# Chalapathi Institute of Technology

### DEPARTMENT OF CSE, CDS, CSC, CAI & CAIML

SUBJECT: P&S          CLASS: IICSE/CDS/CAI/CAIML          SEM: II SEM

## UNIT-1

1. Explain the collection of data and its methods?

2. Explain the Data Science Methodology?

3. Explain the Methods of Sampling?

4. Explain Data Visualisation?

5. Define Measures of Central Tendency. Write merits and demerits for mean, median and mode.

6. From the following, find the mean profits

| Profits for shops | 100-200 | 200-300 | 300-400 | 400-500 | 500-600 | 600-700 | 700-800 |
|---|---|---|---|---|---|---|---|
| No. of Shops | 10 | 18 | 20 | 26 | 30 | 28 | 18 |

7. Calculate the median from the following data

| Marks | 10-25 | 25-40 | 40-55 | 55-70 | 70-85 | 85-100 |
|---|---|---|---|---|---|---|
| frequency | 6 | 20 | 44 | 26 | 3 | 1 |

8. Find the mean deviation from mean of the following data.

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| students | 6 | 5 | 8 | 15 | 7 | 6 | 3 |

9. Find the Standard Deviation and Variance for the given data

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

10. Find Coefficient of Skewness for the following data.

| Marks | 0-8 | 8-16 | 16-24 | 24-32 | 32-40 | 40-48 | 48-56 |
|---|---|---|---|---|---|---|---|
| No. of Students | 9 | 17 | 21 | 35 | 38 | 13 | 2 |

## UNIT-II

1. Fit a Straight line for the following data

| x | 6 | 7 | 7 | 8 | 8 | 8 | 9 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| y | 5 | 5 | 4 | 5 | 4 | 3 | 4 | 3 | 3 |

2. Fit second degree parabola for the following data

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

3.  Find the Exponential curve of the type $y = ae^{bx}$ for the given data

| x | 1 | 5 | 7 | 9 | 12 |
|---|---|---|---|---|----|
| y | 10 | 15 | 12 | 15 | 21 |

4.  Find the Exponential curve of the type $y = ab^x$ for the given data

| x | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| y | 8.3 | 15.4 | 33.1 | 65.2 | 127.4 |

5.  Explain Scatter Diagram?

6.  Find the correlation coefficient for the given data

| x | 12 | 9 | 10 | 8 | 11 | 13 | 7 |
|---|----|---|----|---|----|----|---|
| y | 14 | 8 | 9 | 6 | 11 | 12 | 3 |

7.  Find the Rank Correlation coefficient for the following data

| x | 48 | 33 | 40 | 9 | 16 | 16 | 65 | 24 | 16 | 57 |
|---|----|----|----|---|----|----|----|----|----|----|
| y | 13 | 13 | 24 | 6 | 15 | 4 | 20 | 9 | 6 | 19 |

8.  Define Regression. Write its properties.

9.  Obtain regression equation of X on Y from the following table.

| Age in years | 48 | 33 | 40 | 9 | 16 | 16 | 65 | 24 | 16 | 57 |
|---|----|----|----|---|----|----|----|----|----|----|
| Blood pressure | 13 | 13 | 24 | 6 | 15 | 4 | 20 | 9 | 6 | 19 |

10. The equations of two regression lines $7x-16y+9 = 0$ and $5y-4x-3 = 0$. Find (i) r (ii) means of x, y

## UNIT-III

1.  (a) Explain Random Variables?

    (b) Random variable X has the following probability function

| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| P(x) | 0 | k | 2k | 2k | 3k | $K^2$ | $2k^2$ | $7k^2+k$ |

   Determine (i)k (ii) p(x<6), p(0<x<5), p(0<x<4) ,(iii)determine distribution of x (iv) Mean, (v) variance.

2.  For the continuous probability function $f(x) = k x^2 e^{-x}$ when x≥0, find (i) k (ii) Mean (iii) Variance.

3.  State and Prove Mean and Variance of Binomial Distribution.

4. (a) Ten coins are tossed simultaneously. Find the probability of getting at least seven heads.

   (b) A die is tossed thrice. A success is getting 1 or 6 on a toss. Find the mean and variance of the number of successes.

5. Fit a Binomial distribution to the following data

   | x | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
   |---|---|---|---|---|---|---|---|
   | y | 13 | 25 | 52 | 58 | 32 | 16 | 4 |

6. State and prove Mean and Variance of the Poisson distribution.

7. (a) 2 % of items of a factory are defective. The items are packed in boxes. What is the probability that there will be (i) 2 defective items? (ii) at least three defective items in a box of 100 items.

   (b) A sample of 3 items is selected at random from a box containing 10 Items if which 4 are defective. Find he expected no. of detective items.

8. Fit a Poisson distribution to the following data

   | x | 0 | 1 | 2 | 3 | 4 | **5** | total |
   |---|---|---|---|---|---|---|---|
   | y | 142 | 156 | 69 | 27 | 5 | 1 | 400 |

9. Explain chief-characteristics and Impotence of Normal distribution?

10. (a) For a normally distributed variate with mean 1 and standard deviation 3, Find the probabilities that (i) $3.43 \leq x \leq 6.19$          (ii) $-1.43 \leq x \leq 6.19$

    (b) A manufacturer knows from experience that the resistance of resistors he produces is normal with mean100 ohms and S.D 2 ohm. What percentage of resistors will have resistance between 98 and 102 ohms?

## UNIT-IV

1. Explain the following terms

   (i) Population                    (ii) Sample                (iii) Statistic

   (iv) Parameter                   (v) sampling distribution

2. The efficiency expert of a computer company tested 40 engineers to estimate the 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 appoint 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?

3. Define. Write Properties and Application

   (a) Student's t –Distribution      (b) Chi – Square Distribution

4. A population consists of 5 members 2, 3, 6,8,11. Consider all possible samples of size two. Which can be drawn with replacement from the population?  Find (i) Mean of the population (ii) s.d of the population  (iii) Mean of the sampling distribution mean(iv) s.d of the sampling distribution mean

5. The mean of random sample is an unbiased   estimate of the mean of the Population 3, 6, 9, 15, 27. Find (i) list of all possible samples of size 3   that can be taken without replacement from the finite population. (ii) Calculate mean of each of the samples listed in (i) and assigning each sample of probability of 1/10. Which is equal to the mean of the population?

6. Find the 95% confidence limits for the mean of a normality distributed population from which the following sample was taken 15,17,10,18,16,9,7,11,13,14.

7. What is the maximum error one can we expect to make with probability 0.90 when using the mean of a random sample size n= 64 to estimate the mean of the population with variance    is 2.56.

8. Define standard error and state its uses?

9. (a) Explaining Methods of sampling.

   (b) In a study of an automobile insurance a random sample of 80 body repairs Costs had a mean of 472.76 and s. d 62.35. If   mean is used as a point Estimate to the true average repair costs, with what confidence we can assert that the maximum error doesn't exist 10.

## UNIT-V

1. Define (i) Null Hypothesis (ii) Type 1 and Type 2 Errors. (iii) Level of Significance(iv) critical region

2. Write the General  Procedure for Testing of Hypothesis

3. The mean yield of wheat from a district A was 210 pounds with s.d 10 pounds per acre from a sample of 100 plots. In another district the mean yield was 200 pounds with s.d 12 pounds from sample of 150 plots. Assuming that the s.d of yield in the entire state was 11 pounds, test whether there is any significant difference that mean yield of crops in two districts.

4. Two horses A and B were tested according to the time to run a particular track with the  following results

| Horse A | 28 | 30 | 32 | 33 | 33 | 29 | 34 |
|---------|----|----|----|----|----|----|----|
| Horse B | 29 | 30 | 30 | 24 | 27 | 29 | -  |

Test whether two horses have the same running capacity.

5. Scores obtained in a shooting competition by 10 soldiers and after intensive training are given below

| Before | 67 | 24 | 57 | 55 | 63 | 54 | 56 | 68 | 33 | 43 |
|--------|----|----|----|----|----|----|----|----|----|----|
| After  | 70 | 38 | 58 | 58 | 56 | 67 | 68 | 75 | 42 | 38 |

Test whether the intensive training is useful at 0.05 level of significance.

6. Time taken by workers in performing a job by method−I and method-II is given below

| Method-I  | 20 | 16 | 26 | 27 | 23 | 22 | -  |
|-----------|----|----|----|----|----|----|----|
| Method-II | 27 | 33 | 42 | 35 | 32 | 34 | 38 |

Do the data show that the variances of time distribution from population from which these samples are drawn do not differ significantly?

7. Random sample of 400 men and 600 women were asked whether they would like to have a flyover near their residence 200 men and 325women were in flavour of the proposal. Test the hypothesis that proportions of men and women in favour of the proposal are same at 5% level of significance (5%=1.96).

8. on the basis of their total scores,200 candidates of a civil service examination are divided into two groups, the upper 30% and the remaining 70%. Consider the first question of the examination among the first group 40 had correct answer, where as among the second group 80 had the correct answer on the basis of these results can one conclude that the first question is not good at discriminating ability of the type being examined have?

9. The nicotine contents in milligrams in two samples of tobacco were found to as follow

| Sample A | 24 | 27 | 26 | 21 | 25 | -  |
|----------|----|----|----|----|----|----|
| Sample B | 27 | 30 | 28 | 31 | 32 | 36 |

Can it say that the two samples have from the same normal population?

10. The following figures show the distribution of digits in numbers at random From a telephone directory.

| Digits    | 0    | 1    | 2   | 3   | 4    | 5   | 6    | 7   | 8   | 9   |
|-----------|------|------|-----|-----|------|-----|------|-----|-----|-----|
| Frequency | 1026 | 1107 | 997 | 996 | 1075 | 933 | 1107 | 972 | 964 | 853 |

Test whether the digits may be taken to occur equally frequently in the directory.