

**DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON**

Name Kawale Roshani Vasant
Expt. Title To perform the basic mathematical operations in r programming.
Class SYMCA Batch B2 Performed on _____
Roll No. 48 Expt. No. 1 Submitted on _____
Remarks _____ Returned on _____

What is R?

R is a popular programming language and used for statistical computing and graphical presentation. Its most common use is to analyze & visualize data.

"R is an interpreted computer programming language which was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand".

Features R programming:-

1. It is a simple and effective programming language which has been well developed.
2. It is data analysis software.
3. It provides effective data handling and storage facility.
4. It is an open-source, powerful, and highly extensible software.
5. It provides highly extensible graphical techniques.
6. It allows us to perform multiple calculations using vectors.
7. R is an interpreted language.

Why Use R?

- It is a great resource for data analysis, data visualization, data science and machine learning.
- It provides many statistical techniques (such as classification, clustering & data reduction).
- It works on different platforms (Windows, Linux).
- It is open-source and free.
- It has a large community support.
- It has many packages that can be used to solve different problems.

Incomplete for :

- 1) Algorithm
- 2) Flow Chart
- 3) Programme Listing
- 4) Results
- 5) Comments

Basic Command of R Programming:-

1) Assignment operator:-

The assignment operator is used to assign a value.

Assignment operators in R:

i) Leftward assignment:

$x \leftarrow 3$

$x \llleftarrow 3$

ii) rightward assignment:

value $\rightarrow x$

value $\Rightarrow x$

iii) Is equal to

$x = \text{value}$.

2) Evaluation:-

We can then evaluate the variable by simply typing x at the command line which will return the value of x . Note that prior to the value returned you'll see $\#\#[1]$ in the command line.

3) Case Sensitivity:-

Lastly, note that R is case sensitive programming language. Meaning all variables, functions, and objects must be called by their exact spelling.

4) Basic Arithmetic:-

Arithmetic its most basic function R can be used as a calculator. When applying basic arithmetic, the PEMDAS order of operations applies: parenthesis first followed by exponentiation, multiplication and division, and final addition and subtraction.

5) Miscellaneous Mathematical Functions:-

There are many built-in functions to be aware of. These include $\text{abs}()$, $\text{sqrt}()$, $\text{exp}()$, $\text{log}()$, $\text{cos}()$ functions.

6) Infinite and NaN Numbers:-

When performing undefined calculations, R will produce Inf (infinity) and NaN (not a number) outputs.

• In this way we had understood the basics of R programming

**DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON**

Name Kawale Roshani Vasant
Expt. Title Write program for Creating and Manipulating Objects in R-Vectors
Class SYMCA Batch B2 Performed on _____
Roll No. 48 Expt. No. 2.1 Submitted on _____
Remarks _____ Returned on _____

1] Vectors :-

A vector is the basic data structure in R, or we can say vectors are the most basic R data objects. A vector is a collection of elements which is most commonly of mode character, integer, logical or numeric.

In R, a sequence of elements which share the same data type is known as vector. The elements which are contained in vector known as components of the vector. We can check the type of vector with the help of the `typeof()` function.

creation of vector :-

i) Using the colon(:) operator :-

`a ← 4:10`

This operator creates a vector with elements from 4 to 10 and assign it to a.

ii) Using the `seq()` function :-

A sequence function creates a sequence of elements as a vector.

`seq_vec ← seq(1, 4, by = 0.5)`

`seq_vec ← seq(1, 4, length.out = 6)`

Vector Operations :-

i) combining vectors :-

The `c()` function is not only used to create a vector, but also it is also used to combine two vectors.

ex:- `p ← c(1, 2, 4, 5, 7, 8)`

`q ← c("Shubham", "arpita", "nishka")`

`r ← c(p, q)`

2) Arithmetic Operations:-

We can add, subtract, multiply or divide two vectors.

ex:- $a \leftarrow c(1, 3, 5, 7)$
 $b \leftarrow c(2, 4, 6, 8)$
 $a + b$
 $a - b$
 $a * b$
 a / b
 a

3) Logical Index vector:-

With the help of the logical index vector in R, we can form a new vector from a given vector.

ex:- $a \leftarrow c("shubham", "arpita", "Nishka", "sumit")$
 $b \leftarrow c(TRUE, FALSE, TRUE, TRUE)$
 $a[b]$

4) Numeric Index:-

In R, we specify the index between square braces $[\]$ for indexing a numerical value.

ex:- $a \leftarrow c("shubham", "arpita", "Nishka", "sumit")$
 $a[2]$
 $a[-4]$

5) Duplicate Index:-

An index vector allows duplicate values which means we can access one element twice in one operation.

ex:- $a \leftarrow c("shubham", "arpita", "nishka")$
 $a[c(1, 2, 2, 3)]$

6) Range Indexes:-

Range index is used to slice our vector to form a new vector. For slicing, we used colon $(:)$ operator.

ex:- $a \leftarrow c("shubham", "arpita", "Nishka", "sumit")$
 $b \leftarrow a[2:4]$
 b

7) Out-of-order Index:-

In R, the index vector can be out-of-order.

ex:- $a \leftarrow c("shubham", "arpita", "sumit", "nishka")$
 $a[c(3, 1, 4, 2)]$

**DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON**

Name Kawale Rashani Vasant

Expt. Title Write program for creating & manipulating R objects in R-Matrices

Class SYMCA Batch B2 Performed on _____

Roll No. 48 Expt. No. 2.2 Submitted on _____

Remarks _____ Returned on _____

2] Matrices :-

A matrix is an R object in which the elements are arranged in a two-dimensional rectangular layout. In the matrix, elements of the same atomic types are contained. A matrix is created with the help of the vector input to the matrix function. On R matrices, we can perform addition, subtraction, multiplication, and division operation. In the R matrix, elements are arranged in a fixed number of Rows and columns.

creation of matrix in R :-

R provides the `matrix()` function to create a matrix.

`matrix(data, nrow, ncol, byrow, dim_name)`

Matrix Operations :-

- 1) Addition (+)
- 2) Subtraction (-)
- 3) Multiplication (*)
- 4) Multiplication (By constant)
- 5) Division (/)

Incomplete for :

- 1) Algorithm
- 2) Flow Chart
- 3) Programme Listing

DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name Kawale Roshani Vasant

Expt. Title Write program for creating & manipulating R objects in R-Array.

Class SYMTA

Batch B2

Performed on _____

Roll No. 48

Expt. No. 2-3

Submitted on _____

Remarks _____

Returned on _____

3] Array:-

In R, arrays, the data objects which allow us to store data in more than two dimensions. In R, an array is created with the help of the `array()` function. This `array()` function takes a vector as an input and to create an array it uses vector's value in the `dim` parameter.

R Array Syntax:-

`array_name ← array(data, dim = (row_size, column_size, matrices, dim_names))`

Arrays Operations:-

- 1) How to create?
- 2) Naming columns and rows
- 3) Accessing elements
- 4) Manipulation of elements
- 5) Calculation Across array elements.

Incomplete for :

) Algorithm

Flow Chart

Programme Listing

Results

Comments

**DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON**

Name Kawale Reshmi Vasant
Expt. Title Write program for creating & manipulating Objects in R-Data frames
Class SYNCA Batch B2 Performed on _____
Roll No. 48 Expt. No. 2.4 Submitted on _____
Remarks _____ Returned on _____

4] Data frames :-

A data frame is two-dimensional array-like structure or a table in which a column contains values of one variable, and rows contains one set of values from each column.

A data frame is used to store data table and the vectors which are present in the form of a list in a data frame, are of equal length.

creating the data frames:-

In R, data frames are created with the help of `frame()` function of data.

```
emp.data ← data.frame(  
  employee.id = c(1:5),  
  employee.name = c("shubham", "arpita", "sumit", "nishka",  
                    "runjan"),  
  sal = c(623.3, 915.2, 611.0, 729.0, 843.25)  
)  
print(emp.data)
```

for :

DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name Kawade Ashani Vasant
Expt. Title Write program for creating & manipulating R objects in R-Lists.
Class SYMCB Batch B2 Performed on _____
Roll No. 48 Expt. No. 2.5 Submitted on _____
Remarks _____ Returned on _____

5] Lists:-

In R, lists are the second type of vector lists are the objects of R which contain elements of different types such as number, vectors, string and another list inside it. A list is a data structure which has components of mixed data types. It can also contain a function or a matrix as its elements.

list() function is used to create a list in R.

Lists operations:-

- 1) List creation
- 2) Naming list elements
- 3) Accessing list elements
- 4) Manipulating list elements
- 5) Converting list to vector
- 6) Merging lists.

ete for :

thm

Chart

amme Listing

DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name Kawale Rashani Vasant

Expt. Title While program to demonstrate loops & Vectorization
Listing values

Class SYICA Batch B2 Performed on _____

Roll No. 48 Expt. No. 13 Submitted on _____

Remarks _____ Returned on _____

R Looping :-

Loops are used to repeat the process until the expression (condition) is TRUE. R uses three keywords for, while and repeat for looping purpose.

1) R For loop :-

In R, for loop is a loop which is used to iterate the collection from the data structure like a vector. It uses in keyword to get individual elements from the vector. It is similar to the while loop. There is only one difference between for and while, i.e., in while loop, the condition is checked before the execution of the body, but in for loop condition is checked after the execution of the body.

Syntax:-

for (data in data_collection)
statement 1.

2) R While loop :-

A while loop is a type of control flow statement which is used to iterate a block of code several numbers of times. The while loop terminates when the value of the Boolean expression will be false. In this statement, the condition will be checked $n+1$ time, rather than n times.

Syntax:-

```
while (test-expression) {  
    statement  
}
```

or:

Listing

3) R repeat loop:-

A repeat loop is used to iterate a block of code. It is a special type of loop in which there is no condition to exit from the loop. A repeat loop constructs with the help of the repeat keyword in R. The repeat statement executes the body of the loop until a break occurs.

Syntax:-

```
repeat {  
  commands  
  if (condition) {  
    break  
  }  
}
```

★ Vectorization:-

Most of R's functions are vectorized, meaning that the function will operate on all elements of a vector without needing to loop through and act on each element one at a time.

ex:- $x \leftarrow 1:4$
 $x * 2$

Comparison operators, logical operators, and many functions are also vectorized.

★ Missing Values:-

- A missing value is one whose value is unknown.
- Missing values are represented in R by the NA symbol.

DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name Kawale Rashani Vasant
Expt. Title Demonstrate Importing and Exporting data.
Class SYMCA Batch B2 Performed on _____
Roll No. 48 Expt. No. 4 Submitted on _____
Remarks _____ Returned on _____

★ Importing Data Into R :-

- What is the meaning of importing data?

Data Import lets you upload data from external sources and combine it with data you collect via Analytics. To analyze data using R programming language, data should be first imported in R which can be different formats like txt, csv, or any other delimiter separated files.

- Method 1: Using `read.csv()` method :-

Syntax: `read.csv(path, header=TRUE, sep=",")`

- path: The path of the file to be imported.
- header: By default: TRUE. Indicator of whether to import column headings.
- sep = ",": The separator for the values in each row.

- Method 2: Using `read.table()` method :-

Here we will use `read.table()` method to import csv file into R :-

Syntax :- `x ← read.csv2("C:/Users/Documents/myFile.csv", header=TRUE)`

Import Data from a Text File :-

Syntax:

`x ← read.table("file_name.txt", header=TRUE/
FALSE)`

- Method 3: Importing Data from a delimited File :-

R has a function `read.delim()` to read the delimited files into the list.

Syntax:- `read.delim("file_name.txt", sep=" ", header=TRUE).`

* Exporting Data From R :-

Data export is the extraction and conversion of raw data from their existing format into a format required by another application. Exporting data is also a way of backing up data or moving it between two different versions of programs.

- `write.table()`: The R base function `write.table()` can be used to export a data frame or a matrix to a text file.

Syntax :-

```
write.table(x, file, append = FALSE, sep = " ", dec = ".", row.names =  
TRUE, col.names = TRUE)
```

- `write_tsv()`: This method is also used for to export data to a tab separated ("t") values by using the help of `readr` package.

Syntax: `write_tsv(file, path)`.

- `write_csv()`: This method is also used for to export data to a comma separated (",") values by using the help of `readr` package.

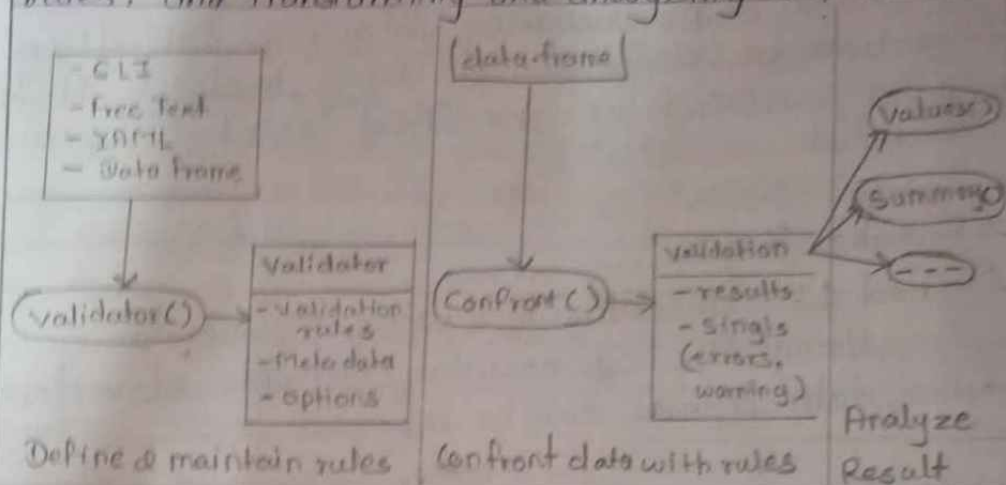
Syntax: `write_csv(file, path)`

**DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON**

Name Kausale Peshani Vasant
 Expt. Title Write program for Validating & Exploring Data Manipulations
 Class Gr 11th Batch B2 Performed on _____
 Roll No. 48 Expt. No. 5 Submitted on _____
 Remarks _____ Returned on _____

★ Validation:-

A data validation task can be split up in three consecutive subtasks: loading the data and the validation rules, confronting the data with the rules, and transforming and analyzing the results.



★ What is data manipulation used for?

It is used in order to make data more understandable or more structured.

proper data analysis involves rearranging, sorting, modifying & shifting data.

★ Which library would you use for data manipulation in R?

Dplyr is mainly used for data manipulation in R.

★ Summarize () method:-

Using the summarize method we can summarize the data in the data frame by using aggregate functions like sum(), mean(), etc.

Syntax:-

Summarize (dataFrame Name, aggregate funⁿ (column Name))

complete for :

algorithm

Flow Chart

Programme Listing

Results

Comments

* Sorting Dataframe in R using dplyr :-

• Sorting in Ascending order :-

Sorting in Ascending order is the default sorting order in `arrange()` function. The attribute to sort by should be given as an argument to this function.

• Sorting in Descending order :-

For sorting our dataframe in descending order, we will use `desc()` function along with the `arrange()` function. we will also use `%>` operator for comparison of the dataframe column which we are taking for sorting purpose.

* Subsetting in R programming :-

In R programming language, subsetting allows the user to access elements from an object. It takes out a portion from the object based on the condition provided.

* R-Subsetting :-

Method 1 :- Subsetting in R using `[]` operator.

Using the `[]` operator, elements of vectors & observations from data frames can be accessed.

Method 2 :- Subsetting in R using `[[]]` operator.

`[[]]` operator is used for subsetting of list-objects. This operator is the same as `[]` operator but the only difference is that `[[]]` selects only one element whereas `[]` operator can select ~~name~~ more than 1 element in a single command.

Method 3 :- Subsetting in R using `$` operator :-

`$` operator can be used for lists and data frames in R. `$` operator is only applicable for recursive objects or list like objects.

Method 4 :- Subsetting in R using `subset()` function :-

`subset()` function in R programming is used to create a subset of vectors, matrices or data frames.

**DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON**

Name Kawale Rishani Vasant
Expt. Title _____
Class SYMCA Batch B2 Performed on _____
Roll No. 48 Expt. No. 5 Submitted on _____
Remarks _____ Returned on _____

*** Merge Dataframes by Row Names in R:**

The merge() function in base R can be used to merge input dataframes by common columns or row names.

*** Joining Data in R with dplyr package:-**

Method 1:- using inner join

In this method of joining data, the user calls the inner-join function, which will result in joined data with the records.

Syntax:-

`inner-join(x, y, by=NULL, on=NULL)`

Method 2:- using left join

In this method of joining data, the user calls the left-join function & this will result in joined data consisting of matching all the rows in the first data frame with the corresponding values on the second.

Syntax:-

`left-join(x, y, by=NULL, on=NULL)`

Incomplete for :

- 1) Algorithm
- 2) Flow Chart
- 3) Programme Listing
- 4) Results
- 5) Comments

**DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON**

Name Kawale Roshani Vasant
Expt. Title Write a program to implement the following analysis techniques using R.
Class SYMCA Batch B2 Performed on _____
Roll No. 48 Expt. No. 6 Submitted on _____
Remarks _____ Returned on _____

Aim:- Write a program to implement following analysis techniques using R. Statistical hypothesis generation and testing

• chi-square test • t-test • Analysis of variance • correlation analysis • Maximum likelihood test • Regression analysis • classification techniques • clustering • Association rules analysis.

What is meant by statistical hypothesis?

A statistical hypothesis is a hypothesis concerning the parameters or form of the probability distribution for a designated population or populations, or more generally, of a probabilistic mechanism which is supposed to generate the observations.

What is the objective of hypothesis testing?

The purpose of hypothesis testing is to making an inference about the population of interest on the basis of a random sample taken from that population.

Seven steps of hypothesis testing:-

step 1:- State the NULL hypothesis.

step 2:- State the Alternative hypothesis.

step 3:- Set

step 4:- Collect data

step 5:- calculate a test statistic

step 6:- Construct Acceptance / Rejection regions

step 7:- Based on step 5 & 6, draw a conclusion about.

Incomplete for :

- 1) Algorithm
- 2) Flow Chart
- 3) Programme Listing
- 4) Results
- 5) Comments

* Importance of statistical hypothesis:-

Hypothesis testing allows the researcher to determine whether the data from the sample is statistically significant.

* What are the five elements of a hypothesis test?

- Specify the NULL Hypothesis
- Specify the Alternative hypothesis.
- set the significance level (α)
- calculate the test statistic & corresponding P-value
- Drawing a conclusion.

* Process of hypothesis Testing:-

1) State the hypothesis:- This step is started by starting null and alternative hypothesis which is presumed as true.

2) Formulate an analysis plan and set the criteria for decision-
In this step, significance level of test is set. The significance level is the probability of a false rejection in a hypothesis test.

3) Analyze sample data:-

In this, a test statistic is used to formulate the statistical comparison between the sample mean ~~se~~ and the mean of the population or standard deviation of the sample and standard deviation of the population.

4) Interpret decision:- The value of the test statistic is used to make the decision based on the significance level.

* One Sample T-testing:-

One sample T-testing approach collects a huge amount of data and tests it on random samples. To perform T-test in R, normally distributed data is required.

Syntax- `t.test(x, mu)`

parameters- x : represents numeric vector of data.

μ : represents true value of the mean.

* Two sample T-testing:-

In two sample T-testing, the sample vectors are compared. If `var.equal = TRUE`, the test assumes that the

parameters: x & y represents numeric data types

DEPARTMENT OF COMPUTER SCIENCE
INSTITUTE OF MANAGEMENT AND RESEARCH, JALGAON

Name Kausale Rashmi Vasant

Expt. Title _____

Class 3YPICB

Batch B2

Performed on _____

Roll No. 48

Expt. No. C

Submitted on _____

Remarks _____

Returned on _____

variance of both are samples are equal.

Syntax:-

$t.test(x, y)$

Parameters: x and y : Numeric vectors.

* Directional Hypothesis:-

using the directional hypothesis, the direction of the hypothesis can be specified like, if the user wants to know the sample mean is lower or greater than another mean sample of the data.

Syntax:- $t.test(x, mu, alternative)$

parameters: x : represents numeric vector data

mu : represents mean against sample data.

$alternative$: sets the alternative hypothesis.

One sample U -test:-

This type of test is used when comparison has to be computed on one sample & the data is non-parametric.

It is performed using " $wilcox.test()$ " function.

Syntax - $wilcox.test(x, y, exact=FALSE)$

Two Sample U -test:-

This test is performed to compare two samples of data.

Correlation Test:-

This test is used to compare the correlation of the two vectors provided in the function call or to test for the association betⁿ the paired samples.

Syntax: $cor.test(x, y)$

parameters: x & y : represents numeric data types

complete for:

Algorithm

Flow Chart

Programme Listing

Results

Comments