Savitribai Phule Pune University

F. Y. B. C. A. (Science)Semester-I

Lab Course – BCA 117

WorkBook

Name:_____

College Name:_____

Roll No.:_____Division:_____

Academic Year:_____

Savitribai Phule Pune University

Section-I

F. Y. B. C. A.(Science)

SEMESTER I

BCA - 117

Lab Course - III
Mathematics Assignments

Editors:

Section-I:

Dr. Yogeshri Gaidhani, MES Abasaheb Garware College, Pune.

Section-II:

Dr. Nanda Rajput, MES Abasaheb Garware College, Pune.

Reviewed By:

Dr. Manisha Bharambe, MES Abasaheb Garware College, Pune.

Introduction

1. **About the workbook:**
   This workbook is intended to be used by F.Y.B.C.A. (Science) students for the Mathematics and Statistics Assignments in Semester–I. This workbook is designed by considering all the practical concepts / topics mentioned in syllabus.

2. **The objectives of this work book are:**
   1) Defining the scope of the course.
   2) To bring the uniformity in the practical conduction and implementation in all colleges affiliated to SPPU.
   3) To have continuous assessment of the course and students.
   4) Providing ready reference for the students during practical implementation.
   5) Provide more options to students so that they can have good practice before facing the examination.
   6) Catering to the demand of slow and fast learners and accordingly providing the practice assignments tothem.

3. **How to use this workbook:**
   The workbook is divided into two sections. Section-I is related to Advance Mathematics assignments and Section-II is related to statistics. Section-I (Mathematics) is divided into Five assignments. Each assignment consists of problems on a particular topic to be solved manually. The instructor should prepare one more assignment on each topic, which will help students understand all the concepts. The problems are designed to enhance understanding of students. Section-II (Statistics) is divided into nine assignments. Each assignment in this section is to be done using R-software.

Instructions to the students

Please read the following instructions carefully and follow them.

1) Students are expected to carry this book every time they come to the lab for applied mathematics practical.
2) Students should prepare oneself beforehand for the Assignment by reading the relevant material.
3) Instructor will specify which problems to solve in the lab during the allotted slot and student should complete them and get verified by the instructor. However student should spend additional hours in Lab and at home to cover as many problems as possible given in this work book.

4) Students will be assessed for each exercise on a scale from0 to5.

Not done	0
Incomplete	1
Late Complete	2
Needs improvement	3
Complete	4
Well Done	5

Instruction to the Instructors

1) Explain the assignment and related concepts in around ten minutes using whiteboard if required or by demonstrating the software.
2) You should evaluate each assignment carried out by a student on a scale of 5 as specified above by ticking appropriate box.
3) The value should also be entered on assignment completion page of the respective Lab course.
4) Average marks of Section I and Section II, out of 15, is to be given as internal assessment marks.

Instructions to the Lab Administrator

You have to ensure appropriate hardware and software is made available to each student.

The operating system and software requirements on server side and also client side areas given below:

1. Server and ClientSide-(Operating System)Linux/Windows
2. R software/ R studio

Certificate

This is to certify that Mr./Ms._____has successfully completed the work for Lab Course- Applied Mathematics (Section I and Section II) and has scored _____Marks out of 15.

Instructor H.O.D. /Coordinator

Internal Examiner External Examiner

Table of Contents for Section-I

Assignment Completion Sheet

Lab Course I			
Mathematics Assignments			
Sr. No.	**Assignment Name**	**Marks (out of 5)**	**Teachers Sign**
1a			
1b			
2a			
2b			
3a			
3b			
4a			
4b			
5a			
5b			
Total (Out of 15)			

Section I - Mathematics

Assignment 1a :- Set Theory

1. Write power sets of A.

i. A={1,2}

ii. A={ø}

2. If A={a,b,c} and B={a,b,g,h,j,i}, find $A \cup B, A \cap B, A - B, A \oplus B, A \times B$.

3.Show that $A \times B \neq B \times A$.

4. If A={1,3,5,7,9} and B={2,4,6,8}, then

i. Describe sets A and B using set builder notation.

ii. Find cardinalities of $A \cup B, A \cap B, A - B, A \oplus B, A \times B$.

5. Prove the following identities using Venn diagram.

i. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

ii. $(\overline{A} \overline{\cap} \overline{B}) = \overline{A} \cup \overline{B}$

iii. $A \cap (B \cap C) = (A \cap B) \cap C$

6. Prove the identities given in problem 5 using membership table.

7. Prove that following sets are countable.

i. Z, ii. N U {-1,-2,-3}

Assignment Evaluation

0: Not Done [ ] 1: Incomplete [ ] 2: Late Complete [ ]

3: Needs Improvement [ ] 4: Complete [ ] 5: Well Done [ ]

Signature of Teacher

Assignment 2a :-Logic

1. Write the following propositions in symbolic form. What is the negation of each of these statements?

i. I will be happy if there is no pollution in the city.

ii. You will not be fined for over speeding if and only if your speed is less than 60 Km/hr.

iii. I will watch a football match on TV or I will study.

2. Write converse, contrapositive, and inverse of each of the statements given in question 1 given above.

3. Construct truth table for each of the following proposition.

i. $(p \vee -q) \rightarrow q$

ii. $(p \wedge -q) \leftrightarrow (-p \vee q)$

iii. $(p \wedge q) \rightarrow (p \vee q)$

4. Check whether the given propositions are logically equivalent.

i. $(p \vee q) \wedge (p \vee r) \equiv p \vee (q \wedge r)$

ii. $-(p \vee (-p \wedge q)) \equiv -p \wedge -q$

5. Test the validity of following arguments.

i. Either I clear my backlogs or I shall not be allowed to go to the next class. If I am not allowed to go to the next class, then I will have to leave my studies. However, I cleared my backlogs. Therefore, I won't have to leave my studies.

ii. If it rains then I wear a raincoat. If it shines then I do not need a sweater. Either it rains or shines. Moreover, I do need a sweater. Hence, I wear a raincoat.

iii. If Sachin runs for office, he will be promoted. If Sachin attends the meeting, then he will run for the office. Either Sachin attends a meeting or goes to Mumbai. But Sachin does not go to Mumbai. Hence, he is promoted.

Assignment Evaluation

0: Not Done [ ] 3: Needs Improvement [ ]

1: Incomplete [ ] 4: Complete [ ]

2: Late Complete [ ] 5: Well Done [ ] **Signature of the teacher**

Assignment 3a :-Relations

1. Check whether the given is an equivalence relation/ partial order.

i. Define a relation r on the set of real numbers as x is related y under the relation r if and only if $x \leq y$.

ii. Define a relation r on the set of integers as x is related y under the relation r if and only if xy is square of an integer.

iii. Define a relation r on the set of real numbers as x is related y under the relation r if and only if $x^2=y^2$.

2. Find matrix and digraph of the following relations.

i. A={1,2,3,4,6}, R={(x,y)/ y is a multiple of x}

ii. A={1,2,3,4,5}, S={(1,2),(3,4),(3,2),(4,5),(5,3),(1,5)}

3. Check whether the relations given in problem 2 are reflexive, symmetric, antisymmetric, transitive. Also find transitive closure of each of them using Warshall's algorithm.

4. Let A={1,2,3,4,5}. Consider two relations on A defined as R={(1,3), (5,2),(4,4)} and S={(1,2),(3,4), (3,2),(4,5), (5,3),(1,5)}. Find the inverse relations R^{-1}, S^{-1}. Also find the composite relations $R \circ S$ and $S \circ R$.

5.Let A={1,2,3,4,6,12}. Consider the partial order 'divisibility' on A i.e. $a \leq b$ if and only if a|b. Draw Hasse diagram for $(A,\leq)$.

6. Let A= set of divisors of 32. Consider the partial order 'divisibility' on A i.e. $a \leq b$ if and only if a|b. Draw Hasse diagram for $(A,\leq)$. Is it a chain?

7. Let X={a,b,c} and A=P(X), power set of X. Consider the partial order 'inclusion' on A. Draw Hasse diagram for $(A,\leq)$.

Assignment Evaluation

0: Not Done [ ] 3: Needs Improvement [ ]

1: Incomplete [ ] 4: Complete [ ]

2: Late Complete [ ] 5: Well Done [ ]

Signature of the teacher

Assignment 4a:-Functions

1. Check whether the given relation is a function.

i. f is a relation from set A to A, where A={1,2,3,4}. f={(1,2),(1,3),(2,4),(3,4)}.

ii. f: R -->R , $f(x)=x^2$.

iii. f: R--> R, f(x)= 2x-5

2. Check whether the given function is 1-1, onto function.

i. Let E be set of all even integers.

g:Z→E, g(x)=2x

ii. h:N→ E, h(x)= 2x

iii. s:N→E^+, s(x) = 2x

3. Find inverse of each of the function given in problems 1 and 2, if it exists.

4. Let f:R→R, f(x)=x+1, g: R→ R, g(x)= $\sqrt{x+1}$. Find composite functions f∘g and g∘f.

Assignment Evaluation

0: Not Done [ ] 3: Needs Improvement [ ]

1: Incomplete [ ] 4: Complete [ ]

2: Late Complete [ ] 5: Well Done [ ]

Signature of the teacher

Assignment 5a: Counting Principles

1. How many 5 letter words are there which begin and end with a vowel?

2.Find the number of binary strings of length 6 which begin or end with 1.

3. Find the number of 4-digit numbers which are even and have different digits.

4. Find the number of 4-digit numbers which are even and repetition of digits is allowed.

5.How many different arrangements of the alphabets of 'MISSISSIPPI' are there?

6. A committee of six members is to be formed such that there is at least one man and one woman representative in it. If there are 10 men and 15 women in a department, then how many ways are there to form the committee?

7. How many numbers greater than 5,00,000 can be formed by arranging the digits 2,3,4, 5,5,6 in a row?

8.If there are 30 students in a class, then show that at least two have their last names beginning with the same letter.

9. Leaflets are to be distributed in a group of 52 people. Prove that if 105 leaflets are distributed in this group, then there is at least one person who has received at least three leaflets.

10. Show that if any 5 numbers from 1 to 8 are chosen, then two of them will add to 9.

Assignment Evaluation

0: Not Done [ ] 3: Needs Improvement [ ]

1: Incomplete [ ] 4: Complete [ ]

2: Late Complete [ ] 5: Well Done [ ]

Signature of the teacher

Savitribai Phule Pune University

Section-II

F. Y. B. C. A.(Science)

SEMESTER I

BCA - 117

Applied Mathematics
Statistics Assignments

Table of Contents

Assignment Completion Sheet

Lab Course III			
Statistics Assignments			
Sr. No.	**Assignment Name**	**Marks (out of 5)**	**Instructor Sign**
1	Download and Install R, understand IDE		
2	Using R execute the basic commands, array, list and frames		
3	Create a Matrix using R and Perform the operations addition, inverse, transpose and multiplication operations		
4	Using R execute the statistical functions : mean, median, mode, quartiles, range, interquartile range and histogram		
5	Using R, drawing scatter Plots, Line Plots, Bar Plots		
6	Using R import the data from Excel/.CSV file and perform operations		
7	Using R import the data from Excel/.CSV file and Calculate the standard deviation		
8	Using R import the data from Excel/.CSV file and draw the skewness		
9	Import the data from Excel/.CSV and perform the Chi-squared Test		
Total (Out of 15)			

Section -II

Assignment No. 1

Objective: Download and Install R, understand IDE

Introduction to R

- **R is an open-source software, hence can be used by students, teachers and researchers for understanding statistics**
- R is a programming language for statistical calculation.It is a free software environment for statistical computing and graphics. It was initially developed by Ross Ihaka and Robert Gentleman in mid 1990's.
- R is an excellent software to use while learning statistics with the help of computer.
- It provides a coherent, flexible system for data analysis which can be extended as per users' requirement
- R is an integrated suit of software facilities for data manipulation, calculation and graphical display
- R as a statistical package
- **Features of R language software**
    - Effective data handling and storage facility.
    - A large collection of intermediate tools for data analysis
    - Excellent graphical facilities for analysis of data and to display the results directly on computer
    - R is a simple and effective programming language which includes conditional loops,user defined functions,Input/Output facilities
    - R has excellent in-built help system

- R has flexibility

Hence R is a useful software for interactive data analysis

- R and R Studio are both free, open-source software, available for commonly used operating systems, including Windows, macOS, and Linux systems. Regardless of your operating system, you should install R before installing R Studio. The home page of R **http://www.r-project.org contains more information about R and instructions for downloading .**

- **Detailed instruction for installing R**

How do I download R programming for Windows?
To Install **R**: Go to http://www.r-project.org/
Click on the download R link
Then select a location closest to you
Click on the "**Download R** for **Windows**" link at the top of the page.
Click on the "install **R** for the first time" link at the top of the page.
Click on base
Click "**Download R** for **Windows**" (In the link,the current version number appears after R) and save the executable file somewhere on your computer.
When the dialog box opens click Run the .exe file and a setup wizard should appear.
Keep clicking Next until the Wizard is finished.
Now you should see an icon on your desktop with a large capital R.

- Starting and Ending R

To start R, double click on R icon appearing on desktop

- This window gives useful information about R such as version of R including instruction for quiting R from the command line.
- When R starts it searches for any saved work in the current directory. If it finds some,that work will be reloaded and we are reminded that it was done.
- The symbol(>) appearing at the bottom is called prompt or R prompt.
- The window showing R prompt is known as R console window.
- At the R prompt> one can directly type commands into R console.
- To end R type q() at r prompt and it displays message box asking whether you want to save the current work or not. If yes then the commands given/executed in current R session are automatically stored so that we can continue our work from session to session.

- **Working of R**
  It is object oriented programming, flexible and simple language, all variables, data, functions, results etc are shared in the active memory of computer in the form of objects with operators and functions which are themselves objects. R can be used interactively. That means we can type any R command at its prompt and that command is executed directly. No need to write complete programming code like in C programming etc.

Note:1.Every command is to be issued at the R prompt(>)
2.Blank space before and after any operator is not needed.
3.More than one commands can be issued on the same line by separating them with (;)
4# symbol can be used for comment about any command

Getting Help in R

R has an excellent in built help facility for getting more information on any named function.Following are various ways of getting help:

- help(mean) # to get help on command mean or ?mean
- Help.start() This launches a web browser which allows the browsing of web pages with hyperlinks.
- Help.search("quantile")
- Example(median)

Gives examples on help topic

R commands and case sensitivity

It is an expression based language with simple syntax. It is a case sensitive language. For example a and A are taken as two different objects.

- **Normally all alpha-numeric symbols and period symbol(.) are allowed with a restriction that name cannot start with numeric digit**
- If an expression is given as a command it is evaluated printed and the value is lost.
  For example
  > 2+8
  [1] 10
  > a+5 #assume that a contains some value
  [1] 43.35052
  The answer is printed starting with[1].
- If an assignment is used as a command then it evaluates an expression and passes its value to a variable but the result is not automatically printed .To view the contents of a variable or a vector type its name
  > x=2+3
  > x
  [1] 5
  > x+2
  [1] 7
  > x=x+2
  > x
  [1] 7
- Note:If a command is not complete till the end of line, Rgives different prompt. By default + is displayed on second and subsequent lines and it continues to read input until the command is syntactically complete.

- R provides a mechanism for recalling and re-executing previous commands. The vertical arrow keys on keyboard↓ and↑ can be used to scroll forward and backward respectively through command history.
- Once the command is located in this way the cursor can be moved within command using horizontal arrow keys on keyboard→ ← to move right or to left characters of a particular command.
- Characters can be deleted by[Del] key or added with other keys as desired.

Accessing Data from External File(Importing data)

Method 1.Data stored in an external file created in Ms-Excel can be accessed in R-software by saving it as a CSV (comma delimited) form.Note that extension ".CSV" of a file is automatically attached to the file.

For example, there existsuch file created in Ms-Excel named input.CSV in current working directory. The contents of this file can be accessed in R using commandSf=read.csv("input.csv") #right click-properties-security-object name(path)-copy path in the bracket.

- Method2.Data stored in an external file(created In Ms-Excel) can be accessed in R software

  Save file created in Ms-Excel as "Text(tab delimited)" form

  File created in Ms-Excel is named def.txt in current working directory, then contents of this file can be accessed using command

  > s2=read.table("def.txt",header=T)

  The file def.txt contains column with headings in 1st row. Now s2 is a data frame containing two columns. Individual heights/weights of students can be accessed as given below:

  > s2$hgt

  [1] 12 23 34

  > s2$wgt

  [1] 50 60 80

  For storing the Heights and weights use the command

  > h=s2$hgt

  > h

  [1] 12 23 34

  > w=s2$wgt

  > w

  [1] 50 60 80

  Note: Command for importing file

  d=read.table(file.choose(),header =T,sep=",")

- **Data Types(Modes)**

  In analysis of statistical data we use different types of data. For manipulating such data R supports 3 data types called modes

  1.Logical: It is Boolean type data whose value can be true or false

  2. Numeric: It can be real or integer

  3. Complex: It consists of real and imaginary numbers

- The entities created and manipulated in R are known as objects. Objects may be a variable,array of numbers, character strings, functions or more general structures developed or built from one or more of these components
- Various objects are created, manipulated and are stored by various names.They can

be displayed using command

- >objects()
- [1] "a" "b12.3" "b13.2" "detr" "h" "m1" "m2" "m3" "R1.23" "R11" "r12" "R12" "r13" "R13" "r23" "s1" "s2" "s3" "w"

Or> ls()

- [1] "a" "b12.3" "b13.2" "detr" "h" "m1" "m2" "m3" "R1.23" "R11" "r12" "R12" "r13" "R13" "r23" "s1" "s2" "s3" "w"
- The collection of objects currently stored is called workspace.
- To remove any object from the work space use command.
- >rm(a,w) #objects named a and w are removed.
- All objects created during R session can be stored permanently in a file for their use in further R sessions.
- Store all the available objects in a file called with extension .R data in the current directory.
- When R is restarted again it reloads the workspace from this file also reloads the associated command history.

Set A Download and Install R,understand IDE

Set BAccessing Data from External File(Importing data)

Assignment Evaluation

0: Not Done [ ]

1: Incomplete [ ]

2: Late Complete [ ]

3: Needs Improvement []

4: Complete [ ]

5: WellDone [ ]

Signature of Teacher

Assignment no 2

Objective: Using R execute the basic commands, array, list and frames

Theory

- **Vectors and Vector Arithmetic**
  - R operates on named data structures. The simplest such structure is numeric vector.
    Numeric vector is a single entity consisting of an ordered collection of numbers.
  - **To create a vector use command**
    >x=c(2,3,4)

  This is an assignment using function c() which combines its arguments. The values are separated by comma.Once vectors are stored with their names, the contents of vector can be displayed by typing the name of vector

```
> x
     [1] 2 3 4
```

  [1] refers to the first observation.

- **Seq() Function**

  Many times it is required to generate vectors with elements in sequence. In such case sequence function is very useful.

```
> x=1:30  # x contains elements 1,2,...30
> x=seq(1,30)
> x=seq(2,15,2) # x contains element 2,4,...,14;here third argument 2 is size of increment
```

  Or named form

```
x=seq(from=2,to=6) #contains 2,3,4,5,6
> x=2:6
> x=seq(to=6,from=2)
>y=seq(-2,2,by=0.5)
> z=seq(0,1,length=11) #z contains 11 values between 0 and1 with same difference
>d=seq(5)
```

- **rep() function**

  It is used to generate vectors having repeated elements in different ways

```
>rep(2,5)  # creates a vector (2,2,2,2,2)
[1] 2 2 2 2 2
> rep(2,times=5)
[1] 2 2 2 2 2
> x=c(1,2)
> y=rep(x,2)
> y
[1] 1 2 1 2
```

- **c(Combine) function**

  Function c is used to create a vector of isolated elements. It is also used for combining one or more vectors and elements which are arguments

```
> x=1:5
> y=c(x,7,3) #y contains elements 1,2,...,5,7,3
> y=c(9,x)
```

> y=c(9,7,x,15,17)

- **Arithmetic operators**

  Existing vectors can be further manipulated using arithmetic operators +,-,*,/,^ to create new vectors

  >x=c(2,6,8,10)
  > y=x+2
  > z=1/x
  > w=y-z
  > y=x # vector x is assigned to vector y
  > y1=c(y,2,5)

- Vectors can be used in arithmetic expressions in which these operations are performed element by element.
- Vectors occurring in the same expression need not all be of same length.If they are not then the resulting value of expression is a vector with same length as the longest vector occurring in expression. Shorter vectors in the expression are recycled as needed until they match the length of longest vector

>x1=c(1,5,2,3,7,9)
> x2=c(1,2,3)
> x3=3*x1+x2
> x3
[1] 4 17 9 10 23 30

- **Numerical Functions**

  Various mathematical and statistical functions are available in R for arithmetic operations. Eg.log,sin,cos,sort,min,max,unique,range,length,var,prod,sum,cor,etc.

  > x=1:5

>y=rep(5,4);
>k=length(x);
>mx=max(x);
>m=mean(x);
>med=median(x); k;mx;m;med;
[1] 5
[1] 5
[1] 3
[1] 3

Vectors are mathematical objects. Standard arithmetic functions and operators apply to vectors on element basis

While applying simple arithmetic functions and operators to vectors proper care should be taken. If the operands are of different lengths then the shorter of the two is extended by repetition. However, if the length of longer is not multiple of length of shorter then warning message is displayed

>x=c(1,4,6,7);y=c(7,8);z=x+y;z;y1=c(3,5,6);z1=x+y1;
[1] 8 12 13 15
Warning message:In x + y1 : longer object length is not a multiple of shorter object length.

- **Accessing Vectors**

- Individual elements of a vector can be accessed by using indices

```
> x=c(1,6,3,8,2)
> x[1]   #first element of vector x is accessed
[1] 1
> x[2:4]
[1] 6 3 8
> x[c(1,3)]
[1] 1 3
> x[4:2]
[1] 8 3 6
> x[x>3]
[1] 6 8
> x[-1]
[1] 6 3 8 2
> x[-1:-3]   #elements except first to third are accessed
[1] 8 2
```

- **Alternative Way to Create Data Vectors**

  Vectors can be created and data can be entered alternatively by using scan function as follows

  ```
  >d1=scan()
  1:
  ```

  It means that enter first data item of vector d1.Then enter data one by one data items separated by<space>,press enter key,it accepts all previously entered elements and are stored in vector d1.

  ```
  > d1=scan()
  1: 2 3 4 5 6
  6:
  Read 5 items
  ```

- **Scan() function has many other arguments**

What-indicating type of data to be accepted, by default it is numeric, for character type data set use what='character'
nmax-indicating maximum number of elements to be accepted

```
>h=scan(what="character")
1: a b c d
5: <enter>
Read 4 items
> f=scan(nmax=3)
1: 1 2 3 4 5
Read 3 items
> f
[1] 1 2 3
```

- **Other Types of objects: Matrices**

  Matrix: A matrix of order mxn is an arrangement of m*n numbers in m rows and n columns

  R supports number of functions for creating and handling matrices.

  matrix() function # command

  It is used to create a matrix of desired order from given elements or from a vector

```
>x=1:6
```

```
> y=matrix(x,byrow=T,ncol=2)
> y #to display matrix y
     [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    5    6
```

This will create a matrix y of order 3 X 2 from vector x containing elements 1, 2,…,6

```
z=c(5,2,7,3,2,1,9,0,8)
> mz=matrix(z,nrow=3,ncol=3);mz
     [,1] [,2] [,3]
[1,]    5    3    9
[2,]    2    2    0
[3,]    7    1    8
```

Matrix of order 3 x 3

R provides number of inbuilt functions for handling them

nrow(),ncol(),dim(),rowSums(),colSums()

```
a=colSums(mz)
> a
[1] 14  6 17
```

- **Data Frames**

  A data frame is a list with class data frame and is similar to data set or data matrix in other statistical packages.It is a list of vectors and /or factors of the same length which are related across such that in the same position come from same experimental unit.Vector structures appearing as variables of the data frame must all have the same length and the matrix structures must all have the same row size.If the vectors included in the data frame are not of same length then vector having less elements is recycled whole number of times.

  Data frames can be created by using command data.frame

```
>x=c(5,7,9,2)
>y=1:4
>z=rep(2,4)
>w=seq(1,12,3)
>m=c(1,5)
>n=5:7
>d1=data.frame(x,y)
>d1
  x y
1 5 1
2 7 2
3 8 3
4 3 4
> d2=data.frame(m,w);d2
  m  w
1 1  1
2 5  4
3 1  7
4 5 10
```

```
> d3=data.frame(z,n);d3
Error in data.frame(z, n) :   arguments imply differing number of rows: 4, 3
```
That is no. of elements in z(ie.4) is not a multiple of number of elements in n(which is 3)

- **Commands for row name and column name of data frame**

```
d4=data.frame("First"=x,"Second"=y)
> d4
  First Second
1     5      1
2     7      2
3     8      3
4     3      4
d5=data.frame(z,w,row.names=c("r1","r2","r3","r4"))
> colnames(d5)=c("one","two")
> d5
   one two
r1   1   1
r2   1   4
r3   1   7
r4   1  10
```

- **Alternative Ways to Creating Data Frame**

  Data frame can be created using inbuilt data editor edit
  d1=edit(as.data.frame(NULL))
  This command opens data editor which is spread sheet like interface(as in Ms-Excel) To enter data click on var1,var2….and set the data type of each variable with its name and enter data in respective columns. After entering complete data close data editor. Data frame can be viewed as d1

```
d1=edit(as.data.frame(NULL))
> d1
  first second
1     1      4
2     2      5
3     3      6
```

Object Data frame d1 can be modified using command fix()

```
> fix(d1)
```

This command reopens data editor again and modifications can be made in d1

- **Creating data frames from external files**

  Large data objects are usually read as values from external files rather than entered during R session through keyboard.R input facilities are simple.For reading and writing in files,R uses working directory. **Such working directory can be viewed using getwd() command**
  Almost all applications used for handling data will export data as delimited file in ASCII text.This is convenient way to get large data in R.R can read data stored in ASCII text file using command read.table
  R can also read files in other format. For example Ms-Excel,SAS etc
  R command >read.table(file.choose())
  Attach()

Create a file in Ms-Excel and save it as (tab delimited) text file.

R can read such file using read.table command

s=read.table("b4.txt") #b4.txt file created in MsExcel

```
> s
  V1
1  x
2  1
3  2
4  3
```

read.table after reading data frame(here s) and variables in data frame will be named as V1,V2……

- **Accessing Data from a Data Frame**

  Once a data frame is created ,data from it can be accessed as follows:

```
 supply=c(120,140,210,100,90)
> demand=c(90,150,200,80,60)
> d=data.frame(supply,demand)
> d
  supply demand
1    120     90
2    140    150
3    210    200
4    100     80
5     90     60
> d$supply[3] #to access third element of vector supply in data frame
[1] 210
> d$demand  # to access vector demand in data frame d
[1]  90 150 200  80  60
> d[3,2]
[1] 200
```

- **attach()**

  If a data frame is placed in the systems search path using command attach() then it is not necessary to specify data frame names

```
>attach(d) # places data frame d in systems search path
The following objects are masked _by_ .GlobalEnv:
    demand, supply
> supply   # to access vector supply
[1] 120 140 210 100  90
```

Systems search path can be viewed using search command

```
> search()
 [1] ".GlobalEnv"        "d"                 "package:stats"
 [4] "package:graphics"  "package:grDevices" "package:utils"
 [7] "package:datasets"  "package:methods"   "Autoloads"
[10] "package:base"
```

d is the entry number 2 in search path;GlobalEnv is the workspace placed as entry number 1

Package base is the library where all standard functions are defined . It is placed at the end.

Note:There may be several objects of same name in different parts of the search path.In such case R searches in Global Env(current workspace) first and then in "d" and so on.Do not give variable name same as name of the data frame.

To view all objects currently in workspace use command

objects() #displays all objects

```
 [1] "a"      "b12.3"  "b13.2"  "d"      "demand" "detr"   "m1"     "m2"
 [9] "m3"     "R1.23"  "R11"    "r12"    "R12"    "r13"    "R13"    "r23"
[17] "s"      "s1"     "s2"     "s3"     "supply"
```

To remove data frame from systems search path command is detach

> detach(d) #to remove data frame d from systems search path

Note: Even if data frame is removed from search path, objects of that frame are still there in work space. Hence object() command will display those objects. To remove objects from work space use command

rm(object name)

objects()

```
 [1] "a"      "b12.3"  "b13.2"  "d"      "demand" "detr"   "m1"     "m2"
 [9] "m3"     "R1.23"  "R11"    "r12"    "R12"    "r13"    "R13"    "r23"
[17] "s"      "s1"     "s2"     "s3"     "supply"
```

> rm(d)

> objects()

```
 [1] "a"      "b12.3"  "b13.2"  "demand" "detr"   "m1"     "m2"     "m3"
 [9] "R1.23"  "R11"    "r12"    "R12"    "r13"    "R13"    "r23"    "s"
[17] "s1"     "s2"     "s3"     "supply"
```

- **<u>Subset and Transform Commands</u>**

- **Command subset is used for extracting data satisfying certain condition from existing data frame or vector**

Command transform is used to add new columns(variables) using columns of existing data frame.Also used to add new columns by giving column name in the command itself

```
>supply=c(120,140,210,100,90)
> demand=c(90,150,200,80,60)
> d=data.frame(supply,demand)
> d
  supply demand
1    120     90
2    140    150
3    210    200
4    100     80
5     90     60
> attach(d)
The following objects are masked _by_ .GlobalEnv:
    demand, supply
> d[1:3,1:2]
  supply demand
1    120     90
2    140    150
3    210    200
```

>**d1=subset(d,demand>100)**

```
> d1
  supply demand
2   140   150
3   210   200
```

- **>d2=transform(d,lack=supply-demand)**

```
> d2
  supply demand lack
1   120    90   30
2   140   150  -10
3   210   200   10
4   100    80   20
5    90    60   30
```

- **>x=c(12,67,32,54,80)**
- **> d3=transform(d2,"X"=x)**

```
> d3
  supply demand  lack  X
1   120    90   30  12
2   140   150  -10  67
3   210   200   10  32
4   100    80   20   54
5    90    60   30   80
> x=c(10,15,20,9,13,22)
> y=subset(x,x<15)
> y
[1] 10  9 13
```

  - **<u>Resident /InbuiltData Sets</u>**
    For accessing existing data sets,data command is used
    >data(data set name)
    >data(Orange) #to access data Orange
    >data() #to view all resident data set names from package base
    For example:Air Passengers Monthly Airline Passenger Numbers 1949-1960
- **To access data set from other packages use command**
- **Data(package=packages(all.available=TRUE))**
  To access data set from a particular package first attach the corresponding package by library commandLibrary(package name)

<u>Set A</u>

1.Create vectors x and y of observations 4,6,2,7,1,8 and4,8respectively.Also create vectors x+y,x/y and x*y.

2. Create a data frame containing state and average yield of wheat(kg) per acre using the following data.

State	Yield of wheat
Punjab	728
Harayana	1753
U.P	1475
Gujarat	1980
Bihar	2210
Karnataka	2270

Obtain the list of states with average yield less than 1500 kg per acre.

3.Create a vector x of following numbers.31,26,36,14,45,28,39,51,,33,34,21,11,12,35,20

(i) Create vector y containing elements numbered 3rd, 7th, 8th and 13th of vector x.

(ii) Create a vector z containing elementsof x >25.

Set B

4.Create two vectors 'a' and 'b' which include marks scored by students in five subjects out of 100 marks:

Marks of A: 65,89,75,95,89

Marks of B: 52,69,70,81,85

Create data frame 'student marks' with column heads as Marks of A, Marks of B

5.Create a vector age with following ages of 8 students:11,10,14,17,19,20,22,21

i) Find the number of students with age>15 and age<20.

ii) Augment age with adding 18,25 in the beginning and 21,24 at the end.

Assignment Evaluation

0: Not Done [ ]

1: Incomplete [ ]

2: Late Complete [ ]

3: Needs Improvement [ ]

4: Complete [ ]

5: WellDone [ ]

Signature of Teacher

Assignment no. 3

Objective: Create a Matrix using R and Perform the operations addition,inverse,transpose and multiplication operations

Theory

Matrix: It is a two dimensional array of numbers.

Order of a matrix: The number of rows and number of columns defines the order of the matrix.

Vectors: It are matrices with either one row(row vector) or one column (column vector).

Scalar: A scalar is a matrix with just one row and one column.

Square matrix: It is a matrix with same number of rows and columns.

Diagonal matrix: It is a square matrix with all non-diagonal elements equal to zero.

Identity matrix: It is a diagonal matrix with all diagonal elements equal to one.

J matrix: It is a matrix with all elements equal to one.

Null matrix: It is a matrix with all elements equal to zero.

Input of matrix in R

Method I: A matrix can be created by combining vectors of equal length using

iii) cbind( ) column bind command

iv) rbind() row bind command

v) direct

Ex.1 For following data create a matrix :

x	1	3	7	11	6
y	1	2	3	4	5

Soln.:

(i) **Using cbind command**

```
>x=c(1,3,7,11,6)
> y=c(1:5);
>A=cbind(x,y);A
```

output :

```
  x y
 [1,] 1 1
[2,] 3 2
```

```
[3,]  7  3
[4,] 11  4
[5,]  6  5
```

(ii) **Using direct method**

```
> A=matrix(c(1,1,3,2,7,3,11,4,6,5),byrow=T,ncol=2)
> A
     [,1] [,2]
[1,]    1    1
[2,]    3    2
[3,]    7    3
[4,]   11    4
[5,]    6    5
```

Input of a diagonal matrix

Ex.2 : Input a diagonal matrix of order 3x3 with diagonal elements 5,20,10

Soln.:> A=diag(c(5,20,10));A

```
     [,1] [,2] [,3]
[1,]    5    0    0
[2,]    0   20    0
[3,]    0    0   10
```

Ex.3: Create a identity matrix of order 2x2

Soln.:> A=diag(rep(1,2)); A

```
     [,1] [,2]
[1,]    1    0
[2,]    0    1
```

- **Matrix operations**

- **Transposition**

Ex.4: Obtain transpose of the matrix given in Ex.1

Soln.:

```
> A=matrix(c(1,1,3,2,7,3,11,4,6,5),byrow=T,ncol=2)
> A
     [,1] [,2]
[1,]    1    1
[2,]    3    2
[3,]    7    3
[4,]   11    4
[5,]    6    5
```

```
> R=t(A)        # command for transpose
> R
  [,1] [,2] [,3] [,4] [,5]
x   1    3    7   11    6
y   1    2    3    4    5
```

- **Scalar multiplication of a matrix**

Ex.5: Multiply the matrix given in Ex.1 by a scalar 5

Soln.:

```
> A=matrix(c(1,1,3,2,7,3,11,4,6,5),byrow=T,ncol=2)
> A
     [,1] [,2]
[1,]    1    1
[2,]    3    2
[3,]    7    3
[4,]   11    4
[5,]    6    5
> G=5*A;G                    # command
      x    y
[1,]  5  5
[2,] 15 10
[3,] 35 15
[4,] 55 20
[5,] 30 25
```

- **Addition of the matrices**

Let A and B be the two matrices of order rxc. Then the matrix A+B is a matrix of order rxc obtained by adding the corresponding elements.

Ex.6: $\mathbf{A}=\begin{bmatrix}2 & 4\\ 5 & 9\\ 3 & 8\end{bmatrix}$ $\mathbf{B}=\begin{bmatrix}3 & 4\\ 5 & 10\\ 39 & 8\end{bmatrix}$

Soln.:

```
>A=matrix(c(2,4,5,9,3,8),byrow=T,ncol=2);A;B=matrix(c(3,4,5,10,39,8),byrow=T,ncol=2);B
     [,1] [,2]
[1,]    2    4
[2,]    5    9
[3,]    3    8
>B=matrix(c(3,4,5,10,39,8),byrow=T,ncol=2);B

     [,1] [,2]
[1,]    3    4
[2,]    5   10
```

```
[3,]  39   8

> D=A+B;D                    # command for addition
   [,1] [,2]
[1,]   5   8
[2,]  10  19
[3,]  42  16
```

- **Multiplication of matrices**

  The two matrices are confirmable for multiplication if ,

  the number of columns in the first matrix = the number of rows in the second matrix.

Ex.7: Obtain product of following pairs of matrices.

$$\mathbf{A}=\begin{bmatrix}3 & 4\\ 5 & 10\\ 6 & 8\end{bmatrix} \quad \mathbf{B}=\begin{bmatrix}4\\ 9\end{bmatrix}$$

Soln.:

```
>A=matrix(c(3,4,5,10,6,8),byrow=T,ncol=2);A;B=matrix(c(4,9),byrow=T,ncol=1);B
   [,1] [,2]
[1,]   3   4
[2,]   5  10
[3,]   6   8
>B=matrix(c(4,9),byrow=T,ncol=1);B

   [,1]
[1,]   4
[2,]   9
> K=A%*%B;K                  #command for multiplication
   [,1]
[1,]  48
[2,] 110
[3,]  96
```

- **Determinant of a square matrix**

Ex.8: Find determinant of the following matrix given in Ex.1.

Soln.:

```
> A=matrix(c(8,20,-1,-16,3,55,9,-4,0),byrow=T,ncol=3);A;det(A)
   [,1] [,2] [,3]
[1,]   8   20   -1
[2,] -16    3   55
[3,]   9   -4    0
```

```
>det(A)                          #command
[1] 11623
```

- **Inverse of a square matrix**

Inverse of square matrix A of order nxn is a square matrix B of same order such that

AB=BA=I,where I is an identity matrix.

Likewise A is inverse of B.

Note: Inverse exists if determinant of the matrix is not equal to zero and when exists it is unique.

Ex.9: Find inverse of matrix

$$A=\begin{bmatrix} 7 & 18 & -2 \\ -16 & 3 & 55 \\ 9 & -4 & 0 \end{bmatrix}$$

Soln.:

```
>A=matrix(c(7,18,-2,-16,3,55,9,-4,0),byrow=T,ncol=3);A;
   [,1] [,2] [,3]
[1,]   7   18   -2
[2,]  -16   3   55
[3,]   9   -4   0
>B=solve(A);                              # command
>B

          [,1]                [,2]           [,3]
[1,] 0.021202776 0.000771010  0.09599075
[2,] 0.047706245 0.001734773 -0.03402082
[3,] 0.003565921 0.018311488  0.02978026

> A%*%B                                   # for verification
          [,1]                [,2]              [,3]
[1,] 1.000000e+00 -1.387779e-17 -9.020562e-17
[2,] 3.903128e-18  1.000000e+00  1.700029e-16
[3,] 2.775558e-17  5.898060e-17  1.000000e+00

> round(A%*%B,4)
   [,1] [,2] [,3]
[1,]   1   0   0
[2,]   0   1   0
[3,]   0   0   1
```

Exercise

1. Multiply the matrix by a scalar 10

$$\mathbf{A}=\begin{bmatrix} 7 & 18 & -2 \\ -16 & 3 & 55 \\ 9 & -4 & 0 \end{bmatrix}$$

2. Add

: $\mathbf{A}=\begin{bmatrix} 6 & 10 \\ 15 & 19 \\ 34 & 81 \end{bmatrix}$ $\mathbf{B}=\begin{bmatrix} 3 & 4 \\ 5 & 10 \\ 39 & 8 \end{bmatrix}$

3.Find Transpose

$$\mathbf{A}=\begin{bmatrix} 17 & 18 & -2 \\ -16 & 20 & 55 \\ 9 & -14 & 10 \end{bmatrix}$$

Set A

1. Create a matrix using R and obtain transpose of following matrix:

$$A=\begin{bmatrix} 7 & 18 & -2 & 11 \\ -18 & 3 & 33 & 1 \\ 9 & -4 & 0 & 21 \end{bmatrix}$$

2. Create a matrix using R and obtain product of following pairs of matrices:

$$A=\begin{bmatrix} 2 & 4 \\ 5 & 9 \\ 3 & 8 \end{bmatrix} \quad B=\begin{bmatrix} 4 \\ 9 \end{bmatrix}$$

1. Create a matrix using R and find the sum of the matrices A and B:

$$A=\begin{bmatrix} 12 & 4 \\ 5 & 8 \\ 23 & 18 \end{bmatrix} B=\begin{bmatrix} 32 & 14 \\ 5 & 7 \\ 3 & 8 \end{bmatrix}$$

Set B

3. Create a matrix using R and find inverse of a square matrix

$$=\begin{bmatrix} 7 & 18 & -2 \\ -16 & 3 & 55 \\ 9 & -4 & 0 \end{bmatrix}$$

4. i)Create a matrix with the cbind( ) function for the following data on :

Age(inyrs)	BMI	BP
60	28	124
72	21	127
45	35	140

ii) Multiply the matrix by a scalar 7.

Assignment Evaluation

0: Not Done [ ]

1: Incomplete [ ]

2: Late Complete [ ]

3: Needs Improvement [ ]

4: Complete [ ]

5: WellDone [ ]

Signature of Teacher

Assignment no.4

Objective: **Using R execute the statistical functions:**

mean, median, mode, quartiles, range, interquartile range and histogram

Theory:

Graphical Representation of data: Histogram

Histogram is commonly used to represent grouped frequency distribution graphically.It can distinguish clearly the class with maximum concentration of frequency. In histogram, classes are represented by series of adjacent rectangles with their area proportional to class frequency. Height of each rectangle is proportional to frequency of that class ,if classes are of uniform width.In case of classes of unequal width height of rectangle is proportional to frequency density .

R command: hist( )

Example1:Monthly sales(in'00Rs) of 10 small shops are given below:

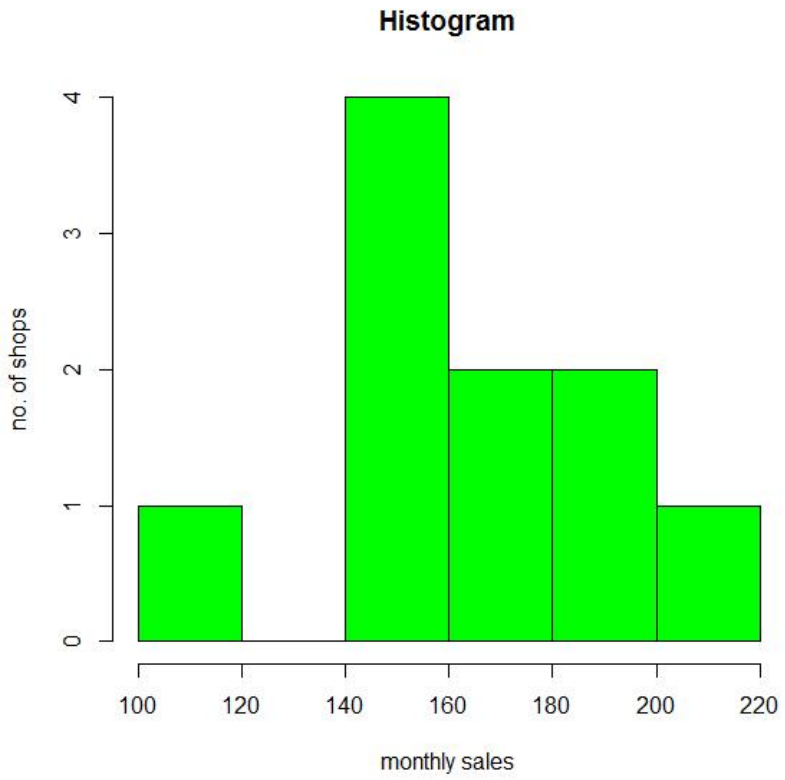
100,190,210,150,170,160,160,180,200,152

i) Draw histogram

ii) Calculate arithmetic mean,Geometric mean,harmonic mean,mode,median,lower quartile, upper quartile,third decile and 42nd percentile,range,interquartile range

Soln.:Histogram:

```
> hist(x,col="green",xlab="monthly sales",ylab="no. of shops",main="Histogram")    # command
```



Interpretation: concentration is observed in the class 140-160.

Central tendency

It is the tendency of the observations to concentrate in central part of the data.Measure of central tendency is a single number which is considered as a representative of the data.

Here we consider following types of measure of central tendency:

i. Arithmetic mean(A.M) or mean
ii. Geometric mean(G.M)
iii. Harmonic mean(H.M)
iv. Median
v. Mode

Computation of mean, geometric mean, harmonic mean, median, mode and partition values for raw data.

Suppose $x_1,x_2,...,x_n$ are n given observations on variable X, then we use the following formulae to calculate different measures of central tendency.

Mean=$\frac{\sum xi}{n}$

G.M=Antilog $(\frac{\sum log x_i}{n})$

H.M =n/ $(\sum \frac{1}{x_i})$

Median: It is the value of middlemost observation when the data is arranged either in ascending or descending order of magnitude.

Mode is that value of the variable which is predominant in the series.

Partition values are those values which divides the arranged data in some number of equal parts. Commonly used partition values are quartiles, deciles and percentiles which divides the arrayed data in four,ten and hundred equal parts respectively.

Computation of mean, geometric mean, harmonic mean, median, mode and partition values for raw data given in example 1.

```
>x=c(100,190,210,150,170,160,160,180,200,152)    # input data by using combine function
> n=length(x)                                     # number of observations
> AM=mean(x)                                      # command  for arithmetic mean
AM
[1] 167.2
>cat("Mean=",AM,"n/");
Mean= 167.2

>lx=log10(x);GM=10^mean(lx);                      #command for geometric mean
>GM
[1] 164.2067

>HM=n/sum(1/x);                                   #command for harmonic mean
>HM
[1] 160.7495

>med=median(x);                                   #command for median
>med
 [1] 165
> q1=quantile(x,0.25);                            # command for quartile
```

```
>q1
25%
154
>q3=quantile(x,0.75);q3;                                  # command for quartile

 75%
187.5
>d3=quantile(x,0.3);d3;                                  # command for decile
 30%
157.6
> P42=quantile(x,0.42);P42;                                  #command  for Percentile
42%
160
>tx=table(x);m=which(tx==max(tx));stx=sort(unique(x));    # command for mode
> mo=stx[m];mo;
[1] 160
```

Dispersion

Average condenses the data into a single value. It is observed that studying only average is not sufficient to describe the distribution completely. Hence for further study and analysis with mean it is essential to measure the extent of variation present in a given data. Observations are dispersed from the central value. This extent of variation is called as dispersion.

Measures of Dispersion

We shall consider the computation offollowing measures of dispersion for raw datagiven in example 1:

(iii) Range=Max.(x)-Min.(x)

```
>xmx=max(x);xmx;                                         #command for maximum value
[1] 210
>xmi=min(x);                                             #command for minimum value

[1] 110
>Range=xmx-xmi;Range;                                    #command for range
 [1]100
```

(ii) Quartile deviation

```
>QD=(q3-q1)/2;QD                                         # command for Quartile deviation
 75%
[1]16.75
```

(iii) **Standard deviation**

```
>v1=var(x);                                              #command for variance
>v=(n-1)*v1/n;v
[1] 874.56

>sd=v^0.5;sd                                             #command for standard deviation
[1] 29.57296
```

Exercise

1.Calculate the mean,mode and median from the following data of the heights in inches of a group of students:

61,62,63,61,63,64,60,65,64,64,64,66,64

Now suppose that a group of students whose heights are 60,96,59,68,67 and 70 inches,is added to the original group. Find the mean, mode and median of the combined group. Also find standard deviation of the combined data.

2.Following are the prices(in Rs.) of shares of a company on six consecutive days. Compute range and quartile deviation.

200,210,208,160,220,250

Set A

1. Following is the size of the 10 consecutive shirts sold in a departmental store. Find the mean size and modal size.

38, 40, 36, 40, 40, 38, 42, 44, 40, 42.

2.Calculate lower quartile,median ,upper quartile,8^{th} decile,45^{th} percentile and quartile deviation for the following observations :

51, 52, 53, 51, 53, 54, 54, 50, 55, 53, 55, 56

3.Construct histogram to represent the data on weekly TV viewing times, in hour for 20 people given below :

25	41	27	32	43	66	35	31	15	5
34	26	32	38	16	30	38	30	20	21

Set B

4.Represent the following data on the lengths of stay in short term hospitals by 21 randomly selected patients using histogram.

4, 4, 12, 18, 9, 6, 12, 3, 6, 15, 7, 3, 55, 1, 10, 13, 5, 7, 1, 23, 9.

5.Find mean,median,mode and 3^{rd} quartile for the following data :

13.2, 13.3, 11.9, 15.7, 11.3, 12.2, 16.7, 10.7, 3.3, 13.6, 14.8, 9.6, 11.6, 8.7, 15.0.

Assignment Evaluation

0: Not Done [ ]

1: Incomplete [ ]

2: Late Complete [ ]

3: Needs Improvement [ ]

4: Complete [ ]

5: WellDone [ ]

Signature of Teacher

Assignment no.5

Objective:Using R, drawing scatter Plots,Line Plots, Bar Plots

Theory

- **Diagrammatic Representation of Categorical data**

Statistical data can be represented in the form of a diagram.Following are the commonly used diagrams used to represent categorical data.

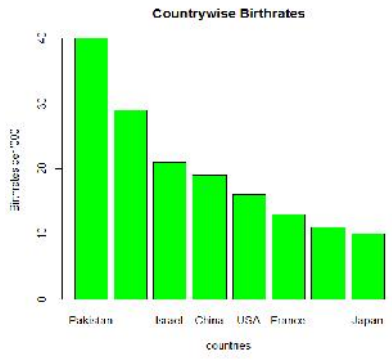
1. Simple Bar diagram
   It is the simplest type of diagram which is used to represent the data related to a single variable. For ex. used to present data like population of different cities etc.
2. Subdivided Bar diagram
   When a single variable is subdivided into two or more components, subdivided bar diagram is used. A bar representing the total value is divided into several components. Each component occupies a part of the bar proportional to its share in the total.The parts are displayed with different shadings ,symbols or colors.
3. Multiple Bar diagram
   It is used for presenting more than one components using multiple bars adjacent to each other. It can be used to represent data on imports and exports of a region for several years.

Example 1: Represent the following information by an appropriate diagram
i)

Country	Pakistan	India	Israel	China	USA	France	Russia	Japan
Birth rates(per'000)	40	29	21	19	16	13	11	10

Soln.: Simple Bar diagram **R code**

```
y=c("Pakistan","India","Israel","China","USA","France","Russia","Japan")
bir=c(40,29,21,19,16,13,11,10)
barplot(bir,xlab="countries",ylab="Birthrates per '000",main="countrywise Birthrates",
col="green",names.arg=y)
```

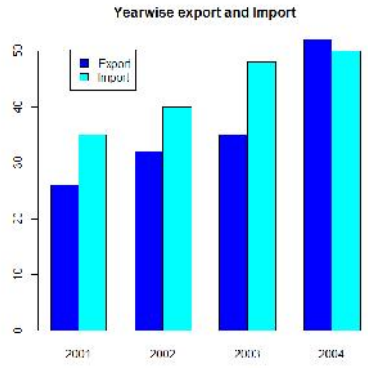


ii)Data on value of exports and imports (in crores of Rs.) are given below:

Year	Export	Import
2001	26	35
2002	32	40
2003	35	48
2004	52	50

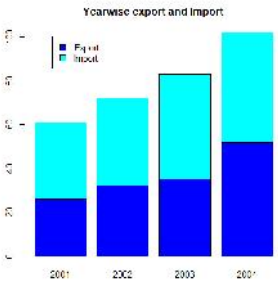
Soln.: Multiple Bar diagram **R-code**

```
y=2001:2004
e=c(26,32,35,52)
i=c(35,40,48,50)
d1=cbind(e,i)
barplot(t(d1),beside=T,main="Yearwise export and Import",col=4:5,names.arg=y)
legend(locator(1),legend=c("Export","Import"),fill=4:5)
```



Soln.: Subdivided Bar diagram **R-code**

```
y=2001:2004
e=c(26,32,35,52)
i=c(35,40,48,50)
d1=cbind(e,i)
barplot(t(d1),main="Yearwise export and Import",col=4:5,names.arg=y)
legend(locator(1),legend=c("Export","Import"),fill=4:5)
```



- **Correlation**

In many situations,we may be interested in observing two concerned variables for every unit in a sample. For example, age of a person and Blood pressure of a person, income and expenditure of a family etc. In such cases we may be interested in studying the relationship between the two

variables. Such interrelated variables are called correlated variables. The extent of linear relationship between them is called as correlation.To study correlation we require data on two concerned variable. Such data is called as bivariate data. Two variables are said to be correlated if the change in the value of one variable is accompanied by proportionate change in the value of other variable.

Types of correlation:

i) Positive correlation

ii) Negative correlation

In positive correlation the change is in the same direction whereas in negative correlation changes are in opposite direction.

Correlation can be studied by following methods:(i) Scatter diagram(ii) Prof. Karl Pearson's coefficient of correlation. Scatter diagram is a simple method of diagrammatic representation of a bivariate data for ascertaining the degree and type of correlation between the variables. Pearsonian coefficient of correlation is a numerical measure of degree of correlation present between two variables.

- **Regression**

In bivariate data, if the two variables are correlated then we express these relationship in a mathematical form which can be used for predicting the value of one variable given the value of other variable. This technique of prediction on the basis of correlation is called as regression analysis. Depending on the type of relationship between two variables, regression can be linear or non-linear.

In linear regression, the relationship between two variables is of the form Y=a+bx,where Y is dependent variable and X is independent variable. This is called as line of regression of dependent variable Y on independent variable X.

Example2: For the following data on heights(in inches) of fathers(X) and their sons(Y):

X	65	66	67	67	68	69	70	72
Y	67	68	65	68	72	72	69	71

Soln.:**R-code**

```
x=c(65,66,67,67,68,69,70,72)
y=c(67,68,65,68,72,72,69,71)
plot(x,y,main="Scatter plot",xlab="Heights(in inches)of father",ylab="Heights(in inches)of
sons") #command
d=data.frame(x,y)
r=cor(d);r;                                  # command for correlation
lr=lm(y~x);                                  #command for regression
sr=summary(lr);sr; abline(lr)
```

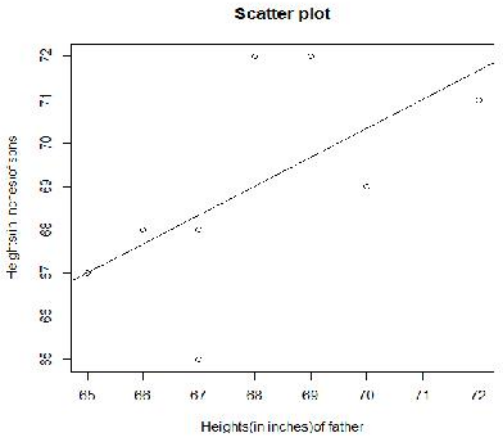
```
Output:       x         y
x 1.0000000 0.6030227
y 0.6030227 1.0000000

Call:
lm(formula = y ~ x)
```

Residuals:
   Min   1Q Median   3Q   Max
-3.3333 -0.8333 -0.1667 0.8333 3.0000

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 23.6667   24.4947   0.966   0.371
x   0.6667   0.3600   1.852   0.114

Residual standard error: 2.16 on 6 degrees of freedom
Multiple R-squared: 0.3636,   Adjusted R-squared: 0.2576
F-statistic: 3.429 on 1 and 6 DF, p-value: 0.1135



Set A

1. The following table shows state-wise details of kharif crop sown and damaged by draught during 2004 :

State	**Area sown (in lakhs of hectares)**	**Area damaged (in lakhs of hectares)**
Punjab	92	17
Madhya pradesh	78	19
Uttar pradesh	61	28
Jharkhand	20	4
Rajasthan	49	45
Karnataka	37	13

Represent the above data using a subdivided bar diagram.

2.The following data give the marks obtained by 10 students in economics and statistics. Draw scatter plot and interpret it.

Economics :	65	70	45	30	40	90	50	85	60	75
Statistics :	70	90	35	40	40	95	60	80	50	80

Set B

3.The following table shows number of farmers insured between Rabi 1999 – 2000 and Rabi 2002 – 03 (7 seasons) of different states.

States	**Number of farmers insured (lakhs)**
Andhra Pradesh	56
Chattisgarh	16
Gujarat	36
Karnataka	21
West Bengal	16
Maharashtra	79

Represent the above data by a suitable bar diagram.

4. From the following information relating to the stock. Exchange Quotations for two shares A and B, ascertain, by using scatter plot, whether shares A and B are correlated, in their prices :

Price share (A) Rs. :	160	164	172	182	166	170	178
Price share (B) Rs. :	292	280	260	234	266	254	230

Assignment Evaluation

0: Not Done [ ] 1: Incomplete [ ] 2: Late Complete [ ]

3: Needs Improvement [ ] 4: Complete [ ] 5: WellDone [ ]

Signature of Teacher

Assignment no.:6

Objective: Using R import the data from Excel/.CSV file and perform operations

Set A :

1. Using R import the following data on marks obtained by a student from Excel/.CSV file and Create two vectors 'a' and 'b' which include marks scored by students in five subjects out of 100 marks

Marks of A: 65,89,75,95,89

Marks of B: 52,69,70,81,85

i) Create two vectors 'a' and 'b' which include marks scored by students in five subjects out of 100 marks

ii)Create data frame 'student marks' with column heads as Marks of A, Marks of B

Set B:

2.Using R import the following data of ages of 8 students from Excel/.CSV file and perform operations:11,10,14,17,19,20,22,21

i) Create a vector 'age'

ii) Find the number of students with age>15 and age<20.

iii) Augment age with adding 18,25 in the beginning and 21,24 at the end.

Assignment Evaluation

0: Not Done [ ]

1: Incomplete [ ]

2: Late Complete [ ]

3: Needs Improvement [ ]

4: Complete [ ]

5: WellDone [ ]

Signature of Teacher

Assignment no.:7

Objective: Using R import the data from Excel/.CSV file and calculate the standard deviation

Theory

Dispersion

Average condenses the data into a single value. It is observed that studying only average is not sufficient to describe the distribution completely. Hence for further study and analysis with mean it is essential to measure the extent of variation present in a given data. Observations are dispersed from the central value. This extent of variation is called as dispersion.

Standard deviation: It is regarded as the best and most powerful absolute measure of dispersion. It is defined as positive square root of arithmetic mean of squares of the deviations taken from arithmetic mean. It is denoted by Greek letter σ and abbreviated as s.d or S.D.

σ =S.D.=Sqrt(variance)=$\sqrt{\left(\frac{\sum(xi-\bar{x})^2}{n}\right)}$

$=\sqrt{(\frac{1}{n}\sum xi^2)-\bar{x}^2}$

Based on standard deviation, a relative measure of dispersion known as coefficient of variation(C.V.) is used for comparison of variability of two or more data sets. It is given as,

C.V=$\frac{S.D}{|\bar{x}|}$*100%

The series which has less CV is said to be more consistent than other.

Accessing Data from External File(Importing data)

Data stored in an external file created in Ms-Excel can be accessed in R-software by saving it as a CSV (comma delimited) form.Note that extension ".CSV" of a file is automatically attached to the file.

For example, there exist such file created in Ms-Excel named input.CSV in current working directory. The contents of this file can be accessed in R using command

Sf=read.csv("input.csv") # right click-properties-security-object name(path)-copy path in the bracket

Standard deviation

```
>v1=var(x);                                  #command for variance
>v=(n-1)*v1/n;v
>sd=v^0.5;sd                                 #command for standard deviation
```

Set A

1.Using R import the data from Excel/.CSV file and calculate the standard deviation for the following observations :

51, 52, 53, 51, 53, 54, 54, 50, 55, 53, 55, 56

Set B

2.Using R import the following data on the lengths of stay in short term hospitals by 21 randomly selected patients from Excel/.CSV file and calculate the standard deviation

4, 4, 12, 18, 9, 6, 12, 3, 6, 15, 7, 3, 55, 1, 10, 13, 5, 7, 1, 23, 9.

Assignment Evaluation

0: Not Done [ ] 1: Incomplete [ ] 2: Late Complete [ ]

3: Needs Improvement [ ] 4: Complete [ ] 5: WellDone [ ]

Signature of Teacher

Assignment no.:8

Objective: Using R import the data from Excel/.CSV file and draw the skewness

Theory

Accessing Data from External File(Importing data)

Data stored in an external file created in Ms-Excel can be accessed in R-software by saving it as a CSV (comma delimited) form.Note that extension ".CSV" of a file is automatically attached to the file.
For example, there exist such file created in Ms-Excel named input.CSV in current working directory. The contents of this file can be accessed in R using command
Sf=read.csv("input.csv") # right click-properties-security-object name(path)-copy path in the bracket

Skewness

The measures of central tendency give us the representative value of a series. Measure of dispersion give us idea about scatteredness in the observation about the central value but it fails to reveal whether the scatter of values on either side of mean is symmetrical or not. These measures do not describe the data completely because there may be two distributions with same mean,variance but still differ from each other regarding shape or pattern. Skewness is the property which describes lack of symmetry in a distribution. A distribution may be skewed or symmetric. There are two types of skewness (i) Positive skewness(ii) Negative skewness. Following are the relative measures of skewness:

(i) Karl Pearson's coefficient of skewness

$$Skp=\frac{(Mean-Mode)}{Standard\ deviation}$$

(ii) Bowley's coefficient of skewness

$Skb=(Q_3+Q_1-2Q_2)/(Q_3-Q_1)$

(iii) Pearsonian coefficient of skewness based on moments

γ_1=sqrt(β_1) where sign of γ_1 is same as μ_3

Example: Using R import the data from Excel/.CSV file and draw the skewness

for the following observations. 10,12,14,17,8,22,19,12,20,30,25,27,12

Find (i) Karl Pearson's coefficient of skewness
(ii)Bowley's coefficient of skewness
(iii)Pearsonian coefficient of skewness based on moments

Soln.:

```
x=c(10,12,14,17,8,22,19,12,20,30,25,27,12);x
n=length(x);
n=10
 [1] 10 12 14 17  8 22 19 12 20 30 25 27 12

> tx=table(x);m=which(tx==max(tx));stx=sort(unique(x));mo=stx[m];mo
[1] 12
> mx=mean(x);mx;mex=median(x);mex;v1=var(x);v=((n-1)*v1)/n;v;sd=sqrt(v);sd;
[1] 17.53846
[1] 17
[1] 47.61495
[1] 6.900359
```

```
> skp=(mx-mo)/sd;skp;q1=quantile(x,0.25);q1;q3=quantile(x,0.75);q3;q2=quantile(x,0.5);q2
[1] 0.8026339
25%
 12
75%
 22
50%
 17
> skb=(q3+q1-2*q2)/(q3-q1);skb;
75%
 0
> cmx3=sum((x-mx)^3)/n;cmx3;beta1=cmx3^2/v^3;beta1;g1=sqrt(beta1);g1
[1] 24.64507
[1] 0.005626394
[1] 0.07500929
> if(cmx3<0){cmx3=-cmx3};cmx3;
[1] 24.64507
>
> cat("Karl Pearson's coefficient of skewness=",skp,"\n")
Karl Pearson's coefficient of skewness= 0.8026339
> cat("Bowley's coefficient of skewness=",skb,"\n")
Bowley's coefficient of skewness= 0
> cat("Karl Pearson's coefficient of skewness based on moments=",g1,"\n")
Karl Pearson's coefficient of skewness based on moments= 0.07500929

> if(cmx3<0){g1=-g1};g1;
[1] 0.07500929
```

Interpretation : Distribution has positive skewness.

Set A:1.Using R import the data from Excel/.CSV file and calculate the standard deviation for the following observations :

51, 52, 53, 51, 53, 54, 54, 50, 55, 53, 55, 56

Set B

2.Using R import the following data on the lengths of stay in short term hospitals by 21 randomly selected patients from Excel/.CSV file and calculate the standard deviation

4, 4, 12, 18, 9, 6, 12, 3, 6, 15, 7, 3, 55, 1, 10, 13, 5, 7, 1, 23, 9.

Assignment Evaluation

0: Not Done [ ] 1: Incomplete [ ] 2: Late Complete [ ]

3: Needs Improvement [ ] 4: Complete [ ] 5: WellDone [ ]

Signature of Teacher

Assignment no. 9

Objective: Import the data from Excel/.CSV and perform the Chi-squared Test Theory

- **Chi-square test for independence of two attributes A and B**

To test H_0: Attributes A and B are independent against H_1: Attributes A and B are not independent.

Test statistic: Under H_0 ,the test statistic is given by

$$\chi^2=\sum\sum\frac{Oij^2}{Eij}- N \sim\chi^2_{(m-1)(n-1)}$$

Where Oij= Observed frequency

Eij= Expected frequency

N= Total of observed frequencies

m= No. of classes of attribute A

n= No. of classes of attribute B

Test procedure:

Reject H_0, if $\chi^2 \geq \chi^2_{(m-1)(n-1)};\alpha$ otherwise accept H_0.

R command for performing the test

chisq.test(y, conf.level=c, correct=F) where y is mxn matrix of given contingency table. C is the confidence coefficient (1-α) for level of significance α.If c is not specified it is taken as 0.95.

Argument correct=F indicates that Yates correction is not to be applied.If any of the cell frequency in contingency table is less than 5 we apply Yates correction by specifying correct=T

Example1: For the data in the following table,test for independence between a person's ability in Mathematics and interest in Economics:

Interest in Economics	Ability in Mathematics		
	Low	Average	High
Low	63	42	15
Average	58	61	31
High	14	47	29

Soln.:

R code

```
>x=scan()
1: 63 42 15 58 61 31 14 47 29
10:
```

Read 9 items
>m=3;n=3
>y=matrix(x,byrow=T,ncol=3)
>cat("Ho:Two attributes are independent\n")
>cat("H1:Two attributes are not independent\n")
>ct=chisq.test(y,correct=F);ct

Output

Pearson's Chi-squared test
data: y
X-squared = 32.1399, df = 4, p-value = 1.791e-06
Conclusion : Reject H0,as p-value is less than 0.05.

- **Chi-square test for goodness of fit**

To test H_0: Fit is good against H_1: Fit is not good
. Test statistic: Under H_0 ,the test statistic is given by

$$\chi^2=\sum \frac{Oi^2}{Ei}-N\sim\chi^2_{k-p-1}$$

Where k=No. of classes after pooling
P= No. of parameters estimated
Test procedure:
Reject H_0, if $\chi^2 \geq \chi^2_{(k-p-1);\alpha}$ otherwise accept H_0.

Example 2: A die is tossed 60 times and the following results are obtained

No turned up : 1 2 3 4 5 6
Frequency : 8 7 12 8 14 11
Test the hypothesis that the die is unbiased.
Soln.:

R code

>f=c(8,7,12,8,14,11)
>ct=chisq.test(f,p=rep(1/6,6));ct

Output:

Chi-squared test for given probabilities

data: f
X-squared = 3.8, df = 5, p-value = 0.5786
Conclusion: Accept H0 as p-value >0.05

Set A:

1. The following table gives the number of aircraft accidents that occurred during the seven days of the week. Find whether the accidents are uniformly distributed over the week.

Days	Mon.	Tue.	Wed	Thurs	Fri	Sat	Sun.
No of accidents	14	18	12	11	15	14	84

Set B:

2. The following data is collected on two characters:

	Cinegoers	Non-cinegoers
Literate	83	57
Illiterate	45	68

Based on this,can you conclude that there is no relation between the habit of cinema going and literacy?

Assignment Evaluation

0: Not Done [ ] 1: Incomplete [ ] 2: Late Complete [ ]

3: Needs Improvement [ ] 4: Complete [ ] 5: WellDone [ ]

Signature of Teacher