CONFIDENTIAL

Name of Examination (Please tick, symbol is given)	**:**	**SUPP LE**		**END**	✓	**Set-1**	✓
Name of the School (Please tick, symbol is given)	**:**	**SOAE**		**SOCS**	✓	**Set-2**	
Programme	**:**	**B.Tech. CSE**					
Semester	**:**	**III**					
Name of the Course	**:**	**Probability and Statistics**					
Course Code	**:**	**CSEG 2036P**					
Name of Question Paper Setter	**:**	**Dr. Mrittunjoy Guha Majumdar**					
Employee Code	**:**	**40003973**					
Mobile	**:**	**8595301465**					
Note: Please mention additional Stationery to be provided, during examination such as Table/Graph Sheet etc. else mention "NOT APPLICABLE":							

FOR SRE DEPARTMENT		
Date of Examination	**:**	
Time of Examination	**:**	
No. of Copies (for Print)	**:**	

Note: - Pl. start your question paper from next page

Model Question Paper (Blank) is on next page

Name: **Enrolment No:**	UPES UNIVERSITY OF TOMORROW

UPES

End Semester Examination, December 2023

Course: Probability and Statistics
Semester: III
Program: B. Tech. CSE
Course Code: CSEG 2036P

Time: 3 hrs.
Max. Marks: 100

SECTION A

(5Qx4M=20Marks)

<table>
<tr><td>S. No.</td><td></td><td>Marks</td><td>CO</td></tr>
<tr><td>Q 1</td><td>Define sample spaces. Identify the set expression as well as Venn diagram representation for the following cases:

1. Among A, B and C, only A occurs.
2. A or C occur but not B

for a sample space S and three events A, B and C.</td><td>4</td><td>CO1</td></tr>
<tr><td>Q 2</td><td>Outline what is meant by random variables. Identify the values of a and b if we have a random variable X with the associated probability density function,

$$f(x) = ax^b, 0 \leq x \leq 1$$

and the mean $E[X]$ is 0.75.</td><td>4</td><td>CO1, CO2</td></tr>
<tr><td>Q 3</td><td>Discuss any two properties of Probability Mass Functions. Apply your understanding of joint probability mass functions to evaluate the value of $(X = 0, Y \leq 1)$ and $P(Y = 1|X = 0)$, if we have random variables X and Y with joint probability mass function given in the table below

<table>
<tr><td></td><td>$Y = 0$</td><td>$Y = 1$</td><td>$Y = 2$</td></tr>
<tr><td>$X = 0$</td><td>$\frac{1}{8}$</td><td>$\frac{1}{6}$</td><td>$\frac{1}{8}$</td></tr>
</table>
</td><td>4</td><td>CO1</td></tr>
</table>

	$X = 1$	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{6}$		
Q 4	**Discuss** covariance of random variables. **Illustrate** that $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$ for a random variable X and that if *continuous random variables* X and Y are independent then $\mathrm{Cov}(X, Y) = 0$.				**4**	**CO2**
Q 5	**Discuss** correlation coefficient. **Illustrate** that if $Y = aX + b$ for random variables X and Y with constants a and b, the absolute value of the correlation coefficient is unity: $$\lvert r_{XY} \rvert = 1$$				**4**	**CO2**

SECTION B

(4Qx10M= 40 Marks)

Q 6	**Define** and **illustrate** the concepts of hypothesis, null hypothesis, alternative hypothesis, significance level and Type I/II errors. **Remember** and **explain** the generalized scheme for Hypothesis Testing. **Apply** and **discuss** Hypothesis Testing for Population Mean for large samples. **Illustrate** if the difference between samples A and B is significant (for $\alpha = 0.05$) if $\bar{X}_A = 75$, $\bar{X}_B = 76.5$, $\sigma_A = 8$, $\sigma_B = 10$, $n_A = 150$ and $n_B = 200$, given that for	**10**	**CO4**

z	-2	-1.75	-1.5	-1.25	-1	-0.5
p	0.04	0.08	0.13	0.21	0.32	0.62

Q 7	**Define** ANOVA and the motivation behind formulating such a statistical characterization. **Discuss** the different kinds of sums of squares, degrees of freedom and mean squares. **Identify** the relevant test statistic (ANOVA coefficient) in terms of mean squares, for ANOVA. **Apply** ANOVA ($\alpha = 0.05$) to the times required by three workers to perform an assembly-line task recorded on five randomly selected occasions.	**10**	**CO4**

Akash	Vasudha	Sunil
8	8	10

	<table><tr><td>10</td><td>9</td><td>9</td></tr><tr><td>9</td><td>9</td><td>10</td></tr><tr><td>11</td><td>8</td><td>11</td></tr><tr><td>10</td><td>10</td><td>9</td></tr></table> Given: F-statistic for degrees of freedom (2,12) at $\alpha = 0.05$ is 3.89.		
Q 8	*Choice 1:* **Define** the Mann Whitney U Test and its test statistic (for sample size $n \leq 20$). **Describe** any three assumptions relevant to this statistical test. **Apply** your understanding of the Mann Whitney U Test for analyzing the scores of two groups of students (Group A and Group B) with <table><tr><td>Group A</td><td>63</td><td>70</td><td>77</td><td>81</td><td>93</td></tr><tr><td>Group B</td><td>66</td><td>76</td><td>81</td><td>78</td><td>85</td></tr></table> Given: For the two-tailed Mann-Whitney U Test with $n_1 = 5$ and $n_2 = 5$ for level of significance $\alpha = 0.05$, the critical value is 2. *Choice 2:* **Define** regression, principle of least squares and residuals**.** **Describe** what is meant by multiple regression model. **Apply** and **discuss** your understanding of non-linear regression to fit a curve of the form $y = ae^{bx}$ for the following data: <table><tr><td>x</td><td>1</td><td>5</td><td>7</td><td>9</td><td>12</td></tr><tr><td>y</td><td>10</td><td>15</td><td>12</td><td>15</td><td>21</td></tr></table> Given: $\log_{10}(9.4795) = 0.976788$ and $\frac{0.0255}{\log_{10} e} = 0.05872$.	**10**	**CO4**
Q 9	**Define** a Decision Tree. **Describe** what is node purity and **highlight** one advantage and one disadvantage of using Decision Trees. **Expand** on the two ways in which Decision Trees can have variable selection criterion for node allocation. **Apply** your understanding of the *Information Gain and Entropy approach* for Decision Trees to analyze 15 students' performance in an online exam. The predictors for this data-set encompass details such as whether the student is enrolled in other online courses, their academic background and whether they are currently employed or not.	**10**	**CO5**

S.No.	Target Variable	Predictor Variables		
	Result	Other Online Courses	Student Background	Working Status
1.	Pass	Yes	Mathematics	Not Working
2.	Fail	No	Mathematics	Working
3.	Fail	Yes	Mathematics	Working
4.	Pass	Yes	CS	Not Working
5.	Fail	No	Other	Working
6.	Fail	Yes	Other	Working
7.	Pass	Yes	Mathematics	Not Working
8.	Pass	Yes	CS	Not Working
9.	Pass	No	Mathematics	Working
10.	Pass	No	CS	Working
11.	Pass	Yes	CS	Working
12.	Pass	No	Mathematics	Not Working
13.	Fail	Yes	Other	Working
14.	Fail	No	Other	Not Working
15.	Fail	No	Mathematics	Working

SECTION-C
(2Qx20M=40 Marks)

Q 10	**Define** a Bernoulli random variable $X \sim \text{Ber}(p)$ and its probability distribution, along with one condition for Bernoulli Distributions. **Derive** the mean and standard deviation of the random variable $X \sim \text{Ber}(p)$. **Define** a Binomial random variable $X \sim \text{Bin}(n, p)$ and its probability distribution. **Highlight** the relation of Binomial distributions with Bernoulli trials. **Derive** the mean and standard deviation of $X \sim \text{Bin}(n, p)$. Given: $k.\binom{n}{k} = n.\binom{n-1}{k-1}$ and $k^2.\binom{n}{k} = k.n.\binom{n-1}{k-1}$. **Define** moment-generating function $M_X(t)$ for a random variable. **Apply** your understanding of Moment-generating function for the case of a Bernoulli random variable. Given $M_{X+Y}(t) = M_X(t)M_Y(t)$ and considering the relation between Bernoulli random variables and Binomial random variables, **apply** your understanding of the moment-generating function for the case of a Binomial random variable.	**20**	**CO3**

<table>
<tr><td></td><td>Considering $p = \frac{\lambda}{n}$ and given the asymptotic behavior $\left(1 + \frac{x}{n}\right)^n \to e^x$, expand on the expression for the Moment-generating function for the Binomial random variable.

Define a Poisson process and its probability distribution. Apply your understanding of the moment-generating function in the case of a Poisson process, considering $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$.

Discuss the logical relation of the Moment-generating functions of the Binomial and Poisson Distributions.</td><td></td><td></td></tr>
<tr><td>Q 11</td><td>Define a normal distribution and a standard normal table. Derive the points of inflection of a normal distribution.

Calculate the probability that a randomly selected student from UPES is shorter than 150 cm, given that the heights of a population of students at UPES follow a normal distribution with a mean height (μ) of 160 cm and a standard deviation (σ) of 10 cm.

<u>Given</u>: The following segment of the standard normal table

<table><tr><td>z</td><td>-3</td><td>-2</td><td>-1</td><td>0</td><td>1</td><td>2</td><td>3</td></tr><tr><td>Value</td><td>0.001</td><td>0.023</td><td>0.159</td><td>0.5</td><td>0.841</td><td>0.977</td><td>0.999</td></tr></table>
Determine the height (in cm) that separates the top 20% of the tallest students from the rest, given that the z-score corresponding to the 80th percentile is approximately 0.84.

Discuss sample statistics and describe the Method of Moments (MoM). Highlight any two properties of a good estimator in sample statistics.

Choice 1: Identify the MoM estimator of the population parameters for n independent and identically distributed samples taken from a normal distribution.

Choice 2: Calculate the probability that the sample mean height of these students (for a sample of 25 students taken from the distribution mentioned above) is greater than 165 cm.

<u>Given</u>: $p_{\alpha=0.05}(z = 2.5) = 0.994$.</td><td>20</td><td>CO3</td></tr>
</table>