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

B.E / B.TECH. DEGREE EXAMINATIONS, MAY 2023

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	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. Find the range and coefficient of range of the weights of 10 students from the following data: 41, 20, 15, 65, 73, 84, 53, 35, 71 & 55. **1** **2**

2. A, B and C are three mutually exclusive and exhaustive events associated with a random experiment. Find $P(A)$ given that $P(B) = (3/2)\,P(A)$ and $P(C) = (1/2)P(B)$. **1** **3**

3. A random variable X has the following probability distribution **2** **2**

$X = x$	-2	-1	0	1	2	3
$P(x)$	0.1	k	0.2	$2k$	0.3	k

Find the value of k.

4. The joint probability density function of the random variable (X, Y) is **2** **2**

$$f(x,y)=\begin{cases} cxy & ,0<x<2\,;0<y<2 \\ 0 & ,otherwise \end{cases}$$

. Find the value of c .

5. Define Type I and Type II errors in sampling. **3** **1**

6. In a random selection of 64 of 600 road crossing in a town, the mean member of automobile accidents per year was found to 4.2 and the sample standard deviation was 0.8. Construct a 95% confidence interval for the mean number of automobile accidents per crossing per year. **3** **3**

7. A sample of size 13 gave an estimated population variance of 3.0, while another Sample of size 15 gave an estimate of 2.5. Calculate the test statistic. **4** **2**

8. What is the expected frequency of 'a' in the following 2 x 2 contingency table? **4** **2**

a	b
c	d

9. The power spectral density function of a zero mean wide-sense stationary process $\{X(t)\}$ is given by $S(\omega) = \begin{cases} 1, & |\omega| < \omega_0 \\ 0, & elsewhere \end{cases}$. Find $R(\tau)$. **5** **3**

10. Can the function $\frac{\omega+7}{\omega^2+5}$ be valid power density spectrum? **5** **2**

PART- B (5 x 14 = 70 Marks)

Marks **CO** **RBT LEVEL**

11.(a) **(i)** Calculate the quartile deviation and its coefficient from the following data: **(7)** **1** **3**

Marks obtained	10	20	30	40	50	60
No. f Students	4	7	15	8	7	2

(ii) Sixty percent of new drivers have had driver education. During their first year, new drivers without driver education have probability 0.08 of having an accident, but new drivers with driver education have only a 0.05 probability of an accident. What is the probability a new driver has had driver education, given that the driver has had no accident the first year? **(7)** **1** **3**

(OR)

(b) Calculate the mean, median and variance of the following data: **(14)** **1** **3**

Height (in cm)	95 – 105	105 – 115	115 – 125	125 – 135	135 – 145
Number of Children	19	23	36	70	52

12.(a) **(i)** The density function of a random variable X is given by $f(x) = kx(2-x),\ \ 0 \le x \le 2$. Find k, mean, variance and r^{th} moment. **(7)** **2** **3**

(ii) The joint probability mass function of (X, Y) is given by $p(x,y) = k(2x+3y)$, $x=0,1,2$, $y=1,2,3$. Find (i) the value of k (ii) the marginal probability mass function of X and Y and (iii) conditional probability distributions X given $Y=2$. **(7)** **2** **3**

(OR)

(b) **(i)** The atoms of radio active element are randomly disintegrating. If every gram of this element, on average, emits 3.9 alpha particles per second, what is the probability during the next second the number of alpha particles emitted from 1 gram is (i) atmost 6 (ii) atleast 2 and (iii) atleast 3 and atmost 6 ? **(7) 2 3**

(ii) A coin is tossed 300 times. Using central limit theorem, find the probability that number of heads obtained is between 140 and 150. **(7) 2 3**

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: $\overline{x}_1 = 72.4$ and $\sigma_1 = 14.8$. A sample of 200 night students yields examination results as under $\overline{x}_2 = 73.9$ and $\sigma_2 = 17.9$. Are the two means statistically equal at 5% level? **(7) 3 3**

(ii) On a certain day, 74 trains were arriving on time at Delhi station during the rush hours and 83 were late. At New Delhi, there were 65 on time and 107 late. Is there any difference in the proportions arriving on time at the two stations? (Test at 5% level of significance) **(7) 3 3**

(OR)

(b) **(i)** To test the claim that the resistance of electric wire can be reduced by more than 0.050 ohm by alloying, 32 values obtained for standard wire yielded $\overline{x_1} = 0.136$ ohm and $s_1 = 0.004$ ohm, and 32 values obtained for alloyed wire yielded $\overline{x_2} = 0.083$ ohm and $s_1 = 0.005$ ohm. At the 0.05 level of significance, does this support the claim? **(7) 3 3**

(ii) 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 3**

14.(a) **(i)** Two independent samples of sizes 8 and 7 contained the following values. **(7) 4 3**

Sample1	19	17	15	21	16	18	16	14
Sample2	15	14	15	19	15	18	16	

Is the difference between the sample means significant?

(ii) The following figures show the distribution of digits in number chosen at random from a telephone directory: **(7) 4 3**

Digits:	0	1	2	3	4	5	6	7	8	9
Frequency:	1026	1107	997	966	1075	933	1107	972	964	853

Test at 5% level whether the digits may be taken to occur equally frequently in the directory.

(OR)

(b) **(i)** Two random samples of sizes 9 and 6 gave the following values of the variable. **(7) 4 3**

Sample 1	15	22	28	26	18	17	29	21	24
Sample 2	8	12	9	16	15	10	-	-	-

Test whether there is any significance difference between the population variances at 5% level of significance.

(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 3**

		I.Q.		
		High	Low	Total
Economic conditions	Rich	100	300	400
	Poor	350	250	600
	Total	450	350	1000

15.(a) Consider two random process $X(t) = 3\cos(\omega t + \theta)$ and $Y(t) = 2\cos(\omega t + \theta - \pi/2)$ where θ is a random variable uniformly distributed in $(0, 2\pi)$. Prove that $\sqrt{R_{XX}(0).R_{YY}(0)} \geq |R_{XY}(\tau)|$. **(14) 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. Let X and Y be random variables having joint density function $f(x,y) = \begin{cases} \frac{3}{2}(x^2+y^2), 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 \quad , \quad otherwise \end{cases}$. Find the correlation coefficient between X and Y. **(10) 2 3**

*****