

Roll No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--



BENGALURU
School of Computer Science and Engineering
Mid - Term Examinations - November 2024

Semester: V

Date: 04-11-2024

Course Code: CSE2027

Time: 02.00pm to 03.30pm

Course Name: Fundamentals of Data Analytics

Max Marks: 50

Program: B. Tech

Weightage: 25%

Instructions:

(i) Read all questions carefully and answer accordingly.

(ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks.

5Qx2M=10M

- | | | | |
|---|---------|----|-----|
| 1. List the Data Analytical tools. | 2 Marks | L1 | CO1 |
| 2. What are the machine generated unstructured data with examples? | 2 Marks | L1 | CO1 |
| 3. Differentiate between Descriptive and Inferential analysis. | 2 Marks | L2 | CO2 |
| 4. Differentiate between random sampling and stratified sampling. Provide examples of scenarios where each method would be appropriate. | 2 Marks | L1 | CO2 |
| 5. Calculate standard error from data and interpret its implications statistically. | 2 Marks | L1 | CO2 |

Part B

Answer ALL Questions. Each question carries 10 marks.

4QX10M=40M

- | | | | |
|--|----------|----|-----|
| 6 Describe briefly about the many V's of data. | 10 Marks | L1 | CO1 |
|--|----------|----|-----|

or

- | | | | |
|---|---------|----|-----|
| 7 7a. Which of the following is the Qualitative and Quantitative variables. And also justify each variable values by the scales of data too. | 5 Marks | L2 | CO1 |
|---|---------|----|-----|

ID	Age	Gender	Height	Blood Group	LDL	Feeling Happy?	Number of Children
1	25	F	1.62	B	150	Agree	0
2	35	F	1.58	0	123	Strongly Agree	1
3	44	M	1.35	A	178	Disagree	3
4	28	F	1.54	AB	205	Disagree	0
5	35	M	1.35	0	229	Indifferent	2

7b. Explain briefly about the sources of data.

5 Marks L1 C01

8 Calculate the Central Tendency measure of Mean, Median, Mode, Variance and Standard Deviation for the following data:

10 Marks L3 C01

Person	X1	X2	X3	X4	X5	X6	X7	X8
Height(in inches)	38	46	54	56	62	64	54	70

or

9 Describe the role of data preprocessing in a data analysis pipeline. Why is it considered a critical step?

10 Marks L1 C01

10 The mean produce of wheat of a sample of 100 fields in 200 lbs. per acre with a standard deviation of 10 lbs. Another samples of 150 fields gives the mean of 220 lbs. with a standard deviation of 12 lbs. Can the two samples be considered to have been taken from the same population whose standard deviation is 11 lbs? Use 5 per cent level of significance. Use the standard table values for test.

10 Marks L3 C02

Standard table values for Z Test:

Level of significance α	Two tailed test	Right tailed test	Left tailed test
90% confidence or $\alpha = 10\% = 0.1$	$z = 1.645$	$z = 1.28$	$z = -1.28$
95% confidence or $\alpha = 5\% = 0.05$	$z = 1.96$	$z = 1.645$	$z = -1.645$
99% confidence or $\alpha = 1\% = 0.01$	$z = 2.58$	$z = 2.33$	$z = -2.33$

or

- 11 The level of calcium in the blood of healthy, young adults varies with a mean of 9.5 mg per deciliter and a SD of 0.4. A clinic in rural Illinois measures the blood calcium level of 180 healthy pregnant women and finds $\bar{x} = 9.57$ mg. Is this an indication that the mean calcium level in this population differs from 9.5mg? Take 5% level of significance. 10 Marks L3 CO2

Standard table values for Z Test:

Level of significance α	Two tailed test	Right tailed test	Left tailed test
90% confidence or $\alpha = 10\% = 0.1$	$z = 1.645$	$z = 1.28$	$z = -1.28$
95% confidence or $\alpha = 5\% = 0.05$	$z = 1.96$	$z = 1.645$	$z = -1.645$
99% confidence or $\alpha = 1\% = 0.01$	$z = 2.58$	$z = 2.33$	$z = -2.33$

- 12 A certain chemical process is said to have produced 15 or less pounds of waste material for every 60 lbs. batch with a corresponding standard deviation of 5 lbs. A random sample of 100 batches gives an average of 16 lbs. of waste per batch. Test at 10 per cent level whether the average quantity of waste per batch has increased. Compute the power of the test for $\mu = 16$ lbs. If we raise the level of significance to 20 per cent, then how the power of the test for $\mu = 16$ lbs. would be affected? 10 Marks L3 CO2

Standard table values for Z Test:

Level of significance α	Two tailed test	Right tailed test	Left tailed test
90% confidence or $\alpha = 10\% = 0.1$	$z = 1.645$	$z = 1.28$	$z = -1.28$
95% confidence or $\alpha = 5\% = 0.05$	$z = 1.96$	$z = 1.645$	$z = -1.645$
99% confidence or $\alpha = 1\% = 0.01$	$z = 2.58$	$z = 2.33$	$z = -2.33$

or

- 13 Tata Motor supplies a rear axle for Postal service mail trucks. These axles must be able to withstand 80000 pounds per square inch in stress tests, but an excessively strong axle raises production costs significantly. Long experience indicates that the standard deviation of the strength of its axles is 4000 pounds per square inch. The 10 Marks L3 CO2

Manufacturer selects a sample of 100 axles from production, tests them, and finds that the mean stress capacity of the sample is 79600 pounds per square inch.

- i. Formulate the null and alternate hypothesis
- ii. If manufacturer uses a significance level of 0.05 in testing, will the axles meet his stress requirements?

Standard table values for Z Test:

Level of significance α	Two tailed test	Right tailed test	Left tailed test
90% confidence or $\alpha = 10\% = 0.1$	$z = 1.645$	$z = 1.28$	$z = -1.28$
95% confidence or $\alpha = 5\% = 0.05$	$z = 1.96$	$z = 1.645$	$z = -1.645$
99% confidence or $\alpha = 1\% = 0.01$	$z = 2.58$	$z = 2.33$	$z = -2.33$