SASTRA DEEMED UNIVERSITY

(A University under section 3 of the UGC Act, 1956)

End Semester Examinations

JULY 2023

Course Code: **MAT244**

Course: **STATISTICAL MODELLING**

QP No. :**UF036-2**   Duration: 3 hours

Max. Marks:100

PART - A

Answer all the questions   **10 x 2 = 20 Marks**

1. Find the correlation coefficient from the following equations:
   $2y - x - 50 = 0$; $3y - 2x - 10 = 0$

2. What methods decompose a time series into its components?

3. A random sample of size 81 was taken, whose variance is 26.25 and the mean is 32. Construct a 98% confidence interval.

4. Define type-I and type-II errors.

5. A random sample of size ten from a central was taken with a standard deviation of 0.03. Find the maximum error with 99% confidence.

6. What is ANOVA? List some of its applications.

7. Write the normal equations to determine the constants a, b, and c in fitting the logarithmic trend equation.

8. List some of the advantages of non-parametric methods.

9. Give an example of a for-loop.

10. Create and display a data frame in R using the following vectors:
    patient ID <- c(1, 2, 3, 4)
    age <- c(25, 34, 28, 52)
    diabetes <- c("Type1", "Type2", "Type1", "Type1")
    status <- c("Poor", "Improved", "Excellent", "Poor")

PART - B

Answer all the questions   **4 x 15 = 60 Marks**

11. a) The prices of a commodity during 2017—2022 are given below. Fit a parabola for the following data. Estimate the price for the year 2023

X:	2017	2018	2019	2020	2021	2022
Y:	100	107	128	140	181	192

(7)

b) The two regression lines are 2Y-X-50=0; 3Y-2X-10=0. Find the mean values of x and y and the correlation coefficient.   (8)

(OR)

12. a) For the AR (2) process $X_t = 1.0X_{t-1} - 0.5X_{t-2} + a_t$, calculate $\rho1$ and $\rho2$. State whether the model is stationary, also calculate $pacf_1$ and $pacf_2$.   (7)

b) Calculate Karl Pearson's coefficient of correlation between expenditure on advertising and sales from the data below.   (8)

Ad.expenses (lakhs)	39	65	62	90	82	75	25	98	36	78
Sales (crores)	47	53	58	86	62	68	60	91	51	84

13. a) Show that the maximum likelihood estimate of the parameter α of a population having density function: $\dfrac{2}{\alpha^2}(\alpha - x), 0 < x < \alpha$ for a

sample of unit size is 2x. x being the sample value and also show that the estimate is biased. (7)

b) Using Neyman Pearson lemma, examine whether a best critical region exists for testing the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta = \theta_1 > \theta_0$ for the parameter θ of the distribution: $f(x,\theta) = \dfrac{1+\theta}{(x+\theta)^2}, 1 \le x < \infty$ (8)

(OR)

14. a) State the sufficient conditions for consistency. (5)

b) A random sample of 100 teachers in a large metropolitan area revealed a mean weekly salary of Rs. 2,000 with a standard deviation of Rs.43. With what degree of confidence can we assert that the average weekly salary of all teachers in the metropolitan area is between Rs.1,985 and Rs.2,015? (10)

15. Set up an ANOVA table for the following information relating to three drug testing to judge the effectiveness in reducing blood pressure for times different groups of people.

Group of people	Drugs		
	X	Y	Z
A	14	10	11
	15	9	11
B	12	7	10
	11	8	11
C	10	11	8
	11	11	7

Do the drugs act differently? Are the different groups of people affected differently? Is the interaction term significant? Assure the above questions take a significant level of 5 percent.

(OR)

16. a) The median age of tourists who have come to India is claimed to be 40 years. A random sample of 18 tourists gives the following

3

ages: 24, 18, 37, 51, 56, 38, 45, 45, 29, 48, 39, 26, 38, 43, 62, 30, 66, 41. Test the hypothesis using 0.05 level of significance. (7)

b) Some items produced by a machine are defective. If the machine follows some pattern where defective items are not randomly produced throughout the process, the machine needs to be adjusted. A quality control engineer wants to determine whether the sequence of defective (D) versus good (G) items is random. The data are GGGGG, DDD, GGGGGG, DDD, GGGGGGGGGG, DDDD, GGGGGGGGGGGG, DDD, GGGGGGGGGG, DDDD. Test whether the distribution of defective and good items is random or not at $\alpha = 0.05$ level of significance. (8)

17. Give the interpretation for the summary of linear regression given below:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	598.919	9552.197	0.063	0.950110	
SqFt	52.994	5.734	9.242	1.10e-15	***
Bedrooms	4246.794	1597.911	2.658	0.008939	**
Bathrooms	7883.278	2117.035	3.724	0.000300	***
Offers	-8267.488	1084.777	-7.621	6.47e-12	***
BrickYes	17297.350	1981.616	8.729	1.78e-14	***
NeighborhoodNorth	1560.579	2396.765	0.651	0.516215	
NeighborhoodWest	22241.616	2531.758	8.785	1.32e-14	***

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

Residual standard error: 10020 on 120 degrees of freedom
Multiple R-squared: 0.8686,   Adjusted R-squared: 0.861
F-statistic: 113.3 on 7 and 120 DF,   p-value: < 2.2e-16

(OR)

18. a) Using the given set of numbers {11, 12, 5, 30, 6, -12, 14,45,67,89} create a vector named 'vec' and write the code in R for the following operation: extract 4th element, extract all the odd elements: extract 3 to 8 elements; except 8th element copy all the elements to another vector; print all the results. (7)

4

b) Write the code for the five variable summaries and discuss the
    output using an example.   (8)

Answer the following   1 x 20 = 20 Marks

9. a) Ten workers were given on-the-job training to shorten their
       assembly time for a certain mechanism. The results of the time (in
       minutes) and motion studies before and after the training program
       are given below:

Worker	1	2	3	4	5	6	7	8	9	10
Before	61	62	55	62	59	74	62	57	64	62
After	59	63	52	54	59	70	67	65	59	71

Using Wilcoxon signed rank test, find evidence that the training
program has shortened the average assembly time.   (10)

b) Find the multiple regression equation of X_3 on X_1 and X_2 from the
    data relating to the three variables given below: (10)

X_1:	4	6	7	9	13	15
X_2:	15	12	8	6	4	3
X_3:	30	24	20	14	10	4

***

5