SVKM's NMIMS
Mukesh Patel School of Technology Management & Engineering

Program: B. Tech. Computer Science & Business Systems		Semester: III
Course/Module:	Computational Statistics	**Module Code:** BTCS03004

Teaching Scheme				Evaluation Scheme	
Lecture (Hours per week)	**Practical (Hours per week)**	**Tutorial (Hours per week)**	**Credit**	**Internal Continuous Assessment (ICA) (Marks - 50)**	**Term End Examinations (TEE) (Marks- 100 in Question Paper)**
3	2	0	4	Marks Scaled to 50	Marks Scaled to 50

Pre-requisite: Statistical Modelling BTCS01002

Objectives:
1. To develop sound knowledge and skills in theoretical, computational and application-oriented statistics.
2. To equip the students with intermediate to advanced level concepts and tools in statistics that help them tackle relevant problems within engineering domain.

Outcomes: After completion of the course, students would be able to:
1. Solve problems involving multivariate normal distribution, multiple and multivariate regression.
2. Demonstrate the ability to understand the concepts of Discriminant analysis, PCA, Factor analysis and Cluster analysis.
3. Analyze data samples using Python programming.

Detailed Syllabus:

Unit	Description	Duration
1	**Multivariate Normal Distribution:** Multivariate Normal Distribution Functions, Conditional Distribution and its relation to regression model, Estimation of parameters.	**7**
2	**Multiple Linear Regression Model:** Standard multiple regression models with emphasis on detection of collinearity, outliers, non-normality and autocorrelation, Validation of model assumptions.	7
3	**Multivariate Regression:** Assumptions of Multivariate Regression Models, Parameter estimation, Multivariate Analysis of variance and covariance	7
4	**Discriminant Analysis:** Statistical background, linear discriminant function analysis, Estimating linear discriminant functions and their properties.	5

_____

Signature

(Prepared by Concerned Faculty/HOD)

5	**Principal Component Analysis:** Principal components, Algorithm for conducting principal component analysis, deciding on how many principal components to retain, H-plot.	4
6	**Factor Analysis:** Factor analysis model, Extracting common factors, determining number of factors, Transformation of factor analysis solutions, Factor scores.	5
7	**Cluster Analysis:** Introduction, Types of clustering, Correlations and distances, clustering by partitioning methods, hierarchical clustering, overlapping clustering, K-Means Clustering-Profiling and Interpreting Clusters.	5
8	**Python Concepts, Data Structures, Classes:** Interpreter, Program Execution, Statements, Expressions, Numeric Types, Flow Controls, Functions, Sequences and Class Definition, Constructors, Text & Binary Files - Reading and Writing **Visualization in Python:** Matplotlib package, Plotting Graphs, Controlling Graph, Adding Text, More Graph Types, Getting and setting values, Patches (To be done in Lab session.) **Multivariate data analysis**: Multiple regression, multi variate regression, cluster analysis with various algorithms, factor analysis, PCA and linear discriminant analysis. (To be done in Lab session.)	**5**
	Total	**45**

Text Books:

1. **T.W. Anderson**, "An Introduction to Multivariate Statistical Analysis",3rd Edition, 2003.
2. **J.D. Jobson,** "Applied Multivariate Data Analysis", Vol I & II, Springer Publication,1st Edition, 1999.
3. **Mark Lutz,** "Programming Python", O'Reilly, 4th Edition,2010.
4. **Tim Hall and J-P Stacey**, "Python 3 for Absolute Beginners", Apress, 1st Edition, 2009.
5. Magnus Lie Hetland, "Beginning Python: From Novice to Professional", Edition, 2005.

Reference Books:

1. D.A. Belsey, E. Kuh and R.E. Welsch," Regression Diagnostics, Identifying Influential Data and Sources of Collinearety" John Wiley & Sons, 1989.
2. J. Neter, W. Wasserman and M.H. Kutner," Applied Linear Regression Models",Mc Graw Hill Education, 4th Edition, 2004.
3. D.C. Montgomery &E. Peck, "Introduction to Linear Regression Analysis", John Wiley and Sons, 5th Edition, 2012
4. M.R. Anderberg," Cluster Analysis for Applications, Springer NY,1st Edition,2000.
5. D.F. Morrison," Multivariate Statistical Methods", Thomson/Brooks/Cole, 2005.

_____
Signature
(Prepared by Concerned Faculty/HOD)

6. Wes Mc Kinney," Python for Data Analysis", O'Reilly, 2nd Edition, 2017.

Any other information:
Total Marks of Internal Continuous Assessment (ICA): 50 Marks
Distribution of ICA Marks:

Description of ICA	Marks
Class Test 1	10
Class Test 2	10
Term Work	30
Total Marks :	**50**

_____
Signature
(Prepared by Concerned Faculty/HOD)