JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR
COLLEGE OF ENGINEERING (AUTONOMOUS) :: ANANTHAPURAMU

ACADEMIC CALENDAR 2024-2025

B. Tech II Year I & II Semesters

(for 2023 Admitted Batch)

I Semester		
Commencement of Class Work	08-07-2024 (Monday)	
Instruction Period for the Semester	08-07-2024 to 09-11-2024	18 Weeks
Examinations		
I Mid-term Examinations	04-09-2024 to 06-09-2024	03 Days
II Mid-term Examinations	11-11-2024 to 13-11-2024	03 Days
End Laboratory Examinations	18-11-2024 to 23-11-2024	01Week
End Theory Examinations	25-11-2024 to 07-12-2024	02 Weeks
Commencement of Class Work for B. Tech II Year II Semester	09-12-2024 (Monday)	
Declaration of Results for II B. Tech. I-Semester	23-12-2024	

II Semester		
Commencement of Class Work	09-12-2024 (Monday)	
Instruction Period for the Semester	09-12-2024 to 12-04-2025	18 Weeks
Examinations		
I Mid -term Examinations	05-02-2025 to 07-02-2025	03 Days
II Mid-term Examinations	15-04-2025 to 17-04-2025	03 Days
End Laboratory Examinations	21-04-2025 to 26-04-2025	01Week
End Theory Examinations	28-04-2025 to 09-05-2025	02 Weeks
Community Service Project (CSP)	12-05-2025 to12-07-2025	02 Months
Commencement of Class Work for B. Tech III Year I Semester	14-07-2025 (Monday)	
Declaration of Results for II B. Tech. II-Semester	26-05-2025(Monday)	

Date: 29-06-2024

PRINCIPAL

Copy to:

a. Vice -Principal
b. All Heads of the departments
c. Placement Officer
d. College Office
e. College Hostel Office
f. Notice Board (Academic Section)
g. File

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, ANANTAPUR
COLLEGE OF ENGINEERING (*Autonomous*), ANANTAPUR
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
B.TECH II YEAR II SEM
TIME TABLE

Academic year: 2023-24 **Commences from: 20/12/2023 To 09/04/2024** **Branch: CSE**
Class Room: LHC-209 **CRC Chairman: Dr. P. Radhika Raju**

DAY/ PERIOD	1 9.30-10.30	2 10.30-11.30	3 11.30-12.30	12.30-1.30	4 1.30-2.30	5 2.30-3.30	6 3.30-4.30
MON	←--------SC-II LAB--------→		SC-II	L U N C H	←----------------DBMS LAB----------------→		
TUE	PSM	PSM	CO		DBMS	OS	DTI
WED	PSM	CO	SE		←------------------SE LAB------------------→		
THUR	←----------------OS LAB---------------→				DBMS	OS	SE
FRI	DTI	CO	SE		DBMS	OS	DTI
SAT	GUEST LECTURE/ WORKSHOP/WEBINAR	MENTORING	NSS/NCC		GUEST LECTURE/WORKSHOP/WEBINAR	LIBRARY	MENTORING

S.NO	SUBJECT	ABBR	STAFF
1	Probability and Statistics Methods	PSM	Ms. S. Srilakshmi
2	Computer Organization	CO	Prof. K. Madhavi
3	Database Management Systems	DBMS	Dr. P. Radhika Raju
4	Operating Systems	OS	Prof. B. Eswar Reddy
5	Software Engineering	SE	Ms. R. Jahnavi
6	Database Management Systems Lab	DBMS LAB	Dr. P. Radhika Raju Mr. K. Surendra Mr. N. Kiran Kumar
7	Operating Systems Lab	OS LAB	Prof. B. Eswar Reddy Mr. G. Pradeep Reddy Mr. K. Varun Kumar
8	Software Engineering Lab	SE LAB	Ms. R. Jahnavi Ram Mr. G. Eswar Ms. J. Sivarani
9	Skill oriented Course – II Exploratory Data Analytics with R	SC-II	Smt. D. Madhuri
10	Skill oriented Course – II Exploratory Data Analytics with R Lab	SC-II LAB	Prof. K. Madhavi Smt. D. Madhuri Mr. M. Vijaya Kanth
11	Mandatory non-credit Course - III (Design Thinking for Innovation)	DTI	Mr. T. Sudhakar

R Bharati

HEAD OF THE DEPARTMENT **VICE-PRINCIPAL** **PRINCIPAL**

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY ANANTAPUR
COLLEGE OF ENGINEERING (AUTONOMOUS) :: ANANTHAPURAMU

ACADEMIC CALENDAR 2023-2024
B.Tech II Year I & II Semesters
(for 2022 Admitted Batch)

I Semester		
Commencement of Class Work	14-08-2023 (Monday)	
Instruction period for the Semester	14-08-2023 to 02-12-2023	16 Weeks
Examinations		
I Mid-term Examinations	04-10-2023 to 06-10-2023	03 Days
II Mid-term Examinations	30-11-2023 to 02-12-2023	03 Days
End Laboratory Examinations	04-12-2023 to 07-12-2023	04 Days
End Theory Examinations	08-12-2023 to 19-12-2023	02 Weeks
Commencement of Class Work for B.Tech II Year II Semester for the Academic year 2023–24	20-12-2023 (Wednesday)	
Declaration of Results for II B.Tech. I-Semester	10-01-2024	

II Semester		
Commencement of Class Work	20-12-2023 (Wednesday)	
Instruction period for the Semester	20-12-2023 to 09-04-2024	16 Weeks
Examinations		
I Mid-term Examinations	07-02-2024 to 09-02-2024	03 Days
II Mid-term Examinations	06-04-2024 to 09-04-2024	04 Days
End Laboratory Examinations	10-04-2024 to 13-04-2024	04 Days
End Theory Examinations	15-04-2024 to 27-04-2024	02 Weeks
Community Service Project (CSP)	29-04-2024 to 22-06-2024	08 Weeks
Commencement of Class Work for B.Tech III Year I Semester for the Academic year 2024–25	24-06-2024 (Monday)	
Declaration of Results for II B.Tech. II-Semester	18-05-2024	

Date: 09-08-2023

PRINCIPAL

Copy to:
a. Vice -Principal
b. All Heads of the departments
c. Placement Officer
d. College Office
e. College Hostel Office
f. Notice Board (Academic Section)
g. File

JNTUA COLLEGE OF ENGINEERING (Autonomous) :: ANANTHAPURAMU

Name of the Department

YEAR & SEMESTER	II B. Tech II Sem
COURSE NAME	Exploratory Data Analytics with R (SOC - II)
COURSE CODE	20A40508

Sample format for the Practical Courses Internal Test Evaluation (15 Marks)

S.No.	Name of the Student	Roll No.	Experiment Title	Evaluation Criteria: Preparation/Setup & Procedure (3 Marks)	Execution (5 Marks)	Observation / Analysis (3 Marks)	Viva - Voce (4 Marks)	Final Marks out of 15
				CO1& C03	CO2& CO4	CO2& CO5	CO3& CO5	
1	GUNDA POOJITHA	22001A0501	a. Viewing and manipulating Data b. Plotting data	3	5	3	3	14
2	R SOMASEKHAR	22001A0503	a. Write Demonstrate Statistical functions in R b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.	3	5	3	3	14
3	THADIKAMALLA JYOTHI	22001A0504	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	4	15
4	B VINOD KUMAR	22001A0505	a. Viewing and manipulating Data b. Plotting data	3	5	3	4	15
5	DABBUDUPALLI ANITHA	22001A0506	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	3	14
6	GIDDALURU SANTHOSH	22001A0507	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	3	5	3	3	14
7	GANGULA VENKATA RISI	22001A0509	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	3	5	3	3	14
8	NITTURU MANOGNA	22001A0510	b. Loops and iterations	3	5	3	3	14
9	CHEBROLU SUNIL	22001A0511	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	3	5	3	3	14
10	KUMPATI JEEVAN	22001A0512	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	3	5	3	3	14

11	KOMMURI MANISHA	22001A0513	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	3	5	3	3	14
12	POLEPALLI VENKATA JA	22001A0514	a. Viewing and manipulating Data b. Plotting data	3	5	3	3	14
13	PASRUGUBBA VENKATA	22001A0515	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	3	5	2	3	13
14	RAGHAVAREDDY VENKA	22001A0516	a. Defining user defined classes and operations, Models and methods in R b. Customizing the user's environment	3	5	3	3	14
15	RAJOLIMUCHI VASUDHA	22001A0517	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	3	5	3	3	14
16	SETHU REVANTHRAJ PAU	22001A0520	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	3	5	3	3	14
17	YENUGA TEJESWAR RED	22001A0521	a. Write Demonstrate Statistical functions in R b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.	3	5	3	3	14
18	AVULA VENKATESH	22001A0522	a. Viewing and manipulating Data b. Plotting data	3	5	3	3	14
19	AMBAPURAM SIREESHA	22001A0528	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	3	5	3	3	14
20	NALLABHAI DHANUSH	22001A0529	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	3	5	3	3	14
21	PATTEM BHANU PRAKAS	22001A0530	a. Viewing and manipulating Data b. Plotting data	3	5	3	3	14
22	B JYOTHISHA	22001A0531	a. Relationship between two continuous variables – covariance, correlation coefficients, comparing multiple correlations. b. Visualization methods – categorical and continuous variables, two categorical variables, two continuous variables.	3	5	3	4	15
23	SANE PRABHAS REDDY	22001A0533	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	4	15
24	MADDUBAIGARI MOHAM	22001A0534	a. Conditional statements b. Loops and iterations	3	5	3	4	15
25	VANKAYALA							

25	VANKAYALAPATI AKHIL	22001A0535	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	3	14
26	BOGINI ANITHA	22001A0536	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	3	5	3	4	15
27	Y PREM SWAROOP REDD	22001A0537	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	4	15
28	POLI SIVA NANDU	22001A0538	a. Viewing and manipulating Data b. Plotting data	3	5	3	3	14
29	PEDDAPALYAM RISHITH	22001A0539	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	3	5	3	3	14
30	SIBYALA PREETHI	22001A0540	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	3	5	3	3	14
31	THOTAPALLI SREEVIDYA	22001A0542	a. Relationship between two continuous variables – covariance, correlation coefficients, comparing multiple correlations. b. Visualization methods – categorical and continuous variables, two categorical variables, two continuous variables.	3	5	3	3	14
32	HANUVARI GRACELILLY	22001A0544	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	4	15
33	HALAHARVI SUMANTH R	22001A0545	a. Viewing and manipulating Data b. Plotting data	3	5	3	4	15
34	CHINTHA JAYANTH	22001A0546	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	3	14
35	BAINENI KEERTHANA	22001A0549	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	3	14
36	CHOPPARA HRISHIKESH I	22001A0550	a. Write Demonstrate Statistical functions in R b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.	3	5	3	3	14
37	CHANAGANI NEELIMA	22001A0551	a. Conditional statements b. Loops and iterations	3	5	3	3	14
38	MALLA THARUNI SRI	22001A0552	a. Perform tests of hypotheses about the mean when the variance is known. b. Compute the p-value. c. Explore the connection between the critical region, the test statistic, and the p-value	3	5	3	3	14

39	RAMPOGU HADASSAH N/	22001A0553	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	3	14
40	DUDEKULA MASOOD VA	22001A0554	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	3	5	3	3	14
41	NEELURU PRAVEEN KUM	22001A0555	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	3	5	3	3	14
42	PALAGIRI PRASANTH REI	22001A0557	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	3	5	3	3	14
43	VELIVELA SRINIKETH	22001A0558	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	3	14
44	TALARI VIJAY KUMAR	22001A0560	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	3	5	3	3	14
45	CHAGANTI KULADEEP	22001A0561	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	3	14
46	SHAIK MOHAMMED ADN	23005A0501	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	3	14
47	K KALYANKUMAR	23005A0502	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	4	15
48	PALAPARTHI LAKSHMI	23005A0503	a. Use the scatter plot to investigate the relationship between two variables	3	5	3	3	14
49	BUJULA KEERTHI	23005A0504	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	3	14
50	SHAIK ABDUL GAFOOR	23005A0505	a. Linear models b. Simple linear regression	3	5	3	3	14
51	NALI KUSUMA	23005A0506	a. Use the scatter plot to investigate the relationship between two variables	3	5	3	3	14
52	GAJJALA POOJITHA	23005A0507	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	3	5	3	3	14
53	MUSALE MANOJ RAO	23005A0508	a. Use the scatter plot to investigate the relationship between two variables	3	5	3	3	14
54	PUTTA RAVI TEJA	23005A0509	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	3	14
55	C GURU MOUNIKA	23005A0510	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	3	5	3	3	14

56	MUDE KIRAN K

				3				
					14			
56	MUDE KIRAN KUMAR NA	23005A0511	a.Conditional statements b.Loops and iterations	3	5	3	3	14

Total No.of Students	56	56	56	56
No.of Students scored more than threshold 70%	56	56	55	56
% of students scored more than threshold marks	100	100	98.214	100
Level Achieved (40%= Level 1 60% = Level 2 & 70% = Level 3)	3	3	3	3

Final CO Average Attainment for Internal test	CO1	CO2	CO3	CO4	CO5
	3	3	3	3	3

	JNTUA COLLEGE OF ENGINEERING (Autonomous) :: ANANTHAPURAMU								
	Name of the Department								
YEAR & SEMESTER	II B. Tech II Sem								
COURSE NAME	Exploratory Data Analytics with R (SOC - II)								
COURSE CODE	20A40508								
Sample format for the Practical Courses Internal Test Evaluation (15 Marks)									
				Evaluation Criteria					
S.No.	Name of the Student	Roll No.	Experiment Title	Preparation/Setup & Procedure (3 Marks)	Execution (5 Marks)	Observation / Analysis (3 Marks)	Viva - Voce (4 Marks)	Day to Day Evaluation 15 Marks
				CO1 & CO3	CO2 & CO4	CO2& CO5	CO3 & CO5	
1	GUNDA POOJITHA	22001A0501	a. Viewing and manipulating Data b. Plotting data	3	5	3	3	14
2	R SOMASEKHAR	22001A0503	a. Write Demonstrate Statistical functions in R b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.	3	5	3	3	14
3	THADIKAMALLA	22001A0504	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	4	15
4	B VINOD KUMAR	22001A0505	a. Viewing and manipulating Data b. Plotting data	3	5	3	4	15
5	DABBUDUPALLI A	22001A0506	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	3	14
6	GIDDALURU SAN	22001A0507	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	3	5	3	3	14
7	GANGULA VENKA	22001A0509	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	3	5	3	3	14
8	NITTURU MANOG	22001A0510	a. Conditional statements b. Loops and iterations	3	5	3	3	14
9	CHEBROLU SUNIL	22001A0511	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	3	5	3	3	14
10	KUMPATI JEEVAN	22001A0512	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	3	5	3	3	14
11	KOMMURI MANIS	22001A0513	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	3	5	3	3	14
12	POLEPALLI VENK	22001A0514	a. Viewing and manipulating Data b. Plotting data	3	5	3	3	14
13	PASRUGUBBA VE	22001A0515	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	3	5	2	3	13

14	RAGHAVAREDDY	22001A0516	a. Defining user defined classes and operations, Models and methods in R b. Customizing the user's environment	3	5	3	3	14
15	RAJOLIMUCHI VA	22001A0517	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	3	5	3	3	14
16	SETHU REVANTH	22001A0520	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	3	5	3	3	14
17	YENUGA TEJESW	22001A0521	a. Write Demonstrate Statistical functions in R b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.	3	5	3	3	14
18	AVULA VENKATE	22001A0522	a. Viewing and manipulating Data b. Plotting data	3	5	3	3	14
19	AMBAPURAM SIR	22001A0528	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	3	5	3	3	14
20	NALLABHAI DHA	22001A0529	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	3	5	3	3	14
21	PATTEM BHANU	22001A0530	a. Viewing and manipulating Data b. Plotting data	3	5	3	3	14
22	B JYOTHISHA	22001A0531	a. Relationship between two continuous variables – covariance, correlation coefficients, comparing multiple correlations. b. Visualization methods – categorical and continuous variables, two categorical variables, two continuous variables.	3	5	3	4	15
23	SANE PRABHAS R	22001A0533	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	4	15
24	MADDUBAIGARI	22001A0534	a. Conditional statements b. Loops and iterations	3	5	3	4	15
25	VANKAYALAPAT	22001A0535	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	3	14
26	BOGINI ANITHA	22001A0536	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	3	5	3	4	15
27	Y PREM SWAROO	22001A0537	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	4	15
28	POLI SIVA NANDU	22001A0538	a. Viewing and manipulating Data b. Plotting data	3	5	3	3	14
29	PEDDAPALYAM R	22001A0539	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	3	5	3	3	14

SIBYALA PREETH 22001A0
30 THOTAPALLI SRE 22001A0

30	SIBYALA PREETH	22001A0540	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	3	5	3	3	14
31	THOTAPALLI SRE	22001A0542	a. Relationship between two continuous variables – covariance, correlation coefficients, comparing multiple correlations. b. Visualization methods – categorical and continuous variables, two categorical variables, two continuous variables.	3	5	3	3	14
32	HANUVARI GRAC	22001A0544	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	4	15
33	HALAHARVI SUM	22001A0545	a. Viewing and manipulating Data b. Plotting data	3	5	3	4	15
34	CHINTHA JAYANT	22001A0546	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	3	14
35	BAINENI KEERTH	22001A0549	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	3	14
36	CHOPPARA HRISH	22001A0550	a. Write Demonstrate Statistical functions in R b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.	3	5	3	3	14
37	CHANAGANI NEE	22001A0551	a. Conditional statements b. Loops and iterations	3	5	3	3	14
38	MALLA THARUNI	22001A0552	a. Perform tests of hypotheses about the mean when the variance is known. b. Compute the p-value. c. Explore the connection between the critical region, the test statistic, and the p-value	3	5	3	3	14
39	RAMPOGU HADAS	22001A0553	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	3	14
40	DUDEKULA MASC	22001A0554	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	3	5	3	3	14
41	NEELURU PRAVEI	22001A0555	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	3	5	3	3	14
42	PALAGIRI PRASAI	22001A0557	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	3	5	3	3	14
43	VELIVELA SRINIK	22001A0558	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	3	14
44	TALARI VIJAY KU	22001A0560	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	3	5	3	3	14

45	CHAGANTI KULAI	22001A0561	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	3	14
46	SHAIK MOHAMMI	23005A0501	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	3	14
47	K KALYANKUMAI	23005A0502	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	4	15
48	PALAPARTHI LAK	23005A0503	a. Use the scatter plot to investigate the relationship between two variables	3	5	3	3	14
49	BUJULA KEERTHI	23005A0504	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	3	5	3	3	14
50	SHAIK ABDUL GA	23005A0505	a. Linear models b. Simple linear regression	3	5	3	3	14
51	NALI KUSUMA	23005A0506	a. Use the scatter plot to investigate the relationship between two variables	3	5	3	3	14
52	GAJJALA POOJITH	23005A0507	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	3	5	3	3	14
53	MUSALE MANOJ I	23005A0508	a. Use the scatter plot to investigate the relationship between two variables	3	5	3	3	14
54	PUTTA RAVI TEJA	23005A0509	a. Multiple regression b. Bias-variance trade-off – cross-validation	3	5	3	3	14
55	C GURU MOUNIKA	23005A0510	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	3	5	3	3	14
56	MUDE KIRAN KUN	23005A0511	a.Conditional statements b.Loops and iterations	3	5	3	3	14
			Total No.of Students	56	56	56	56	
			No.of Students scored more than threshold 70%	56	56	55	56	
			% of students scored more than threshold marks	**100.0**	**100.0**	**98.2**	**100.0**	
			Level Achieved (40%= Level 1 60% = Level 2 & 70% = Level 3)	**3**	**3**	**3**	**3**	

Final CO Average Attainment for Day to Day Evaluation	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
	3	**3**	**3**	**3**	**3**

JNTUA COLLEGE OF ENGINEERING (Autonomous) :: ANANTHAPURAMU									
Name of the Department									
YEAR & SEMESTER		II B. Tech II Sem							
COURSE NAME		Exploratory Data Analytics with R (SOC - II)							
COURSE CODE		20A40508							
Sample format for the Practical Courses External Test Evaluation (70 Marks)									
S.No.	Name of the Student	Roll No.	Experiment Title	Evaluation Criteria				Mapped COs	Final Marks out of 70
				Preparation (20 Marks)	Execution (15 Marks)	Observation / Analysis (15 Marks)	Viva - Voce (20 Marks)		
1	GUNDA POOJITH	22001A0501	a. Viewing and manipulating Data b. Plotting data	15	12	13	15		55
2	R SOMASEKHAR	22001A0503	a. Write Demonstrate Statistical functions in R b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.	10	10	10	10		40
3	THADIKAMALL	22001A0504	a. Reading the data from console, file (.csv) local disk and web b. Working with larger datasets	17	13	13	15		58
4	B VINOD KUMA	22001A0505	a. Viewing and manipulating Data b. Plotting data	17	13	13	16		59
5	DABBUDUPALLI	22001A0506	c. Reading the data from console, file (.csv) local disk and web d. Working with larger datasets	17	13	13	13		56
6	GIDDALURU SA	22001A0507	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	12	13	12	15		52
7	GANGULA VEN	22001A0509	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	17	13	13	15		58
8	NITTURU MANO	22001A0510	c. Conditional statements d. Loops and iterations	18	13	12	17		60

9	CHEBROLU SUN	22001A0511	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	18	13	12	13		56
10	KUMPATI JEEVA	22001A0512	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	18	13	12	11		54
11	KOMMURI MAN	22001A0513	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	18	13	12	15		58
12	POLEPALLI VEN	22001A0514	a. Viewing and manipulating Data b. Plotting data	18	13	12	10		53
13	PASRUGUBBA V	22001A0515	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	18	13	12	16		59
14	RAGHAVAREDD	22001A0516	a. Defining user defined classes and operations, Models and methods in R b. Customizing the user's environment	18	12	12	13		55
15	RAJOLIMUCHI V	22001A0517	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	18	13	15	16		62
16	SETHU REVANT	22001A0520	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	18	13	12	14		57
17	YENUGA TEJESV	22001A0521	a. Write Demonstrate Statistical functions in R b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.	18	13	15	15		61
18	AVULA VENKAT	22001A0522	a. Viewing and manipulating Data b. Plotting data	18	13	15	14		60
19	AMBAPURAM SI	22001A0528	a. Least Squares Estimates b. The R Function lm c. Scrutinizing the Residuals	18	13	15	16		62

20	NALLABHAI DH.	22001A0529	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	18	14	15	17		64
21	PATTEM BHANU	22001A0530	a. Viewing and manipulating Data b. Plotting data	18	13	15	16		62
22	B JYOTHISHA	22001A0531	a. Relationship between two continuous variables – covariance, correlation coefficients, comparing multiple correlations. b. Visualization methods – categorical and continuous variables, two categorical variables, two continuous variables.	18	15	15	17		65
23	SANE PRABHAS	22001A0533	a. Multiple regression b. Bias-variance trade-off – cross-validation	18	15	15	16		64
24	MADDUBAIGAR	22001A0534	c. Conditional statements d. Loops and iterations	18	14	15	18		65
25	VANKAYALAPA	22001A0535	c. Reading the data from console, file (.csv) local disk and web d. Working with larger datasets	18	13	15	11		57
26	BOGINI ANITHA	22001A0536	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	18	13	15	17		63
27	Y PREM SWARO	22001A0537	c. Multiple regression d. Bias-variance trade-off – cross-validation	15	15	15	15		60
28	POLI SIVA NANI	22001A0538	a. Viewing and manipulating Data b. Plotting data	12	13	12	17		54
29	PEDDAPALYAM	22001A0539	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	18	13	15	15		61
30	SIBYALA PREET	22001A0540	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	18	13	15	16		62

31	THOTAPALLI SR	22001A0542	a. Relationship between two continuous variables – covariance, correlation coefficients, comparing multiple correlations. b. Visualization methods – categorical and continuous variables, two categorical variables, two continuous variables.	19	14	15	17		65
32	HANUVARI GRA	22001A0544	c. Reading the data from console, file (.csv) local disk and web d. Working with larger datasets	19	14	15	14		62
33	HALAHARVI SU	22001A0545	a. Viewing and manipulating Data b. Plotting data	19	14	15	13		61
34	CHINTHA JAYA	22001A0546	c. Multiple regression d. Bias-variance trade-off – cross-validation	16	14	15	13		58
35	BAINENI KEERT	22001A0549	c. Multiple regression d. Bias-variance trade-off – cross-validation	15	15	15	15		60
36	CHOPPARA HRIS	22001A0550	a. Write Demonstrate Statistical functions in R b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.	12	13	12	18		55
37	CHANAGANI NE	22001A0551	c. Conditional statements d. Loops and iterations	15	14	15	13		57
38	MALLA THARUN	22001A0552	a. Perform tests of hypotheses about the mean when the variance is known. b. Compute the p-value. c. Explore the connection between the critical region, the test statistic, and the p-value	15	10	10	16		51
39	RAMPOGU HAD	22001A0553	c. Reading the data from console, file (.csv) local disk and web d. Working with larger datasets	18	13	15	17		63

40	DUDEKULA MAS	22001A0554	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	15	10	10	15		50
41	NEELURU PRAV	22001A0555	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	16	10	10	15		51
42	PALAGIRI PRAS	22001A0557	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	15	11	12	15		53
43	VELIVELA SRINI	22001A0558	c. Reading the data from console, file (.csv) local disk and web d. Working with larger datasets	15	14	15	15		59
44	TALARI VIJAY K	22001A0560	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset	15	10	10	15		50
45	CHAGANTI KUL	22001A0561	c. Multiple regression d. Bias-variance trade-off – cross-validation	14	10	10	12		46
46	SHAIK MOHAMM	23005A0501	c. Multiple regression d. Bias-variance trade-off – cross-validation	18	15	13	12		58
47	K KALYANKUM	23005A0502	c. Reading the data from console, file (.csv) local disk and web d. Working with larger datasets	18	15	13	13		59
48	PALAPARTHI LA	23005A0503	c. Use the scatter plot to investigate the relationship between two variables	15	11	11	12		49
49	BUJULA KEERTI	23005A0504	c. Reading the data from console, file (.csv) local disk and web d. Working with larger datasets	15	10	10	12		47
50	SHAIK ABDUL G	23005A0505	a. Linear models b. Simple linear regression	13	12	13	12		50
51	NALI KUSUMA	23005A0506	c. Use the scatter plot to investigate the relationship between two variables	18	15	13	10		56

52	GAJJALA POOJI	23005A0507	a. Tables, charts and plots. b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. c. Multivariate data, relationships between a categorical and a continuous variable	15	15	13	15		58
53	MUSALE MANOJ	23005A0508	c. Use the scatter plot to investigate the relationship between two variables	15	15	13	18		61
54	PUTTA RAVI TE	23005A0509	c. Multiple regression d. Bias-variance trade-off – cross-validation	12	12	13	10		47
55	C GURU MOUNI	23005A0510	a. Null hypothesis significance testing b. Testing the mean of one sample c. Testing two means	15	15	13	15		58
56	MUDE KIRAN KU	23005A0511	d. Loops and iterations	15	12	13	15		50
		Total No.of Students		**56**	**56**	**56**	**56**		
		No.of Students scored more than		**50**	**49**	**49**	**38**		
		% of students scored more than threshold marks		**89.2857**	**87.5**	**87.5**	**67.857**		
		Level Achieved (40%= Level 1 50% = Level 2 & 60% = Level 3)		**3**	**3**	**3**	**3**		

Final CO Average Attainment for End Exam	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
	3	**3**	**3**	**3**	**3**

...loratory Data Analy...

IE A...

COs

Assessment Tool

Exploratory Data Analytics with R (SOC - II) of CO Direct Attainment

COs	IE Assessment			EE Assessment	Direct CO attainment (0.3*IE Assessment +0.7*EE Assessment)
	Assessment Tool	Attainment Level	Average Level	End Exam	
CO1	Internal Test Evaluation	3.00	3.00	3	3.00
	Day to Day Evaluation	3.00			
CO2	Internal Test Evaluation	3.00	3.00	3	3.00
	Day to Day Evaluation	3.00			
CO3	Internal Test Evaluation	3.00	3.00	3	3.00
	Day to Day Evaluation	3.00			
CO4	Internal Test Evaluation	3.00	3.00	3	3.00
	Day to Day Evaluation	3.00			
CO5	Internal Test Evaluation	3.00	3.00	3	3.00
	Day to Day Evaluation	3.00			

Students Answered Level 1	1	2	1	1	2
Students Answered Level 2	8	9	15	46	9
Students Answered Level 3	47	45	40	9	45
Total Students	56	56	56	56	56
:tainment	2.82	2.77	2.70	2.14	2.77

Co Overall Attainment					
Attainment	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
Direct Attainment	3	3	3	3.00	3.00
Indirect Attainment	2.82	2.77	2.70	2.14	2.77
Overall Attainment	**2.96**	**2.95**	**2.94**	**2.83**	**2.95**

Exploratory Data Analytics with R (SOC - II) of PO & PSO Direct Attainment

Course	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
Exploratory Data Analytics with R Weighted Average	3.00	2.60	2.40	2.40	3.00	2.00	0.00	0.00	2.25	2.00	2.25	2.60	3.00	2.80	2.40
CO Attained Average	2.93	2.93	2.93	2.93	2.93	2.93	2.93	2.93	2.93	2.93	2.93	2.93	2.93	2.93	2.93
PO/PSO attinment Value	2.93	2.93	2.93	2.93	2.93	2.93	0	0	2.93	2.93	2.93	2.93	2.93	3	2.93

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2
CO1	3	2	1	1	3							2	3	2
CO2	3	2	2	2	3				2	2	2	2	3	3
CO3	3	3	3	3	3				2	2	2	3	3	3
CO4	3	3	3	3	3				2	2	2	3	3	3
CO5	3	3	3	3	3	2			3	2	3	3	3	3
	3	2.6	2.4	2.4	3	2	0	0	2.25	2	2.25	2.6	3	2.8

EXPLORATORY DATA ANALYTICS WITH R (SOC II)II COURSE SURVEY						
FACULTY NAME : Smt.D.Madhuri						
Admission Number	CO1: Understanding Basic R Programming Concepts	CO2: Data Handling and Visualization	CO3: Analysis of Data Relationships and Distributions	CO4: Application of Statistical Tests and Probability Distributions	CO5: Model Building and Regression Analysis	Timestamp
22001A0501	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	SATISFACTORY	12-06-2024 17:04
22001A0503	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 16:51
22001A0504	EXCELLENT	SATISFACTORY	EXCELLENT	GOOD	EXCELLENT	12-06-2024 13:06
22001A0505	EXCELLENT	EXCELLENT	EXCELLENT	SATISFACTO	EXCELLENT	12-06-2024 15:23
22001A0506	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 16:06
22001A0507	EXCELLENT	EXCELLENT	SATISFACTOR	GOOD	EXCELLENT	12-06-2024 15:25
22001A0509	SATISFACTORY	GOOD	EXCELLENT	GOOD	EXCELLENT	12-06-2024 12:02
22001A0510	EXCELLENT	EXCELLENT	GOOD	GOOD	SATISFACTORY	12-06-2024 17:17
22001A0511	EXCELLENT	GOOD	EXCELLENT	GOOD	EXCELLENT	12-06-2024 12:41
22001A0512	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 11:36
22001A0513	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 14:42
22001A0514	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	EXCELLENT	12-06-2024 13:25
22001A0515	GOOD	GOOD	EXCELLENT	GOOD	GOOD	12-06-2024 17:42
22001A0516	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 09:05
22001A0517	EXCELLENT	EXCELLENT	GOOD	GOOD	EXCELLENT	12-06-2024 17:19
22001A0520	EXCELLENT	EXCELLENT	EXCELLENT	EXCELLENT	EXCELLENT	12-06-2024 09:44
22001A0521	GOOD	GOOD	EXCELLENT	GOOD	GOOD	12-06-2024 12:28
22001A0522	EXCELLENT	EXCELLENT	GOOD	GOOD	EXCELLENT	12-06-2024 17:23
22001A0528	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 17:20
22001A0529	EXCELLENT	GOOD	GOOD	EXCELLENT	EXCELLENT	12-06-2024 13:13
22001A0530	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	GOOD	12-06-2024 13:59
22001A0531	GOOD	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 10:11
22001A0533	EXCELLENT	SATISFACTORY	GOOD	EXCELLENT	EXCELLENT	12-06-2024 17:19
22001A0534	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 17:22
22001A0535	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	GOOD	12-06-2024 17:20
22001A0536	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	EXCELLENT	12-06-2024 16:18
22001A0537	GOOD	GOOD	EXCELLENT	GOOD	EXCELLENT	12-06-2024 13:12
22001A0538	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 14:29
22001A0539	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	GOOD	12-06-2024 17:55
22001A0540	GOOD	GOOD	GOOD	EXCELLENT	EXCELLENT	12-06-2024 10:35
22001A0542	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 15:25
22001A0544	EXCELLENT	EXCELLENT	GOOD	GOOD	EXCELLENT	12-06-2024 15:23
22001A0545	EXCELLENT	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	12-06-2024 09:00
22001A0546	GOOD	GOOD	EXCELLENT	GOOD	EXCELLENT	12-06-2024 11:02
22001A0549	EXCELLENT	EXCELLENT	GOOD	GOOD	EXCELLENT	12-06-2024 12:54
22001A0550	EXCELLENT	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	12-06-2024 14:19
22001A0551	GOOD	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 16:57
22001A0552	EXCELLENT	GOOD	GOOD	GOOD	EXCELLENT	12-06-2024 13:17
22001A0553	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	GOOD	12-06-2024 12:24
22001A0554	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 17:57
22001A0555	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	EXCELLENT	12-06-2024 11:20
22001A0557	GOOD	EXCELLENT	EXCELLENT	GOOD	GOOD	12-06-2024 17:41
22001A0558	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 16:33

Faculty Sign

Performanc
Completed i

22001A0560	EXCELLENT	EXCELLENT	GOOD	GOOD	EXCELLENT	12-06-2024 10:47
22001A0561	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 15:23
23005A0501	EXCELLENT	EXCELLENT	GOOD	GOOD	EXCELLENT	12-06-2024 12:30
23005A0502	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 09:43
23005A0503	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 12:49
23005A0504	EXCELLENT	EXCELLENT	GOOD	GOOD	EXCELLENT	12-06-2024 13:35
23005A0505	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 14:05
23005A0506	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 12:50
23005A0507	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 12:58
23005A0508	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 14:52
23005A0509	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 14:48
23005A0510	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 12:30
23005A0511	EXCELLENT	EXCELLENT	EXCELLENT	GOOD	EXCELLENT	12-06-2024 16:19

	CO1	CO2	CO3	CO4	CO5
SATISFACTOR	1	2	1	1	2
GOOD	8	9	15	46	9
EXCELLENT	47	45	40	9	45

JNTUA College Of Engineering (Autonomous),Ananthapuramu

Computer Science & Engineering

Skill Oriented Course-II

Exploratory Data Analytics with R

Semester IV(R20) **L T P C : 1 0 2 2**

Course Code:

Course Objectives:

- How to manipulate data within R and to create simple graphs and charts used in introductory statistics.
- The given data using different distribution functions in R.
- The hypothesis testing and calculate confidence intervals; perform linear regression models for data analysis.
- The relevance and importance of the theory in solving practical problems in the real world.

Course Outcomes:

CO1: Install and use R for simple programming tasks.
CO2: Extend the functionality of R by using add-on packages
CO3: Extract data from files and other sources and perform various data manipulation tasks on them.
CO4: Explore statistical functions in R.
CO5: Apply the knowledge of R gained to data Analytics for real-life applications.

List of Experiments:

1: INTRODUCTION TO COMPUTING
a. Installation of R
b. The basics of R syntax, workspace
c. Matrices and lists
d. Subsetting
e. System-defined functions; the help system
f. Errors and warnings; coherence of the workspace

2: GETTING USED TO R: DESCRIBING DATA
a. Viewing and manipulating Data
b. Plotting data
c. Reading the data from console, file (.csv) local disk and web
d. Working with larger datasets

3: SHAPE OF DATA AND DESCRIBING RELATIONSHIPS
a. Tables, charts and plots.
b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape.
c. Multivariate data, relationships between a categorical and a continuous variable,
d. Relationship between two continuous variables – covariance, correlation coefficients, comparing multiple correlations.
e. Visualization methods – categorical and continuous variables, two categorical variables, two continuous variables.

4: PROBABILITY DISTRIBUTIONS

rformance Evaluation Criteria (R-RECOR
pleted in the same lab: 10 (M); with in w

a. Sampling from distributions – Binomial distribution, normal distribution
b. tTest, zTest, Chi Square test
c. Density functions
d. Data Visualization using ggplot – Box plot, histograms, scatter plotter, line chart, bar chart, heat maps

5: EXPLORATORY DATA ANALYSIS Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset.

6: TESTING HYPOTHESES
a. Null hypothesis significance testing
b. Testing the mean of one sample
c. Testing two means

7: PREDICTING CONTINUOUS VARIABLES
a. Linear models
b. Simple linear regression
c. Multiple regression
d. Bias-variance trade-off – cross-validation

8: CORRELATION
a. How to calculate the correlation between two variables.
b. How to make scatter plots.
c. Use the scatter plot to investigate the relationship between two variables

9: TESTS OF HYPOTHESES
a. Perform tests of hypotheses about the mean when the variance is known.
b. Compute the p-value.
c. Explore the connection between the critical region, the test statistic, and the p-value

10: ESTIMATING A LINEAR RELATIONSHIP Demonstration on a Statistical Model for a Linear Relationship
a. Least Squares Estimates
b. The R Function lm
c. Scrutinizing the Residuals

11: APPLY-TYPE FUNCTIONS
a. Defining user defined classes and operations, Models and methods in R
b. Customizing the user's environment
c. Conditional statements
d. Loops and iterations

12: STATISTICAL FUNCTIONS IN R
a. Write Demonstrate Statistical functions in R
b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.

References:

1. SandipRakshit, "Statistics with R Programming", McGraw Hill Education, 2018.
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, "AN Introduction to Statistical Learning: with Applications in R", Springer Texts in Statistics, 2017.
3. Joseph Schmuller, "Statistical Analysis with R for Dummies", Wiley, 2017.
4. K G Srinivasa, G M Siddesh, ChetanShetty, Sowmya B J, "Statistical Programming in R", Oxford Higher Education, 2017.

28 - 7/8 - 9
29 - 9/10 - 8
27 - 8/7/6

JNTUA COLLEGE OF ENGINEERING (AUTONOMOUS): Ananthapuramu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

II B.Tech II SEMESTER

DAY TO DAY PERFORMANCE EVALUATION

Name of the Laboratory: Exploratory Data Analytics with R

Sr No	Roll No	Lab No : 1 Date: 8/1				Lab No : 2 Date: 22/1				Lab No : 3 Date: 29/1				Lab No : 4 Date: 5/2				Lab No: 5 Date: 12/2				Lab No : 6 Date: 19/2				Total Average
		R 3	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	
1	22001A0501	3	1	2	2	3	1	2	2	3	1	1	2	2	1	2	2	3	1	2	2	3	1	2	2	8
2	22001A0503	A				2	1	2	3	3	1	2	2	A				2	1	1	2	2	1	2	2	7
3	22001A0504	3	2	1	2	2	2	2	2	3	1	2	2	2	2	1	2	2	1	2	2	3	1	2	2	9
4	22001A0505	2	2	2	2	2	2	2	3	3	2	1	2	3	2	1	2	3	1	1	2	2	2	1	2	9
5	22001A0506	3	1	2	2	3	1	2	2	2	1	1	3	3	2	1	1	2	2	2	2	2	2	1	2	8
6	22001A0507	2	2	2	2	2	1	2	2	3	1	2	2	3	2	1	2	3	1	1	2	2	2	1	3	7
7	22001A0509	3	1	1	2	2	1	1	2	3	1	2	3	3	1	2	3	3	1	2	2	2	1	2	3	8
8	22001A0510	2	1	1	3	3	2	1	3	2	1	2	2	3	1	2	2	3	2	1	2	3	1	2	2	8
9	22001A0511	3	2	1	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2	1	2	7
10	22001A0512	3	2	2	2	A				3	1	2	2	2	2	1	3	2	1	2	3	A				8
11	22001A0513	3	1	2	2	2	2	1	3	3	2	1	2	2	1	2	3	3	1	2	2	2	1	2	3	8
12	22001A0514	2	1	2	1	3	1	1	2	3	2	1	2	2	2	2	2	A				2	2	1	3	7
13	22001A0515	2	1	1	1	2	2	2	3	2	1	2	3	3	1	2	2	3	1	2	2	3	2	1	1	6
14	22001A0516	3	1	1	2	2	2	1	2	2	2	2	2	3	2	1	2	A				3	2	2	2	7
15	22001A0517	2	2	2	2	3	2	2	3	2	2	1	3	2	2	2	2	2	1	2	3	2	2	2	2	8
16	22001A0519	3	1	2	3	A				2	1	2	3	3	1	2	2	A				2	1	2	3	7
Remarks (if any)																										
Faculty Sign																										

Performance Evaluation Criteria (R-RECORD, O – OBSERVATION, V-VIVA VOCE),E-EXECUTION:

Completed in the same lab: 10 (M); with in week: 7-9 (M); next one week: 5-7 (M); more than two weeks: 1-4 (M); complete entire semester: 0 (M)

JNTUA COLLEGE OF ENGINEERING (AUTONOMOUS): Ananthapuramu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

II B.Tech II SEMESTER

DAY TO DAY PERFORMANCE EVALUATION

Name of the Laboratory: Exploratory Data Analytics with R

Sr No	Roll No	Lab No: 1 Date: 8/1				Lab No: 2 Date: 22/1				Lab No: 3 Date: 29/1				Lab No: 4 Date: 5/2				Lab No: 5 Date: 12/2				Lab No: 6 Date: 19/2				Total Average
		R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	
17	22001A0520	2	1	2	3	A				A				3	1	2	3	2	2	2	2	2	1	2	3	7
18	22001A0521	3	1	2	2	3	1	2	2	3	1	2	2	2	2	2	2	2	2	1	3	A				8
19	22001A0522	3	2	1	2	2	1	2	3	2	2	2	3	2	1	2	3	3	2	1	2	2	2	1	3	7
20	22001A0528	2	2	2	2	3	2	1	2	2	2	2	2	2	2	1	3	3	1	2	2	3	1	2	2	8
21	22001A0529	2	1	2	3	3	1	1	2	2	1	2	3	A				A				3	2	1	2	8
22	22001A0530	2	2	1	2	A				2	2	1	3	3	1	2	2	A				3	1	2	2	7
23	22001A0531	2	2	2	2	3	1	2	2	3	1	2	2	3	2	1	2	2	2	2	2	2	2	2	2	9
24	22001A0532	3	1	2	1	3	1	2	2	3	2	1	2	3	2	2	2	2	1	2	3	3	1	2	2	8
25	22001A0533	3	1	1	2	2	1	2	3	3	1	2	2	2	2	2	2	3	1	2	2	3	1	2	3	9
26	22001A0534	2	1	2	2	A				3	2	2	2	2	1	2	3	A				A				8
27	22001A0535	2	2	2	2	A				3	1	2	2	2	2	1	3	A				3	1	2	2	7
28	22001A0536	2	1	2	3	2	2	1	3	A				3	1	2	2	2	2	2	2	A				9
29	22001A0537	3	1	1	2	3	2	1	2	2	2	2	2	3	2	1	2	2	1	2	3	3	1	2	2	8
30	22001A0538	3	1	2	2	2	1	2	3	2	1	2	3	2	2	2	2	A				2	2	2	2	8
31	22001A0539	2	1	2	3	2	2	2	2	2	2	1	3	2	1	2	3	A				2	1	2	3	7
32	22001A0540	2	2	1	3	3	1	2	2	3	1	2	2	2	2	1	3	2	2	1	3	2	2	1	3	8
Remarks (if any)																										
Faculty Sign																										

Performance Evaluation Criteria (R-RECORD, O – OBSERVATION, V-VIVA VOCE),E-EXECUTION:
Completed in the same lab: 10 (M); with in week: 7-9 (M); next one week: 5-7 (M); more than two weeks: 1-4 (M); complete entire semester:

JNTUA COLLEGE OF ENGINEERING (AUTONOMOUS): Ananthapuramu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

II B.Tech II SEMESTER

DAY TO DAY PERFORMANCE EVALUATION

Name of the Laboratory: Exploratory Data Analytics with R

Sr No	Roll No	Lab No : 1 Date: 8/1				Lab No : 2 Date: 22/1				Lab No : 3 Date: 29/1				Lab No : 4 Date: 5/2				Lab No: 5 Date: 12/2				Lab No : 6 Date: 19/2				Total Average
		R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	
33	22001A0541	2	2	2	2	A				3	1	2	2	2	2	2	2	2	1	2	2	2	1	2	3	8
34	22001A0542	2	1	2	3	2	1	2	3	3	1	2	2	2	1	2	3	A				2	2	1	3	7
35	22001A0543	3	1	2	2	2	2	2	2	3	2	1	2	3	1	2	2	3	1	2	2	2	2	2	2	8
36	22001A0544	3	2	1	2	2	2	1	2	3	2	1	2	2	2	2	3	3	1	2	3	3	2	1	3	9
7	22001A0545	2	1	1	3	3	1	2	2	2	2	2	2	2	2	2	2	A				3	2	1	2	8
8	22001A0546	3	1	2	2	3	2	1	2	2	2	2	2	3	1	2	2	A				3	1	2	2	7
9	22001A0549	3	2	1	2	2	2	2	2	2	1	2	3	3	2	1	2	3	1	2	2	2	2	2	2	8
	22001A0550	2	2	2	2	2	1	2	3	2	2	1	3	3	1	2	2	A				2	1	2	3	7
	22001A0551	2	1	2	3	2	2	1	3	2	2	1	3	2	1	2	2	3	1	2	3	2	2	1	3	8
	22001A0552	2	2	1	3	3	1	2	2	3	1	2	2	2	2	1	2	3	1	2	2	3	1	2	2	8
	22001A0553	3	1	2	3	3	2	1	2	3	1	2	2	2	1	2	2	2	2	2	3	3	2	1	2	7
	22001A0554	3	2	1	3	A				3	2	1	2	2	1	2	2	A				2	2	2	2	7
	22001A0555	2	2	2	2	2	2	2	2	3	1	2	2	2	2	1	2	2	2	2	2	2	3	2	2	8
	22001A0556	2	1	2	3	2	1	2	3	2	2	2	3	2	2	2	2	2	1	2	3	2	2	2	2	8
	22001A0557	2	2	1	3	2	2	1	3	2	2	1	3	2	2	2	3	A				2	1	2	3	7
	22001A0558	3	1	2	2	A				A				A				A				2	2	1	3	8
arks (if any)																										
lty Sign																										

formance Evaluation Criteria (R-RECORD, O – OBSERVATION, V-VIVA VOCE),E-EXECUTION:

npleted in the same lab: 10 (M); with in week: 7-9 (M); next one week: 5-7 (M); more than two weeks: 1-4 (M); complete entire semester: 0 (M)

JNTUA COLLEGE OF ENGINEERING (AUTONOMOUS): A[illegible]napuramu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

II B.Tech II SEMESTER

DAY TO DAY PERFORMANCE EVALUATION

Name of the Laboratory: Exploratory Data Analytics with R

Sr No	Roll No	Lab No: 1 Date: 8/1				Lab No: 2 Date: 22/1				Lab No: 3 Date: 29/1				Lab No: 4 Date: 5/2				Lab No: 5 Date: 12/2				Lab No: 6 Date: 19/2				Total Average
		R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	
49	22001A0559	2	1	2	3	2	1	2	3	2	1	2	2	2	1	2	3	A				A				8
50	22001A0560	2	2	2	2	2	2	1	2	3	1	2	2	A				3	1	2	2	2	1	2	2	8
51	22001A0561	3	1	2	2	A				A				3	1	2	2	3	2	1	2	3	1	2	2	7
52	22001A0571	3	2	1	2	A				A				3	2	1	2	2	2	2	2	A				7
53	22001A0572	3	1	1	2	2	1	2	1	A				2	1	2	3	A				3	1	2	3	9
54	22001A0573	3	2	2	1	A				3	1	2	2	2	2	1	2	A				3	1	2	2	8
55	22001A0574	3	1	2	2	A				3	2	1	2	2	1	2	2	A				3	2	1	2	7
56	22001A0575	A				2	1	2	1	2	2	2	2	3	2	1	2	2	1	2	3	2	2	2	2	8
57	22001A0576	2	1	1	3	2	2	1	1	2	1	2	2	3	1	1	2	3	1	2	2	2	1	2	3	8
58	21001A0574	A				A				A				3	2	1	2	2	2	1	2	2	1	2	2	7
59	23005A0501	2	1	2	2	A				A				3	2	1	2	2	1	2	2	2	2	1	2	8
60	23005A0502	2	2	2	2	2	1	2	1	3	1	2	2	2	1	2	3	2	2	2	2	3	1	2	3	9
61	23005A0503	3	1	2	3	A				2	2	1	3	A				3	1	2	2	2	1	2	2	8
62	23005A0504	3	1	1	2	2	2	2	2	2	2	2	2	2	1	2	3	3	2	1	2	2	1	2	1	7
63	23005A0505	A				2	1	2	1	2	1	1	3	2	2	2	2	2	2	2	2	3	2	2	2	[illegible]
64	23005A0506	A				3	1	2	2	3	1	2	2	2	1	2	3	2	1	2	2	3	1	1	1	[illegible]
Remarks (if any)																												
Faculty Sign																												

Performance Evaluation Criteria (R-RECORD, O – OBSERVATION, V-VIVA VOCE), E-EXECUTION:
Completed in the same lab: 10 (M); with in week: 7-9 (M); next one week: 5-7 (M); more than two weeks: 1-4 (M); complete entire sem

JNTUA COLLEGE OF ENGINEERING (AUTONOMOUS): Ananthapuramu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

II B.Tech II SEMESTER

DAY TO DAY PERFORMANCE EVALUATION

Name of the Laboratory: Exploratory Data Analytics with R

Sr No	Roll No	Lab No : 1 Date: 8/1				Lab No : 2 Date: 22/1				Lab No : 3 Date: 29/1				Lab No : 4 Date: 5/2				Lab No: 5 Date: 12/2				Lab No : 6 Date: 19/2				Total Average
		R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	
65	23005A0507	3	1	2	2	3	1	2	3	2	1	2	3	2	1	2	3	A				2	2	2	2	8
66	23005A0508	A				3	1	2	2	2	2	1	3	3	1	2	2	2	1	2	2	2	1	2	2	7
67	23005A0509	2	2	2	2	A				3	2	1	2	3	2	1	3	2	2	1	2	2	2	1	2	7
68	23005A0510	2	1	2	3	2	2	2	3	3	1	2	2	3	2	1	2	3	1	1	2	3	1	2	2	8
69	23005A0511	2	2	1	3	2	1	2	3	2	1	2	2	3	2	1	1	3	1	1	2	3	2	1	2	8
Remarks (if any)																										
Faculty Sign																										

Performance Evaluation Criteria (R-RECORD, O – OBSERVATION, V-VIVA VOCE),E-EXECUTION:

Completed in the same lab: 10 (M); with in week: 7-9 (M); next one week: 5-7 (M); more than two weeks: 1-4 (M); complete entire semester: 0 (M)

JNTUA COLLEGE OF ENGINEERING (AUTONOMOUS): Ananthapuramu
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
II B.Tech II SEMESTER
DAY TO DAY PERFORMANCE EVALUATION

Name of the Laboratory: Exploratory Data Analytics with R

Sr No	Roll No	Lab No : 7 Date: 26/2				Lab No : 8 Date: 4/3				Lab No : 9 Date: 11/3				Lab No : 10 Date: 18/3				Lab No: 11 Date: 1/4				Lab No : 12 Date: 6/4				Total Average
		R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	
1	22001A0501	2	1	2	3	2	2	2	2	2	1	2	3	2	1	2	3	2	1	2	3	2	1	2	2	8
2	22001A0503	A								2	2	1	3	2	2	2	2	2	2	1	3	2	2	1	2	7
3	22001A0504	3	1	2	3	A				3	1	2	2	3	1	2	3	3	1	2	3	3	1	2	3	9
4	22001A0505	3	2	1	3	2	1	2	3	A				3	2	1	3	3	2	1	2	3	2	1	3	9
5	22001A0506	2	1	2	3	3	1	2	2	3	2	1	2	3	1	2	2	A				3	2	1	2	8
6	22001A0507	A								3	2	1	2	3	2	1	2	3	1	2	2	2	2	1	2	7
7	22001A0509	2	2	1	3	3	2	1	2	3	1	2	2	2	1	2	3	2	1	2	3	3	1	2	2	[illegible]
8	22001A0510	2	2	2	2	2	2	1	2	2	2	2	2	2	2	1	3	2	2	1	3	3	2	1	2	[illegible]
9	22001A0511	2	1	2	3	2	2	2	2	A				2	2	1	2	3	2	1	2	2	2	1	2	7
10	22001A0512	2	2	1	3	2	2	1	3	2	1	2	3	2	2	1	3	A				3	2	2	2	[illegible]
11	22001A0513	3	1	2	2	2	1	2	3	2	2	1	3	2	1	2	3	3	1	2	2	3	2	1	2	[illegible]
12	22001A0514	3	2	1	2	3	1	2	2	A				2	1	2	2	2	2	1	2	2	2	1	2	[illegible]
13	22001A0515	2	2	2	2	3	2	1	2	3	1	2	2	2	1	2	3	A				2	1	1	2	[illegible]
14	22001A0516	2	1	2	3	2	1	2	3	3	1	2	3	2	1	2	3	2	1	2	2	2	1	2	2	[illegible]
15	22001A0517	2	2	1	3	2	2	2	2	3	1	2	2	3	2	1	2	A				3	1	2	2	[illegible]
16	22001A0519	3	1	2	2	A				3	2	1	2	3	1	2	2	2	2	1	2	2	1	2	2	[illegible]
Remarks (if any)																										
Faculty Sign																										

Performance Evaluation Criteria (R-RECORD, O – OBSERVATION, V-VIVA VOCE),E-EXECUTION:

Completed in the same lab: 10 (M); with in week: 7-9 (M); next one week: 5-7 (M); more than two weeks: 1-4 (M); complete entire sem

Total Average
8
4
9
9
8
7
8
8
7
8
8
7
6
7
8
7

JNTUA COLLEGE OF ENGINEERING (AUTONOMOUS): Ananthapuramu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

II B.Tech II SEMESTER

DAY TO DAY PERFORMANCE EVALUATION

Name of the Laboratory: Exploratory Data Analytics with R

Sr No	Roll No	Lab No: 7 Date: 26/2				Lab No: 8 Date: 4/3				Lab No: 9 Date: 11/3				Lab No: 10 Date: 18/3				Lab No: 11 Date: 1/4				Lab No: 12 Date: 6/4				Total Average
		R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	
17	22001A0520	2	1	2	3	2	1	2	3	2	1	2	2	2	1	2	3	3	1	2	2	3	1	2	2	7
18	22001A0521	2	2	1	3	A				2	2	2	2	2	2	1	3	3	1	2	2	3	2	1	2	8
19	22001A0522	2	2	2	2	2	2	1	3	2	1	2	3	3	1	2	2	3	2	1	2	2	2	2	2	7
20	22001A0528	A				A				2	1	2	3	3	2	1	2	A				2	1	2	3	8
21	22001A0529	2	2	1	3	3	1	2	2	2	2	1	3	3	2	1	3	2	2	2	2	2	2	1	3	8
22	22001A0530	A				3	2	1	2	3	1	2	3	3	1	2	2	2	1	2	3	2	2	2	2	7
23	22001A0531	A				2	2	2	2	3	1	2	2	3	2	1	2	A				3	1	2	3	9
24	22001A0532	A				2	1	2	3	3	2	1	2	2	2	2	2	A				2	1	2	3	8
25	22001A0533	2	1	2	3	2	2	1	3	2	2	2	2	2	1	2	3	2	2	1	3	3	2	1	3	9
26	22001A0534	2	2	1	3	3	1	2	2	1	2	2	2	2	2	1	3	3	1	2	2	2	1	2	3	8
27	22001A0535	3	2	1	2	2		1	2	2	1	2	2	2	1	2	3	2	1	2	3	2	2	1	3	7
28	22001A0536	A				A				3	1	2	3	2	2	1	3	2	2	1	3	3	1	2	3	9
29	22001A0537	A				2	2	2	2	3	1	2	2	2		1	3	3	2	1	2	2	3	1	3	8
30	22001A0538	3	1	2	2	2	1	1	2	2	1	2	3	1		2	2	3	1	2	2	2	2	1	3	8
31	22001A0539	A				2	1	2	3	1	2	3	3	2		1	2	2	2	2	2	3	2	1	2	7
32	22001A0540	2	2	2	2	2	2	1	3	2	1	3	3	2		1	2	2	2	2	2	2	2	1	3	8
Remarks (if any)																										
Faculty Sign																										

Performance Evaluation Criteria (R-RECORD, O – OBSERVATION, V-VIVA VOCE),E-EXECUTION:

Completed in the same lab: 10 (M); with in week: 7-9 (M); next one week: 5-7 (M); **more than two weeks: 1-4 (M); complete entire semester: 0 (M)**

JNTUA COLLEGE OF ENGINEERING (AUTONOMOUS): Ananthapuramu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

II B.Tech II SEMESTER

DAY TO DAY PERFORMANCE EVALUATION

Name of the Laboratory: Exploratory Data Analytics with R

Sr No	Roll No	Lab No : 7 Date: 26/2				Lab No : 8 Date: 4/3				Lab No : 9 Date: 11/3				Lab No : 10 Date: 18/3				Lab No: 11 Date: 1/4				Lab No : 12 Date: 6/4				Total Average
		R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	
33	22001A0541	3	1	2	3	3	2	1	2	2	2	2	2	2	2	2	2	3	1	2	2	2	2	2	2	8
34	22001A0542	A				2	2	1	2	2	1	2	3	3	1	2	3	3	2	1	2	2	1	2	3	7
35	22001A0543	3	1	2	2	2	1	2	2	3	1	2	2	2	2	1	3	2	2	2	2	2	2	1	3	8
36	22001A0544	3	2	1	2	3	1	2	2	2	2	1	3	3	1	2	3	2	1	2	3	3	1	2	2	9
37	22001A0545	A				3	1	2	2	3	2	1	2	3	2	1	2	2	2	1	3	3	2	1	2	8
38	22001A0546	2	2	2	2	3	2	1	2	3	1	1	2	3	1	2	2	A				3	2	1	2	7
39	22001A0549	2	1	2	3	2	3	1	2	3	1	2	2	A				3	1	2	2	2	1	2	3	8
40	22001A0550	2	2	1	3	2	2	1	2	3	2	1	2	2	1	2	3	3	1	2	2	2	1	2	2	7
41	22001A0551	3	1	2	2	2	1	2	2	2	3	1	2	2	2	1	3	3	2	1	2	2	1	2	3	8
42	22001A0552	3	2	1	2	2	1	2	3	2	2	2	2	3	1	2	2	3	2	1	2	3	1	2	2	8
43	22001A0553	2	2	2	2	2	2	1	3	3	1	2	2	3	2	1	2	2	2	2	2	2	2	1	2	7
44	22001A0554	2	1	2	3	3	1	2	2	3	2	1	2	3	2	1	2	A				2	1	2	2	7
45	22001A0555	A				3	1	2	2	2	1	2	3	2	2	2	2	2	1	2	3	3	1	2	3	8
46	22001A0556	2	2	1	3	3	2	1	1	2	2	1	3	A				2	2	1	3	3	1	2	2	8
47	22001A0557	A				3	2	1	2	3	1	2	2	2	1	2	3	2	2	2	2	2	2	1	2	7
48	22001A0558	3	1	2	2	3	1	2	2	3	2	1	2	3	1	2	2	A				2	2	2	2	8
Remarks (if any)																										
Faculty Sign																										

Performance Evaluation Criteria (R-RECORD, O – OBSERVATION, V-VIVA VOCE),E-EXECUTION:
Completed in the same lab: 10 (M); with in week: 7-9 (M); next one week: 5-7 (M); more than two weeks: 1-4 (M); complete entire seme

JNTUA COLLEGE OF ENGINEERING (AUTONOMOUS): Ananthapuramu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

II B.Tech II SEMESTER

DAY TO DAY PERFORMANCE EVALUATION

…e of the Laboratory: Exploratory Data Analytics with R

Roll No	Lab No : 7 Date: 26/2				Lab No : 8 Date: 4/3				Lab No : 9 Date: 11/3				Lab No : 10 Date: 18/3				Lab No: 11 Date: 1/4				Lab No : 12 Date: 6/4				Total Average
	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	
22001A0559	3	1	2	2	2	1	2	2	2	1	2	3	2	1	2	3	2	1	2	2	2	2	2	2	8
22001A0560	A				2	1	2	3	3	1	2	2	A				3	1	2	2	2	1	2	3	8
22001A0561	A				2	2	1	3	2	2	1	3	2	1	2	3	2	2	1	2	2	1	2	2	7
22001A0571	3	2	1	2	3	1	2	2	A				2	1	2	2	A				2	2	1	2	7
22001A0572	3	1	2	2	3	2	1	2	3	2	1	2	3	1	2	3	A				3	1	2	3	9
22001A0573	A				2	1	2	3	3	1	2	2	A				—				3	1	2	2	8
22001A0574	2	2	2	2	2	2	1	3	3	2	1	2	A				—				2	1	2	3	7
22001A0575	2	1	2	3	2	1	2	3	2	2	1	3	2	1	2	3	2	1	2	3	3	1	2	2	8
22001A0576	2	2	1	3	2	2	1	3	2	1	2	3	2	2	1	3	2	2	1	3	3	2	1	2	8
21001A0574	3	1	2	2	3	1	2	2	3	1	2	2	3	1	2	2	A				2	1	2	2	7
23005A0501	A				3	2	1	2	3	1	2	2	3	2	1	2	3	1	2	3	3	1	2	2	8
23005A0502	3	2	1	2	2	1	2	3	3	2	1	2	3	2	1	2	3	1	2	3	3	1	2	3	9
23005A0503	3	2	1	2	2	2	1	3	2	1	2	3	2	2	1	3	3	1	2	2	3	1	2	2	8
23005A0504	A				3	1	2	2	A				2	2	2	2	3	2	1	2	2	1	2	2	7
23005A0505	A				3	2	1	2	2	2	1	3	2	2	2	2	A				2	1	2	3	8
23005A0506	2	1	2	3	3	2	1	2	2	1	2	3	2	1	2	3	2	1	2	2	2	2	1	2	7
…marks (if any)																									
…ulty Sign																									

…rformance Evaluation Criteria (R-RECORD, O – OBSERVATION, V-VIVA VOCE), E-EXECUTION:

…mpleted in the same lab: 10 (M); with in week: 7-9 (M); next one week: 5-7 (M); more than two weeks: 1-4 (M); complete entire semester: 0 (M)

JNTUA COLLEGE OF ENGINEERING (AUTONOMOUS): Ananthapuramu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

II B.Tech II SEMESTER

DAY TO DAY PERFORMANCE EVALUATION

Name of the Laboratory: Exploratory Data Analytics with R

Sr No	Roll No	Lab No : 7 Date: 26/2				Lab No : 8 Date: 4/3				Lab No : 9 Date: 11/3				Lab No : 10 Date: 18/3				Lab No: 11 Date: 1/4				Lab No : 12 Date: 6/4				Total Average
		R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	R	O	V	E	
65	23005A0507	A				2	2	2	2	A				3	1	2	2	2	1	2	3	3	1	2	2	8
66	23005A0508	A				2	1	2	3	2	1	2	3	3	2	1	2	2	2	1	3	2	1	2	2	7
67	23005A0509	A				3	1	2	2	A				3	1	2	2	A				2	2	1	2	7
68	23005A0510	A				3	2	1	2	3	1	2	2	A				2	2	2	2	3	1	2	2	8
69	23005A0511	A				2	1	2	3	3	1	2	2	2	2	2	2	3	1	2	2	3	2	1	2	8
Remarks (if any)																										
Faculty Sign																										

Performance Evaluation Criteria (R-RECORD, O – OBSERVATION, V-VIVA VOCE),E-EXECUTION:
Completed in the same lab: 10 (M); with in week: 7-9 (M); next one week: 5-7 (M); more than two weeks: 1-4 (M); complete entire semester

JNTUA College Of Engineering (Autonomous),Ananthapuramu
Computer Science & Engineering
Skill Oriented Course-II
Exploratory Data Analytics with R **L T P C : 1 0 2 2**
Semester IV(R20)

Course Code:
Course Objectives:

- How to manipulate data within R and to create simple graphs and charts used in introductory statistics.
- The given data using different distribution functions in R.
- The hypothesis testing and calculate confidence intervals; perform linear regression models for data analysis.
- The relevance and importance of the theory in solving practical problems in the real world.

Course Outcomes:

CO1: Install and use R for simple programming tasks.
CO2: Extend the functionality of R by using add-on packages
CO3: Extract data from files and other sources and perform various data manipulation tasks on them.
CO4: Explore statistical functions in R.
CO5: Apply the knowledge of R gained to data Analytics for real-life applications.

List of Experiments:

1: INTRODUCTION TO COMPUTING
a. Installation of R
b. The basics of R syntax, workspace
c. Matrices and lists
d. Subsetting
e. System-defined functions; the help system
f. Errors and warnings; coherence of the workspace

2: GETTING USED TO R: DESCRIBING DATA
a. Viewing and manipulating Data
b. Plotting data
c. Reading the data from console, file (.csv) local disk and web
d. Working with larger datasets

3: SHAPE OF DATA AND DESCRIBING RELATIONSHIPS
a. Tables, charts and plots.
b. Univariate data, measures of central tendency, frequency distributions, variation, and Shape.
c. Multivariate data, relationships between a categorical and a continuous variable,
d. Relationship between two continuous variables – covariance, correlation coefficients, comparing multiple correlations.
e. Visualization methods – categorical and continuous variables, two categorical variables, two continuous variables.

4: PROBABILITY DISTRIBUTIONS
a. Sampling from distributions – Binomial distribution, normal distribution

b. tTest, zTest, Chi Square test
c. Density functions
d. Data Visualization using ggplot – Box plot, histograms, scatter plotter, line chart, bar chart, heat maps

5: EXPLORATORY DATA ANALYSIS Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset.

6: TESTING HYPOTHESES
a. Null hypothesis significance testing
b. Testing the mean of one sample
c. Testing two means

7: PREDICTING CONTINUOUS VARIABLES
a. Linear models
b. Simple linear regression
c. Multiple regression
d. Bias-variance trade-off – cross-validation

8: CORRELATION
a. How to calculate the correlation between two variables.
b. How to make scatter plots.
c. Use the scatter plot to investigate the relationship between two variables

9: TESTS OF HYPOTHESES
a. Perform tests of hypotheses about the mean when the variance is known.
b. Compute the p-value.
c. Explore the connection between the critical region, the test statistic, and the p-value

10: ESTIMATING A LINEAR RELATIONSHIP Demonstration on a Statistical Model for a Linear Relationship
a. Least Squares Estimates
b. The R Function lm
c. Scrutinizing the Residuals

11: APPLY-TYPE FUNCTIONS
a. Defining user defined classes and operations, Models and methods in R
b. Customizing the user's environment
c. Conditional statements
d. Loops and iterations

12: STATISTICAL FUNCTIONS IN R
a. Write Demonstrate Statistical functions in R
b. Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.

JNTUA College Of Engineering (Autonomous),Ananthapuramu
Department of Computer Science & Engineering
Mandatory non-credit Course-III
Design Thinking for Innovation
Semester IV(R20)

Course Code: **L T P C : 2 1 0 0**

Course Objectives:

- The objective of this course is to familiarize students with design thinking process as a tool for breakthrough innovation. It aims to equip students with design thinking skills and ignite the minds to create innovative ideas, develop solutions for real-time problems.

Course Outcomes:

CO1: Define the concepts related to design thinking.
CO2: Apply the design thinking techniques for solving problems in various sectors.
CO3: Analyze to work in a multidisciplinary environment
CO4: Evaluate the value of creativity
CO5: Formulate specific problem statements of real time issues

UNIT – I: Introduction to Design Thinking
Introduction to elements and principles of Design, basics of design-dot, line, shape, form as fundamental design components. Principles of design.Introduction to design thinking, history of Design Thinking, New materials in Industry.

UNIT – II: Design Thinking Process
Design thinking process (empathize, analyze, idea & prototype), implementing the process in driving inventions, design thinking in social innovations. Tools of design thinking - person, costumer, journey map, brain storming, product development
Activity: Every student presents their idea in three minutes, Every student can present design process in the form of flow diagram or flow chart etc. Every student should explain about product development.

UNIT – III: Innovation
Art of innovation, Difference between innovation and creativity, role of creativity and innovation in organizations.Creativity to Innovation.Teams for innovation, Measuring the impact and value of creativity.
Activity: Debate on innovation and creativity, Flow and planning from idea to innovation, Debate on value-based innovation.

UNIT – IV: Product Design
Problem formation, introduction to product design, Product strategies, Product value, Product planning, product specifications. Innovation towards product design Case studies.
Activity: Importance of modelling, how to set specifications, Explaining their own product design.

References:

1. SandipRakshit, "Statistics with R Programming", McGraw Hill Education, 2018.
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, "AN Introduction to Statistical Learning: with Applications in R", Springer Texts in Statistics, 2017.
3. Joseph Schmuller, "Statistical Analysis with R for Dummies", Wiley, 2017.
4. K G Srinivasa, G M Siddesh, ChetanShetty, Sowmya B J, "Statistical Programming in R", Oxford Higher Education, 2017.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, ANANTAPUR
COLLEGE OF ENGINEERING, (Autonomous)
ANANTHAPURAMU - 515 002 (A.P.) INDIA

ACADEMIC RECORD BOOK

Academic Year : 2023-24

Class and Semester : II B.Tech - II Sem (R20)

Course : Exploratory Data Analytics with R

Branch : Computer Science & Engineering

Name and Designation of Staff Member handling the course }
1. Prof K. Madhavi
2. Smt D. Madhuri
3. Mr M. Vijaya Kanth.

1. 2. [signature] 3.

Signature of Concerned staff Member :

Class : II B.Tech
Semester : IInd
Branch : CSE

ATTENDANCE

Academic Year : 2023-24
Starting : 27-12-2024
Ending :
Teacher :

T: 1 P: 2

S.l. No.	Admn.No.	Name	1² (8/1)	2² (22/1)	3¹ (8/1)	4¹ (22/1)	5² (29/1)	6¹ (29/1)	7² (5/2)	8¹ (5/2)	9² (12/2)
1.	22001A0501	Gunda Poojitha	2	4	5	6	8	9	11	12	14
2.	503	R. Somasekhar	A	2	A	3	5	6	A	A	8
3.	504	T. Jyothirmayee	2	4	5	6	8	9	11	12	14
4.	505	B. Vinod Kumar	2	4	5	6	8	9	11	12	14
5.	506	D. Anitha	2	4	5	6	8	9	11	12	14
6	507	G. Santhosh	2	4	5	6	8	9	11	12	14
7	509	G. Venkata Rishitha	2	4	5	6	8	9	11	12	14
8	510	N. Manogna	2	4	5	6	8	9	11	12	14
9	511	C. Sunil	2	4	5	6	8	9	11	12	14
10	512	K. Jeevan	2	A	3	A	5	6	8	9	11
11	513	K. Manisha	2	4	5	6	8	9	11	12	14
12	514	P.V. Jagadeesh	2	4	5	6	8	9	11	12	A
13	515	P. Venkata Ravicharan	2	4	5	6	8	9	11	12	14
14	516	R.V. Sesha Reddy	2	4	5	6	8	9	11	12	A
15	517	R. Vasudha	2	4	5	6	8	9	11	12	14
16	519	S. Sujith	2	A	3	A	5	6	8	9	A
17	520	S. Revanth Raj Paul Yadav	2	A	3	A	A	A	5	6	8
18	521	Y. Tejeswara Reddy	2	4	5	6	8	9	11	12	14
19	522	A. Venkatesh	2	4	5	6	8	9	11	12	14
20	528	A. Sireesha	2	4	5	6	8	9	11	12	14
21	529	N. Dhanush	2	4	5	6	8	9	A	A	A
22	530	P. Bhanu Prakash	2	A	3	A	5	6	8	9	A
23	531	B. Jyothisha	2	4	5	6	8	9	11	12	14
24	532	H. Purandeswara Chowdary ~~Reddy~~	2	4	5	6	8	9	11	12	14
25	533	S. Prabas Reddy	2	4	5	6	8	9	11	12	14
26	534	M. Mohammed Irfan	2	A	3	A	5	6	8	9	A
27	535	V. Akhil	2	A	3	A	5	6	8	9	A
28	536	B. Anitha	2	4	5	6	A	A	8	9	11
29	537	Y. Prem Swaroop Reddy	2	4	5	6	8	9	11	12	13
30	538	P. Siva Nandu	2	4	5	6	8	9	11	12	A
31	539	P. Rishitha Sai	2	4	5	6	8	9	11	12	A
32	540	S. Preethi	2	4	5	6	8	9	11	12	14
33	541	N. Naga Swetha	2	A	3	A	5	6	8	9	11
34	542	T. Sreevidya	2	4	5	6	8	9	11	12	A
35	543	N. Amrutha Sai.	2	4	5	6	8	9	11	12	14

REGISTER

	10	11	12	13	14	15	16	17	18	19	20	Quiz	Assignment		Final Sessional Marks	Remarks
	12/2	19/2	19/2	26/2	26/2	4/3	4/3	11/3	11/3	18/3	1/4					
1	15	17	18	20	21	23	24	26	27	29	30				28	
3	9	11	12	A	A	A	A	14	15	17	18				27	
4	15	17	18	20	21	A	A	23	24	26	27				29	
5	15	17	18	20	21	23	24	A	A	26	27				29	
6	15	17	18	20	21	23	24	26	27	29	A				28	
7	15	17	18	A	A	A	A	20	21	23	24				28	
9	15	17	18	20	21	23	24	26	27	29	30				28	
10	15	17	18	20	21	23	24	26	27	29	30				28	
11	15	17	18	20	21	23	24	A	A	26	27				28	
12	12	A	A	14	15	17	18	20	21	23	A				28	
13	15	17	18	20	21	23	24	26	27	29	30				28	
14	13	14	15	17	18	20	21	A	A	23	24				27	
15	15	17	18	20	21	23	24	26	27	29	A				26	
16	A	14	15	17	18	20	21	23	24	26	27				27	
17	15	17	18	20	21	23	24	26	27	29	A				28	
19	A	11	12	14	15	A	A	17	18	20	21				27	
20	9	11	12	14	15	17	18	20	21	23	24				27	
21	15	A	A	17	18	A	A	20	21	23	24				28	
22	15	17	18	20	21	23	24	26	27	29	30				28	
28	15	17	18	A	A	A	A	20	21	23	A				28	
29	A	11	12	14	15	17	18	20	21	23	24				28	
30	A	11	12	A	A	14	15	17	18	20	21				27	
31	15	17	18	A	A	20	21	23	24	26	A				29	
32	15	17	18	A	A	20	21	23	24	26	A				28	
33	15	17	18	20	21	23	24	26	27	29	30				29	
34	A	11	12	14	15	17	18	20	21	23	24				29	
35	A	11	12	14	15	17	18	20	21	23	24				27	
36	12	A	A	A	A	14	15	17	18	20	21				29	
37	14	16	17	A	A	19	20	22	23	25	26				29	
38	A	14	15	17	18	A	A	20	21	23	24				28	
39	A	14	15	A	A	17	18	20	21	23	24				28	
40	15	17	18	20	21	23	24	26	27	29	30				28	
41	12	14	15	17	18	20	21	23	24	26	27				28	
42	13	15	16	A	17	19	20	22	23	24	25				28	
43	15	17	18	20	21	23	24	26	27	29	30				28	

RB
23/2/24

Class : II B.Tech
Semister : IInd
Branch : CSE

ATTENDANCE

S.I. No.	Admn.No.	Name		21	22	23	24	25	26	27	28	29	30
				1/4	6/4	6/4					C	L	
1.	22001A0501	Gunda Poojitha	1	32	39	35					11	24	
2.	503	R. Somasekhar	3	20	A	A					6	14	
3.	504	T. Jyothirmayee	4	29	30	32					10	22	
4.	505	B. Vinod kumar	5	29	A	A					9	20	
5.	506	D. Anitha	6	A	A	A					9	20	
6	507	G. Santhosh	7	26	27	29					9	20	
7	509	G. Venkata Rishitha	9	32	33	35					11	24	
8	510	N. Manogna	10	32	33	35					11	24	
9	511	C. Sunil	11	29	30	33					9	22	
10	512	K. Teerun	12	A	24	26					8	18	
11	513	K. Manisha	13	32	33	35					11	24	
12	514	P.V. Jagadeesh	14	26	27	29					8	20	
13	515	P. Venkata Ravichan	15.	A	A	A					9	20	
14	516	R.V. Sesha Reddy	16	29	30	32					10	22	
15	517	R. Vasudha	17	A	A	A					9	20	
16	519	S. Sujith	19	23	A	A					7	16	
17	520	S. Revanth Raj Pauly	20	26	A	A					8	18	
18	521	Y. Tejeswara Reddy	21	26	A	A					8	18	
19	522	A. Venkatesh	22	32	A	A					10	22	
20	528	A. Sireesha	28	A	A	A					7	16	
21	529	N. Dhanush	29	26	27	29					9	20	
22	530	P. Bhanu Prakash	30	23	24	26					8	18	
23	531	B. Jyothisha	31	A	A	A					8	18	
24	532	H. Purandareswara Reddy Chou	32	A	A	A					8	18	
25	533	S. Prabas Reddy	33	32	A	A					10	22	
26	534	M. Mohammed Irfan	34	26	A	A					8	18	
27	535	V. Akhil	35	26	A	A					8	18	
28	536	B. Anitha	36	23	A	A					7	16	
29	537	Y. Prem Swaroop Red	37	28	29	31					9	22	
30	538	P. Siva Nandu	38	26	27	29					9	20	
31	539	P. Rishitha Sai	39	26	27	29					9	20	
32	540	S. Preethi	40	32	33	35					11	24	
33	541	N. Naga Swetha	41	29	30	32					10	22	
34	542	T. Sreevidya	42	27	A	19					10	20	
35	543	N. Amrutha Sai.	43	32	33	35					11	24	

REGISTER

	Unit Test						Total of Best Units	Quiz	Assign-ment		Final Sessional Marks	Remarks
	I	II	III	IV	V	VI						
1											28	
3											27	
4											29	
5											29	
6											28	
7											28	
9											28	
10											28	
11											28	
12											28	
13											28	
14											27	
15											26	
16											27	
17											28	
19											27	
20											27	
21											28	
22											28	
28											28	
29											28	
30											27	
31											29	
32											28	
33											29	
34											29	
35											27	
36											29	
37											29	
38											28	
39											28	
40											28	
41											28	
42											28	
43											28	

Class : II B.Tech
Semister : IInd
Branch : CSE

Academic Year : 2023-24
Starting :
Ending :
Teacher :

ATTENDANCE

S.l. No.	Admn.No.	Name	No	1^{2}	2^{2}	3^{1}	4^{1}	5^{2}	6^{1}	7^{2}	8^{1}	9^{2}
			Date	8/1	22/1	8/1	22/1	29/1	29/1	5/2	5/2	12/2
36	22001A0544	H. Gracelilly		2	4	5	6	8	9	11	12	14
37	545	H. Sumanth Raj		2	4	5	6	8	9	11	12	A
38	546	C. Jayanth		2	4	5	6	8	9	11	12	A
39	549	B. Keerthana		2	4	5	6	8	9	11	12	14
40	550	C. Hrishikesh Kumar		2	4	5	6	8	9	11	12	A
41	551	C. Neelima		2	4	5	6	8	9	11	12	14
42	552	H. Tharuni Sri		2	4	5	6	8	9	11	12	14
43	553	B. H. Nathanyah		2	4	5	6	8	9	11	12	14
44	554	Dudekula Masoodvali		2	4	5	6	A	A	8	10	A
45	555	N. Praveen kumar Reddy		2	4	5	6	8	9	11	12	14
46	556	M. Kavya Sudha		2	4	5	6	8	9	11	12	14
47	557	P. Prasanth Reddy		2	4	5	6	8	9	11	12	A
48	558	V. Saiiketh		2	A	3	A	A	A	5	6	A
49	559	V. V. L. Nanndeeswar		2	4	5	6	8	9	11	12	1[illegible]
50	560	T. Vijay Kumar		2	4	5	6	8	9	A	A	10
51	561	C. Kuladeep		2	A	3	A	A	A	5	6	8
52	571	G. Abhilash		2	A	3	A	A	A	5	6	A
53	572	Y. Nikhil Sauri		2	4	5	6	A	A	8	9	11
54	573	S. Ibtihajulla Sha		2	A	3	A	5	6	8	9	A
55	574	S. Mahammed Sameer		2	A	3	A	5	6	8	9	A
56	575	G. S. Nanda krishna		A	2	A	3	5	6	8	9	11
57	576	B. Ajay kumar Reddy		2	4	5	6	8	9	11	12	14
58	21001A0574	R. Venu Madhava Naidu		A	2	A	3	4	5	7	8	10
59	23005A0501	S. Mohammed Adnan		2	A	3	A	5	6	8	9	10
60	502	K. kalyankumar		2	4	5	6	8	9	11	12	14
61	503	P. Lakshmi		2	A	3	A	5	6	A	A	8
62	504	B. Keerthi		2	4	5	6	8	9	11	12	14
63	505	S. Abdul Gafoor.		A	2	A	3	5	6	8	9	11
64	506	N. Kusuma		A	2	A	3	5	6	8	9	11
65	507	G. Poojitha		2	4	5	6	8	9	11	12	A
66	508	M. Manoj Rao		A	2	A	3	5	6	8	9	11
67	509	P. Ravi Teja		2	A	3	A	5	6	8	9	11
68	510	C. Guru Mounika		2	4	5	6	8	9	11	12	14
69	511	M. ~~Durga Naik~~ Kiran Kumar Naik.		2	4	5	6	8	9	11	12	14
70												

REGISTER

	10	11	12	13	14	15	16	17	18	19	20	Quiz	Assignment		Final Sessional Marks	Remarks
	12/2	19/2	19/2	26/2	26/2	4/3	4/3	11/3	11/3	18/3	1/4					
44	15	17	18	20	21	23	24	26	27	29	30				29	
45	13	15	16	A	A	18	19	21	22	24	25				29	
46	12	14	15	17	18	20	21	23	24	26	A				28	
47	15	17	18	20	21	23	24	26	27	A	28				28	
50	A	14	15	17	18	20	21	23	24	26	27				27	
51	15	17	18	20	21	23	24	26	27	29	30				28	
52	15	17	18	20	21	23	24	26	27	29	30				28	
53	15	17	18	20	21	23	24	26	27	29	30				28	
54	A	11	12	14	15	17	18	20	21	23	A				27	
55	15	17	18	A	A	20	21	23	24	26	27				28	
56	15	A	16	18	19	21	22	23	24	A	25				28	
57	A	14	15	A	A	17	18	20	21	23	24				28	
58	9	[illegible]	12	A	A	14	15	17	19	20	A				28	
59	15	A	A	17	18	20	21	23	24	26	27				28	
60	13	15	16	A	A	18	19	20	22	A	23				28	
61	9	11	12	A	A	14	15	17	18	20	21				27	
72	A	A	A	8	9	11	12	A	A	14	A				27	
72	12	14	15	17	18	20	21	23	24	26	A				29	
73	A	11	12	A	A	14	15	17	18	20	21				28	
74	A	11	12	14	15	17	18	20	21	A	A				28	
75	12	14	15	17	18	20	21	23	24	26	27				28	
76	15	17	19	20	21	A	A	22	23	26	27				28	
21.74	11	13	14	16	17	19	20	22	23	25	26				26	
1	11	13	14	A	A	16	17	19	20	22	23				27	
2	15	17	18	20	21	23	24	26	27	29	30				29	
3	9	11	12	14	15	17	18	20	21	23	24				28	
4	15	17	18	A	A	20	21	A	A	23	24				28	
5	12	14	15	A	A	17	18	20	21	23	A				28	
6	12	14	15	17	18	20	21	23	24	26	27				28	
7	A	14	15	A	A	17	18	A	A	20	21				28	
8	12	14	15	A	A	17	18	20	21	23	24				27	
9	12	14	15	A	A	17	18	A	A	20	A				27	
10	15	17	18	A	A	20	21	23	24	A	25				27	
11	15	17	18	A	A	20	21	23	24	26	27				28	

23/9/24

Class : II B.Tech
Semester : IInd
Branch : CSE

ATTENDANCE

S.l. No.	Admn.No.	Name		21	22	23	24	25	26	27	28	29	30
				2	1	2							
				4	6/4	6/4					c	L	
36	22001A0544	H. Gracelilly	44	32	33	35					11	24	
37	545	H. Sumanth Raj	45	27	A	29					9	20	
38	546	C. Jayanth	46	A	A	A					8	18	
39	549	B. Keerthana	49	30	A	A					10	20	
40	550	C. Hrishikesh Kuma	50	29	A	A					9	20	
41	551	C. Neelima	51	32	A	A					10	22	
42	552	M. Tharuni Sri	52	32	33	35					11	24	
43	553	B. H. Nathanyah	53	32	33	35					11	24	
44	554	Dudekula Masood	54	A	A	A					7	16	
45	555	N. Praveen Kumar	55	29	30	32					10	22	
46	556	M. Kavya Sudha	56	27	A	A					9	18	
47	557	P. Prasanth Reddy	57	26	A	A					8	18	
48	558	V. Saiketh	58	A	A	A	+6				6	14	=20
49	559	V. V. L. Nanneeswa	59	29	A	A					9	20	
50	560	T. Vijay Kumar	60	25	26	28					10	18	
51	561	C. Kuladeep	61	23	A	A					7	16	
52	571	G. Abhilash	71	A	A	A					4	10	
53	572	Y. Nikhil Souri	72	A	27	29					9	20	
54	573	S. Ibtihajulla Sha	73	22	A	A					7	16	
55	574	S. Mohammed Sam	74	A	A	A					7	14	
56	575	G. S. Nanda Krisha	75	29	30	32					10	22	
57	576	B. Ajay Kumar Redd	76	28	30	32					10	22	
58	21001A0574	R. Venu Madhava Na	21-74	28	29	31					8	12	992
59	23005A0501	S. Mohammed Adn	23-1	25	A	A					9	16	
60	502	K. Kalyankumar	2	32	33	35					11	24	
61	503	P. Lakshmi	3	26	27	29					9	20	
62	504	B. Keerthi	4	26	27	29					9	20	
63	505	S. Abdul Gafoor	5	A	24	26					8	18	
64	506	N. Kusuma	6	29	30	33					10	22	
65	507	G. Poojitha	7	23	24	26					8	18	
66	508	M. Manoj Rao	8	26	A	A					8	18	
67	509	P. Ravi Teja	9	A	21	23					7	16	
68	510	C. Guru Mounika	10	27	28	30					10	20	
69	511	M. ~~Durga~~ Kiran Kumar Naik	11	29	A	A					9	20	
~~70~~						1968							

REGISTER

	31	32	33	34	35	36	37	38	39	40

ISTER

Quiz	Assign-ment		Final Sessional Marks	Remarks
			29	
			29	
			28	
			28	
			27	
			28	
			28	
			28	
			27	
			28	
			28	
			28	
			28	
			28	
			28	
			27	
			27	
			29	
			28	
			28	
			28	
			28	
			26	
			27	
			29	
			28	
			28	
			28	
			28	
			28	
			27	
			27	
			27	
			28	

Class : II B.Tech
Semester : IInd
Branch : CSE

ATTENDANCE

S.l. No.	Admn.No.	Name		81	82	83	84	85	86	87	88	89	90
36	22001A0544	H. Gracelilly	44										
37	545	H. Sumanth Raj	45										
38	546	C. Jayanth	46										
39	549	B. Keerthana	49										
40	550	C. Hrishikesh Kuma	50										
41	551	C. Neelima	51										
42	552	M. Tharuni Sri	52										
43	553	R. H. Nathanyah	53										
44	554	Dudekula Masood	54										
45	555	N. Praveen kumar	55										
46	556	M. Kavya Sudha	56										
47	557	P. Prasanth Reddy	57										
48	558	V. Sriniketh	58										
49	559	V. V. L. Nanndeeswa	59										
50	560	T. Vijay Kumar	60										
51	561	C. Kuladeep	61										
52	571	G. Abhilash	71										
53	572	Y. Nikhil Sauri	72										
54	573	S. Ibtihajulla Sha	73										
55	574	S. Mohammed Sam	74										
56	575	G. S. Nanda Krishn	75										
57	576	B. Ajay Kumar Redd	76										
58	21001A0574	R. Venu Madhava N	21..										
59	23005A0501	S. Mohammed Adn	23.										
60	502	K. Kalyankumar											
61	503	P. Lakshmi	3										
62	504	B. Keerthi	4										
63	505	S. Abdul Gafoor	5										
64	506	N. Kusuma	6										
65	507	G. Poojitha	7										
66	508	M. Manoj Rao	8										
67	509	P. Ravi Teja	9										
68	510	C Guru Mounika	1										
69	511	M. ~~Durga~~ Kiran Kumar Naik.	1										
70													

REGISTER

	Unit Test						Total of Best Units	Quiz	Assign-ment		Final Sessional Marks	Remarks
	I	II	III	IV	V	VI						
44											29	
45											29	
46											28	
49											28	
50											27	
51											28	
52											28	
53											28	
54											27	
55											28	
56											28	
57											28	
58											28	
59											28	
60											28	
61											27	
71											27	
72											29	
73											28	
74											28	
75											28	
76											28	
21..74											26	
23..1											27	
2											29	
3											28	
4											28	
5											28	
6											28	
7											28	
8											27	
9											27	
10											27	
11											28	

DAIRY OF LECTURER CLASSES

S.L.No.	Date	Period(s)	Topic Covered
1)	8-1-2024	4,5,6	Introduction to computing a. Installation of R b. The basics of R syntax, workspace c. Matrices & lists d. Subsetting
2)	22-1-2024	4,5,6	e. system-defined functions, the help system. f: errors and warnings, coherence of the workspace.
3)	29-1-2024	4,5,6	Getting used to R: Describing data a. Viewing and manipulating data. b. Plotting data c. Reading the data from the console, file (.csv) local disk & web. d. working with larger datasets.
4)	5-2-2024	4,5,6	Shape of data and describing relationships. a. Tables, charts & plots. b. Univariate data, measures of central tendency, frequency distributions, variations & shape. c. Multivariate data, relationship b/w categorical & continuous variable. d. Relationship b/w two continuous variables - covariance, correlation coefficients, comparing multiple correlations. e. Visualization methods - categorical and continuous variable, two categorical variables, two continuous variables.

DAIRY OF LECTURE CLASSES

S.L.No.	Date	Periods	Topic Covered
5)	12-2-2024	4,5,6	Probability distributions:
			a. Sampling from distributions - Binomial distribution, normal distribution.
			b. tTest, zTest, chi Square Test
			c. Density function.
6)	19-2-2024	4,5,6	d. Data visualization using ggplot - Box plot, histograms, scatter plot, line chart, bar chart, heat maps.
7)	26-2-2024	4,5,6	Demonstrate the range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population data set.
8)	4-3-2024	4,5,6	Null hypothesis significant test
			Testing the mean of one sample
			Testing two mean
9)	11-3-2024	4,5,6	Predicting continuous variables
			Linear model, simple linear regression, Multiple regression, Bias variance trade off cross-validation.
10)	18-3-2024	4,5,6	Correlation
			How to calculate correlation b/w the two variables.
			How to make scatter plots.
			Use the scatter plots to investigate the relationship between two variables.
11), 12)	1,6-4-2024	4,5,6	Test of hypothesis.
			Perform the test of hypothesis about the mean when the variance is known.
			Compute the P-value.
			Explore the connection b/w critical region, the test statistics & P values.
			[illegible] 23/3/24

JNTUA College of Engineering (Autonomous)
Anantapuramu

Department of Computer Science and Engineering

Exploratory Data Analytics with R

(SKILL ORIENTED COURSE) (R20)

Lab Manual

Prepared by
D. MADHURI
Assistant Professor(C)
Department of CSE,
JNTUACEA, ATP.

JNTUA College of Engineering (Autonomous)
Ananthapuramu Department of Computer Science and Engineering
B.Tech(R20)
Exploratory Data Analytics with R

Semester IV **LTPC:1 0 2 2**

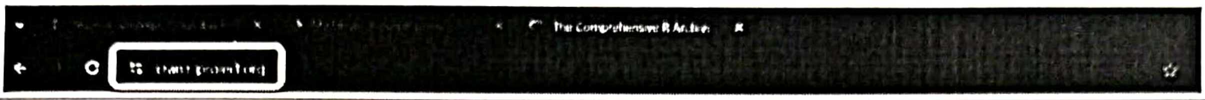
S.NO	PAGE NO	Name of the Experiment
1	1-19	1:INTRODUCTIONTOCOMPUTING Installation of R The basics of R syntax, workspace Matrices and lists Sub setting System-defined functions; the help system Errors and warnings ; coherence of the workspace
2	20-24	2:GETTING USED FOR:DESCRIBING DATA Viewing and manipulating Data Plotting data Reading the data from console,file (.csv) local disk and web Working with larger datasets
3	26-36	3:SHAPE OF DATA AND DESCRIBING RELATIONSHIPS Tables, charts and plots. Univariate data, measures of central tendency, frequency distributions, variation, and Shape. Multivariate data, relationships between acategorical and a continuous variable, Relationship between two continuous variables– covariance, correlation coefficients, comparing multiple correlations. e. Visualization methods–categorical and continuous variables, two categorical variables, two continuous variables.
4	37-59	4:PROBABILITYDISTRIBUTIONS Sampling from distributions–Binomial distribution, normal distribution T-Test ,zTest, Chi Square test Density function sd.Data Visualizationus in gggplot–Box plot, histograms, scatter plotter, line chart, bar chart, heat maps
5	60-62	5:EXPLORATORYDATAANALYSISDemonstrate The range, summary, mean, variance, median, standard deviation, histogram, box plot, scatter plot using population dataset.
6	63-64	6:TESTING HYPOTHESES Null hypothesis significance testing Testing the mean of one sample. Testing two means
7	65-67	7:PREDICTINGCONTINUOUSVARIABLES Linear models Simple linear regression Multiple regression Bias-variance trade-off–cross-validation
8	68-70	8:CORRELATION How to calculate the correlation between two variables How to make scatter plots. Use the scatter plot to investigate the relationship between two variables

9	71-72	9:TESTSOFHYPOTHESES Perform tests of hypotheses about the mean when the variance is known Compute the p-value.
10	73-74	10:ESTIMATINGALINEARRELATIONSHIP Demonstration on a Statistical Model for a Linear Relationship Least Squares Estimates The R Function lmc. Scrutinizing the Residuals
11	75-80	11:APPLY-TYPEFUNCTIONS Defining user defined classes and operations, Models and methods in R Customizing the user's environment Conditional statements Loops and iterations
12	81-86	12:STATISTICALFUNCTIONSINR Write Demonstrate Statistical functions in R Statistical inference, contingency tables, chi-square goodness of fit, regression, generalized linear models, advanced modeling methods.

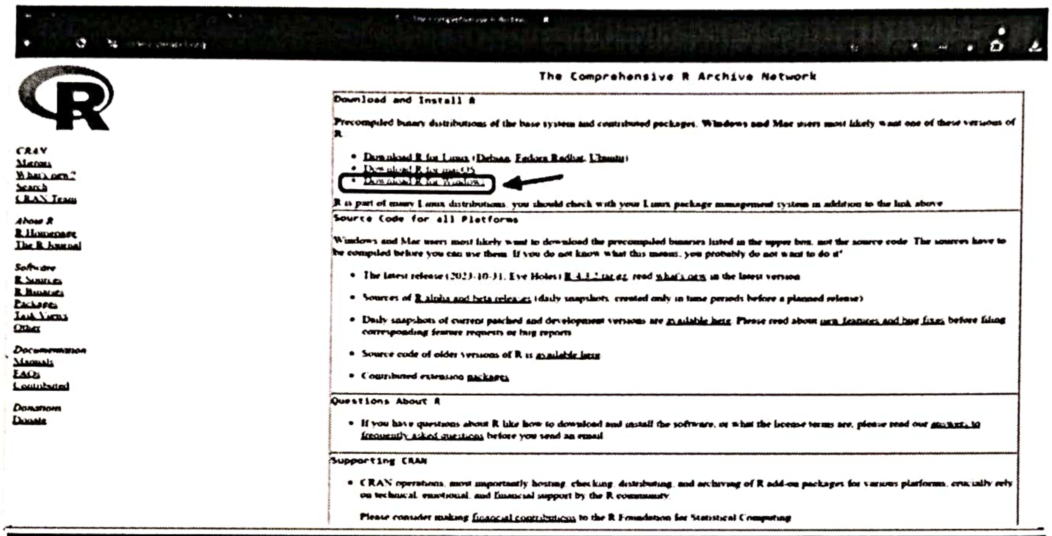
Week-1:Introduction to computing

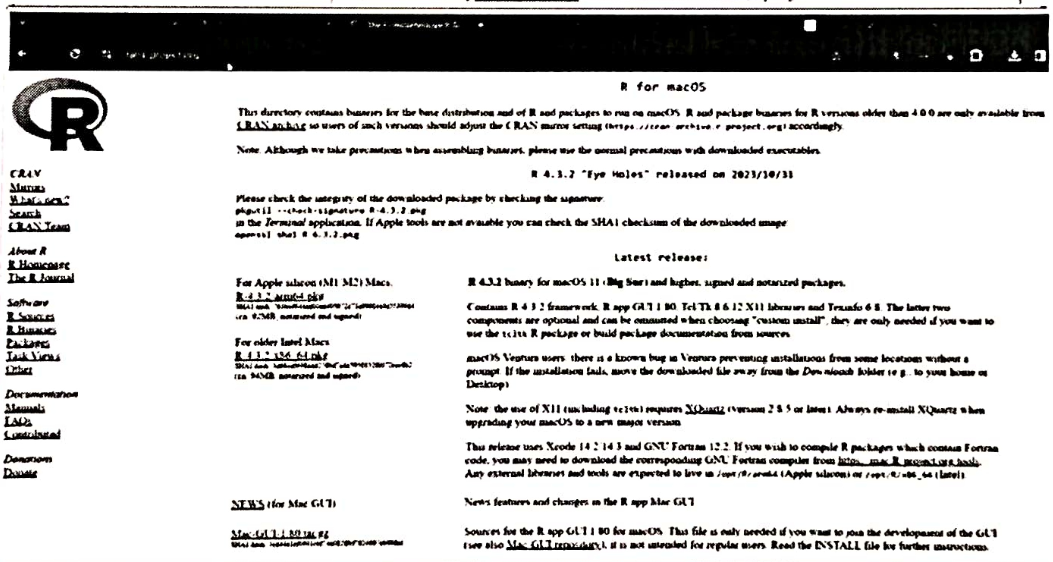
a.installation of R

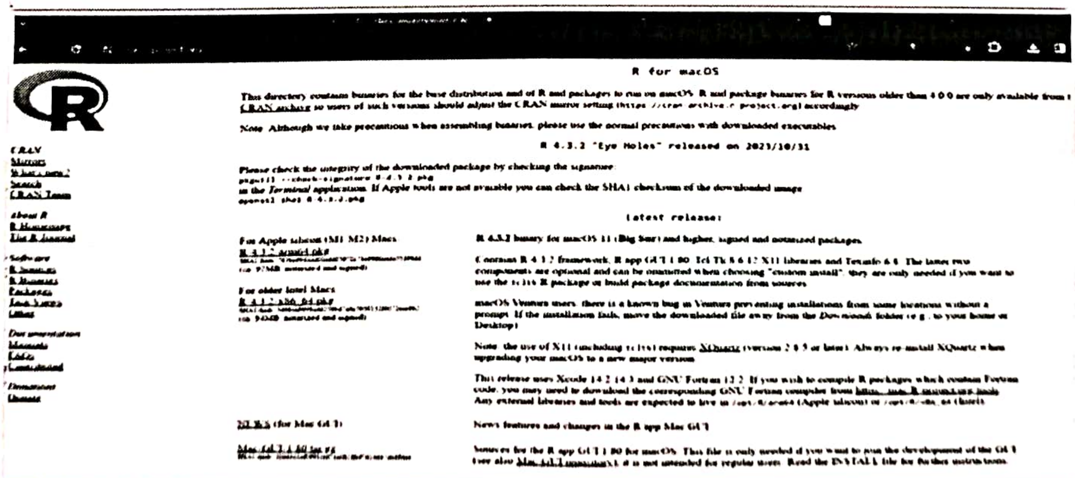
1. To install R, go to cran.r-project.org

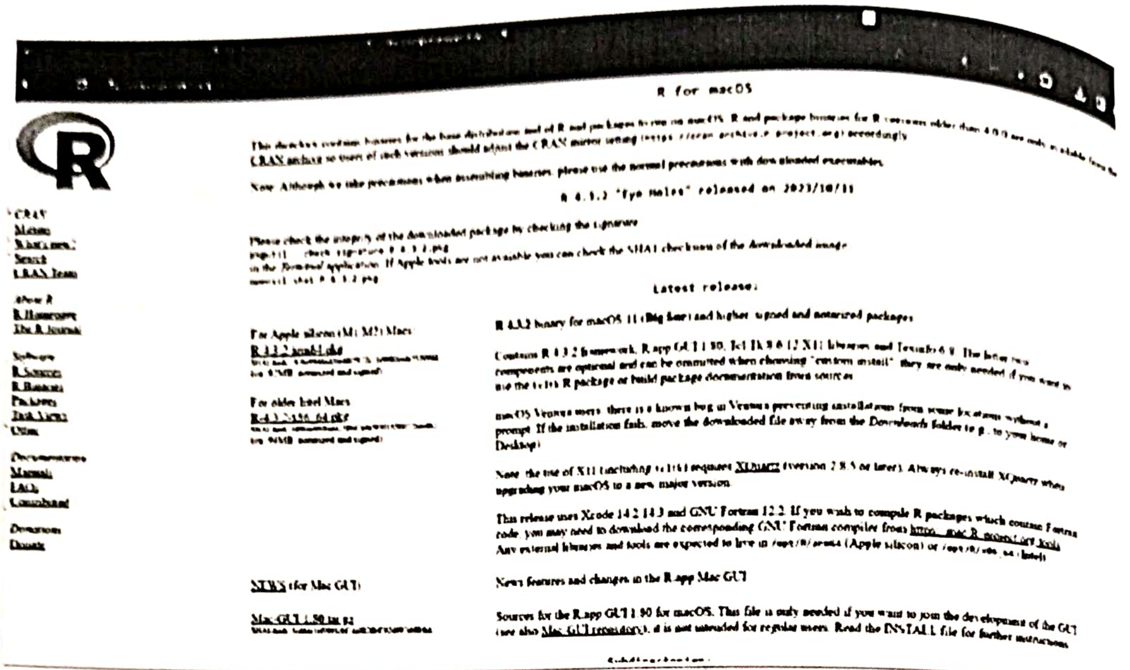


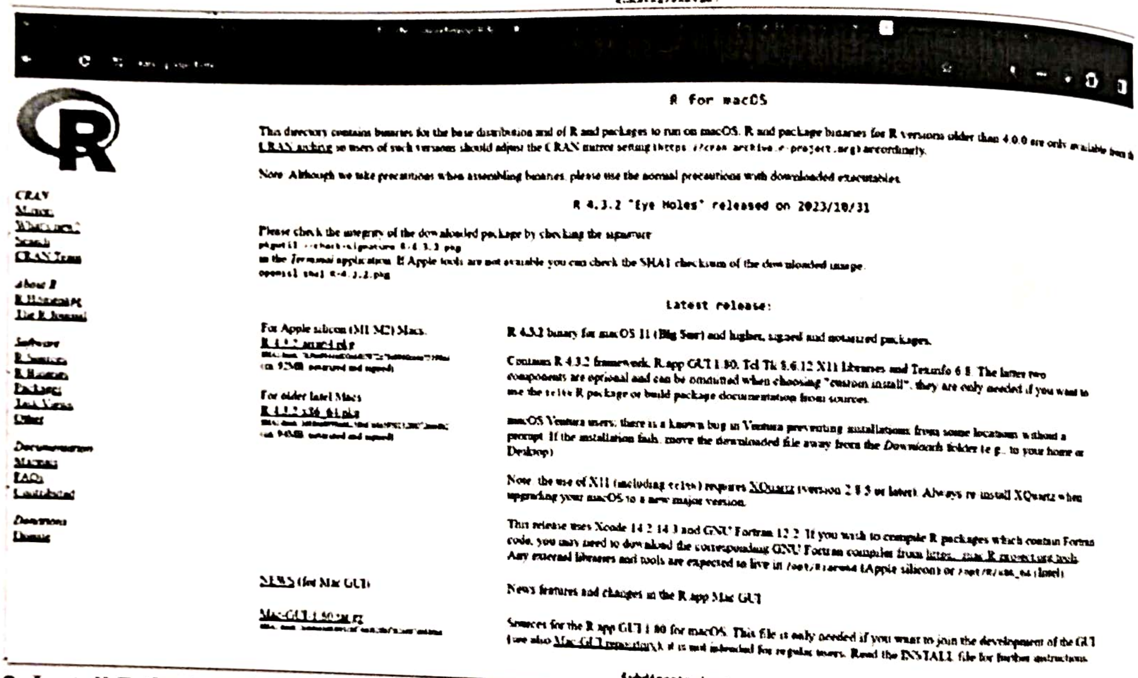
2. Click Download R for Windows.



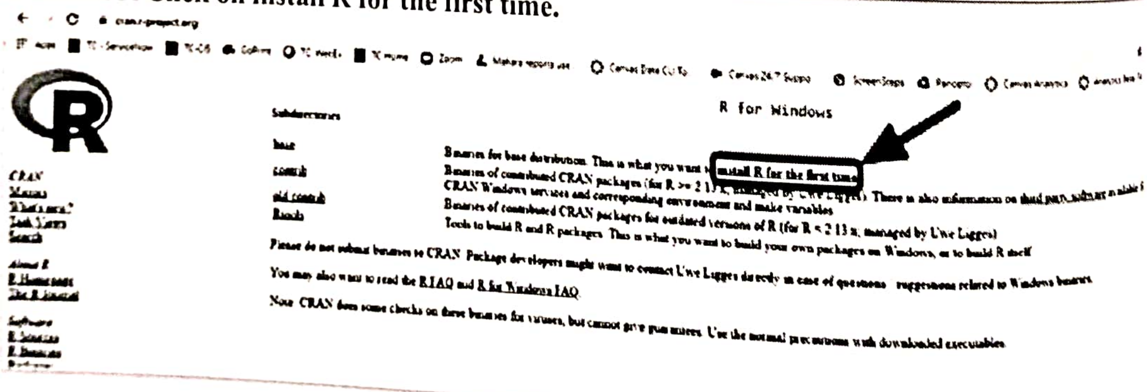




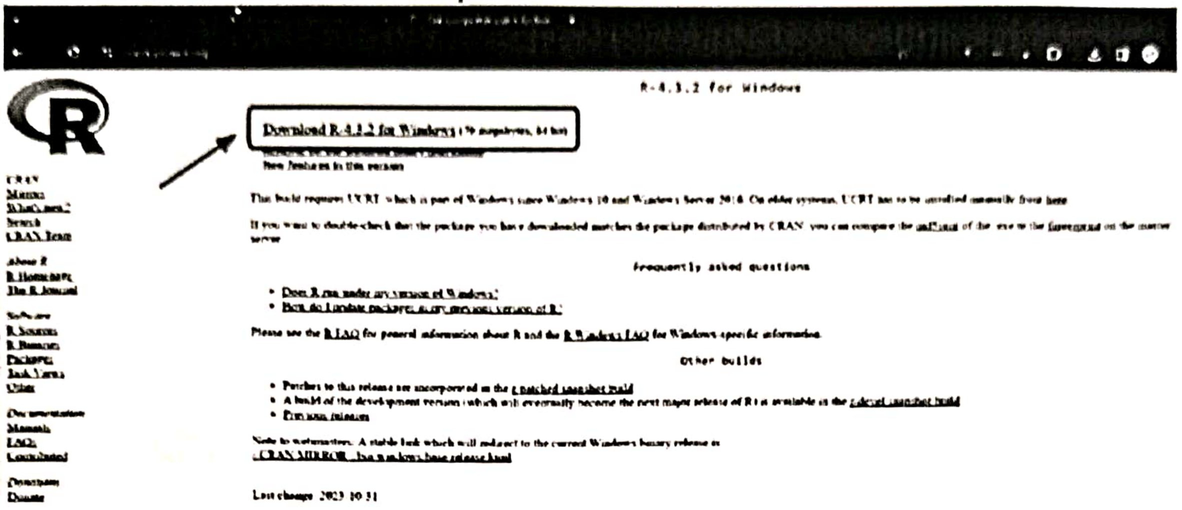




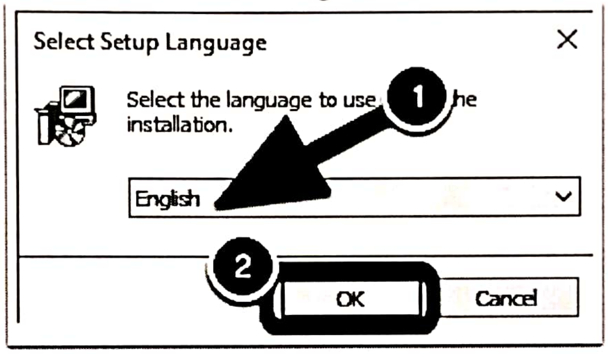
3. Install R Click on install R for the first time.



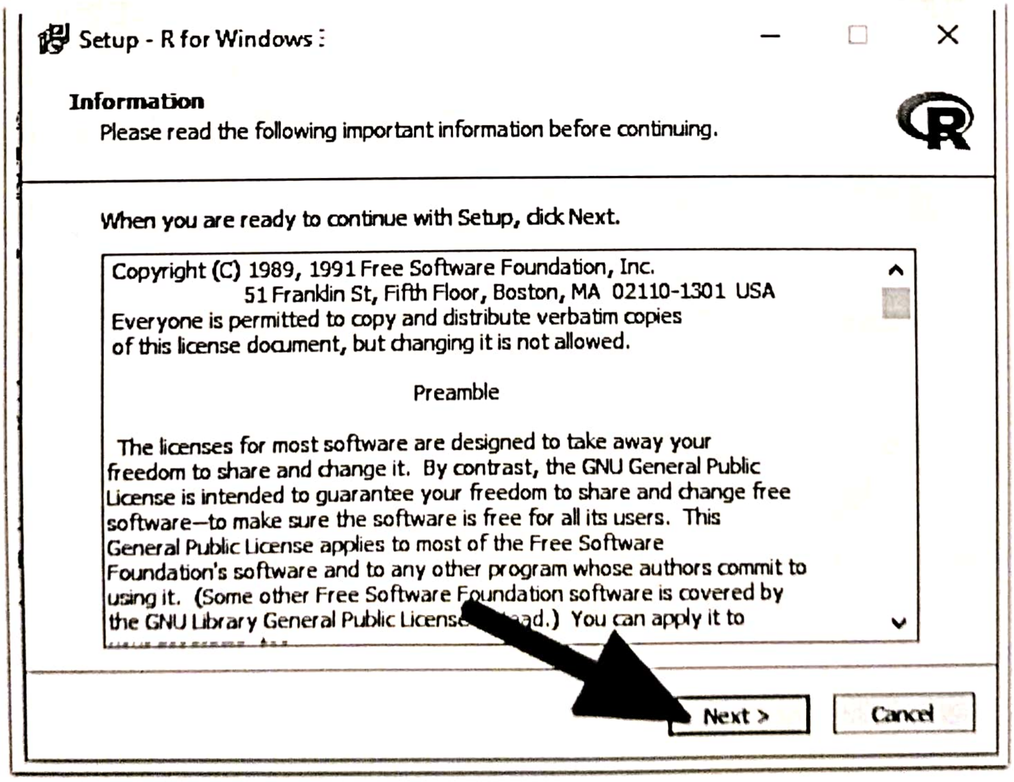
4. Click Download R for Windows. Open the downloaded file.



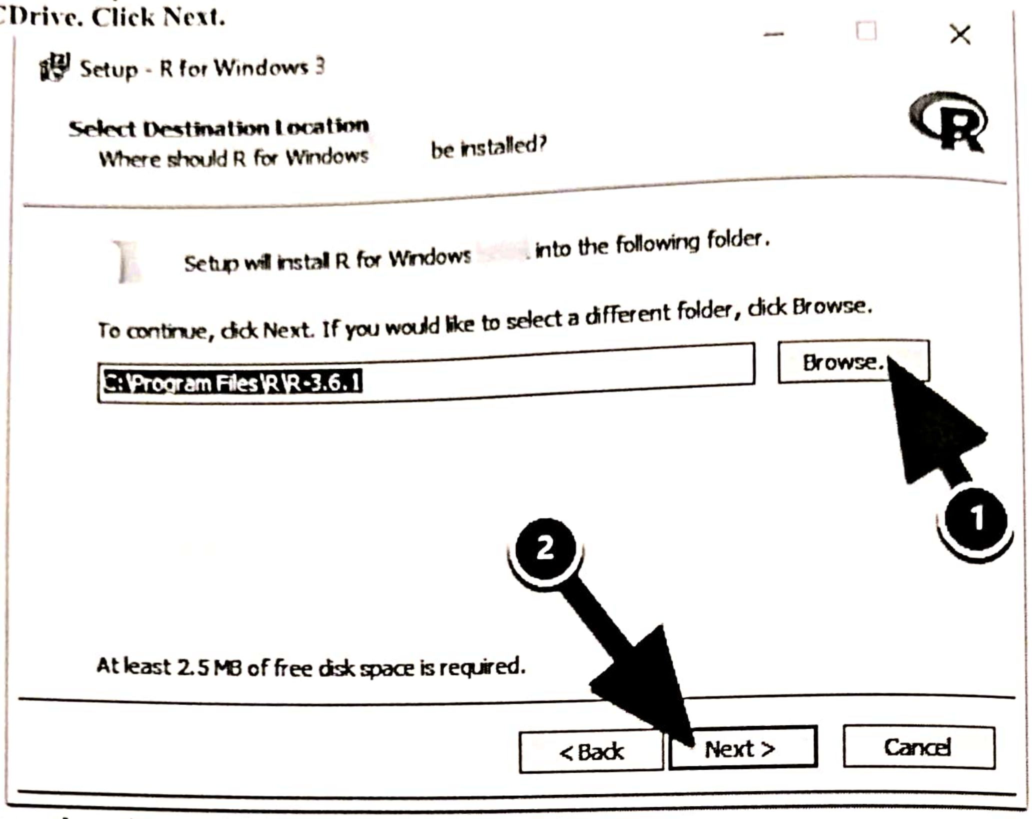
5. Select the language you would like to use during the installation. Then click OK.



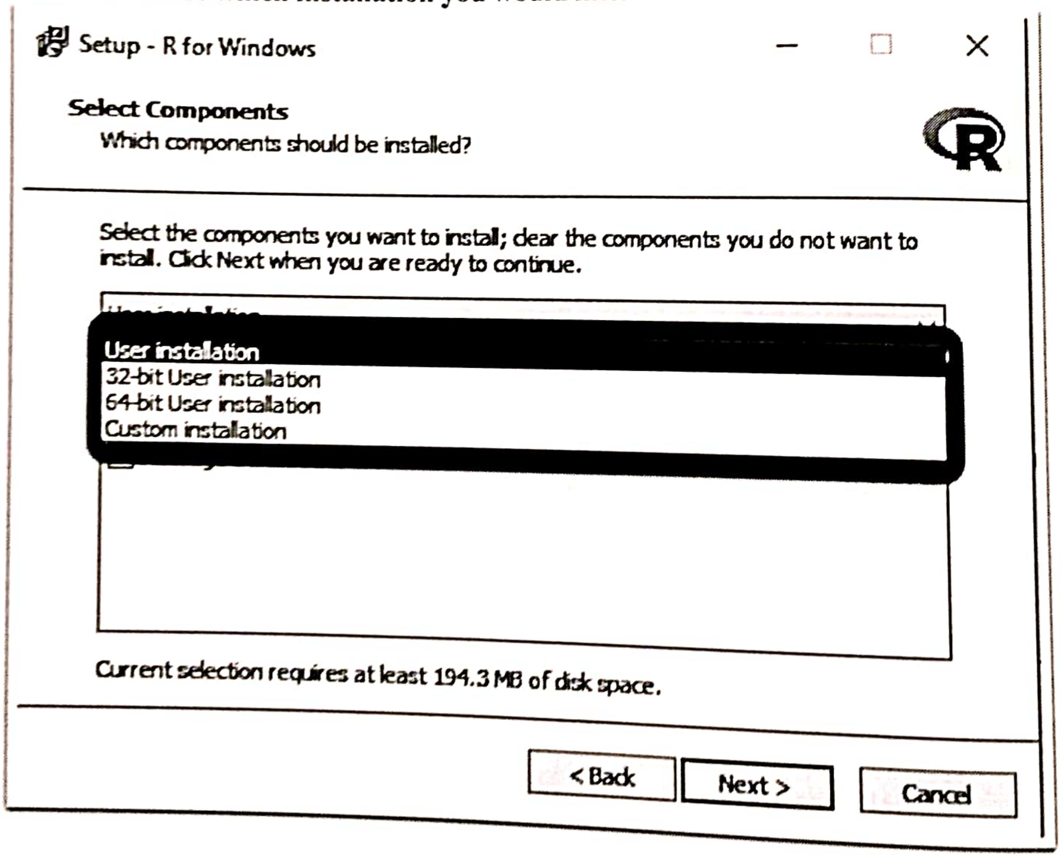
6. Click Next.



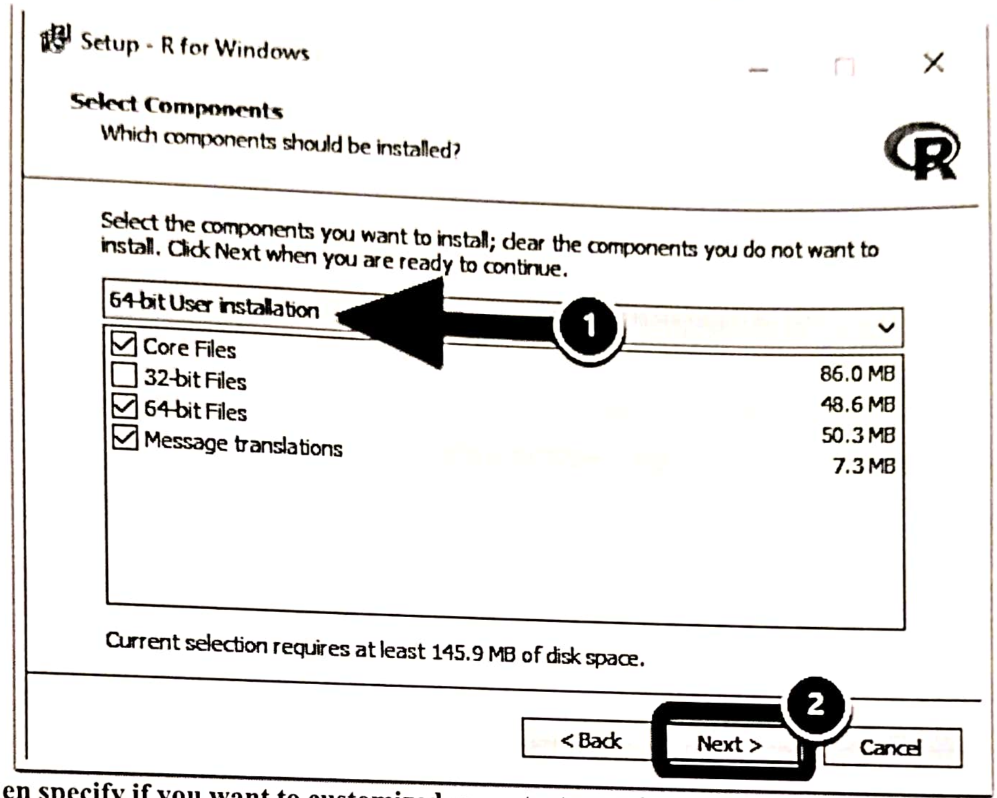
7. **Select where you would like R to be installed. It will default to your Program Files on your CDrive. Click Next.**



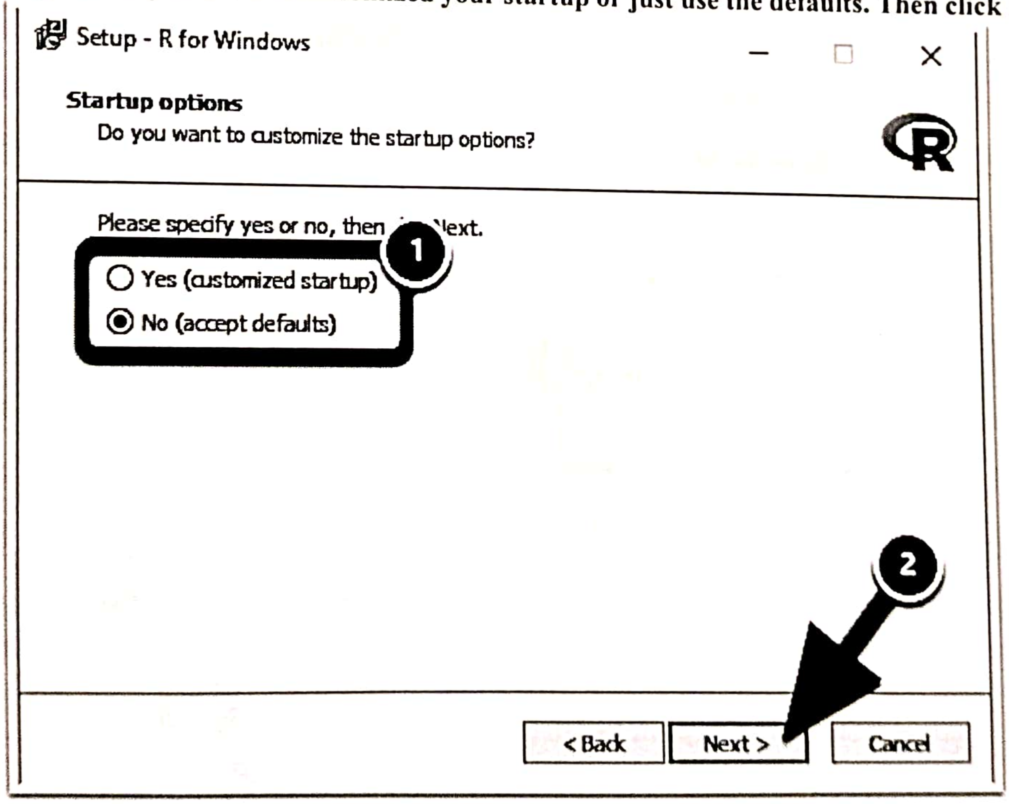
8. **You can then choose which installation you would like.**



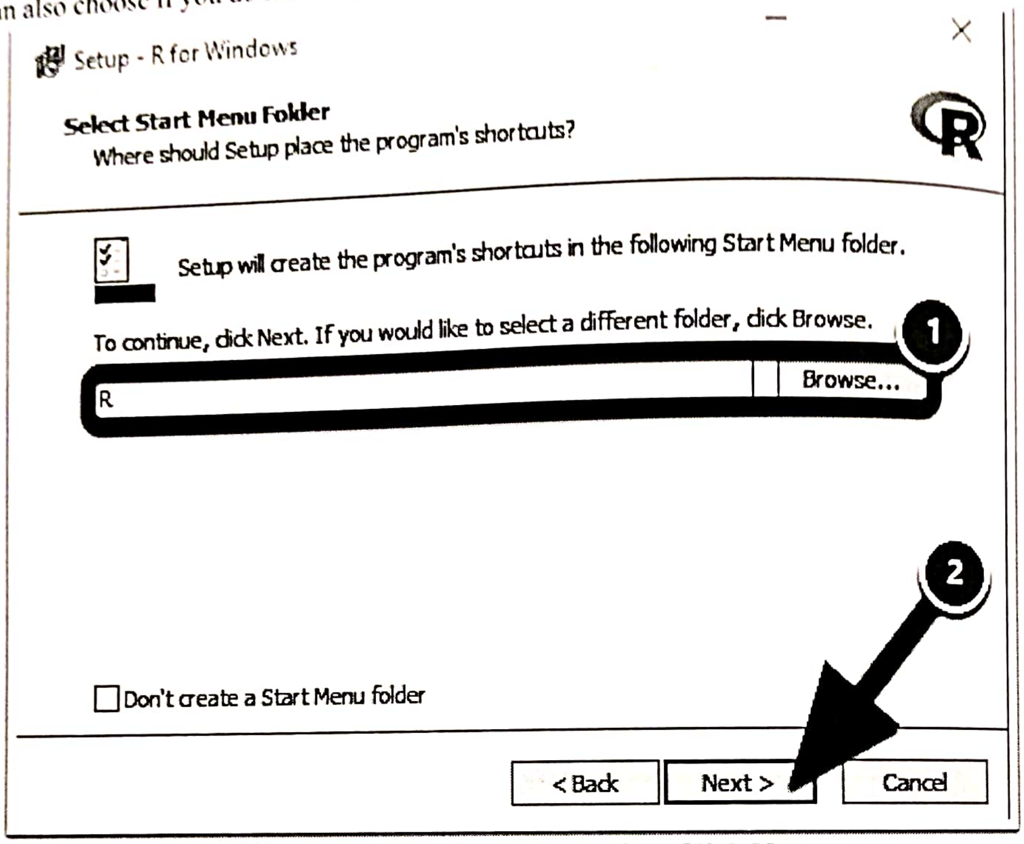
9. (Optional) If your computer is a 64-bit, you can choose the 64-bit User Installation. Then click Next.



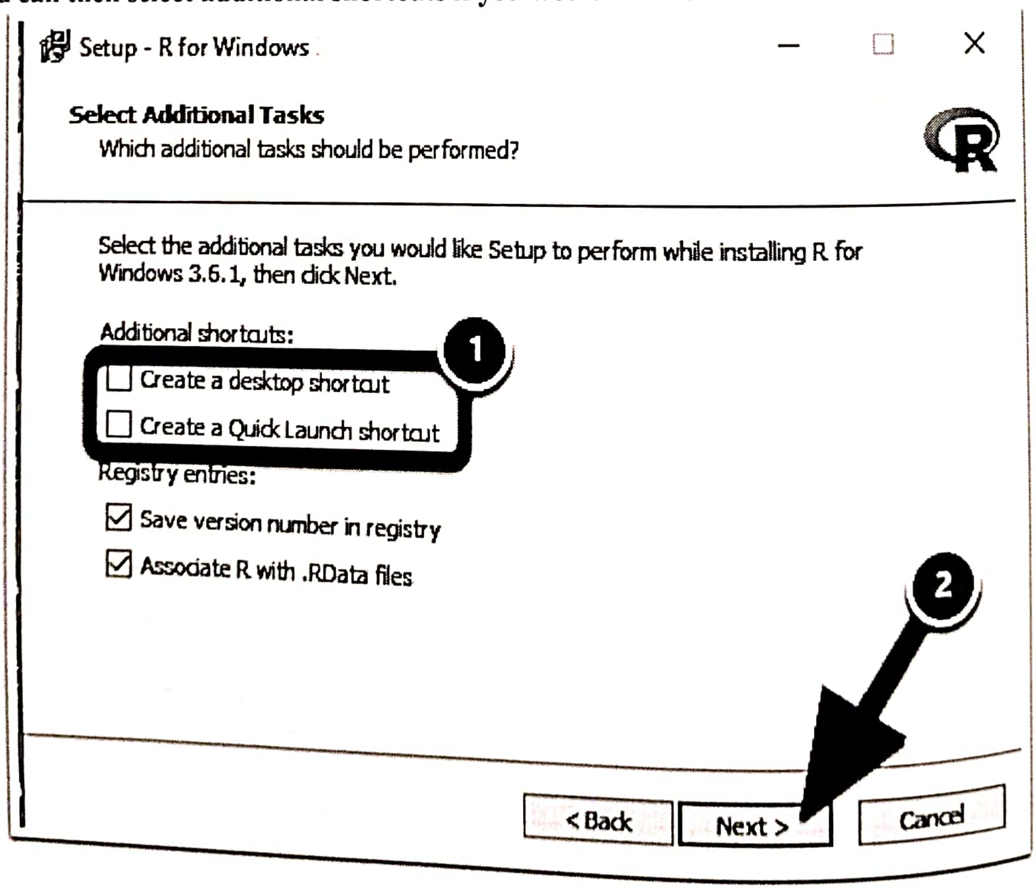
10. Then specify if you want to customized your startup or just use the defaults. Then click Next.



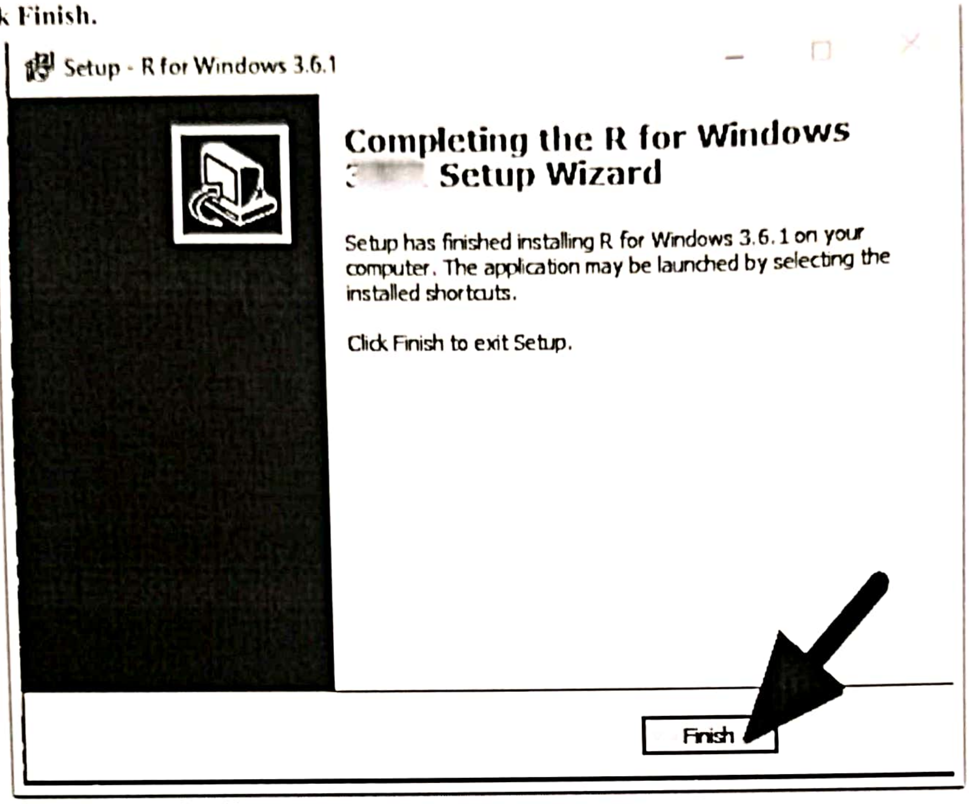
11. Then you can choose the folder that you want R to be saved within or the default if the R folder that was created. Once you have finished, click Next.
You can also choose if you do not want a Start Menu folder at the bottom.



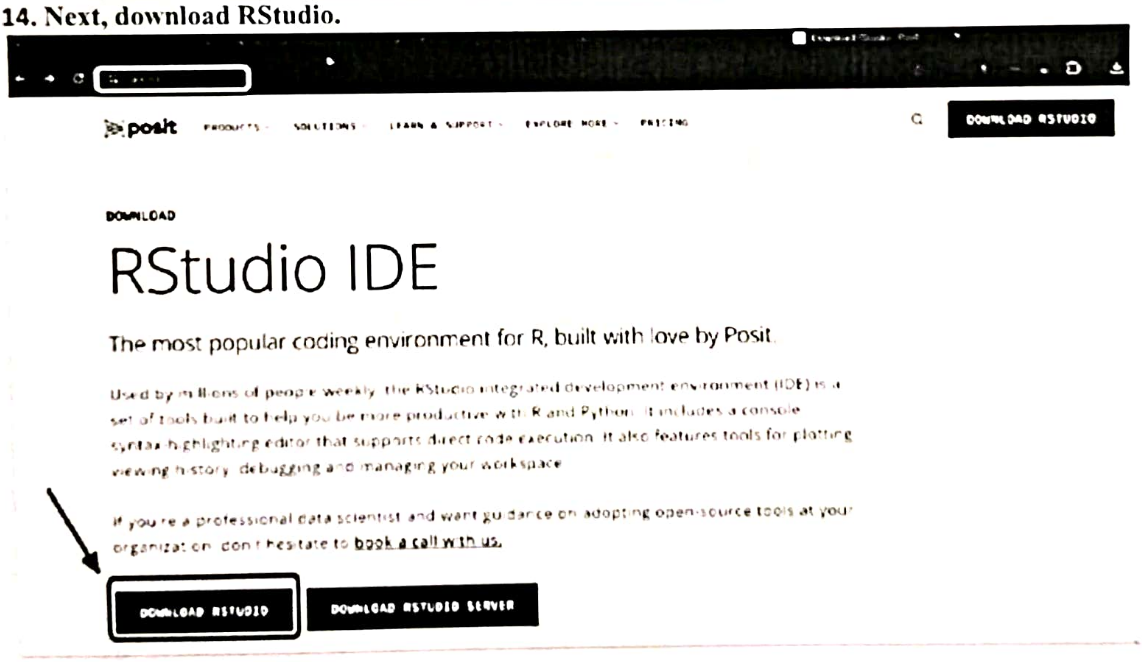
12. You can then select additional shortcuts if you would like. Click Next.



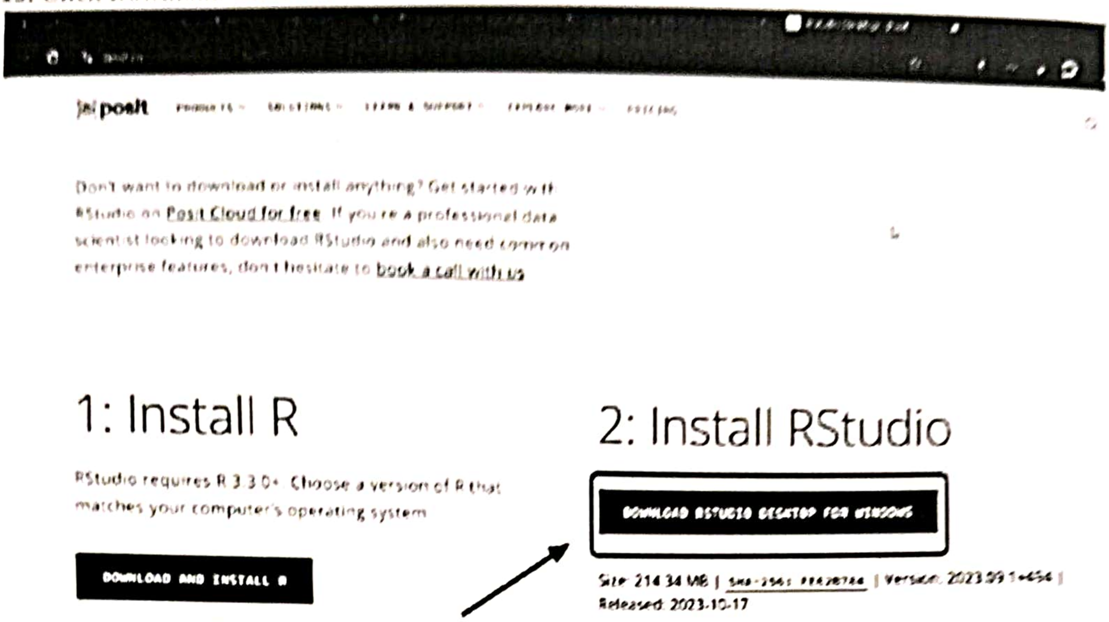
13. Click Finish.



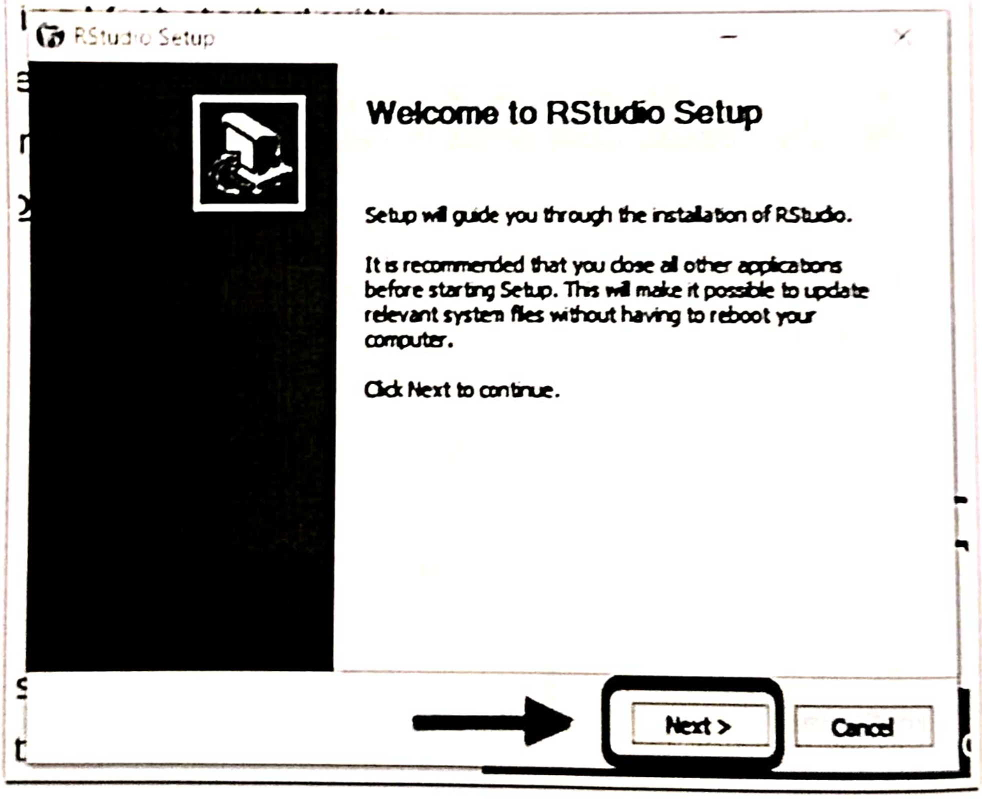
14. Next, download RStudio.



15. Click Download RStudio.



16. Once the packet has downloaded, the Welcome to RStudio Setup Wizard will open. Click Nextand go through the installation steps



17. After the Setup Wizard finishing the installation,

B.The basics of R syntax,workspace.

a.= (Simple Assignnt)

b.<- (Leftward Assignment) c.-> (Rightward Assignment)

Example:

```r
var1 = "Simple Assignment"

var2 <- "Leftward Assignment!"

"Rightward Assignment" ->  var3

print(var1)
print(var2)
print(var3)
```

Output:

“Simple Assignment" "Leftward Assignment!""Rightward Assignment"

2. **Data Types**: R has several data types, including numeric, character, logical, integer, and complex.
3. **Vectors**: Create vectors using **c()**. For example, **v <- c(1, 2, 3).**
4. **Lists**: Create lists using **list()**. For example, **mylist <- list(a = 1, b = "hello", c = TRUE).**
5. **Data Frames**: Create data frames using **data.frame()**. For example, **df <- data.frame(x = 1:3, y = c("a", "b", "c")).**
6. **Functions**: Define functions using **function()**. For example, **myfunc <- function(x) { x * 2 }.**
7. **Control Structures**: Use control structures like **ifelse, for,** and **while** for flow control.
8. *Comments in R*

comments are a way to improve your code’s readability and are only meant for the user so the interpreter ignores it. Only single-line comments are available in R but we can also use multiline comments by using a simple trick which is shown below. Single line comments can be written by using # at the beginning of the statement.

Example:

```
print("This is fun!")

(FALSE)
{
  "This is multi-line comment which should be put inside either a
      single or a double quote"
}
```

Output:

[1] "This is fun!"

Regarding workspace management, R uses the concept of a workspace or environment to store objects (variables, functions, etc.). You can manage your workspace using the following commands:

- **ls()** : List objects in the workspace.
- **rm(x)**: Remove object **x** from the workspace.
- **rm(list = ls())** : Remove all objects from the workspace.
- **save.image("filename.RData")** : Save the current workspace to a file.

C.Matrices and lists

Matrices

A matrix is a two dimensional data set with columns and rows.

A column is a vertical representation of data, while a row is a horizontal representation of data.

A matrix can be created with the matrix() function. Specify the nrow and ncol parameters to get the amount of rows and columns:

Example:

```
# Create a matrix
thismatrix <- matrix(c(1,2,3,4,5,6), nrow = 3, ncol = 2)

# Print the matrix
thismatrix
```

Output:

```
     [,1] [,2]
[1,]    1    4
[2,]    2    5
[3,]    3    6
```

Note: Remember the c()function is used to concatenate items together.

Access Matrix Items

You can access the items by using [ ]brackets. The first number "1" in thebracket specifies the row-position, while the second number "2" specifies the column-position:

Example

thismatrix <- matrix(c("apple", "banana", "cherry", "orange"), nrow = 2, ncol = 2)

thismatrix[1, 2]

output:

```
[1] "cherry"
```

Add Rows and Columns

Use the cbind() function to add additional columns in a Matrix:

Example

Thismatrix=matrix(c("apple", "banana", "cherry", "orange","grape", "pineapple", "pear", "melon", "fig"), nrow = 3, ncol = 3)

newmatrix=cbind(thismatrix,c("strawberry", "blueberry", "raspberry"))# Print the new matrix newmatrix

Output:

```
     [,1]     [,2]        [,3]    [,4]
[1,] "apple"  "orange"    "pear"  "strawberry"
[2,] "banana" "grape"     "melon" "blueberry"
[3,] "cherry" "pineapple" "fig"   "raspberry"
```

rbind() function to add additional rows in a Matrix.

To find out if a specified item is present in a matrix, use the %in% operator.

dim() function to find the number of rows and columns in a Matrix. length()

function to find the dimension of a Matrix.

LISTS:

Lists are the R objects which contain elements of different types like – numbers, strings, vectors and another list inside it. A list can also contain a matrix or a function as its elements. List is created using **list()** function.

Creating a list:

Example

```
# List of strings
thislist <- list("apple", "banana", "cherry")

# Print the list
thislist
```

Output:

```
[[1]]
[1] "apple"

[[2]]
[1] "banana"

[[3]]
[1] "cherry"
```

Access Lists items:

You can access the list items by referring to its index number, inside brackets. The first item has index 1, the second item has index 2, and so on.

Example:

```
thislist <- list("apple", "banana", "cherry")

thislist[1]
```

output:

```
[[1]]
[1] "apple"
```

Change Item Value

To change the value of a specific item, refer to the index number:

Example

thislist <- list("apple", "banana", "cherry")thislist[1] <-
"blackcurrant"

Print the updated listthislist

output:

```
[[1]]
[1] "blackcurrant"

[[2]]
[1] "banana"

[[3]]
[1] "cherry"
```

d) subsetting:

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. There are 4 ways of subsetting in R programming. Each of the methods depends on the usability of the user and the type of object.

Method 1: Subsetting in R Using [ ] Operator

Using the '[ ]' operator, elements of vectors and observations from data frames can be accessed. To neglect some indexes, '-' is used to access all other indexes of vector or data frame.

EXAMPLE:

```
# Create vector
x <- 1:15
# Print vector
cat("Original vector: ", x, "\n")
# Subsetting vector
cat("First 5 values of vector", x[1:5], "\n")
cat("Without values present at index 1, 2 and 3: ", x[-c(1, 2, 3)], "\n")
```

Output:

```
Original vector:  1 2 3 4 5 6 7 8 9 10 11 12 13 14
First 5 values of vector:  1 2 3 4 5
Without values present at index 1, 2 and 3
```

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 selects more than 1.

EXAMPLE:

```
# Create list
ls <- list(a = 1, b = 2, c = 10, d = 20)
# Print list
cat("Original List: \n")
print(ls)
# Select first element of list
cat("First element of list: ", ls[[1]], "\n")
```

OUTPUT:

```
Original list:

[1] 1

$b

1] 2

c

1] 10
```

Method 3: Subsetting in R Using $ Operator:

$ operator can be used for lists and data frames in R. Unlike [ ] operator, it selects only a singleobservation at a time. It can be used to access an element in named list or a column in data frame.
$ operator is only applicable for recursive objects or list-like objects.

Example:

```
# Create list
ls <- list(a = 1, b = 2, c = "Hello", d =
"CSE")cat("Original list:\n")
print(ls)
cat("Using $ operator:\n")
print(ls$d)
```

OUTPUT:

Original list:

$a

[1] 1

$b

[1] 2

$c

[1] "Hello"

$d

[1] "CSE"

Using $ operator:

[1] "CSE"

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 based on the conditions provided in the parameters.
Syntax: *subset(x, subset, select)*
Parameters:

- ***x:*** *indicates the object*
- ***subset:*** *indicates the logical expression on the basis of which subsetting has to be done*
- ***select:*** *indicates columns to select.*

e) system defined functions;the help sysem:

A function is a set of statements organized together to perform a specific task. R has a large number of in-built functions and the user can create their own functions.In R, a function is an object so the R interpreter is able to pass control to the function, along with arguments that may be necessary for the function to accomplish the actions.

Function Definition:

An R function is created by using the keyword **function**.Syntax:

```
function_name <- function(arg_1, arg_2, ...) {
 Function body
}
```

Function Components

The different parts of a function are −

- **Function Name** − This is the actual name of the function. It is stored in R environment as an object with this name.
- **Arguments** − An argument is a placeholder. When a function is invoked, you pass a value to the argument. Arguments are optional; that is, a function may contain no arguments. Also arguments can have default values.

Function Body – The function body contains a collection of statements that defines what the 0function does.

Return Value – The return value of a function is the last expression in the function body to be evaluated.

R has many **in-built** functions which can be directly called in the program without defining them first. We can also create and use our own functions referred as **user defined** functions.

Built-in Function:

Simple examples of in-built functions are **seq()**, **mean()**, **max()**, **sum(x)** and **paste(...)** etc. They aredirectly called by user written programs.

Example:

```
# Create a sequence of numbers from 32 to 44.
print(seq(32,44))

# Find mean of numbers from 25 to 82.
print(mean(25:82))

# Find sum of numbers from41 to 68.
print(sum(41:68))
```

output:

[1] 32 33 34 35 36 37 38 39 40 41 42 43 44

[1] 53.5

[1] 1526

User-defined Function:

We can create user-defined functions in R. They are specific to what a user wants and once created they can be used like the built-in functions.

The Help system:

The **help() function** in R is used to get help on any given R function passed to

it.Syntax:

help(function name)

we'll use the help function to get help on the following R functions:

- `eval()` function
- `dump()` function

we use the help() function to provide the official documentation page for the R function eval().

Now, let's use the help() function to get help on the dump() function. In other words, we use the help() function to provide the official documentation page for the R function dump.

f) Errors and warnings;coherence of the workspace:

Handling Errors in R

Error Handling is a process in which we deal with unwanted or anomalous errors which may cause abnormal termination of the program during its execution. In R there are basically two ways in which we can implement an error handling mechanism. Either we can directly call the functions like stop() or warning(), or we can use the error options such as "warn" or "warning.expression". The basic functions that one can use for error handling in the code :

- **stop(...)**: It halts the evaluation of the current statement and generates a message argument. The control is returned to the top level.
- **waiting(...)**: Its evaluation depends on the value of the error option warn. If the value of the warning is negative then it is ignored. In case the value is 0 (zero) they are stored and printed only after the top-level function completes its execution. If the value is 1 (one) then it is printed as soon as it has been encountered while if the value is 2 (two) then immediately the generated warning is converted into an error.
- **tryCatch(...)**: It helps to evaluate the code and assign the exceptions.

Example:

```
# Function to divide two numbers

divide <- function(x, y) {

 if (y == 0) {

  stop("Division by zero is not

 allowed.")}return(x/y)}

# Error handling

result <- tryCatch(

 { divide(6, 0)

 }

 error = function(e) {

  message("An error occurred: ", e$message)
```

```
    return(NA)
  } finally = {
    message("Division attempt completed.")
  }
)print(result)
```

OUTPUT:

An error occurred: Division by zero is not

allowedNA

WEEK – 2:Getting used to R :Describing Data:

a)Viewing and manipulating data

Viewing data:

Data can be viewed in R using data frames, a data structure that supports the storage and view of data in the form of tables. It is used to manage data in a concise manner i.e., in a tabular form enabling easy understandability and readability.

To view data, we must load the data in to a data frame. That is done by the code snippet given below.

```
name <- c(     ,     ,     ,     )
age <- c( , , , )
gender <- c(     ,     ,     ,     )
weight <- c( , , , )
height <- c( , , , )
```

```
data <- data.frame(Name=name, Age=age, Gender=gender, Weight=weight, Height=height)
```

Let's view the data that is present in the data frame "data"

```
   Name Age Gender Weight
   John       Male
  Alice     Female
    Bob       Male
  Emily     Female
```

Manipulating the data:

To manipulate the data of a data frame we access the elements of it and assign desired values to it. That is done by the code snippet below.

```
data[ ,      ] <-
```

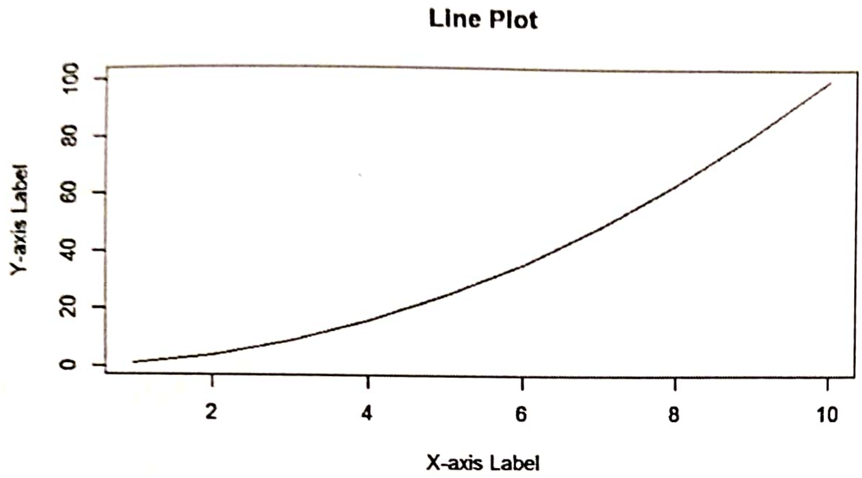
Output after manipulating:

```
   Name Age Gender Weight
   John       Male
  Alice     Female
    Bob       Male
  Emily     Female
```

b)Plotting data:

In R, the data can be plotted by using plot() function. This is demonstrated by the code snippet below.

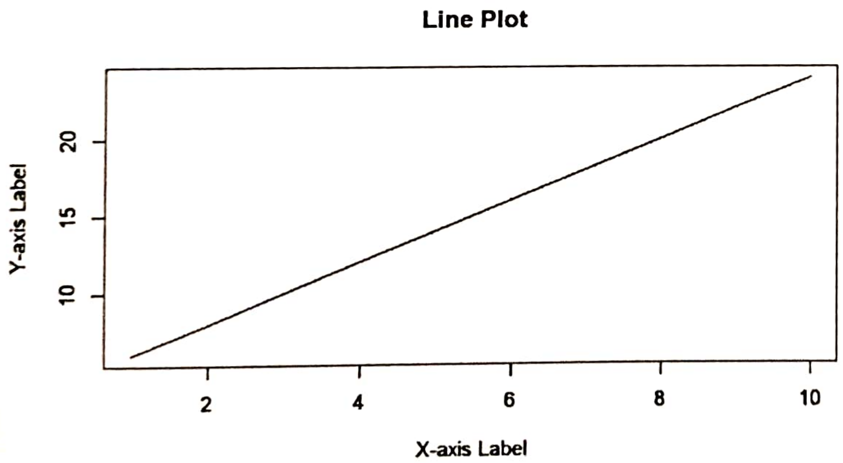
```
x<- :
y<-x^
plot(x,y,type=   ,main=              ,xlab=              ,ylab=              ,col=    )
```



Another example for plotting the values:

Loading data,

```
x<-seq(-  ,  ,by=  )
y<- *x+
plot(x,y,type=   ,main=              ,xlab=              ,ylab=              )
```



c)Reading data from console, file (.csv) local disk and web.

Reading data from console:

Now, let's see how the data can be entered using the **console**.

This can be done by using the readline () function in R. It is demonstrated by the code snippet below. It accepts a prompt that has to be displayed when the input is being taken from the console.

Syntax: var1<- readline ("Prompt to be

displayed")String input:

```
> var1<- readline("Enter the data: ")
Enter the data: Hello, This is R Programming.
> print(var1)
[1] "Hello, This is R Programming."
> typeof(var1)
[1] "character"
```

Integer input:

This can be implemented in R using as.integer () function.

```
> var1<- as.integer(readline("Enter the data: "))
Enter the data: 21
> print(var1)
[1] 21
> typeof(var1)
[1] "integer"
```

Character input:

This can be implemented by using as.character () function.

```
> var1<- as.character(readline("Enter the data: "))
Enter the data: Hello, This is R Programming.
> print(var1)
[1] "Hello, This is R Programming."
> typeof(var1)
[1] "character"
```

Reading data from a csv file:

Let's load the data from a **csv file**. This can be done by using the function [read.csv ()] which accepts the name of the file as a parameter.

Let the CSV file be:

```
Name,Age,Gender,Weight,
John,  ,Male,  ,
Alice,  ,Female,  ,
Bob,  ,Male,  ,
Emily,  ,Female,  ,
```

Let's load this data in to a data frame named data.

```
data <- read.csv(        )
```

After loading the data, we need to use print "data" to view the contents of it.

```
print(data)
```

The output of the above code snippet would be,

```
   Name Age Gender Weight
   John       Male
  Alice     Female
    Bob       Male
  Emily     Female
```

Reading data from the web:

From web, we can directly load the csv files in to a variable and hence by doing it, we read and display the data using read.csv () and print () functions.

This is implemented by the code snippet below:

```
url <- "https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv"
data <- read.csv(url)
(head(data))
```

Output:

```
> url <- "https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv"
> data <- read.csv(url)
> head(data)
  PassengerId Survived Pclass
1           1        0      3                                                Name    Sex Age
2           2        1      1                             Braund, Mr. Owen Harris   male  22
3           3        1      3 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female  38
4           4        1      1                              Heikkinen, Miss. Laina female  26
5           5        0      3        Futrelle, Mrs. Jacques Heath (Lily May Peel) female  35
6           6        0      3                            Allen, Mr. William Henry   male  35
  SibSp Parch           Ticket    Fare Cabin Embarked          Moran, Mr. James   male  NA
1     1     0        A/5 21171  7.2500              S
2     1     0         PC 17599 71.2833   C85        C
3     0     0 STON/O2. 3101282  7.9250              S
4     1     0           113803 53.1000  C123        S
5     0     0           373450  8.0500              S
6     0     0           330877  8.4583              Q
```

d) Working with the larger datasets

Let's take the dataset we considered above as a large dataset.

```
url <- "https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv"
data <- read.csv(url)
head(data)
```

Let's take a look at the first part of the file.

```
> print(head(data))
  PassengerId Survived Pclass                                                Name    Sex Age
1           1        0      3                             Braund, Mr. Owen Harris   male  22
2           2        1      1 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female  38
3           3        1      3                              Heikkinen, Miss. Laina female  26
4           4        1      1        Futrelle, Mrs. Jacques Heath (Lily May Peel) female  35
5           5        0      3                            Allen, Mr. William Henry   male  35
6           6        0      3                                    Moran, Mr. James   male  NA
  SibSp Parch           Ticket    Fare Cabin Embarked
1     1     0        A/5 21171  7.2500              S
2     1     0         PC 17599 71.2833   C85        C
3     0     0 STON/O2. 3101282  7.9250              S
4     1     0           113803 53.1000  C123        S
5     0     0           373450  8.0500              S
6     0     0           330877  8.4583              Q
```

To find the number of rows in this dataset:

```
> print(nrow(data))
[1] 891
```

To find the dimensions of the dataset:

```
> print(dim(data))
[1] 891  12
```

This shows that there are 891 rows and 12 columns in this dataset. To know the column names of the dataset:

```
> print(names(data))
 [1] "PassengerId" "Survived"    "Pclass"      "Name"        "Sex"         "Age"
 [7] "SibSp"       "Parch"       "Ticket"      "Fare"        "Cabin"       "Embarked"
```

To view a particular part of a dataset:

To display the first 10 rows and 3 columns of the dataset.

```
> print(data[1:10,1:3])
   PassengerId Survived Pclass
1            1        0      3
2            2        1      1
3            3        1      3
4            4        1      1
5            5        0      3
6            6        0      3
7            7        0      1
8            8        0      3
9            9        1      3
10          10        1      2
```

d) Working with the larger datasets

Let's take the dataset we considered above as a large dataset.

```
url <- "https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv"
data <- read.csv(url)
head(data)
```

Let's take a look at the first part of the file.

```
> print(head(data))
  PassengerId Survived Pclass                                                Name    Sex Age
1           1        0      3                             Braund, Mr. Owen Harris   male  22
2           2        1      1 Cumings, Mrs. John Bradley (Florence Briggs Thayer) female  38
3           3        1      3                              Heikkinen, Miss. Laina female  26
4           4        1      1        Futrelle, Mrs. Jacques Heath (Lily May Peel) female  35
5           5        0      3                            Allen, Mr. William Henry   male  35
6           6        0      3                                    Moran, Mr. James   male  NA
  SibSp Parch           Ticket    Fare Cabin Embarked
1     1     0        A/5 21171  7.2500              S
2     1     0         PC 17599 71.2833   C85        C
3     0     0 STON/O2. 3101282  7.9250              S
4     1     0           113803 53.1000  C123        S
5     0     0           373450  8.0500              S
6     0     0           330877  8.4583              Q
```

<u>To find the number of rows in this dataset:</u>

```
> print(nrow(data))
[1] 891
```

<u>To find the dimensions of the dataset:</u>

```
> print(dim(data))
[1] 891  12
```

This shows that there are 891 rows and 12 columns in this dataset.<u>To know the column names of the dataset:</u>

```
> print(names(data))
 [1] "PassengerId" "Survived"    "Pclass"      "Name"        "Sex"         "Age"
 [7] "SibSp"       "Parch"       "Ticket"      "Fare"        "Cabin"       "Embarked"
```

<u>To view a particular part of a dataset:</u>

To display the first 10 rows and 3 columns of the dataset.

```
> print(data[1:10,1:3])
   PassengerId Survived Pclass
1            1        0      3
2            2        1      1
3            3        1      3
4            4        1      1
5            5        0      3
6            6        0      3
7            7        0      1
8            8        0      3
9            9        1      3
10          10        1      2
```

Week-3:shape of data and describing relationships:

b) Univariate data, measures of central tendency, frequency distributions, variation, shape

Univariate analysis explores each variable in a data set, separately. It looks atthe range of values, as well as the central tendency of the values. It describes the pattern of responseto the variable. It describes each variable on its own.

Key points in Univariate analysis:

1. **No Relationships** : Univariate analysis focuses solely on describing and summarizing the distribution of the single variable. It does not explore relationships between variables or attempt to identify causes.
2. **Descriptive Statistics** : Descriptive statistics, such as measures of central tendency (mean, median, mode) and measures of dispersion (range, standard deviation), are commonly used in the analysis of univariate data.
3. **Visualization** : Histograms, box plots, and other graphical representations are often used to visually represent the distribution of the single variable.

R Code:

```
# Load the iris dataset

data(iris)

# Display the first few rows of the dataset

cat("First few rows of the iris dataset:\n")

print(head(iris))

# Calculate measures of central tendency

mean_value <- apply(iris[, 1:4], 2, mean)  # Calculate mean for each column

median_value <- apply(iris[, 1:4], 2, median)  # Calculate median for each column

mode_value <- apply(iris[, 1:4], 2, function(x) { names(table(x))[table(x) == max(table(x))]

})# Calculate mode for each column

# Print measures of central tendency

print("Measures of Central Tendency:")

print(paste("Mean:", mean_value))

print(paste("Median:", median_value))

print(paste("Mode:", mode_value))

# Calculate variance

variance_value <- apply(iris[, 1:4], 2, var)  # Calculate variance for each column

# Print variance

print("Variance:")
```

```
print(variance_value)
# Plotting histograms for each variable
par(mfrow = c(2, 2)) # Arrange plots in a 2x2
gridfor (i in 1:4) {
 hist(iris[, i], main = paste("Histogram of", names(iris)[i]), xlab = names(iris)[i], col = "lightblue")
}
```

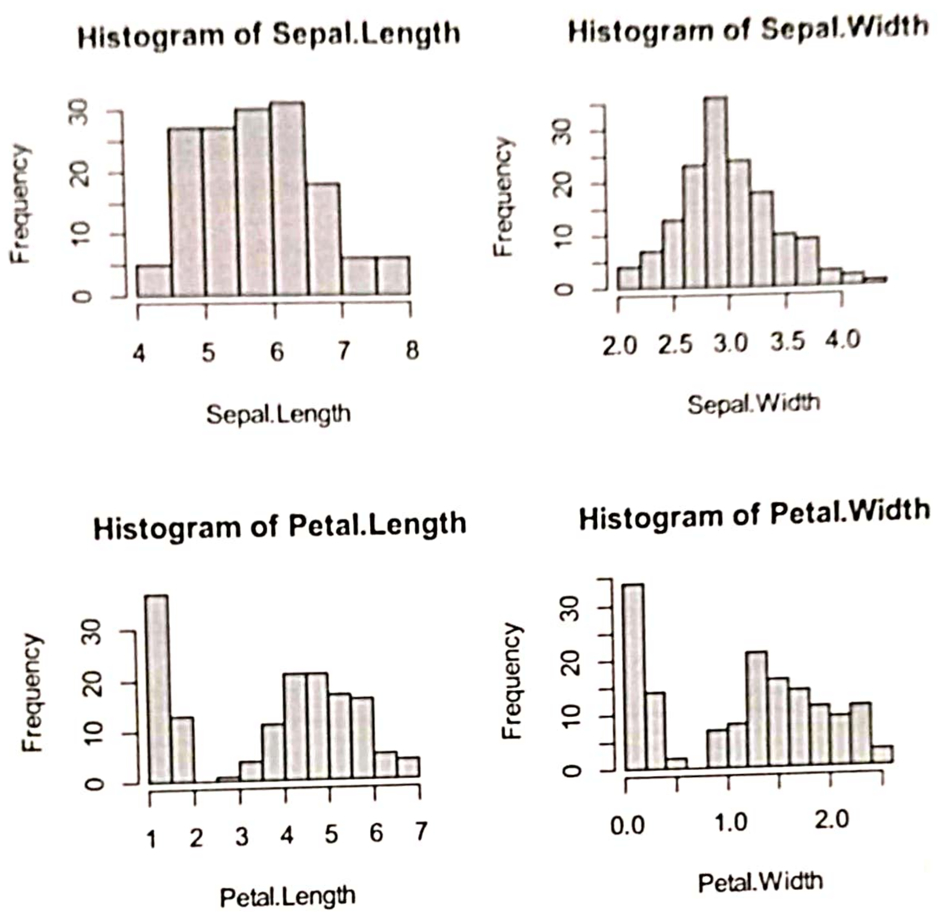
output :

First few rows of the iris dataset:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

Measures of Central Tendency:

```
[1] "Mean: 5.84333333333333"    "Mean: 3.05733333333333"
[3] "Mean: 3.75800000000000"    "Mean: 1.19933333333333"
[5] "Median: 5.8"               "Median: 3"
[7] "Median: 4.35"              "Median: 1.3"
[9] "Mode: 5.1"                 "Mode: 3"
[11] "Mode: 1.5"                "Mode: 0.2"
[13] "Variance:"
Sepal.Length Sepal.Width Petal.Length Petal.Width
   0.6856935   0.1899794    3.1162779   0.5810063
```



c)Multivariate data, relationships between a categorical and a continuous variable.

Multivariate data refers to datasets where each observation or sample point consists of multiple variables or features. These variables can represent different aspects, characteristics, or measurements related to the observed phenomenon. When dealing with three or more variables, thedata is specifically categorized as multivariate.

Analyzing multivariate data involving a categorical variable and a continuous variable means understanding the relationship between these two types of variables. There are several statistical techniques one can employ for this analysis:

1. Descriptive Statistics: Start by examining summary statistics for the continuous variable within each category of the categorical variable. This can include measures like means, medians, standard deviations, and ranges.

2. Boxplots: Create boxplots for the continuous variable, with each boxplot corresponding to a different category of the categorical variable. This visual representation allows you to compare the distributionsof the continuous variable across categories.

3. ANOVA (Analysis of Variance): If you have multiple categories in your categorical variable, you can use ANOVA to test whether there are statistically significant differences in the means of the continuousvariable across these categories. ANOVA assesses whether there are significant variations between groups while accounting for variations within groups.

4. Correlation Analysis: While correlation is typically used for assessing relationships between two continuous variables, you can also calculate correlations between a continuous and a binary categorical variable using techniques like point-biserial correlation or rank-biserial correlation. Thesemeasures indicate the strength and direction of the relationship between the two variables.

R Code :

```
# Load necessary libraries
library(ggplot2)
library(dplyr)
# View the first few rows of the mtcars dataset
head(mtcars)
# Summary statistics for mpg by cyl
summaryByCyl <- mtcars %>%
group_by(cyl) %>%
 summarise(mean_mpg = mean(mpg),
      median_mpg = median(mpg),
      sd_mpg = sd(mpg),
      min_mpg = min(mpg),
      max_mpg =
      max(mpg))
print(summaryByCyl
)# Boxplot
ggplot(mtcars, aes(x = factor(cyl), y = mpg)) +
 geom_boxplot() +
 labs(title = "Boxplot of MPG by Number of
   Cylinders",x = "Number of Cylinders",
   y = "Miles per
Gallon")# ANOVA
anova_result <- aov(mpg ~ factor(cyl), data = mtcars)
print(summary(anova_result))
# Correlation analysis
correlation <- cor.test(x = mtcars$mpg, y = as.numeric(mtcars$cyl))
print(correlation)
```

Output :

 # View the first few rows of the mtcars dataset

```
mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
```

```
Mazda RX4 Wag  21.0  6 160 110 3.90 2.875 17.02  0  1  4  4
Datsun 710        22.8  4  108  93 3.85 2.320 18.61  1  1  4  1
Hornet 4 Drive    21.4  6  258 110 3.08 3.215 19.44  1  0  3  1
Hornet Sportabout 18.7  8  360 175 3.15 3.440 17.02  0  0  3  2
Valiant        18.1  6  225 105 2.76 3.460 20.22  1  0  3  1
# Summary statistics for mpg by cyl
  cyl mean_mpg median_mpg sd_mpg min_mpg max_mpg
 <dbl>  <dbl>   <dbl>  <dbl>  <dbl>
<dbl>1 4        26.7  26      4.51  21.4  33.9
2   6   19.7     19.7  1.45   17.8   21.4
3   8   15.1     15.2  2.56   10.4   19.2
# ANOVA
       Df Sum Sq Mean Sq F value Pr(>F)
factor(cyl)  2  824.8  412.4  39.7 4.98e-09 ***
Residuals  29  301.3   10.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# Correlation analysis
Pearson's product-moment correlation
data:  mtcars$mpg and as.numeric(mtcars$cyl)t
= -8.9197, df = 30, p-value = 6.113e-10
alternative hypothesis: true correlation is not equal to 095
percent confidence interval:
 -0.9257694 -0.7163171
sample estimates:    cor
 -0.852162
```

Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4

Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1

Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1

Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2

Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

 # Summary statistics for mpg by cyl

cyl mean_mpg median_mpg sd_mpg min_mpg max_mpg

<dbl> <dbl> <dbl> <dbl> <dbl>

<dbl>1 4 26.7 26 4.51 21.4 33.9

2 6 19.7 19.7 1.45 17.8 21.4

3 8 15.1 15.2 2.56 10.4 19.2

 # ANOVA

Df Sum Sq Mean Sq F value Pr(>F)

factor(cyl) 2 824.8 412.4 39.7 4.98e-09 ***

Residuals 29 301.3 10.4

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 # Correlation analysis

Pearson's product-moment correlation

data: mtcars$mpg and as.numeric(mtcars$cyl)t

= -8.9197, df = 30, p-value = 6.113e-10

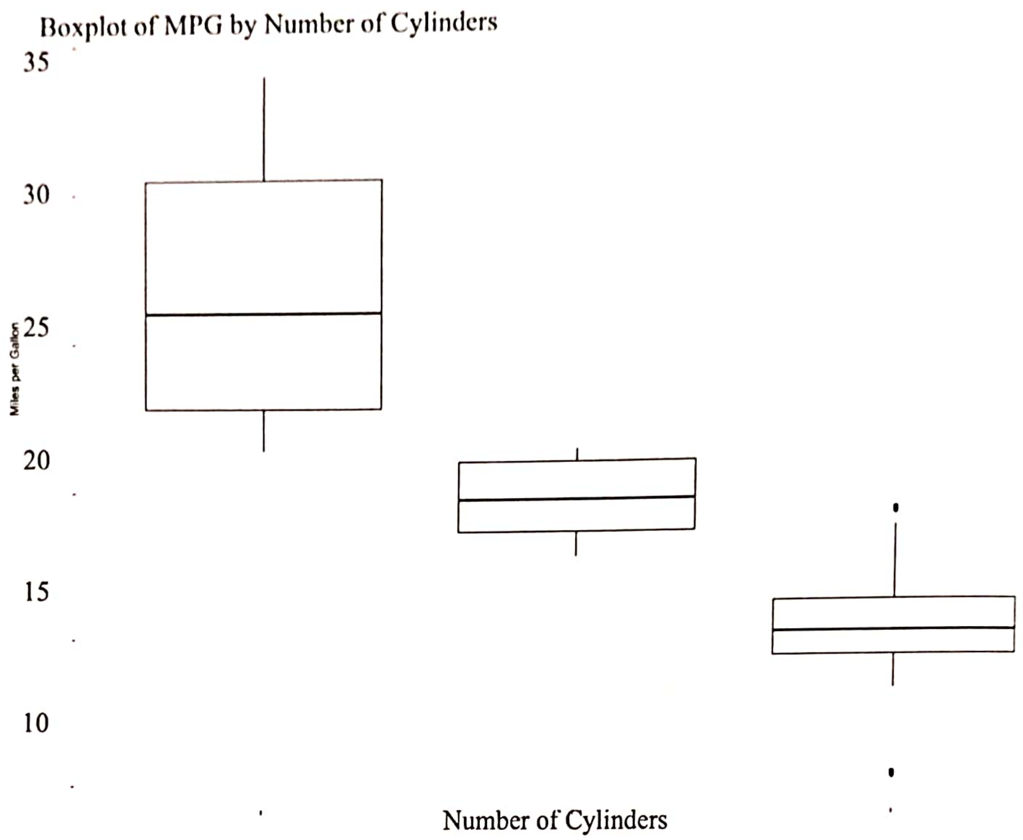
alternative hypothesis: true correlation is not equal to 095

percent confidence interval:

-0.9257694 -0.7163171

sample estimates: cor

-0.852162



D. Relationship between two continuous variables – covariance, correlation coefficients, comparing multiple correlations.

Covariance :

Covariance is a measure of how two variables change together. It indicates the direction of the linear relationship between the variables (positive, negative, or none) and the strength of that relationship. Here's a detailed explanation of covariance and its usage in R:

Definition: Covariance between two variables X and Y is calculated as:

$$s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Where:

- **n** is the number of data points

- X_i & Y_i are the individual data points

- ($\bar{X}$ and ($\bar{Y}$) are the means of **X** and **Y**, respectively

Interpretation:

- If (cov(X, Y) > 0),it indicates a positive relationship, meaning that as (X) increases, (*Y*) also tendsto increase.

- If (cov(X, Y) < 0),it indicates a negative relationship, meaning that as (*X*) increases, (*Y*) tends to decrease.

- If (cov(X, Y) = 0) , it indicates no linear relationship between (X) and (Y). However, it's importantto note that two variables can be related in a non-linear way even if their covariance is 0.

Correlation:

Correlation is a standardized measure of the strength and direction of the linear relationship between two variables. Unlike covariance, correlation coefficients are bound between -1 and 1, where -1 indicates a perfect negative linear relationship, 1 indicates a perfect positive linear relationship, and0 indicates no linear relationship.

Pearson Correlation Coefficient:

The Pearson correlation coefficient (r) between two variables X and Y is calculated as:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Where:

- **Sxy** is the covariance between x and y.

- **Sx** and **Sy** are the standard deviations of x and y respectively.

-**rxy** is the correlation coefficient.

R code :

```
# Sample data
set.seed(123) # for reproducibility
x <- rnorm(100)
y <- 2*x + rnorm(100)
# Calculate covariance
covariance_xy <- cov(x,
y)# Compute correlations
correlation_xy <- cor(x,
y)
correlation_x2y <- cor(x^2,
y)correlation_x3y <- cor(x^3,
y)# Fisher z transformation
fisher_z_xy <- 0.5 * log((1 + correlation_xy) / (1 - correlation_xy))
fisher_z_x2y <- 0.5 * log((1 + correlation_x2y) / (1 - correlation_x2y))
fisher_z_x3y <- 0.5 * log((1 + correlation_x3y) / (1 - correlation_x3y))
# Compare correlations using ANOVA
```

```
p_value <- anova(lm(fisher_z_xy ~ 1, data.frame(x = 1)))$`Pr(>F)`[1]
p_value_x2y <- anova(lm(fisher_z_x2y ~ 1, data.frame(x =
1)))$`Pr(>F)`[1]p_value_x3y <- anova(lm(fisher_z_x3y ~ 1, data.frame(x
= 1)))$`Pr(>F)`[1]# Print results
print("Covariance:")
print(covariance_xy)
print("P-values:")
print(p_value)
print(p_value_x2y)
print(p_value_x3y)
```

Output :

```
[1] "Covariance:"
      [,1]      [,2]
[1,] 0.94387197 1.85675756
[2,] 1.85675756 4.08266999
[1] "P-values:"
[1] 2.220446e-16
[1] 3.159168e-05
[1] 4.707527e-13
```

e. Visualization methods – Categorical & Continuous variable , Two Categorical Variables , Two Continuous variables

Categorical variables :

Categorical variables represent data that can be divided into groups or categories. Unlike numerical (quantitative) variables, which represent a measurable quantity, categorical variables represent characteristics or attributes. Here's a more detailed explanation of categorical variables:

Types of Categorical Variables:

Nominal:

Categories with no inherent order or ranking. For example, eye color (blue, brown, green) or gender (male, female, other).

Ordinal:

Categories with a clear order or ranking. For example, education level (high school, college, graduate) or income level (low, medium, high).

Continuous variables :

Continuous variables, unlike categorical variables, can take on any value within a certain range. They are used to represent measurements or quantities that can be expressed numerically. Here's a detailed explanation:

Characteristics:

- Continuous variables can take on an infinite number of values within a range.
- They are typically measured, not counted.
- Examples include height, weight, temperature, and time.

Visualization Methods :

BAR PLOT :

There are two types of bar plots- horizontal and vertical which represent data points as horizontal or vertical bars of certain lengths proportional to the value of the data item. They are generally used for continuous and categorical variable plotting. By setting the horiz parameter totrue and false, we can get horizontal and vertical bar plots respectively.

HISTOGRAM :

A histogram is like a bar chart as it uses bars of varying height to represent data distribution. However, in a histogram values are grouped into consecutive intervals called bins. In a Histogram, continuous values are grouped and displayed in these bins whose size can be varied.

BOX PLOT :

The statistical summary of the given data is presented graphically using a boxplot. A boxplot depicts information like the minimum and maximum data point, the median value, first and third quartile, and interquartile range.

SCATTER PLOT :

A scatter plot is composed of many points on a Cartesian plane. Each point denotes the value taken by two parameters and helps us easily identify the relationship between them.

HEAT MAP :

Heatmap is defined as a graphical representation of data using colors to visualize the value of the matrix. heatmap() function is used to plot heatmap.
Syntax: heatmap(data)

Parameters: data: It represent matrix data, such as values of rows and columns

Return: This function draws a heatmap.

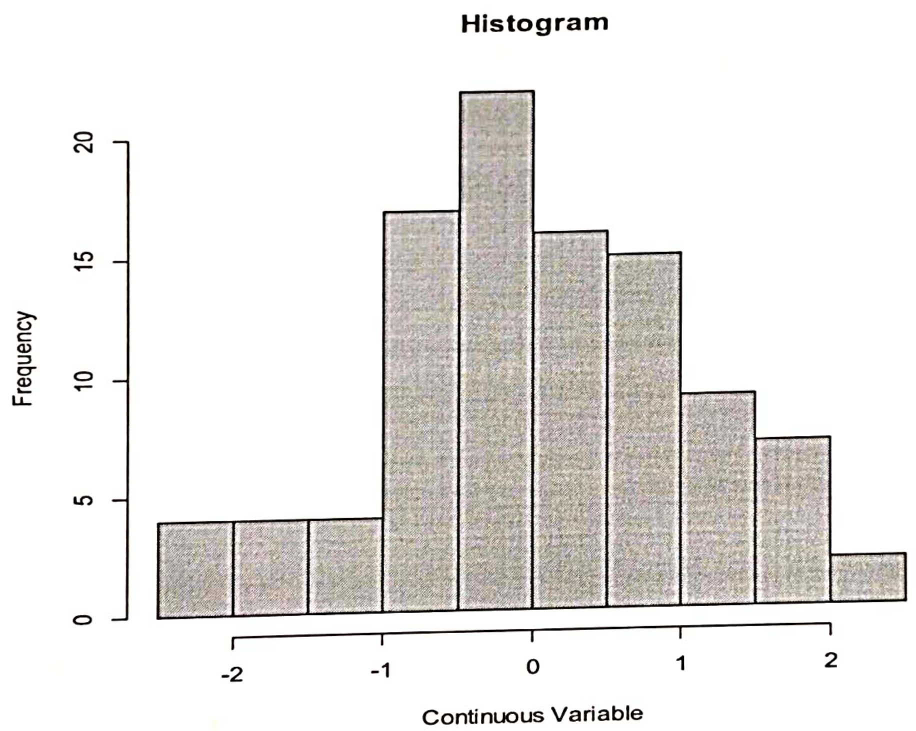
R Code :

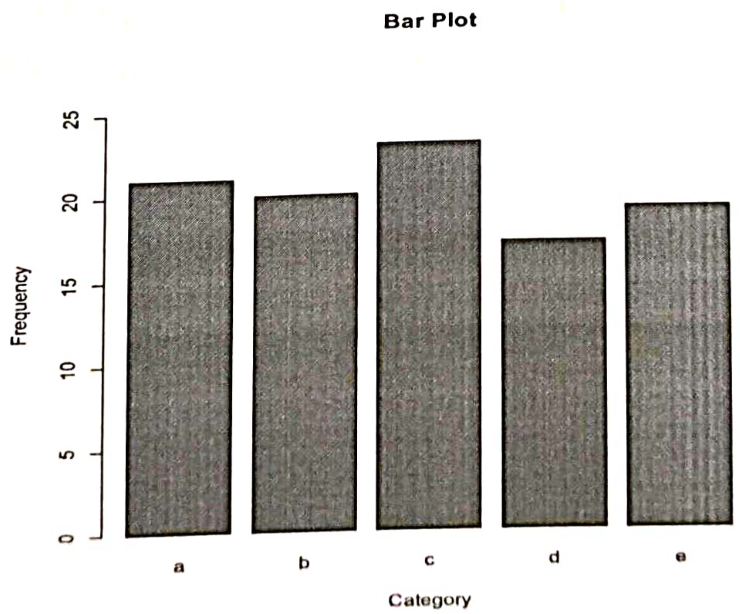
```
# Sample data

set.seed(123)

data <- data.frame(

  category = sample(letters[1:5], 100, replace =
```

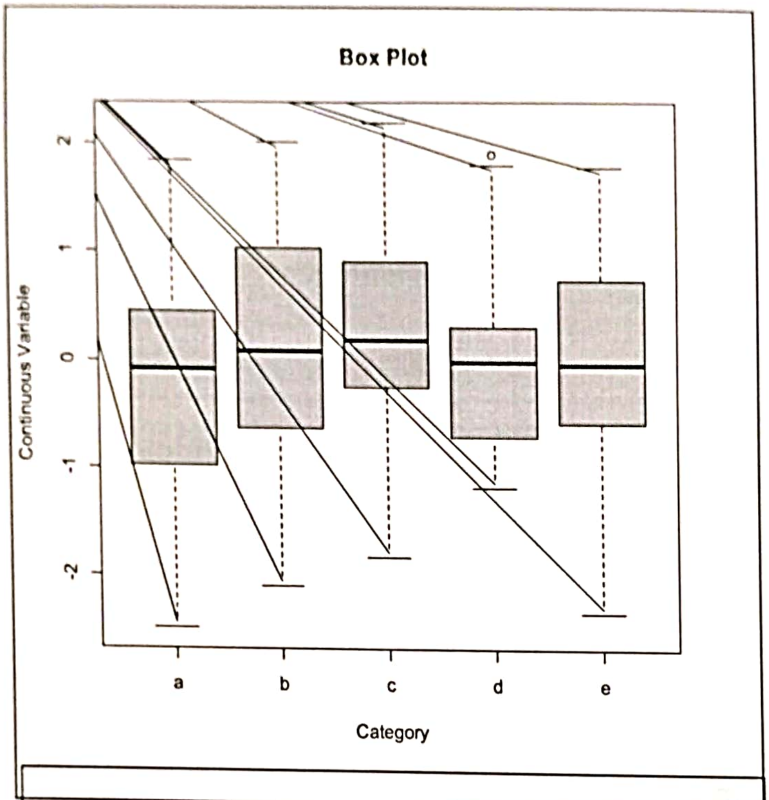
```
 TRUE),continuous = rnorm(100)

)

# Bar plot

barplot(table(data$category),

       xlab = "Category",

       ylab =

       "Frequency",main

       = "Bar Plot", col =

       "skyblue",

       ylim = c(0, max(table(data$category)) +

5))# Histogram

hist(data$continuous,

     breaks = 10,

     col = "lightgreen",

     xlab = "Continuous

     Variable",ylab =

     "Frequency",

     main = "Histogram")
# Box plot

boxplot(data$continuous ~ data$category,

       xlab = "Category",

       ylab = "Continuous

       Variable",main = "Box Plot",

       col =

"lightblue")# Scatter

plot

plot(as.numeric(factor(data$category)),

    data$continuous,

    xlab = "Category",

    ylab = "Continuous
```

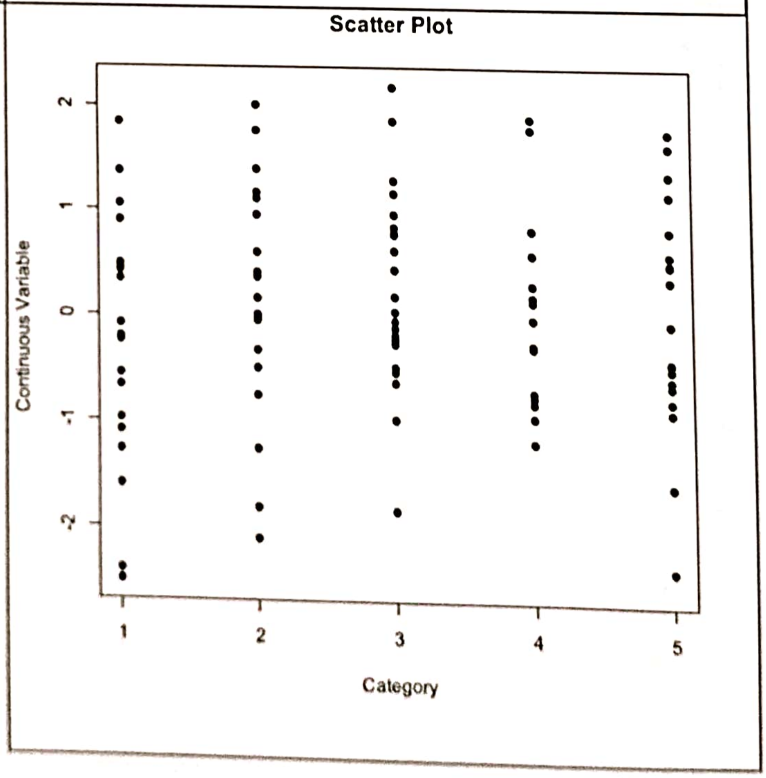
```
Variable",main = "Scatter
Plot",
col = "darkred",
pch
```

Output :






Box Plot
Continuous Variable
2
1
0
-1
-2
a
b
c
d
e
Category


Scatter Plot
Continuous Variable
2
1
0
-1
-2
1
2
3
4
5
Category

Week-4:PROBABILITY DISTRIBUTIONS:

a) Sampling from distributions – Binomial distribution, normaldistribution.

>Binomial distributions:

It is a probability distribution used in statistics. It is a discrete distribution and has only two outcomes i.e., success or failure all its trials are independent, the probability of success remains the same and the previous outcome does not affect the next outcome. It helps to findthe individual probabilities as well as cumulative probabilities over a certain range.

>Formula:

$P(X = k) = nC_r p^r q^{n-r}$, where r = 0, 1, 2, 3,..., n

p is the probability of success

q is the probability of failure

$p + q = 1$

>Functions for Binomial Distribution:

We have four functions for handling binomial distribution in R namely

1) dbinom() Function:

This function is used to find probability at a particular value for a data that follows binomial distribution i.e. it finds:

P(X = k)

>Syntax:

dbinom(k, n, p)

>Example:

```
dbinom(3, size = 13, prob = 1 / 6)

probabilities <- dbinom(x = c(0:10), size = 10, prob = 1 / 6)

data.frame(x, probs)

plot(0:10, probabilities, type = "l")
```

>Output:

```
> dbinom(3, size = 13, prob = 1/6)
[1] 0.2138454
> probabilities = dbinom(x = c(0:10), size = 10, prob = 1/6)
> data.frame(probabilities)
   probabilities
1  1.615056e-01
2  3.230112e-01
3  2.907100e-01
4  1.550454e-01
```

5 5.426588e-02

6 1.302381e-02

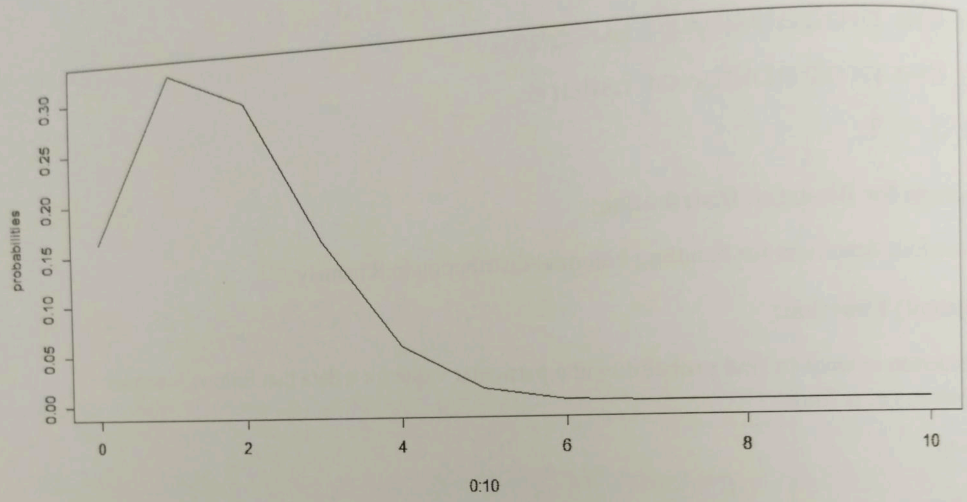
7 2.170635e-03

8 2.480726e-04

9 1.860544e-05

10 8.269086e-07

11 1.653817e-08



2) pbinom() Function:

The function pbinom() is used to find the cumulative probability of a data following binomial distribution till a given value i.e., it finds P(X <= k)

>Syntax:

```
pbinom(k, n, p)
```

>Example:

```
pbinom(3, size = 13, prob = 1 / 6)
plot(0:10, pbinom(0:10, size = 10, prob = 1 / 6), type = "l")
```

>Output :

```
> pbinom(3, size = 13, prob = 1/6)[1]
0.8419226
```

5 5.426588e-02

6 1.302381e-02

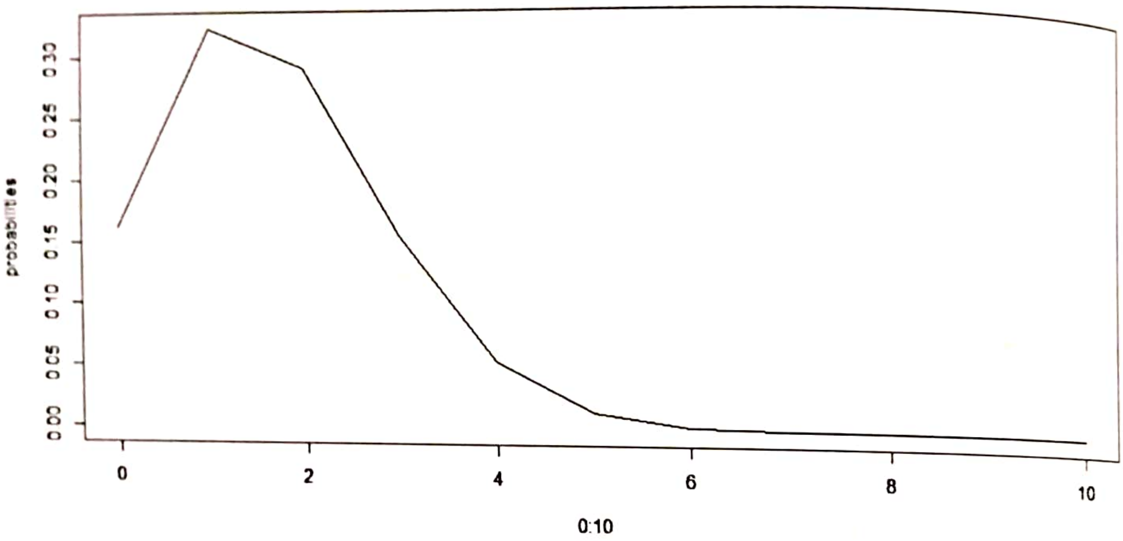
7 2.170635e-03

8 2.480726e-04

9 1.860544e-05

10 8.269086e-07

11 1.653817e-08



2) pbinom() Function:

The function pbinom() is used to find the cumulative probability of a data following binomial distribution till a given value i.e., it finds P(X <= k)

>Syntax:

pbinom(k, n, p)

>Example:

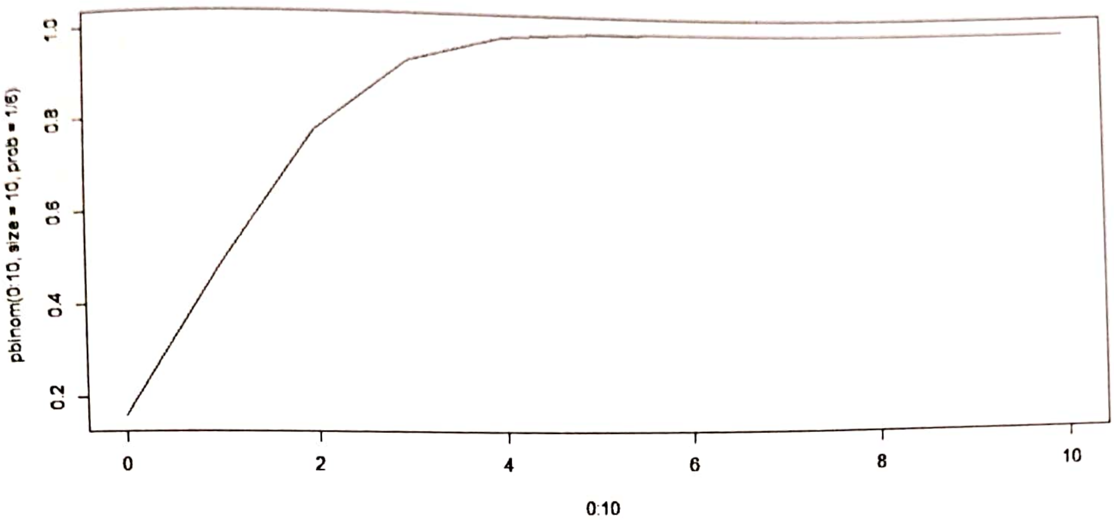
pbinom(3, size = 13, prob = 1 / 6)

plot(0:10, pbinom(0:10, size = 10, prob = 1 / 6), type = "l")

>Output :

> pbinom(3, size = 13, prob = 1/6)[1]

0.8419226



3) qbinom() Function:

This function is used to find the nth quantile, that is if P(x <= k) is given, it finds k

>Syntax:

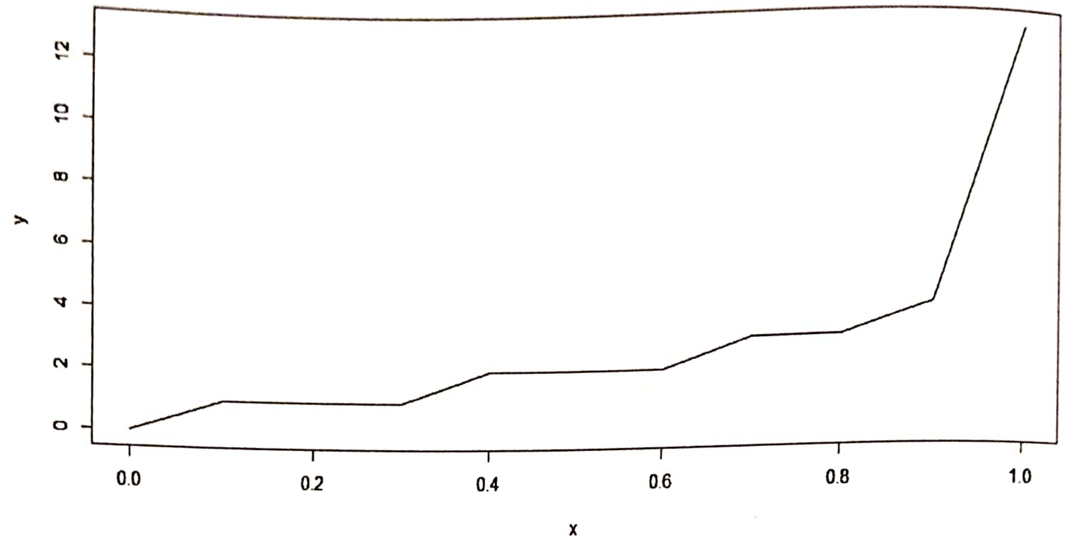
qbinom(P, n, p)

>Example:

```
qbinom(0.8419226, size = 13, prob = 1 / 6)
x <- seq(0, 1, by = 0.1)
y <- qbinom(x, size = 13, prob = 1 / 6)
plot(x, y, type = 'l')
```

>Output :

```
> qbinom(0.8419226, size = 13, prob = 1/6)
[1] 3
```



4) rbinom() Function:

This function generates n random variables of a particular probability.

> Syntax:

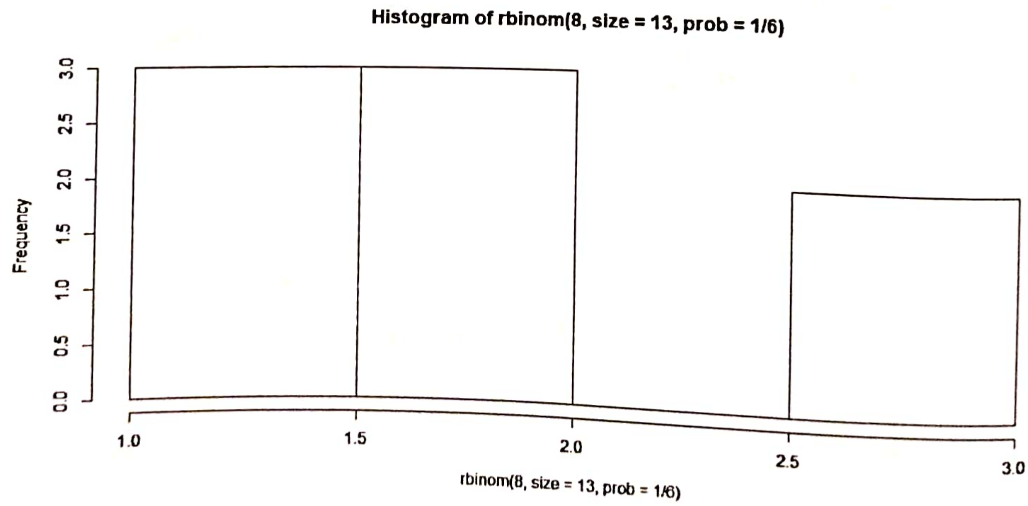
rbinom(n, N, p)

>Example:

rbinom(8, size = 13, prob = 1 / 6)

hist(rbinom(8, size = 13, prob = 1 / 6))

>Output:



> rbinom(8, size = 13, prob = 1/6)[1]

1 1 2 1 4 0 2 3

>Normal Distribution:

Normal Distribution is a probability function used in statistics that tells about how the data values are distributed. It is the most important probability distribution function used in statistics because of its advantages in real case scenarios. For example, the height of the population, shoe size, IQ level, rolling a dice, and many more. It is generally observed that data distributionis normal when there is a random collection of data from independent sources.

The graph is symmetric distribution. In R, there are 4 built-in functions to generate normal distribution:

- dnorm()

dnorm(x, mean, sd)

- pnorm()

pnorm(x, mean, sd)

- qnorm()

qnorm(p, mean, sd)

- rnorm()

rnorm(n, mean, sd)

where,

– x represents the data set of values – mean(x) represents the mean of data set x. It'sdefault value is 0.

– sd(x) represents the standard deviation of data set x. It's default value is 1.m

– n is the number of observations. – p is vector of probabilities.

>Functions To Generate Normal Distribution in R:

1)dnorm():

dnorm() function in R programming measures density function of distribution.

>Syntax :

dnorm(x, mean, sd)

>Example:

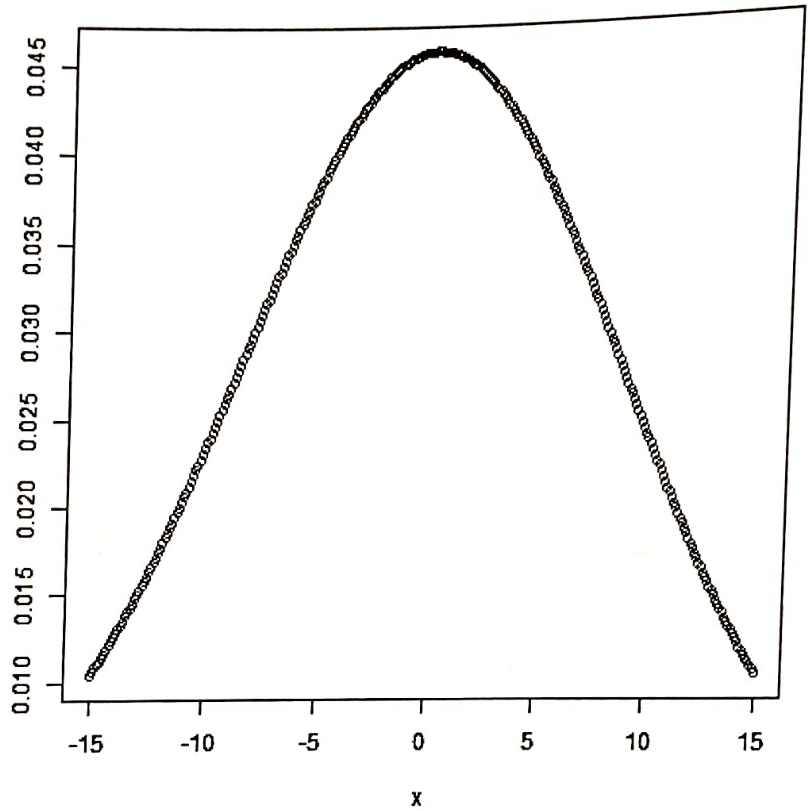
```
# creating a sequence of values
# between -15 to 15 with a difference of 0.1x =
seq(-15, 15, by=0.1)
y = dnorm(x, mean(x), sd(x))
# output to be present as PNG file
png(file="dnormExample.png")
# Plot the graph.
```

plot(x, y)

saving the file

dev.off()

>**Output:**



2) pnorm():

pnorm() function is the cumulative distribution function which measures the probability thata random number X takes a value less than or equal to x .

>**Syntax:**

pnorm(x, mean, sd)

creating a sequence of values

between -10 to 10 with a difference of 0.1x <-

seq(-10, 10, by=0.1)

y <- pnorm(x, mean = 2.5, sd = 2)#

output to be present as PNG file

png(file="pnormExample.png")#

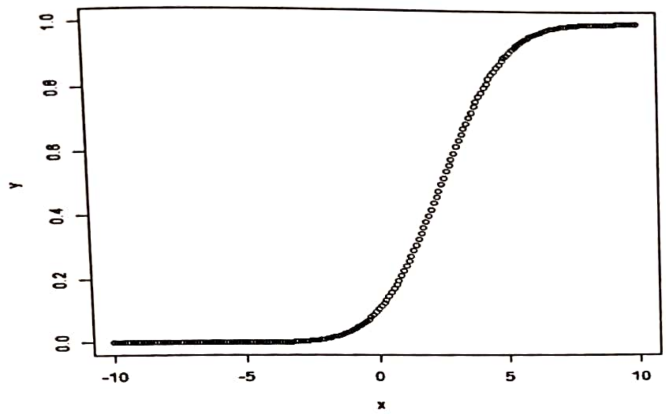
Plot the graph.

plot(x, y)

saving the file

dev.off()

>Output:



3) qnorm():

qnorm() function is the inverse of pnorm() function. It takes the probability value and gives output which corresponds to the probability value. It is useful in finding the percentiles of a normal distribution.

>Syntax:

qnorm (p, mean, sd)

>Example:

Create a sequence of probability values#

incrementing by 0.02.

x <- seq(0, 1, by = 0.02)

y <- qnorm(x, mean(x), sd(x))

output to be present as PNG file

png(file = "qnormExample.png")

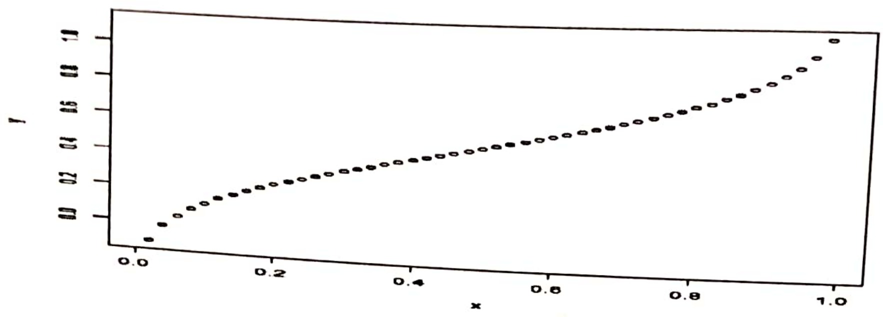
Plot the graph.

plot(x, y)

Save the file.

dev.off()

>**Output:**



4) rnorm():

rnorm() function in R programming is used to generate a vector of random umbers which arenormally distributed.

>**Syntax:**

>**Example:**

rnorm(x, mean, sd)

Create a vector of 1000 random numbers#

with mean=90 and sd=5

x <- rnorm(10000, mean=90, sd=5)#

output to be present as PNG file

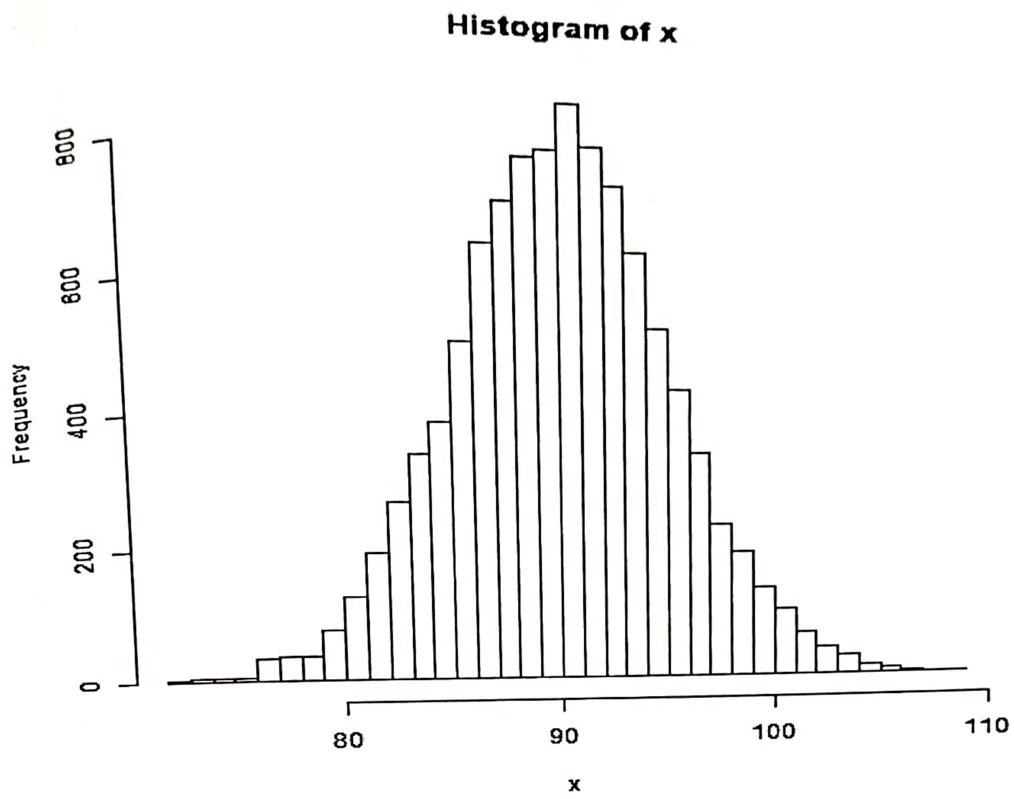
png(file = "rnormExample.png")

Create the histogram with 50 bars

hist(x, breaks=50) # Save the file.

dev.off()

>**Output :**



b) tTest, zTest, Chi Square Test:

>T-Test Approach:

Mathematically, what the t-test does is, take a sample from both sets and establish theproblem assuming a null hypothesis that the two means are the same.

Classification of T-tests

- One Sample T-test
- Two sample T-test
- Paired sample T-test

One Sample T – Test Approach:

The One-Sample T-Test is used to test the statistical difference between a sample meanand a known or assumed/hypothesized value of the mean in the population.

>Syntax:

t.test(y, mu = 0)

where x is the name of the variable of interest and mu is set equal to the mean specifiedby the null hypothesis.

>**Example:**

set.seed(0)

sweetSold <- c(rnorm(50, mean = 140, sd = 5))

mu=The hypothesized mean difference between the two groups.

t.test(sweetSold, mu = 150)

>**Output:**

One Sample t-test

data: sweetSold

t = -15.249, df = 49, p-value < 2.2e-16

alternative hypothesis: true mean is not equal to 15095

percent confidence interval:

138.8176 141.4217

sample estimates:

mean of x

140.1197

Two sample T-Test Approach:

It is used to help us to understand whether the difference between the two means is realor simply by chance.

>**Syntax:**

t.test(y1, y2, paired=FALSE)

By default, R assumes that the variances of y1 and y2 are unequal, thus defaulting toWelch's test. To toggle this, we use the flag var.equal=TRUE.

>**Example:**

set.seed(0)

shopOne <- rnorm(50, mean = 140, sd = 4.5)

shopTwo <- rnorm(50, mean = 150, sd = 4)

t.test(shopOne, shopTwo, var.equal = TRUE)

>**Output:**

Two Sample t-test

data: shopOne and shopTwo

t = -13.158, df = 98, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 095

percent confidence interval:

-11.482807 -8.473061

sample estimates:

mean of x mean of y

140.1077 150.0856

Paired Sample T-test:

This is a statistical procedure that is used to determine whether the mean difference between two sets of observations is zero. In a paired sample t-test, each subject is measuredtwo times, resulting in pairs of observations.

>Syntax:

t.test(y1, y2, paired=TRUE)

>Example:

```
set.seed(2820)
sweetOne <- c(rnorm(100, mean = 14, sd = 0.3))
sweetTwo <- c(rnorm(100, mean = 13, sd = 0.2))
t.test(sweetOne, sweetTwo, paired = TRUE)
```

>Output:

```
    Paired t-test
data:  sweetOne and sweetTwo
t = 29.31, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean difference is not equal to 095
percent confidence interval:
 0.9892738 1.1329434
sample estimates:
mean difference
      1.061109
```

>Z Test Approach:

Z test is a popular parametric test used for hypothesis testing. Z test is a statistical methodused to determine if there is a significant difference between sample and population means orbetween the means of two samples. It is used when there is a large sample size and the population. It is to be noted that Z Test follows normal distribution. The Z value acts as a threshold. Based on its value it is decided whether to accept the hypothesis or reject thehypothesis. This test is applicable where the sample size is greater than 30.

There are two types of Z tests based on samples:

- One Sample Z-test
- Two Sample Z-test

>Syntax:

z.test(x, y, alternative='two.sided', mu=0, sigma.x=NULL, sigma.y=NULL,conf.level=.95)

>One Sample Z test:

Here Z Test is applicable on one sample that has been taken from the population.

>Example:

```
library(BSDA)

# Sample data

sample_data <- c(26, 25, 10, 34, 30, 23, 28, 29, 25, 27)

# One-sample Z-test

z_test <- z.test(sample_data, mu = 24, sigma.x=10)#

Print the result

print(z_test)
```

>Output:

```
   One-sample z-Test
data:  sample_data
z = 0.53759, p-value = 0.5909
alternative hypothesis: true mean is not equal to 2495 percent
confidence interval:
 19.50205 31.89795
sample estimates:
mean of x
    25.7
```

Two sample Z test:

Here Z Test is applicable on two samples that has been taken from the population

>Example:

```
# Two vectors of sample data data1

<- c(27, 24, 18, 29, 30,27)

data2 <- c(23, 28, 20, 19, 35,23)

 # Two-sample Z-test
```

threshold. Based on its value it is decided whether to accept the hypothesis or reject thehypothesis. This is applicable where the sample size is greater than 30.

There are two types of Z tests based on samples:

- One Sample Z-test
- Two Sample Z-test

>Syntax:

z.test(x, y, alternative='two.sided', mu=0, sigma.x=NULL, sigma.y=NULL,conf.level=.95)

>One Sample Z test:

Here Z Test is applicable on one sample that has been taken from the population.

>Example:

```
library(BSDA)

# Sample data

sample_data <- c(26, 25, 10, 34, 30, 23, 28, 29, 25, 27)

# One-sample Z-test

z_test <- z.test(sample_data, mu = 24, sigma.x=10)#

Print the result

print(z_test)
```

>Output:

```
    One-sample z-Test
data:  sample_data
z = 0.53759, p-value = 0.5909
alternative hypothesis: true mean is not equal to 2495 percent
confidence interval:
 19.50205 31.89795
sample estimates:
mean of x
    25.7
```

Two sample Z test:

Here Z Test is applicable on two samples that has been taken from the population

>Example:

```
# Two vectors of sample data data1

<- c(27, 24, 18, 29, 30,27)

data2 <- c(23, 28, 20, 19, 35,23)

# Two-sample Z-test
```

```
z_test_result <- z.test (data1, data2, mu=26, sigma.x=10, sigma.y =15)# Print
the result
print(z_test_result)
```

>Output:

```
Two-sample z-Test data:
data1 and data2
z = -3.3742, p-value = 0.0007403
alternative hypothesis: true difference in means is not equal to 2695
percent confidence interval:
 -13.25828  15.59161
sample estimates:
mean of x mean of y
 25.83333  24.66667
```

>Chi-Square Test:

The chi-square test of independence evaluates whether there is an association between the categories of the two variables. There are basically two types of random variables and they yield two types of data: numerical and categorical. In R Programming Language Chi- square statistics is used to investigate whether distributions of categorical variables differ from one another.

>Syntax:

chisq.test(data)

>Parameters:

data: data is a table containing count values of the variables in the table.

>Example:

```
library(MASS)
# Create a data frame from the main data set. stu_data =
data.frame(survey$Smoke,survey$Exer)
# Create a contingency table with the needed variables.
stu_data = table(survey$Smoke,survey$Exer) print(stu_data)
# applying chisq.test() function
print(chisq.test(stu_data))
```

>Output

```
          Freq None Some
  Heavy     7     1    3
```

Never	87	18	84
Occas	12	3	4
Regul	9	1	7

Pearson's Chi-squared testdata:
stu_data
X-squared = 5.4885, df = 6, p-value = 0.4828

c) Density functions:

>Density Plots:

A **density plot** is a representation of the distribution of a numeric variable that uses a kernel density estimate to show the probability density function of the variable. In R Language we use the density() function which helps to compute kernel density estimates. And further with its return value, is used to build the final density plot.

>Syntax: density(x)

>Parameters:

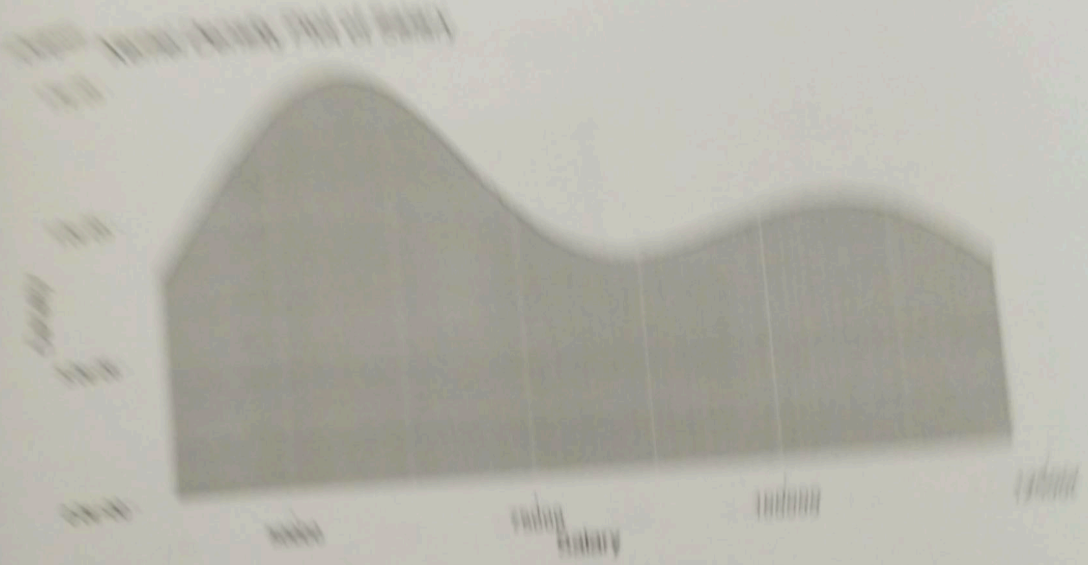
- **x:** the data from which the estimate is to be computed

>Returns:

It will return the kernel density.

>Example:

```
library(readxl)

library(ggplot2)


Salary_Data <- read_excel("Salary_Data.xls")


den <- density(Salary_Data$YearsExperience)library(ggplot2)


ggplot(Salary_Data, aes(x = Salary)) +

  geom_density(fill = "skyblue", alpha = 0.7) +

  labs(title = "Kernel Density Plot of Salary",

      x = "Salary", y

      = "Density")
```

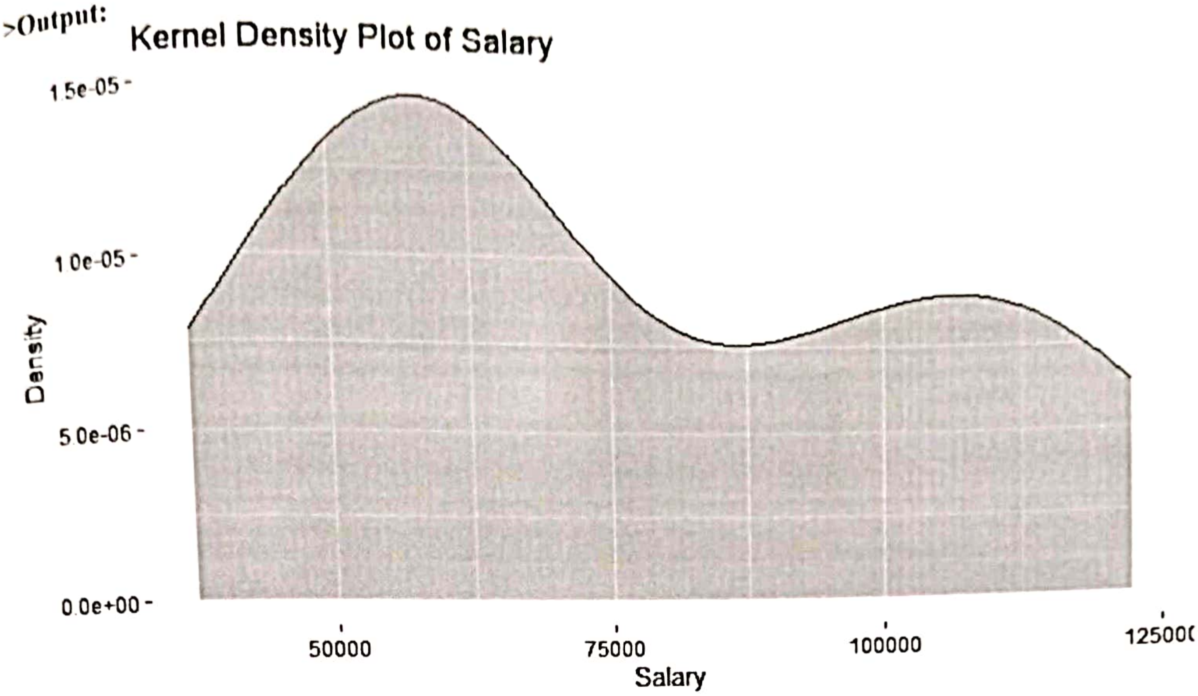


- Create a histogram and a density plot in the same frame:

- Example:

```
hist(beaver1$temp,
    col="green",
    border="black", prob
    = TRUE, xlab =
    "temp", main =
    "GFG")
lines(density(beaver1$temp),lwd = 2,
     col = "chocolate3")
```

>Output:

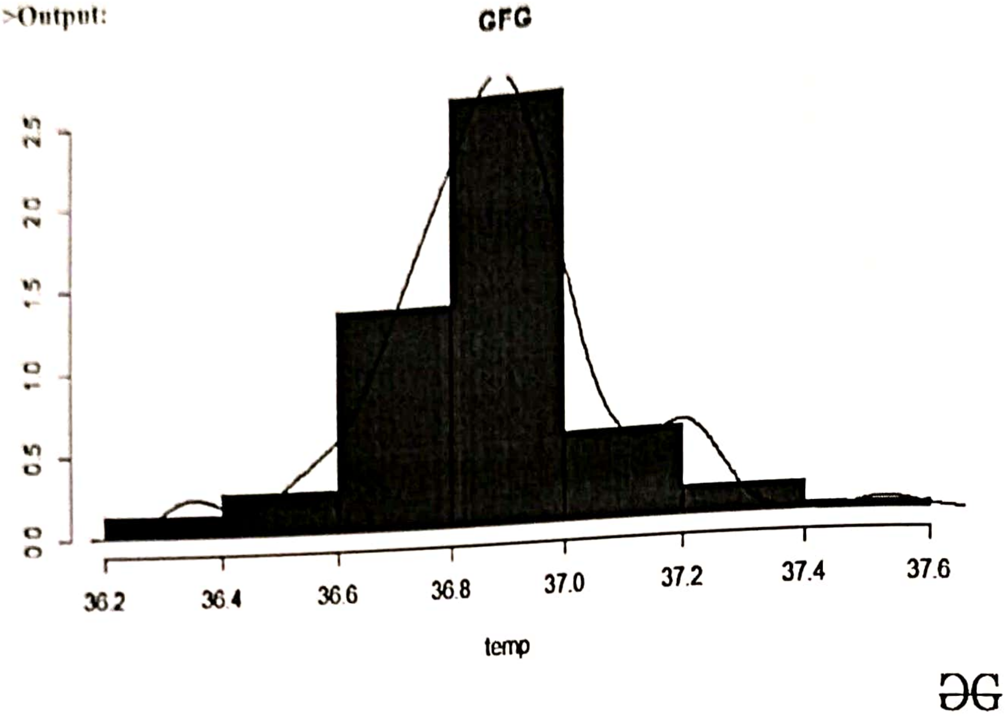


>Create a histogram and a density plot in the same frame:

>Example:

```
hist(beaver1$temp,
     col="green",
     border="black", prob
     = TRUE, xlab =
     "temp", main =
     "GFG")
 lines(density(beaver1$temp),lwd = 2,
      col = "chocolate3")
```

>Output:



d) **Data Visualization using ggplot - Box plot, histograms, scatterplotter, line chart, bar chart, heat maps:**

>Data visualization with R and ggplot2:

Data visualization with R and ggplot2 in R Programming Language also termedas Grammar of Graphics is a free, open-source, and easy-to-use visualization package widely used in R Programming Language. It is the most powerful visualization package written by Hadley Wickham.

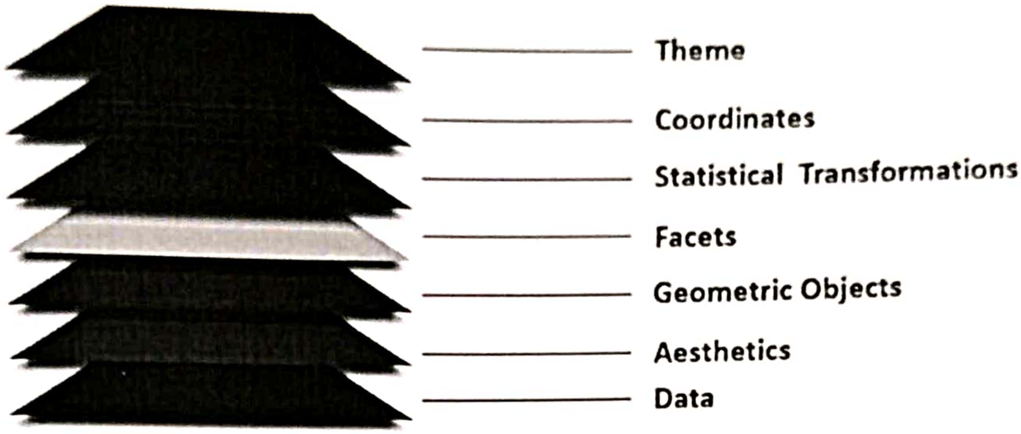
It includes several layers on which it is governed. The layers are as follows:

>Building Blocks of layers with the grammar of graphics:

- **Data:** The element is the data set itself
- **Aesthetics:** The data is to map onto the Aesthetics attributes such as x-axis, y-axis, color, fill, size, labels, alpha, shape, line width, line type
- **Geometrics:** How our data being displayed using point, line, histogram, bar,boxplot
- **Facets:** It displays the subset of the data using Columns and rows
- **Statistics:** Binning, smoothing, descriptive, intermediate
- **Coordinates:** the space between data and display using Cartesian, fixed,polar, limits.

- **Themes:** Non-data link

Main Components of the Grammar of Graphics



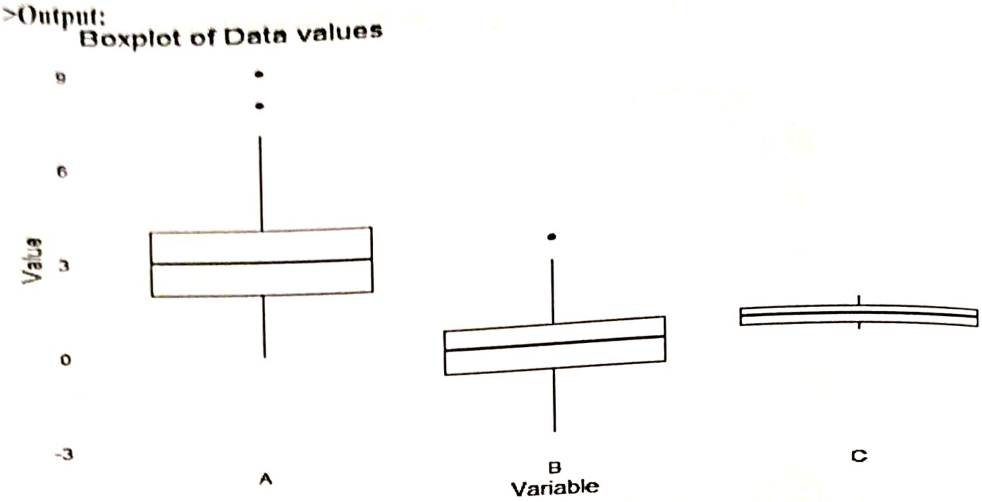
>Box Plots in R using ggplot2:

It is a visual representation of the spread of data for a variable and displays the range of values along with the median and quartiles. The **geom_boxplot()** function is used to create box plots in ggplot2.

>Example:

```
library(ggplot2)
set.seed(20000)
data <- data.frame(A = rpois(900, 3),
                   B = rnorm(900),
                   C = runif(900))
# Creating the boxplot using ggplot2
boxplot_plot <- ggplot(data_long, aes(x = variable, y = value)) +
  geom_boxplot() +
  labs(title = "Boxplot of Data values", x = "Variable", y = "Value") +
  theme_minimal()
print(boxplot_plot)
```

>Output:



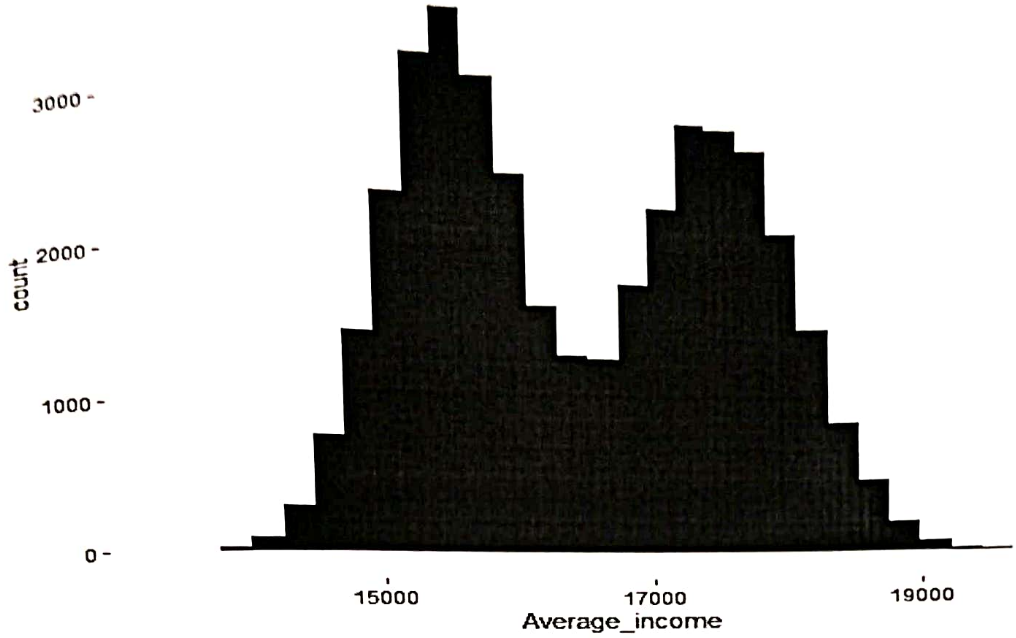
>Histogram in R using ggplot2:

Histograms plot quantitative data with ranges of the data grouped into intervals while bar charts plot categorical data. **geom_histogram()** function is an in-builtfunction of the ggplot2 module.

>Example:

```
set.seed(123)

df <- data.frame(

  gender=factor(rep(c(

    "Average Female income ", "Average Male income"), each=20000)),

  Average_income=round(c(rnorm(20000, mean=15500, sd=500),

                  rnorm(20000, mean=17500, sd=600)))

)

# if already installed ggplot2 then use library(ggplot2)

library(ggplot2)


# Basic histogram

ggplot(df, aes(x=Average_income)) + geom_histogram()
```

>**Output:**



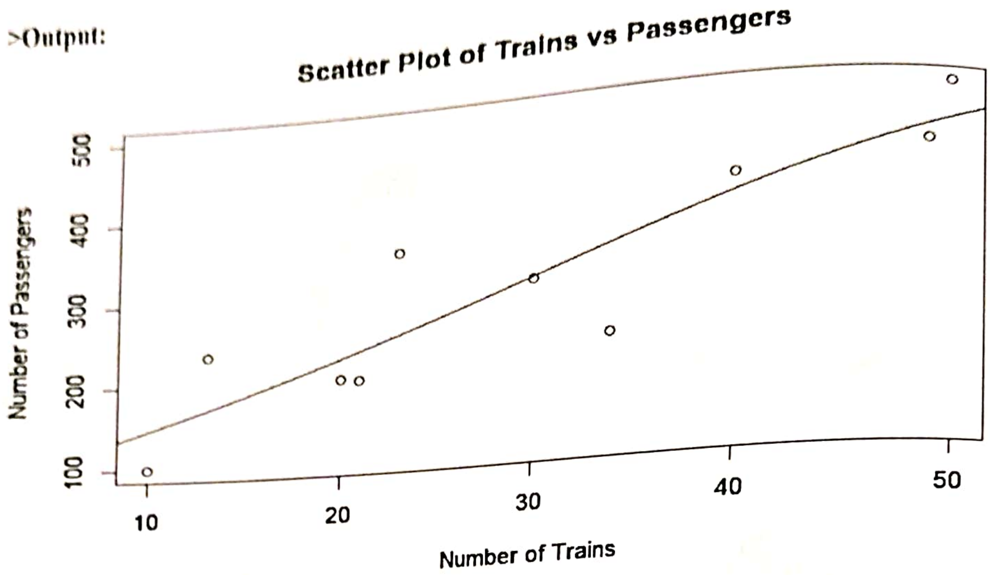
>Scatter plots in R using ggplot2:

Scatter plots can be used to visualize the relationship between two variables and identify trends, patterns, and outliers in the data. To create a scatter plot in R, you can use the **plot()** or **ggplot** function.

>Example:

```
trains <- c(10, 20, 30, 40, 50, 34, 23, 49, 21, 13)

passengers <- c(100, 200, 300, 400, 500,

            229, 346, 432, 198, 235)


plot(trains, passengers,

    xlab = "Number of Trains",

    ylab = "Number of Passengers",

    main = "Scatter Plot of Trains vs Passengers")

abline(lm(passengers~trains), col = "red")
```

>Output:



>Line Plot in R using ggplot2:

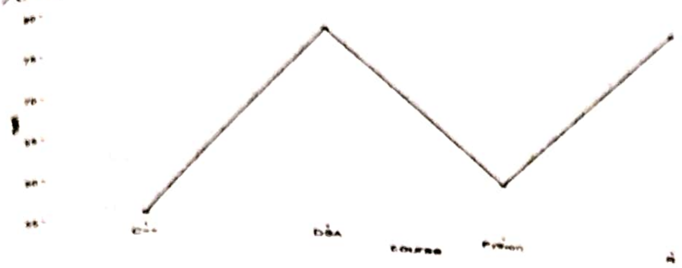
In a line graph, we have the horizontal axis value through which the line will be ordered and connected using the vertical axis values. We are going to use the R package **ggplot2** which has several layers in it.

First, you need to install the ggplot2 package if it is not previously installed in R Studio.

>Example:

```
library(ggplot2)

# Create data for chart

val <-data.frame(course=c('DSA','C++','R','Python'),num=c(77,55,80,60))

# Basic Line

ggplot(data=val, aes(x=course, y=num, group=1)) +
 geom_line()+
 geom_point()
```

>**Output:**



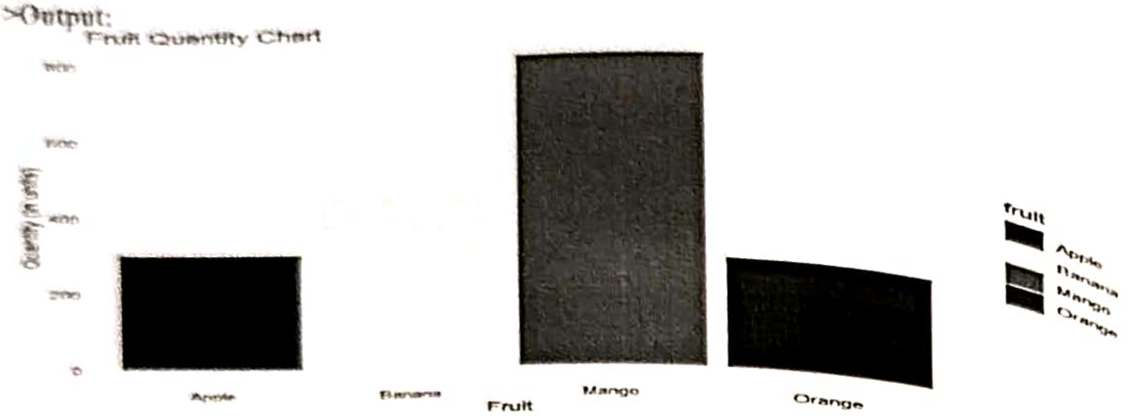
>Bar chart in R using ggplot2:

A bar chart is a representation of the dataset in the format of a rectangular bar. Respectively, its height depends on the values of variables in a dataset. You canuse **geom_bar()** to create bar charts with ggplot2.

>**Example:**

```
library(ggplot2)


data <- data.frame(
  fruit = c("Apple", "Banana", "Orange", "Mango"),
  quantity = c(300, 450, 280, 800),
  color = c("red", "yellow", "orange", "green")
)
bar_chart <- ggplot(data, aes(x = fruit,
                              y = quantity,
                              fill = fruit,
                              color = fruit)) +
   geom_bar(stat = "identity") + labs(title
   = "Fruit Quantity Chart",
      x = "Fruit", y = "Quantity (in units)") +
   scale_fill_manual(values = data$color) +
   scale_color_manual(values = data$color) +
   theme_minimal()
 bar_chart
```

>**Output:**



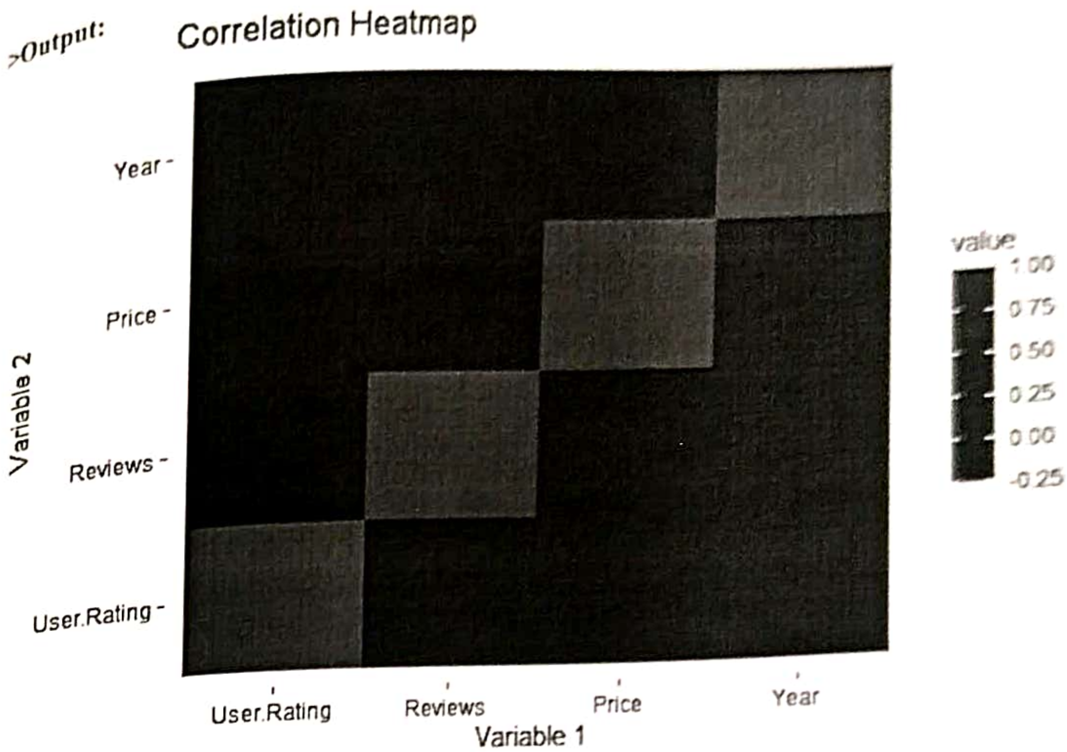
>Heatmap in R Using ggplot2:

They are especially beneficial for displaying and examining relationships and patterns in tabular data. The ggplot2 package in R, a robust and adaptable data visualization library, can be used to make heatmaps.

>Example:

```
library(ggplot2)

library(reshape2)


df <- read.csv("bestsellers.csv")


data <- cor(df[sapply(df,is.numeric)])


data1 <- melt(data)


ggplot(data1, aes(x = Var1, y = Var2, fill = value)) +geom_tile()
  +
  labs(title = "Correlation Heatmap",x
      = "Variable 1",
      y = "Variable 2")
```

>Output:



Correlation Heatmap
Year
Price
Reviews
User.Rating
Variable 2
User.Rating
Reviews
Price
Year
Variable 1
value
1.00
0.75
0.50
0.25
0.00
-0.25

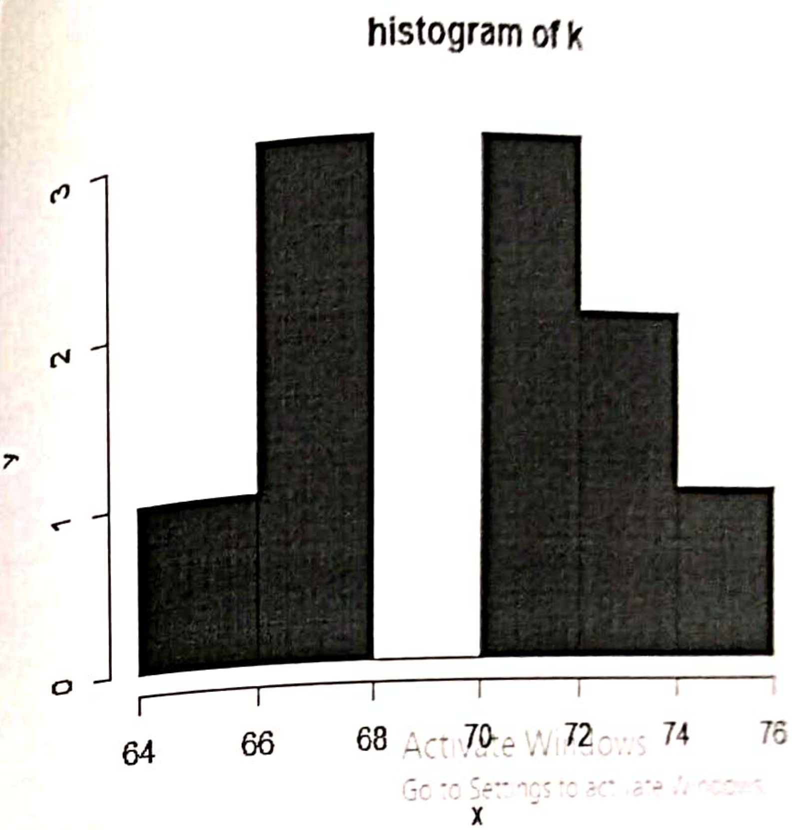
Week-5:Exploratory data analysis

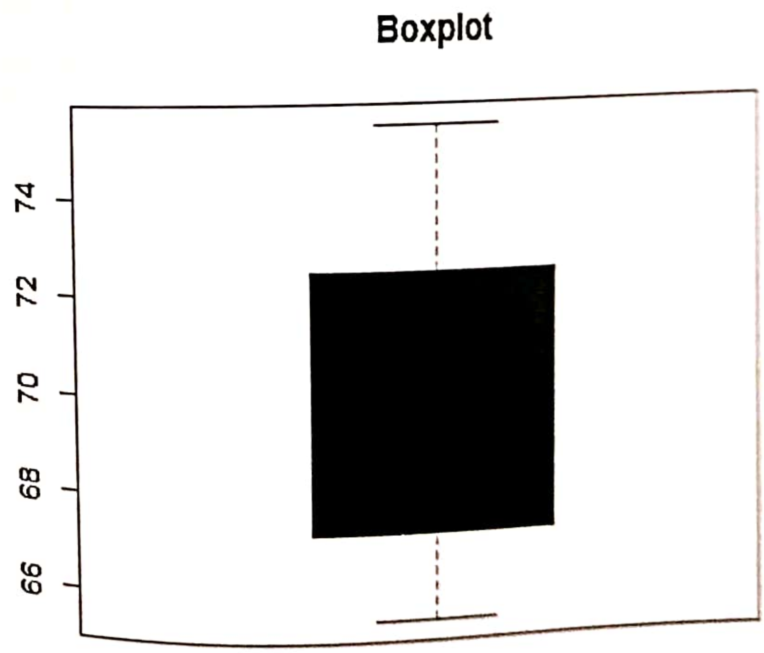
Demonstrate the range, summary, mean, variance, standard deviation, histogram, box
population dataset.

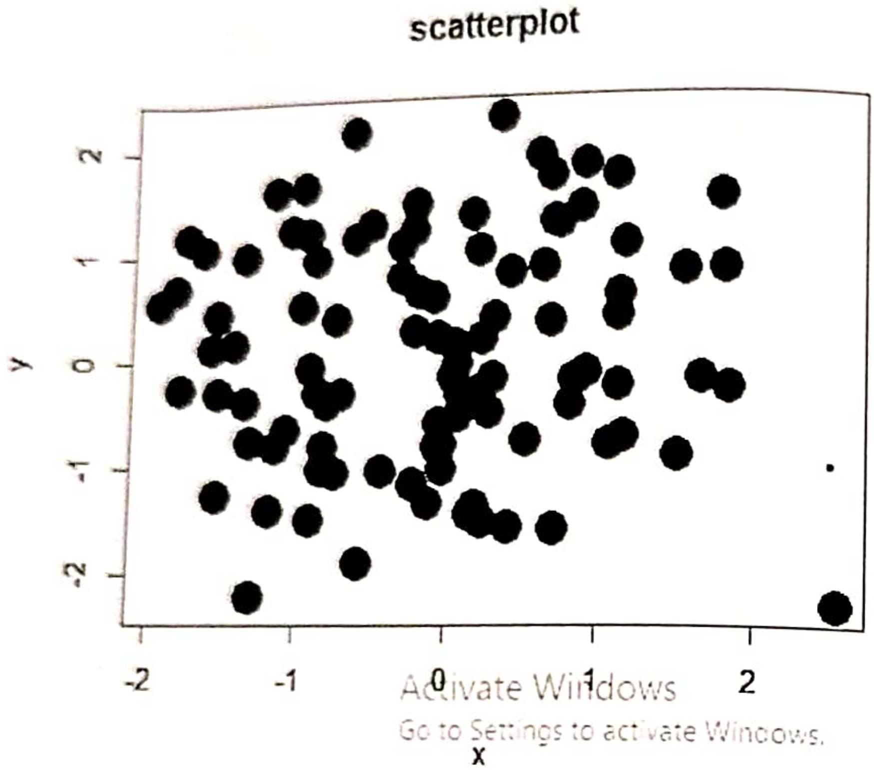
```
k=rnorm(10,mean=70,sd=3)
pop_range=range(k)
cat("range:",pop_range)
r=mean(k)
cat("mean:",r)
v=var(k)
cat("varience:",v)
s=sd(k)
cat("standard deviation::",s)
sumofk=sum(k)
cat("sum of k:",sumofk)
hist(k,main="histogram of
k",xlab='x',ylab='y',col="green",border="black")summ=summary(k)
cat("summary:",summ)
boxplot(k,main="Boxplot",col="blue",border="red")
scatter_data=data.frame(x=k,y=k+rnorm(2,mean=0,sd=5
plot(scatter_data$x,scatter_data$y,main="scatterplot",col="black",xlab='x',ylab='y',pch=2
```

output:

```
> k=rnorm(10,mean=70,sd=3)
> pop_range=range(k)
> cat("range:",pop_range)
range: 65.41912 75.54528> r=mean(k)
> cat("mean:",r)
mean: 70.09853> v=var(k)
> cat("varience:",v) varience:
11.29644> s=sd(k)
> cat("standard deviation::",s)
standard deviation:: 3.361018> sumofk=sum(k)
> cat("sum of k:",sumofk)sum of
k: 700.9853
> summ=summary(k)
> cat("summary:",summ)
summary: 65.41912 67.32304 70.70319 70.09853 72.42869 75.54528
```


histogram of k
y
0
1
2
3
64
66
68
70
72
74
76
x


Boxplot
66
68
70
72
74


scatterplot
y
2
1
0
-1
-2
-2
-1
0
1
2
x
Activate Windows
Go to Settings to activate Windows.

Week-6:.TESTING HYPOTHESIS:

a) Null hypothesis significance testing:

- **Null Hypothesis (H0)**: There is no significant difference between the sample mean and the population mean.
- **Alternative Hypothesis (Ha)**: There is a significant difference between the sample mean and the population mean.

p-value

The p-value is a crucial output of the one-sample t-test. It represents the probability of observing the sample mean (or something more extreme) if the null hypothesis is true. A small p-value (typically less than 0.05) suggests strong evidence against the null hypothesis, indicating that the sample mean is significantly different from the population mean.

State the Hypotheses – Stating the null and alternative hypotheses.

Formulate an Analysis Plan – The formulation of an analysis plan is a crucial step in this stage.

Analyze Sample Data – Calculation and interpretation of the test statistic, as described in the analys

Interpret Results – Application of the decision rule described in the analysis plan.

Hypothesis testing ultimately uses a p-value to weigh the strength of the evidence or in other words what the data.are about the population. The p-value ranges between 0 and 1. It can be interpreted in the followingway:

A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so you reject it.

A large p-value (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject it.

A p-value very close to the cutoff (0.05) is considered to be marginal and could go either way.

Decision Errors in R

The two types of error that can occur from the hypothesis testing:

Type I Error – Type I error occurs when the researcher rejects a null hypothesis when it is true. The term significance

level is used to express the probability of Type I error while testing the hypothesis. The significance level is

represented by the symbol α (alpha).

Type II Error – Accepting a false null hypothesis H0 is referred to as the Type II error. The term power of the test is

used to express the probability of Type II error while testing hypothesis. The power of the test is represented by the

symbol β (beta).

B)TESTING THE MEAN OF ONE SAMPLE

One Sample T-test

The One-Sample T-Test is used to test the statistical difference between a sample mean and a known or

assumed/hypothesized value of the mean in the population.

So, for performing a one-sample t-test in R, we would use the

syntax t.test(y, mu = 0)

example:

```
# Sample data: test scores of 15 students
student_scores <- c(68, 74, 80, 77, 82, 72, 76, 70, 88, 81, 66, 78, 75, 79, 84)
# Perform the one-sample t-test
t_test_result <- t.test(student_scores, mu = known_mean)
# Print the results
print(t_test_result)
```

output:

```
One Sample t-test

data: student_scores

t = 1.066, df = 14, p-value = 0.3045
alternative hypothesis: true mean is not equal to 75
95 percent confidence interval:
73.31335 80.01998
sample estimates:
mean of x
76.66667
```

c)Testing two samples:

t.test(group1, group2, var.equal=TRUE)

Note: By specifying **var.equal=TRUE**, we tell R to assume that the variances are equal between thetwo samples.

Example:

```
group1 <- c(8, 8, 9, 9, 9, 11, 12, 13, 13, 14, 15, 19)
group2 <- c(11, 12, 13, 13, 14, 14, 14, 15, 16, 18, 18, 19)

t.test(group1, group2, var.equal=TRUE)
```

output:

```
        Two Sample t-test

data:  group1 and group2
t = -2.5505, df = 22, p-value = 0.01823
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.5904820 -0.5761847
sample estimates:
mean of x mean of y
11.66667  14.75000
```

a)Linear models:

Linear models are a statistical technique for determining the relationship between two or more than two variables. There are two types of variables in Linear models – independent variable and dependent variable. Independent variables are also known as predictor variables. These are the variables that donot change. On the other side, the variables whose values change are known as dependent variables. These variables depend on the independent variables. Dependent variables are also known as response variables

b. Simple linear regression

Simple linear regression is used for finding the relationship between the dependent variable Y and the independent or predictor variable X. Both of these variables are continuous in nature. While performing simple linear regression, we assume that the values of predictor variable X are controlled. Furthermore, they are not subject to the measurement error from which the corresponding value of Y is observed.

To predict the weight of new persons, use the predict() function in R.

lm() Function

This function creates the relationship model between the predictor and the response variable.

Syntax

The basic syntax for lm() function in linear regression is –

lm(formula,data)

example:

```
x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)

y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)

# Apply the lm()

function.relation <-

lm(y~x)

print(relation)
```

output:

```
Call:

lm(formula = y ~

x)Coefficients:

(Intercept) x

-38.4551 0.6746
```

c) Multiple regression:

The value is dependent upon more than one exploratory varable in multiple regression.

Using the two explanatory variables, we can delineate the equation of multiple linear regression as follows: $yi = \beta0 + \beta1x1i + \beta2x2i + \varepsilon i$

General Multiple regression models can be represented as: $yi = \Sigma\beta1x1i + \varepsilon i$

lm() Function

This function creates the relationship model between the predictor and the response variable.Syntax

The basic syntax for lm() function in multiple regression is –

lm(y ~ x1+x2+x3...,data)

example:

```
# Sample data
set.seed(123)
x1 <-
rnorm(100)x2 <-
rnorm(100)
y <- 2*x1 + 3*x2 + rnorm(100)
# Fit the multiple linear regression model
model <- lm(y ~ x1 + x2)
# Print the summary of the model
summary(model)
```

output:

```
Call:
lm(formula = y ~ x1 +
x2)Residuals:
   Min     1Q  Median    3Q    Max
-2.49768 -0.71382 -0.01973 0.68135 2.21461
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.06804 0.10647 -0.639 0.524
x1        1.92474  0.10651 18.068 <2e-16 ***
x2        2.98729  0.10667 28.009 <2e-16 **
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.035 on 97 degrees of freedom

Multiple R-squared: 0.8723,   Adjusted R-squared: 0.8698

F-statistic: 344.7 on 2 and 97 DF, p-value: < 2.2e-16

d)Bias-varience trade-off-cross-validation:

The bias-variance trade-off is a fundamental concept in machine learning that refers to the balance between a model's ability to capture the underlying patterns in the data (bias) and its sensitivity to variations in the training data (variance).

Cross-validation is a technique used to estimate the performance of a machine learning model. It helps to assess how well the model generalizes to an independent dataset by simulating the process of training and testing on multiple subsets of the data.

In the context of the bias-variance trade-off, cross-validation can be used to find the optimal level of model complexity that minimizes the overall error.

- **High Bias (Underfitting)**: A model with high bias pays little attention to the training data and oversimplifies the problem. It has high error on both the training and test sets.
- **High Variance (Overfitting)**: A model with high variance fits the training data too closely and fails to generalize to new data. It has low error on the training set but high error on the testset.

Cross-validation helps in understanding and mitigating these issues by providing insights into the model's performance on unseen data, thus aiding in the selection of an optimal model complexity that balances bias and variance.

Week-8:Correlation

a) How to calculate correlation between two variables.

Correlation is a statistical measure of finding out linear dependence between the two variables. The values of the correlation coefficient range between -1 to +1. There can be two types of correlation – Positive Correlation and Negative Correlation.

Example:

```
res <- cor.test(my_data$wt, my_data$mpg,
 method = "pearson")
print(res)
cor(x, y, method = 'kendall')
```

output :

Pearson's product-moment correlation

data: my_data$wt and my_data$mpg t

= -9.559, df = 30, p-value = 1.294e-10

alternative hypothesis: true correlation is not equal to 095

percent confidence interval:

-0.9338264 -0.7440872

sample estimates:

cor

-0.8676594

[1] -0.03070707

b) How to make scatter plot:

A scatter plot is a type of data visualization that displays the relationship between two continuous variables.It consists of a series of points, each representing the value of one variable plotted on the horizontal (x- axis) against the value of another variable plotted on the vertical (y-axis). The position of each point on theplot corresponds to the intersection of the values of the two variables for a particular observation in the dataset.

Key features of scatter plots include:

1. **Horizontal Axis (X-axis)**: This axis represents the values of one variable (usually the independentvariable) being analyzed.

2. **Vertical Axis (Y-axis)**: This axis represents the values of the other variable (usually the dependent variable) being analyzed.

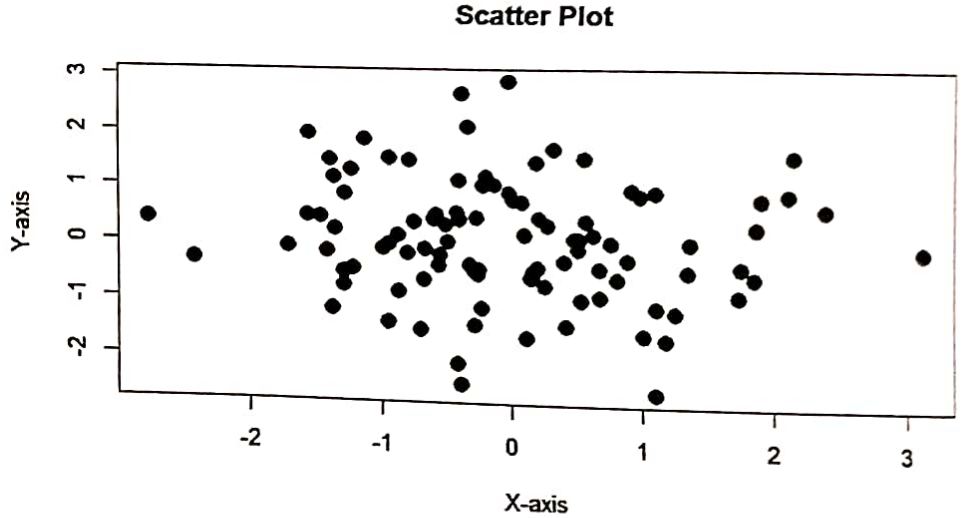
3. **Points**: Each point on the plot represents the intersection of the values of the two variables for a particular observation in the dataset.

4. Trend Line: In some cases, a trend line (or regression line) may be added to the plot to indicate the overall trend or relationship between the variables. This line can help identify whether the relationship between the variables is positive, negative, or neutral.

Let's take a look at a code snippet that creates a scatter plot in R.

```
> x <- rnorm(100)
> y <- rnorm(100)
> plot(x, y,
+      main = "Scatter Plot",
+      xlab = "X-axis",
+      ylab = "Y-axis",
+      col = "blue",
+      pch = 16,
+      cex = 1.5)
```

Output:



C) Use the scatter plot to investigate the relationship between twovariables:

1. Install and load necessary packages (if not already installed):

install.packages("ggplot2")

library(ggplot2)

2. Prepare your data. Make sure you have a data frame with the two variables you want to investigate.

3. Create a scatter plot using the `plot()` function or `ggplot2` package.
If using base R:

Assuming your data frame is named "data" and the variables are "x" and "y"

plot(data$x, data$y, xlab = "X Variable", ylab = "Y Variable", main = "Scatter Plot")If using ggplot2:

Assuming your data frame is named "data" and the variables are "x" and "y"

ggplot(data, aes(x = x, y = y)) +

geom_point() +

labs(x = "X Variable", y = "Y Variable", title = "Scatter Plot")

4. Interpret the plot:

- If the points on the plot are scattered randomly without any clear pattern, it suggests that there isno relationship between the two variables.

- If the points form a linear pattern (either positively or negatively sloped), it suggests a linear relationship between the variables.

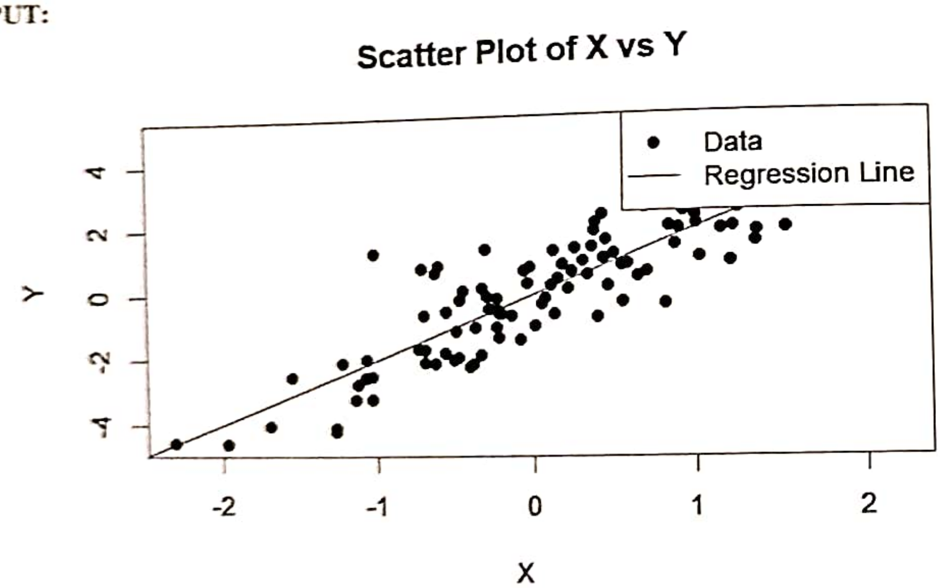
- If the points form a curve, it suggests a nonlinear relationship between the variables.

- You can also compute correlation coefficient and regression lines to quantify and visualize the relationship if needed.

Example Code:

```
# Generate some sample data
set.seed(123)
x <- rnorm(100)
y <- 2*x + rnorm(100)
# Create scatter plot
plot(x, y, main = "Scatter Plot of X vs Y", xlab = "X", ylab = "Y", col = "blue", pch = 16)
# Add a regression line
abline(lm(y ~ x), col = "red")
# Add a legend
legend("topright", legend = c("Data", "Regression Line"), col = c("blue", "red"), pch = c(16, NA), lty =c(NA, 1))
```

OUTPUT:



WEEK-9: TESTS OF HYPOTHESIS:

a) perform tests of hypothesis about the mean when the variance is known

SOURCE CODE:

```
data <- rnorm(100, mean = 5, sd = 2)
mu <- 5
x_bar <- mean(data)
n <- length(data)
sigma <- 2
t_stat <- (x_bar - mu) / (sigma / sqrt(n))
p_value <- 2 * pt(-abs(t_stat), df = n - 1)
cat("Test Statistic:", t_stat, "\n")
cat("P-value:", p_value, "\n")
```

OUTPUT:

```
        One Sample t-test
data: data
t = -0.125, df = 99, p-value = 0.9006
alternative hypothesis: true mean is not equal to 5
95 percent confidence interval:
4.717465 5.202096
sample estimates:
mean of x
4.959781
Test Statistic: -0.1249982
P-value: 0.9005892
```

a. Compute p value:

SOURCE CODE:

```
set.seed(123)
sample_data <- rnorm(20, mean = 10, sd = 2)
t_test_result <- t.test(sample_data, mu = 10)
p_value <- t_test_result$p.value
cat("P-value:", p_value, "\n")
```

OUTPUT:

Test Statistic: 0.6376906

P-value: 0.5251472

B) Explore the connection between the critical region, the test statistic and the p value

The connection between the critical region, the test statistic, and the p-value is fundamental in hypothesis testing. Understanding this relationship helps in making decisions about the null hypothesis.

1. Critical Region:
- The critical region is a set of possible values of the test statistic that would lead to rejection of the null hypothesis at a given significance level (\(\alpha \)).
- It represents the extreme values of the test statistic for which the evidence against the null hypothesis is strong enough to warrant rejection.

2. Test Statistic:
- The test statistic is a numerical value calculated from sample data, which measures the degree of agreement between the sample and the null hypothesis.
- It serves as the basis for deciding whether to reject the null hypothesis or not.

3. P-value:
- The p-value is the probability of observing a test statistic as extreme as, or more extreme than, the one actually observed, under the assumption that the null hypothesis is true.
- It measures the strength of evidence against the null hypothesis.
- A small p-value indicates strong evidence against the null hypothesis, suggesting that the observed data are unlikely to occur if the null hypothesis is true.

The connection between these three elements can be summarized as follows:

- If the test statistic falls into the critical region (i.e., its value is extreme enough), the p-value associated with that te statistic will be small.
- Conversely, if the test statistic falls outside the critical region (i.e., its value is not extreme enough), the p-value associated with that test statistic will be large.
- In decision-making:
  - If the p-value is less than the chosen significance level (\(\alpha \)), typically 0.05, then the evidence against the null hypothesis is considered strong enough to reject it.
  - If the p-value is greater than or equal to \(\alpha \), then there is not enough evidence to reject the null hypothes

Week-10

Estimating a linear relationship demonstration on a statistical model for a linear relationship

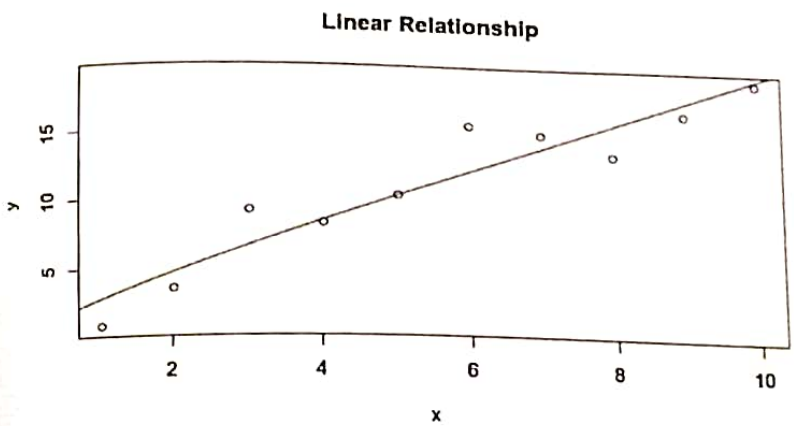
a)Least square estimates

Source code:

```
set.seed(123)
x <- 1:10
y <- 2 * x + rnorm(10, mean = 0, sd =
2)model <- lm(y ~ x)
summary(model)
plot(x, y, main = "Linear Relationship", xlab = "x", ylab = "y")
abline(model, col = "red")
```

output:

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

Residuals:
Min      1Q  Median      3Q     Max
-2.2695 -1.1248 -0.2785  0.7707  3.3628
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.0509     1.3346  0.787  0.454
x             1.8361     0.2151  8.537 2.73e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.954 on 8 degrees of freedom
Multiple R-squared:  0.9011,    Adjusted R-squared:  0.8887
F-statistic: 72.87 on 1 and 8 DF,  p-value: 2.729e-05
```



OUTPUT:

Test Statistic: 0.6376906

P-value: 0.5251472

B) Explore the connection between the critical region, the test statistic and the p value

The connection between the critical region, the test statistic, and the p-value is fundamental in hypothesis testing. Understanding this relationship helps in making decisions about the null hypothesis.

1. Critical Region:
- The critical region is a set of possible values of the test statistic that would lead to rejection of the null hypothesis at a given significance level (\(\alpha \)).
- It represents the extreme values of the test statistic for which the evidence against the null hypothesis is strong enough to warrant rejection.

2. Test Statistic:
- The test statistic is a numerical value calculated from sample data, which measures the degree of agreement between the sample and the null hypothesis.
- It serves as the basis for deciding whether to reject the null hypothesis or not.

3. P-value:
- The p-value is the probability of observing a test statistic as extreme as, or more extreme than, the one actually observed, under the assumption that the null hypothesis is true.
- It measures the strength of evidence against the null hypothesis.
- A small p-value indicates strong evidence against the null hypothesis, suggesting that the observed data are unlikely to occur if the null hypothesis is true.

The connection between these three elements can be summarized as follows:

- If the test statistic falls into the critical region (i.e., its value is extreme enough), the p-value associated with that test statistic will be small.
- Conversely, if the test statistic falls outside the critical region (i.e., its value is not extreme enough), the p-value associated with that test statistic will be large.
- In decision-making:
  - If the p-value is less than the chosen significance level (\(\alpha \)), typically 0.05, then the evidence against the null hypothesis is considered strong enough to reject it.
  - If the p-value is greater than or equal to \(\alpha \), then there is not enough evidence to reject the null hypothesis.

Week-10

Estimating a linear relationship demonstration on a statistical model for a linear relationship

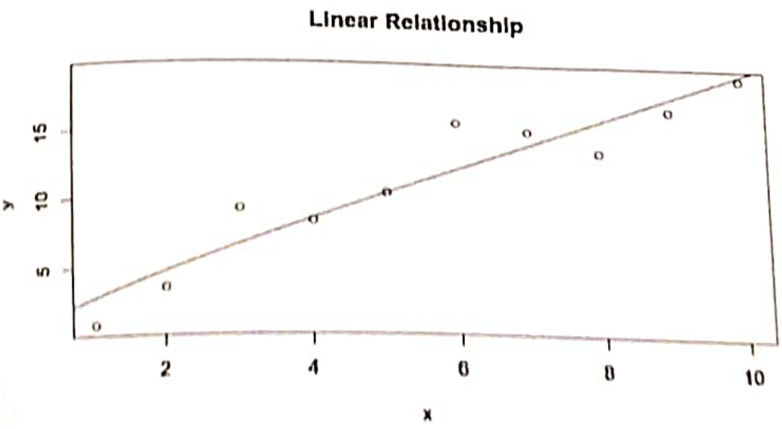
a)Least square estimates

Source code:

```
set.seed(123)
x <- 1:10
y <- 2 * x + rnorm(10, mean = 0, sd =
2)model <- lm(y ~ x)
summary(model)
plot(x, y, main = "Linear Relationship", xlab = "x", ylab = "y")
abline(model, col = "red")
```

output:

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

Residuals:
Min      1Q  Median      3Q     Max
-2.2695 -1.1248 -0.2785  0.7707  3.3628
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.0509     1.3346  0.787  0.454
x            1.8361     0.2151  8.537 2.73e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.954 on 8 degrees of freedom
Multiple R-squared:  0.9011,    Adjusted R-squared:  0.8887
F-statistic: 72.87 on 1 and 8 DF,  p-value: 2.729e-05
```



b)The R function lm:

```
x <- c(1, 2, 3, 4, 5)

y <- c(2, 3, 4, 5, 6)

model <- lm(y ~

x)

summary(model

)
```

output:

```
Call:

lm(formula = y ~

x)Residuals:

    1     2     3     4     5

-8.5e-16  1.0e+00 -2.0e-15 -1.0e+00  1.0e-15

Coefficients:

         Estimate Std. Error t value Pr(>|t|)

(Intercept)  1.0000     1.1952     0.837

0.453

x            1.0000     0.3410     2.932  0.064 .

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1 on 3 degrees of freedom

Multiple R-squared:  0.6923,   Adjusted R-squared:  0.6154

F-statistic: 8.592 on 1 and 3 DF,  p-value: 0.06395
```

WEEK 11:Apply -Type Function

A) Defining user defined classes and operations model in r.

USER-DEFINED CLASS:-

In r we can create user defined class using various system varioussystems s3 s4 or refernce classes r6.

S3 classes:-

S3 classes are the simple and flexible allowing you to defined classes quicklythen are often used for simple object oriented programming in r.

Example:-

```
if (!require Namespace ("methods", quickly=TRUE))
{
install Packages ("methods")
}
library (methods)
setclass ("Person", slots=C(
name=" character",
age =" numeric")
)

Person1 <-new("Person", name= "john",age=30)
print(Person1@name)
print(person1@age)
OUTPUT:-
"john"
30
```

Various operations:-

Once we defined a class in r you can perform various on objects of that class1.Accessing slots

2. Updating slots
3. Defining methods
4. Inheritance
5. Overriding methods
6. validity checking

Accessing slots:- we can access slots of an object using @operator

Ex:-

```
Person1@name
```

Person1@age

Updating slots:-

You can update the values of slots using the at operator or directly assigning name newvalues

Ex:-

Person1@name<-"Alice"

Person1@age<-35

Defining Methods:-

We can defined method specific to your class using the "setMethod()"

Ex:-

```
setMethods("print",signature="person",function(object){ cat("Name:",object@name,"\n")
 cat("Age:",object@age,"\n")
})
```

Inheritance:-

We can define subclass that inheritance from super classEx:-

```
setclass(
"student",
slot=c(name="charcter",age="numeric",student_id="charcter"), contains="person")
student1<-new("Student",name="Emma",
age=20,student_id="512345");
```

Overriding methods:-

We can override methods for your custom class by defining method with the same nameas there in the generic functions but to your class.

EX:-

```
setMethod("print",signature="Persin",function(object){
cat("person Information:\n")
cat("Name:",object@name,"\n")
cat(Age:",object@age,"\n")
})
print(person1)
```

validity checking:-

To perform validity checking in R, you can use various techniques depending on your specific requirements and the type of data you're working with. Here's a general approach:

1. *Missing Values*: Check for missing values using functions like is.na() or complete.cases().

Check for missing values

any(is.na(data))

2. Data Type Ensure variables are of the correct data type using functions like class() or typeof().

```
# Check data type
class(data)
```

3. *Range*: Check if values fall within a certain range using logical operators like <, >, <=, >=.

```
# Check range
any(data < min_value | data > max_value)
```

4. Pattern Matching: Validate strings or patterns using regular expressions with functionslike grep() or grepl().

```
# Check pattern matching
any(grepl("pattern", data))
```

5. Custom Checks: Implement custom functions tailored to your specific validationcriteria using if statements or packages like assertthat.

```
# Custom validity check
custom_check <- function(x) {if
(condition) {
   # Handle invalid data
  } else {
   # Data is valid

  }
 }

 # Apply custom check
 custom_check(data)
```

Combine these techniques as needed to perform comprehensive validity checking on your data. Adjust the conditions and criteria according to your specific requirements andthe nature of your data.

MODELS:-

When creating user-defined models in R, you have the flexibility to define various types of models based on your specific requirements and the nature of your data. Here are some common types of models you can define:

1. Linear Regression Model: A simple model that represents a linear relationship between independent variables and a continuous dependent variable.

2. Logistic Regression Model: Used for binary classification problems, where the outcomevariable is categorical with two levels.

3. Generalized Linear Model (GLM): An extension of linear regression that allows for non-normal error distributions and link functions, such as Poisson regression for count data orlogistic regression

for binary data.

4.Decision Tree Model: A tree-like model that recursively splits the data into subsets basedon t values of predictor variables, used for both classification and regression tasks.

METHODS:-

we can defined our own methods using rExamplesare:-

1. Method Dispatch
2. Defining methods
3. Functionality Extension4 Method overrinding.

b) Customising the users environment.

In r we can customize the users environment by setting various options users environmentbyse various options modifying the search path defining path and controlling the behaviours of r sessions

The various oprations of modifying the user environment is

1.setting option

2.Defining custom functions

3.customizing the search path

4.customizing start-up behaviour

5.customizing output

6.setting working directory

Settiing options:-

We can set global options that affect the behaviour of r using the options() functionEX:-

Options(digit=4)

Options(warm=1)

Defining custom Function:-

We can define your own functions and the store them In the user enviroments this allowsyou to create reusuable code and customize behavior.

EX:-

My_function<-function(x)

{

}

Customizing the search paths:-

We can modify the search path to control where r looks for objects

Ex:-

Attach("~/mydirectory")

Setting working Directory:-

We can set the working directory using the setwd() function to control where R reads andwrites files by defaults

Ex:-setwd("~/mydirectory")

C)CONDITIONAL STATEMENTS:-

Conditional statements in R allow the control of flow in programs based on certain conditions. Using these statements different blocks of code can be executeddepending on whether a condition is true or false.

If Statement

The if statement is used to execute a block of code if a condition is true. The general syntax is:

```
if (condition) {
  # Code to be executed if condition is true
}
```

If-Else Statement

The if-else statement allows the execution of a block of code if a condition is true, and another block of code if the condition is false. The general syntax is:

```
if (condition) {
  # Code to be executed if condition is true
} else {
  # Code to be executed if condition is false
}
```

Example:-

```
x <- 10

if (x > 5) {
  print("x is greater than 5")
}
x <- 3

if (x > 5) {
  print("x is greater than 5")
```

```
} else {
print("x is less than or equal to 5")
}
```

d)Loops and Iterations:-

R For-Loop on a List

```
# Create a list of numbers
my_list <- list(1, 2, 3, 4, 5)
# Loop through the list and print each elementfor (i in
seq_along(my_list)) {
  current_element <- my_list[[i]]
  print(paste("The current element is:", current_element))

}
```

```
while
  (test_expression)
  {# Block of code
}
```

Example:-i

```
<- 1
while (i
  < 6)
  {
  print
  (i)
  i <- i + 1
}
```

Output:-

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

WEEK – 12 : statistical function in R:

a)write Demonstrate statistical functions in R

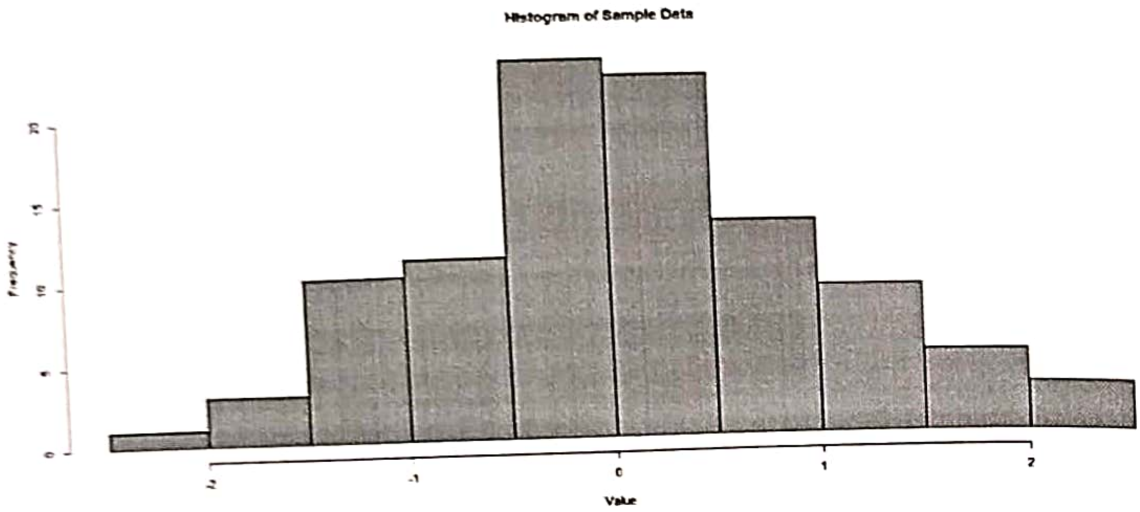
Statistical functions are mathematical operations or procedures used to analyze and summarize data in statistics. They are used to describe, summarize, and interpret data to gain insights into the underlying patterns and relationships. Some common statistical functions include measures of central tendency (e.g., mean, median, mode), measures of dispersion (e.g., standard deviation, variance), measures of distribution shape (e.g., skewness, kurtosis), and summary statistics (e.g., minimum, maximum, quartiles).

Example:

```
# Generate some sample data
set.seed(123)
data <- rnorm(100)

# Mean
mean_value <- mean(data)
print(paste("Mean:", mean_value))

# Median
median_value <- median(data)
print(paste("Median:", median_value))

# Standard Deviation
sd_value <- sd(data)
print(paste("Standard Deviation:", sd_value))

# Variance
var_value <- var(data)
print(paste("Variance:", var_value))

# Summary statistics
summary_stats <- summary(data)
print(summary_stats)

# Histogram
hist(data, main = "Histogram of Sample Data", xlab = "Value", ylab = "Frequency", col = "skyblue")
```

output:

```
"Median: 0.0617563090775401"
"Standard Deviation: 0.912815879680979"
"Variance: 0.833232830197759"
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-2.30917 -0.49385  0.06176  0.09041  0.69182  2.18733
```



b)statistical inference , contingency tables chi-square goodness of fit ,regression , generalized linear models and advanced modelling methods.

Statistical inference:

Statistical inference is the process of making predictions, decisions, or conclusions about a population based on sample data. It involves using statistical methods to draw inferences about population parameters, such as meansor proportions, from sample statistics. Common techniques include hypothesis testing, confidence intervals, and regression analysis.

statistical inference can be performed using a variety of built-in functions and packages. Commonly used packages for statistical inference include:

1. stats
2. dplyr and tidyr

Contingency tables:

Contingency tables, also known as cross-tabulations, are used to summarize the relationship between two categorical variables. In R,

Example:

```
# Create some example data
gender <- c("Male", "Female", "Female", "Male", "Male", "Female")
smoker <- c("Yes", "No", "Yes", "No", "Yes",
"Yes")# Create a contingency table
cont_table <- table(gender, smoker)

# Display the contingency table
```

```
cont_table
```

Output:

```
    smoker
gender  No Yes
  Female  1  2
  Male    1  2
```

This will produce a contingency table showing the frequency counts for each combination of categories in the two variables (gender and smoker). You can also compute row percentages, column percentages, or overall percentages using additional functions like **prop.table()**.

Example:

```
# Row percentages
row_percentages <- prop.table(cont_table, margin = 1)
print(row_percentages)

# Column percentages
col_percentages <- prop.table(cont_table, margin = 2)
print(col_percentages)

# Overall percentages
overall_percentages <- prop.table(cont_table)
print(overall_percentages)
```

Output:

```
    smoker
gender        No       Yes
  Female 0.1666667 0.3333333
  Male   0.1666667 0.3333333
    smoker
gender  No Yes
  Female 0.5 0.5
  Male   0.5 0.5

    smoker
gender        No       Yes
  Female 0.1666667 0.3333333
  Male   0.1666667 0.3333333
```

Chi Square goodness of fit:

you can perform a Chi-square goodness-of-fit test using the **chisq.test()** function. This test is used to assess whether the observed frequency distribution of categorical data differs from the expected frequency distribution.

Example:

```
# Example data: observed frequencies
observed <- c(20, 30, 25, 15)

# Expected frequencies (in this example, assuming equal probabilities)expected
<- rep(sum(observed) / length(observed), length(observed))

# Perform the Chi-square goodness-of-fit test
chi_square_test <- chisq.test(observed, p = expected)

# Print the results
print(chi_square_test)
```

Output:

```
Chi-squared test for given probabilitiesdata:
observed
X-squared = 2.5, df = 3, p-value = 0.4757
```

Regression:

you can perform regression analysis using various functions depending on the type of regression you want to conduct. Here's a basic example of linear regression using the built-in lm() function:

Example:

```
# Example data
x <- c(1, 2, 3, 4, 5)
y <- c(2, 3, 4, 5, 6)

# Perform linear regression
model <- lm(y ~ x)

# Summary of the regression model
summary(model)
```

Output:

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

```
         1          2          3          4          5
 7.015e-16 -9.462e-16 -2.452e-17  8.156e-17  1.876e-16

Coefficients:
            Estimate Std. Error  t value Pr(>|t|)
(Intercept) 1.000e+00  7.241e-16 1.381e+15   <2e-16 ***
x           1.000e+00  2.183e-16 4.581e+15   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.904e-16 on 3 degrees of freedom
Multiple R-squared:      1,  Adjusted R-squared:      1
F-statistic: 2.098e+31 on 1 and 3 DF,  p-value: < 2.2e-16
```

Generalized linear models:

eTo perform generalized linear regression in R, you can use the **glm()** function, which allows you to specify a variety of distributions and link functions suitable for different types of response variables. Here's an example of fitting a generalized linear model:

Example:

```
# Example data
x <- c(1, 2, 3, 4, 5)
y <- c(0, 1, 0, 1, 0)  # Binary response variable
# Perform logistic regression (binomial family with logit link)
model <- glm(y ~ x, family = binomial(link = "logit"))

#Summary of the generalized linear model
summary(model)
```

Output:

```
Call:
glm(formula = y ~ x, family = binomial(link =
"logit"))Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.055e-01  2.141e+00  -0.189     0.85
x            1.433e-16  6.455e-01   0.000     1.00
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6.7301  on 4  degrees of freedom
```

Residual deviance: 6.7301 on 3 degrees of freedom

AIC: 10.73

Number of Fisher Scoring iterations: 4

Advanced modeling methods:

Advanced modeling methods in R encompass a wide range of techniques beyond simple linear regression or generalized linear models. Some popular advanced modeling methods include:

Tree-based methods:

- Decision Trees: Build a tree structure to make predictions based on feature splits.
- Random Forest: Ensemble method that constructs multiple decision trees and combines their prediction
- Gradient Boosting Machines (GBM): Sequentially build trees to minimize errors of the previous trees.

Support Vector Machines (SVM): Supervised learning models used for classification and regression analysis. **Neural Networks**:

- Multilayer Perceptrons (MLP): Basic feedforward neural networks with multiple layers.
- Convolutional Neural Networks (CNN): Designed for image recognition tasks.
- Recurrent Neural Networks (RNN): Suitable for sequence data like time series or natural language processing.

Cluster Analysis:

K-Means Clustering: Partition data into clusters based on similarity.

Hierarchical Clustering: Build a hierarchy of clusters.

Dimensionality Reduction:

- Principal Component Analysis (PCA): Reduce the dimensionality of the data while preserving varia
- t-Distributed Stochastic Neighbor Embedding (t-SNE): Non-linear dimensionality reduction tech often used for visualization.

Time Series Analysis:

- ARIMA (AutoRegressive Integrated Moving Average): Suitable for univariate time series forecastin
- Prophet: Developed by Facebook for forecasting time series data with trend, seasonality, and ho effects.