Q. Code:389560

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

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

Third Semester

MA22356 – STATISTICS FOR DATA SCIENCE

(Artificial Intelligence and Data Science)

(Regulation 2022)

(Use of statistical tables are permitted)

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

COURSE OUTCOMES	STATEMENT	RBT LEVEL
CO 1	The student will be able to acquire fundamental knowledge of the concepts of probability.	**3**
CO 2	The student will be able to achieve an understanding of standard distributions which is more relevant to Data Science and its applications	**3**
CO 3	The student will be able to test a hypothesis by measuring and examining a random sample of the population.	**3**
CO 4	The student will be able to classify and apply the related analysis of variance techniques in all fields of scientific experimentation.	**3**
CO 5	The student will be able to apply statistical quality control theory on real time problems.	**3**

PART- A (20 x 2 = 40 Marks)

(Answer all Questions)

		CO	RBT LEVEL
1.	Find out mode of the following distribution:	**1**	**2**

Class interval	0-10	10-20	20-30	30-40	40-50
Frequency	8	10	15	7	10

2.	Find the value of k for the following distribution whose mean is 16.6.					**1**	**2**

x	8	12	15	k	20	25	30
f	12	16	20	24	16	8	4

3.	Ten cards numbered 1 to 10 are placed in a box, mixed up thoroughly and then one card is drawn randomly. If it is known that the number on the drawn card is more than 3, what is the probability that it is an even number?	**1**	**2**
4.	An instructor has a question bank consisting of 300 easy True / False questions, 200 difficult True / False questions, 500 easy multiple choice questions and 400 difficult multiple choice questions. If a question is selected at random from the question bank, what is the probability that it will be an easy question given that it is a multiple choice question?	**1**	**2**

5.	A random variable X has the probability density function *f(x)* given by $f(x) = \begin{cases} Cx\, e^{-x}, & x \geq 0 \\ 0, & otherwise \end{cases}$. Find the value of C.	**2**	**2**
6.	The cumulative distribution function of a random variable X is $F(X) = 1-(1+x)\, e^{-x},\ x > 0$. Find the probability density function of X.	**2**	**2**
7.	Find the moment generating function of the random variable X whose probability density function is $f(x) = \begin{cases} \frac{1}{10}, & 0<x<10 \\ 0, & elsewhere \end{cases}$.	**2**	**2**
8.	The mean of Binomial distribution is 20 and standard deviation is 4. Determine the parameters of the distribution.	**2**	**2**
9.	A sample of 100 students is taken from a large population. The mean height of the students in this sample is 160 cm. If in the population, the mean height is 165 cm, and the S.D. is 10 cm, compute the test statistic.	**3**	**3**
10.	In a sample of 1000 people in Maharashtra, 540 are rice eaters and the rest are wheat eaters. Compute the test statistic.	**3**	**2**
11.	Explain Type – I error and Type – II error in sampling.	**3**	**2**
12.	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 mean 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. Calculate the standard error.	**3**	**2**
13.	What is the aim of the design of experiments?	**4**	**2**
14.	Write down the format of ANOVA table for two factor of classification.	**4**	**2**
15.	Compare and contrast the Completely Randomised Design with the Randomised Block Design.	**4**	**2**
16.	Why a 2×2 Latin square design is not possible? Explain	**4**	**2**
17.	What is control chart? Name the types of control charts.	**5**	**2**
18.	15 tape-recorders were examined for quality control test. The number of defects in each tape-recorder is recorded below. 2, 4, 3, 1, 1, 2, 5, 3, 6, 7, 3, 1, 4, 2, 1.find the central line.	**5**	**2**
19.	Explain the term tolerance limits in statistical quality control.	**5**	**2**
20.	Explain the term chance variation in statistical quality control.	**5**	**2**

PART- B (5 x 10 = 50 Marks)

		Marks	CO	RBT LEVEL
21(a)	Calculate mean and variance for the following distribution:	**(10)**	**1**	**3**

Classes	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	3	7	12	15	8	3	2

(OR)

(b) Of the students in a college, it is known that 60% reside in hostel and 40% are day scholars (not residing in hostel). Previous year results report that 30% of all students who reside in hostel attain A grade and 20% of day scholars attain A grade in their annual examination. At the end of the year, one student is chosen at random from the college and he has an A grade, what is the probability that the student is a hostlier? **(10) 1 3**

22.(a) i Out of 800 families with 4 children each, how many families would be expected to have 2 boys and 2 girls? **(4) 2 3**

ii Let $X_1, X_2, \ldots, X_{100}$ be independent and identically distributed random variables with mean $\mu = 2$ and $\sigma^2 = \frac{1}{4}$. Find $P(192 < X_1 + X_2 + \ldots + X_{100} < 210)$. **(6) 2 3**

(OR)

(b) Find correlation coefficient between $X \wedge Y$ for the following discrete bivariate distribution. **(10) 2 3**

Y \ X	5	15
10	0.2	0.4
20	0.3	0.1

23.(a) Test the significance of the difference between the means of the samples, drawn from two normal populations with the same S.D. from the following data: **(10) 3 3**

	Size	Mean	S.D
Sample 1	100	61	4
Sample 2	200	63	6

(OR)

(b) Two random samples of sizes 9 and 6 gave the following values of the variable. **(10) 3 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.

24.(a) Three samples A, B, C have been obtained from normal populations with equal variances. Test whether the population means are equal at 5% level. **(10) 4 3**

Sample A	12	14	12	9	13
Sample B	9	9	5	7	10
Sample C	7	8	10	11	14

(OR)

(b) In a Latin square experiment noted below, the yields in quintals per acre on the paddy crop carried out for testing the effect of five fertilizers A, B, C, D, E are given. Analyse the data for variations. **(10) 4 3**

B25	A18	E27	D30	C27
A19	D31	C29	E26	B23
C28	B22	D33	A18	E27
E28	C26	A20	B25	D33
D32	E25	B23	C28	A20

25.(a) The measurements of pitch diameter of thread in air-craft fitting is checked with 5 samples each containing 5 items at equal intervals of time. The measurements are given below. Construct mean chart and state your inference from the charts. **(10) 5 3**

Sample Number	Measurements				
1	46	45	44	43	42
2	41	41	44	42	40
3	40	40	42	40	42
4	42	43	43	42	45
5	43	44	47	47	45

(OR)

(b) In an integrated circuit production line, samples of 100 units are checked to electrical specifications on alternate days of a month and the results declared as number of defectives are tabulated below. Draw a p-chart and comment on the results. **(10) 5 3**

Sample No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. of defective	24	38	62	34	26	36	38	52	33	44	44	52	45	30	34

PART- C (1 x 10 = 10 Marks)

(Q.No.26 is compulsory)

	Marks	CO	RBT LEVEL

26. The mean height of 50 male students who showed above average participation in college athletics was 68·2 inches with a standard deviation of 2·5 inches; while 50 male students who showed no interest in such participation had a mean height of 67·5 inches with a standard deviation of 2·8 inches. Test the hypothesis that male students who participate in college athletics are taller than other male students. **(10) 3 3**

****