

| Subject Code | Subject Name (Lab oriented Theory Courses) | Category | L | T | P | C |
|--------------|--|----------|---|---|---|---|
| CB19342 | COMPUTATIONAL STATISTICS | BS | 3 | 0 | 2 | 4 |

| | |
|--------------------|--|
| Objectives: | |
| ● | To study the mean, variance, linear regression models and error term for use in Multivariate data analysis. |
| ● | To understand the relationship of the data collected for decision making. |
| ● | To know the concept of principal components, factor analysis and cluster analysis for profiling and interpreting the data collected. |

| | | |
|---|---|-------------|
| UNIT-I | MULTIVARIATE NORMAL DISTRIBUTION | 9 |
| Multivariate Normal Distribution Functions, Conditional Distribution and its relation to regression model, Estimation of parameters. | | |
| UNIT-II | DISCRIMINANT ANALYSIS | 9 |
| Statistical background, linear discriminant function analysis, Estimating linear discriminant functions and their properties. | | |
| UNIT-III | PRINCIPAL COMPONENT ANALYSIS | 9 |
| Principal components, Algorithm for conducting principal component analysis, deciding on how many principal components to retain, H-plot. | | |
| UNIT-IV | FACTOR ANALYSIS | 9 |
| Factor analysis model, Extracting common factors, determining number of factors, Transformation of factor analysis solutions, Factor scores. | | |
| UNIT-V | CLUSTER ANALYSIS | 9 |
| Introduction, Types of clustering, Correlations and distances, clustering by partitioning methods, hierarchical clustering, overlapping clustering, K-Means Clustering-Profiling and Interpreting Clusters. | | |
| Contact Hours | | : 45 |

| | | | |
|----------------------------|---|----------------------------|-------------|
| List of Experiments | | | |
| 1. | Python Concepts, Data Structures, Classes: Interpreter, Program Execution, Statements, Expressions, Flow Controls, Functions, Numeric Types, Sequences and Class Definition, Constructors, Text & Binary Files – Reading and Writing | | |
| 2. | Visualization in Python: Matplotlib package, Plotting Graphs, Controlling Graph, Adding Text, More Graph Types, Getting and setting values, Patches. | | |
| 3. | Multivariate data analysis: Multiple regression, multivariate regression, cluster analysis with various algorithms, factor analysis, PCA and linear discriminant analysis. Various datasets should be used for each topic. | | |
| | | Contact Hours | : 30 |
| | | Total Contact Hours | : 75 |

| | |
|---|--|
| Course Outcomes: | |
| On completion of the course, the students will be able to | |
| ● | Analyze means and variances of the individual variables in a multivariate set and also the correlations between those variables. |
| ● | To find discriminants, rules to optimally assign new objects to the labelled classes. |
| ● | Apply the principal component techniques to reduce data and to interpret. |
| ● | To reduce the number of variables in regression models using Factor analysis |
| ● | Apply the techniques of clustering methods for massive amounts of data. |

| Text Books: | |
|--------------------|--|
| 1 | T.W. Anderson.”An Introduction to Multivariate Statistical Analysis”. Wiley, Third edition, 2003 |
| 2 | J.D. Jobson,”Applied Multivariate Data Analysis”, Volume I & II, Springer texts in statistics, New York, Fourth Edition 1999. |
| 3 | Python 3 for Absolute Beginners, Tim Hall and J-P Stacey. Beginning Python: From Novice to Professional, Magnus Lie Hetland. Edition, 2005. |
| 4 | Mark Lutz.,”ProgrammingPython”O'Reilly Media ,Germany, Fourth edition, 2011. |

| Reference Books / Web links: | |
|-------------------------------------|---|
| 1 | D.A. Belsey, E. Kuh and R.E. Welsch ,”Regression Diagnostics , Identifying Influential Data and Sources of Collinearity” |
| 2 | Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, ”Introduction to Linear Regression Analysis”, Fifth Edition, Wiley, 2012. |
| 3 | Johnson R.A. & Wichern, D.W, “Applied Multivariate Statistical Analysis “, Sixth Edition, Pearson, 2018. |
| 4 | Magnus Lie Hetland, “Beginning Python: From Novice to Professional”, Third Edition, Apress, 2005. |
| 5 | M.R. Anderberg, “Cluster Analysis for Applications”, Academic Press. |

CO - PO – PSO matrices of course

| PO/PSO | PO 1 | PO 2 | PO 3 | PO 4 | PO 5 | PO 6 | PO 7 | PO 8 | PO 9 | PO 10 | PO 11 | PO 12 | PSO 1 | PSO 2 | PSO 3 |
|------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|
| CO | | | | | | | | | | | | | | | |
| CB19342.1 | 3 | 2 | 3 | 3 | 2 | 1 | - | - | - | 2 | 2 | - | 2 | 2 | 2 |
| CB19342.2 | 3 | 3 | 2 | 3 | 2 | 1 | - | - | - | 1 | 2 | - | 2 | 2 | 2 |
| CB19342.3 | 3 | 3 | 2 | 3 | 3 | 1 | - | - | - | 1 | 2 | - | 2 | 3 | 2 |
| CB19342.4 | 3 | 3 | 2 | 3 | 3 | 2 | - | - | - | 2 | 2 | - | 2 | 3 | 3 |
| CB19342.5 | 3 | 3 | 2 | 3 | 3 | 2 | - | - | - | 2 | 2 | - | 2 | 3 | 3 |
| Average | 3.0 | 2.8 | 2.2 | 3.0 | 2.6 | 1.4 | - | - | - | 1.6 | 2.0 | - | 2.0 | 2.6 | 2.4 |

Correlation levels 1, 2 or 3 are as defined below:

1: Slight (Low) 2: Moderate (Medium) 3: Substantial (High)

No correlation: “-”