Reg. No. : E N G G T R E E . C O M

Question Paper Code : 51329

B.E./B.Tech. DEGREE EXAMINATIONS, APRIL/MAY 2024.

Third / Fourth Semester

Environmental Engineering

MA 3391 — PROBABILITY AND STATISTICS

(Common to : Artificial Intelligence and Data Science / Biotechnology and Biochemical Engineering / Computer Science and Business Systems / Plastic Technology)

(Regulations 2021)

Time : Three hours Maximum : 100 marks

(use of statistical table is pemitted)

Answer ALL questions.

PART A — (10 × 2 = 20 marks)

1. During off-peak hours a commuter train has five cars. Suppose a commuter is twice as likely to select the middle car (#3) as to select either adjacent car (#2 or #4), and is twice as likely to select either adjacent car as to select either end car (#1 or #5). Find the probability that one of the three middle cars is selected.

2. The probability that a certain kind of component will survive a shock test is 3/4. Find the probability that exactly 2 of the next 4 components tested survive.

3. A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is $f(x,y) = k(2x+3y)$; $0 \le x \le 1$, $0 \le y \le 1$. Find the value of k.

4. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours. What is the probability that a random sample of 16 bulbs will have an average life of less than 775 hours?

5. Data pertaining to height of 5 school students are given as 149, 150,151, 138,148 cms. Obtain a point estimate for the mean μ.

6. A random sample of size $n = 100$ is taken from a population with $\sigma = 5.1$. Given that the sample mean is $\bar{x} = 21.6$, construct a 95% confidence interval for the population mean μ.

7. Write the benefit of applying non-parametric tests compared to parametric tests.

8. What is the Kolmogorov-Smirnov Test? Is it non parametric?

9. Which control chart is not influenced by the sample size? What probability distribution does it follow?

10. What is the advantage of using control charts for attributes compared to variable control charts?

PART B — (5 × 16 = 80 marks)

11. (a) (i) Customers are used to evaluate preliminary product designs. In the past, 95% of highly successful products received good reviews, 60% of moderately successful products received good reviews, and 10% of poor products received good reviews. In addition, 40% of products have been highly successful, 35% have been moderately successful, and 25% have been poor products.

(1) What is the probability that a product attains a good review?

(2) If a new design attains a good review, what is the probability that it will be a highly successful product?

(3) If a product does not attain a good review, what is the probability that it will be a highly successful product? (8)

(ii) Assume that each of your calls to a popular radio station has a probability of 0.02 of connecting, that is, of not obtaining a busy signal. Assume that your calls are independent.

(1) What is the probability that your first call that connects is your tenth call?

(2) What is the probability that it requires more than five calls for you to connect?

(3) What is the mean number of calls needed to connect? (8)

Or

(b) (i) The thickness of a flange on an aircraft component is uniformly distributed between 0.95 and 1.05 millimeters.

(1) Determine the cumulative distribution function of flange thickness.

(2) Determine the proportion of flanges that exceeds 1.02 millimeters.

(3) What thickness is exceeded by 90% of the flanges?

(4) Determine the mean and variance of flange thickness. (8)

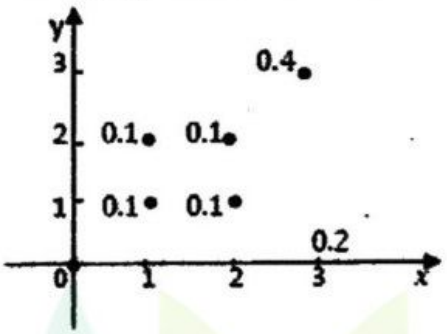
(ii) The life of a semiconductor laser at a constant power is normally distributed with a mean of 7000 hours and a standard deviation of 600 hours.

(1) What is the probability that a laser fails before 5000 hours?

(2) What is the life in hours that 95% of the lasers exceed?

(3) If three lasers are used in a product and they are assumed to fail independently, what is the probability that all three are still operating after 7000 hours? (8)

12. (a) For the discrete random variables X and Y with the joint distribution shown in the following figure: Determine the following :



(i) $P(X < 2, Y < 3)$

(ii) $P(1 < X < 2.5)$

(iii) $P(0 < Y < 2.5)$

(iv) $E(X), E(Y), V(X)$ and $V(Y)$

(v) Marginal probability distribution of the random variable X and Y.

(vi) Conditional probability distribution of Y given that $X = 1$.

(vii) Covariance and Correlation

(viii) Are X and Y independent? (16)

Or

(b) The joint pdf of X amount of almonds and Y amount of cashews were $f(x,y) = \begin{cases} 24xy, 0 \le x \le 1, 0 \le y \le 1, x + y \le 1 \\ 0 \quad otherwise \end{cases}$. Find the covariance and correlation between the random variables X and Y. (16)

13. (a) (i) Let X be exponentially distributed with parameter λ. Using maximum likelihood estimation, find an estimate for the parameter λ. (8)

(ii) If 83 male students are randomly chosen and yield an average of 6.6 hours of sleep with a standard deviation of 1.8 and 65 females are randomly selected with an average of 6.9 hours of sleep with a standard deviation of 1.5. Construct a 95% confidence interval for the difference between the two mean sleep hours for males and females. (8)

Or

(b) (i) For random sampling from normal population $N(\mu, \sigma^2)$, find the maximum likelihood estimators for μ and σ^2. (8)

(ii) Find the estimator for λ by the method of moments for the exponential distribution whose probability density function is given by $f(x,\lambda) = \frac{1}{\lambda} e^{-x/\lambda}$, $x > 0, \lambda > 0$. (8)

14. (a) (i) A manufacturer of electric irons, wishing to test the accuracy of the thermostat control at the 500°F setting, instructs a test engineer to obtain actual temperatures at that setting for 15 irons using a thermocouple. The resulting measurements are as follows :

494.6	510.8	487.5	493.2	502.6	485.0	495.9	498.2
501.6	497.3	492.0	504.3	499.2	493.5	505.8	

The engineer believes it is reasonable to assume that a temperature deviation from 500° of any particular magnitude is just as likely to be positive as negative (the assumption of symmetry) but wants to protect against possible non normality of the actual temperature distribution. Use signed-rank test to see whether the data strongly suggests incorrect calibration of the iron. (8)

(ii) The urinary fluoride concentration (parts per million) was measured both for a sample of livestock grazing in an area previously exposed to fluoride pollution and for a similar sample grazing in an unpolluted region :

Pollutted	21.3	18.7	23.0	17.1	16.8	20.9	19.7
Unpollutted	14.2	18.3	17.2	18.4	20.0		

Does the data indicate strongly that the true average fluoride concentration for livestock grazing in the polluted region is larger than for the unpolluted region? Use the Wilcoxon rank-sum test at level $\alpha = 0.01$. (8)

Or

(b) (i) The effectiveness of advertising for two rival products (Brand X and Brand Y) was compared. Market research at a local shopping centre was carried out, with the participants being shown adverts for two rival brands of coffee, which they then rated on the overall likelihood of them buying the product (out of 10, with 10 being "definitely going to buy the product"). Half of the participants gave ratings for one of the products, the other half gave ratings for the other product. Is there is a highly significant difference between the ratings given to each brand in terms of the likelihood of buying the product. Use U-test (take $\alpha = 0.05$) (8)

Brand X		Brand Y	
Participant	Rating	Participant	Rating
1	3	1	9
2	4	2	7
3	2	3	5
4	6	4	10
5	2	5	6
6	5	6	8

(ii) Four group of students were randomly assigned to taught with four different techniques and their achievement test scores were recorded. Are the distributions of test scores the same or do they differ in location? (take $\alpha = 0.05$). (8)

1	2	3	4
65	75	59	94
87	69	78	89
73	83	67	80
79	81	62	88

15. (a) A hard-bake process is used in conjunction with photolithography in semiconductor manufacturing. Fifteen samples, each of size five wafers, have been taken when we think the process is in control. The interval of time between samples or subgroups is one hour. The flow width measurement data (in x microns) from these samples are shown in Table. Establish statistical control of the flow width of the resist in this process using $\bar{x}$ and R charts. (16)

	Wafers				
Sample No.	1	2	3	4	5
1	1.3235	1.4128	1.6744	1.4573	1.6914
2	1.4314	1.3592	1.6075	1.4666	1.6109
3	1.4284	1.4871	1.4932	1.4324	1.5674
4	1.5028	1.6352	1.3841	1.2831	1.5507
5	1.5604	1.2735	1.5265	1.4363	1.6441
6	1.5955	1.5451	1.3574	1.3281	1.4198
7	1.6274	1.5064	1.8366	1.4177	1.5144
8	1.4190	1.4303	1.6637	1.6067	1.5519
9	1.3884	1.7277	1.5355	1.5176	1.3688
10	1.4039	1.6697	1.5089	1.4627	1.5220
11	1.4158	1.7667	1.4278	1.5928	1.4181

	Wafers				
Sample No.	1	2	3	4	5
12	1.5821	1.3355	1.5777	1.3908	1.7559
13	1.2856	1.4106	1.4447	1.6398	1.1928
14	1.4951	1.4036	1.5893	1.6458	1.4969
15	1.3589	1.2863	1.5996	1.2497	1.5471

Or

(b) (i) Frozen orange juice concentrate is packed in 6-oz cardboard cans. These cans are formed on a machine by spinning them from cardboard stock and attaching a metal bottom panel. By inspection of a can, we may determine whether, when filled, it could possibly Leak either on the side seam or around the bottom joint. Such a nonconforming can has an improper seal on either the side seam or the bottom panel. To establish the control chart, 15 samples of $n = 50$ cans each were selected at half-hour intervals over a three-shift period in which the machine was in continuous operation.

The data are shown in Table. Set up a control chart for the fraction of nonconforming cans produced by this machine. (8)

Sample Number	Number of Nonconforming Cans, D_i
1	12
2	15
3	8
4	10
5	4
6	7
7	16
8	9
9	14
10	10
11	5
12	6
13	17
14	12
15	22

(ii) Table presents the number of nonconformities observed in 26 successive samples of 100 printed circuit boards. Note that, for reasons of convenience, the inspection unit is defined as 100 boards. Set up a c chart for these data. (8)

Sample Number	Number of Nonconformities	Sample Number	Number of Nonconformities
1	21	14	19
2	24	15	10
3	16	16	17
4	12	17	13
5	15	18	22
6	5	19	18
7	28	20	39
8	20	21	30
9	31	22	24
10	25	23	16
11	20	24	19
12	24	25	17
13	16	26	15