

Note: Marks are indicated against each question. Parts of a question must be answered together.

Q1. Attempt any four questions out of the following: [4 * 5marks =20]

- a) Create an array of dimension 3,3,2 from two numeric vectors and set dimension names. Display the first and third column of second matrix.
- b) Explain `rbind()` and `cbind()` functions with an example each.
- c) (i) Load `diamonds` dataset from `ggplot2` library and plot a graph to show relationship between `diamonds$price` and `diamonds$carat`.
(ii) Explain the usage of `which` function with an example.
- d) Create a function `oddeven(m)` to check and then print whether the given number `m` is an odd number or an even number.
- e) Load `mtcars` dataset and write `R` code for the following:
 - (i) Display all columns from `mpg` to `hp`.
 - (ii) Filter those records from `mtcars` dataset whose `cylinder=6`.
 - (iii) Arrange the records on the basis of ascending order of `mpg`.

Q2. Attempt any four questions out of the following: [4* 10marks=40]

- a) Load `UScrime` dataset and `car` and `psych` packages and determine multicollinearity in the dataset. Determine the number of principal components required. Write `R` code to perform principal component analysis on the dataset and obtain factor loadings.

- b) Load *HouseVotes84* dataset and *caret* and *e1071* package and perform naïve bayes classification to determine the *class* of the US House of Representatives Congressmen. Use 70:30 *train:test* ratio to partition the dataset and compute the accuracy of the classifier.
- c) Briefly explain the utility of `apply()`, `tapply()`, `lapply()`, `sapply()` and `mapply()` functions in R using examples.
- d) Write R code to discuss sub-setting (element(/s) referencing) for the following datatypes: vectors, dataframes, lists, arrays, matrices with the help of examples.
- e) Write R code to show that *airquality* dataset contains missing values and to calculate total number of observations with and without missing values. Write a custom function which will replace all missing values in a vector with the mean value. Use that function to perform missing value imputation on *Ozone* column.
