DATA ANALYTICS

Paper Code: ETCE-419	**L**	**T/P**	**C**
Paper: Data Analytics	**3**	**1**	**4**

INSTRUCTIONS TO PAPER SETTERS: **MAXIMUM MARKS: 75**

1. Question No. 1 should be compulsory and cover the entire syllabus. This question should have objective or short answer type questions. It should be of 25 marks.
2. Apart from Question No. 1, rest of the paper shall consist of four units as per the syllabus. Every unit should have two questions. However, student may be asked to attempt only 1 question from each unit. Each question should be of 12.5 marks

Objective*: This course is aimed at providing in-depth understanding of data analysis based on statistical techniques. The approach to data analysis involves exploratory methods, continuous distributions such as normal, lognormal distribution, probability plotting for normal distributions, hypothesis testing etc. The subject deals with model estimation and testing using parametric and non-parametric methods, identification and accommodation of outliers, frequency analysis of extreme events like flood, storms, droughts etc and use of simulation techniques such as monte-carlo simulation.*

UNIT – I

Preliminary Data Analysis: Graphical representation-line diagram or Bar Chart, Dot diagram, Histogram, Exploratory methods- stem and leaf plot, Box plot. Random events- sample space and events, the null event, Intersection and Union, Venn Diagram and Event space. Continuous Distributions- Normal Distribution, Lognormal Distribution, Bivariate Normal Distribution.

[T1][No. of Hrs. 10]

UNIT – II

Model Estimation and Testing: Properties of Estimators- Unbiasedness, Consistency, Minimum Variance, Efficiency, Sufficiency. Estimation of Confidence Intervals. Hypothesis testing- Procedure for testing, Probabilities of Type I and Type II Errors and the power function, Tests of Hypothesis involving the Variance, the F Distribution and its use. Nonparametric methods- Wilcox on Signed- Rank Test for Association of Paired Observations.

[T2][No. of Hrs. 10]

UNIT – III

Goodness of Fit Tests: Chi-squared Goodness of Fit test, Kolmogorov- Smirnov Goodness of Fit test, Kolmogorov- Smirnov Two- sample test, Anderson- Darling Goodness of Fit test, Other methods for testing the Goodness of Fit to a Normal Distribution.

Analysis of Variance: One-Way Analysis of Variance, Two-way analysis of Variance.

Probability Plotting for Normal Distribution, Probability Plotting for Type I Extreme Value Distribution.

Identification and Accommodation of Outliers: Hypothesis Tests, Test Statistics for Detection of Outliers, Dealing with Non-normal Data.

Estimation of Probabilities of Extreme events when outliers are present. Multivariate Analysis- Principle Components Analysis, Factor Analysis, Cluster analysis.

Spatial Correlation: The Estimation problem, Spatial Correlation and the Semivariogram, some Semivariogram Models and Physical Aspects, Spatial Interpolations and Kriging.

[T1,T2] [No. of Hrs. 12]

UNIT – IV

Frequency Analysis of Extreme Events: Order Statistics- Functions of Order Statistics, Expected value and Variance of Order Statistics, Expected Value and Variance of Order Statistics. Extreme Value Distributions- Basic Concepts, Gumbel Distribution, Weibull Distribution as an Extreme Value Model, General Extreme Value Distribution. Analysis of Natural Hazards: Floods, storms and Droughts, Earthquakes and volcanic eruptions, winds, sea levels and Highest sea waves.

Simulation techniques for Design: MonteCarlo Simulation- Statistical Experiments, Probability Integral Tranform, Sample size and accuracy of Monte Carlo Experiments.

[T1,T2] [No. of Hrs. 12]

Text Books:

[T1] Kottegoda N.T. and Rosso R., "Probability, Statistics and Reliability for Civil and Environmental Engineers", McGraw Hill, USA.

[T2] Azzalini A., Scarpa B., "Data Analysis and Data Mining- An Introduction", Oxford University Press, New York.

Reference Books:

[R1] Stokes M.E., Davis C.S., Koch G.G., "Categorical Data Analysis Using the SAS System", SAS Publishing, North Carolina.

[R2] Ruppert D., "Statistics and Data Analysis for Financial Engineering", Springer, New York.