# PS01 Homework

**Title:** Problem Set 1

**Date:** 3rd of October 2022

**Author:** Brendan Walsh – 22334595

## PROBLEM 1: EDUCATION

A school counsellor 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 = 25 randomly selected school students

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)

### Question 1 – Find a 90% confidence interval for the average student IQ in the school

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

$\bar{y} = \frac{\sum y}{n} = \frac{2461}{25}$

mean(y) = **98.44**

$\sigma y = \sqrt{\frac{\sum_{i=0}^{n}(yi - \bar{y})^2}{n-1}}$

sd(y) = **13.09287**

$\sigma\bar{y} = \frac{\sigma y}{\sqrt{n}} = \frac{13.09287}{\sqrt{25}}$

Standard_error_y <- sd(y) / sqrt(length(y))

Standard_error_y = **2.618575**

t-distribution for y <- **1.711** # (from page 41 of the State Examinations Commission's formulae and tables booklet

Confidence interval for y = $\bar{y} \pm t(\sigma\bar{y})$ = 98.44 $\pm$ 1.711(2.618575) = 98.44 $\pm$ 4.480072833

$\mu$ = Average IQ of all the school's students

Confidence interval for y = 93.96 $\leq \mu \leq$ 102.92

***Result: We can say with 90% confidence that the school's mean IQ falls between 93.96 and 102.92***

## Question 2 – A Hypothesis Test of the Data

2. Next, the school counsellor 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: Assumptions:

The IQ data is continuous, randomly selected, and randomly selected, with n = 25

#2 Hypotheses:

H0: $\mu \leq 100$

Ha: $\mu > 100$

#3 Test statistic:

One sample t-test = t = $\frac{\bar{y}-\mu}{\left(\frac{\sigma y}{\sqrt{n}}\right)} = \frac{98.44-100}{2.618575}$

One_sample_t_test_y <- (mean(y) – 100 / (Standard_error_y)

format(round(One_sample_t_test_y, 2), nsmall = 3)

One_sample_t_test_y = -0.59

#4 P-value:

P_value_y <- 2*pt(abs(One_sample_t_test_y), 24, lower.tail = FALSE)

P_value_y <- format(round(P_value_y, 3), nsmall=3)

P_value_y = 0.557

#5 Conclusion:

***Result: Because the P-value of Y is 0.557, we can't reject the null hypothesis - i.e., we can't say the average IQ of the school's students is higher than the national average of 100. The results of the hypothesis test aren't statistically significant.***

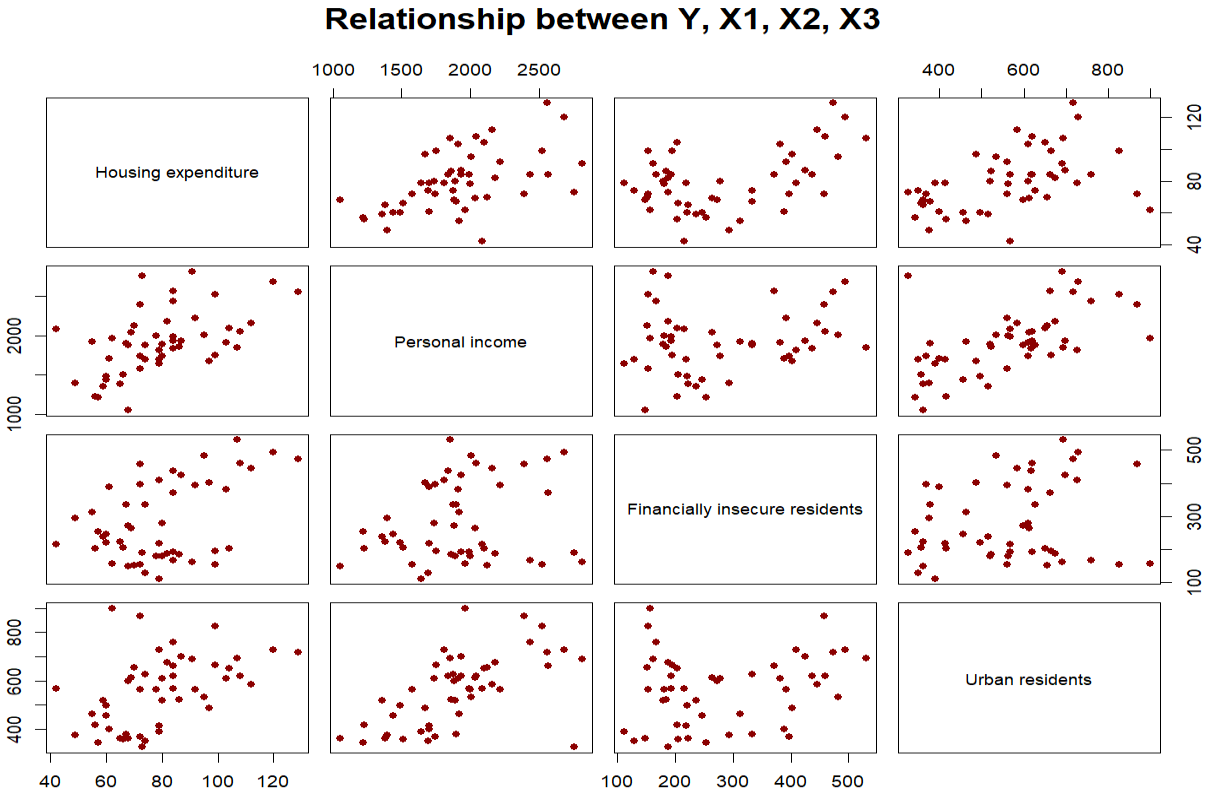
## PROBLEM 2: POLITICAL ECONOMY

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

• 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)?

```
pairs(~expenditure$Y + expenditure$X1 + expenditure$X2 + expenditure$X3,
      col = "darkred", pch = 19,
      labels = c("Housing expenditure",
                 "Personal income",
                 "Financially insecure residents",
                 "Urban residents"),
      main = "Relationship between Y, X1, X2, X3")
```



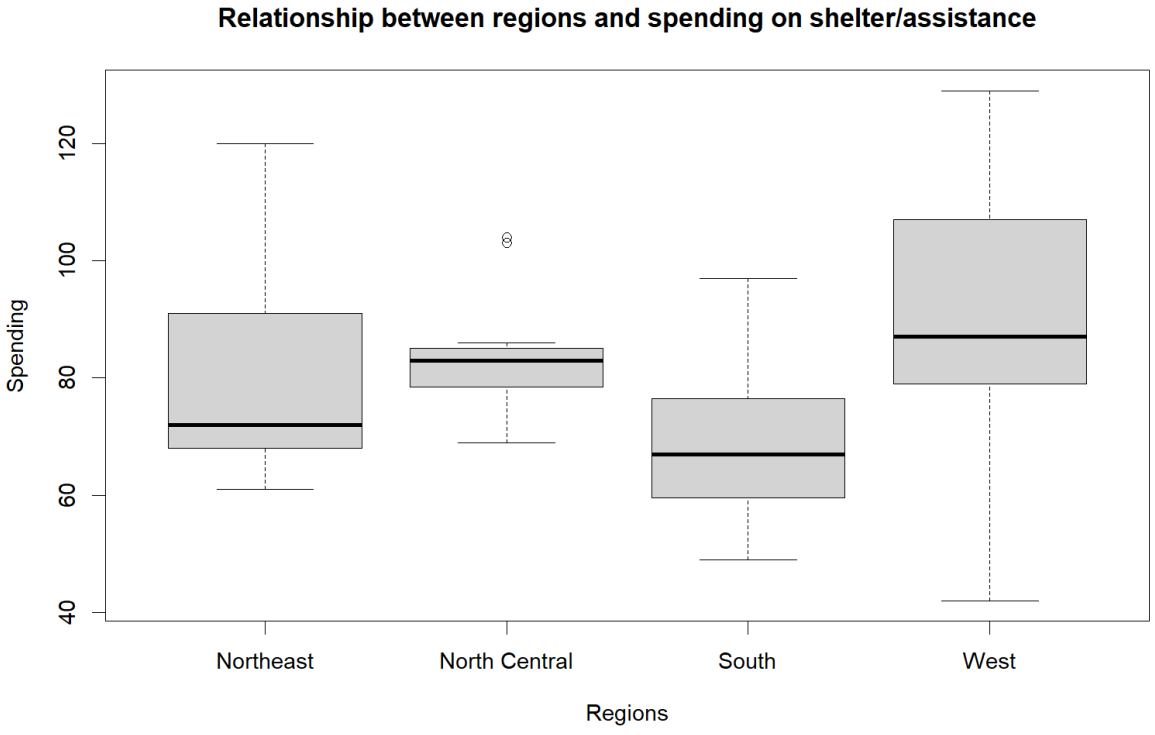
*Result: From this graph, we can interpret:*

- *Per capita housing expenditure on shelters/housing assistance in state (Y) as being a weak predictor of the other three variables (X1, X2, X3)*
- *Personal income (X1) weakly predicts Per capita expenditure on shelters/housing assistance in state (Y), and the Number of residents per 100,000 that are financially insecure in state (X2) the Number of people per thousand in urban areas in state (X3)*
- *the Number of residents per 100,000 that are financially insecure in state (X2) holds no predictive value of the other three variables (Y, X1, X3)*
- *the Number of people per thousand residing in urban areas in state (X3) weakly predicts Per capita personal income in state (X1), and even more weakly predicts Per capita expenditure on shelters/housing assistance in state (Y), and holds no predictive value for Number of residents per 100,000 that are financially insecure in state (X2)*

*Overall, we can see the Number of residents per 100,00 that are financially insecure in state (X2) is of minimal predictive or correlative relevance in this study as it neither predicts nor is predicted by any variables.*

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

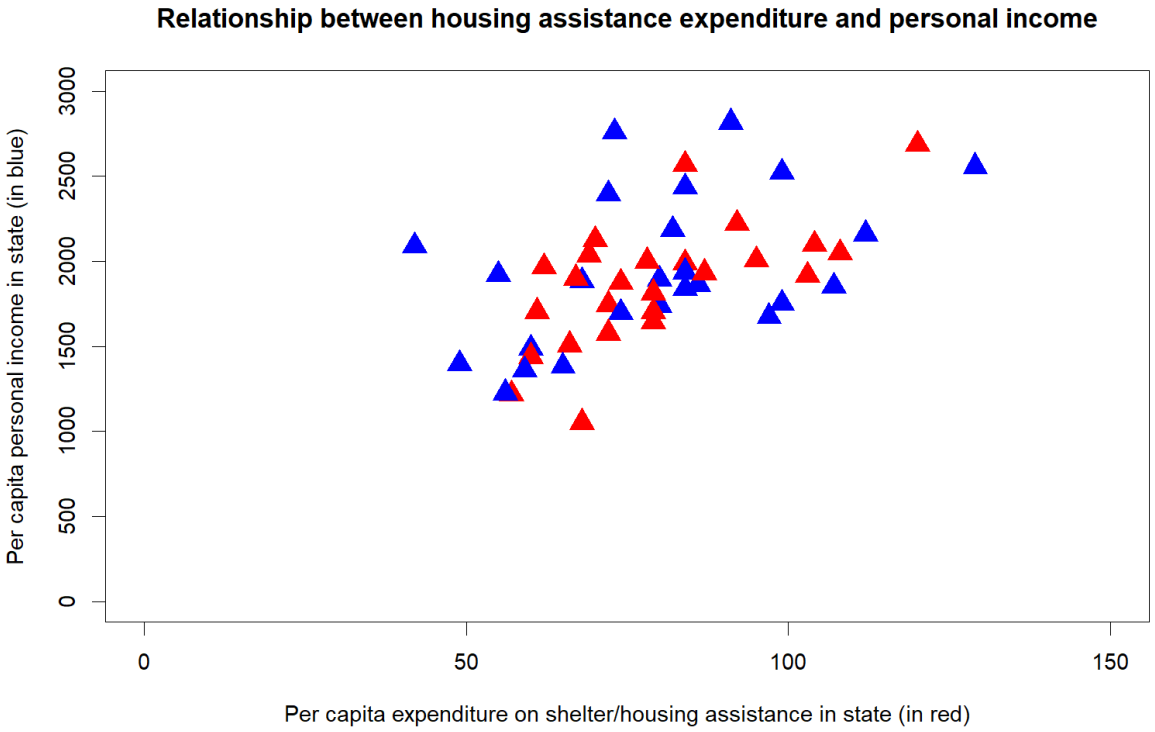
```
boxplot(Region_1_Northeast_expenditure, Region_2_NorthCentral_expenditure,
        Region_3_South_expenditure, Region_4_West_expenditure,
        main="Relationship between regions and spending on shelter/assistance",
        xlab = "Regions", ylab = "Spending",
        names = c("Northeast", "North Central", "South", "West"))
```




*Result: The region with the highest average per capita expenditure on housing is the West region.*

• 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 colours.

```
StatePerCapitaIncome_and_PerCapitaHousingAssistanceExpenditure <-
  plot(expenditure$Y, expenditure$X1, main = "Relationship between housing
assistance expenditure and personal income",
       xlab = "Per capita expenditure on shelter/housing assistance in state (in red)",
       ylab = "Per capita personal income in state (in blue)", pch = 17, cex = 2, type =
"p",
       col=c("red", "blue"), xlim = c(0, 150), ylim = c(0, 3000))
```




```
cor(expenditure$Y, expenditure$X1)
R_Squared_Y_X1 <- cor(expenditure$Y, expenditure$X1) ^ 2

c(round(cor(expenditure$Y, expenditure$X1), 2))
R_Squared_Y_X1 <- c(round(cor(expenditure$Y, expenditure$X1) ^ 2, 2))
R_Squared_Y_X1 = 0.28
```

$r$ = 0.53
$r^2$ = 0.28

***Result: More than 28% of the variance in per capita expenditure on shelter/housing assistance (Y) correlates with per capita personal income in individual states, which we can see in the regression plot. Thus, we can tentatively interpret these data as suggesting personal income plays a significant role in shelter/housing expenditure.***