Q. Code:829906

Reg. No. | | | | | | | | | | | | | | |

B.E / B.TECH. DEGREE EXAMINATIONS, MAY 2024

Third Semester

MA18353 – PROBABILITY AND STATISTICS FOR DATA SCIENCE

(Artificial Intelligence and Data science)

(Regulation 2018 / Regulation 2018A)

Use of Normal, t, F, and Chi-square tables are permitted

TIME: 3 HOURS **MAX. MARKS: 100**

CO 1 The students will have a fundamental knowledge of the concepts of probability.

CO 2 The students will have knowledge of standard distributions which is more relevant to Data Science and its applications.

CO 3 The students will have a notion of sampling distributions and statistical techniques used in Data science.

CO 4 The students will be able to analyse and interpret the data based on the sample tests.

CO 5 The students will acquire knowledge on Random processes and its applications.

PART- A (10 x 2 = 20 Marks)

(Answer all Questions)

		CO	RBT LEVEL
1.	Given that $P(A)=3/8, P(B)=5/8$ and $P(A \cup B)=3/4$. Verify whether A and B are independent.	**1**	**2**
2.	A die is thrown twice and the sum of the numbers appearing is observed to be 6. What is the conditional probability that the number 4 has appeared at least once?	**1**	**2**
3.	A continuous random variable X has the probability function $f(x) = k(1+x),\ 2 \le x \le 5$. Find the value of k.	**2**	**2**
4.	The no. of monthly breakdowns of a computer is a random variable having Poisson distribution with mean 1.8. Find the probability that this computer will function for a month with only one breakdown.	**2**	**2**
5.	A random sample of 500 apples is taken from a large consignment and 60 are found to be bad. Obtain the 95% confidence limits for the percentage of bad apples in the sample.	**3**	**2**
6.	Define type – I and type – II errors in sampling distribution.	**3**	**2**
7.	The mean life-time of a sample of 25 bulbs is found as 1550 hours with a standard deviation of 120 hours and the population mean is 1600. Calculate t – value.	**4**	**2**
8.	Write any two uses of chi-square test.	**4**	**1**
9.	A stationary random process $\{X(t)\}$ has autocorrelation function given by $R_{xx}(\tau)=\frac{25\tau^2+36}{6.25\tau^2+4}$. Find $E[x^2(t)]$	**5**	**2**

10. State any two properties of cross correlation functions. **5** **1**

PART- B (5 x 14 = 70 Marks)

			Marks	CO	RBT LEVEL
11. (a)	**(i)**	An incomplete frequency distribution is given as follows:	**(7)**	**1**	**3**

Variable	10-20	20-30	30-40	40-50	50-60	60-70	70-80	Total
Frequency	12	30	f_1	65	f_2	25	18	229

Given that the median value is 46, determine the missing frequencies f_1 and f_2, using the median formula.

	(ii)	The probability that doctor A will diagnose a disease X correctly is 60%. The probability that a patient will die by his treatment after correct diagnosis is 40% and the probability of death by wrong diagnosis is 70%. A patient of doctor A, who had disease X, died. What is the probability that his disease was diagnosed correctly?	**(7)**	**1**	**3**

(OR)

(b)		Find (i) Inter-quartile Range (ii) Quartile Deviation and (iii) Coefficient of Quartile deviation for the following distribution:	**(14)**	**1**	**3**

Class Interval	0-15	15-30	30-45	45-60	60-75	75-90	90-100
f	8	26	30	45	20	17	4

12. (a)	**(i)**	The scores secured by a candidate in a certain test are normally distributed with mean 500 and standard deviation 100. What percentage of candidate receives the scores between 400 and 600?	**(7)**	**2**	**3**
	(ii)	The joint probability density function of (X, Y) is given by $f(x,y)=\begin{cases}\frac{6}{5}(x+y^2), 0\le x\le 1, 0\le y\le 1\\ 0, elsewhere\end{cases}$. Obtain the marginal probability density functions of X and Y.	**(7)**	**2**	**3**

(OR)

(b)		The joint probability mass function of X and Y is given below:	**(14)**	**2**	**3**

		X	
		-1	1
Y	0	$\frac{1}{8}$	$\frac{3}{8}$
	1	$\frac{2}{8}$	$\frac{2}{8}$

(i) Find the correlation coefficient between X and Y and (ii) Are X and Y independent?

13. (a) **(i)** A college conducts both day and night classes intended to be identical. A sample of 100 day students yields examination results as under: $\bar{x}_1=72.4$ and $\sigma_1=14.8$. A sample of 200 night students yields examination results as under $\bar{x}_2=73.9$ and $\sigma_2=17.9$. Are the two means statistically equal at 5% level? **(7) 3 4**

(ii) In a sample of 300 units of a manufactured product, 65 units were found to be defective and in another sample of 200 units, there were 35 defectives. Is there significant difference in the proportion of defectives in the samples at the 5% level of significance? **(7) 3 4**

(OR)

(b) **(i)** A study shows that 16 of 200 tractors produced on one assembly line required extensive adjustments before they could be shipped, while the same was true for 14 of 400 tractors produced on another assembly line. At the 0.01 level of significance, does this support the claim that the second production line does superior work? **(7) 3 4**

(ii) The store keeper wanted to buy a large number of light bulbs of two brands A and B. He bought 100 bulbs from each brand and found by testing that brand A had been life time of 1120 hours and standard deviation of 75 hours; brand B had mean lifetime of 1060 hours and standard deviation of 80 hours. Examine if the difference is significant. (Test at 1% level of significance) **(7) 3 4**

14. (a) **(i)** The heights of 8 males participating in an athletic championship are found to be 175 cm, 168 cm, 165 cm, 170 cm, 167 cm, 173 cm, 170 cm and 168 cm. Can we conclude that the average height is greater than 165 cm? Test at 5% level of significance. **(7) 4 4**

(ii) One thousand girls in a college were graded according to their I.Q. and the economic conditions of their homes. Use Chi-square test to find out whether there is any association between economic conditions at home and I.Q. of girls: **(7) 4 4**

	I.Q.		
	High	Low	Total

Economic conditions	Rich	100	300	400
	Poor	350	250	600
	Total	450	350	1000

(OR)

(b) **(i)** In a laboratory experiment, two samples gave the following results: **(7)** **4** **4**

Sample	Size	Sample mean	Sum of squares of deviation from the mean
1	10	15	90
2	12	14	108

Test the equality of sample variances at 5% level of significance.

(ii) The means of two random samples of size 9 and 7 are 196.42 and 198.82 respectively. The sums of the squares of the deviation from the mean are 26.94 and 18.73 respectively. Can the sample be considered to have been drawn from the same normal population? **(7)** **4** **4**

15. (a) **(i)** The cross power spectrum of real random processes $X(t)$ and $Y(t)$ is given by $S_{XY}(\omega)=\begin{cases} a+ib\omega \; if \; |\omega|<1 \\ 0, elsewhere \end{cases}$. Find the cross correlation function. **(7)** **5** **3**

(ii) The autocorrelation function of an Ergodic process $X(t)$ is $R_{xx}(\tau)=\begin{cases} 1-|\tau| if |\tau|\le 1 \\ 0, otherwise \end{cases}$. Obtain the spectral density of $X(t)$. **(7)** **5** **3**

(OR)

(b) **(i)** Show that the random process $X(t)=A\cos(\omega_0 t+\theta)$ is wide-sense stationary, if A and ω_0 are constants and θ is a uniformly distributed random variable in $(0, 2\pi)$. **(7)** **5** **3**

(ii) Find the power spectral density of a wide sense stationary process with autocorrelation function $R(\tau)=e^{-\alpha\tau^2}$. **(7)** **5** **3**

PART- C (1 x 10 = 10 Marks)

(Q.No.16 is compulsory)

		Marks	CO	RBT LEVEL
16.	If the probability density function of a continuous random variable X is given by $f(x)=k(1+x)e^{-2x}$, in $-1\le x<\infty$, find the value of k, mean, variance and the moment generating function of X.	**(10)**	**2**	**3**

****