****

**USN No:**

**I Semester M. Tech Internal Assessment -II**

**Computer Science and Engineering**

**Course Title:** **DATA SCIENCE**  **Course Code:** 23CSE5107

**Duration**: 75 Mins. **Date:** 25-02-2025

**Time:**10.00AM-11.15AM **Max Marks:**40

**Instructions to Candidates:**

* Answer all the four Questions
* Each question carries 10 marks

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Marks | BT Level | CO’s | PO’s/  PSO’s |
| Q. 1(a) | What is A/B testing, and how is it used in real-world decision-making? Explain with an example. | 10 | L3 | CO-3 | PO-1 |
|  |  |  |  |  |  |
| Q. 2(a) | Define sample mean, sample variance, and sample moments. Explain their importance in statistical analysis with suitable examples. | 10 | L2 | CO-3 | PO-3 |
|  |  |  |  |  |  |
| Q. 3(a) | A university wants to estimate the average GPA of its students. They take a sample of 200 students from a total of 10,000 students.  Explain how the sample mean and standard error can be used to estimate the true mean GPA of the entire student population. | 10 | L3 | CO-3 | PO-4 |
|  |  |  |  |  |  |
| Q. 4(a) | A national bank wants to analyze the average daily withdrawal amount from ATMs across the country. The bank has 10,000 ATMs, but it is impractical to collect data from all of them.  **Scenario:**  •The bank randomly selects 100 ATMs and records the average daily withdrawal amounts for each.  •The true population distribution of daily withdrawals is unknown and may not be normal.  •The sample means are calculated for these 100 ATMs.  **Questions:**  1.Explain how the Central Limit Theorem (CLT) justifies using the sample mean to estimate the population mean, even if the withdrawal amounts are not normally distributed.  2.If the standard deviation of daily withdrawals is known to be ₹1,500, what is the expected standard deviation of the sample mean for a sample size of 100 ATMs?  3.Suppose the sample mean withdrawal amount is ₹5,000. Construct a 95% confidence interval for the true population mean using CLT.  4.If the bank decides to increase the sample size from 100 to 400 ATMs, how will this affect the sampling distribution of the mean? Explain in terms of standard error and accuracy.  5.Discuss a real-world implication of using CLT in this study. Why is it useful for the bank’s decision-making process? | 10 | L3 | CO-4 | PO-4 |

|  |  |
| --- | --- |
| Course Outcomes Being Assessed | |
| CO3 | **Utilize** mathematical and statistical techniques to test hypothesis and to identify covariance with A/B testing and Analysis of Variance. |
| CO4 | **Employ** central limit theorem and confidence interval enabling them to model real-world phenomena and make accurate predictions. |