17. The infant mortality rate of the 8 north eastern states of India are as follows:

State	IMR
Assam	36
Arunachal Pradesh	24
Meghalaya	29
Manipur	3
Nagaland	17
Mizoram	4
Sikkim	5
Tripura	18

Represent the data by dot plot in a order such that the best state appears at the top & the values are also displayed around the corresponding dots.

18. Fit the following data to log-normal distribution and test the goodness of fit using the K-S test.

7.21, 5.63, 4.78, 11.45, 6.17, 9.32, 8.5, 12.61, 10.24, 5.91, 7.88, 6.45, 13.28, 4.39, 9.75

Also represent the fitting graphically.

Section - C

19. Note Book. 10

20. Viva-Voce. 10

★★★

PG Even Semester Examination, May 2025

STATISTICS

2nd Semester

Course No. : STS - 555

R-Programming (ALIF)

Full Marks : 70

Pass Marks : 28

Time : 3 hours

The figures in the margin indicate full marks for the questions

Section - A

(Answer any 5 questions from this unit)

2x5=10

1. What are the assumptions of multiple regression?
2. Define odds-ratio. What is its importance?
3. What is a correlation matrix?
4. What is ANCOVA? When is it used?
5. What are the assumptions of binary-logistic regression?
6. If y is the dependent and x is the independent variable then write the
   (i) cubic equation of y on x
   (ii) exponential equation of y on x.
7. What is the difference between scatter plot and dot plot?

(Turn Over)

8. What do you mean by data ink ratio in connection to graphs?

9. What is a bubble plot?

10. Explain the concept of multicollinearity.

Section - B

Answer any four (4) questions from this section.

4x10 = 40

11. For the data provided below perform lowess smoothing in R and plot the graph along with original values.

Time :	1	2	3	4	5	6	7	8	9	10
Value :	0.8	1.05	0.53	0.63	1	0.58	1.06	1.25	0.8	1
Time :	11	12	13	14	15	16	17	18	19	20
Value :	1.16	1.08	1.25	0.75	0.75	0.3	0.63	0.91	0.72	0.9

12. For the following values of x and y fit the linear function and quadratic function with y as the dependent variable and x as the independent variable in R. Also fit the lowess curve in R.

Compare between the linear and quadratic functions in R and hence derive the better fit.

x :	3	7	8	9	11	15	13
y :	12	18	21	27	32	49	36

13. Fit linear, quadratic and exponential models to the data set and comment on the best fit. Also compare the curve in the scatter plot.

Overseas - Spinner		Overseas - Fastbowler	
Balls	E.R	Balls	E.R
330	6.69	320	8.3
209	8.32	102	8.32
262	8.4	201	10.17
126	10	366	9.27
301	9.14	246	10.14
		144	10.62
		211	8.61
		206	9.26

16. Perform appropriate non-parametric test to check if the running speed of 12 players have changed before and after a training program.

Player No.	Running Before	Speed (km/hr) After
1	12	14
2	14	15
3	16	16
4	13	12
5	12	14
6	15	16
7	13	13
8	17	16
9	16	15
10	14	19
11	12	14
12	13	15

(*Turn Over*)

Day (x)	New-infected-cases
1	73
2	81
3	84
4	107
5	114
6	137
7	151
8	173
9	223
10	283
11	360
12	434
13	519
14	606
15	694
16	724

14. The data provided below pertains to the height weight, age and average time to daily newspaper rending of a group of people.

Height	Weight	Age	Time
6.5	76	8	0.5
9.5	32	9	2.5
9	29	17	3
8.9	19	16	2.2
8	40	7	0
10	47	16	2.5
8	42	14	2
7.9	23	12	3

(Turn Over)

6.9	47	7	2.5
7.3	42	19	0.4
5.2	23	16	0.5
8.4	47	8	1
7.2	52	7	0.3
6.3	44	14	1.5
7.3	36	18	1.2
3.1	72	21	0.4
6.4	29	20	1

Create a .csv file of the above data and call it in R and compute the correlation matrix. Also represent the correlation matrix by a heatmap.

15. The economy rate of four category of bowlers who participated in IPL-2024 is provided below along with the number of balls they bowled (to be considered as an commitment variables). Perform ANCOVA to test if the economy rate of the four category of bowlers differ significantly.

Indian - Spinners		Indian - Fastbowler	
Balls	E.R	Balls	E.R
330	8.49	300	9.58
348	9.41	180	9.2
210	8.91	302	10.03
282	7.85	329	9.59
258	8.65	311	6.48
198	7.72	141	8.59
		288	8.83
		253	9.08
		324	9.18