

STELLA MARIS COLLEGE (AUTONOMOUS) CHENNAI – 600 086
(For candidates admitted from the academic year 2024 – 2025 & thereafter)

M.Sc. DEGREE EXAMINATION – APRIL 2026
DATA SCIENCE
SECOND SEMESTER

COURSE : CORE
PAPER : STATISTICS - II
SUBJECT CODE : 24DS/PC/ST24
TIME : 3 HOURS

MAX. MARKS: 100

Q. No.	SECTION A Answer all the questions (10 x 2=20)	CO	KL
1.	Define stratified sampling.	CO1	K1
2.	State any two properties of F-distribution.	CO1	K1
3.	Define regression.	CO1	K1
4.	What is ANOVA? When is it preferred over multiple t-tests?	CO1	K1
5.	What are nonparametric tests?	CO1	K1
6.	What is meant by a test of a null hypothesis?	CO2	K2
7.	Define chi-square distribution.	CO2	K2
8.	What is a standard error of estimate?	CO2	K2
9.	What is the null hypothesis in a one-way ANOVA test?	CO2	K2
10.	What is the Mann-Whitney U test used for?	CO2	K2
Q. No.	SECTION B Answer all the questions (4 x 5=20)	CO	KL
11.	a) A machine that automatically controls the amount of ribbon on a tape has recently been installed. This machine will be judged to be effective if the standard deviation σ of the amount of ribbon on a tape is less than 0.15 cm. If a sample of 20 tapes yields a sample variance of $S^2 = 0.025 \text{ cm}^2$, are you justified in concluding that the machine is ineffective? (OR) b) Explain significance levels.	CO3	K3
12.	a) State the properties of t-distribution. (OR) b) In one sample of 8 observations, the sum of the squares of deviations of the sample values from the sample mean was 84.4 and in the other sample of 10 observations it was 102.6. Test whether this difference is significant at 5 percent level.	CO3	K3
13.	a) List out the uses of regression analysis. (OR) b) From the following data calculate the correlation coefficient and the standard deviation of Y: $b_{xy} = 0.85; \quad b_{yx} = 0.89; \quad \sigma_x = 3$	CO3	K3

14.	a) Suppose we want to test the hypothesis that a given population distribution is exponential with mean 100; that is, $F(x) = 1 - e^{-x/100}$. If the (ordered) values from a sample of size 10 from this distribution are 66, 72, 81, 94, 112, 116, 124, 140, 145, 155, what conclusion can be drawn? (OR) b) The following data shows the breaking strength (in lbs) of random samples of two kinds of 2 inch cotton ribbons. Use median test to examine the null hypothesis that there is no difference in the average strength of the two kinds of ribbons. Type I ribbon: 1, 5, 7, 9, 15, 17, 21, 23 Type II ribbon: 2, 6, 10, 12, 18, 20, 26, 28, 32	CO4	K4																		
Q. No.	SECTION C Answer all the questions (6 x 10=60)	CO	KL																		
15.	a) It is known that if a signal of value μ is sent from location A, then the value received at location B is normally distributed with mean μ and standard deviation 2, i.e., the random noise added to the signal is an $N(0, 4)$ random variable. There is reason for the people at location B to suspect that the signal value $\mu = 8$ will be sent on a particular day. Test this hypothesis if the same signal value is independently sent five times and the average value received at location B is $\bar{X} = 9.5$. (OR) b) A public health official claims that the mean home water use is 350 gallons a day. To verify this claim, a study of 20 randomly selected homes was instigated with the result that the average daily water uses of these 20 homes were as follows: 340, 344, 362, 375, 356, 386, 354, 364, 332, 402, 340, 355, 362, 322, 372, 324, 318, 360, 338, 370 Do the data contradict the official's claim?	CO3	K3																		
16.	a) When $v_1 = 2$, show that the significance level of F corresponding to a significant probability p is $F = \frac{v_2}{2} [p^{-(2/v_2)} - 1]$ where v_1 and v_2 have their usual meanings. (OR) b) An IQ test was administered to 5 persons before and after they were trained. The results are given below. Test whether there is any change in IQ after the training programme. <table border="1" style="margin-left: auto; margin-right: auto;"><tr><td>Candidates</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>IQ before training</td><td>110</td><td>120</td><td>123</td><td>132</td><td>125</td></tr><tr><td>IQ after training</td><td>120</td><td>118</td><td>125</td><td>136</td><td>121</td></tr></table>	Candidates	1	2	3	4	5	IQ before training	110	120	123	132	125	IQ after training	120	118	125	136	121	CO4	K4
Candidates	1	2	3	4	5																
IQ before training	110	120	123	132	125																
IQ after training	120	118	125	136	121																

17.	a) Calculate the two regression equations of X on Y and Y on X from the data given below, taking deviations from the actual means of X and Y: <table border="1"><tr><td>Price(Rs.)</td><td>10</td><td>12</td><td>13</td><td>12</td><td>16</td><td>15</td></tr><tr><td>Amount Demanded</td><td>40</td><td>38</td><td>43</td><td>45</td><td>37</td><td>43</td></tr></table> Estimate the likely demand when the price is Rs. 20. (OR) b) Given the following data, estimate the marks in Mathematics obtained by a students who secured 60 marks in English: <table border="1"><tr><td>Mean marks in Mathematics</td><td>80</td></tr><tr><td>Mean marks in English</td><td>50</td></tr><tr><td>S.D. of marks in Mathematics</td><td>15</td></tr><tr><td>S.D. of marks in English</td><td>10</td></tr><tr><td>Coefficient of correlation</td><td>0.4</td></tr></table>	Price(Rs.)	10	12	13	12	16	15	Amount Demanded	40	38	43	45	37	43	Mean marks in Mathematics	80	Mean marks in English	50	S.D. of marks in Mathematics	15	S.D. of marks in English	10	Coefficient of correlation	0.4	CO4	K4
Price(Rs.)	10	12	13	12	16	15																					
Amount Demanded	40	38	43	45	37	43																					
Mean marks in Mathematics	80																										
Mean marks in English	50																										
S.D. of marks in Mathematics	15																										
S.D. of marks in English	10																										
Coefficient of correlation	0.4																										
18.	a) An auto rental firm is using 15 identical motors that are adjusted to run at a fixed speed to test 3 different brands of gasoline. Each brand of gasoline is assigned to exactly 5 of the motors. Each motor runs on 10 gallons of gasoline until it is out of fuel. The following represents the total mileages obtained by the different motors: Gas 1: 220, 251, 226, 246, 260 Gas 2: 244, 235, 232, 242, 225 Gas 3: 252, 272, 250, 238, 256 Test the hypothesis that the average mileage obtained is not affected by the type of gas used. Use the 5 percent level of significance. (OR) b) Three experimenters determine the moisture content of samples or a powder, each man taking a sample from each of 4 Consignments. The results are given below. <table border="1"><tr><th rowspan="2">Experimenters</th><th colspan="4">Consignments</th></tr><tr><th>I</th><th>II</th><th>III</th><th>IV</th></tr><tr><td>A</td><td>9</td><td>10</td><td>9</td><td>10</td></tr><tr><td>B</td><td>12</td><td>11</td><td>9</td><td>11</td></tr><tr><td>C</td><td>11</td><td>12</td><td>10</td><td>12</td></tr></table> Perform an analysis of variance on these data and discuss whether there is any significant difference between consignments or between experimenters.	Experimenters	Consignments				I	II	III	IV	A	9	10	9	10	B	12	11	9	11	C	11	12	10	12	CO5	K5
Experimenters	Consignments																										
	I	II	III	IV																							
A	9	10	9	10																							
B	12	11	9	11																							
C	11	12	10	12																							

19.	a) Suppose the weekly number of accidents over a 30-week period is as follows: 8, 0, 0, 1, 3, 4, 0, 2, 12, 5, 1, 8, 0, 2, 0, 1, 9, 3, 4, 5, 3, 3, 4, 7, 4, 0, 1, 2, 1, 2. Test the hypothesis that the number of accidents in a week has a Poisson distribution. (OR) b) A randomly chosen group of 20,000 nonsmokers and one of 10,000 smokers were followed over a 10-year period. The following data relate the numbers of them that developed lung cancer during that period. <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th><th>Smokers</th><th>Nonsmokers</th><th>Total</th></tr> </thead> <tbody> <tr> <td>Lung cancer</td><td>62</td><td>14</td><td>76</td></tr> <tr> <td>No lung cancer</td><td>9938</td><td>19986</td><td>29924</td></tr> <tr> <td>Total</td><td>10000</td><td>20000</td><td>30000</td></tr> </tbody> </table> Test the hypothesis that smoking and lung cancer are independent. Use the 1 percent level of significance.		Smokers	Nonsmokers	Total	Lung cancer	62	14	76	No lung cancer	9938	19986	29924	Total	10000	20000	30000	CO5	K5
	Smokers	Nonsmokers	Total																
Lung cancer	62	14	76																
No lung cancer	9938	19986	29924																
Total	10000	20000	30000																
20.	a) The following data shows the weights in kgs of a random sample of 38 students: 46, 49, 57, 64, 46, 57, 50, 65, 66, 50, 59, 47, 47, 59, 53, 67, 60, 53, 48, 67, 54, 48, 69, 61, 54, 48, 61, 54, 48, 62, 49, 55, 63, 56, 49, 64, 50, 51. Apply ordinary sign test and Wilcoxon signed test to examine whether the median weight in the population is 50. (OR) b) The following are the prices of a commodity in rupees per kg from two random samples of shops from two cities A and B. Apply run test to examine if the prices of the commodity in the two cities is the same. City A: 6.9, 11.2, 14.0, 13.2, 9.1, 13.9, 16.1, 9.3, 2.4, 6.4, 18.0, 11.5 City B: 15.5, 11.1, 16.0, 15.8, 18.2, 13.7, 18.3, 9.0, 17.2, 17.8, 13.0, 15.1	CO5	K6																
