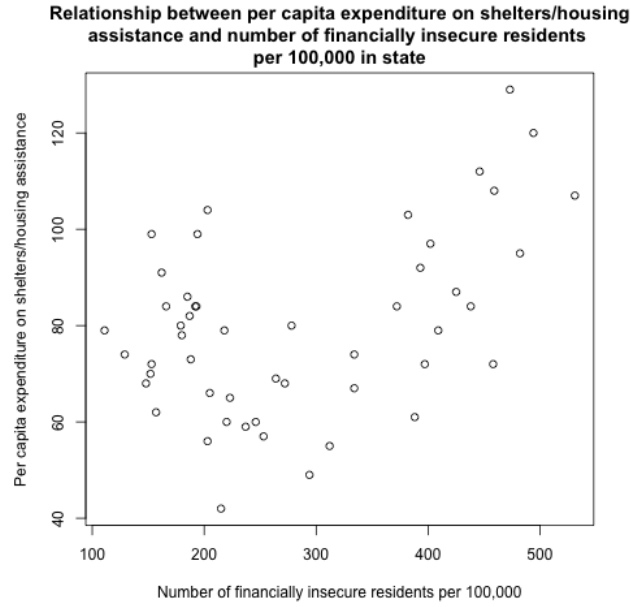


Figure 3 below shows a weak positive association between state expenditure on shelter-s/housing and the share of urban population: state expenditure increases with increasing number of urban residents.

Figure 3: Expenditure and number of people residing in urban areas

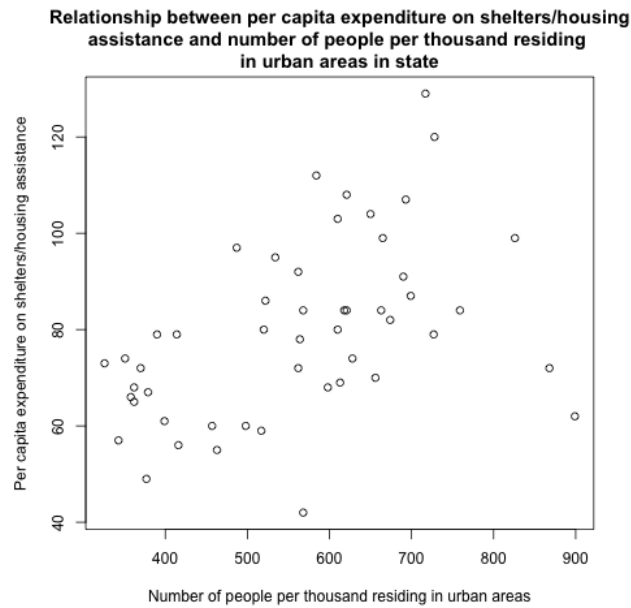


Figure 4 shows the relationship between personal income and the number of financially insecure people in state. The visual investigation shows no association between these two variables.

Figure 4: Personal income and number of financially insecure residents

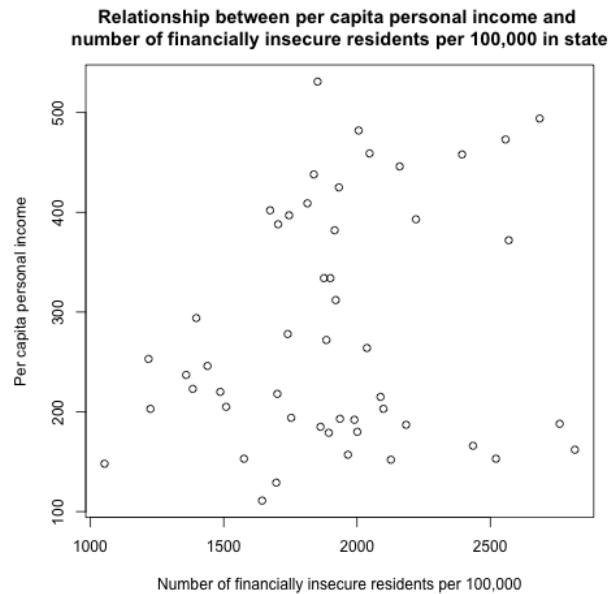


Figure 5 below shows a relatively strong positive association between per capita personal income and the number of people per thousand residing in urban areas, meaning with an increase in the number of urban residents in state there is an increase in the personal income. The data contains one extreme outlier which should be explored.

Figure 5: Personal income and number of people residing in urban areas

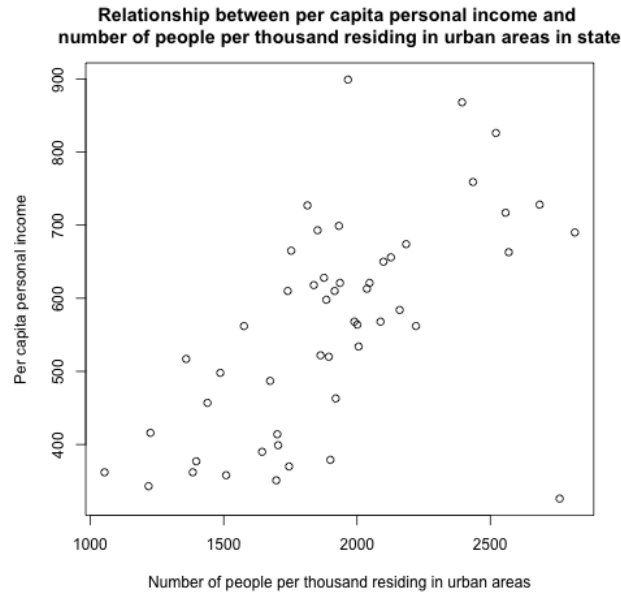
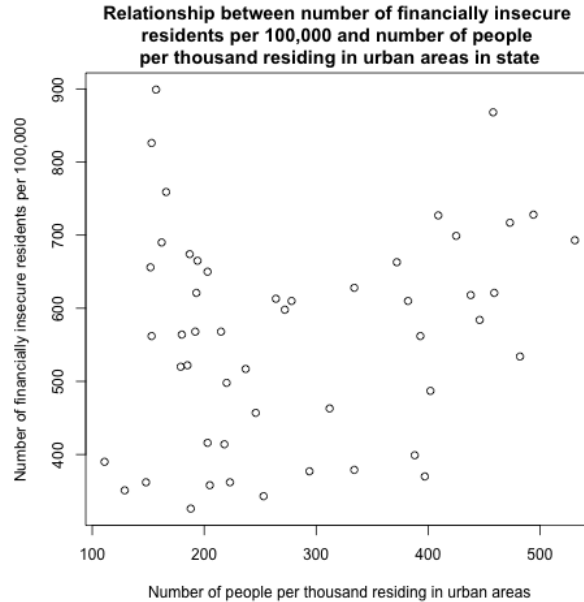


Figure 6 below shows a weak U-shaped distribution of the relationship between number of financially insecure residents and urban population in state or no association at all. The visual analysis of the plot shows that below a certain margin (around 300 urban residents) there is no association between the variables, however, above the margin there is a possible weak positive relationship.



Figure 6: Number of financially insecure residents and number of people residing in urban areas



Next, we investigate the difference in state expenditure on shelters/housing assistance across regions.

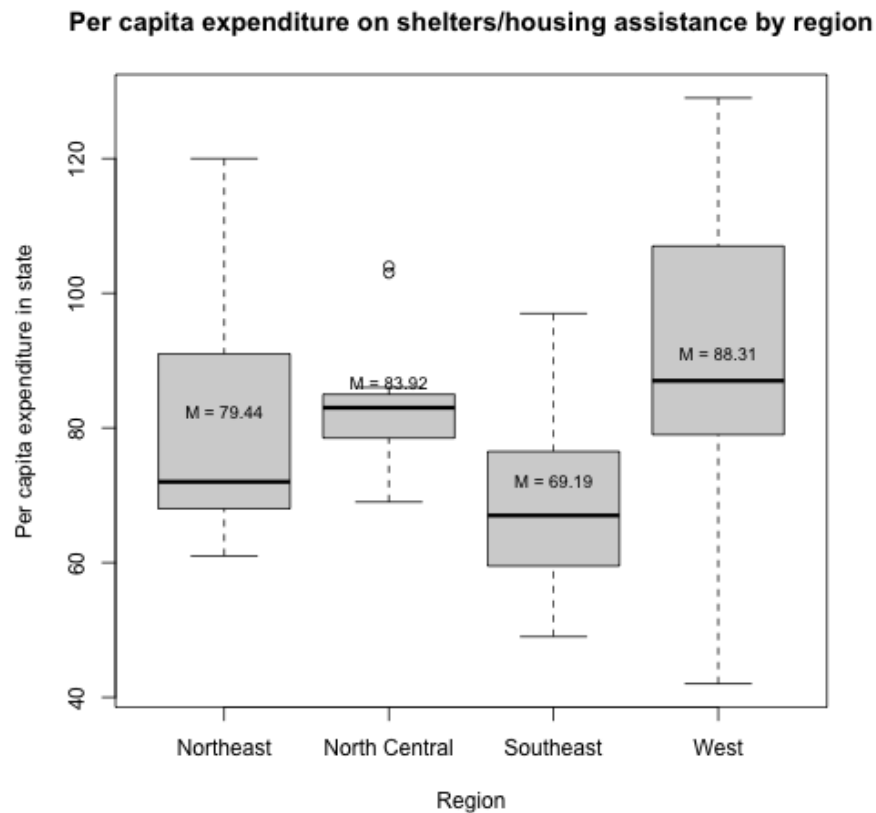
```

1 # Plotting Y and Region
2 png( file="/Users/ellakaragulyan/Documents/StatsI_Fall2024/problemSets/PS01_EK/
  boxplot_y_r.png")
3 boxplot(expenditure$Y ~ expenditure$Region,
4   main = "Per capita expenditure on shelters/housing assistance by region",
5   ylab = "Per capita expenditure in state",
6   xlab = "Region",
7   names = c("Northeast", "North Central", "Southeast", "West")
8 )
9 for (i in 1:length(means)) {
10   text(i, means[i], labels = paste0("M = ", round(abs(means[i]), 2)), pos = 3,
11     col="black", cex=0.8)
12 }
13 dev.off()

```

Figure 7 below shows the distribution of per capita expenditure shelters/housing assistance in state across the four regions. The states located in the "West" region have the highest per capita spending on shelters/housing assistance (88.31), while those located in the "Southeast" have the lowest (69.19).

Figure 7: Expenditure on shelters/housing assistance by region



Lastly, we explore the relationship between expenditure on shelters/housing assistance and personal income across the four regions.

```

1 # Plotting Y and X1 by region
2 png(file = "/Users/ellakaragulyan/Documents/StatsI_Fall2024/problemSets/PS01_
  EK/scatterplot_y_x1_r.png")
3 plot(expenditure$X1, expenditure$Y,
4       col=c("aquamarine3", "coral", "cornflowerblue", "darkolivegreen"),
5       pch = 19, # Solid circle for points
6       ylab = "Per capita expenditure on shelters/housing assistance",
7       xlab = "Per capita personal income",
8       main = "Relationship between expenditure on shelters/housing assistance
  and personal income in state")
9       legend("topright", legend = levels(as.factor(expenditure$Region)),
10            col=c("aquamarine3", "coral", "cornflowerblue", "darkolivegreen"),
11            pch = 19, title = "Region")
12 dev.off()

```

Figure 8: Relationship between expenditure on shelters/housing assistance and personal income in state

