

B.Sc. (Data Science)
Discipline Specific Elective (DSE)
Semester II
BSDB31202T: Statistical Foundation

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 40%

Objective: This course will enable students to understand the fundamentals of statistics to apply descriptive measures and probability for data analysis. Students will be able to infer the concept of correlation and regression for relating two or more related variables and probabilities for various events.

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER

1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES

Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

Section A

Unit I: Origin and development of Statistics, Scope, limitation and misuse of statistics. Types of data: primary, secondary, quantitative and qualitative data. Types of Measurements: nominal, ordinal, discrete and continuous data.

Unit II: Presentation of data by tables: construction of frequency distributions for discrete and continuous data, graphical representation of a frequency distribution by histogram and frequency polygon, cumulative frequency distributions. Classification and Graphical representation of data (Pie Chart, Bar Diagram, Histogram, Frequency Polygon, Ogive Curve, etc.).

Unit III: Measures of Central Tendency – Arithmetic Mean, Median and Mode and its Graphical representation, Measures of dispersion – range, variance, mean deviation, standard deviation and coeff. of variation, Concepts and Measures of Skewness and Kurtosis.

Unit IV: Descriptive Statistics, Exploratory data analysis, Coefficient of variation, Data **visualization**, Scatter diagram, Grouped data,

Section B

Unit V: Correlation: Scatter plot, Karl Pearson coefficient of correlation, Spearman's rank correlation coefficient, multiple and partial correlations (for 3 variates only). Regression: Introduction to regression analysis: Modelling a response, overview and applications of regression analysis, Simple linear regression (Two variables)

Unit VI: Mathematical and Statistical probability, Elementary events, Sample space, Compound events, Types of events, Random experiment, sample point and sample space, event, algebra of events.

Unit VII: Definition of Probability: classical, empirical and axiomatic approaches to probability, properties of probability. Theorems on probability, conditional probability and independent events

Unit VIII: Statistical inference, Concept of Random Variable, Probability Mass Function & Density Function, Mathematical Expectation (meaning and properties), Moments, Moment Generating Function and Characteristic Function

Suggested Readings

1. Gupta, S.C. and Kapoor, V.K.: Fundamentals of Mathematical Statistics, Sultan & Chand & Sons, New Delhi, 11th Ed,
2. Hastie, Trevor, et al. The elements of Statistical Learning, Springer,
3. Ross, S.M., Introduction to Probability and Statistics, Academic Foundation,
4. Papoulis, A. and Pillai, S.U. , Probability, Random Variables and Stochastic Processes, TMH,