Name:

Enrolment No:

UPES

End Semester Examination, December 2023

Course: Statistical Modelling and Simulation **Semester: III**

Program: B. Tech Sustainability Engineering **Time: 03 hrs.**

Course Code: SUEN 2004 **Max. Marks: 100**

Instructions: Attempt all the questions;

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Draw a frequency polygon following data: Marks: <20, <40, <60, <80, <100 No of Students: 10, 40, 80, 100, 110	**4**	**CO1**
Q 2	The following are the monthly evaporation data (Jan.-Dec.) at a Dam in a certain year in cm: **16.7, 14.3, 17.8, 25.0, 28.6, 21.4, 16.7, 16.7, 16.7, 21.4, 16.7, 16.7** Calculate the **kurtosis** for the data and interpret its physical significance.	**4**	**CO1**
Q 3	The foreman of ABC mining company has estimated the average quantity of iron ore extracted to be 36.8 tons per shift and the sample standard deviation to be 2.8 tons per shift, based upon a random selection of 4 shifts. Construct a 90 per cent confidence interval around this estimate.	**4**	**CO2**
Q 4	Define the following terms: a) Seasonality b) Stochasticity c) Periodicity d) Cyclicity	**4**	**CO1**
Q 5	Differentiate between the following: a) Precision and accuracy b) Causation and correlation	**4**	**CO1**

SECTION B
(4Qx10M= 40 Marks)

Q 6	A sample of 10 is drawn randomly from a certain population. The sum of the squared deviations from the mean of the given sample is 50. Test the hypothesis that the variance of the population is 5 at 5 per cent level of significance.	**10**	**CO2**

<table>
<tr><td colspan="4">OR</td></tr>
<tr><td>Q 6</td><td>There are 100 students in a university college and in the whole university, inclusive of this college, the number of students is 2000. In a random sample study 20 were found smokers in the college and the proportion of smokers in the university is 0.05. Is there a significant difference between the proportion of smokers in the college and university? Test at 5 per cent level.</td><td>10</td><td>CO2</td></tr>
<tr><td>Q 7</td><td>Write a python program to calculate for multiple-linear regression to predict y using x_1 and x_2 and validate the model. Assume any arbitrary data for the code.</td><td>10</td><td>CO1</td></tr>
<tr><td>Q 8</td><td>Demonstrate the applicability of one-way ANOVA classification with the help of an example.</td><td>10</td><td>CO3</td></tr>
<tr><td>Q 9</td><td>What are the different methods of estimating a missing data value? Explain with the help of an example.</td><td>10</td><td>CO1</td></tr>
<tr><td colspan="4">SECTION-C
(2Qx20M=40 Marks)</td></tr>
<tr><td>Q 10</td><td>The stage-discharge data of a river are given below. Establish the stage-discharge relationship to predict the discharge for a given stage. Assume the value of stage for zero discharge as 35.00 m. (2) What is the correlation coefficient of the relationship established above? (3) Estimate the discharge corresponding to stage values of 42.50 m and 48.50 m respectively.

<table>
<tr><th>Stage (m)</th><th>Discharge (m^3/s)</th><th>Stage (m)</th><th>Discharge (m^3/s)</th></tr>
<tr><td>35.91</td><td>89</td><td>39.07</td><td>469</td></tr>
<tr><td>36.90</td><td>230</td><td>41.00</td><td>798</td></tr>
<tr><td>37.92</td><td>360</td><td>43.53</td><td>2800</td></tr>
<tr><td>44.40</td><td>3800</td><td>48.02</td><td>5900</td></tr>
<tr><td>45.40</td><td>4560</td><td>49.05</td><td>6800</td></tr>
<tr><td>46.43</td><td>5305</td><td>49.55</td><td>6900</td></tr>
<tr><td></td><td></td><td>49.68</td><td>6950</td></tr>
</table>

The relationship between the discharge (Q) and gauge reading (G) is non-linear and given as:

$$Q = C_r (G - a)^b$$

a = constant which represents the gauge reading corresponding to zero discharge, C_r and b are rating curve constants.</td><td>20</td><td>CO4</td></tr>
<tr><td colspan="4">OR</td></tr>
<tr><td>Q 10</td><td>a) In a random selection of 64 of the 2400 intersections in a small city, the mean number of scooter accidents per year was 3.2 and the sample standard deviation was 0.8.
i. Make an estimate of the standard deviation of the population from the sample standard deviation.</td><td>15+5</td><td>CO4</td></tr>
</table>

	ii. Work out the standard error of mean for this finite population. iii. If the desired confidence level is .90, what will be the upper and lower limits of the confidence interval for the mean number of accidents per intersection per year? b) Explain the properties of Poisson distribution.		
Q 11	For the data given below, plot the series and construct ARMA model to predict the future values.	**20**	**CO4**

Month	Passengers	Month	Passengers
Jan-49	112	Jan-50	115
Feb-49	118	Feb-50	126
Mar-49	132	Mar-50	141
Apr-49	129	Apr-50	135
May-49	121	May-50	125
Jun-49	135	Jun-50	149
Jul-49	148	Jul-50	170
Aug-49	148	Aug-50	170
Sep-49	136	Sep-50	158
Oct-49	119	Oct-50	133
Nov-49	104	Nov-50	114
Dec-49	118	Dec-50	140