[EID-758]
M.Tech. Degree Examination

Computer Science & Technology
I SEMESTER

STATISTICS FOR DATA SCIENCE
(Effective from the admitted batch 2017–18)

Time: 3 Hours
Max.Marks: 60

Instructions: Each Module carries 12 marks.
Answer all modules choosing one question from each module
All parts of the module must be answered in one place only.
Figures in the right hand margin indicate marks allotted.

MODULE-I

1. a) Explain the concept of conditional probability 6
b) A company produces 1,000 refrigerators a week at three plants. Plant A produces 350 refrigerators a week, plant B produces 250 refrigerators a week, and plant C produces 400 refrigerators a week. Production records indicate that 5% of the refrigerators at plant A will be defective, 3% of those produced at plant B will be defective, and 7% of those produced at plant C will be defective. All refrigerators are shipped to a central warehouse. If a refrigerator at the warehouse is found to be defective, what is the probability that it was produced at plant A? 6

OR

2. a) Explain the importance of the central limit theorem 6
b) The achievement scores for a college entrance examination are normally distributed with mean 75 and standard deviation 10. What fraction of the scores lies between 80 and 90? 6
MODULE-II

3. A firm wishes to compare four programs for training workers to perform a certain manual task. Twenty new employees are randomly assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee, with the following results:

- Program 1: 9, 12, 14, 11, 13
- Program 2: 10, 6, 9, 9, 10
- Program 3: 12, 14, 11, 13, 11
- Program 4: 9, 8, 11, 7, 8

Using $\alpha = .05$, determine whether the treatments differ in their effectiveness.

OR

4. a) Explain the level of significance and power of the test
   b) Discuss the various steps involved in the testing of hypothesis problem

MODULE-III

5. a) Distinguish between correlation and regression
   b) The data below shows the number of absences, ($X$) in Statistics course and the final exam grade, ($Y$), for 7 students. Find the correlation coefficient and interpret your result.

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>0</th>
<th>2</th>
<th>6</th>
<th>4</th>
<th>3</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>95</td>
<td>90</td>
<td>90</td>
<td>55</td>
<td>70</td>
<td>80</td>
<td>85</td>
</tr>
</tbody>
</table>

OR

6. What is a discriminant analysis? Explain the linear discriminant function

MODULE- IV

7. What do you mean by cross validation in regression analysis?
   Discuss its advantages and limitations

OR

8. Explain the importance of resampling methods in regression analysis with suitable examples
MODULE- V

9. What do you mean by subset regression models? Discuss the stepwise regression procedure.

OR

10. Explain the dimensional reduction procedures in regression analysis.

[2/IS/217]