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:

```
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.

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 $\alpha = 0.05$.

Solution - Question 1: Education

1. Confidence Interval

- Calculate the mean

```
1 mean_y = mean(y)
2 mean_y
```

    [1] 98.44

- Calculate the sum of the squared errors to get the variance, and then the standard deviation

```
1 demeanedSum <- y - mean_y
2 squaredError <- demeanedSum ^ 2
3 variance <- sum(squaredError)/(length(y)-1)
4 stdev = sqrt(variance)
5 stdev
```

    [1] 13.09287

- Calculate the standard error

```
1 sterror = stdev/sqrt(length(y))
2 sterror
```

    [1] 2.618575

- Find the t-score for the desired confidence level, with 24 degrees of freedom

```
1 t = qt(p = (1 - 0.9)/2 , df = 24, lower.tail = FALSE)
2 t
```

    [1] 1.710882

- Construct the confidence interval by calculating its lower and upper limits

```
1 lowerlimit = mean_y - t * sterror
2 upperlimit = mean_y + t * sterror
3 lowerlimit
4 upperlimit
```

    > lowerlimit
    [1] 93.95993
    > upperlimit
    [1] 102.9201

2. Hypothesis Testing

- **Step 1:** assumptions

  - the data is quantitative
  - the sampling method is random
  - the sample size is smaller than 30, so we cannot assume it follows a normal distribution - therefore, I used the t-score and not the Z-score as the test statistic

- **Step 2:** hypotheses

  - **Null hypothesis:** the mean is lower or equal to 100
  - **Alternative hypothesis:** the mean is higher than 100

- **Step 3:** calculate the test statistic

```
1 ts = (mean_y − 100)/sterror
2 ts
```

```
[1] -0.5957439
```

- **Step 4:** calculate the p-value

```
1 p = pt(ts, df = 24, lower.tail = FALSE)
2 p
```

```
[1] 0.7215383
```

- **Step 5:** draw a conclusion
  The p-value is higher than alpha, so we cannot reject the null hypothesis that the mean is lower or equal to 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`.

```
1  t = qt(p = (1 − 0.9)/2 , df = 24, lower.tail = FALSE)
```

- 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)?

- 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.

Solution - Question 2: Political Economy
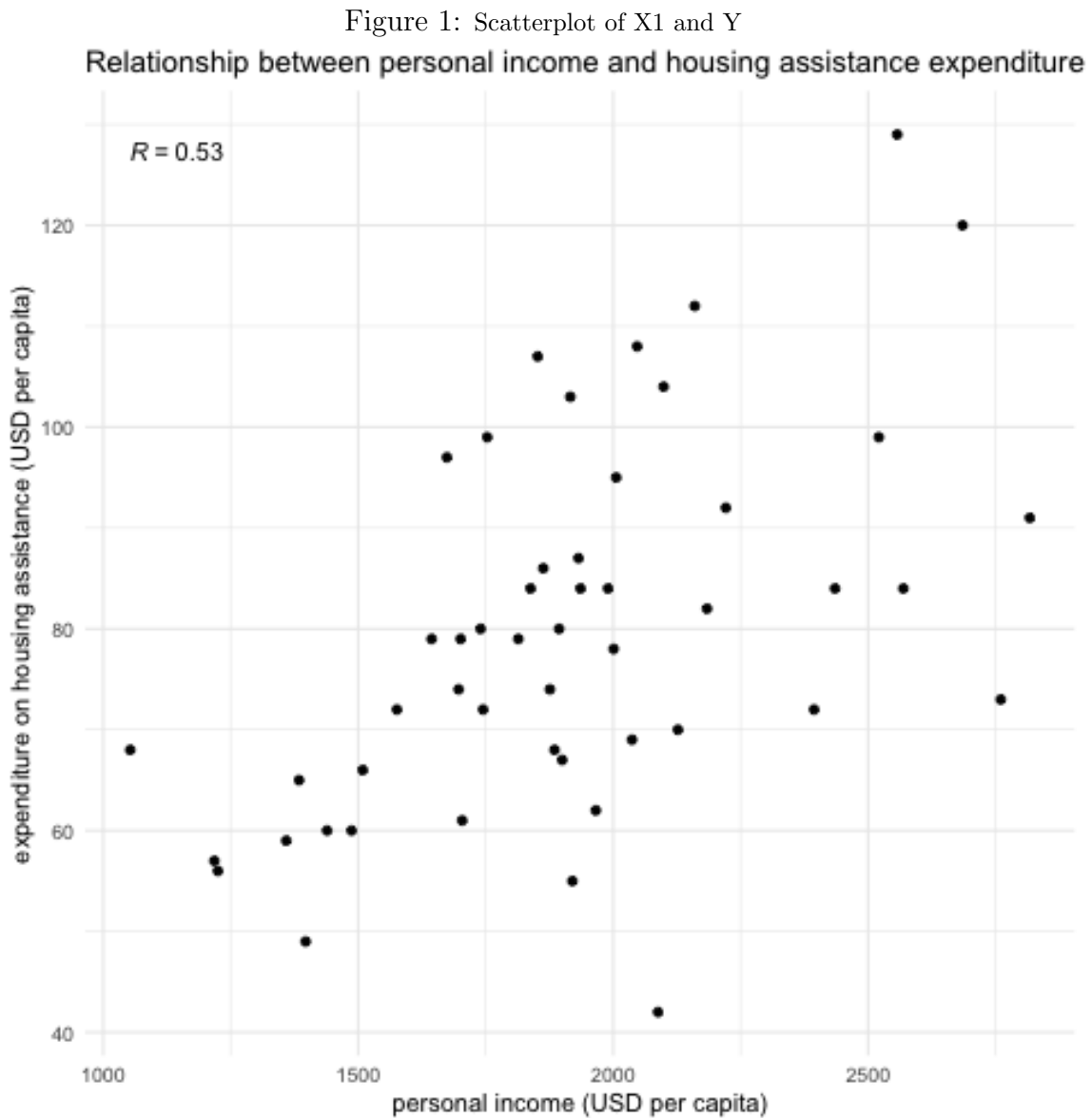
Figure 1: Scatterplot of X1 and Y

Relationship between personal income and housing assistance expenditure
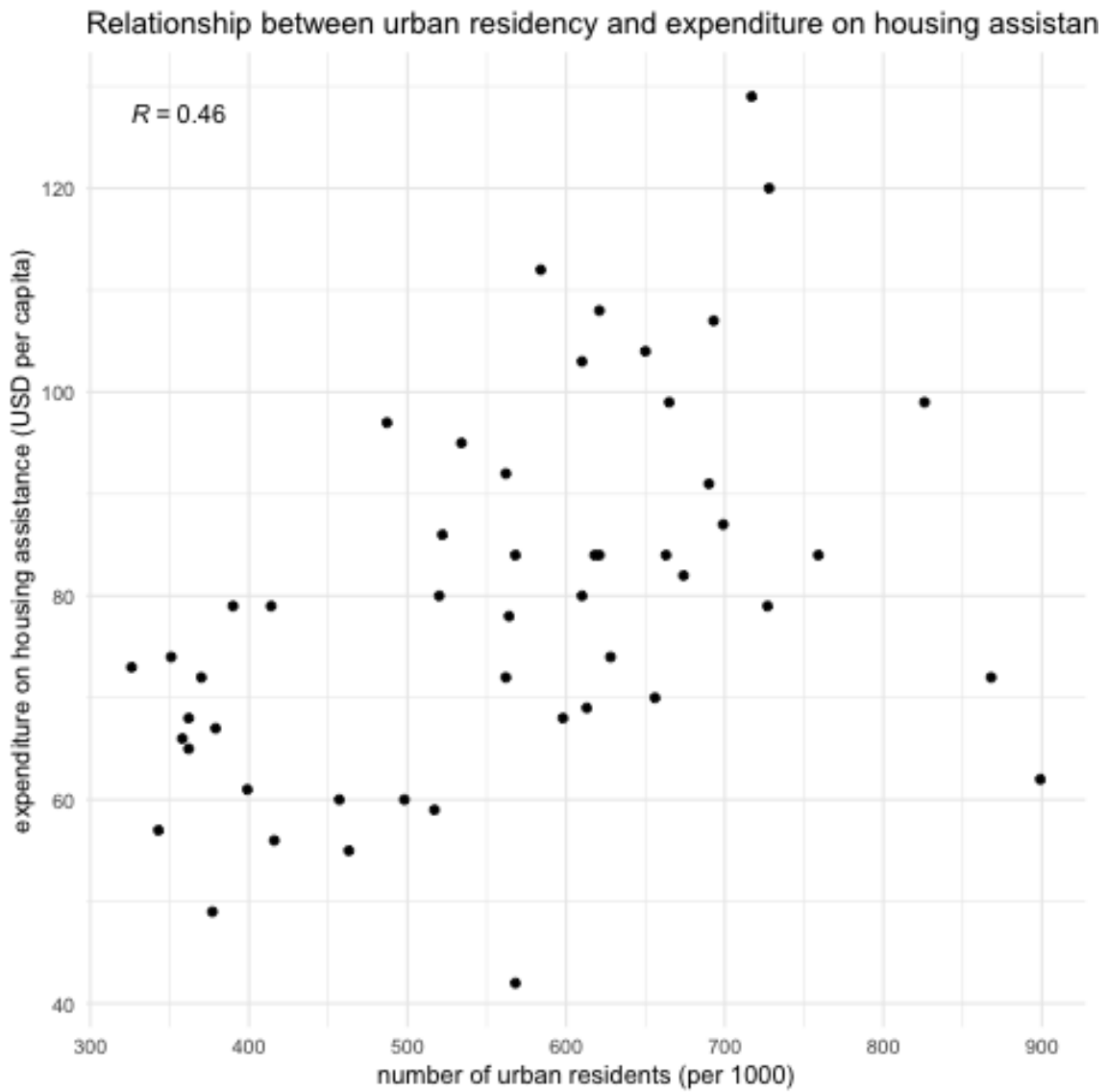
Interpretation:
There is a moderate positive association between the personal income and the expenditure on housing assistance in a state, both measured in USD per capita. The Pearson correlation coefficient points towards the same conclusion, as its value is above 0, but not close enough to 1 to suggest a strong association between the two variables.
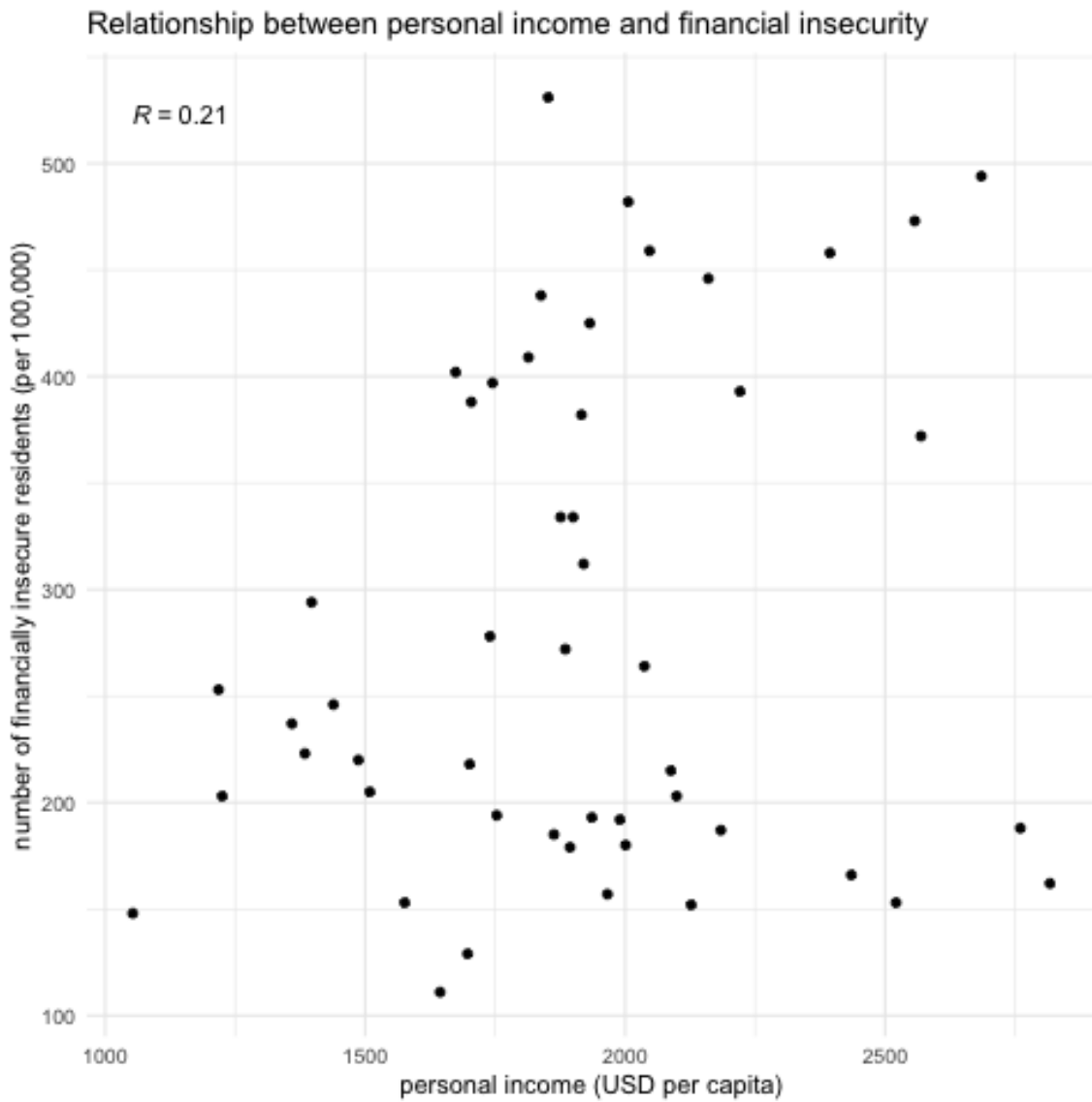
Figure 2: Scatterplot of X2 and Y
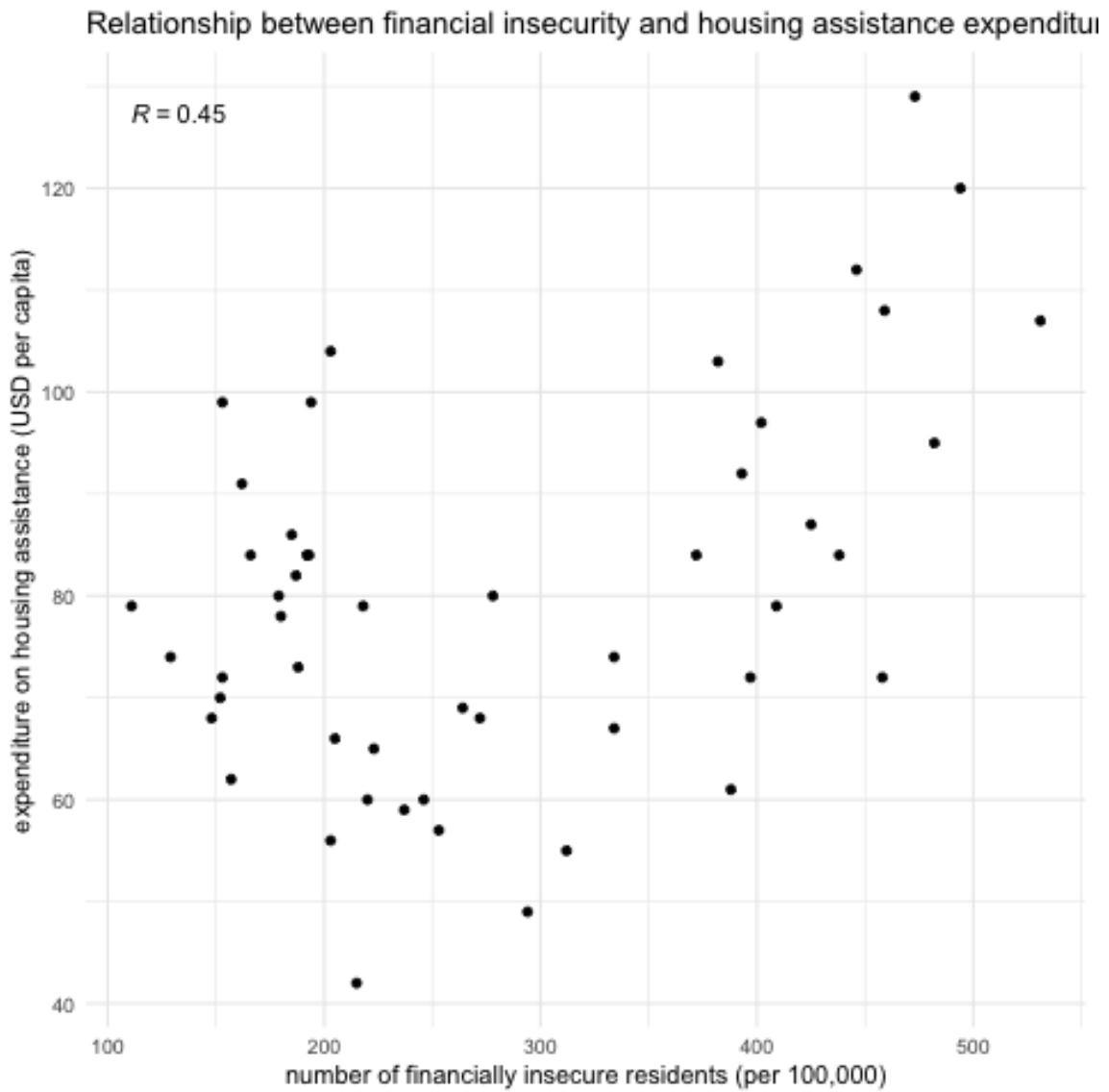
Relationship between financial insecurity and housing assistance expenditure

Interpretation: There is a moderate positive association between the personal income and the expenditure on housing assistance in a state, both measured in USD per capita. The Pearson correlation coefficient points towards the same conclusion, as its value is above 0, but not close enough to 1 to suggest a strong association between the two variables.
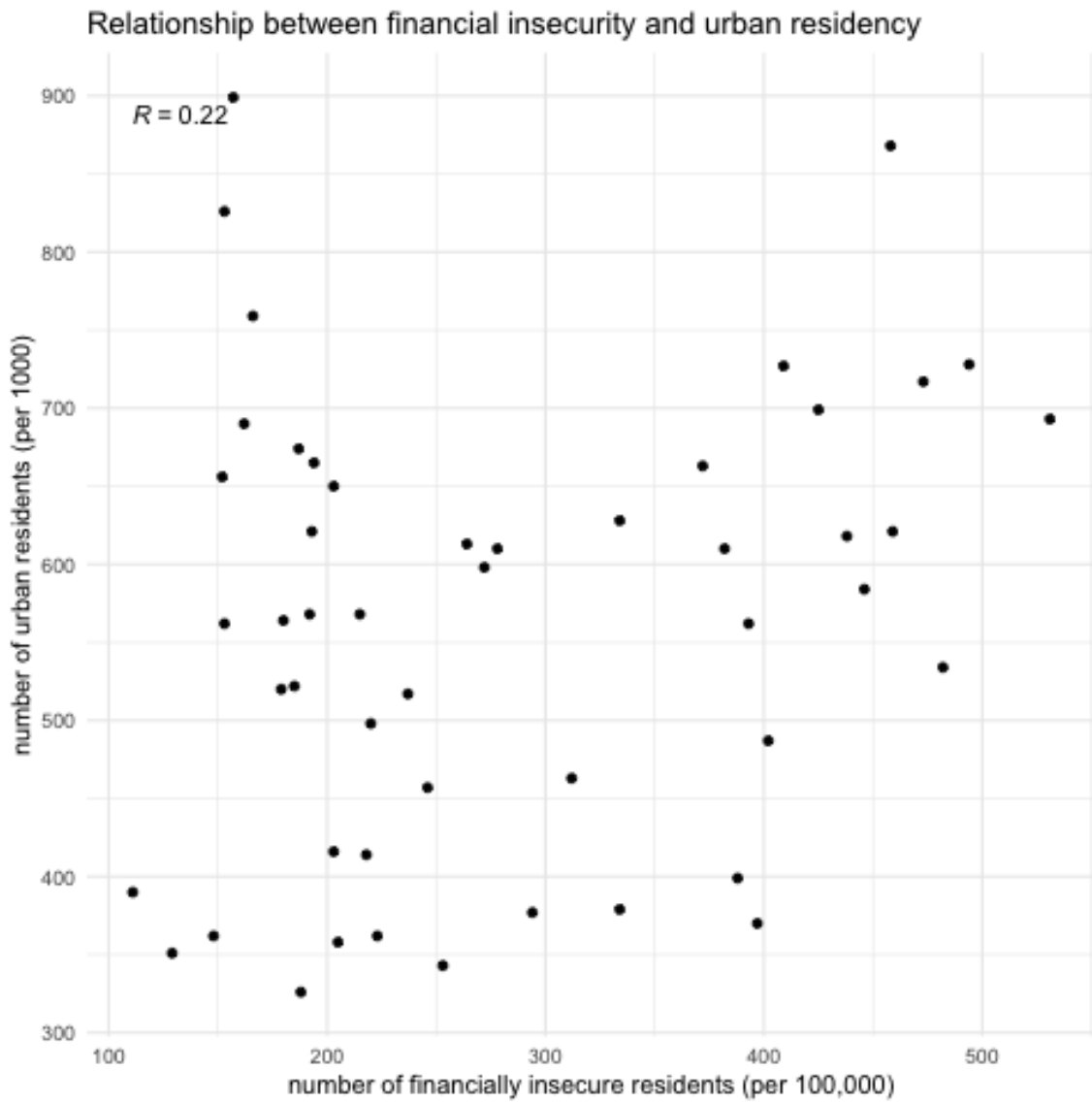
Figure 3: Scatterplot of X3 and Y

Relationship between urban residency and expenditure on housing assistan

Interpretation: There is a moderate positive association between the personal income and the expenditure on housing assistance in a state, both measured in USD per capita. The Pearson correlation coefficient points towards the same conclusion, as its value is above 0, but not close enough to 1 to suggest a strong association between the two variables.

Figure 4: Scatterplot of X1 and X2

Relationship between personal income and financial insecurity
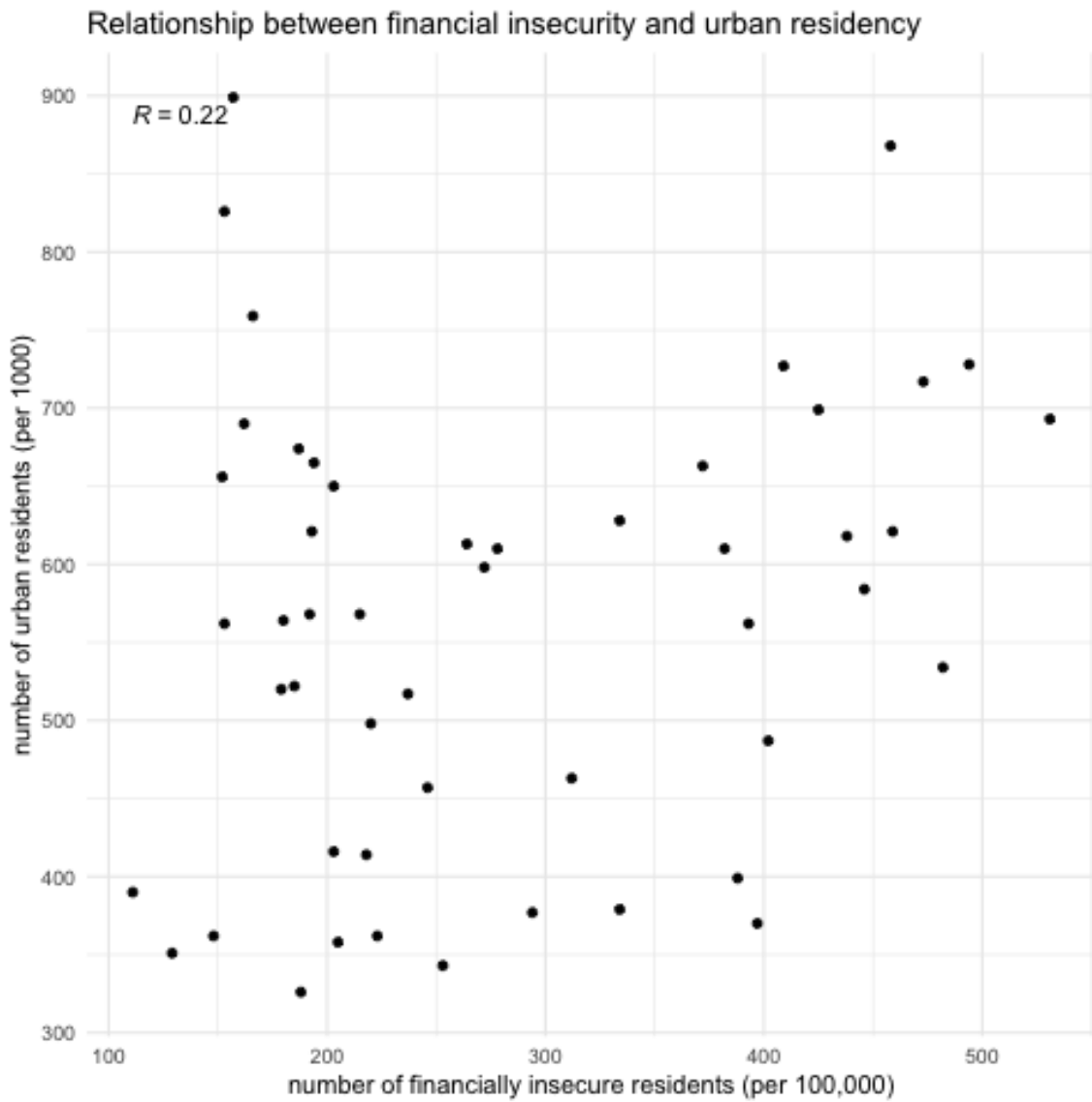
$R = 0.21$

Interpretation: There is a moderate positive association between the personal income and the expenditure on housing assistance in a state, both measured in USD per capita. The Pearson correlation coefficient points towards the same conclusion, as its value is above 0, but not close enough to 1 to suggest a strong association between the two variables.

Figure 5: Scatterplot of X2 and X3

Relationship between financial insecurity and urban residency

Interpretation: There is a moderate positive association between the personal income and the expenditure on housing assistance in a state, both measured in USD per capita. The Pearson correlation coefficient points towards the same conclusion, as its value is above 0, but not close enough to 1 to suggest a strong association between the two variables.
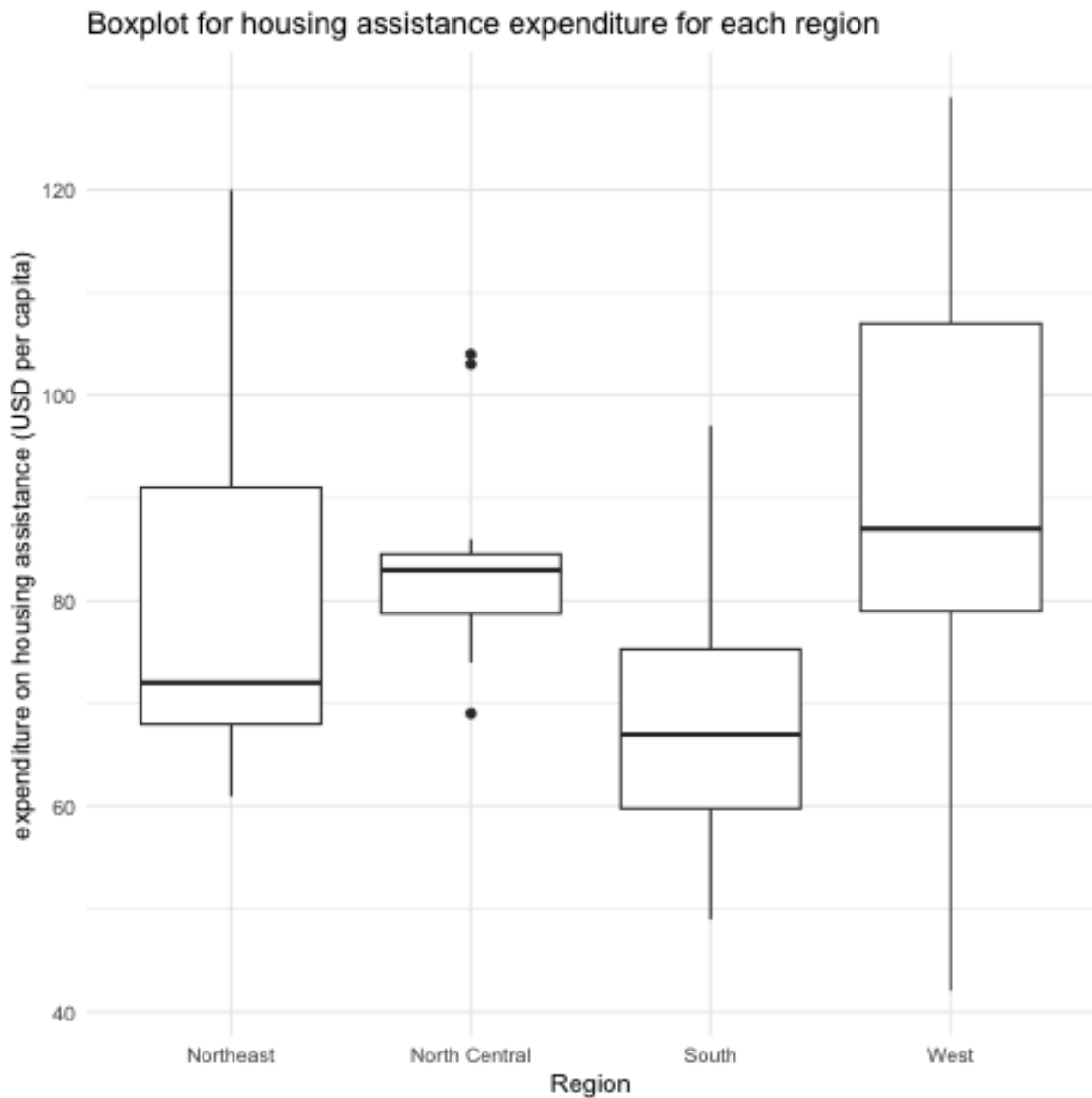
Figure 6: Scatterplot of X2 and X3

Relationship between financial insecurity and urban residency

Interpretation: There is a moderate positive association between the personal income and the expenditure on housing assistance in a state, both measured in USD per capita. The Pearson correlation coefficient points towards the same conclusion, as its value is above 0, but not close enough to 1 to suggest a strong association between the two variables.

Figure 7: Scatterplot of X1 and X2

Boxplot for housing assistance expenditure for each region

Interpretation: There is a moderate positive association between the personal income and the expenditure on housing assistance in a state, both measured in USD per capita. The Pearson correlation coefficient points towards the same conclusion, as its value is above 0, but not close enough to 1 to suggest a strong association between the two variables.