[Mar-21]

[19EMA110]
M.Tech. Degree Examination

I Semester
Data Science

STATISTICAL MODELING
(For the admitted batch 2020-21)

Time: 3 Hours **Max.Marks: 60**

Instructions: All parts of the unit must be answered in one place only.
Figures in the right hand margin indicate marks allotted.

SECTION-A

1. **Answer All the Questions:** **10x2=20M**
   a. Is there correlation between rainfall and education? Explain.
   b. Interpret ANOVA.
   c. Explain alternative Hypothesis with an example.
   d. Inspect Type-I error.
   e. Analyse non-parametric tests.
   f. Define two drawbacks of nonparametric methods over parametric methods.
   g. Define time series.
   h. Explain ARIMA (0,1,0)?
   i. Analyse the various types of R-objects?
   j. Explain missing values and impossible values represented in R language?

Section-B

Answer the following **5x8=40M**

UNIT-I

2. Find the Most likely price in Chennai corresponding to price of rupees 30 at Bengaluru from the following data. 8

	Price at Bengaluru (X)	Price at Chennai (Y)
Average	73	77
Standard deviation	3.5	4.5
Coefficient of correlation	0.9	

Scanned with CamScanner

OR

3. An experiment was designed to study the performance of 4 different detergents for cleaning fuel injectors. The following "cleanness" readings were obtained with specially designed equipment for 12 tanks of gas distributed over 3 different models of engines:

	Engine 1	Engine 2	Engine 3	Totals
Detergent A	45	43	51	139
Detergent B	47	46	52	145
Detergent C	48	50	55	153
Detergent D	42	37	49	128
Totals	182	176	207	565

Looking at the detergents as treatments and the engines as blocks, obtain the appropriate analysis of the variance table and test at the 0.01 l.o.s whether there are differences in the detergents or in the engines.
$F(3, 6)_{0.01}$ d.f = 9.78 and $F(2, 6)_{0.01}$ d.f = 10.92. 8

UNIT-II

4. Write down the general procedure for testing of hypothesis. 8

OR

5. From the following data, find whether there is any significant liking in the habit of taking soft drinks among the categories of employees. [χ^2 – 4 d.f at 5 % l.o.s is 9.488]. 8

soft drinks	clerks	teachers	officers
Pepsi	10	25	65
Thumps up	15	30	65
Fanta	50	60	30

UNIT-III

6. Distinguish between parametric and non-parametric methods. 8

OR

7. The win or lose record of a certain basketball team for their last 50 consecutive games was as follows:
WWWWWW L WWWWWW L W L WWWW LL WWWW L WWW LL WWWWWW LL WW LLL W L WWW.
Apply run test to test that sequence of wins and losses is random. 8

Scanned with CamScanner

UNIT-IV

8. Explain the four phase Business cycle. 8

OR

9. What is ARIMA model? What does the p, d and q in ARIMA model mean? 8

UNIT-V

10. Suppose x is any vector as x = c(10,70,90,101:200, NA) then what is the outcome of the command Var(x)? 8

OR

11. Discuss about vectors and matrices in R. 8

Scanned with CamScanner