U.S.N. | | | | | | | | | | |

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

June 2025 Semester End Main Examinations

Programme: B.E. **Semester: IV**
Branch: Chemical Engineering **Duration: 3 hrs.**
Course Code: 23MA4BSSAP / 22MA4BSSAP **Max Marks: 100**
Course: STATISTICS AND PROBABILITY

Instructions: 1. Each unit has an internal choice; answer one complete question from each unit.
2. Missing data, if any, may be suitably assumed.
3. Use of Statistical tables are permitted.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

<table>
<tr><th></th><th></th><th>UNIT - 1</th><th>CO</th><th>PO</th><th>Marks</th></tr>
<tr><td>1</td><td>a)</td><td>Fit a curve of the form $y = ae^{bx}$ to the following data:
<table><tr><td>x</td><td>10</td><td>25</td><td>40</td><td>60</td></tr><tr><td>y</td><td>50</td><td>110</td><td>250</td><td>600</td></tr></table></td><td>1</td><td>1</td><td>6</td></tr>
<tr><td></td><td>b)</td><td>Suppose a number of fecal bacteria/ml of milk follows a Poisson distribution with mean 3 bacteria. If a random sample of one ml. of milk is examined, find the probability that it contains (i) no bacteria (ii) at least one bacterium (iii) at most two bacteria.</td><td>1</td><td>2</td><td>7</td></tr>
<tr><td></td><td>c)</td><td>A researcher is investigating the relationship between the number of hours a student spends studying for an exam and their score on that exam. Data for six students is provided below:
<table><tr><td>Hours Studied</td><td>3</td><td>5</td><td>2</td><td>4</td><td>6</td><td>3</td></tr><tr><td>Exam Score</td><td>65</td><td>80</td><td>60</td><td>75</td><td>85</td><td>70</td></tr></table>
Calculate the correlation coefficient for this data.</td><td>1</td><td>2</td><td>7</td></tr>
<tr><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td>2</td><td>a)</td><td>The braking distance D (in meters) of a car is hypothesized to depend on its initial speed S (in km/h) according to a quadratic law: $D = a + bS + cS^2$. Determine the best values for the coefficients a, b, and c based on the following experimental data:
<table><tr><td>Speed(S)</td><td>20</td><td>40</td><td>60</td><td>80</td><td>100</td></tr><tr><td>Braking Distance(D)</td><td>12</td><td>35</td><td>68</td><td>110</td><td>165</td></tr></table></td><td>1</td><td>1</td><td>6</td></tr>
<tr><td></td><td>b)</td><td>The daily electricity consumption in a small-town during winter is normally distributed with a mean of 85 MWh (megawatt-hours) and a standard deviation of 12 MWh. The town's power plant has a maximum daily output capacity of 100 MWh.
(i) What is the probability that on a given winter day, the electricity demand will exceed the power plant's maximum output capacity?
(ii) What power plant capacity would be needed so that the probability of it being exceeded is only 2.5%?</td><td>1</td><td>2</td><td>7</td></tr>
</table>

	c)	In a partially destroyed laboratory record, only the lines of regression of y on x and x on y are available as $4x-5y+33=0$ and $20x-9y=107$ respectively. Calculate $\bar{x}$, $\bar{y}$ and the co-efficient of correlation between x and y.	*1*	*2*	**7**
		UNIT - 2			
3	a)	Let *X* and *Y* be the number of defective items produced on two different assembly lines, Line A and Line B, in an hour. The joint probability distribution of *X* and *Y* is given by $P(X=x,Y=y)=c(x+y)$, where $x\in\{0,\ 1,\ 2\}$ and $y\in\{0,\ 1\}$. i) Construct the joint probability distribution table. ii) Compute $E[X],\ E[Y],\ E[XY]$.	*1*	*2*	**6**
	b)	A small town has three main political parties: Party A, Party B, and Party C. A political scientist models the shifts in voter allegiance between these parties over a year as a Markov chain. The transition probability matrix (where rows are current party and columns are next year's party) is given by $\begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.2 & 0.7 & 0.1 \\ 0.3 & 0.2 & 0.5 \end{pmatrix}$. The initial distribution of voters across the political parties is $P^{(0)}=(0.5\quad 0.3\quad 0.2)$. Find: (i) the probability that a randomly chosen voter belongs to Party C after two years and (ii) Steady state distribution.	*1*	*2*	**7**
	c)	A student studies either in the library or in their dorm room. They never study in the library on two consecutive days. However, if they study in their dorm room on a given day, there's a 70% chance they will study in the dorm room again the next day, and a 30% chance they will go to the library. (i) Find the transition matrix for the Markov chain of the student's study location, (ii) If the student studies in their dorm room on Monday, find the probability that they study in the library on Wednesday.	*1*	*2*	**7**
		OR			
4	a)	A small coffee shop records the number of hot beverages *(X)* and cold beverages *(Y)* sold per hour during a specific morning period. The joint probability distribution of *X* and *Y* is given by $P(X=x,Y=y)=c(2x+y)$, where $x\in\{1,\ 2\}$ and $y\in\{0,\ 1\}$. (i) Construct the joint probability distribution table. (ii) Check whether hot beverages *(X)* and cold beverages *(Y)* are stochastically dependent.	*1*	*2*	**6**
	b)	A city's public transport system operates with buses and trams. If a commuter takes a bus on a particular day, they are 80% likely to take a bus again the next day and 20% likely to take a tram. However, if a commuter takes a tram on a given day, they will never take a tram again the next day; they will always take a bus. (i) Find the transition matrix for the Markov chain of the commuter's chosen mode of transport. (ii) Find the steady-state vector for this Markov chain.	*1*	*2*	**7**

		c)	A small ecosystem can be in one of three states based on its water level: Low (State 1), Medium (State 2), or High (State 3). The transition probability matrix describing the change in water level from one month to the next is given by $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$. Currently, the probabilities of the ecosystem being in a Low, Medium, or High-water level state are given by the initial distribution $P^{(0)} = \begin{pmatrix} 0.6 & 0.3 & 0.1 \end{pmatrix}$. i) Check whether the Markov chain is irreducible and ii) Find the probability that the ecosystem is in a Low water level state after two months.	*1*	*2*	**7**
			UNIT - 3			
	5	a)	A manufacturer claims that its new energy-efficient light bulbs have a longer average lifespan than their previous model. The previous model was known to have an average lifespan of 800 hours with a standard deviation of 60 hours. To test if the new light bulbs have a significantly longer lifespan, a random sample of 40 new light bulbs was tested, and their lifespans were recorded. The experiment yielded a sample mean lifespan of 825 hours and a sample standard deviation of 55 hours. Is there sufficient evidence in the sample to indicate, at the 5% level of significance, that the new light bulbs do have a longer average lifespan?	*2*	*2*	**6**
		b)	In an infantile paralysis epidemic 500 persons contracted the disease. 300 received no serum treatment and of them 75 became paralysed. Of those who received serum treatment 65 became paralysed. Was the serum treatment effective at 1% level of significance?	*2*	*2*	**7**
		c)	The standard deviation of the height of Honor students of a college is 4*ft*.Two samples are taken. The standard deviation of the height of 100 B.Com Honor students is 3.5*ft* and 50 B.A. Honor students is 4.5*ft*. Test the significance of the difference of standard deviations of the samples at 5% level of significance?	*2*	*2*	**7**
			OR			
	6	a)	Historically, the proportion of customers who prefer product A from a certain electronics company is 30%. The company has recently launched a new advertising campaign, and they want to see if this has changed customer preferences. To test this, a random sample of 800 customers was surveyed. It was found that 260 of these customers now prefer product A. Determine whether there is sufficient evidence, at the 5% level of significance, to support the belief that the proportion of customers preferring product A has changed.	*2*	*2*	**6**
		b)	A comparative study of variation in weights (in pound) of Army soldiers and Navy- sailors was made. The sample variance of the weight of 120 soldiers was 60 pound^2 and the sample variance of the weight of 160 sailors was 70 pound^2. Test whether the soldiers and sailors have equal variation in their weights. Use 5% level of significance.	*2*	*2*	**7**

<table>
<tr><td></td><td></td><td>c)</td><td>A nutritionist is interested in whether two proposed diets, <i>A</i> and <i>B</i> work equally well in providing weight-loss for customers. In order to assess a difference between the two diets, she puts 50 customers on <i>Diet A</i> and 60 other customers on the <i>Diet B</i> diet for two weeks. Those on the former had weight losses with an average of 11 pounds and a standard deviation of 3 pounds, while those on the latter lost an average of 8 pounds with a standard deviation of 2 pounds. Do the average diets differ in terms of their weight loss at 1% level of significance?</td><td><i>2</i></td><td><i>2</i></td><td>7</td></tr>
<tr><td></td><td></td><td></td><td>UNIT - 4</td><td></td><td></td><td></td></tr>
<tr><td></td><td>7</td><td>a)</td><td>A group of 5 patients treated with medicine “A” weigh 42, 39, 48, 60, and 41 kgs. A second group of 7 patients from the same hospital treated with medicine “B” weigh 38, 42, 56, 64, 68, 69, and 62 kgs. Do you agree with the claim that the medicine “B” increases the average weight significantly at 1% level of significance?</td><td><i>2</i></td><td><i>2</i></td><td>6</td></tr>
<tr><td></td><td></td><td>b)</td><td>A major fast-food chain conducted a survey of 200 randomly selected customers, recording their preferred main dish category. The company's historical assumed probability distribution of customer preferences and the observed frequencies from the survey are as follows:
<table>
<tr><th>Main Dish Category</th><th>Assumed Distribution</th><th>Observed Frequency</th></tr>
<tr><td>Burgers</td><td>0.50</td><td>110</td></tr>
<tr><td>Chicken</td><td>0.25</td><td>45</td></tr>
<tr><td>Salads</td><td>0.15</td><td>25</td></tr>
<tr><td>Sides</td><td>0.10</td><td>20</td></tr>
</table>
Applying Chi-Square distribution, check whether the customer’s preferences are as per the assumed probability distribution.</td><td><i>2</i></td><td><i>2</i></td><td>7</td></tr>
<tr><td></td><td></td><td>c)</td><td>Measurements on the length of a copper wire were taken in 2 experiments A and B as under:
<table>
<tr><td>A</td><td>12.29</td><td>12.25</td><td>11.86</td><td>12.13</td><td>12.44</td><td>12.78</td></tr>
<tr><td>B</td><td>12.39</td><td>12.46</td><td>12.34</td><td>12.22</td><td>11.98</td><td>12.46</td></tr>
</table>
Test whether B’s measurements are more accurate than A’s at 5% level of significance.</td><td><i>2</i></td><td><i>2</i></td><td>7</td></tr>
<tr><td></td><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td></td><td>8</td><td>a)</td><td>In a study of usefulness of yoga in weight reduction, a random sample of 8 persons undergoing yoga were examined for their weight before and after yoga with the following results
<table>
<tr><td>before</td><td>209</td><td>178</td><td>169</td><td>212</td><td>180</td><td>192</td><td>158</td><td>180</td></tr>
<tr><td>after</td><td>196</td><td>171</td><td>170</td><td>207</td><td>177</td><td>190</td><td>159</td><td>180</td></tr>
</table>
Test whether yoga is useful in weight reduction at $\alpha = 1\%$</td><td><i>2</i></td><td><i>2</i></td><td>6</td></tr>
</table>

<table>
<tr><td></td><td></td><td>b)</td><td>Two sample polls of votes for 2 candidates A and B are taken from residents of different areas. The results are given below. Examine whether the nature of the area is related to voting preferences in this election using Chi Square distribution at 5% level of significance.
<table>
<tr><td>Votes
Area</td><td>A</td><td>B</td></tr>
<tr><td>Rural</td><td>620</td><td>380</td></tr>
<tr><td>Urban</td><td>550</td><td>450</td></tr>
</table></td><td>2</td><td>2</td><td>7</td></tr>
<tr><td></td><td></td><td>c)</td><td>Three nutrients were examined for the height of seedlings and the following data is recorded:
<table>
<tr><td></td><td colspan="6">Height of seedlings</td></tr>
<tr><td>A</td><td>22</td><td>20</td><td>21</td><td>18</td><td>16</td><td>14</td></tr>
<tr><td>B</td><td>12</td><td>14</td><td>15</td><td>10</td><td>9</td><td>-</td></tr>
<tr><td>C</td><td>7</td><td>9</td><td>7</td><td>6</td><td>-</td><td>-</td></tr>
</table>
Perform one way ANOVA at 1% level of significance.</td><td>2</td><td>2</td><td>7</td></tr>
<tr><td></td><td></td><td></td><td>UNIT - 5</td><td></td><td></td><td></td></tr>
<tr><td></td><td>9</td><td>a)</td><td>An agricultural research institute is testing the effectiveness of three different fertilizers (Fertilizer X, Fertilizer Y, and Fertilizer Z) on the yield of a specific crop. They selected 21 identical plots of land. Each fertilizer was randomly applied to 6 different plots. After the growing season, the crop yield (in kilograms per plot) was recorded. Test whether there is any significant difference in the mean crop yields among the three fertilizers at 1% level of significance.
<table>
<tr><td>Fertilizer X</td><td>15.2</td><td>14.8</td><td>16.1</td><td>15.5</td><td>14.5</td></tr>
<tr><td>Fertilizer Y</td><td>16.8</td><td>17.5</td><td>16.0</td><td>17.2</td><td>16.5</td></tr>
<tr><td>Fertilizer Z</td><td>18.5</td><td>19.2</td><td>18.0</td><td>19.5</td><td>17.8</td></tr>
</table></td><td>2</td><td>2</td><td>10</td></tr>
<tr><td></td><td></td><td>b)</td><td>A pharmaceutical company is testing the effectiveness of four different formulations of a new pain medication (Formulation A, B, C, D). To control for variability, they designed an experiment using 4 different patient age groups (G1, G2, G3, G4) and administered the medication at 4 different times of day (T1, T2, T3, T4). Each patient in a specific age group received each formulation once, and each formulation was administered at each time of day once. The arrangement of age groups, times of day, and formulations, along with the reported pain relief score (on a scale of 0-100) after 2 hours, is shown in the following table.
<table>
<tr><td></td><td></td><td colspan="4">Age Groups</td></tr>
<tr><td></td><td></td><td>G1</td><td>G2</td><td>G3</td><td>G4</td></tr>
<tr><td rowspan="4">Time of the day</td><td>T1</td><td>B(75)</td><td>C(88)</td><td>A(65)</td><td>D(70)</td></tr>
<tr><td>T2</td><td>D(68)</td><td>B(78)</td><td>C(72)</td><td>A(80)</td></tr>
<tr><td>T3</td><td>A(92)</td><td>D(70)</td><td>B(85)</td><td>C(68)</td></tr>
<tr><td>T4</td><td>A(80)</td><td>C(70)</td><td>D(65)</td><td>B(72)</td></tr>
</table>
Test whether there is any significant difference in the mean pain relief scores due to the three factors: age groups, times of day, and medication formulations at 5% level of significance.</td><td>2</td><td>2</td><td>10</td></tr>
<tr><td></td><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
</table>

<table>
<tr><td></td><td>10</td><td>a)</td><td>A consumer research company conducted an experiment to compare the cleaning effectiveness of four different laundry detergents (Brands A, B, C, D). Five identical loads of laundry were prepared, each with similar stains and dirt levels. Each load was randomly assigned one of the four detergents, and the cleanliness score (on a scale of 1 to 100, with 100 being perfectly clean) was recorded.
<table>
<tr><td></td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>A</td><td>75</td><td>78</td><td>76</td><td>79</td><td>77</td></tr>
<tr><td>B</td><td>80</td><td>83</td><td>85</td><td>82</td><td>84</td></tr>
<tr><td>C</td><td>88</td><td>90</td><td>89</td><td>91</td><td>87</td></tr>
<tr><td>D</td><td>90</td><td>92</td><td>91</td><td>93</td><td>89</td></tr>
</table>
Test whether there is any significant difference in the mean cleanliness score according to the detergent brand and the laundry load at 1% level of significance.</td><td><i>2</i></td><td><i>2</i></td><td>10</td></tr>
<tr><td></td><td></td><td>b)</td><td>A pharmaceutical company conducted a study to compare the effectiveness of four new pain relief creams (Cream P, Cream Q, Cream R, Cream S) on muscle soreness. A 4x4 Latin square design was used, with four different patient groups (Rows) and four different application times (Columns). The pain relief score (on a scale of 0-10, where 10 is complete relief) was recorded for each patient after treatment.
<table>
<tr><td>R(7.5)</td><td>P(6.0)</td><td>Q(8.2)</td><td>S(7.8)</td></tr>
<tr><td>P(6.8)</td><td>Q(7.0)</td><td>S(7.3)</td><td>R(8.0)</td></tr>
<tr><td>Q(8.5)</td><td>S(6.5)</td><td>R(7.9)</td><td>P(6.2)</td></tr>
<tr><td>S(7.0)</td><td>R(6.9)</td><td>P(5.8)</td><td>Q(8.8)</td></tr>
</table>
Test whether there is any significant difference in the mean pain relief scores due to the patient groups, application times, and cream types (treatment) at 5% level of significance.</td><td><i>2</i></td><td><i>2</i></td><td>10</td></tr>
</table>

******