

# Problem Set 1

Valeriia Babaian

Due: October 1, 2023

## Question 1 (40 points): Education

1. Find a 90% confidence interval for the average student IQ in the school.

*Answer:* [93.95993; 102.92007].

```
1 t_score <- qt(0.95, df=length(y)-1)
2 lower_90_t <- mean(y)-(t_score)*(sd(y)/sqrt(length(y)))
3 upper_90_t <- mean(y)+(t_score)*(sd(y)/sqrt(length(y)))
4
5 # Confidence interval boundaries: 93.95993; 102.9201
6 lower_90_t
7 mean(y) #98.44
8 upper_90_t
9
10 #double checking:
11 t.test(y, conf.level = 0.9, alternative = "two.sided")
12 # 90 percent confidence interval: 93.95993; 102.92007
```

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.

*Answer:* It is not higher:

```
1 t.test(y, mu = 100, alternative = "greater")
2 # We cannot reject the null hypothesis that the average student IQ in the
   school is not higher than the average IQ score across all schools in
   the country (p-value = 0.7215)
```

Using the same sample, conduct the appropriate hypothesis test with  $\alpha = 0.05$ .

*Answer:* Expectedly, the answer is the same as the average student IQ score in the given school was not greater than average IQ across the country on any of conventional levels of confidence.

```
1 ##Using the same sample, conduct the appropriate hypothesis test with \  
    alpha = 0.05.  
2 t.test(y, mu = 100, conf.level = 0.95) # we cannot first reject the H0  
    that the averages are equal on 0.05 significance (p-value = 0.5569)
```

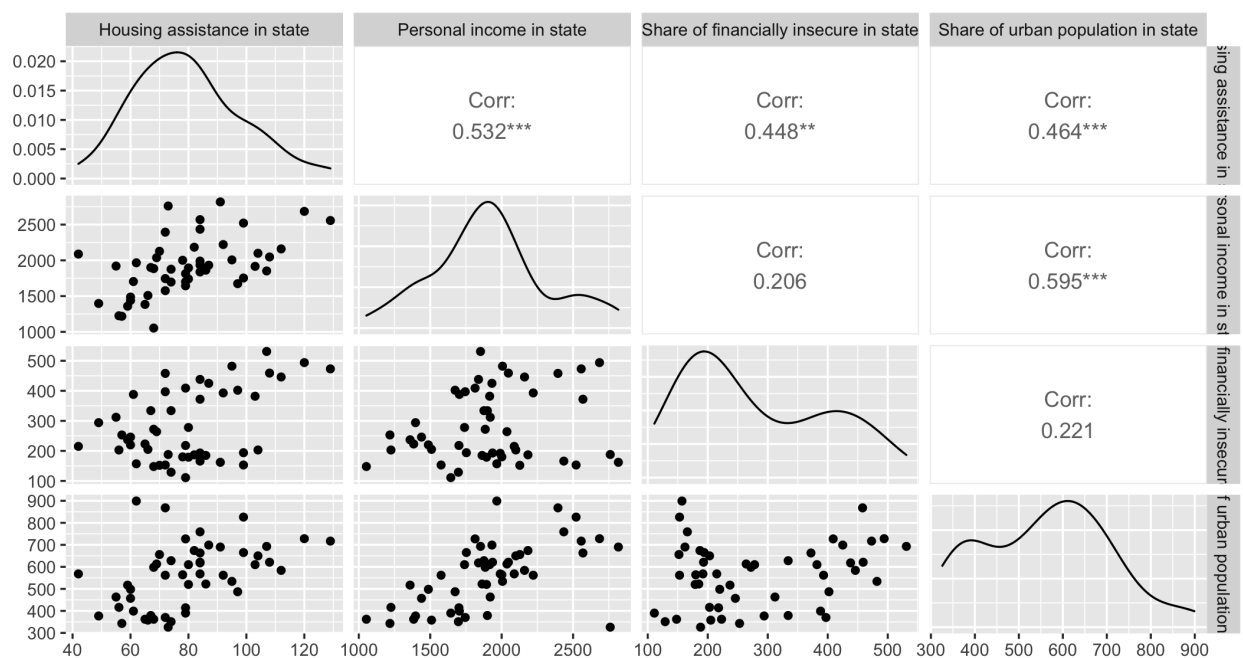
## Question 2 (40 points): Political Economy

Explore the `expenditure` data set and import data into R.

```
1 expenditure <- read.table("https://raw.githubusercontent.com/ASDS-TCD/StatsI_Fall2023/main/datasets/expenditure.txt", header=T)
2 str(expenditure)
```

- 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 ggpairs(expenditure,
2         columns = 2:5,
3         columnLabels=c("Housing assistance in state",
4                         "Personal income in state",
5                         "Share of financially insecure in state",
6                         "Share of urban population in state"))
```



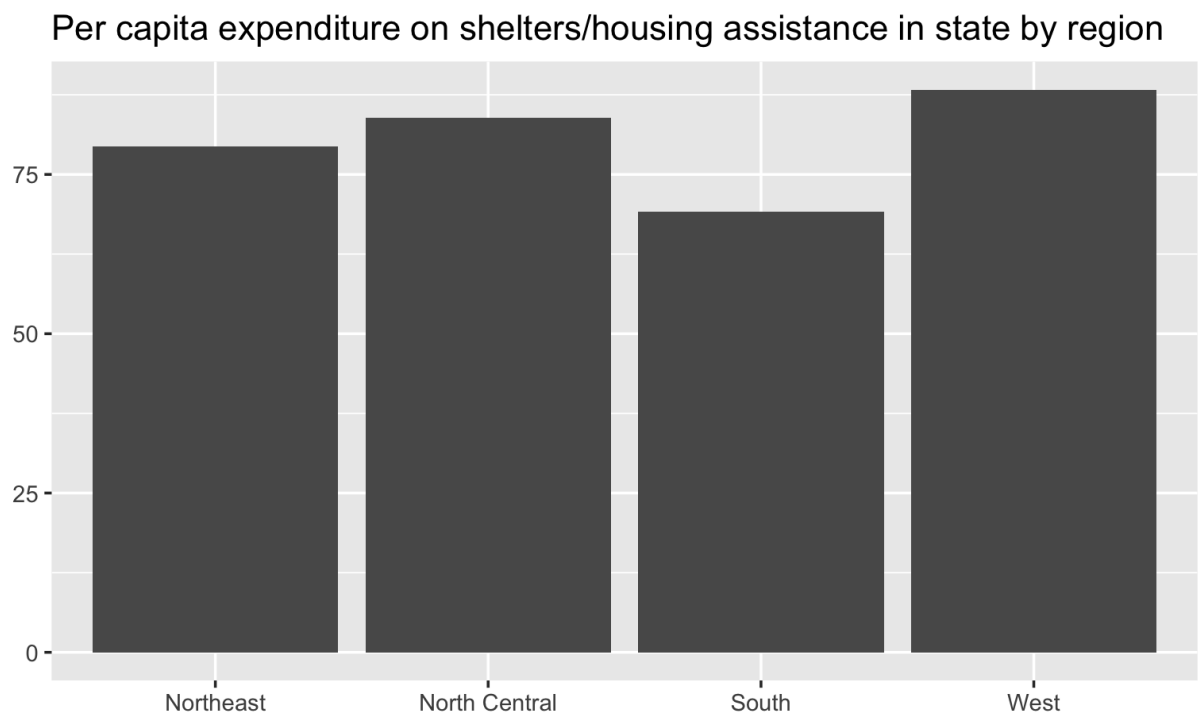
According to the visualized relationships between all meaningful quantities in the data, the variables are positively related, but there are no strong correlations between any of them. The upper-right part of the graphics proves this point.

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

```

1 #transforming Region to a factor for meaningful visualization:
2 expenditure$Region <- factor(expenditure$Region, labels = c("Northeast",
3   "North Central", "South", "West"))
4
5 library(ggplot2)
6 ggplot(data=expenditure, mapping=aes(x=Region, y=Y)) +
7   stat_summary(fun.data=mean_sdl, geom="bar") +
8   labs(title="Per capita expenditure on shelters/housing assistance in
9     state by region",
10    x="", y="")

```



The highest per capita expenditure on housing assistance is in the Western states.

- 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.

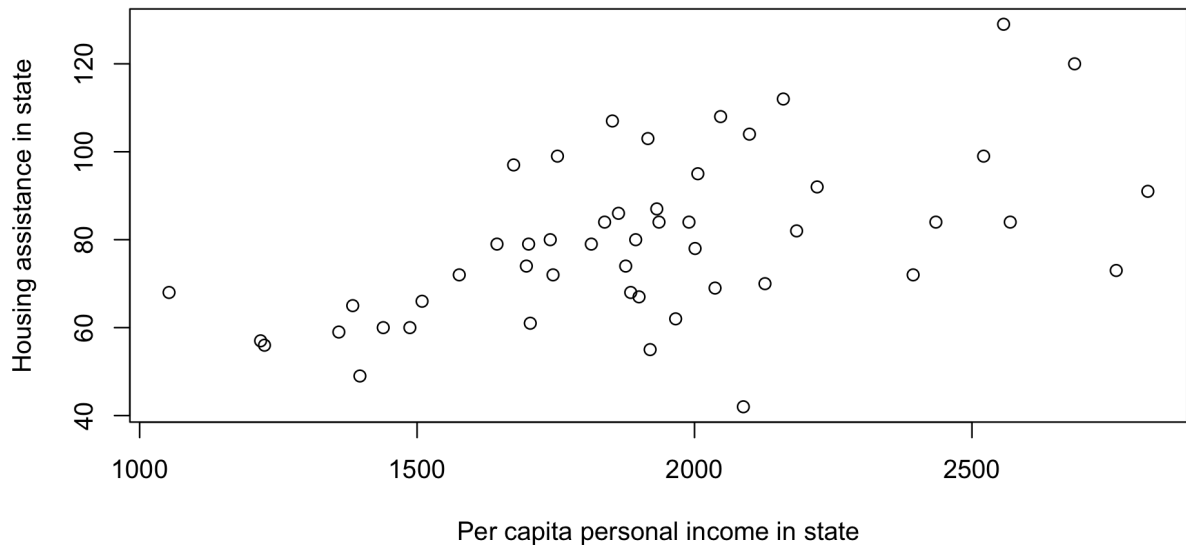
```

1 plot(expenditure$X1, expenditure$Y,
2       xlab="Per capita personal income in state",
3       ylab="Housing assistance in state",

```

```
4 main="The Relationship between state-level personal income and
housing assistance")
```

### The Relationship between state-level personal income and housing assistance



There is a positive relationship between per capita personal income in state and its housing assistance expenditures, however, there is no strong correlation as there is a significant amount on deviations on both sides of the distributions.

```
1 plot(expenditure$X1, expenditure$Y,
2       col= expenditure$Region,
3       pch=c(21,22,23,24)[as.numeric(expenditure$Region)],
4       xlab="Per capita personal income in state",
5       ylab="Housing assistance in state",
6       main="The Relationship between state-level personal income and
7             housing assistance by regions")
8 legend("topleft", cex=0.5, legend=levels(expenditure$Region), pch=unique(
9       expenditure$Region), col=unique(expenditure$Region))
```

**The Relationship between state-level personal income and housing assistance by regions**

