**Homework 4**

**The code**

rm(list=ls()) #Clears the global environment

setwd("C:/Users/hp/OneDrive - Al Akhawayn University in Ifrane/Desktop/QUANTITATIVE MTHDS") #setting the working directory

set.seed(1995) #set seed

#Load the tidyverse package

library(tidyverse)

#Load the Pairviz package

library(PairViz)

#load the data

data\_shape <- read.csv("morocco\_data\_regions.csv", header = TRUE)

#Organize the data means and variances according to regions

region\_info <- data\_shape %>%

group\_by(region) %>%

summarize(mean\_hdi = mean(HDI\_04), var\_hdi= var(HDI\_04))

#Visualization of the HDI means and variability of all the regions:

data\_shape %>%

ggplot(aes(x = region , y = HDI\_04 , color = region))+

geom\_violin()+

theme\_grey()+

ylab("HDI in 2004")+

theme(legend.title = element\_blank())+

theme(legend.position = "none")+

theme(axis.title.x = element\_blank())+

theme(axis.text.x = element\_text(angle = 60 , hjust = 1))+

stat\_summary(fun.y = mean ,geom = "point" , shape = 23 , size = 2)

#Annova test assuming that the variances of the populations are equal:

res.aov <- aov (HDI\_04 ~ region, data\_shape)

summary (res.aov)

#Annova test not assuming that the variances of the populations are equal:

res.aov <- oneway.test(HDI\_04~ region, data\_shape)

res.aov

#Testing two population means of all 12 regions together (testing all the possible pairs)

test<- pairwise.t.test(data\_shape$HDI\_04, data\_shape$region,

p.adjust.method = "BH", pool.sd = FALSE)

test

#get the test results and create a p-value matrix(I had to install ‘graph’)

region <- with(data\_shape, split(HDI\_04, region))

regionNames <- names(region)

pvals <- test$p.value

weights <- pvals[!is.na(pvals)]

weights <- edge2dist(weights)

weights <- as.matrix(weights)

rownames(weights) <- regionNames

colnames(weigths) <- rownames(weights)

view(weights)

**Questions:**

1. According to the violon plot, the Dakhla-Oued eddahab region violon indicates the highest mean HDI. As its upper part is wider, the most communes in Dakhla has the highest HDI. However, Laayoune-Sakia elhamra region’ violon indicates high HDI values too, but with a lower density. Yet, the mean of the latter is lower than Dakhla-Oued eddahab’ mean. Other regions have communes, with lower mean HDI, have some HDI values higher than Dakhla’s peak HDI value. Statistically speaking, we aren’t sure that Dakhla region has the highest HDI values, as Laayoune region isn’t statistically proven to have different HDI means, but we are sure that, Dakhla has the highest HDI values compared to other regions left.
2. According to the violon plot, the region with the lowest HDI values is Marrakech as it has the lowest mean. Most communes in Marrakesh-Safi region have a low HDI, as its violon indicates a wider lower part (majority of communes have low HDI). Statistically speaking, the p-value matrix indicates that all regions have different means, thus a lower HDI mean; so, we are 95% sure that Marrakech-Safi region has the lowest HDI values.