FYUG Even Semester Exam., 2024

STATISTICS

(2nd Semester)

Course No. : STASEC–151T

(Statistical Data Analysis using R)

Full Marks : 50
Pass Marks : 20

Time : 2 hours

The figures in the margin indicate full marks for the questions

SECTION—A

Answer any *fifteen* questions : 1×15=15

1. Write the use of with( ) command in *R*.

2. Write the use of by( ) command in *R*.

3. Write one feature of *R*.

4. How to remove duplicate rows in *R*?

5. Which package is used to remove duplicate rows in *R*?

6. Write the use of scan( ) function in *R*.

7. How to add labels in a plot?

8. How to export a plot in *R*?

9. What is normal probability plot?

10. Which function is used to draw QQ plot?

11. What is the relation between median and percentile?

12. Which function is used to find variance in *R*?

13. Write the name of the package which is used to compute skewness in *R*?

14. Correlation coefficient lies between _____.
(Fill in the blank)

15. Which function is used to compute covariance in *R*?

16. Write a note on scatterplot.

17. Write down the equation for simple linear regression with explanation.

18. Write the use of max( ) function in *R*.

19. What is the use of '$' operator in *R*?

20. Which function is used to fit best linear line in *R*?

SECTION—B

Answer any *five* questions : 2×5=10

21. What is *R*? Write its two advantages.

22. What is *R*-Studio? How to install *R*-Studio in your system?

23. Define xlab, ylab, col, main with respect to data visualization in *R*.

24. Define Boxplot with explaining all of its parts.

25. Define quartiles.

26. Write *R*-code to find 60th and 85th percentile for the following data :

5, 7, 9, 15, 23, 10

27. Write a note on kurtosis explaining its different types.

28. Write *R*-code to find covariance for the following data :

x :	10	15	13	17	21
y :	9	7	5	12	11

29. Write *R*-code for computing range for the following data :

8, 9, 3, 10, 7, 2, 3, 8, 12, 15, 17, 11

30. Write a note on R^2 and adjusted R^2.

SECTION—C

Answer any *five* questions : 5×5=25

31. Write a note on different operators in *R*.

32. Write down the steps followed in *R* for data cleaning. Also write any two disadvantages of using *R*.

33. The following table shows the wage distribution of certain company :

Weakly wage	*No. of employees*
0–10	17
10–20	25
20–30	13
30–40	33
40–50	15
50–60	19
60–70	21

Write down the *R*-code for drawing—

(a) bar diagram;

(b) pie diagram;

(c) histogram;

(d) frequency polygon.

34. Draw an ogive of both less than type and more than type using *R* for the following data :

CI	:	0–5	5–10	10–15	15–20	20–25
Frequency	:	7	2	3	5	1

35. Write *R*-code to find mean, median, variance, standard deviation and standard error for the following data with writing all their formulas :

7, 3, 5, 9, 22, 7, 11, 13, 15, 10

36. Write *R*-code to generate random number from—

(a) '5' random number from uniform distribution;

(b) '7' random number between 50 and 100;

(c) '7' random number from normal distribution.

Also calculate weighted mean and geometric mean from the following data :

x	:	3	7	11	13	2	7
Weight	:	0·3	0·1	0·1	0·2	0·2	0·1

37. Write *R*-code to find Karl Pearson and Spearman rank correlation for the following data :

80, 120, 45, 33, 37, 11, 15, 21

38. Write a note on skewness. Also write *R*-code to find skewness for the following data :

7, 3, 5, 11, 17, 21, 23

39. Write down the *R*-code for performing multiple linear regression to the following data :

Total sales (in ₹'000)	*Hours worked By Employees*	*Years of Experience*	*Training attended by the employees*
2,700	54	7	5
2,500	57	5	5
1,500	47	8	3
750	51	4	2
1,250	59	6	2

40. Define simple and multiple linear regression model. For the model $y = a + bx$, write *R*-code to find the values of a and b for the following data :

y (dependent variable)	:	7	5	3	2	11
x (independent variable)	:	4	2	9	3	2

★★★