Final Exam (R Code)

STAT380

2023-11-27

########################################################################### # United Arab Emirates University   
# # College of Business and Economics   
# # Final Exam (R Codes)  
# # STAT380 - Fall 2023   
# # # ########################################################################### ########################################################################### # Instructions # #   
## 1. The exam should be submitted in Word format (save as pdf) using Markdown.   
## 2. You need to submit A .RMD file and the correspondind output  
## 3. The maximum grade is 50 #

# Name and ID

###########################################################################   
# # Name:   
# # ID :   
######################################################################### ########################################################################### # Set your working directory

# Problem1

## Set the seed using your student ID ## set.seed(ID)   
# Problem 1 [20 points] : Students in high school have passed their exam on   
# 6 topics, Gaelic (X1), English(X2), History(X3), Arithmetic(X4),  
# Algebra(X5), and Geometry(X6). The standardized grades are saved   
# in Grades\_Fin.csv.

# 1. Calculate the correlation matrix, plot it and comment on any possible # relationship between the variables.[5 points]

# 2. Run the principal component analysis and answer the following questions # ## a. How many eigenvalues are larger than 1 [3 point]   
## b. How many components you will choose? [3 point]   
## c. What percentage of the total variance is explained by the first two components? ## [3 point]   
## d. Create a plot to Visualize the percentage of variances explained by each ## PC. comment[3]   
## e. Find the factor loadings for all components by selecting the appropriate   
## cutoff. Comment on your findings [3 point]

# Problem2

# Problem 2:[30 points] The Human Development Index is used to classify   
# countries based on different variables. In this problem, we use 4   
# variables for 189 countries, Life Expectancy at Birth (LEF), Expected   
# Years of Schooling (EYS), Mean years of Schooling (MYS), Gross National   
# Income per capita (GNI).   
# The data includes also the HDI for all all countries and their   
# classification, using the variable DEV. The data is stored in HDI\_Fin.csv.   
  
  
## 1. Read the data and obtain a summary on the variables.[2 points]   
## 2. Calculate the mean and standard deviation for all numerical variables. [2 points]   
## 3. Describe the summary statistic for the categorical variable DEV. [1 points]   
## 4. Calculate and plot the correlation matrix between the numerical   
## variables. Comment. [3 points]  
  
## In the following questions, do not use the variable HDI.  
## 5. Calculate and plot the distance matrix using both Euclidean and ## Manhattan distances. [5 points]   
  
## 7. Use a Hierarchical agglomerative cluster and produce a dendogram. [4 points]  
  
##8. By visual inspection of the dendogram how many clusters would you choose for the data? # [1 points]   
  
## 8. Using the optimal number of clusters in the previous question, run a K-means cluster analysis and # comment. [4 points]   
  
## 9. How many observations (countries) are in each cluster? [2 point]   
  
## 10. Visualize the observations in each cluster. Comment. [2 points]   
  
   
## 11. Run a regression tree analysis for the response variable HDI. [4 points]   
## # Note: The analysis should include also the pruning.