Problem set 1

Luna Goldstein

1 October 2021

Question 1: Education

Part 1

The 90% confidence interval for the average IQ of the counselor's students lies between 93.96 and 102.92. Those values were calculated using a one-sample t-test:

```
t.test(y, conf.level = 0.9)
```

The output of the t-test:

t-score = 37.593,
degrees of freedom = 24,
p-value < 2.2e-16
The alternative hypothesis proposes that the true mean does not equal 0.
Mean of x = 98.44

Part 2

We cannot reject the null hypothesis that the average student IQ us lower or equal to the average IQ of schools across the country.

```
t.test(y, mu = 100, alternative = 'greater')
```

To tackle this question a t-test was conducted, utilizing the sample y and setting the population average to 100. As the teacher is investigating whether her students' average is higher (one-sided test), we insert 'alternative = 'greater'.

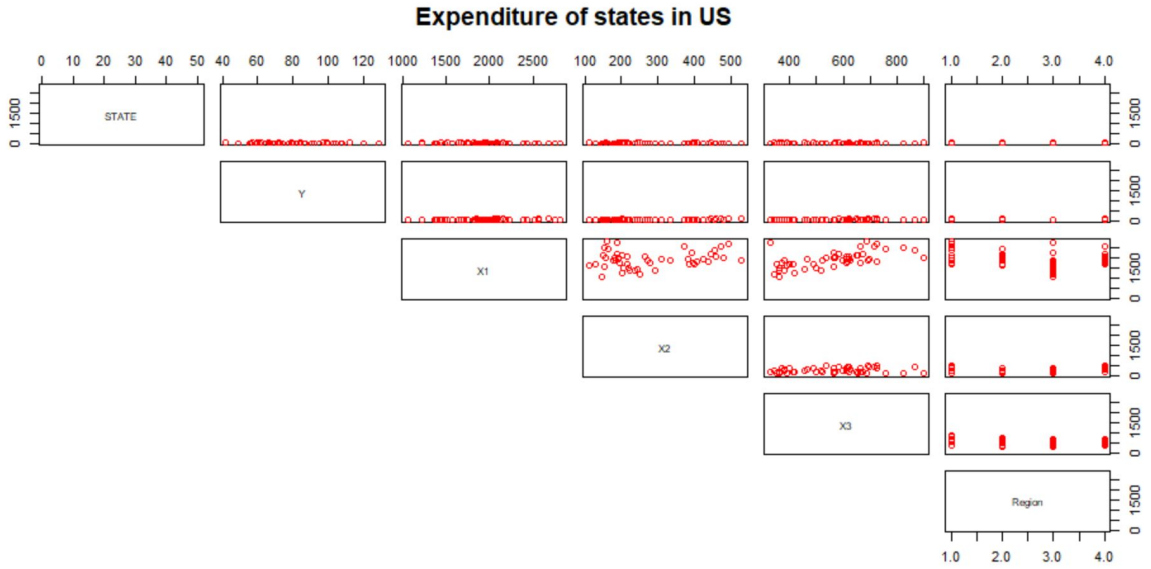
The t-test's output returns a p-value of 0.7215. Those findings indicate that we cannot reject the null hypothesis (it appears although the students' average IQ does not exceed the average IQ of the country's students).

Question 2: Political Economy

Part 1

State	50 states in US
Y	per capita expenditure on shelters/housing assistance in state
X1	per capita personal income in state
X2	X2 Number of residents per 100,000 that are "financially insecure" in state
X3	X3 Number of people per thousand residing in urban areas in state
Region	Region 1=Northeast, 2= North Central, 3= South, 4=West

The relationship between Y, X1, X2, and X3 and their correlation is described in a graph below.



Conclusions drawn from the interpretation of the matrix scatter plot:

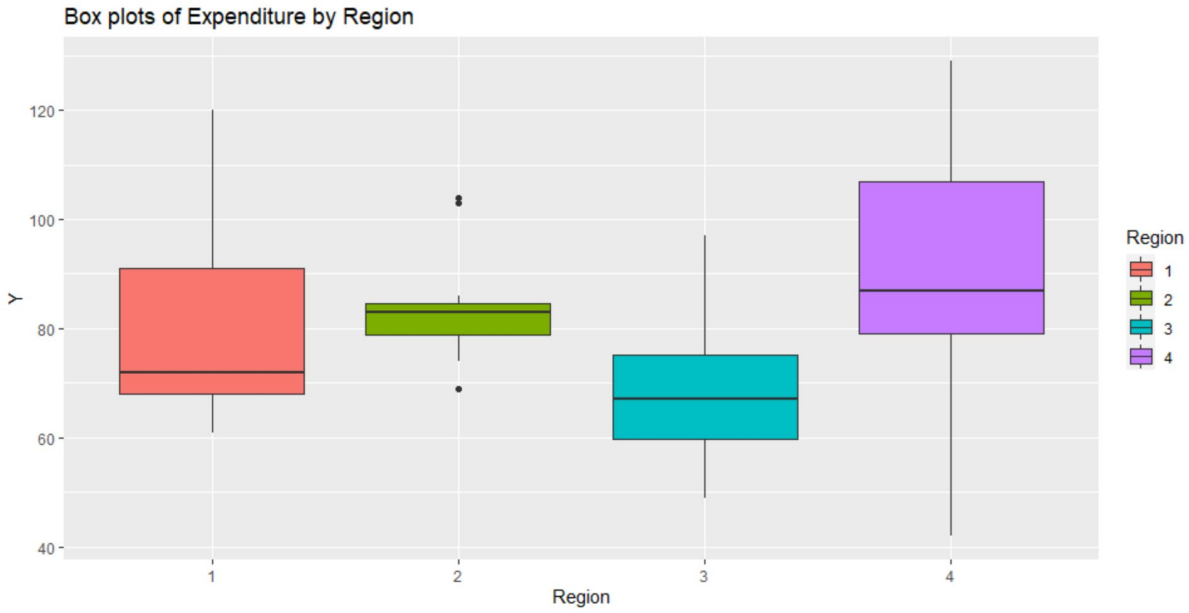
- There is little to no relationship between the state and any of the other variables.
- There is little to no relationship between housing assistance per capita and any of the other variables.
- There is a strong positive relationship between X1 (personal income) and X2 (financial security).

The code:

plot(expenditure, ylim=range(expenditure$Y$, $expenditure$X1,expenditure$X2$, $expenditure$X3), col='red', main = "Expenditure of states in US", lower.panel = NULL)

Part 2

The relationship between Y and region was plotted below in order to investigate which region on average has the highest expenditure on housing assistance:



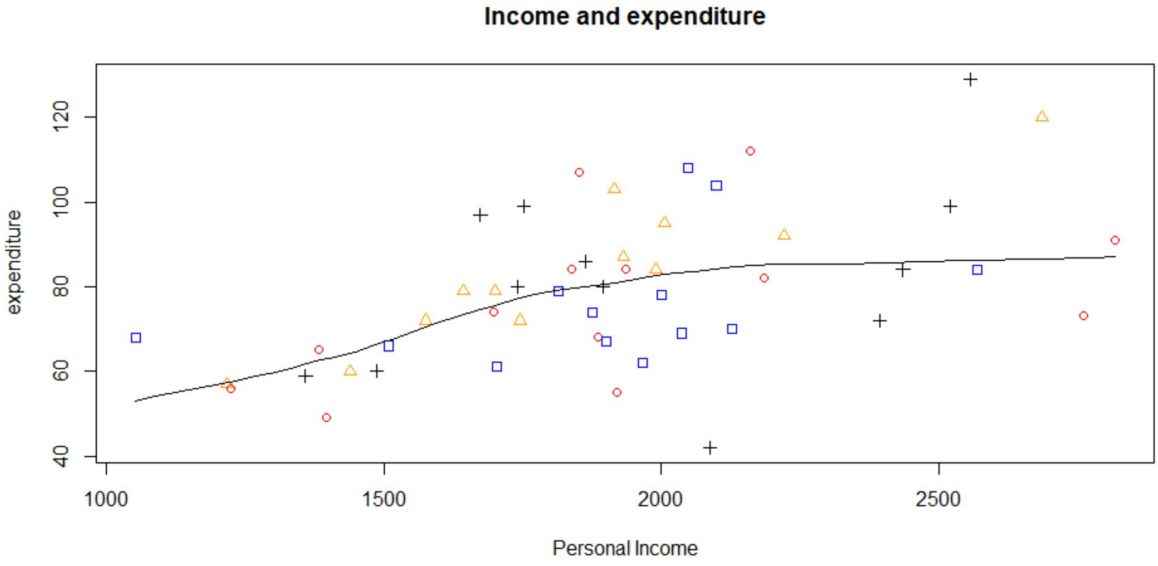
On average, the results suggest that the states in the West of the US have the highest per capita expenditure on housing assistance.

The code:

- data=as.data.frame(expenditure[,c(2,6)])
- data$Region = as.factor(data$Region)
- mode(dataRegion)
- ggplot(aes(y = Y, x = Region, fill=Region), data = data)+ geombox-plot()+ggtitle("Box plots of Expenditure by Region")

Part 3

The relationship between personal income in different states with the per capita expenditure on shelters/housing assistance illustrated below:



We interpret the following from the scatter plot: It appears as if the more the personal income increases the more per capita expenditure on shelters and housing increases. However, this relationship stagnates once the personal income reaches a value of 2000.