

PRACTICE TASKS

VECTOR

1. Create a vector x13 with values 2, 3, 4, 5, 6
2. Create a vector x14 with values 2.0, 2.1, 2.2, 2.3, 2.4, ..., 4
3. Create a vector x15 with 10 random values between 4 and 6
4. Create a vector x16 with repeated values 3, 4, 5, 3, 4, 5, 3, 4, 5
5. Create x17 with repeated values 7,7,7,8,8,8,9,9,9
6. Create a vector x18 with 10 random values between 20 and 30
7. Create a vector x19 with 10 normally distributed random values
8. Create a vector x20 with values of vectors x13 and x16 followed by 3, 5,10
9. Create a vector x21 with values 33,55,66,88,99. Fetch its 3rd, 5th and 2nd values
10. Fetch values of x21 from 1 to 4
11. Fetch values of x21 vector excluding 2nd and 3rd elements
12. Fetch last element of x21 using length()
13. Create a vector x23 with values 5, 7, 6, 8, 1, 4. Delete 1st and last element. Reset the value of second element to 12. Add value 0 at the beginning of a vector x
14. Write the arithmetic expression to calculate variance of a vector. Cross check your result using var() function.
Formula: $\text{Variance} = \frac{\sum((x - \text{mean}(x))^2)}{n-1}$ where n is total number of elements.
15. Given x, y, z coordinates of two atoms.
Atom1 (1.2, 2.3, 3.4) and Atom2 (4.5, 5.6, 6.7). Find distance between 2 atoms.
Formula: $\text{sqrt}((x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2)$

MATRIX

17. Create a matrix from a vector consisting of numbers from 1 to 12 with 3 columns
 18. Fetch 2nd row. Fetch 3rd column
 19. Fetch the value 6
 20. Fetch the value 8 and 12
 21. Fetch the value 7, 8, 11 and 12
-

DATA FRAME

22. Create a data frame of gene expression data such that

First column, "Genes", is a character vector of 6 gene names (G1, G2, ..., G6).

Second column, "C1" is a numeric vector of 6 random values from 3 to 5.

Hint: Generate random numbers using function `sample()`. Use R help to see the syntax of `sample`.

Third column, "C2" is a numeric vector of 6 random values from 3 to 5.

Fourth column, "T1" is a numeric vector of 6 random values from 5 to 7.

Fifth column, "T2" is a numeric vector of 6 random values from 5 to 7.

Sixth column, "Pathway" is a character vector of which first 3 represent one pathway "P1" and other 3 represent pathway "P2".

23. Fetch values of column T2
24. Fetch values of gene G2
25. Fetch value of gene G3 from C2
26. Delete column C2
27. Insert column C3, which should be numeric vector of 6 random values from 3 to 5.
28. Find mean of column C1

Tasks on Iris data set

1. Open iris data set file using `read.table()` and store in a variable names “iris_data”
2. Check the structure of “iris_data”. Note the column names. How many categories are there in column named as “Species”. Note the names of species.
3. How many rows and columns are there?
4. How many observations (rows) are there for each species?
 - a. Number of rows with Species as setosa
 - b. Number of rows with Species as virginica
 - c. Number of rows with Species as versicolor
5. Find the mean of all sepal lengths.
6. Find the mean of sepal lengths in Species setosa.
7. What is the overall correlation between Sepal length and Petal length?
8. What is the correlation between Sepal width and Petal width of Species virginica?
9. Find the difference between sepal lengths of species setosa and versicolor. What is the mean difference between them?
10. Carry out the t-test between sepal lengths of species setosa and versicolor. Is it statistically significant?