**QUIZ 2**

**(Deadline: November 30,2022,23.59)**

Please copy and paste your codes and outputs from R Console to word file, if you do not know how to use R Notebook, R Markdown or Quarto. The name of the solution files has to be your name and surname. Then, send it to ozanstat@gmail.com.

This quiz is mandatory to get a certificate from the course organizers. You should solve at least 3 parts from each question.

**1.** . Install "mlbench" library in R. Then, load dataset PimaIndiansDiabetes that is available in mlbench library. The expressions for the variable in this dataset are given below.

- pregnant Number of times pregnant
- glucose Plasma glucose concentration (glucose tolerance test)
- pressure Diastolic blood pressure (mm Hg)
- triceps Triceps skin fold thickness (mm)
- insulin 2-Hour serum insulin (mu U/ml)
- mass Body mass index (weight in kg/(height in m)\^2)
- pedigree Diabetes pedigree function
- age Age (years)
- diabetes Class variable

Answer the following questions related with this dataset.

**a.** What is the average value (mean) of 2-Hour serum insulin?

**b.** What is the average value of 2-Hour serum insulin in terms of level of diabetes variable which are negative and positive?

**c.** Which level in diabetes variable has more variability (standard deviation) around it's average value for 2-Hour serum insulin?

**d.** Measure the correlation coefficient between age and glucose of participants?

**e.** Create a table with summary command for all variables.

**2. ship.txt is a dataset including measurements of ship size, capacity, crew, and age for 158 cruise. Please read the dataset with an appropriate function. Then, answer the following questions.**

**a.** Please find the name of the ships that has maximum age, maximum tonnage, maximum passenger, maximum length, maximum cabin and maximum crew separately.

**b.** Obtain the summary of the dataset.

**c.** What is the association between length of the ship and number of passengers? (You can use either base R commands or ggplot2 package.)

Ozdemir, O. Room No: 234
e-mail: ozancan@metu.edu.tr

**d.** Draw the histogram of age.

**e.** Draw a bar plot of cruise lines. Then, write the name of most frequent three lines.

**f.** Consider the three cruise lines that you found in part d. Then, subset the dataset that contains these three cruise lines and corresponding number of passengers for these lines. Having a data, draw a box plot of passengers to compare these three lines.

**g.** Create a new variable and call it class of the passenger by using logical operator and for loop with regard to the following conditions. If number of passenger is less than 19, then class them as 0, if it is between 19 and 24, then class them as 1, and if it is greater than 24, then class them as 2. After that, draw a bar plot then write the class that has highest frequency.

**3.** Please use apply family functions to solve the following questions.

Load the library 'ggplot2', and dataset 'diamonds'.

**a.** For observations whose row index are between 10000 to 11000, get the mean of columns whose index number are 8, 9, 10.

**b.** Same as 'a' but round the results to one digit. (Hint: use round function)

**c.** Sort the rounded results in ascending order.

**d.** Calculate the median of table by the cut.

**e.** Use 'apply' to perform a modulo division by 2 on each value in the x,y and z columns of the matrix.

Ozdemir, O. Room No: 234
e-mail: ozancan@metu.edu.tr