Shri Vile Parle Kelavani Mandal's

MITHIBAI COLLEGE OF ARTS, CHAUHAN INSTITUTE OF SCIENCE & AMRUTBEN JIVANLAL COLLEGE OF COMMERCE AND ECONOMICS (AUTONOMOUS)

NAAC Reaccredited 'A' grade. CGPA: 3.57,
Granted under RUSA, FIST-DST & Star College Scheme of DBT, Government of India,
Best College (2016-17), University of Mumbai

Affiliated to the
UNIVERSITY OF MUMBAI

Program: Master of Science

Statistics

M. Sc. PART II

Semester III & IV

Choice Based Credit System (CBCS) with effect from the Academic year 2022-23

A.C. No: 13
Agenda No: 3 (xvi)

Evaluation Pattern

The performance of the learner will be evaluated in two components. The first component will be a Continuous Assessment with a weightage of 25% of total marks per course. The second component will be a Semester End Examination with a weightage of 75% of the total marks per course. The allocation of marks for the Continuous Assessment and Semester End Examinations is as shown below:

a) Details of Continuous Assessment (CA)

25% of the total marks per course:

Continuous Assessment	Details	Marks
Component 1 (CA-1)	Test	15 marks
Component 2 (CA-2)	Test	10 marks

b) Details of Semester End Examination

75% of the total marks per course. Duration of examination will be THREE HOURS.

Question Number	Description	Marks	Total Marks
Q.1 to Q.4	A. Attempt any ONE sub questions out of Two sub questions.	10	15 Marks 15 x 4 = 60 Marks
	B. One Compulsory sub question	5	
Q.5	Attempt any THREE sub questions out of Four Sub Questions. (1 Sub question from Each Unit).	5	15
		Total Marks	75

c) The Period for the internship or the Statistical project will be 6 weeks. At the end of each week, the learner has to meet the coordinator and submit the weekly report having mentioned the date and time of submission of the weekly report. Questions (if any) raised by the coordinator have to be answered by the learner and a record of this Q & A session should be maintained. At the end of the internship or completion of the Statistical Project, the learner will have to give a presentation of his work and satisfactorily answer the questions asked by the examiners.

Signature
HOD

Signature
Approved by Vice –Principal

Signature
Approved by Principal

Program: M.Sc.		Semester: III	
Course: Multivariate Analysis II		Course Code: PSMAST301	
Teaching Scheme		Evaluation Scheme	
Lecture (per week)	Credits	Continuous Assessment (CA) (Marks - 25)	Semester End Examinations (SEE) (Marks - 75 in Question Paper)
4-	4	25	75

Objectives:

To learn the technique of compression and classification of large data.

To reduce the number of variables by using lesser number of surrogate variables (factors) while retaining the variability.

To understand the nature of measurement error and it impact on multivariate analysis.

To define specific techniques included in Multivariate Analysis.

To explore the relationships between two multivariate sets of variables (vectors).

Outcomes:

(CO1: Remember)

The nature of measurement scales and their relationship to Multivariate Techniques.

The similarities and differences between multiple regression, discriminant analysis, factor analysis, and canonical correlation.

(CO2: Understand)

The guidelines for application and interpretation of Multivariate Analysis

What Multivariate analysis is, and explain where and when it can be used.

(CO3: Apply)

Carry out classification of given multivariate data Analyze and analysis of the given multivariate data.

To determine which Multivariate Technique can be used for a specific research project.

(CO4: Analyse)

Perform the statistical inference procedures using the data from multivariate distribution.

(CO5: Evaluate)

Perform an extensive exploratory multivariate analysis for a given multivariate data.

Solve problems involving multivariate normal distribution.

Outline of Syllabus: (per session plan)

Module	Description	No of Hours
1	Principal Components	15
2	Factor Analysis.	15
3	Canonical Correlation	15
4	Cluster Analysis	15
	Total	60

Module	Topic	No. of Lectures /Creditss 60/4
1	Introduction, Population Principal Components Principal Components Obtained from Standardized Variables, Principal Components for Covariance Matrices with Special Structures. Summarizing Sample Variation by Principal Components, The Number of Principal Components, Interpretation of the Sample Principal Components, Standardizing the Sample Principal Components. Graphing the Principal Components. Large Sample Inferences Large Sample Properties of $\widehat{\lambda_\iota}$ and $\widehat{e_\iota}$, Testing for the Equal Correlation Structure, Monitoring Quality with Principal Components, checking a Given Set of Measurements for Stability, Controlling Future Values, The Geometry of the Sample Principal Component Approximation; The p-Dimensional Geometrical Interpretation; The n-Dimensional Geometrical Interpretation.	15
2	Factor Analysis: Introduction, The Orthogonal Factor Model. Methods of Estimation; The Principal Component (and Principal Factor) Method; A Modified Approach-the Principal Factor Solution, The Maximum Likelihood Method, A Large Sample Test for the Number of Common Factors. Factor Rotation, Oblique Rotations, Factor Scores, The Weighted Least Squares Method, The Regression Method. Perspectives and a Strategy for Factor Analysis. Some Computational Details for Maximum Likelihood Estimation, Computational Scheme, maximum likelihood Estimators of $\rho = L_z L_z' + \varphi_z$.	15
3	Introduction, Canonical Variates and Canonical Correlations. Interpreting the Population Canonical Variables, Identifying the Canonical Variables, Canonical Correlations as Generalizations of Other Correlation Coefficients, The First r Canonical Variables as a Summary of Variability. A Geometrical Interpretation of the Population Canonical Correlation Analysis. The Sample Canonical Variates and Sample Canonical Correlations. Additional Sample Descriptive Measures. Matrices of Errors of Approximations. Proportions of Explained Sample Variance. Large Sample Inferences.	15

4	Introduction, Similarity Measures; Distances and Similarity Coefficients for Pairs of Items, Similarities and Association Measures for Pairs of Variables; Concluding Comments on Similarity. Hierarchical Clustering Methods; Single Linkage, Complete Linkage, Average Linkage, Wards Hierarchical Clustering Method, Hierarchical Procedures, Nonhierarchical Clustering Methods, K-means Method, Non-hierarchical Procedures. Clustering Based on Statistical Models, Multidimensional Scaling The Basic Algorithm. Correspondence Analysis, Algebraic Development of Correspondence Analysis, Inertia, Interpretation in Two Dimensions, Biplots for Viewing Sampling Units and Variables, Constructing Biplots. Procrustes Analysis: A Method for Comparing Configurations, Constructing the Procrustes Measure of Agreement.	15

Essential Readings:

1. Johnson, R. A. and Wichern, D. W. (2015): Applied Multivariate Statistical Analysis. 6th Edition. PHI Learning Private Limited.
2. Anderson, T. W. (2003): An Introduction to Multivariate Statistical Analysis. John Wiley. 3rd edition.

Supplementary Readings:

1. Giri, N. C. (2003): Multivariate Statistical Analysis. CRC Press. 2nd edition.
2. Hardle, W. K. and Hlavka, Z. (2015): Multivariate Statistics: Exercise and solutions. Springer.
3. Kshirsagar, A. M. (1979): Multivariate Analysis, Marcel Dekker Inc. New York.
4. Mukhopadhyay, P. (2008): Multivariate Statistical Analysis. World Scientific.

Srivastava, M. S. (2002): Methods of Multivariate Statistics. John Wiley.

Program: M.Sc.		Semester: III	
Course: Time Series Analysis		Course Code: PSMAST305	
Teaching Scheme		Evaluation Scheme	
Lecture (per week)	Credits	Continuous Assessment (CA) (Marks - 25)	Semester End Examinations (SEE) (Marks - 75 in Question Paper)
4	4	25	75

Course Objectives:

- To equip various forecasting techniques, get familiarize to modern statistical methods for analyzing time-series data.
- Learn the intellectual facts of the time series data and to implement the same.
- To link time-dependent analytical tools and building the models by extracting real-time data.

Course Outcomes:

(CO1: Remember)

the theory related to linear time series models;
the theory related to estimation and forecasting;
the concepts of stationarity of a time series;
the theory related to ARCH/GARCH models

(CO2: Understand)

the theory related to linear time series models, estimation and forecasting, ARCH/GARCH models
the concepts of stationarity of a time series;

(CO3: Apply)

use information criteria for the selection of models;
the concepts of exploratory analysis.

(CO4: Analyse)

Perform an exploratory analysis of time series;
Test the stationarity of a time series;
Fit an appropriate linear time series model for the data;
Data using ARCH/GARCH models, Count data.

(CO5: Evaluate)

Solve various problems related to a Time series.

Outline of Syllabus: (per session plan)

Module	Description	No of Hours
1	Time Series Models	15
2	Auto Regressive Models	15
3	Forecasting and Estimation in Time Series Models	15
4	Estimation of ARIMA Models	15
	Total	60

Module	Topic	No. of Lectures /Credits 60/4
1	Real life examples of time series, types of variation in time series, exploratory time series analysis, tests of randomness, tests for trend, seasonality. Auto-covariance and auto-correlation functions and their properties, Exponential and Moving average smoothing. Holt -Winters smoothing. Forecasting based on smoothing, adaptive smoothing. Time - series as a discrete parameter stochastic process. Portmanteau tests for noise sequences, transformation to obtain Gaussian series. General linear processes.	15
2	Auto regressive (AR), Moving average (MA) and Autoregressive moving average (ARMA), Stationarity and invertibility conditions. Nonstationary and seasonal time series models: Auto regressive integrated moving average (ARIMA) models, Seasonal ARIMA (SARIMA) models, Transfer function models (Time series regression).	15
3	Estimation of mean, auto covariance and autocorrelation functions, Yule-Walker estimation. Forecasting in time series models, Durbin-Levinson algorithm, innovation algorithm.	15
4	Estimation of ARIMA model parameters, Choice of AR and MA periods, FPE, AIC, BIC, residual analysis and diagnostic checking, Unit-root non stationarity, unit-root tests, ARCH and GARCH models.	15

Essential Readings:

1. Brockwell, P. J. and Davis, R. A. (2003): Introduction to Time Series Analysis, Springer
2. Fuller, W. A. (1996): Introduction to Statistical Time Series, 2nd Ed. Wiley.

Supplementary Readings:

1. Chatfield, C. (2001): Time Series Forecasting, Chapman &Hall.
2. Hamilton, N. Y. (1994): Time Series Analysis, Princeton University press.
3. Kendall, M. and Ord, J. K. (1990): Time Series, 3rd Ed. Edward Arnold.
4. Lutkepohl, H. (2005): New Introduction to Multiple Time Series Analysis, Springer
5. Shumway, R. H. and Stoffer, D. S. (2010): Time Series Analysis & Its Applications, Springer.
6. Tsay, R. S. (2010): Analysis of Financial Time Series, Wiley.

Program: M.Sc.		Semester: III	
Course: Statistics Using R and Python		Course Code: PSMAST306	
Teaching Scheme		Evaluation Scheme	
Lecture (per week)	Credits	Continuous Assessment (CA) (Marks - 25)	Semester End Examinations (SEE) (Marks - 75 in Question Paper)
4	4	25	75

Course Objectives:

To make the learner:

- Aware of R as a Statistical Software and language, R preliminaries, Method of data input, Data accessing or indexing, Data frames and lists, Functions, Graphics with R, Saving, Storing and retrieving work, work space and files, using scripts, using packages.
- use the R software to learn different inbuilt/library function sof R for standard statistical calculations.
- Aware of the principles of statistical thinking and interpretation byexample, exercises and discussion about a range of problems. The examples will be used to present a variety of statistical concepts and techniques.
- To introduce the learner to Python as a programming language.

Course Outcomes:

After completion of the course, learners would be able to:

(CO1: Remember)

i) Remember and implement the statistical theory in R/Python.

(CO2: Understand)

i) Understand how to Assign vectors using different functions, make use of operators, perform basic calculations using inbuilt functions, create diagrams and plots.

ii) summarize and graph data.

(CO3: Apply)

i) handle data, perform basic data analysis procedures

ii) assess whether different variables are linked, using correlation and regression analysis.

(CO4: Analyse)

i) choose the right method to summarize a dataset, graphically and numerically

ii) perform basic hypothesis tests on a data set.

(CO5: Evaluate)

Calculate confidence intervals, test hypotheses, assess goodness-of-fit, and perform linear regression.

Outline of Syllabus: (per session plan)

Module	Description	No of Hours
1	R fundamentals, Measures of Central Tendency and Dispersion. R Graphics.	15
2	Correlation and Regression; Distribution Theory, Inference. Principal Components and Discriminant Analysis	15
3	Python Programming Basics	15

4	Automating Tasks	15
	Total	60
Module	Topic	No. of Lectures /Credits 60/4
Module 1	Introduction to R, features of R, installation of R, starting and ending R session, getting help in R, R commands and case sensitivity Data types: Logical, numeric and complex Vectors and vector arithmetic. Creation of vectors using functions c, assign, seq, rep Arithmetic operations on vectors using operators +, -, *, /, ^. Numerical functions: log10, log, sort, max, min, unique, range, length, var, prod, sum, summary etc. Accessing vectors Alternative ways to create vectors by scan function Data frames: creation using data.frame, subset and transform commands Resident data sets: Accession and summary Measures of Central Tendency: Mean, Mode, Median, Quartiles, Deciles, Percentiles, G.M and H.M. Measures of Dispersion: Range, Variance, Standard deviation, coefficient of variation, Mean deviation. Skewness: Bowley's coefficient and Karl Pearson's coefficient of skewness Moments: Computations of raw and central moments, measure of skewness and kurtosis based on it. Graphics using R: a. High level plotting functions b. Low level plotting functions c. Interactive graphic functions Diagrams and Graphs. (Using R) Simple bar diagram, Subdivided bar diagram, multiple bar diagram, Piediagram, Stem and leaf diagram. Box plot, rod or spike plot, histogram (both equal and unequal classintervals), frequency polygon, Ogive curves.	15
Module 2	Fitting of lines of regression, computation of correlation coefficient, Fitting of parabola. Multiple Regression: Fitting of regression plane for trivariate data. Computation of probabilities, Model Sampling, fitting random number generation for Hyper geometric, Binomial, Poisson, Geometric, Negative Binomial, Normal, Exponential, Gamma, Beta distributions	15

	Plots to check normality. Goodness of fit, One/Two Way ANOVA. Principal Components and Discriminant Analysis.	
Module 3	Python basics, Flow Control, Functions, Lists, Dictionaries and Structuring Data, String Manipulation, Regular Expressions.	15
Module 4	Pattern Matching, input validation, Reading, Writing Organising Files, Working with Google spreadsheets. Working with Excel spreadsheets - Pivot, vlookup-hloopup, formatting, shortcuts, etc.	15

Essential Readings:

1. Dr. Marsh Gardner, Beginning R, Wiley. (2012)
2. Micheal J Crawley, The R book, 2nd Ed., Wiley.
3. AI Sweigart, Automate the Boring Stuff with Python, 2nd Ed, No Starch Press.
4. David Beazley, Brian K. Jones, Python Cookbook, 3rd Ed., O'Reilly.

Supplementary Readings:

1. Richard Cotton, Learning R, Shroff/ O'Reilly.
2. Maria. L. Rizzo (2007), Statistical Computing with R - (Chapman& Hall/CRC)
3. John Braun, Duncan J. Murdoch, A first course in Statistical Programming with R. Cambridge.
4. Wes McKinney, Python for Data Analysis, 2nd Edition, O'Reilly.
5. Sameer Madhavan, Mastering Python for Data Science. (2015), Packt Publishing (Open Source).

Program: M.Sc. (2018-19)		Semester: III	
Course: ELEMENTS OF DATA SCIENCE.		Course Code: PSMAST304	
Teaching Scheme		Evaluation Scheme	
Lecture (per week)	Credits	Continuous Assessment (CA) (Marks - 25)	Semester End Examinations (SEE) (Marks - 75 in Question Paper)
4	4	25	75

Objectives:

To introduce the learner to

- The structure of analytics/data science/machine learning and its business use cases
- The regularization techniques (Ridge, Lasso)
- The basics of machine learning (bias-variance, under fitting-overfitting)
- The detailed working of decision trees, its advantages and disadvantages
- bagging, boosting, random forest, GBM
- the basic structure of artificial neural networks
- Basics of SQL - Select, where, group by, having, order by, joins.

Outcomes:

(CO1: Remember)

The concepts related to supervised and unsupervised learning methods, text mining, SVM, Neural Networks

Random Forests, Regression trees.

A general classification of task to integrate data mining system.

(CO2: Understand)

The concept of data base technology which has led to the emergence of data mining and its applications

Classification methods for data;

Concepts of Supervised and unsupervised learning methods, text mining and SVM, Neural Networks; Dimension and reduction.

(CO3: Apply)

Supervised and unsupervised learning methods to different data sets.

Clustering algorithms and related methods;

Text mining in various contexts; Decision trees

Apply statistical methods for any given raw data.

(CO4: Analyse)

the concepts of feature selection and feature extraction;

the concepts of Regression Trees, Random Forests, Bagging and boosting.

Online analytical data processing.

(CO5: Evaluate)

Regression Trees, Random Forests, Bagging and boosting.

appropriate data mining algorithms and apply, interpret and report the output appropriately.

Outline of Syllabus: (per session plan)

Module	Description	No of Hours

1	Introduction to Data Mining.	15
2	Reduction and Visualization Techniques.	15
3	AI and Machine Learning.	15
4	Neural Networks.	15
	Total	60
Module	Topic	No. of Lectures /Credits 60/4
1	Introduction to Data Mining, Classification techniques, CART, Random forests, Bayesian classification and learning rules. Introduction to Big Data. Large dimension small size multivariate data analysis, tackling the problems of estimation and inference. Classification of Big Data, Screening and Variable Selection.	15
2	Dimension Reduction and Visualization Techniques, Algorithms for data-mining using multiple nonlinear and nonparametric regression, Lasso Regression, Projection Methods, penalty, ridge regression, Bootstrap methods.	15
3	Introduction to Nonlinear regression, Introduction to Nonparametric regression, generalized additive models, kernel methods, neural network, Artificial Intelligence, machine learning. Introduction to Structured Data and Structural Equation Modeling.	15
4	Neural Networks: Multi-layer perceptron, predictive ANN model building using back propagation algorithm, Exploratory data analysis using Neural Networks – self organizing maps. Genetic Algorithms, Neuro-genetic model building. SQL Basics.	15

Essential Readings:

1. Hastie, T., Tibshirani, R. and Friedman, J. H. (2001): The Elements of Statistical Learning: Data Mining, Inference & Prediction, Springer Series in Statistics, Springer-Verlag.
2. Hastie, T., Tibshirani, R. and Wainwright, M. (2015): Statistical Learning with Sparsity: The Lasso and generalizations.

Supplementary Readings:

1. Breiman, L., Friedman, J. H., Olschen, R. A. and Stone, C.J. (1984): Classification of Regression Trees, Wadsowrth Publisher.
2. Hand, D. J. , Mannila, H. and Smith, P. (2001): Principles of Data Mining, MIT Press, Cambridge.

3. Hassoun, M. H. (1998): Fundamentals of Artificial Neural Networks, Prentice-Hall of India, New Delhi.
4. Hardle, W.(1990): Applied Nonparametric Regression, Cambridge University Press.
5. Hastie, T. and Tibshirani, R.(1990): Generalized Additive Models, Chapman and Hall, London.
6. Seber, G. A. F. and Wild, C. J. (1989): Nonlinear Regression, John Wiley.

Program: M.Sc.		Semester: III	
Course: Statistics Practical V		Course Code: PSMAST P3A	
Teaching Scheme		Evaluation Scheme	
Practical (per week)	Credits	Continuous Assessment (CA) (Marks - 20)	Semester End Examinations (SEE) (Marks - 80 in Question Paper)
4	4	--	100

Outline of Syllabus: (per session plan)

STATISTICS PRACTICAL-V

PSMAST P3A	Practicals Based on Multivariate Analysis II (PSMAST 301) and Time Series Analysis (PSMAST302). All Practicals should be performed using R software.	No. of Lectures /Credits 60/4
1	Principal Component Analysis 1	
2	Principal Component Analysis 2	
3	Factor Analysis 1	
4	Factor Analysis 2	
5	Canonical Correlation 1	
6	Canonical Correlation 2	
7	Cluster Analysis 1	
8	Cluster Analysis 2	
9	Exponential and Moving average smoothing. Holt -Winters smoothing.	
10	Auto regressive (AR), Moving average (MA) and Autoregressive moving average (ARMA).	
11	Nonstationary and seasonal time series models: Auto regressive integrated moving average (ARIMA) models, Seasonal ARIMA (SARIMA) models (Time series regression.	
12	Estimation of ARIMA model parameters. AR and MA periods.	
13	FPE, AIC, BIC, residual analysis and diagnostic checking, ARCH and GARCH models.	

Program: M.Sc.		Semester: III	
Course: Statistics Practical VI		Course Code: PSMAST P3B	
Teaching Scheme		Evaluation Scheme	
Practical (per week)	Credits	Continuous Assessment (CA) (Marks - 20)	Semester End Examinations (SEE) (Marks- 80 in Question Paper)
4	4	--	100

Outline of Syllabus: (per session plan)

STATISTICS PRACTICAL-VI

PSMAST P3B	Practicals Based on Statistics Using R and Python (PSMAST 306) and Elements of Data Science (PSMAST 304).	No. of Lectures /Creditss 60/4
1	Practicals on R software 1	
2	Practicals on R software 2	
3	Practicals on R software 3	
4	Practicals on R software 4	
5	Practicals on Python 1	
6	Practicals on Python 2	
7	Practicals on Python 3	
8	Elements of Data Science 1	
9	Elements of Data Science 2	
10	Elements of Data Science 3	

Program: M.Sc.		Semester: IV	
Course: Stochastic Processes		Course Code: PSMAST401	
Teaching Scheme		Evaluation Scheme	
Lecture (per week)	Credits	Continuous Assessment (CA) (Marks - 25)	Semester End Examinations (SEE) (Marks - 75 in Question Paper)
4	4	25	75

Course Objective

- To be aware of the basic concepts of the theory of stochastic processes;
- To introduce the most important types of stochastic processes;
- To learn the various properties and characteristics of processes;
- Methods for describing and analyzing complex stochastic models.
- To introduce the learner to the important types of stochastic processes (Poisson, Markov, Gaussian, Wiener processes and others).
- To finding the most appropriate process for modelling in particular situations arising in economics, engineering and other fields;
- To know the idea of ergodicity, stationarity, stochastic integration; application of these terms in context of stochastics processes.
- To make the learner aware of the fact that Stochastic Processes are a necessary theoretical basis for studying other programs such as financial mathematics, quantitative finance, stochastic modeling and the theory of jump - type processes.

Course Outcomes

(CO1: Remember)

The difference differential equations of different stochastic processes.

(CO2: Understand)

The concepts related to birth-death processes, Poisson processes, Renewal processes, Markov chain, Branching processes.

(CO3: Apply)

The knowledge of stochastic processes in different fields.

(CO4: Analyse)

To work with problems of stochastic analysis for modeling in different application areas such as financial mathematics.

(CO5: Evaluate)

solve some stochastic differential equations.

solve problems related to different types of stochastic processes.

Outline of Syllabus: (per session plan)

Module	Description	No of Hours
1	Introduction to stochastic processes	15
2	Continuous time Processes	15
3	Renewal Process	15
4	Branching Process	15
	Total	60

Module	Topic	No. of Lectures /Credits 60/4
1	Introduction to stochastic processes, specification of stochastic processes, real life applications of stochastic processes, introduction to different types of stochastic processes. Markov chain, real life examples of Markov chain, order of a Markov chain, transition probabilities, Chapman-Kolmogorov equations, classification of states, periodicity, closed class, minimal closed class, stationary distribution of a Markov chain. Gamblers ruin problem, random walk. Concept of absorption probabilities, Statistical inference for Markov chains.	15
2	Continuous time Processes: Poisson process, Generalizations of Poisson process, birth and Death process. Brownian Motion, Wiener process, Kolmogorov equations.	15
3	Renewal Process: Renewal process in continuous time, renewal equation, stopping time, renewal theorem. Real life applications.	15
4	Branching Process: Introduction to branching process, probability generating function of branching process, moments, classification of states, identification of criticality parameter, extinction probability, relationship between criticality parameter and extinction probability of the process, Expression for mean and variance of the process. Extinction probability, Some applications.	15

Essential Readings:

1. Basu, S (2012): Applied Stochastic Processes. New Central book agency.
2. Hoel, P. G., Port, S. C. and Stone, C. J. (1972): Introduction to Stochastic Processes, Houghton.

Supplementary Readings:

1. Bhat, B.R. (2000). Stochastic Models: Analysis and Applications, New Age International.
2. Bhat, U. N. and Miller, G. K. (2002): Elements of Applied Stochastic Processes. 3rd Edition. Wiley.
3. Durrett, R. (1999): Essentials of Stochastic Process. Mifflin
4. Karlin, S. and Taylor, H. M. (1975): First Course in Stochastic Processes second edition.
5. Kulkarni, V. G. (2011): Modeling and Analysis of Stochastic Systems, Chapman and Hall, London.
6. Medhi, J. (1994): Stochastic Processes Second edition, Wiley Eastern.
7. Ross, S. M. (2004): Introduction to Probability Models, Wiley Eastern.

Program: M.Sc.		Semester: III	
Course: Statistical Quality Control		Course Code: PSMAST405	
Teaching Scheme		Evaluation Scheme	
Lecture (per week)	Credits	Continuous Assessment (CA) (Marks - 25)	Semester End Examinations (SEE) (Marks - 75 in Question Paper)
4	4	25	75

Objectives:

- To Define and discuss quality and quality improvement, the different dimensions of quality, product liability.
- To Discuss the evolution of modern quality improvement methods
- To Discuss the role that variability and statistical methods play in controlling and improving quality.
- To Discuss chance and assignable causes of variability in a process
- To Explain the statistical basis of the Shewhart control chart, including choice of sample size, control limits, and sampling interval.
- To design control charts for variables and attributes.
- To set up and use control charts for individual measurements
- To understand the importance of the normality assumption for individuals control charts and know how to check this assumption.
- To Understand the role of acceptance sampling in modern quality control systems
- To understand the difference between attributes and variables sampling plans, and the major types of acceptance-sampling procedures.

Outcomes:

(CO1: Remember)

the links between quality and productivity and between quality and cost.

the three functions: quality planning, quality assurance, and quality control and improvement.

(CO2: Understand)

the rational subgroup concept for variables control charts

the basic tools of SPC: the histogram or stem-and-leaf plot, the check sheet, the Pareto chart, the cause-and-effect diagram, the defect concentration diagram, the scatter diagram, and the control chart.

the effects of the sampling plan parameters on sampling plan performance

(CO3: Apply)

how average run length is used as a performance measure for a control chart.

how to design single-sampling, double-sampling, and sequential-sampling plans for attributes

(CO4: Analyse)

phase I and phase II use of control charts.

sensitizing rules and pattern recognition are used in conjunction with control charts.

How single, double, and sequential-sampling plans are used to determine the OC curve for a single-sampling plan for attributes.

The Use of rectified inspection.

(CO5: Evaluate)

Determine the average run length for variables control chart. The structure and use of MIL STD 105E and its civilian counterpart plans The structure and use of the Dodge–Romig system of sampling plans.		
Outline of Syllabus: (per session plan)		
Module	Description	No of Hours
1	Quality Improvement & Statistical Process Control	15
2	Control Charts	15
3	Statistical Process Monitoring And Control Techniques.	15
4	Acceptance Sampling	15
	Total	60
Module	Topic	No. of Lectures /Credits 60/4
Module 1	*Quality Improvement In The Modern Business Environment* The Meaning of Quality and Quality Improvement; Dimensions of Quality; Quality Engineering Terminology; A Brief History of Quality Control and Improvement; Statistical Methods for Quality Control and Improvement; Management Aspects of Quality Improvement; Quality Philosophy and Management Strategies; The Link Between Quality and Productivity; Supply Chain Quality Management; Quality Costs; Legal Aspects of Quality; Implementing Quality Improvement. *The DMAIC Process* Overview of DMAIC; The Define Step; The Measure Step; The Analyze Step; The Improve Step; The Control Step; Examples of DMAIC; Litigation Documents; Improving On-Time Delivery; Improving Service Quality in a Bank. *Methods And Philosophy Of Statistical Process Control.* Introduction Chance and Assignable Causes of Quality Variation; Statistical Basis of the Control Chart; Basic Principles; Choice of Control Limits; Sample Size and Sampling Frequency; Rational Subgroups; Analysis of Patterns on Control Charts; Discussion of Sensitizing Rules for Control Charts; Phase I and Phase II of Control Chart Application; The Magnificent Seven; Implementing SPC in a Quality Improvement Program; An Application of SPC; Applications of Statistical Process Control and Quality Improvement Tools in Transactional and Service Businesses.	15

Module 2	*Control Charts For Variables:* Introduction; Control Charts for $\bar{x}$ and *R;* Statistical Basis of the Charts; Development and Use of $\bar{x}$ and *R* Charts; Charts Based on Standard Values; Interpretation of $\bar{x}$ and *R* Charts; The Effect of Non-normality on $\bar{x}$ and *R* Charts; The Operating-Characteristic Function; The Average Run Length for the $\bar{x}$ Chart; Control Charts for $\bar{x}$ and *c*onstruction and Operation of $\bar{x}$ and *s* Charts; The $\bar{x}$ and *s* Control Charts with Variable Sample Size; The *s*2 Control Chart; The Shewhart Control Chart for Individual Measurements; Summary of Procedures for $\bar{x}$, *R*, and *s* Charts; Applications of Variables Control Charts. *Control Charts For Attributes* Introduction; The Control Chart for Fraction Nonconforming; Development and Operation of the Control Chart; Variable Sample Size; Applications in Transactional and Service Businesses; The Operating-Characteristic Function and Average Run Length Calculations; Control Charts for Nonconformities (Defects); Procedures with Constant Sample Size; Procedures with Variable Sample Size; Demerit Systems; The Operating-Characteristic Function; Dealing with Low Defect Levels; Nonmanufacturing Applications; Choice Between Attributes and Variables Control Charts; Guidelines for Implementing Control Charts.	15
Module 3	*Process And Measurement System Capability Analysis:* Introduction; Process Capability Analysis Using a Histogram, a Probability Plot. Process Capability Ratios: Use and Interpretation of *Cp;* Process Capability Ratio for an Off-Center Process; Normality and the Process Capability Ratio; More about Process Centering; Confidence Intervals and Tests on Process Capability Ratios. Process Capability Analysis Using a Control Chart, Process Capability Analysis with Attribute Data. Setting Specification Limits on Discrete Components: Linear Combinations; Nonlinear Combinations. Estimating the Natural Tolerance Limits of a Process: Tolerance Limits Based on the Normal Distribution; Nonparametric Tolerance Limits. The Cumulative Sum Control Chart: Basic Principles: The CUSUM Control Chart for Monitoring the Process Mean; The Tabular or Algorithmic CUSUM for Monitoring the Process Mean; Recommendations for CUSUM Design; The Standardized CUSUM; Improving CUSUM Responsiveness for Large Shifts; The Fast Initial Response or Headstart Feature; One-Sided CUSUMs; A CUSUM for Monitoring Process Variability;	15

	Rational Subgroups; CUSUMs for Other Sample Statistics; The V-Mask Procedure; The Self-Starting CUSUM. The Exponentially Weighted Moving Average Control Chart: The Exponentially Weighted Moving Average Control Chart for Monitoring the Process Mean; Design of an EWMA Control Chart: Robustness of the EWMA to Nonnormality; Rational Subgroups; Extensions of the EWMA; Moving Average Control Chart. Statistical Process Control for Short Production Runs: $\bar{x}$ and R Charts for Short Production Runs; Attributes Control Charts for Short Production Runs; Other Methods. Modified and Acceptance Control Charts; Modified Control Limits for the $\bar{x}$ Chart; Acceptance Control Charts .457 Control Charts for Multiple-Stream Processes: Multiple-Stream Processes; Group Control Charts; Other Approaches 460 SPC With Autocorrelated Process Data: Sources and Effects of Autocorrelation in Process Data. Model-Based Approaches; A Model-Free Approach.	
Module 4	The Acceptance-Sampling Problem; Advantages and Disadvantages of Sampling; Types of Sampling Plans; Lot Formation; Random Sampling; Guidelines for Using Acceptance Sampling. Single-Sampling Plans for Attributes: Definition of a Single-Sampling Plan; The OC Curve; Designing a Single-Sampling Plan with a Specified OC Curve; Rectifying Inspection. Double, Multiple, and Sequential Sampling: Double-Sampling Plans; Multiple-Sampling Plans; Sequential-Sampling Plans. Military Standard 105E (ANSI/ASQC Z1.4, ISO 2859): Description of the Standard; Procedure; Discussion. The Dodge–Romig Sampling Plans: AOQL Plans, LTPD Plans, Estimation of Process Average. Acceptance Sampling by Variables: Advantages and Disadvantages of Variables Sampling; Types of Sampling Plans Available; Caution in the Use of Variables Sampling; Designing a Variables Sampling Plan with a Specified OC Curve. MIL STD 414 (ANSI/ASQC Z1.9): General Description of the Standard; Use of the Tables; Discussion of MIL STD 414 and ANSI/ASQC Z1.9; Other Variables Sampling Procedures; Sampling by Variables to Give Assurance Regarding the Lot or Process Mean; Sequential Sampling by Variables; Chain Sampling; Continuous Sampling; CSP-1 701; Other Continuous-Sampling Plans; Skip-Lot Sampling Plans.	15

Essential Readings:

1. Douglas C Montgomery; Introduction to Statistical Quality Control. 6th Edition, (2009) John Wiley and sons Inc.

Supplementary Readings:

1) Duncan, A. J. (1986). *Quality Control and Industrial Statistics,* 5th ed., Irwin, Homewood, IL.
2) Grant, E. L., and R. S. Leavenworth (1980). *Statistical Quality Control,* 5th ed., McGraw-Hill, New York.
3) Bhisham C. Gupta - Statistical Quality Control_ Using MINITAB, R, JMP and Python-Wiley (2021).
4) Irving W Burr, Statistical Quality Control Methods, (2018),Routledge.

Program: M.Sc.		Semester: IV	
Course: Reliability and Survival Analysis.		Course Code: PSMAST403	
Teaching Scheme		Evaluation Scheme	
Lecture (per week)	Credits	Continuous Assessment (CA) (Marks - 25)	Semester End Examinations (SEE) (Marks- 75 in Question Paper)
4	4	25	75

Course Objective:

- To introduce the learner to the analysis of time-to-event data.
- To make the learner aware that Survival methods are considered appropriate to incorporate the variations from both uncensored and censored observations.
- Learners will learn some parametric, nonparametric and semiparametric approaches to analyse survival data.
- To introduce Descriptive methods, Kaplan-Meiers curves, regression models for survival data, and proportional hazard models.
- To model and test differences in survival times of two or more groups of interest together with the effect of one or more variable on survival time.
- It is expected that the course will develop and nurture the learners interest to explore more about time-to-event data.

Course Outcomes:

(CO1: Remember)

To distinguish between different types of systems and evaluate the reliability of such systems

The concepts related to the estimation of survival function under a parametric regression set up.

(CO2: Understand)

The concepts related to the estimation of survival function under a parametric regression set up

The concepts of ageing of systems and classify them based on ageing properties

(CO3: Apply)

The concepts related to the point-process approach of survival function estimation and analysis.

The theory related to competing risk models and apply them for the estimation of survival function.

(CO4: Analyse)

Learn the concepts related to frailty modeling and Apply them for survival data.

Learn the concepts related to estimation of survival function under a semi-parametric regression set up (Cox PH model)

(CO5: Evaluate)

To estimate the survival function parametrically using various parametric models from the given survival data.

To compute the survival probability with respect to various ageing models.

Outline of Syllabus: (per session plan)

Module	Description	No of Hours
1	Survival Functions	15
2	System Reliability	15

3	Survival Analysis	15
4	Regression models in Survival analysis and Frailty Models.	15
	Total	60
Module	Topic	No. of Lectures /Credits 60/4
1	Survival function, Hazard function, cumulative hazard function, reversed hazard function, nature of hazard function, bath-tub shape hazard function, class of increasing failure rate distributions, decreasing failure rate distributions, theorems. Relations between survival function, probability function, hazard function, cumulative hazard function, reversed hazard function.	15
2	Reliability of the system: structure function, standard systems: series system, parallel system, k-out-of-n system, coherent system, path sets and path vectors, minimal path sets, cut sets and cut vector, minimal cut sets, reliability of coherent system, reliability bounds.	15
3	Introduction to Survival Analysis: need of survival analysis, censoring: left censoring, right censoring, interval censoring, random censoring, times censoring, order censoring, hybrid censoring. Kaplan-Meier estimator of survival function, properties of Kaplan-Meier estimator, Nelson-Aalen estimator of cumulative hazards function. Linear and log-transformed confidence interval for survival function and cumulative hazard function. Q-Q plot, hazards plot for lifetime distributions. Competing risk models.	15
4	Regression models in Survival analysis: proportional hazards model, Accelerated failure time model, Cox proportional hazards model, residual analysis of proportional hazards model. Frailty models: Univariate frailty, multivariate frailty models, shared frailty, correlated frailty, additive frailty models. Using Weibull as baseline and gamma as frailty distribution.	15

Essential Readings:

1. Barlow, R. E. and Proschan, F. (1965): Mathematical theory of reliability
2. Barlow, R. E. and Proschan, F. (1975): Statistical theory of reliability and life testing. Holt, Reinhart and Winston.

Supplementary Readings:

1. Deshpande, J. V. and Purohit, S. G. (2005). Life Time Data: Statistical Models and Methods, World Scientific.
2. Hanagal, D. D. (2011). Modeling Survival Data Using Frailty Models. CRC Press.

3. Hosmer, D. and Lemeshow, S. (1999). Applied Survival Analysis: Regression Modeling of Time to Event Data, Wiley, New York.
4. Kalbfleisch, J. D. and Prentice, R.L. (1986): The Statistical Analysis of Failure Time Data, John Wiley.
5. Kleinbaum, D. G. and Klein, M. (2012). Survival Analysis: A Self-Learning Text, 3rd Ed, Springer, New York.
6. Lawless, J.F.(1982): Statistical models and methods for life time data. John Wiley.
7. Lee, E. T. and Wang, J. W. (2003). Statistical Methods for Survival Data Analysis, 3rd Edition. John Wiley.
8. Liu, X. (2012). Survival Analysis: Models and Applications, Wiley, New York.
9. Ross S. M. (2014): Introduction to Probability Models. Elsevier. 11th Edition.
10. Smith, P.J. (2002): Analysis of Failure and Survival data. CRC.
11. Wienke, A. (2011). Frailty Models in Survival Analysis, CRC..

Program: M.Sc.		Semester: IV	
Course: Statistics Practical VII		Course Code: PSMAST P4A	
Teaching Scheme		Evaluation Scheme	
Practical (per week)	Credits	Continuous Assessment (CA) (Marks - 20)	Semester End Examinations (SEE) (Marks- 80)
4	4	--	100

PRACTICALS: STATISTICS PRACTICAL-VII		
PSMAST P4A	Practicals Based on Stochastic Processes (PSMAST401), Statistical Quality Control (PSMAST405) and Reliability and Survival Analysis (PSMAST403) All Practicals should be performed using R software.	No. of Lectures /Credits 60/4
1	Introduction to stochastic processes	
2	Continuous time Processes	
3	Renewal Process	
4	Branching Process	
5	*Methods Statistical Process Control*	
6	Control Charts For Variables	
7	Control Charts For Attributes	
8	Process And Measurement System Capability Analysis	
9	Acceptance-Sampling	
10	Survival Functions	
11	System Reliability	
12	Survival Analysis	
13	Regression models in Survival analysis and Frailty Models.	

Program: M.Sc.		Semester: IV	
Course: Internship or Statistical Project .		Course Code: PSMAST P4B	
Teaching Scheme		Evaluation Scheme	
Internship or Statistical Project	Credits	Continuous Assessment (CA) (Marks 80)	Semester End Examinations (SEE) (Marks- 120)
	8	--	200

PSMAST P4B WEEK	Course: Internship or Statistical Project .	No. of hours /Credits 240/8
1	Report of 040 hours of Internship or Statistical Project.	
2	Report of 080 hours of Internship or Statistical Project.	
3	Report of 120 hours of Internship or Statistical Project.	
4	Report of 160 hours of Internship or Statistical Project.	
5	Report of 200 hours of Internship or Statistical Project.	
6	Report of 240 hours of Internship or Statistical Project.	
7	Final submission and Presentation.	