

Amrita Vishwa Vidyapeetham  
Amrita School of Computing, Coimbatore  
Lab Evaluation -2, Fifth Semester  
Computer Science and Engineering  
19CSE304 Foundations of Data Science

**Course Outcomes (COs):**

CO	Course Outcomes
CO01	Understand the statistical foundations of data science.
CO02	Apply pre-processing techniques over raw data so as to enable further analysis.
CO03	Conduct exploratory data analysis and create insightful visualizations to identify patterns.
CO04	Identify machine learning algorithms for prediction/classification and to derive insights
CO05	Analyse the degree of certainty of predictions using statistical test and models

Duration: One hour

Maximum: 20 Marks

**Important Instructions:**

1. There are 2 parts: Part A(17marks) and Part B(3marks). Submit solutions for both parts in a single ipynb notebook.
2. Assume:
  - (a) Number of Purple flowering plants should be (705 plus “last 2 digits of your Roll Number”). E.g., Roll No 20135 should take  $705 + 35 = 740$
  - (b) Number of Simulations should be (2000 minus “last 3 digits of your Roll Number”).

**Part A (17 marks):**

- **Narrative:** Gregor Mendel (1822-1884) was an Austrian monk widely recognized as the founder of the modern field of genetics. Many of his experiments were on varieties of pea plants. *He formulated sets of assumptions about each variety; these were his models. He then tested the validity of his models by growing the plants and gathering data.* You are required to analyze the data from one such experiment to see if Mendel's model was good. In a particular variety, each plant has either purple flowers or white. The color in each plant is unaffected by the colors in other plants.
- **Mendel's hypothesis:** plants should bear purple or white flowers at random, in the ratio 3:1. i.e., For every plant, there is a 75% chance that it will have purple flowers, and a 25% chance that the flowers will be white, regardless of the colors in all the other plants.
- **Mendel's experiment:** Mendel recorded the number of purple and white flowering plants. Among the 929 plants that he grew, assume (705 + “last 2 digits of your Roll Number”), were purple flowering.
- **Hypothesis Test:** Carry out a hypothesis test to ascertain Mendel's model.

**Specific Requirements:**

1. State the Null and Alternative Hypothesis (2).
2. What statistic do you recommend to apply for this case(2)?
3. State the observed statistic(2)
4. Simulate the test statistic under the Null that will help make the decision(5). The number of simulations should be equal to (2000 - “Last 3 digits of your Roll Number”).
5. Draw a histogram to depict what the model of random selection predicts about the statistic.(4). Depict the observed statistic as a red dot on the histogram(1).
6. What is the outcome of your analysis? (1)

**Part B(3 marks)**

Carry out a hypothesis test for “mean” of a population. The population standard deviation is **13.46**, and the test is for a “mean” different from **60**. Given: a sample size of **30**, with “sample mean” **65.10**. Consider  $\alpha = 0.05$ . Demonstrate your solution using either the critical-value or p-value method.

-----