**RV College of Engineering**

Academic Year 2023-2024 (Odd Semester)

USN | 1 | R | V | 2 | 2 | A | J 2 0 7

# Department of Artificial Intelligence and Machine Learning

Course Code: AI235AT

Marks: 50 Marks

Date : 21-03-2024

Duration: 90 Mins

## CIE 3

## Statistics for Data Science

| SL. No | Questions | M | BT | CO |
|---|---|---|---|---|
| 1 | A cable company plans to survey potential customers in a small city currently served by satellite dishes. Two sampling methods are being considered. Method A is to randomly select a sample of 25 city blocks and survey every family living on those blocks. Method B is to randomly select a sample of families from each of the five natural neighborhoods making up the city. <br> (i) Which among the methods A and B correspond to cluster sampling and why? <br> (ii) Which among the two methods A and B correspond to stratified sampling and why? <br> **Your reason must not be more than a sentence or two each for each subdivision.** | 12 | 3 | 1 |
| 2 | (i) For a one-sided test, what should the alternate hypothesis be? <br> (ii) Last year, the mean monthly rent for an apartment in a certain city was Rs. 15,000. A real estate agent believes that the mean rent is higher this year. State the appropriate null and alternative hypotheses. <br> (iii) The mean caffeine content per cup of regular coffee served at a certain coffee shop is supposed to be 100 milligrams. A test is made of H0: $\mu = 100$ versus H1: $\mu \neq 100$. The null hypothesis is rejected. State an appropriate conclusion. | 12 | 3 | 3 |
| 3 | The mean number of pages that can be printed by a sample of 25 printers is 1550 pages with a standard deviation of 120 pages. The company manufacturing the printers claims that the average number of pages printed by their printer is 1600 pages. Is the claim acceptable at 5% level of significance? | 8 | 3 | 3 |
| 4 | In the past the average length of an outgoing telephone call from a business office has been 143 seconds. A manager wishes to check whether that average has decreased after the introduction of policy changes. A sample of 100 telephone calls produced a mean of 133 seconds, with a standard deviation of 35 seconds. Perform the relevant test at the 1% level of significance. | 8 | 3 | 3 |
| 5 | Which of the following statements is **TRUE** about the null and alternative hypothesis? Support your choice with a logical reason of *not more than 2 sentences.* <br> **Statement 1:** *Null and Alternative hypotheses are statements about sample statistics* <br> **Statement 2:** *Null and Alternative hypotheses are statements about population parameters.* | 5 | 2 | 1 |
| 6 | The mean test score for a simple random sample of n = 100 students was xbar = 67.30. The population standard deviation of test scores is $\sigma = 15$. Construct a 98% confidence interval for the population mean test score $\mu$. <br> xbar is the sample mean, denoted by $\overline{x}$ | 5 | 3 | 3 |

| Course Outcomes: After completing the course, the students will be able to | |
|---|---|
| CO1 | Apply the knowledge of statistics in providing solutions to some common real-life and business problems. |
| CO2 | Visualize data better, make logical inferences about the data in real-world scientific/business use cases, and present the analysis results. |
| CO3 | Make inferences about a population from samples through various statistical techniques. |
| CO4 | Use statistical tools to illustrate the principles of data distribution, data sampling, and data visualization. |
| CO5 | Appraise the knowledge of statistics in data science to build a successful career as an AI&ML engineer, work in teams, and communicate their ideas effectively. |

M-Marks, BT-Blooms Taxonomy Levels, CO-Course Outcomes

| | Particulars | CO1 | CO2 | CO3 | CO4 | CO5 | L1 | L2 | L3 | L4 | L5 | L6 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Marks Distribution | Max Marks CIE | 17 | - | 33 | - | - | - | 5 | 45 | | | |