COURSE 4: DESCRIPTIVE STATISTICS

Theory   Credits: 3   3 hrs/week

Course Learning Outcomes: Students will acquire:
- knowledge of Statistics and its implementation through practical understanding for various domains related to data science.

- knowledge of various types of data, their organization and evaluation of summary measures such as measures of central tendency and dispersion etc.

- knowledge of other types of data reflecting quality characteristics including concepts of independence and association between two attributes,

- insights into preliminary exploration of different types of data.

- Knowledge of correlation, regression analysis, regression diagnostics, partial and multiple correlations.

UNIT I:
Introduction to Statistics: Importance of Statistics. Scope of Statistics in different fields. Concepts of primary and secondary data. Diagrammatic and graphical representation of data: Histogram, frequency polygon, Pie. Measures of Central Tendency: Mean, Median, Mode, Geometric Mean and Harmonic Mean. Median and Mode through graph.

UNIT II:
Measures of Dispersion: Range, Quartile Deviation, Mean Deviation and Standard Deviation, Variance. Central and Non-Central moments and their interrelationship. Skewness and kurtosis.

UNIT III:
Curve fitting: Bi- variate data, Principle of least squares, fitting of degree polynomial. Fitting of straight line, Fitting of Second degree polynomial or parabola, Fitting of power curve and exponential curves.

Correlation: Meaning, Types of Correlation, Measures of Correlation: Scatter diagram, Karl Pearson's Coefficient of Correlation, Rank Correlation Coefficient (with and without ties), Bi-variate frequency distribution, correlation coefficient for bi-variate data and simple problems. Concept of multiple and partial correlation coefficients (three variables only) and properties

UNIT IV:
Regression: Concept of Regression, Linear Regression: Regression lines, Regressioncoefficients and it's properties, Regressions lines for bi-variate data and simple problems. Correlation vs regression, sigmoid curve, derivation from linear regression to logistic regression.

UNIT-V

Attributes : Notations, Class, Order of class frequencies, Ultimate class frequencies, Consistency of data, Conditions for consistency of data for 2 and 3 attributes only , Independence of attributes , Association of attributes and its measures, Relationship between association and colligation of attributes, Contingency table: Square contingency, Mean square contingency, Coefficient of mean square contingency,