Submission for Problem Set 1

Applied Stats/Quant Methods 1

Duc Minh, VU
TCD StudentID: 22996761 / UCD StudentID: 19211157

Question 1: Education

Data will first be loaded or in this case manually constructed:

```
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. **Calculation of 90% confidence interval for student IQ**

   (a) *Step 1: Calculation of the sample mean:*
   • Calculation of sample mean manually:

   ```
   ybar_mnl <- sum(y)/length(y)
   ```

   The sample mean for the IQ score calculated manually is 98.44.
   • Calculation of sample mean using R:

   ```
   ybar_r_cal <- mean(y)
   ```

   The sample mean for the IQ score calculated using R is also 98.44.
   • Using R to double-check if the manual score is the same as the R-calculated score:

   ```
   ybar_r_cal == ybar_mnl
   ```

   ```
   [1] TRUE
   ```

   R output confirms that the manually caculated mean score is equal to/the same as the R-calculated mean score.

   (b) *Step 2: Calculation of the sample variance and standard deviation:*
   • Manual calculation:

   ```
   var_mnl <- (sum((y - ybar_mnl)^2))/(length(y)-1)
   sd_mnl <- sqrt(var_mnl)
   ```

The manually caclulated value for the sample variance is 171.42333, and for the standard deviation is 13.0929.

• Calculation using R:

```
var_r_cal <- var(y)
sd_r_cal <- sd(y)
```

The R-caclulated value for the sample variance and standard deviation is also 171.42333, and 13.0929 respectively.

• Using R to double-check the results between the manual and the R calculation:

```
ybar_r_cal == ybar_mnl
```

```
[1] TRUE
```

R output confirms that the manually caculated score is equal to/the same as the R-calculated score for variance and standard deviation.

(c) *Step 3: Finding the associated t-score :*
Since the number of observation in the dataset is less than 30, t-distribution is more appropriated and will be used.

```
t10 <- qt(p=0.1/2, df = length(y) - 1, lower.tail = F) #t-score
```

The t-score at 90% confidence level is 1.71

(d) *Step 4: Caclculating the confidence interval*
The confidence interval will be calculated as:

$$\bar{y} \pm \text{t-value} \times \text{Standard Errors}$$

Based on the results from the two previous steps, the ... observations from the IQ test scores are summarized by $\bar{y} = 98.44$ and $s = 13.0929$. The estimated standard error of the sampling distribution of $\bar{y}$ can be calculated as:

$$SE(StandardErrors) = \frac{s}{\sqrt{n}}$$

```
SE <- sd_mnl/sqrt(length(y))
```

Which gives the result of 2.6186 in R.

```
#T-score
lower_90_t <- ybar_mnl - (t10 * sd_mnl/sqrt(length(y)))
upper_90_t <- ybar_mnl + (t10 * sd_mnl/sqrt(length(y)))
```

```
> conf_int_t
[1]  93.95993 102.92007
```

2. **Hypothesis testing for the average student IQ versus the country average IQ score**

(a) *Assumptions*
It will be assumed that the school counselor's sample is randomly selected and normally distributed. The IQ scores itself are quantitative data.

(b) *Generation of hypotheses*
Let μ be the population mean for the school's average IQ score. Since the school counselor is interested in whether her school's average score is HIGHER than the country's average score of 100, the alternative hypothesis will be one-sided:

$$\mathbf{H_1 > 100}$$

And thus the null hypothesis will be:

$$\mathbf{H_0 \leq 100}$$

(c) *Calculation of test statistic*
Based on the value of the sample mean and estimated standard error from the previous part, the test statistic is

$$t = \frac{\bar{y} - \mu}{se}$$

```
test_stat <- (ybar_mnl-100)/(SE)
```

which gives the result of -0.5957.

(d) *Calculation of P-value*
For n = 25, the degree freedom is n - 1 = 24. Hence, the probability (P-value) of a t-score above the observed t-score or the right-tail probability abolve -0.5967 is:

```
p.value_t_val <- pt(test_stat,df=length(y)-1,lower.tail = F)
```

which gives the result P = 0.7215.

(e) *Conclusion*
With sample mean $\bar{y} = 98.44$, the P-value is 0.7215 which is very large. If μ= 100, it would not be unusual to observe $\bar{y} = 98.44$. Since P-value is larger than α (0.7215 >0.05), we fail to reject the null hypothesis H_0. Students in the counselor's chool doesn't have higher average IQ scores than students among all the schools in the country. Check P.199 in textbook

Question 2: Political Economy

1. **The relationships between Y, X1, X2 and X3**
Figure 1 contains the scatterplots for all of the relationships between Y, X1, X2 and X3. The next section will go into detail the relationship for each pair of the variables.

Figure 1: Scatterplots of all the relationships between Y, X1, X2 and X3 .

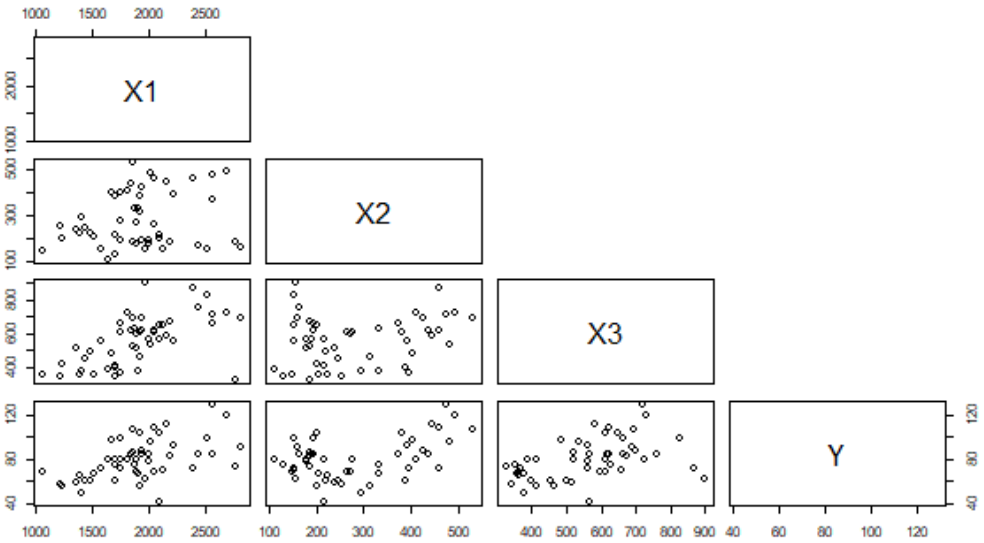
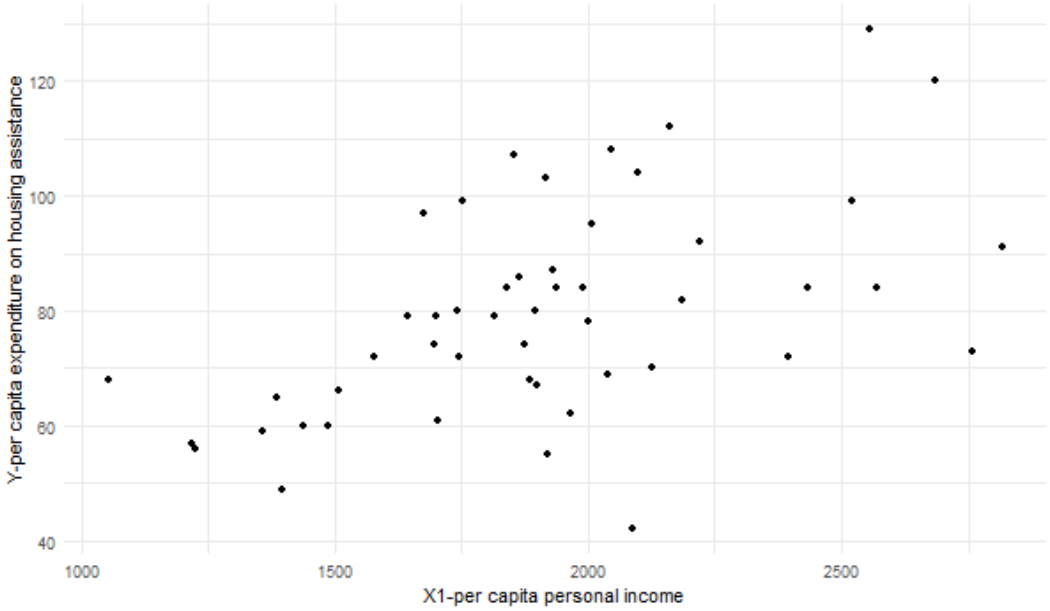


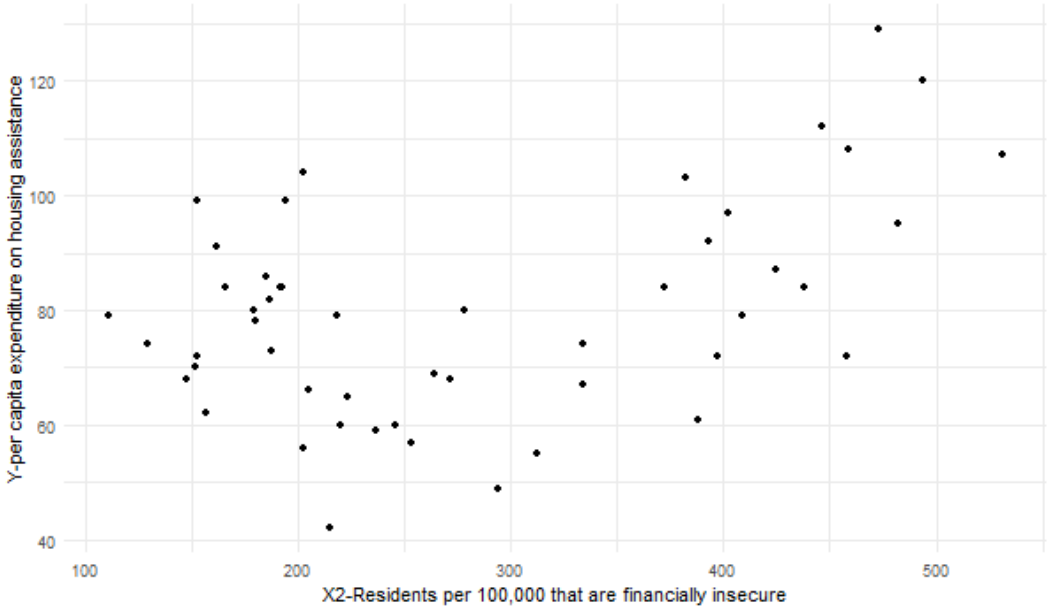
Figure 2: Scatterplot between Y and X1.



(a) The relationship between Y and X1
Figure 2 displays the relationship between Per capita expenditure on housing assistance (Y) and Per capita income (X1). From the graph, there seems to be a positive relationship between the two varaibles, as Y increases and X1 increases

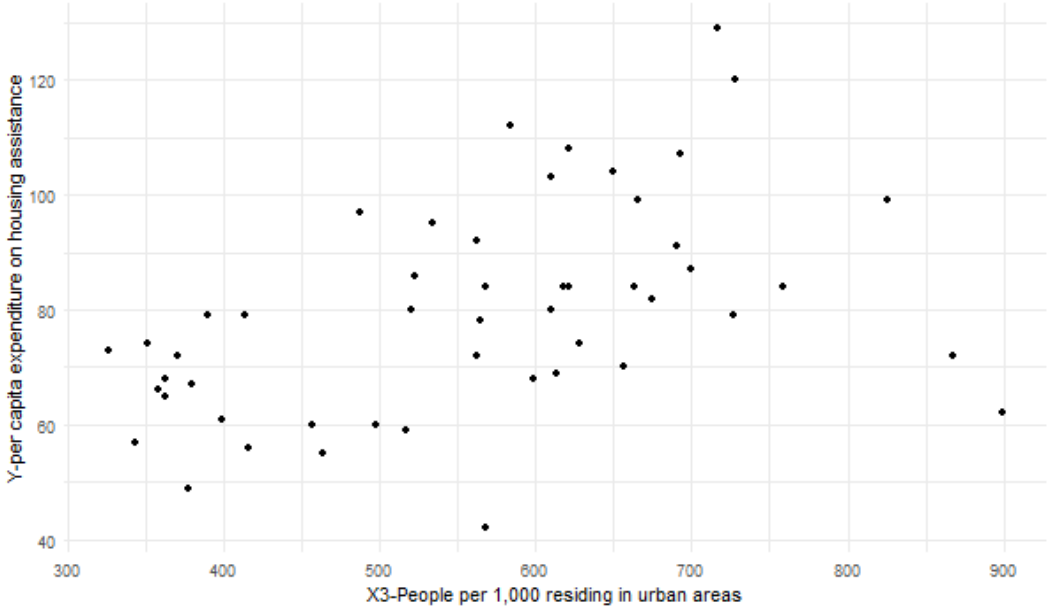
(b) The relationship between Y and X2

Figure 3: Scatterplot between Y and X2.



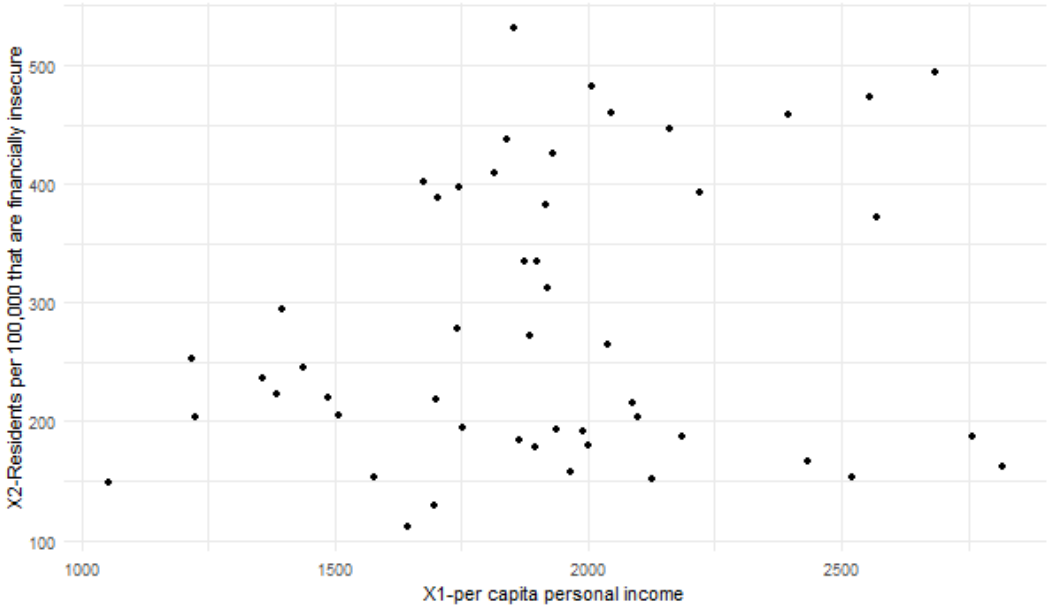
(c) The relationship between Y and X3

Figure 4: Scatterplot between Y and X3.



(d) The relationship between X1 and X2

Figure 5: Scatterplot between X1 and X2.



(e) The relationship between X1 and X3

Figure 6: Scatterplot between X1 and X3.