Punyashlok Ahilyadevi Holkar Solapur University, Solapur

NAAC Accredited-2022
'B^{++}'Grade (CGPA 2.96)

Name of the Faculty: Science &Technology

(As per New Education Policy 2020)

Syllabus: Biostatistics

Name of the Course: M. Sc. II (Sem. III &IV)

(Syllabus to be implemented from June 2024)

A Four Semester M.Sc. Biostatistics Course Structure as per NEP-2020

M. Sc. program in Biostatistics consists of 88 credits. Credits of a course are specified against the title of the course.

M.Sc. Biostatistics Semester-I			
Course Type	**Course code**	**Course Title**	**No. of Credits**
Major Mandatory	DSC-1	Basic Epidemiology	4
	DSC-2	Statistical Inference-I	4
	LAB-1	Biostatistics Practical-I	2
	LAB-2	Biostatistics Practical-II	2
Major Elective	DSE-1(A)	Probability Distributions	4
	DSE-1(B)	Statistical Ecology	
	LAB-3	Biostatistics Practical-III	2
Research Methodology	RM	Research Methodology in Biostatistics	4
M.Sc. Biostatistics Semester-II			
Course Type	**Course code**	**Course Title**	**No. of Credits**
Major Mandatory	DSC-3	Design of Experiments and Bioassay	4
	DSC-4	Statistical Inference-II	4
	LAB-4	Biostatistics Practical-IV	2
	LAB-5	Biostatistics Practical-V	2
Major Elective	DSE-2 (A)	Stochastic Processes	4
	DSE-2 (B)	Nonparametric Methods	
	LAB-6	Biostatistics Practical-VI	2
On Job Training/ Field Project	OJT/FP	-----	4
M.Sc. Biostatistics Semester-III			
Course Type	**Course code**	**Course Title**	**No. of Credits**
Major Mandatory	DSC-5	Statistical Genetics	4
	DSC-6	Demography and Health Statistics	4
	LAB-7	Biostatistics Practical-VII	2
	LAB-8	Biostatistics Practical-VIII	2
Major Elective	DSE-3 (A)	Multivariate Statistical Methods	4
	DSE-3 (B)	Regression Analysis	
	LAB-9	Biostatistics Practical-IX	2

Research Project	RP-1	Research Project-I	4
M.Sc. Biostatistics Semester-IV			
Course Type	**Course code**	**Course Title**	**No. of Credits**
Major Mandatory	DSC-7	Clinical Trials	4
	DSC-8	Survival Analysis	4
	LAB-10	Biostatistics Practical-X	2
Major Elective	DSE-4 (A)	Data Mining	4
	DSE-4 (B)	Time Series Analysis	
	LAB-11	Biostatistics Practical-XI	2
Research Project	RP-2	Research Project-II	6

Punyashlok Ahilyadevi Holkar Solapur University Solapur

M. Sc. Part-II (Biostatistics) NEP-2020 Structure w.e.f 2024-25

M.Sc. (Biostatistics) Semester –III											
Paper Code	**Title of the Paper**	**Credits**	**Contact hours/week**			**Distribution of Marks for Examination**					
			Th (L)	**Pr**	**Total**	**Internal**		**External**		**Total**	
						Th	**Pr**	**Th**	**Pr**	**Th**	**Pr**
DSC-5	Statistical Genetics	4	4	---	4	40	---	60	---	100	---
DSC-6	Demography and Health Statistics	4	4	---	4	40	---	60	---	100	---
DSE-3	(A) Multivariate Statistical Methods / (B) Regression Analysis	4	4	---	4	40	---	60	---	100	---
LAB-7	Biostatistics Practical-VII	2	---	4	4	---	20	---	30	---	50
LAB-8	Biostatistics Practical-VIII	2	---	4	4	---	20	---	30	---	50
LAB-9	Biostatistics Practical-IX	2	---	4	4	---	20	---	30	---	50
RP-I	Research Project-I	4		8	8	---	40	---	60	---	100
Total for Semester-III		**22**	**12**	**20**	**32**	**120**	**100**	**180**	**150**	**300**	**250**
M.Sc. (Biostatistics) Semester -IV											
Code	**Title of the Paper**	**Credits**	**Contact hours / week**			**Distribution of Marks for Examination**					
			Th (L)	**Pr**	**Total**	**Internal**		**External**		**Total**	
						Th	**Pr**	**Th**	**Pr**	**Th**	**Pr**
DSC-7	Clinical Trials	4	4	---	4	40	---	60	---	100	---
DSC-8	Survival Analysis	4	4	---	4	40	---	60	---	100	---
DSE-4	(A) Data Mining / (B) Time Series Analysis	4	4	---	4	40	---	60	---	100	---
LAB-10	Biostatistics Practical-X	2	---	4	4	--	20	---	30	---	50
LAB-11	Biostatistics Practical-XI	2	---	4	4	--	20	---	30	---	50
RP-II	Research Project-II	6	---	12	12	---	60	---	90	---	150
Total for Semester-II		**22**	**12**	**20**	**32**	**120**	**100**	**180**	**150**	**300**	**250**

DSC: Discipline Specific Course **DSE:** Discipline Specific Elective **RP:** Research Project

Evaluation Scheme

Each theory paper will have 100 marks, out of which 60 marks will be for Term End examination and 40 marks for Internal Assessment. The candidate has to appear for internal evaluation of 40 marks and external evaluation (University Examination) of 60 marks for each theory paper.

Each practical paper will have 50 marks, out of which 30 marks will be for Term End examination and 20 marks for Internal Assessment. The candidate has to appear for internal evaluation of 20 marks and external evaluation (University Examination) of 30 marks for each practical paper.

Internal Evaluation

- In case of theory papers, internal examinations will be conducted by department / school.
- In case of practical papers, 5 marks shall be for day-to-day journal and 15 marks shall be for internal test, which will be conducted by the department / school.

External Evaluation (End of Term University Examination)

I) Nature of Theory question paper:

1) Each Theory paper is of 60 marks.
2) Each Theory paper will be of 2.5 hours.

II) Nature of Practical question paper: (End of Term Examination)

Sem-III and IV: Practical examination will be conducted for 30 marks and is of two hours duration. There shall be 05 questions each of 8 marks, of which student has to attempt any 03 questions. VIVA will be for 6 marks.

Equivalence for Theory Papers:

Semester No.	Old Syllabus		New Syllabus	
	Paper Code	**Title of the Paper**	**Paper Code**	**Title of the Paper**
III	HCT 3.1	Statistical Inference-II	DSC-4	Statistical Inference-II
	HCT 3.2	Micro-array Data Analysis	---	No equivalence
	SCT 3.1	Multivariate Statistical Methods	DSE-3 (A)	Multivariate Statistical Methods
	SCT 3.2	Research Ethics	RM	Research Methodology in Biostatistics
	OET 3.1	Applied Statistics	---	No equivalence
	OET 3.2	Modeling and Simulation	---	No equivalence
IV	HCT 4.1	Demography and Health Statistics	DSC-6	Demography and Health Statistics
	HCT 4.2	Clinical Trials	DSC-7	Clinical Trials
	HCT 4.3	Survival Analysis	DSC-8	Survival Analysis
	SCT 4.1	Time Series Analysis	DSE-4 (B)	Time Series Analysis
	SCT4.2	Data Mining	DSE-4 (A)	Data Mining

M.Sc. (Biostatistics) Semester -III

Major Mandatory	DSC-5: STATISTICAL GENETICS	Credits: 04

Course Outcomes:
Upon successful completion of this course, the student will be able to:

1. Understand the basics of genetics.
2. Understand the random and non-random mating.
3. Understand and analyze family data.
4. Analyze population genetics using statistical techniques.

Unit-1: Basic biological concepts in genetics, Mendel's law, Hardy Weinberg equilibrium, estimation of allele frequency (dominant/co-dominant cases), Approach to equilibrium for X-linked gene. Law of natural selection, mutation, genetic drift. (15L)

Unit-2: Non-random mating ,inbreeding, phenotypic assortative mating. I. T. O. matrices, identity by descent. (15L)

Unit-3: Family data-estimation of segregation ratio under ascertainment bias, pedigree data: Elston - Stewart algorithm for calculation of likelihoods. Linkage, estimation of re-combination fraction, inheritance of quantitative traits. (15L)

Unit-4: Population Genetics: Random mating, genetical variance and correlations, multiple alleles and blood types, maximum likelihood method of estimation, Sex linked genes, Autopolyploid. (15L)

Reference Books:

1. Sharad Gogate: Preventive Genetics- JAYPEE Publishers.
2. Agarwal B. L. and Agarwal S. P. : Statistical Analysis of Quantitative Genetics- New Age International Publishers
3. Liu B.H. (1998). Statistical Genomics, CRC Press, New York.
4. Falconer D.S. (1970). Introduction to Genetics, Oliver and Boyd.

Major Mandatory	DSC-6: DEMOGRAPHY AND HEALTH STATISTICS	Credits: 04

Course Outcomes:

Upon successful completion of this course, the student will be able to:

1) Understand basic demographic concepts, components of population change.
2) Identify different sources for obtaining demographic data.
3) Learn and differentiate among different population theories.
4) Model the population growth and obtain projection for future population.

Unit-1: Definition and Scope of Demography: Demography as a scientific discipline; Development of demography as a discipline; Basic demographic concepts; Components of population change. Contribution of fertility, mortality and migration to population change in the past; major sources of data about the population in the past; major explanations of population change in the past; Components of population changes fertility mortality and migration. (15L)

Unit-2: Sources of Demographic Data: Population census; Uses and limitations; Indian Censuses; Vital registration system, National Sample Survey; Sample Registration System; Demographic Health Surveys (DHS), and other sample surveys. Demographic health surveys National family health survey first and second, National Rural Health Mission, Longitudinal Ageing Study in India (LASI). (15L)

Unit-3: Population theories: Malthusian and Neo-Malthusian theories, optimum theory of population, biological theories of population, social and economic theories of population: Dumont's theory Karl Marx theory of surplus population Leibenstein's motivation theory of population growth, criticism on all above theories. Theories on migration Lewis and Fei-Ranis model of rural urban migration. (15L)

Unit-4: Population Estimates and Projections: Concepts of population projections; population estimates, forecasts and projections, uses of projections; Methods of interpolation, extrapolation using linear, exponential, polynomial, logistics and Gompertz curves; Cohort component method: basic methodology; Life tables, projection of mortality, fertility and migration components; Population projections of United Nations, World Bank and Expert Committees of Government of India; Methods of rural-urban and sub-national population projections; Methods of related socio-economic projections: labour force, school-enrolment, health personnel and households. Population policies, India's population policies. (15L)

Reference Books:

1. Shryoch, Henry S, Jacob S, Siegel and Associates (1964):Methods and materials of demography (condensed edition) Academic press, London.
2. Barclay, George W.(1968):Techniques of population analysis, John Wiley and sons, New-York.
3. Keyfitz N. (1968):Introduction to Mathematics of Population. Addision-Wesley Publishing Co, Reading , Messachusetts.
4. Ramkumar R. (1986): Technical Demography, Wiley Eastern, New Delhi.
5. Sudhendu Biswas (1988): Stochastic processes in Demography and Applications, Wiley Eastern, New Delhi.

Major Elective	DSE-3(A): MULTIVARIATE STATISTICAL METHODS	Credits: 04

- Course Outcomes:

Upon successful completion of this course, the student will be able to:

1. Decide the distribution of linear combinations of multivariate normal distribution.
2. Use the tools of multivariate normal distribution for inference on population means.
3. Derive the important properties of multivariate normal distribution.
4. Analyze the multivariate data using data reduction techniques like principle component analysis and factor analysis.

Unit-1: **Multivariate Normal Distribution:** Eigen value & Eigen vector of a matrix. Multivariate data, sample mean vector, sample dispersion matrix, correlation matrix, partial and multiple correlations, correlation of linear transformation, multivariate normal distribution, random sampling from a multivariate normal distribution. Maximum likelihood estimators of the parameters of the multivariate normal distribution, sampling distributions of mean vector. (15L)

Unit-2: Hotelling's T^2 and Wishart Distribution: Hotelling's $T^2$and Mahalanobis D^2 statistics, applications in tests on mean vector for one and more multivariate normal populations and also on equality of the components of a mean vector in a multivariate normal population. Wishart distribution and its properties, distribution of generalized variance. (15L)

Unit-3: **Classification**: Discrimination procedures for discrimination between two multivariate normal populations, sample discriminant function, tests associated with discriminant functions, classification into more than two multivariate normal populations. (15L)

Unit-4: Principal Components Analysis: Introduction to principal component analysis as dimension reduction technique, canonical correlation, canonical variables, Cluster analysis: Hierarchical and Non-hierarchical clustering, single, complete, average linkage methods and k-means clustering. Introduction to factor analysis: Orthogonal factor model, estimation of factor loading, MLE and principal component method, rotation of factors. (15L)

Reference Books:

1. Anderson T. W. (1984): An Introduction to Multivariate Analysis, 2nd Ed., John Wiley.
2. Kshirsagar A. M. (1972): Multivariate Analysis, Marcel Dekker
3. Johnson and Dean W. Wichern (2002): Applied Multivariate Analysis, John Wiley.
4. Rao C. R. (1973): Linear Statistical Inference and Its Applications, 2nd Ed. Wiley.
5. Sharma S. (1996): Applied Multivariate Techniques, Wiley.
6. Srivastava M. S. and Khatri C. G. (1979): An introduction to multivariate statistics, North Holland.
7. Bhuyan, K. C. (2005): Multivariate Analysis and its Applications, New central Book Agency (P) Ltd. Kolkata.
8. Giri, N. C. (1977): Multivariate Statistical Inference, Academic Press.

Major Elective	DSE-3(B): REGRESSION ANALYSIS	Credits: 04

Course outcomes:

Upon successful completion of this course, the student will be able to:

1. Students learnt about different linear and non-linear regression models and their appropriate computational procedures.
2. They know R^2, adjusted R^2 and Cp criteria for model selection.
3. Implement variable selection methods to identify appropriate model for further analysis.
4. They will get the knowledge of building and fitting linear regression models with software.
5. They also learn about the theory underlying point estimation, hypothesis and confidence intervals for linear regression models.

Unit-1: Multiple Linear Regression Model: Least squares estimates (LSE), Properties of LSE, Hypothesis testing, Confidence and prediction intervals, General linear hypothesis testing. Residuals and their properties, Residual diagnostics. Transformation of variables: VST and Box-Cox power transformation. (15L)

Unit-2: Variable Selection Procedures: R-square, adjusted R-square, Mallow's Cp, forward, backward selection, stepwise selection methods. Multicollinearity: Consequences, detection and remedies, Ridge regression. Autocorrelation: causes, consequences, detection: Durbin-Watson test, Estimation of parameters in presence of autocorrelation: Cochrane-Orcutt method. (15L)

Unit-3: Nonlinear Regression Models: Difference between linear and nonlinear regression models, transformation to linear model, intrinsically linear and nonlinear models, Parameter estimation using the Newton-Gauss method. Polynomial models in one and two variables, orthogonal polynomials, smoothing splines: linear, quadratic, cubic, cubic-B. (15L)

Unit-4: Generalized Linear Models: Exponential families, Definition of Generalized linear model (GLM), Link function, Estimation of parameters and inference in GLM. Logistic Regression Model: Link function, Logit, probit, complementary log-log, estimation of parameters. Odds ratio, Hypothesis testing using model deviance. (15L)

Reference Books:

1) Draper N. R. and Smith H. (1998): Applied Regression Analysis, 3rd Ed. Wiley.
2) Wiesberg S. (1985): Applied Linear Regression, Wiley.
3) Kutner, Neter, Nachtsheim and Wasserman (2003): Applied Linear Regression Models, 4th Ed. McGraw-Hill.
4) Montgomery, D. C., Peck E. A. and Vining, G. (2001): Introduction to Linear Regression Analysis, 3rd Ed. Wiley.
5) Cook, R. D. and Wiesberg, S. (1982): Residuals and Influence in Regression, Chapman and Hall.
6) Seber, G. A. , Wild, C. J. (2003), Nonlinear Regression, Wiley.

LAB-7	BIOSTATISTICS PRACTICAL-VII (Based on DSC-5 and DSC-6)	Credits: 02

Course outcomes:

Upon successful completion of this course, the student will be able to:

1. Understand the genetics and heredity.
2. Learn about statistical tools applied in genetics.
3. Understand the difference between phenotype and genotype.
4. Analyse the population data and the projection of population growth.
5. Understand different factors affecting population change.

LAB-8	STATISTICS PRACTICAL-VIII (Practicals based on DSE-3)	Credits: 02

Course outcomes:

Upon successful completion of this course, the student will be able to solve numerical problems related to Multivariate Statistical Methods and Regression Analysis.

LAB-9	STATISTICS PRACTICAL-IX (Design of Experiments)	Credits: 02

Course Outcomes:

Upon successful completion of this course, the student will be able to:

1. Design and analyze two-level full factorial experiments.
2. Analyze un-replicated factorial experiments.
3. Design and analyze two-level confounded factorial experiments.

Unit-1: Analysis of variance: one way classification, two way classification without interaction and with interaction with equal number of observations per cell. Estimation and related tests of hypothesis, Analysis of Covariance: Estimation of parameters, related tests of hypothesis. General theory and application to one- way and two-way set up. Two way classification with unequal number of observations per cell (without interaction), connectedness, balancedness, orthogonality, related tests of hypothesis. BIBD: Definition, parametric relationship, inter and intra block analysis, Symmetric BIBD. (15L)

Unit-2: 2^n Factorial Experiments: Concepts of main effects, interaction: graphical representation, Analysis of full 2^n replicated and unreplicated factorial design. Confounding: Total and partial confounding, construction and analysis of confounded design. (15L)

RP-1	RESEARCH PROJECT-I	Credits: 04

Course Objective:

1. To familiarize students with the fundamentals of research.

2. To help students to make appropriate grammatical and lexical choices while writing research articles and organize information effectively.

3. To integrate theoretical research knowledge with practical skills that will help students to undertake research.

Guidelines regarding Literature Review:

1. Group of students: Maximum four students in a group.
2. Selection of topic for literature review: Student has to select a topic/area for literature review with the help of a guide allotted to the student. The topic must be relevant to current trends and advancements in Statistics.
3. Conducting Literature Review: Student has to use academic databases, journals, conference proceedings, reputed online sources etc. for searching the research articles / materials related to the selected topic / area. Student has to study at least 15 recent research articles related to the selected topic/area.
4. Prepare review articles: While studying these research articles, student has to consider the objectives, methodologies, comparative study of these methodologies, research gaps etc. used by various researchers. Student needs to submit a project report.

Evaluation Scheme:

i) Internal Evaluation: Mid-term presentation of project. 40 marks

ii) University Evaluation: University evaluation will be done by internal and external examiners for 60 marks, out of which, evaluation of 40 marks will be made based on the submitted project report and remaining 20 marks are for project presentation.

M.Sc. (Statistics) Semester-IV

Major Mandatory	DSC-7: CLINICAL TRIALS	Credits: 04

Course outcomes: Upon successful completion of the requirements for this course, students will be able to:

1. Understand the basic statistical principles, methods for clinical data analysis and reporting.
2. Demonstrate an understanding of the essential principles of modern biostatistical methods and statistical software and how to apply them.

Unit-1: Introduction to Clinical Trials: Introduction to Clinical Trials, need of ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of phase I-IV clinical trials. Classification of clinical trials, Multicenter clinical trials, Active control trials, Combination trials equivalence trials. Data Management: Data definition, case report forms database design, data collection system for good clinical practice, Clinical Data reliability and data validity. (15L)

Unit-2: Design of Clinical Trials: Parallel Vs cross-over designs, cross-sectional Vs longitudinal designs, review of factorial design, objective and endpoints of clinical trials, design of Phase I trials, design of single-stage and multi-stage Phase II trials, design and monitoring of Phase III trials with sequential stopping, design and analysis of bioequivalence trials. (15L)

Unit-3: Reporting and Analysis: Power and sample size calculation for Phase I-III trials, qualitative and quantitative data analysis, and time to event data analysis in clinical trials. (15L)

Unit-4: Surrogate endpoints: Selection and design of trials with surrogate endpoints, analysis of surrogate endpoint data. Meta-analysis of clinical trials. (15L)

Reference Books:

1. Piantadosi S. (1997): Clinical Trials: A Methodological Perspective. Wiley and Sons.
2. Wang D. and Bakhai A. (2006).Clinical Trials: A Practical Guide to Design, Analysis and Reporting, Andrew
3. Friedman L. M., Furburg C. and Demets D. L. (1998). Fundamentals of Clinical Trials, Springer Verlag.
4. Fleiss J. L.(1989): The Design and Analysis of Clinical Experiments. Wiley and Sons.
5. Marubeni E. and Valsecchi M. G. (1994). Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley and Sons.

Major Mandatory	DSC-8: SURVIVAL ANALYSIS	Credits: 04

Course outcomes:

Upon successful completion of the requirements for this course, students will be able to:

1. Understand the elements of reliability, hazard function and its applications.
2. Compute reliability of coherent systems, bounds on system reliability, Execute modular decomposition of coherent systems.
3. Students learnt about and survival analysis with their related distributions, relationships, non-parametric methods for computing survival analysis.
4. Explain the concept of censoring and know various types of censoring. Handle the censored data, techniques and tools to obtain survival probability.
5. Perform parametric analysis of different types of censored data and non-parametric estimate of survival function based on censored data.
6. Estimate nonparametric survival function of the data.

Unit-1: Structure function, dual of a structure, cuts and paths, components and systems, coherent systems, pivotal decomposition, coherent modules, modular decomposition, reliability concepts and measures, reliability of coherent systems, bounds on system reliability, Burnham's measure of structural importance, Associated random variables and their properties. (15L)

Unit-2: Life time distributions, survival functions, hazard rate, cumulative hazard function, residual life time, survival function of residual life time, mean residual life time. Computation of these functions for common life time distributions: Exponential, Weibull, Gamma, Pareto, Rayleigh, Lognormal distributions.
Notion of Ageing; IFR, IFRA, DMRL, NBU, NBUE, NWUE classes, ageing properties of common life time distributions, closures of these classes under formation of coherent structures, convolutions and mixtures of these classes. (15L)

Unit-3: Estimation and testing for Exponential, Gamma, Weibull, Lognormal, Pareto and Linear failure rate distributions for complete life data. Concept of censoring, various types of censoring, Estimation and Testing of parameters of exponential distribution under various types of censoring. (15L)

Unit-4: Estimation of survival function: Actuarial estimator, Greenwood's formula, Kaplan-Meier estimator, Estimation under the assumption of IFR/DFR., Test for exponentiality against nonparametric classes, Total time on test, Deshpande's test. Two-sample problem: Gehen's test, Log rank test, Mantel-Haenszel test, Tarone-Ware tests. (15L)

Reference Books:

1) Barlow R. E. and Proschan F. (1975): Statistical Theory of Reliability & Life Testing, Holt, Reinhart and Winston.
2) Lawless J. F. (1982): Statistical Models and Methods of Life Time Data, John Wiley.
3) Miller R. C. (1981): Survival Analysis, John Wiley.
4) Bain L.J. and Engelhardt (1991): Statistical Analysis of Reliability and Life testing Models, Marcel Dekker.
5) Deshpande, J. V. and Purohit, S. G. (2005): Life Time Data: Statistical Models and Methods, Word Scientific.
6) Lawless J. F. (1982): Statistical models and methods for failure time data, John Wiley.
7) Nelson W. (1982): Applied Life Data Analysis, John Wiley and Sons, Inc.

Major Elective	DSE-4 (A): DATA MINING	Credits: 04

Course outcomes:

Upon successful completion of this course, the student will be able to:

1. Differentiate between classical techniques and data oriented techniques.
2. Explain supervised and unsupervised learning.
3. Construct classifiers namely, decision tree, naïve Bayes, and k-nearest neighbor(s).
4. Compare different classifiers and employ techniques to improve their performance.
5. Apply artificial neural network model for classification and prediction.
6. Explain support vector machine (SVM) for classification and regression.
7. Generate association rules using apriori algorithm.
8. Apply clustering techniques, k-mediods, CLARA, DBSCAN, DENCLUE, probability model based clustering algorithm to form meaningful clusters.

Unit-1: Data preparation for knowledge discovery: Data understanding and data cleaning tools, Data transformation, Data Discretization, Data Visualization, Imbalanced data, Data Mining Process: CRISP and SEEMA; Concept of training data, testing data and validation of model. (15L)

Unit-2: Supervised Learning techniques: Problem of classification, classification techniques: k-nearest neighbor, decision tree, Naïve Bayesian, Classification based on logistic regression. (15L)

Unit-3: Artificial Neural Network (ANN): Introduction to ANN, types of activation function, McCulloch-Pitts AN model, single layer network, multilayer feed forward network model, training methods, ANN and regression models. Support vector machine: Introduction to support vector machine, loss functions, soft margin, optimization hyper plane, support vector classification, support vector regression, linear programming support vector machine for classification and regression. (15L)

Unit-4: Unsupervised learning: Density based methods and grid based methods for clustering. Market Basket Analysis, Association rules and prediction. Apriori algorithm, data attributes, applications to electronic commerce. (15L)

Reference Books:

1) Berson and Smith S. J. (1997): Data warehousing, Data mining and OLAP, McGraw Hill.
2) Breiman J. H., Friedman R. A., Olshen and Stone, C. J. (1984): Classification and Regression Trees, Wadsworth and Books/Cole.
3) Han and Kamber (2000): Data Mining: Concepts and Techniques, Morgan Gaufmann.
4) Mitchell T. M. (1997): Machine Learning, McGraw-Hill.
5) Ripley B. D. (1996): Pattern Recognition and Neural Networks, Cambridge University Press.

Major Elective	DSE-4(B): TIME SERIES ANALYSIS	Credits: 04

- **Course Outcomes:** Upon successful completion of the requirements for this course, students will be able to:
  1. Understand the concept of stationarity to the analysis of time series (TS) data in various contexts (such as actuarial studies, climatology, economics, finance, geography, meteorology, political science, and sociology).
  2. Identify stationarity/non-stationarity status of observed TS.
  3. Identify and isolate non deterministic components of observed TS; learn to translate an observed non-stationary series to stationarity TS series using an appropriate transformation.
  4. Model, estimate, interpret and forecast observed TS through ARMA and ARIMA

Unit-1: Time series as discrete parameter stochastic process, Auto-covariance, Auto-correlation functions and their properties. Exploratory Time series Analysis, Tests for trend and seasonality, Exponential and moving average smoothing. Holt-Winter smoothing, forecasting based smoothing. (15L)

Unit-2: Wold representation of linear stationary processes, Detailed study of the linear time series models: Autoregressive (AR), Moving average (MA), Autoregressive Moving Average (ARMA) models. Concept of causality, invertibility, Computation of π-weights and Ψ-weights, computation of ACVF and ACF. Partial auto covariance function, Autoregressive Integrated Moving Average (ARIMA) models. (15L)

Unit-3: Estimation of ARMA models: Yule-Walker estimation for ARMA processes, Discussion (without proof) of estimation of mean, Auto-covariance and auto-correlation function under large sample theory, Residual analysis and diagnostic checking, Forecasting using ARIMA models. (15L)

Unit-4: Analysis of seasonal models: Parsimonious models for seasonal time series, SARIMA models, forecasting, identification, estimation and diagnosis methods for seasonal time series. Introduction to ARCH and GARCH models. (15L)

Reference Books:

1. Box, G. E. P. and Jenkins, G. M. (1976): Time Series Analysis-Forecasting and control, Hodlen-day, San Francisco.
2. Brockwell, P. J. and Davis R. A. (1987): Time Series: Theory and Methods, 2nd Ed., Springer-Veriag.
3. Chatfield, C. (2004): The Analysis of Time Series-An Introduction, 6th Ed., Chapman and Hall.
4. Kendall, M. G. (1978): Time Series, Charler Graffin.
5. Montgomery, D. C. and Johnson, L. A. (1977): Forecasting and Time Series Analysis, McGraw Hill.
6. Fuller, W. A. (1996): Introduction to Statistical Time Series, John Wiley, New York.

LAB-10	STATISTICS PRACTICAL-X (Based on DSC-7 and DSC-8)	Credits: 02

Course outcomes:

Upon successful completion of this course, the student will be able to solve numerical problems related to Clinical Trials and Survival Analysis.

LAB-11	STATISTICS PRACTICAL-XI (Based on DSE-4)	Credits: 02

Course outcomes:

Upon successful completion of this course, the student will be able to solve numerical problems related to Data Mining and Time Series Analysis.

RP-2	Research Project-II	Credits:06

Course Objective:

- To enhance the practical knowledge and result analysis skills.
- To enable the students experience a real-life problem solving under the supervision of faculty members.
- To prepare the students perform functions that demand higher competence in national/international organizations.
- To train the students in scientific research.
- Develop research/ experimentation skills as well as enhancing project writing and oral presentation skills.

Evaluation Scheme:

i) Internal Evaluation: Mid-term presentation of project. 60 marks

ii) University Evaluation: University evaluation will be done by internal and external examiners for 90 marks, out of which, evaluation of 60 marks will be made based on the submitted project report and remaining 30 marks are for project presentation.

Guideline to prepare the Dissertation

An acceptable M.Sc. thesis in Statistics should attempt to satisfy one or more of the following criteria:

- Original research results are explained clearly and concisely.

- The thesis explains a novel exploratory implementation or a novel empirical study whose results will be of interest to the Statistics community in general and to a portion of the Statistics community in particular.
- Novel implementation techniques are outlined, generalized, and explained.
- Theoretical results are obtained, explained, proven, and (worst, best, average) case analysis is performed where applicable.
- The implementation of a practical piece of nontrivial software whose availability could have some impact on the Statistics community.

A good methodology to follow, immediately upon completion of the required courses, is to keep a paper or electronic research notebook and commit to writing research-oriented notes in it every day. From time to time, organize or reorganize your notes under headings that capture important categories of your thoughts. This journal of your research activities can serve as a very rough draft of your thesis by the time you complete your research. From these notes to a first M.Sc. thesis draft is a much less painful experience than to start a draft from scratch many months after your initial investigations. To help structure an M.Sc. thesis, the following guide may help.

Chapter 1. Introduction: This chapter contains a discussion of the general area of research which you plan to explore in the thesis. It should contain a summary of the work you propose to carry out. Describe the general problem that you are working towards solving and the specific problem that you attempt to solve in the thesis.

Chapter 2. Theory/Solution/Algorithm/Program: This chapter outlines your proposed solution to the specific problem described in Chapter 1. The solution may be an extension to, an improvement of, or even a disproof of someone else's theory / solution / method / ...).

Chapter 3 Description of Implementation or Formalism: This chapter describes your implementation or formalism. Depending on its length, it may be combined with Chapter 2. Not every thesis requires an implementation. Prototypical implementations are common and quite often acceptable although the guiding criterion is that the research problem must be clearer when you've completed your task than it was when you started!

Chapter 4 Results and Evaluation: This chapter should present the results of your thesis. You should choose criteria by which to judge your results, for example, the adequacy, coverage, efficiency, productiveness, effectiveness, elegance, user friendliness, etc., and then clearly, honestly and fairly adjudicate your results according to fair measures and report those results.

Chapter 5 Conclusions and future scope: This chapter should summarize the achievements of your thesis and discuss their impact on the research questions you raised in Chapter 1. If you solved the specific problem described in Chapter 1, you should explicitly say so here. If you did not, you should also make this clear. You should indicate open issues and directions for further or future work in this area with your estimates of relevance to the field, importance and amount of work required.

References : Complete references for all cited works. This should not be a bibliography of everything you have read in your area.

Appendices include technical material (program listings, output, graphical plots of data, detailed tables of experimental results, detailed proofs, etc.) which would disrupt the flow of the thesis but should be made available to help explain or provide details to the curious reader.

*****