INTRODUCTION TO R Studio

Q1. How to Install R Studio? What is the latest version of R. Give details?
Q2. RStudio Layout with snapshot. Explain the purpose of all panes
Q3. Who designed and developed the R language?
Q4. Are variables 'H' and 'h' same in R language?
Q5. What is a package? What are two major parts of R language?
Q6. How is a package installed and accessed?
Q7. What is CRAN?
Q8. What do you mean by Object Assignment? Elucidate difference between Left-side and Right-side Assignment with output. What is Assignment operator in Rstudio
Q9. Explain The c( ) function. Write a sample command for the same
Q10. What is paste( ) function used for? Write a sample command for the same
Q11. What is %>% operator used for? Write a sample command for the same
Q12. What is meant by ">", "+" and [1]in R console?
Q13. Write the code to identify odd or even numbers using IF statement
Q14. Write the code to identify minimum number among three numbers using Nested IF statement
Q15. Display grade of student using nested if command for following criterion (customize student name). Output example: Kritika has scored "A" Grade

IF PERCENTAGE	THEN GRADE
>= 90	A+
>=75<90	A
>=50<75	B
<50	F

Q16. How to Import of Data Sheet in Excel
Q17. What are Data Frames, Matrices, Vectors
Q18. Write a command to create Data Frames, Matrices, Vectors
Q19. Name some built-in functions with their description.
Q20. Create a function for multiplication but no return value
Q21. Write a command for Accessing Rows and Columns
Q22. Create a data frame by your surname of 12 rows and 8 columns.
Q23. Write a command to access non-consecutive rows or columns, use 'c() '. For example, to obtain rows 1 to 5, 7 and 11 and columns 3 to 4 and 7.
Q24. Add one new column and drop two existing columns 4 and 5
Q25. Drop rows 1, 3 and 4
Q26. Write a command to calculate the number of columns and number of rows.
Q27. What is the command to access built in datasets? What is the command to get description of a built-in datasets

Q28. Access Titanic dataset and Execute commands to evaluate whether the evacuation strategy was fair or not. If biased, state which gender, age group and class was most favored. Analyze using cross tabulations

Q29. Calculate correlation by importing data from excel. Determine whether there is a positive or a negative correlation in Advertisement in month and Sales in crores.

Advertisement	Sales
32	5
54	10
67	15
65	20
98	24
112	34
62	25
34	34

Packages in R Programming

The tidyr Package

Q30.Apply Important Functions (gather, separate, unite, spread, fill, full_seq, drop_na, and replace_na) in “tidyr Package” for following dataset

S.No	Group.1	Group.2	Group.3
1	23	117	29
2	345	89	101
3	76	66	239
4	212	334	289
5	88	90	176
6	199	101	320
7	72	178	89
8	35	233	109
9	90	45	199
10	265	200	56

The dplyr Package

Q31.Apply Important Functions (filter, arrange, select, rename, mutate and transmute, sample_n and sample_frac) for following column heads with 5 data rows:

Name	Semester 1 marks	Semester 2 marks	Subject maths in 12th (Y/N)

Data Visualization in R Studio

Quick plot with ggplot2

Q32. Generate BCOM marks data containing the sections and overall percentage (5 sections ranging from A to E), with 60 students in each section

Q33. Create following Quick plots with customized labels (with your name and DOB) for both the axis and Main title of the chart

Histogram plot

- Histogram fill color by group (Section)
- Basic density plot
- Density plot line color by group (Section) and change line type
- Draw a plot using data from numeric vectors where X contains values ranging from 10 to 20 and Y is square of X
- Add to the dot plot for X & Y

Q34. Activate Motor Trend Car Road Tests dataset. Using the given data set prepare following quick plots:

- Scatter plots with smoothed line for Miles/(US) gallon on y axis and Weight (lb/1000) on x axis
- Scatter plots (for Miles/(US) gallon on y axis and Weight (lb/1000) on x axis) with Smoothed line by groups (Number of cylinders)
- Scatter plots with colors for Miles/(US) gallon on y axis and Weight (lb/1000) on x axis
- Scatter plots (for Miles/(US) gallon on y axis and Weight (lb/1000) on x axis) with colors by groups (Number of gears)
- Scatter plots (for Miles/(US) gallon on y axis and Weight (lb/1000) on x axis) with Smoothed line and colors by groups (Number of gears)
- Scatter plots (for Miles/(US) gallon on y axis and Weight (lb/1000) on x axis) with Smoothed line and the point shape by groups (Number of gears)

Q35. Provide 5 commands for Descriptive statistics

Q36. Provide summary statistics for the MTCARS dataset while displaying a count summary of categorical variables.

HYPOTHESIS TESTING using R studio

For all test import excel with the given data saved by your name_test

T-TEST

A. One Sample t- Test using dummy (One- Tailed)

File name example: kritika_ttest

Problem 1:

To determine that the population mean of age is equal to 40 at α=0.05.

Age
18
24
56
78
67
24
65
89
25
23
45
65
78
55
32
33
44
26
56
89
44
34
3
4
56
56
76

B. Two Sample t- Test

File name example: kritika_ttest2

Problem 1:

To analyze that the time spent by full time students in studying statistics is different as time spent by part time students.

Full time	Part time
3.2	3.1
1.5	3.4
6.5	4.6
0.2	2.8
3.7	2.3
3.3	1.5
1.7	3.8
3.6	9.5
3.8	4.3
5.3	2.7
6.9	3.4
3.6	1.6
1.7	3.2
2.2	4.2
7.2	3.9
3.9	1.2
1.9	
5.3	

C. Two Sample t- Test

Problem 1:

Is there sufficient evidence to suggest that the mean time to exhaustion is greater after chocolate milk than after carbohydrate replacement drink? Use a significance level of 0.05. (Use μ_{CM}-μ_{CD} in hypothesis statements)

Cyclist	Chocolate Milk	Carbohydrate Replacement Drink
1	50.46	32.9
2	47.08	20.1
3	57.51	41.67
4	46.6	32.69
5	49.1	46.33
6	27.5	31.63
7	23.87	50.61
8	28.65	14.99
9	35.37	20.11

D. Paired t- Test

Problem 1:

Coaching was given to students for Statistical software after their result was evaluatedin January in order to improve their performance in April exams. Determine if the coaching was successful. (α = 0.05%)

Jan	May
45	56
54	57
44	32
56	67
34	44
45	34
34	34
56	76
45	56
54	45
67	55
56	87
56	66
56	65
76	45
45	76

E. Two Sample t Test

Problem 1:

To analyse that there is a significant difference between the marks scored by class groups A & B in mathematics at α=10%.

Group A	Group B
76	95
55	97
76	87
76	89
89	56
65	98
76	76
88	56
78	76
87	56
87	87
65	76
76	87
89	88
65	76
78	66
69	45
65	76
89	77

F. F Test

Problem 1:

Determine whether or not there is a significant difference between variances of two data sets.

Group 1	Group 2
150	170
125	165
160	130
130	155
160	125
125	150

G. One Way Anova

Problem 1:

The marks for 3 different groups in Economics, Science, History are given.

Determine whether there is a significant difference between the means of population.

Economics	Science	History
45	69	75
53	54	20
54	58	45
53	64	42
43	64	50
44	55	39
56	45	55
52		39
20		20

H. Chi Square Test

Problem 1:

Determine whether brand preference is independent of age group.

Age/ Brand	Brand 1	Brand 2	Brand 3
15-25	75	56	72
26-35	60	40	64
36-45	45	52	50
46-55	55	35	45