ECONOMETRICS REPORT

ASSIGNMENT 1

CANDIDATE NUMBER: 702115

BS2280 - ECONOMETRICS 1

DR. YICHEN ZHU

1. Descriptive Statistics Data Analysis

1.1. Descriptive Statistics Summary

```
> prop.table(table(X2011lifeexpectancy$Status))

        0         1
0.8304094 0.1695906
```

Analyzing the "Status" column, it is possible to see that the majority of countries in the 2011 sample were developing countries (83.04%).

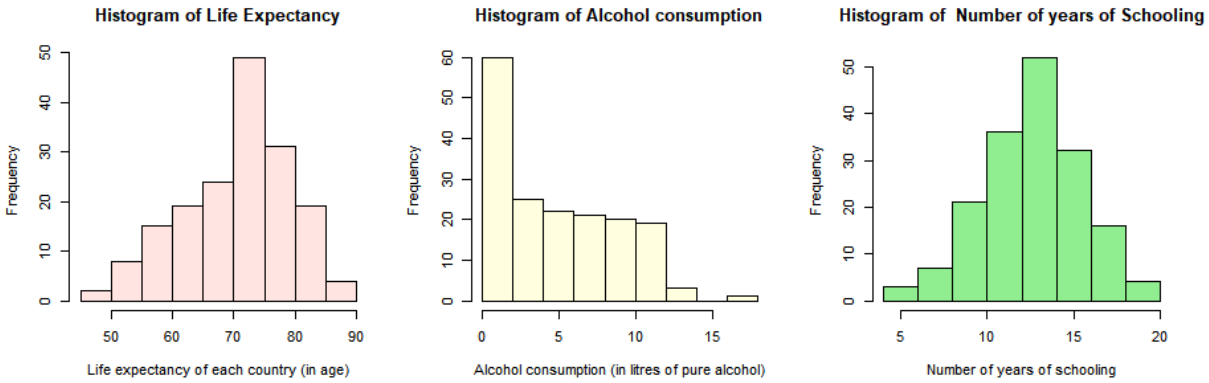
```
> summary(X2011lifeexpectancy[, 3:7])
 Life_Expectancy    Alcohol            BMI        Total_Expenditure   Schooling
 Min.   :48.90   Min.   : 0.010   Min.   : 2.50   Min.   : 0.760    Min.   : 4.80
 1st Qu.:64.75   1st Qu.: 1.160   1st Qu.:22.20   1st Qu.: 4.040    1st Qu.:10.70
 Median :73.40   Median : 4.090   Median :47.10   Median : 5.650    Median :12.90
 Mean   :70.88   Mean   : 4.804   Mean   :40.44   Mean   : 5.898    Mean   :12.68
 3rd Qu.:76.05   3rd Qu.: 7.930   3rd Qu.:57.85   3rd Qu.: 7.520    3rd Qu.:14.50
 Max.   :88.00   Max.   :17.310   Max.   :75.70   Max.   :13.760    Max.   :19.80
> apply(X2011lifeexpectancy[, 3:7],2,sd)
  Life_Expectancy           Alcohol               BMI Total_Expenditure         Schooling
         8.645899          3.901778         20.022287          2.602877          2.837944
```

```
> summary(developed$Life_Expectancy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  72.80   77.00   82.00   80.69   82.60   88.00
> summary(developing$Life_Expectancy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  48.90   62.83   71.45   68.88   74.88   85.00
```

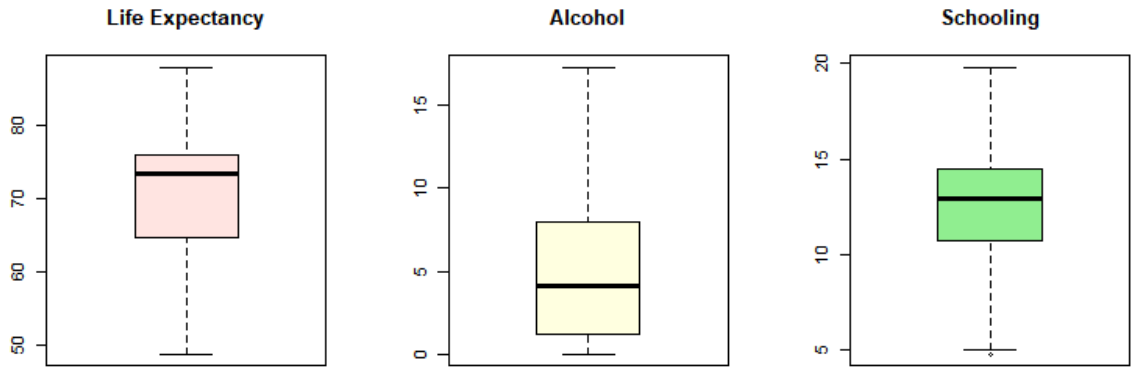
Regarding life expectancy, the median and maximum values are relatively close, both when looking at the entire dataset and when considering developed and developing countries separately. Therefore, there may be a notable clustering of countries with relatively high life expectancies, which is consistent with the positive trend in global life expectancy during this period (Roser, Ortiz-Ospina, and Ritchie, 2013).

Regarding the average BMI, the standard deviation (20.02) is high compared to the range of values for this variable, indicating that there is a high dispersion between the countries. This is the opposite of what occurs in the variables Alcohol, Total Expenditure, and Schooling, in which the mean and median are values close to each other, as well as the standard deviation, is relatively low, indicating that the distribution of values in the data set can be approximately symmetrical and that there is not much variability.

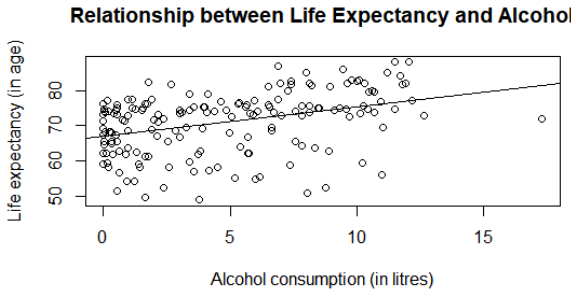
1.2. Analyzing the frequency of Life Expectancy, Alcohol and Schooling

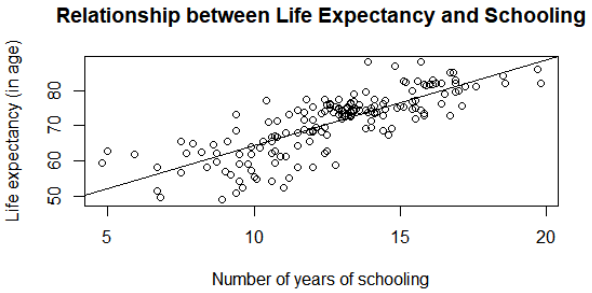


The histogram of the number of years of schooling has a symmetrical format, which may indicate that there is a certain regularity in the data, and in this case, there is a greater frequency of countries whose value of schooling is between 10 and 15 years. Besides, the life expectancy histogram is skewed to the right, in agreement with the statement presented in the previous topic. On the other hand, the histogram of alcohol consumption is skewed to the left, with a large concentration of countries with low consumption, more specifically between 0 and 5 litres. Also, just the “Schooling” variable has an outlier (which is Niger with 4.8 years).



1.3. Analyzing the relationship between some variables





The scatterplots above present positive slopes, so the greater the alcohol consumption and the years of schooling, the greater the life expectancy. However, in the case of the relationship between alcohol consumption and life expectancy, the values are more dispersed and the slope is not accentuated, presenting a weak and less predictable correlation. In the second graph, the points are closer to a straight line and the slope is greater, indicating that the correlation is stronger and that there is less variability among the variables.

```
> cor(X2011lifeexpectancy$Alcohol, X2011lifeexpectancy$Life_Expectancy)
[1] 0.3778416
> cor(X2011lifeexpectancy$Schooling, X2011lifeexpectancy$Life_Expectancy)
[1] 0.7986761
```

By calculating the correlation coefficients, it is possible to verify the points observed in the previous scatterplots. Thus, the correlation between life expectancy and alcohol consumption is weak and positive, whereas the correlation between life expectancy and years of schooling is high and positive, which may make sense as people with greater education tend to have greater access to health information and a greater understanding of healthy lifestyle practices and habits. However, it is important to state that correlation does not imply causation (Rohrer, 2018).

2. References

Rohrer, J.M. (2018). Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. Advances in Methods and Practices in Psychological Science, [online] 1(1), pp.27–42. doi:https://doi.org/10.1177/2515245917745629.

Roser, M., Ortiz-Ospina, E. and Ritchie, H. (2013). Life Expectancy. [online] Our World in Data. Available at: https://ourworldindata.org/life-expectancy [Accessed 21 Oct. 2023].

3. Appendix

 # Read xls file

```
install.packages("readxl")
library(readxl)
X2011lifeexpectancy <- read_xls("C:/Users/maria/OneDrive/Documentos/3. Econometrics 1/BS2280/Coursework1/2011lifeexpectancy.xls")
```

 # Task 1. Generate a descriptive statistics table to summarize your sample.

```
summary(X2011lifeexpectancy[, 3:7])
apply(X2011lifeexpectancy[, 3:7],2,sd)
```

```
> summary(X2011lifeexpectancy[, 3:7])
 Life_Expectancy    Alcohol            BMI        Total_Expenditure   Schooling
 Min.   :48.90   Min.   : 0.010   Min.   : 2.50   Min.   : 0.760    Min.   : 4.80
 1st Qu.:64.75   1st Qu.: 1.160   1st Qu.:22.20   1st Qu.: 4.040    1st Qu.:10.70
 Median :73.40   Median : 4.090   Median :47.10   Median : 5.650    Median :12.90
 Mean   :70.88   Mean   : 4.804   Mean   :40.44   Mean   : 5.898    Mean   :12.68
 3rd Qu.:76.05   3rd Qu.: 7.930   3rd Qu.:57.85   3rd Qu.: 7.520    3rd Qu.:14.50
 Max.   :88.00   Max.   :17.310   Max.   :75.70   Max.   :13.760    Max.   :19.80
> apply(X2011lifeexpectancy[, 3:7],2,sd)
  Life_Expectancy           Alcohol               BMI Total_Expenditure         Schooling
         8.645899          3.901778         20.022287          2.602877          2.837944
```

```
prop.table(table(X2011lifeexpectancy$Status))
```

```
> prop.table(table(X2011lifeexpectancy$Status))

        0         1
0.8304094 0.1695906
```

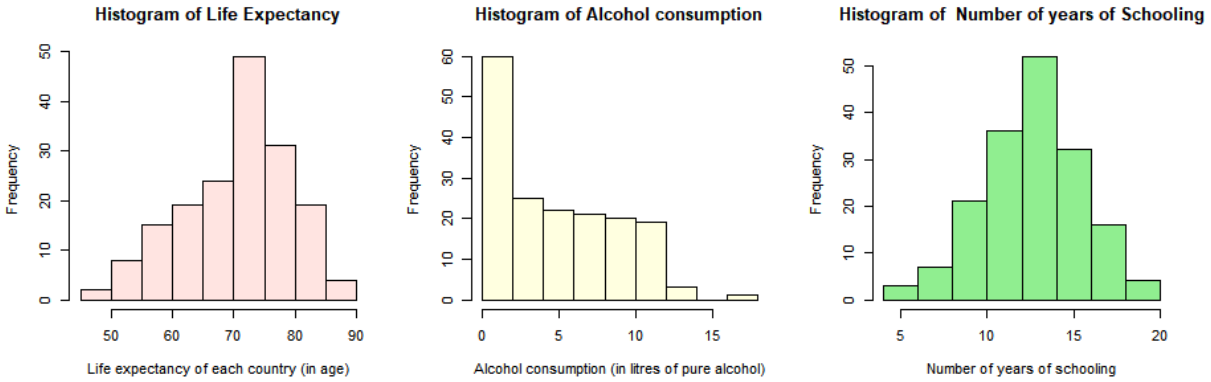
```
developed <- subset(X2011lifeexpectancy, Status == 1)
developing <- subset(X2011lifeexpectancy, Status == 0)
summary(developed$Life_Expectancy)
summary(developing$Life_Expectancy)
```

```
> summary(developed$Life_Expectancy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  72.80   77.00   82.00   80.69   82.60   88.00
> summary(developing$Life_Expectancy)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  48.90   62.83   71.45   68.88   74.88   85.00
```

 # Task 2. Generate histograms for Life Expectancy, Alcohol, and Schooling.

```
par(mfrow = c(1, 3))
hist(X2011lifeexpectancy$Life_Expectancy,
```

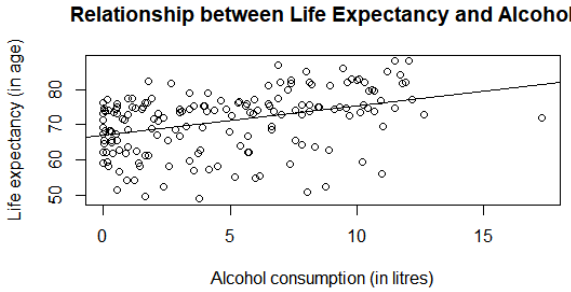
```
    main = "Histogram of Life Expectancy",
    xlab = "Life expectancy of each country (in age)",
    col = "mistyrose")
hist(X2011lifeexpectancy$Alcohol,
    main = "Histogram of Alcohol consumption",
    xlab = "Alcohol consumption (in litres of pure alcohol)",
    col = "lightyellow")
hist(X2011lifeexpectancy$Schooling,
    main = "Histogram of  Number of years of Schooling",
    xlab = "Number of years of schooling",
    col = "lightgreen")
par(mfrow = c(1, 1))
```

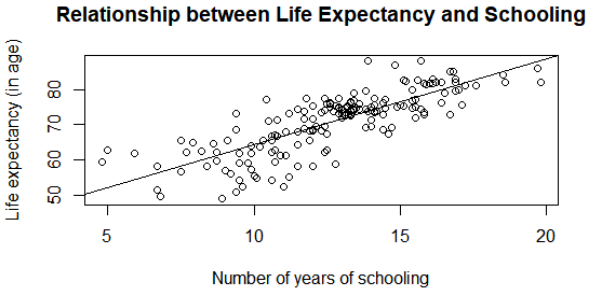


 # Task 3. Plot the relationship between Life Expectancy and Alcohol as well as Life Expectancy and Schooling

```
alcoholfit <- lm(X2011lifeexpectancy$Life_Expectancy ~ X2011lifeexpectancy$Alcohol,
data=X2011lifeexpectancy)
schoolingfit <- lm(X2011lifeexpectancy$Life_Expectancy ~
X2011lifeexpectancy$Schooling, data=X2011lifeexpectancy)

par(mfrow = c(1, 2))

plot(X2011lifeexpectancy$Life_Expectancy ~ X2011lifeexpectancy$Alcohol,
    main = "Relationship between Life Expectancy and Alcohol",
    ylab = "Life expectancy (in age)",
```

```
    xlab = "Alcohol consumption (in litres)")
abline(alcoholfit)

plot(X2011lifeexpectancy$Life_Expectancy ~ X2011lifeexpectancy$Schooling,
    main = "Relationship between Life Expectancy and Schooling",
    ylab = "Life expectancy (in age)",
    xlab = "Number of years of schooling")
abline(schoolingfit)

par(mfrow = c(1, 1))
```





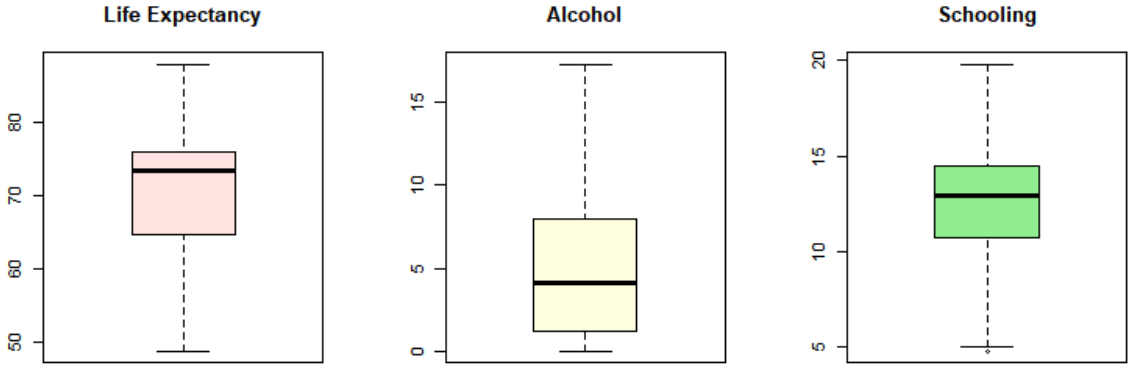
 # Task 4. Calculate the correlation coefficient between (i) Life Expectancy and Alcohol, and (ii) Life Expectancy and Schooling.

```
cor(X2011lifeexpectancy$Alcohol, X2011lifeexpectancy$Life_Expectancy)
cor(X2011lifeexpectancy$Schooling, X2011lifeexpectancy$Life_Expectancy)
```

```
> cor(X2011lifeexpectancy$Alcohol, X2011lifeexpectancy$Life_Expectancy)
[1] 0.3778416
> cor(X2011lifeexpectancy$Schooling, X2011lifeexpectancy$Life_Expectancy)
[1] 0.7986761
```

 # Extra. Boxplot to analyze outliers

```
par(mfrow = c(1, 3))
boxplot(X2011lifeexpectancy$Life_Expectancy, main = "Life Expectancy", col =
"mistyrose")
boxplot(X2011lifeexpectancy$Alcohol, main = "Alcohol", col = "lightyellow")
boxplot(X2011lifeexpectancy$Schooling, main = "Schooling", col = "lightgreen")
par(mfrow = c(1, 1))
```


Life Expectancy
Alcohol
Schooling
50
60
70
80
0
5
10
15
5
10
15
20