

II B. Tech II Semester Regular Examinations, June/July - 2022

PROBABILITY AND STATISTICS

(Common to CSE, CST, CSE(AIML), CSE(AI), CSE(DS), CSE(AIDS), CSE(CS), CSE(IOTCSIBCT),
CSE(CSBS), CSE(IOT), AIDS, & AIML)

Time: 3 hours

Max. Marks: 70

Answer any **FIVE** Questions each Question from each unit
All Questions carry **Equal** Marks

- ~~~~~
- 1 a) What are the main methods of collecting primary data? State briefly the advantages and disadvantages of each. [7M]
b) What do you understand by skewness. What are the various methods of measuring skewness? [7M]

Or

- 2 a) What do you understand by a measure of dispersion ? What purpose does a measure of dispersion serve ? [6M]
b) Calculate the mean and standard deviation for the following table giving the age distribution of 542 members. [8M]

| | | | | | | | |
|----------------|---------|---------|---------|---------|---------|---------|---------|
| Age (in years) | 20 - 30 | 30 - 40 | 40 - 50 | 50 - 60 | 60 - 70 | 70 - 80 | 80 - 90 |
| No. of members | 3 | 61 | 132 | 153 | 140 | 51 | 2 |

- 3 a) Calculate correlation coefficient to the following data ; [7M]

| | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|----|
| x | 10 | 15 | 12 | 17 | 13 | 16 | 24 | 14 | 22 | 20 |
| y | 30 | 42 | 45 | 46 | 33 | 34 | 40 | 35 | 39 | 38 |

- b) Fit an exponential curve of the form $y = ab^x$ to the following data: [7M]

| | | | | | | | | |
|---|-----|-----|-----|-----|-----|-----|-----|-----|
| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| y | 1.0 | 1.2 | 1.8 | 2.5 | 3.6 | 4.7 | 6.6 | 9.1 |

Or

- 4 a) Obtain the equations of two lines of regression for the following data. Also obtain the estimate of X for Y = 70. [10M]

| | | | | | | | | |
|---|----|----|----|----|----|----|----|----|
| X | 65 | 66 | 67 | 67 | 68 | 69 | 70 | 72 |
| Y | 67 | 68 | 65 | 68 | 72 | 72 | 69 | 71 |

- b) State the principle of least squares and write the normal equations for the fitting of the Straight line. [4M]

- 5 a) The probability that a turbine will have a defective coil is 0.10, the probability that it will have defective blades is 0.15, and the probability that it will have both defects is 0.04. [7M]

- (i) What is the probability that a turbine will have one of these defects?
(ii) What is the probability that a turbine will have either of these defects?

- b) A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives. [7M]

Or

- 6 a) The diameter of an electric cable, say X , is assumed to be a continuous random variable with p.d.f. $f(x) = 6x(1-x)$, $0 \leq x \leq 1$. [8M]
(i) Check that $f(x)$ is p.d.f., and
(ii) Determine a number b such that $P(X < b) = P(X > b)$.
- b) Given a standard normal distribution, find the area under the curve that lies [6M]
(i) to the right of $z = 1.84$ and
(ii) between $z = -1.97$ and $z = 0.86$.

- 7 a) Define Population and sample with examples. [6M]
- b) For a chi-squared distribution, find [8M]
(i) $\chi^2_{0.025}$ when $\nu = 15$;
(ii) $\chi^2_{0.01}$ when $\nu = 7$.

Or

- 8 The pulse rate of 50 yoga practitioners decreased on the average by 20.2 [14M] beats/minute with s.d. of 3.5. (a) If $\bar{x} = 20.2$ is used as a point estimate of the true average decrease in the pulse rate, what can we assert with 95% confidence about the maximum error E . (b) Construct 99% confidence intervals for the true average decrease in pulse rate.
- 9 a) A random sample of 100 recorded deaths in the United States during the past [7M] year showed an average life span of 71.8 years. Assuming a population standard deviation of 8.9 years, does this seem to indicate that the mean life span today is greater than 70 years? Use a 0.05 level of significance.
- b) An urban community would like to show that the incidence of breast cancer is [7M] higher in their area than in a nearby rural area. If it is found that 20 of 200 adult women in the urban community have breast cancer and 10 of 150 adult women in the rural community have breast cancer, can we conclude at the 0.05 level of significance that breast cancer is more prevalent in the urban community?

Or

- 10 a) Past experience indicates that the time required for high school seniors to [7M] complete a standardized test is a normal random variable with a mean of 35 minutes. If a random sample of 20 high school seniors took an average of 33.1 minutes to complete this test with a standard deviation of 4.3 minutes, test the hypothesis, at the 0.05 level of significance, that $\mu = 35$ minutes against the alternative that $\mu < 35$ minutes.
- b) A manufacturer claims that the average tensile strength of thread A exceeds the [7M] average tensile strength of thread B by at least 12 kilograms. To test this claim, 50 pieces of each type of thread were tested under similar conditions. Type A thread had an average tensile strength of 86.7 kilograms with a standard deviation of 6.28 kilograms, while type B thread had an average tensile strength of 77.8 kilograms with a standard deviation of 5.61 kilograms. Test the manufacturer's claim using a 0.05 level of significance.



II B. Tech II Semester Regular Examinations, June/July - 2022

PROBABILITY AND STATISTICS

(Common to CSE, CST, CSE(AIML), CSE(AI), CSE(DS), CSE(AIDS), CSE(CS), CSE(IOTCSIBCT), CSE(CSBS), CSE(IOT), AIDS, CS, & AIML)

Time: 3 hours

Max. Marks: 70

Answer any **FIVE** Questions each Question from each unitAll Questions carry **Equal** Marks

- ~~~~~
- 1 a) Distinguish between primary and secondary data and discuss the various methods of collecting primary data. [7M]
- b) What do you understand by skewness and kurtosis? Point out their role in analyzing a frequency distribution. [7M]

Or

- 2 a) What do you understand by dispersion? Explain briefly the various methods used for measuring dispersion. [7M]
- b) Calculate the coefficient of skewness based on mean and median from the following distribution [7M]

| | | | | | | | | |
|----------------|--------|---------|---------|---------|---------|---------|---------|---------|
| Class interval | 0 - 10 | 10 - 20 | 20 - 30 | 30 - 40 | 40 - 50 | 50 - 60 | 60 - 70 | 70 - 80 |
| Frequency | 6 | 12 | 22 | 48 | 56 | 32 | 18 | 6 |

- 3 a) Obtain the correlation coefficient for the following data : [7M]

| | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|
| x | 48 | 60 | 72 | 62 | 56 | 40 | 39 | 52 | 30 |
| y | 62 | 78 | 65 | 70 | 38 | 54 | 60 | 32 | 31 |

- b) Fit a parabola $y = a + bx + cx^2$ to the following data: [7M]

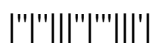
| | | | | | |
|---|------|-------|-------|-------|-------|
| x | 2 | 4 | 6 | 8 | 10 |
| y | 3.07 | 12.85 | 31.47 | 57.38 | 91.29 |

Or

- 4 a) In a partially destroyed laboratory, record of an analysis of correlation data, the following only are legible: Variance of $X = 9$, Regression equations: $8X - 10Y + 66 = 0$, $40X - 18Y = 214$. What are: (i) the mean values X and Y , (ii) the correlation coefficient between X and Y , and (iii) the standard deviation of Y ? [8M]
- b) The ranks of same 16 students in Mathematics and Physics are as follows. Two numbers within brackets denote the ranks of the students in Mathematics and Physics: (1,1) (2,10) (3,3) (4,4) (5,5) (6,7) (7,2) (8,6) (9,8) (10,11) (11,15) (12,9) (13,14) (14,12) (15,16) (16,13). Calculate the rank correlation coefficient for proficiencies of this group in Mathematics and Physics. [6M]
- 5 a) The probability that a construction company will get the tender for constructing a flyover is 0.33, the probability that it will get the tender for constructing an underpass is 0.28, and the probability that it will get both tenders is 0.13. (i) What is the probability that it will get at least one tender? (ii) What is the probability that it will get neither tender? [7M]
- b) Find the mean and the variance of the uniform probability distribution given by [7M]

$$f(x) = \frac{1}{n} \text{ for } x = 1, 2, 3, \dots, n.$$

Or

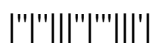


- 6 a) Find the probabilities that a random variable having the standard normal distribution will take on a value [7M]
(i) between 0.87 and 1.28;
(ii) between -0.34 and 0.62 ;
(iii) greater than 0.85 ;
(iv) greater than -0.65 .
- b) In a certain industrial facility, accidents occur infrequently. It is known that the probability of an accident on any given day is 0.005 and accidents are independent of each other. [7M]
(i) What is the probability that in any given period of 400 days there will be an accident on one day?
(ii) What is the probability that there are at most three days with an accident?
- 7 a) State the Central limit theorem. [4M]
- b) For an F -distribution, find [10M]
i) $F_{0.05}$ with $v_1 = 7$ and $v_2 = 15$;
ii) $F_{0.01}$ with $v_1 = 24$ and $v_2 = 19$;
iii) $F_{0.95}$ with $v_1 = 19$ and $v_2 = 24$;
iv) $F_{0.99}$ with $v_1 = 28$ and $v_2 = 12$.

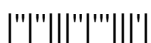
Or

- 8 a) Assuming that the population standard deviation is 0.3 , calculate the (i) 95% and (ii) 99% confidence intervals for the mean lead concentration in a river if the mean lead concentration recovered from a sample of lead measurements in 36 different locations is 2.6 gms/ml. [7M]
- b) The contents of seven similar containers of sulfuric acid are $9.8, 10.2, 10.4, 9.8, 10.0, 10.2,$ and 9.6 liters. Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution. [7M]
- 9 a) A manufacturer of sports equipment has developed a new synthetic fishing line that the company claims has a mean breaking strength of 8 kilograms with a standard deviation of 0.5 kilogram. Test the hypothesis that $\mu = 8$ kilograms against the alternative that $\mu \neq 8$ kilograms if a random sample of 50 lines is tested and found to have a mean breaking strength of 7.8 kilograms. Use a 0.01 level of significance. [7M]
- b) A commonly prescribed drug for relieving nervous tension is believed to be only 60% effective. Experimental results with a new drug administered to a random sample of 100 adults who were suffering from nervous tension show that 70 received relief. Is this sufficient evidence to conclude that the new drug is superior to the one commonly prescribed? Use a 0.05 level of significance. [7M]

Or



- 10 a) A vote is to be taken among the residents of a town and the surrounding county [7M]
to determine whether a proposed chemical plant should be constructed. The
construction site is within the town limits, and for this reason many voters in the
county believe that the proposal will pass because of the large proportion of town
voters who favor the construction. To determine if there is a significant difference
in the proportions of town voters and county voters favoring the proposal, a poll is
taken. If 120 of 200 town voters favour the proposal and 240 of 500 county residents
favor it, would you agree that the proportion of town voters favoring the proposal is
higher than the proportion of county voters? Use an $\alpha = 0.05$ level of significance.
- b) It is claimed that automobiles are driven on average more than 20,000 kilometers per [7M]
year. To test this claim, 100 randomly selected automobile owners are asked to keep
a record of the kilometers they travel. Would you agree with this claim if the random
sample showed an average of 23,500 kilometers and a standard deviation of 3900
kilometers?



II B. Tech II Semester Regular Examinations, June/July - 2022

PROBABILITY AND STATISTICS

(Common to CSE, CST, CSE(AIML), CSE(AI), CSE(DS), CSE(AIDS), CSE(CS), CSE(IOTCSIBCT), CSE(CSBS), CSE(IOT), AIDS, CS, & AIML)

Time: 3 hours

Max. Marks: 70

Answer any **FIVE** Questions each Question from each unitAll Questions carry **Equal** Marks

~~~~~

- 1 a) Explain the various methods that are used in the collection of primary data pointing out their merits and demerits. [7M]
- b) The following table shows the marks obtained by 100 candidates in an examination. Calculate the mean and standard deviation: [7M]

|                   |        |         |         |         |         |         |
|-------------------|--------|---------|---------|---------|---------|---------|
| Marks obtained    | 1 - 10 | 11 - 20 | 21 - 30 | 31 - 40 | 41 - 50 | 51 - 60 |
| No. of candidates | 3      | 16      | 26      | 31      | 16      | 8       |

Or

- 2 a) Define the various measures of central tendency. What purposes do these measurements serve. [6M]
- b) Obtain Karl Pearson's measure of skewness for the following data: [8M]

|           |        |         |         |         |         |         |         |
|-----------|--------|---------|---------|---------|---------|---------|---------|
| Values    | 5 - 10 | 10 - 15 | 15 - 20 | 20 - 25 | 25 - 30 | 30 - 35 | 35 - 40 |
| Frequency | 6      | 8       | 17      | 21      | 15      | 11      | 2       |

- 3 a) Two random variables have the regression lines with equations  $3x + 2y = 26$  and  $6x + y = 31$ . Find the means values and the correlation co-efficient between  $x$  and  $y$ . [7M]
- b) Fit a polynomial of the second degree to the form  $y = a + bx + cx^2$  to the following data by the method of least squares: [7M]

|     |   |   |   |    |    |
|-----|---|---|---|----|----|
| $x$ | 0 | 1 | 2 | 3  | 4  |
| $y$ | 1 | 0 | 3 | 10 | 21 |

Or

- 4 a) A sample of 12 fathers and their eldest sons gave the following data about their height in inches: [7M]

|        |    |    |    |    |    |    |    |    |    |    |    |    |
|--------|----|----|----|----|----|----|----|----|----|----|----|----|
| Father | 65 | 63 | 67 | 64 | 68 | 62 | 70 | 66 | 68 | 67 | 69 | 71 |
| Son    | 68 | 66 | 68 | 65 | 69 | 66 | 68 | 65 | 71 | 67 | 68 | 70 |

Calculate coefficient of rank correlation.

- b) Write the properties of regression coefficients. [7M]
- 5 a) Two cards are drawn at random from an ordinary deck of 52 playing cards. What is the probability of getting two aces if [7M]
- (i) the first card is replaced before the second card is drawn;
- (ii) the first card is not replaced before the second card is drawn?
- b) It is known that 5% of the books bound at a certain bindery have defective bindings. Find the probability that 2 of 100 books bound by this bindery will have defective bindings using [7M]
- (i) the formula for the binomial distribution;
- (ii) the Poisson approximation to the binomial distribution.

Or

- 6 a) A continuous random variable  $X$  has a p.d.f.  $f(x) = 3x^2$ ,  $0 \leq x \leq 1$ . Find  $a$  and  $b$  [7M]  
such that (i)  $P(X \leq a) = P(X > a)$ , and (ii)  $P(X > b) = 0.05$ .
- b) In a certain city district, the need for money to buy drugs is stated as the reason for [7M]  
75% of all thefts. Find the probability that among the next 5 theft cases reported in  
this district,  
(i) exactly 2 resulted from the need for money to buy drugs;  
(ii) at most 3 resulted from the need for money to buy drugs.

- 7 a) Find the value of  $F_{0.95}$  for  $\nu_1 = 10$  and  $\nu_2 = 20$  degrees of freedom. [7M]
- b) Determine 99% confidence interval for the mean of contents of soft drink bottles if [7M]  
contents of 7 such soft drink bottles are 10.2, 10.4, 9.8, 10.0, 9.8, 10.2, 9.6 ml.

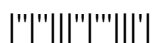
Or

- 8 The efficiency expert of a computer company tested 40 engineers to estimate the [14M]  
average time it takes to assemble a certain computer component, getting a mean of  
12.73 minutes and s.d. of 2.06 minutes. (a) If  $\bar{x} = 12.73$  is used as a point estimate  
of the actual average time required to perform the task, determine the maximum  
error with 99% confidence (b) construct 98% confidence intervals for the true  
average time it takes to do the job (c) with what confidence can we assert that the  
sample mean does not differ from the true mean by more than 30 seconds.

- 9 a) A random sample of 64 bags of white cheddar popcorn weighed, on average, 5.23 [7M]  
ounces with a standard deviation of 0.24 ounce. Test the hypothesis that  $\mu = 5.5$   
ounces against the alternative hypothesis,  $\mu < 5.5$  ounces, at the 0.05 level of  
significance.
- b) An experiment was performed to compare the abrasive wear of two different [7M]  
laminated materials. Twelve pieces of material 1 were tested by exposing each  
piece to a machine measuring wear. Ten pieces of material 2 were similarly tested.  
In each case, the depth of wear was observed. The samples of material 1 gave an  
average (coded) wear of 85 units with a sample standard deviation of 4, while the  
samples of material 2 gave an average of 81 with a sample standard deviation of 5.  
Can we conclude at the 0.05 level of significance that the abrasive wear of material  
1 exceeds that of material 2 by more than 2 units? Assume the populations to be  
approximately normal with equal variances.

Or

- 10 a) A commonly prescribed drug for relieving nervous tension is believed to be only [7M]  
60% effective. Experimental results with a new drug administered to a random  
sample of 100 adults who were suffering from nervous tension show that 70  
received relief. Is this sufficient evidence to conclude that the new drug is superior  
to the one commonly prescribed? Use a 0.05 level of significance.
- b) Test the hypothesis that the average content of containers of a particular lubricant [7M]  
is 10 liters if the contents of a random sample of 10 containers are 10.2, 9.7, 10.1,  
10.3, 10.1, 9.8, 9.9, 10.4, 10.3, and 9.8 liters. Use a 0.01 level of significance and  
assume that the distribution of contents is normal.



## II B. Tech II Semester Regular Examinations, June/July - 2022

## PROBABILITY AND STATISTICS

(Common to CSE, CST, CSE(AIML), CSE(AI), CSE(DS), CSE(AIDS), CSE(CS), CSE(IOTCSIBCT), CSE(CSBS), CSE(IOT), AIDS, CS, &amp; AIML)

Time: 3 hours

Max. Marks: 70

Answer any **FIVE** Questions each Question from each unit  
All Questions carry **Equal** Marks

- 1 a) Distinguish between primary source and secondary source of statistical data. What precautions would you take before using data from a secondary source? [7M]

- b) Calculate the quartile coefficient of skewness from the following data: [7M]

|                |         |         |          |           |           |           |           |           |
|----------------|---------|---------|----------|-----------|-----------|-----------|-----------|-----------|
| Weight (lbs)   | 70 - 80 | 80 - 90 | 90 - 100 | 100 - 110 | 110 - 120 | 120 - 130 | 130 - 140 | 140 - 150 |
| No. of persons | 12      | 18      | 35       | 42        | 50        | 45        | 20        | 8         |

Or

- 2 a) What is meant by measures of central tendency? What are the characteristics of a good measure of central tendency? [7M]

- b) Assume that a firm has selected a random sample of 100 from its production line and has obtain the data shown in the table below: [7M]

|                |         |         |         |         |         |         |         |
|----------------|---------|---------|---------|---------|---------|---------|---------|
| Class interval | 130-134 | 135-139 | 140-144 | 145-149 | 150-154 | 155-159 | 160-164 |
| Frequency      | 3       | 12      | 21      | 28      | 19      | 12      | 5       |

Compute the following: (i) The arithmetic mean, (ii) The standard deviation.

- 3 a) Calculate the correlation coefficient for the following heights(in inches) of fathers (X) and their sons (Y): [7M]

|   |    |    |    |    |    |    |    |    |
|---|----|----|----|----|----|----|----|----|
| X | 65 | 66 | 67 | 67 | 68 | 69 | 70 | 72 |
| Y | 67 | 68 | 65 | 68 | 72 | 72 | 69 | 71 |

- b) Find the normal equations for fitting of the parabola  $y = a + bx + cx^2$  to a set of  $n$  points. [7M]

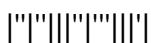
Or

- 4 a) Fit the curve of the form  $y = ae^{bx}$  to the following data: [8M]

|   |     |     |     |      |      |
|---|-----|-----|-----|------|------|
| x | 77  | 100 | 185 | 239  | 285  |
| y | 2.4 | 3.4 | 7.0 | 11.1 | 19.6 |

- b) Explain the difference between the correlation coefficient and rank correlation coefficient. [6M]

- 5 a) Among 40 condensers produced by a machine, 6 are defective. If we randomly check 5 condensers, what are the probabilities that [6M]  
(i) none are defective;  
(ii) all are defective?





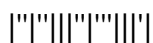
- b) A total of 46 percent of the voters in a certain city classify themselves as Independents, whereas 30 percent classify themselves as Liberals and 24 percent say that they are Conservatives. In a recent local election, 35 percent of the Independents, 62 percent of the Liberals, and 58 percent of the Conservatives voted. A voter is chosen at random. Given that this person voted in the local election, what is the probability that he or she is (i) an Independent?(ii) a Liberal?(iii) a Conservative?.

**Or**

- 6 a) If the probability density of a random variable is given by [7M]
- $$f(x) = \begin{cases} (k+2)x^3, & 0 < x < 10 \\ 0, & \text{elsewhere} \end{cases}$$
- find the value  $k$  and the probability that the random variable takes on a value (a) greater than  $3/4$ ; (b) between  $1/3$  and  $2/3$ .
- b) 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. [7M]
- 7 a) We know that silk fibers are very tough but in short supply. Engineers are making breakthroughs to create synthetic silk fibers that can improve everything from car bumpers to bullet-proof vests or to make artificial blood vessels. One research group reports the summary statistics  $n=18$ ,  $\bar{x}=22.6$ ,  $s=15.7$  for the toughness (MJ/m<sup>3</sup>) of processed fibers. Construct a 95% confidence interval for the mean toughness of these fibers. Assume that the population is normal. [7M]
- b) Find the values of [7M]
- (i)  $F_{0.95}$  for 15 and 12 degrees of freedom;
- (ii)  $F_{0.99}$  for 5 and 20 degrees of freedom.

**Or**

- 8 a) Using the mean of a random sample of size 150 to estimate the mean mechanical aptitude of mechanics of a large workshop and assuming  $\sigma=6.2$ , what can we assert with 0.99 probability about the maximum size of the error. [7M]
- b) The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter. [7M]
- 9 a) An electrical firm manufactures light bulbs that have a lifetime that is approximately normally distributed with a mean of 800 hours and a standard deviation of 40 hours. Test the hypothesis that  $\mu=800$  hours against the alternative,  $\mu \neq 800$  hours, if a random sample of 30 bulbs has an average life of 788 hours. [7M]



- b) In a study to estimate the proportion of residents in a certain city and its suburbs [7M]  
who favor the construction of a nuclear power plant, it is found that 63 of 100 urban  
residents favor the construction while only 59 of 125 suburban residents are in  
favor. Is there a significant difference between the proportions of urban and  
suburban residents who favor construction of the nuclear plant at 1% level ?

**Or**

- 10 a) In a study conducted by the Department of Human Nutrition and Foods at Virginia [7M]  
Tech, the following data were recorded on sorbic acid residuals, in parts per  
million, in ham immediately after dipping in a sorbate solution and after 60 days of  
storage:

| Sorbic Acid Residuals in Ham |                |               |
|------------------------------|----------------|---------------|
| Slice                        | Before Storage | After Storage |
| 1                            | 224            | 116           |
| 2                            | 270            | 96            |
| 3                            | 400            | 239           |
| 4                            | 444            | 329           |
| 5                            | 590            | 437           |
| 6                            | 660            | 597           |
| 7                            | 1400           | 689           |
| 8                            | 680            | 576           |

Assuming the populations to be normally distributed, is there sufficient evidence, at  
the 0.05 level of significance, to say that the length of storage influences sorbic acid  
residual concentrations?

- b) A builder claims that heat pumps are installed in 70% of all homes being [7M]  
constructed today in the city of Richmond, Virginia. Would you agree with this  
claim if a random survey of new homes in this city showed that 8 out of 15 had heat  
pumps installed? Use a 0.10 level of significance.

