 <b>SASTRA</b> <small>AN ISO 9001:2015 CERTIFIED ORGANIZATION</small> DEEMED TO BE UNIVERSITY <small>(U.S.A of the UGC Act 1956)</small> THANAMETHU   THINK TRANSPARENCY   THIRU SASTRA THANJAVUR KUMBakonam CHENNAI	<b>SASHE</b> <b>First CIA Exam – March 2023</b> Course Code: <b>MAT244</b> Course Name: <b>STATISTICAL MODELLING</b> Duration: 90 minutes <span style="float: right;">Max Marks: 50</span>
---	--

### PART - A

**Answer all the questions**

**(5 x 2 = 10)**

1. Distinguish between linear and non-linear correlation
2. From the following information, calculate line of regression of y on x:

	x	y
Mean	40	60
Standard Deviation	10	15
Correlation co-efficient	0.7	

3. Distinguish between additive model and multiplicative model in the analysis of time series.
4. What are the components of a time series? Give an example of each component.
5. Find the autocorrelation function of the process  $X_t = X_{t-1} - 0.25X_{t-2} + a_t$  and obtain  $\rho_1$  and  $\rho_2$ .

### PART B

**Answer all the questions**

**(4 x 10 = 40)**

6. Calculate the coefficient of correlation between the age of cars and annual maintenance cost and comment.

Age of Cars in Year	: 2	4	6	7	8	10	12
Maintenance cost ('00):	16	15	18	19	17	21	20


7. Fit a second-degree parabola to the data below, and forecast the sales for 2023 and 2024.

Year	:	2018	2019	2020	2021	2022
Sales ('000)	:	16	18	19	20	24

8. Compute the multi-linear regression line of X1 on X2 and X3 from the following data and estimate the value of X1 when X2=6 and X3=10

X1:	4	6	7	9	13	15
X2:	15	12	8	6	4	3
X3:	30	24	20	14	10	4

9. (i) The three autocorrelation coefficients are  $r_1 = 0.756$ ;  $r_2 = 0.314$ ;  $r_3 = 0.141$ , calculate the pacfs. (6)
- (ii) For the model  $(1 - 0.7B)(1 - B)X_t = (1 - 0.8B)a_t$ ; find p, d, q and express it as ARIMA (p, d, q). (4)

 <b>SASTRA</b> <small>INDEPENDENT. TRANSPARENT. QUALITY EDUCATION.</small> DEEMED TO BE UNIVERSITY <small>ESTABLISHED IN 1983</small> THINK WORTH   THINK TRANSPARENCY   THINK SASTRA HANJAVUR KUMBakonam CHENNAI	<b>SASHE</b> <b>Second CIA Exam – May 2023</b> Course Code: <b>MAT244</b> Course Name: <b>STATISTICAL MODELLING</b> Duration: 90 minutes Max Marks:50
---	---

### PART - A

**Answer all the questions**

**(5 x 2 = 10)**

1. What are the uses of ANOVA?
2. Say true or false: (i) ANOVA cannot be used when there are sample of unequal sizes (ii) Analysis of variance is a statistical method of comparing the means of several populations.
3. What is the Difference Between a Parametric and Non-Parametric Test?
4. What are the Advantages of a Non-Parametric Test?
5. Write the four properties of good estimator.

### PART B

**Answer any FOUR questions**

**(4 x 10 = 40)**

6. Three varieties of coal were tested and the ash contents form the varieties are given below:

Variety	Ash contents				
	I	II	III	IV	V
A	9	7	6	5	8
B	7	4	5	4	5
C	6	5	6	7	6

Do the varieties differ significantly in their ash content?

7. A certain company had four salesmen *A*, *B*, *C*, and *D* each of whom was sent for a month to three types of areas country side *K*, outskirts of a city *O*, and shopping centre of a city *S*. The sales in hundreds of rupees per month are as follows:

Salesmen	Distributers			
	A	B	C	D
K	30	70	30	30
O	80	50	40	70
S	100	60	80	80

Perform an ANOVA test and interpret your results.

8. The survey shows that differences in heart rate of men and women whilst waiting for an interview. Use a Mann-Whitney U test to test if heart rate differs between men and women at the 95% level for the following data:

Men	80	74	73	72	78	75	70	74	69
Women	84	81	80	70	72	69	65	74	80


9. A professor believes that students in his class are not getting enough sleep to perform to the best of their ability on tests. To test this theory, he asks 10 students in his class to keep sleep diaries over a four-week period. He tells the students to ensure they sleep for 8 hours per night. At the beginning and end of the four weeks, all the students take a standard IQ test. Scores on the test at the beginning and end of the four-week period are below. Is there any evidence to support the researcher's theory that longer periods of sleep improve performance?

Test-1	92	97	76	87	80	79	99	111	103	93
Test-2	102	100	74	85	83	89	100	112	99	97

10. (i) Prove that for the binomial population with density function  $P(X = x) = nC_x p^x q^{n-x}$ ;  $x = 1, 2, \dots, n$ ,  $q = 1 - p$ ; the maximum likelihood estimator for  $p$  is  $X / n$ .

(ii) If  $X, X, \dots, X_n$  is a random sample of size  $n$  taken from a population whose pdf is  $f(x, \theta) = e^{-(x-\theta)}$ ;  $\theta \leq x < \infty$ ;  $-\infty < \theta < \infty$  then show that the sample mean is an unbiased estimator of  $(1 + \theta)$ .



 <b>SASTRA</b> DEEMED TO BE UNIVERSITY TIRU VALLUVAR UNIVERSITY TRINAJAYAM KUMBakonam CHENNAI	<b>SASHE</b> <b>Third CIA Exam – June 2023</b> Course Code: <b>MAT244</b> Course Name: <b>STATISTICAL MODELLING</b>	
	<b>Duration: 90 minutes</b>	
	<b>Max Marks: 50</b>	
	<b>PART - A</b>	

**Answer all the questions**

(5 x 2 = 10)

1. Define one-tailed and two-tailed tests.
2. State the condition for Neyman-Pearson's lemma to test the best critical region.
3. Write the syntax of if else in R
4. Write the syntax for loading and printing a .csv file in R?
5. Write the syntax of correlation and plotting of any two vectors.

**PART B**

**Answer all the questions**

(2 x 15 = 30)

6. (i) Based on the following data, calculate multiple correlation coefficients of  $R1.23$ ,  $R3.12$ , and  $R2.13$ .

$\bar{x}_1 = 6.8$	$\bar{x}_2 = 7.0$	$\bar{x}_3 = 74$
$s_1 = 1.0$	$s_2 = 0.8$	$s_3 = 9$
$r_{12} = 0.6$	$r_{13} = 0.7$	$r_{23} = 0.65$

- (ii) The mean diameter of a random sample of 200 ball bearings made by a certain machine during one week is 0.824 inches and standard deviation is 0.042 inches. Find (a) 95% (b) 99% confidence interval for the mean diameter of all the ball bearings.

**OR**

7. (i) Perform Kendall Concordance test at 5% level to test the qualities of mangoes.

	Ranking of qualities of Mangoes							
A	8	4	2	3	5	1	6	7
B	7	3	1	4	5	2	6	8
C	8	2	5	6	7	1	4	3
D	8	3	4	2	5	1	6	7

- (ii) Using Neyman Pearson lemma, obtain the best critical region for testing Null hypothesis  $H_0 : \theta = \theta_0$  Vs Alternative hypothesis  $H_1 : \theta = \theta_1 (\leq \theta_0)$  given that  $f(x, \theta) = \theta e^{-\theta x}$

8. The following table gives the number of units of production per day turned out by four different types of machines :

Employee	Types of Machines			
	$M_1$	$M_2$	$M_3$	$M_4$
$E_1$	40	36	45	30
$E_2$	38	42	50	41
$E_3$	36	30	48	35
$E_4$	46	47	52	44

Using analysis of variance (a) test the hypothesis that the mean production is the same for the four machines and (b) test the hypothesis that the employees do not differ with respect to mean productivity.

OR

9. Analyze the summary of linear regression given below and give suitable interpretations:

Call:

```
lm(formula = ROLL ~ UNEM + HGRAD + INC, data = datavar)
```

Residuals:

Min	1Q	Median	3Q	Max
-1148.840	-489.712	-1.876	387.400	1425.753

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-9.153e+03	1.053e+03	-8.691	5.02e-09 ***
UNEM	4.501e+02	1.182e+02	3.809	0.000807 ***
HGRAD	4.065e-01	7.602e-02	5.347	1.52e-05 ***
INC	4.275e+00	4.947e-01	8.642	5.59e-09 ***

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

Residual standard error: 670.4 on 25 degrees of freedom  
Multiple R-squared: 0.9621, Adjusted R-squared: 0.9576  
F-statistic: 211.5 on 3 and 25 DF, p-value: < 2.2e-16

## PART - C

Answer all the questions

(1 x 10 = 10)

10. (i) Create a data frame with  $a = c(1, 2, 3)$ ,  $b = c(4, 5, 6)$ ,  $c(7, 8, 9)$  and find the value of the following (a) How to select the  $c(4, 5, 6)$ ? (b) How to select the 1 (c) How to select the 5? (d) What is  $df[ , 3]$ ? (e) What is  $df[1, ]$ ? (f) What is  $df[2, 2]$ ?

(ii) Write the R-code to create a simple data frame of five COVID-19 patient details. Your data frame should contain the columns "ID"; "AGE"; "MOBILE NO"; "GENDER" and "TREATMENT STATUS".