

MSc in Bioinformatics and Applied Biotechnology

Practicals Assignment 1

For each question, write an R-script whose name is mentioned below the question.

1. The data sets $V1, V2, V3, V4$ and $V5$ are given below:

$V1 = \{ "S1", "S2", "S3", "S4", "S5", "S6", "S7" \}$

$V2 = \{ 233, 255, 199, 265, 200, 215, 207 \}$

$V3 = \{ 10, 20, 30, 40, 50, 60, 70 \}$

$V4 = \{ 12.22, 17.55 \}$

$V5 = \{ "experiment - 1", "Method - standard" \}$

Write an R-script "Question-1.r" which does the following:

- (a) For each of the five data sets, create a vector whose name is the same as that of the data set.
- (b) Create a data frame called "table-1" from the vectors $V1, V2$ and $V3$. Print the columns of this data frame separately.
- (c) Multiply the columns $V2$ and $V3$ and print the resulting column vector.
- (d) Create a list called "alldata" that contains the data frame "table-1" and the vectors $V4, V5$.
Access the vector $V5$ from the list and print its elements.

R-script name : Question-1.r

2. The given data file "crop_table.txt" contains the Indian agricultural data for the fractional land cultivation of various crops in the country (expressed in percentage of total cultivation area in that year) over the years.

- (a) Read the given data into a data frame called "indat".
- (b) Create a column consisting of the sum of cultivation areas of Rice, wheat and Cereals over the years.
- (c) Create a subset of data in which cereals were cultivated in more than 10 percent of total area and print this subset.

R-script name : Question-2.r

3. Read the given data file "crop_table.txt" into a data frame called "mydata".

- (a) Create a bar plot comparing the cultivation area of rice and wheat during 1980-81 and 2012-13. Label the entities properly.
- (b) Create a Pie-chart depicting the percentage of cultivation areas of all the crops. Write the legends properly.

R-script name : Question-3.r

4. (a) Write a R function called "string_location" which takes two strings "strbig" and "strsmall" with the condition that $length(strbig) > length(strsmall)$.

The function checks whether strsmall is a substring of strbig. If so, it returns the start location of strsmall in strbig. If not, it returns zero.

Demonstrate this with example DNA strings.

(b) Write a R function called "VectorOp" which takes two vector of numbers V1 and V2 as parameters. It returns a vector which is the square root of the product of two vectors. Demonstrate this function with example.

R-script name : Question-4.r

5. Write a R-script for performing the following tasks:

(a) Plot a Poisson distribution with a mean 5 in the range 0 to 12. The title of the plot should be "Poisson distribution : mean = 5".

(b) Generate 10000 points randomly drawn from a Gaussian distribution whose mean is 6.0 and standard deviation is 1.6. Plot the histogram of these data points

(c) For this data, find the number of data points more than 2 standard deviations above the mean value and print this number.

R-script name : Question-5.r

6. In a clinical trial, the placebo and treatment groups consisting of 12 patients each were subjected to a test. The results are presented here:

Placebo : 3.1,5.2,5.3,4.7,5.4,5.7,3.8,6.2,6.9,5.5,4.1,7.8

Treatment : 4.9,6.9,7.1,4.9,4.5,6.1,6.4,6.2,6.3,7.4,5.4,4.4

Write an R script that

(A) plots their bar plots side by side.

(B) Add standard error on mean as error bar in the plot

R-script name : Question-6.r

7. In order to correlate the salary earned by an employee to his work experience in a particular company, 9 employees were randomly selected and the data on their work experience and salary was collected.

It is presented here:

Work Experience (years) = 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5

Monthly Salary (Thousand of Rupees) = 23.7, 25.6, 34.0, 35.8, 44.0, 46.0, 54.2, 55.9, 62.5

Write an R script to plot Monthly salary as a function of work experience.

Give X, Y titles properly.

Plot should have filled circles in blue color as points, also joined by a line.

R-script name : Question-7.r

8. Expression levels of a gene is measured in a particular experiment. A random sample of 15 experiments gave the following result:

95,90,99,98,88,86,92,95,95,87,88,89,87,87

Assuming the experimental measurement of gene expression follows a normal distribution, find a 95% confidence interval for the population mean expression level.

Print the mean, standard deviation and confidence interval.

R-script name : Question-8.r