M.S.Ramaiah Institute of Technology

(Autonomous Institute, Affiliated to VTU)

Department of Computer Science & Engineering

QUESTION BANK FOR VII SEMESTER (Term: Aug-Dec 2016)

Data Analytics Laboratory (CSL1542)

I.A. Marks : 15 **Exam Hours: 02**

1. From the library MASS, use Cars93 data and perform the following.
    a. Using the Cars93 data and the `t.test()` function, run a t-test to see if average `MPG.highway` is different between US and non-US vehicles.
    b. What is the confidence interval for the difference? Interpret this confidence interval.
    c. Repeat part (a) using the `wilcox.test()` function.
    d. Are your results for (a) and (c) very different?

2. Consider the data as shown in the table and perform the following.
    a. Create a table by name 'smoking' with rbind() and list() function.
    b. Use `fisher.test()` to test if there's an association between smoking and lung cancer. Interpret the results of the same.
    c. What is the odds ratio? Interpret this quantity.
    d. Write an inline code chunk that determines whether your findings are statistically significant?

has.smoked	lung.cancer	Freq
yes	yes	688
no	yes	21
yes	no	650
no	no	59

3. Create a data frame based on the data shown in the table and perform the following.
    a. Plot a single scatterplot() that describes the relationships between all the variables in the dataset. What do you infer from the plot i.e. dependent variable?
    b. Apply linear regression model for the same using the dependent variable as income and age,education, gender as the independent variables. Interpret the results of the model? Is it over-fitted?
    c. If the model is overfit in (b), apply linear regression model once again with suitable independent variables.

ID	Income	Age	Gender
1	113	69	1
2	91	52	0
3	121	65	0
4	81	58	1
5	68	31	1

4. Load the in-built dataset mtcars() and perform the following.
    a. Dot plot of mpg for each car model
    b. Create a colored histogram of 12 bins with x-axis as 'Miles per gallon' and y-axis as 'frequency'.
    c. Create kernel density plots of mpg by number of cylinders with legends as 4 cylinders, 6 cylinders and 8 cylinders. Interpret the results obtained in (a) & (b).
    d. Generate a box plot of car mileage versus transmission type and number of cylinders.
5. Examine the built in ChickWeight data and perform the following.
    a. Construct a plot of weight against time for chick number 34.
    b. For chicks in diet group 4, display box plots for each time point.
    c. Compute the mean weight for chicks in group 4, for each time point. Plot this mean value against time.
    d. Repeat the previous computation for group 2. Add the mean for group 2 to the existing plot.

6. Create the following patterned matrices. In each case, your solution should make use of the special form of the matrix—this means that the solution should easily generalise to creating a larger matrix with the same structure and should not involve typing in all the entries in the matrix.

(a)
$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 0 \\ 2 & 3 & 4 & 0 & 1 \\ 3 & 4 & 0 & 1 & 2 \\ 4 & 0 & 1 & 2 & 3 \end{pmatrix}$$

(b)
$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 8 & 9 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 9 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{pmatrix}$$

(c)
$$\begin{pmatrix} 0 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 1 & 0 & 8 & 7 & 6 & 5 & 4 & 3 & 2 \\ 2 & 1 & 0 & 8 & 7 & 6 & 5 & 4 & 3 \\ 3 & 2 & 1 & 0 & 8 & 7 & 6 & 5 & 4 \\ 4 & 3 & 2 & 1 & 0 & 8 & 7 & 6 & 5 \\ 5 & 4 & 3 & 2 & 1 & 0 & 8 & 7 & 6 \\ 6 & 5 & 4 & 3 & 2 & 1 & 0 & 8 & 7 \\ 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 & 8 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 \end{pmatrix}$$

7. (a) Suppose matA is a matrix containing some occurrences of NA. Pick out the submatrix which consists of all columns which contain no occurrence of NA. So the objective is to write a function which takes a single argument which can be assumed to be a matrix and returns a matrix.
   (b) Now write a function which takes a single argument which can be assumed to be a matrix and returns the submatrix which is obtained by deleting every row and column from the input matrix which contains an NA.