



# Problem Set 1 Response

Ariana Antunes

October 1st, 2022

## Question 1 (50 points): Education

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

```
1 CI_lower <- qnorm(0.05,
2                   mean = mean(y),
3                   sd = (sd(y)/sqrt(length(y))) # the equation for the standard
4                   error of the mean
5 )
6 CI_upper <- qnorm(0.95,
7                   mean = mean(y),
8                   sd = (sd(y)/sqrt(length(y)))
9 )
10 matrix(c(CI_lower, CI_upper), ncol = 2,
11         dimnames = list("", c("Lower", "Upper")))
12
13 mean(y)
```

Lower	Upper
94.13283	102.7472

```
mean(y)[1] 98.44
```

The hypothesis test with  $\alpha = 0.05$ .

```
1 t.test(y,
2       country_schools_IQ = country_schools_IQ,
3       mu = 0,
4       var.equal = FALSE,
5       alternative = "two.sided",
6       conf.level = 0.05)
```

One Sample t-test

data: y  
t = 37.593, df = 24, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0

5 percent confidence interval: 98.27407 98.60593  
sample estimates: mean of x  
98.44

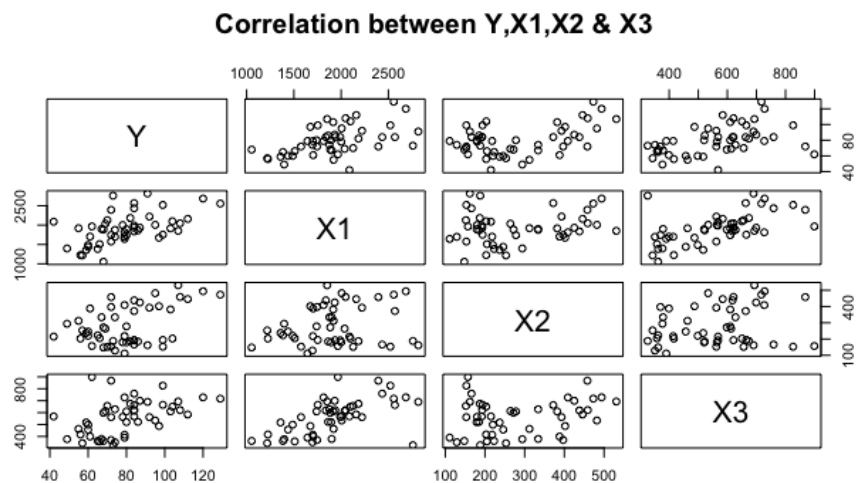
## Question 2 (50 points): Political Economy

### The correlation plot between Y, X1, X2 and X3

It seems the correlation appears to be much similar when comparing the different variables.

```
1 pairs(~ Y + X1 + X2 + X3, data = expenditure, rowlattice = FALSE, bg = "blue",  
2      main = "Correlation between Y,X1,X2 & X3")
```

Figure 1: Correlation between Y, X1, X2 and X3.

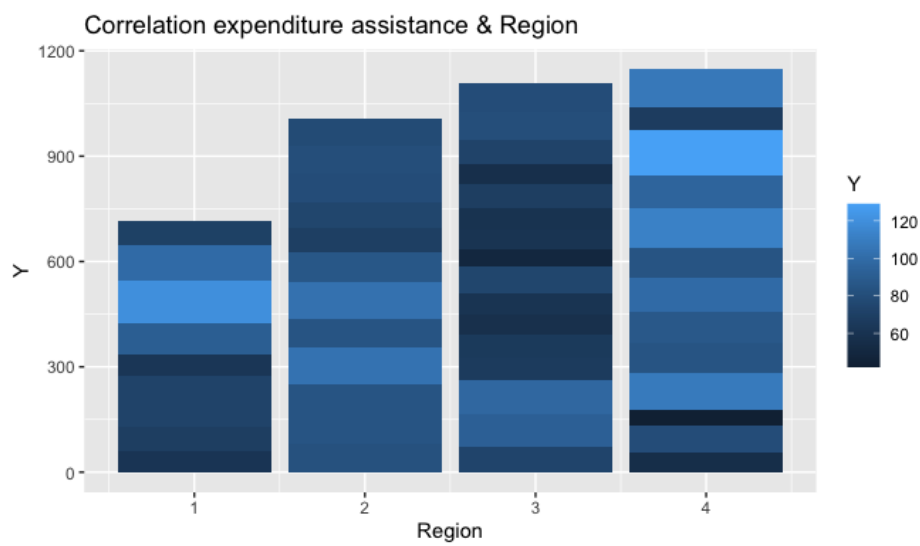


## The correlation plot between Y and Region

On average west region have the highest per capita expenditure on housing assistance.

```
1 ggplot(data = expenditure, aes(x=Region, y=Y)) +  
2   geom_col(aes(fill = Y)) + labs(title = "Correlation expenditure assistance &  
   Region")
```

Figure 2: correlation plot between Y and Region.



## The correlation plot between Y, X1 and Region

It seems there's a clear line between the regions, but region 3 is the lowest per capita expenditure on shelters/housing assistance in state and the lowest income per capita as well, region 4 is the highest income per capita, but it seems region 1 has the highest expenditure on housing assistance.

```
1 ggplot(data = expenditure, aes(x=X1, y=Y)) +  
2   geom_point(aes(color=Region,  
3               shape=as.factor(Region)))
```

Figure 3: correlation plot between Y and Region.

