# STATISTICAL METHODOLOGIES IN DATA SCIENCE

| | | | | | |
|---|---|---|---|---|---|
| **Programme** | **B.Sc. Statistics** | | | | |
| **Course Code** | **STA1MN103 (P)** | | | | |
| **Course Title** | **Introductory Statistics with R** | | | | |
| **Type of Course** | **Minor** | | | | |
| **Semester** | **I** | | | | |
| **Academic** | **100-109** | | | | |
| **Level** | | | | | |
| **Course Details** | **Credit** | **Lecture per week** | **Tutorial per week** | **Practical per week** | **Total Hours** |
| | **4** | **3** | **-** | **2** | **75** |
| **Pre-requisites** | Basic knowledge about data, basic mathematical knowledge | | | | |
| **Course Summary** | This course covers data types, distributions, graphs, and statistical measures using R programming. Students learn to analyze data effectively for informed decision-making across diverse domains. | | | | |

**Course Outcomes (CO):**

| CO | CO Statement | Cognitive Level* | Knowledge Category# | Evaluation Tools used |
|---|---|---|---|---|
| CO1 | Identify data types and construct frequency distributions. | Ap | C | Instructor-created exams / Quiz |
| CO2 | Create diverse graphical representations effectively and critically evaluate ethical implications of statistical methods aligning with human values. | Ap | F | Practical Assignment / Observation of Practical Skills/ Instructor-created exams |
| CO3 | Use measures of central tendency to summarize and describe data, demonstrating the ability to communicate the findings in both written and graphical formats. Between different data sets | Ap | C | Seminar Presentation / Group Tutorial Work/ Instructor-created exams |
| CO4 | Master R programming basics and descriptive statistics. | Ap | F | Instructor-created exams / Home Assignments |

| | | | | |
|---|---|---|---|---|
| CO5 | Implement R for practical data analysis and graphical representation. | Ap | P | Viva Voce/ Instructor-created exams |

* - Remember (R), Understand (U), Apply (Ap), Analyse (An), Evaluate (E), Create (C)
# - Factual Knowledge(F) Conceptual Knowledge (C) Procedural Knowledge (P) Metacognitive
Knowledge (M)

## Detailed Syllabus:

| Module | Unit | Content | Hrs (45+30) | Marks |
|---|---|---|---|---|
| **I** | **Data** | | **12** | **15** |
| | 1 | Types of data: Primary data, Secondary data, Quantitative data, Qualitative data, discrete data, continuous data | 4 | |
| | 2 | Frequency distribution: Ungrouped and grouped | 4 | |
| | 3 | Cumulative frequency distribution | 4 | |
| | **Sections from References:**<br>Unit 1:2.2,11.1,2.1 Ref[1]<br>Unit 2: 2.2 Ref[1]<br>Unit 3: 3.5 Ref[3] | | | |
| **II** | **Graphical Representation of Data** | | **9** | **15** |
| | 4 | Line diagram, Bar diagram | 3 | |
| | 5 | Pictogram, Pie diagram, Histogram | 3 | |
| | 6 | Frequency Polygon, Frequency curve, Ogives. | 3 | |
| | **Sections from References:**<br>Unit 4: 4.3.3 Ref[3]<br>Unit 5:4.3.4, 4.3.6 Ref[3]<br>Unit 6: 4.4.3 Ref[3] | | | |
| **III** | **Measures of Central Tendency** | | **10** | **25** |
| | 7 | Arithmetic Mean | 2 | |
| | 8 | Median | 2 | |
| | 9 | Mode | 2 | |
| | 10 | Geometric Mean | 2 | |
| | 11 | Harmonic Mean | 2 | |
| | **Sections from References:**<br>Unit 7: 5.4 Ref[3]<br>Unit 8: 5.6.1 Ref[3]<br>Unit 9: 5.7.1 Ref[3]<br>Unit 10: 5.9 Ref[3]<br>Unit 11: 5.10 Ref[3] | | | |
| **IV** | **Introduction to R Programming** | | **14** | **15** |
| | 12 | Installing R | 1 | |
| | 13 | Objects in R | 1 | |
| | 14 | Using functions in R | 1 | |

| | | | | |
|---|---|---|---|---|
| | 15 | Importing data | 1 | |
| | 16 | Exporting data | 1 | |
| | 17 | Simple base R plots | 2 | |
| | 18 | Multiple graphs | 2 | |
| | 19 | R Packages | 1 | |
| | 20 | Exporting plots | 2 | |
| | 21 | Getting help | 1 | |
| | 22 | Saving stuff in R | 1 | |
| | **Sections from References:**<br>Unit 12: 1.1 Ref[2]<br>Unit 13: 2.2 Ref[2]<br>Unit 14: 2.3 Ref[2]<br>Unit 15: 3.3 Ref[2]<br>Unit 16: 3.6 Ref[2]<br>Unit 17: 4.2 Ref[2]<br>Unit 18: 4.4 Ref[2]<br>Unit 19: 1.5 Ref[2]<br>Unit 20: 4.5 Ref [2]<br>Unit 21: 2.5 Ref[2]<br>Unit 22: 2.6 Ref[2] | | | |
| **V** | **PRACTICUM** | | **30** | |
| | Do practice problems in R software from any 5 units of the given list and one additional problem decided by the teacher-in-charge, related to the content of the course. Other units listed here may be used as demonstrations of the<br>concepts taught in the course. | | | |
| | 1 | Functions in R— data.frame | | |
| | 2 | multiply_columns() | | |
| | 3 | return() | | |
| | 4 | identical() | | |
| | 5 | Conditional statements-if and else | | |
| | 6 | Combining logical operators | | |
| | 7 | For loop | | |
| | 8 | While loop | | |
| | **Sections from References:**<br>Unit 1: 7.2 Ref[2]<br>Unit 2: 7.2Ref[2]<br>Unit 3: 7.2Ref[2]<br>Unit 4: 7.2Ref[2]<br>Unit 5: 7.3Ref[2]<br>Unit 6: 7.4 Ref[2]<br>Unit 7: 7.5.1 Ref[2]<br>Unit 8: 7.5.2 Ref[2] | | | |

**Books and References:**

1. Gupta, S.C. and Kapoor, V.K. (1997) Fundamentals of Mathematical Statistics. Sultan Chand and
Sons, New Delhi

2. Douglas, Alex, Deon Roos, Francesca Mancini, Ana Couto, and David Lusseau. (2020), *An Introduction to R*. https://intro2r.com/index.html.
3. S.C. Gupta, Fundamentals of Statistics, Himalaya Publishing house,

**Mapping of COs with PSOs and POs :**

| | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 |
|---|---|---|---|---|---|---|---|
| CO1 | 2 | 2 | 0 | 0 | 0 | 0 | 0 |
| CO2 | 0 | 3 | 1 | 0 | 3 | 0 | 0 |
| CO3 | 3 | 1 | 0 | 0 | 0 | 0 | 0 |
| CO4 | 2 | 0 | 2 | 3 | 0 | 0 | 1 |
| CO5 | 1 | 0 | 1 | 3 | 0 | 0 | 1 |

## Correlation Levels:

| Level | Correlation |
|---|---|
| - | Nil |
| 1 | Slightly / Low |
| 2 | Moderate/Medium |
| 3 | Substantial / High |

## Assessment Rubrics:

- Quiz / Assignment/ Quiz/ Discussion / Seminar
- Midterm Exam
- Programming Assignments (20%)
- Final Exam (70%)

## Mapping of COs to Assessment Rubrics:

| | **Internal Exam** | **Assignment** | **Project Evaluation** | **End Semester Examinations** |
|---|---|---|---|---|
| **CO1** | ✓ | ✓ | | ✓ |
| **CO2** | ✓ | ✓ | | ✓ |
| **CO3** | ✓ | ✓ | | ✓ |
| **CO4** | ✓ | ✓ | | ✓ |
| **CO5** | ✓ | | | |