PUNYASHLOK AHILYADEVI HOLKAR SOLAPUR UNIVERSITY, SOLAPUR

पुण्यश्लोक अहिल्यादेवी होळकर
सोलापूर विद्यापीठ
।। विद्यया संपन्नता ।।

NAAC Accredited-2015
'B' Grade (CGPA 2.62)

Name of the Faculty: Science & Technology

CHOICE BASED CREDIT SYSTEM

Syllabus: BIOSTATISTICS

Name of the Course: M.Sc. Part-II

(Syllabus to be implemented from w.e.f. June 2020)

PUNYASHLOK AHILYADEVI HOLKAR SOLAPUR UNIVERSITY

M.Sc. Part-II (BIOSTATISTICS) CBCS Structure (w.e.f. June 2020-21)

Semester	Code	Title of the Paper	Semester Examination			L	T	P	Credits
			Theory	IA	Total				
Sem-III	**Hard Core**								
	HCT 3.1	Statistical Inference-II	70	30	100	4	--	--	4
	HCT 3.2	Micro-array Data Analysis	70	30	100	4	--	--	4
	Soft Core Theory (Any one)								
	SCT 3.1	Multivariate Statistical Methods	70	30	100	4	--	--	4
	SCT 3.2	Research Ethics	70	30	100	4	--	--	4
	Open Elective Theory (Any One)								
	OET 3.1	Applied Statistics	70	30	100	4	--	--	4
	OET 3.2	Modeling and Simulation							
	Practical and Project Work								
	HCP 3.1	Statistical Data Analysis-IV (Based on HCT3.1 and HCT3.2)	35	15	50	--	--	3	2
	HCP 3.2	Statistical Data Analysis-V (Based on SCT3.1/SCT3.2)	35	15	50	--	--	3	2
	HCP 3.3	PROJECT WORK-I	35	15	50	--	--	3	2
	OEP 3.1	Statistical Computing (Based on OET3.1/OET3.2)	35	15	50	--	--	3	2
	Total for Semester-III		**420**	**205**	**625**	**--**	**--**	**--**	**25**
Sem-IV	**Hard Core**								
	HCT4.1	Demography and Health Statistics	70	30	100	4	--	--	4
	HCT4.2	Clinical Trials	70	30	100	4	--	--	4
	HCT4.3	Reliability and Survival Analysis	70	30	100	4	--	--	4
	Soft Core (Any one)								
	SCT4.1	Time Series Analysis	70	30	100	4	--	--	4
	SCT4.2	Data Mining	70	30	100	4	--	--	4
		Seminar/Tutorial/ Industrial Visit/ Field Tour	---	25	25	--	1	--	1
	Practical and Project Work								
	HCP4.1	Statistical Data Analysis-VI (Based on HCT4.1)	35	15	50	--	--	3	2
	HCP4.2	Statistical Data Analysis-VII (Based on HCT4.2)	35	15	50	--	--	3	2
	HCP4.3	Statistical Data Analysis-VIII (Based on SCT4.1 /SCT4.2)	35	15	50	--	--	3	2
	HCP4.4	PROJECT WORK-II	35	15	50	--	--	3	2
	Total for Semester-IV		**420**	**205**	**625**	**--**	**--**	**--**	**25**

HCT: Hard Core Theory **HCP: Hard Core Practical** **SCT: Soft Core Theory**
OET: Open Elective Theory **OEP: Open Elective Practical** **IA: Internal Assessment**
L: Lecture **T: Tutorials** **P: Practical**

Evaluation Scheme:

Each theory paper will have 100 marks out of which 70 marks will be for Term End examination and 30 marks for Internal Assessment. The candidate has to appear for internal evaluation of 30 marks and external evaluation (University Examination) of 70 marks for each paper. Each practical paper will have 50 marks out of which 35 marks will be for Term End examination and 15 marks for Internal Assessment. The candidate has to appear for internal evaluation of 15 marks and external evaluation (University Examination) of 35 marks for each practical paper.

Project Work :

The objective of the project work is to train the students to undertake projects individually. The projects shall enable the students to take up their own statistical study and to understand the application of statistical methods that they learned during the course. The project requires the students to synthesize the topics from the course into some theme of practical use. The project work will be assessed for 50 marks. There should be an internal assessment of 15 marks and external assessment of 35 marks. Internal assessment of project is based on primary preparation for the project like selection of topic, preparation of questionnaire, synopsis presentation, day-to-day reporting of the project work and project presentation before term end examination. The external evaluation of the Project work is followed by presentation of work including Project Report and Viva-Voce. Project report will be evaluated for 25 marks and 10 marks are reserved for project based VIVA.

Internal Evaluation:

- For theory papers, internal examination will be conducted by department / school.
- For practical papers, 05 marks shall be for day-to-day journal and 10 marks shall be for internal test, which will be conducted by the department / school.
- **Nature of Practical paper: (End of Term Examination)**

Practical examination of each paper will be conducted for 35 marks and is of two hours duration. There shall be 05 questions each of 10 marks, of which student has to attempt any 03 questions and 05 marks are reserved for VIVA.

M.Sc. Part-II (BIOSTATISTICS)
SEMESTER-III

Hard Core	HCT 3.1: STATISTICAL INFERENCE-II	No. of credits: 04

Unit-1: Introduction to hypothesis testing, simple and composite hypotheses, null and alternative hypotheses, two types of errors, critical region, significance level, size and power of the test, p-value and its interpretation. Large sample tests for single mean, single proportion, two means and two proportions. Small sample tests for means and variances based on chi-square, t and F distributions. Test of significance for correlation coefficient ($\rho = 0,\ \rho = \rho_0$) (one and two sample problem). (15L)

Unit-2: Neyman-Pearson Lemma (Statement only), Most powerful (MP) test for simple null against simple alternative hypothesis. Uniformly most powerful (UMP) tests for simple null hypothesis against one-sided alternatives and for one-sided null against one-sided alternatives in one parameter exponential family. Non-existence of UMP test for simple null against two sided alternatives in one parameter exponential family. Tests based on Binomial, Poisson and Normal distributions. (15L)

Unit-3: Sequential Tests: Basics of sequential testing, Wald's SPRT with illustrations, OC and ASN functions. Applications to Binomial, Poisson and Normal distributions. Likelihood ratio test (LRT) and its application to standard distributions. Goodness of fit test based on chi-square distribution and application to contingency tables. Non-parametric Tests: One and two sample problem; one sample tests: Sign test, Wilcoxon signed-rank test, run test. Two sample tests: Wald-Wolfowitz runs test, Mann-Whitney U test, Median test, Kolmogorov-Smirnov test, Spearman's rank correlation test, Kendall's rank correlation test, Kruskal-Wallis Test. (15L)

Unit-4: Interval Estimation: Concepts of confidence interval estimation, confidence coefficient, length of confidence interval, definition of pivot, construction of confidence intervals using pivots, confidence interval for the parameters of univariate normal, proportion, mean, difference of means. Large sample confidence intervals for binomial and Poisson parameters, Shortest length confidence intervals. (15L)

Reference Books:

1. Kale B.K. (1999): A first Course on Parametric Inference-Narosa
2. Rohatgi V.K. (1988): Introduction to Probability and Mathematical Statistics, Wiley Eastern Ltd. New Delhi. Student Edition.
3. Dudewicz E.J. and Mishra S.N.(1988): Modern Mathematical Statistics, Wiley Series
4. Lehman E.L. (1987): Theory of Testing of Hypotheses. Student Edition.
5. Srivastava M. and Srivastava N. (2009): Statistical Inference-Testing of Hypotheses, PHI Learning Pvt. Ltd.
6. Shanthakumaran A. (2001): Fundamentals of Testing of Hypotheses, Atlantic Publishers and Distributors.

Hard Core	HCT 3.2: MICRO-ARRAY DATA ANALYSIS	No. of credits: 04

Unit-1: Microarrays and Normalization techniques, Introduction to Biology relevant to microarray experiment, Microarray experimental set up and quantification of information available from microarray experiments, Data cleaning, transformation of data. Between array and within array normalization, quantile and LOWESS normalization, stage wise normalization, Concordance coefficient and its role in normalization. (15L)

Unit-2: Statistical Inference procedures in comparative experiments, Inference procedures for single channel microarray data, application of two sample t –test, Tests for validating assumptions of two sample t-test. Application of Welch test and Wilcoxon rank sum test, Inference procedures for two channel microarray data. Paired t–test, Tests for validating assumptions of paired t-test. Wilcoxon signed-rank test, Comparison of more than two types of mRNA samples in single channel or two channel microarray experiments. One way ANOVA F test, one way ANOVA Welch F test, Kruskal-Wallis test, pairwise t-test, pairwise Welch test and pairwise Wilcoxon rank sum test, Strip charts and its role to decide the profile of differentially expressed genes. (15L)

Unit-3: Multiple hypotheses testing and Principal component analysis Multiple hypotheses testing, Adjustments for multiple hypotheses testing, adjusted p-values, false discovery rate and its application to microarray data analysis. Principal component analysis for microarray data, scree plot, plot of scores to rectangular matrix and the concept of ballot, its application to microarray. (15L)

Unit-4: Cluster analysis and Logistic regression, Hierarchical cluster analysis of microarray data, K - means cluster analysis of microarray data, Application of logistic regression for microarray data, Concept of AIC and BIC and its role to identify marker genes. (15 L)

Reference Books:

1. Amartunga D. and Cabrera J. (2004): Exploration and Analysis of DNA Microarray and Protein Array Data, Wiley.
2. Deshmukh S.R. and Purohit S.G. (2007): Microarray Data: Statistical Analysis Using R, Narosa.
3. Draghici S. (2003): Data Analysis Tools for DNA Microarrays, Chapman and Hall/CRC.
4. Dov S. (2003). Microarray Bioinformatics, Cambridge University Press,
5. McLachlan G. J., Do K. A. and Ambroise C. (2004): Analyzing Microarray Gene Expression Data, Wiley.
6. Simon R.M , Korn E.L., McShane L.M., Radmacher M.D., Wright G.W. and Zhao Y. (2003): Design and Analysis of DNA Microarray Investigations, Springer.
7. Speed T. (2003): Statistical Analysis of Gene Expression Microarray Data, Chapman and Hall/CRC.

Soft Core	SCT 3.1: MULTIVARIATE STATISTICAL METHODS	No. of credits: 04

Unit-1: 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 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 and 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: 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.

Soft Core	SCT 3.2: RESEARCH ETHICS	No. of credits: 04

Unit-1: Introduction to Ethics and Professionalism, Importance of ethics in research. Ethical principles for scientific research, Principles of Ethics and Professionalism in Biostatistics, Ethical dilemmas, Ethical analysis to address ethical dilemmas. Illustration with case studies. (15 L)

Unit-2: Protecting Human and Animal Research Participants, Ethical Decision Making, Defining "Research Misconduct" and Conflicts of Interest.
Illustration with case studies. (15L)

Unit-3: Ethics in Clinical Trial Design, Ethical Issues in Data Analysis, Reproducibility in Science, Ethics in Big Data and Genetic Research. (15L)

Unit-4: Data Integrity and Data Stewardship (Managing, Storing, Sharing, and Securing Data and Research Records), Working in Industry and Commercializing Science. (15L)

Reference Books:

1. Shamoo A.E. and Resnik D. B. (2015): Responsible conduct of research (3rd edition), Oxford University Press, New York.
2. American Statistical Association. Ethical Guidelines for Statistical Practice (1999). Available at: http://www.amstat.org/about/ethicalguidelines.cfm
3. International Statistical Institute. Declaration on professional ethics. Available at: http://www.isi-web.org/images/about/Declaration-EN2010.pdf

Open Elective	OET 3.1: APPLIED STATISTICS	No. of credits: 04

Unit-1: Time series: Introduction, component of time series, measurement of trend by moving average and progressive average methods and graphical representation. Sampling Theory: Introduction, advantages of sampling. Simple random sampling and stratified random sampling. Sampling and non-sampling errors. (15L)

Unit-2: Index Numbers: Meaning and utility of index numbers, problems in construction of index numbers; Unweighted price and quantity index numbers using Aggregate method and Average of relatives method (A. M. and G. M. to be used as average); Weighted price and quantity index numbers using aggregate method: Laspeyre's, Paasche's, Fisher's Formulae, cost of living index numbers; Tests of Index numbers (time reversal and factor reversal tests); Illustrative Examples. (15L)

Unit-3: Vital Statistics: Introduction and need of vital statistics. Mortality rates: Crude Death Rate, Specific Death rate, Standardized Death rate, Fertility rates: Crude Birth Rate, General Fertility Rate, Age Specific Fertility Rate, Total Fertility Rate. Reproduction rates: Gross Reproduction Rate, Net Reproduction Rate, Simple Examples. (15L)

Unit-4: Statistical Quality Control (SQC): Concept and need of SQC, chance and assignable causes, control charts for variables (mean and range charts), control chart for attributes (p-chart), control chart for number of defectives (np chart), control chart for number of defects per unit (c-chart), demerit control chart. (15L)

Reference Books:

1. Gupta S.C. and Kapoor V.K.: Fundamentals of Applied statistics - Sultan Chand and Sons Publication
2. Agarwal B.L. : Programmed Statistics- New Age International Publishers.
3. Gupta S.C.: Fundamentals of Statistics- Himalaya Publishing House, Mumbai.
4. Goon A.M., Gupta A.K. and Dasgupta: Fundamentals of Statistics (vol. I & II), The world Calcutta Press Pvt. Ltd., Calcutta.

Open Elective	OET 3.2: MODELING AND SIMULATION	No. of credits: 04

Unit-1: Stochastic Models: Introduction, Poisson Process, Markov chains and applications. Queuing Models: Introduction, Queuing System, Elements of queuing system, birth and death process model, Queuing Models M/ M/1, M/M/C. (15L)

Unit-2: Network Analysis**:** Applications of PERT and CPM techniques, Network diagram representation, Rules for constructing the network diagram, Determination of critical path. (15L)

Unit-3: Simulation: Introduction, Uses of simulation, Steps in simulation study, Advantages and disadvantages of simulation, Simulation models- continuous and discrete simulations. Simulation Models: (Flow chart and/or algorithms): Monte-Carlo simulation, Simulation of queuing system, Simulation of PERT problems. (15L)

Unit-4: Random Number Generation **:** Introduction, Types of random numbers, Pseudo random number generator, Tests for random numbers, Techniques for generating random numbers, Inverse transformation technique, Generating random variates from Uniform, Bernoulli, Binomial, Exponential and Normal distributions. (15L)

Reference Books:

1. Allen Arnold O. (1978): Probability, Statistics and Queuing with Computer Science Applications, Academic Press.
2. Kishore Trivedi. (1982): Probability and Statistics with Reliability, Queuing with computer science Applications, Prentice Hall.
3. Geofrey Gordon (1999): System Simulation, PHI, Second Ed.
4. Narsingh Deo (1979): System Simulation with Digital Computer, PHI.
5. Fred Maryanski (1987): Digital Computer, Simulation, CBSPD.
6. Jerry Banks, John Carson, B. L. Nelson (1998): Discrete-Event Simulation. PHI, 2nd ed.
7. Taylor and Karlin, Stochastic Modeling, Academic Press.
8. Sharma J. K. (2003): Operations Research Theory and Applications, 2nd Ed. Macmillan

M.Sc. Part-II (BIOSTATISTICS) SEMESTER-III (Practical Papers)

Hard Core	HCP 3.1: STATISTICAL DATA ANALYSIS-IV	No. of credits: 02

This paper includes practical problems from papers HCT3.1 and HCT3.2. There will be at least eight practicals. Each practical should consist of problems to be solved using each of the following software EXCEL/ R/ MINITAB/ MATLAB/ SYSTAT wherever applicable.

Hard Core	HCP 3.2: STATISTICAL DATA ANALYSIS-V	No. of credits: 02

This paper includes practical problems from papers SCT3.1/ SCT3.2. There will be at least eight practicals. Each practical should consist of problems to be solved using each of the following software EXCEL/ R/ MINITAB/ MATLAB/ SYSTAT wherever applicable.

Hard Core	HCP 3.3: PROJECT WORK-I	No. of credits: 02

Project work consists of understanding the domain of the problem, formulation of the problem, collection of the relevant data, analysis of the data and report writing. Students are expected to design computer programs or use statistical software for computational purpose.

OPEN ELECTIVE	OEP 3.1: STATISTICAL COMPUTING	No. of credits: 02

This paper includes practical problems from papers OET3.1/OET3.2. There will be at least eight practicals. Each practical should consist of problems to be solved using Microsoft Excel or Scientific Calculator.

*****

M.Sc. Part-II (BIOSTATISTICS)
SEMESTER-IV

Hard Core	HCT 4.1: DEMOGRAPHY AND HEALTH STATISTICS	No. of credits: 04

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; relation between population change and other social and economic changes at the national and local levels. (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. Population Theories: By Malthus and Marx; Optimum population. (15L)

Unit-3: Logistic Regression: Logit, Probit and cloglog model for dichotomous data with single and multiple explanatory variables, ML estimation, large sample tests about parameters. Hosmer-Lemshow test, ROC curve. Multilevel logistic regression, Logistic regression for Nominal response: Baseline Category model and ordinal response: Proportional odds model. (15L)

Unit-4: Population Estimates and Projections : Concepts of population projections; population estimates, forecasts and projections, uses of population projections; Methods of interpolation, extrapolation using linear, exponential, polynomial, logistics and Gompertz curves; Cohort component method: basic methodology; 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. (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.

Hard Core	HCT 4.2: CLINICAL TRIALS	No. of credits: 04

Unit-1: 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. (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.

Hard Core	HCT 4.3: RELIABILITY AND SURVIVAL ANALYSIS	No. of credits: 04

Unit-1: Reliability concepts and measures, components and systems, coherent systems, reliability coherent systems, cuts and paths, modular decomposition, bounds on system reliability, structural and reliability importance components. Life time distributions, survival function, hazard rate, hazard function, Residual life time, mean residual life time, one-one correspondence of these functions, Computation of these functions for common life time distributions: Exponential, Weibull, Gamma, Pareto, Rayleigh, Lognormal distributions. (15L)

Unit-2: Notions of Ageing, Increasing failure rate (IFR), Increasing failure rate average (IFRA), New better than used (NBU), Decreasing mean residual life (DMRL) and New better than used in expectation (NBUE) classes and their duals, Bathtub failure rate, Ageing properties of common life time distributions, closures of these classes under formation of coherent systems, convolutions and mixtures. (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. (15 L)

Unit-4. Estimation of survival function: Actuarial estimator, Kaplan-Meier estimator, properties: self-consistency and MLE, estimation under the assumption of IFR / DFR. Concept of TTT transform and its applications, Test for exponentiality against alternatives IFRA, NBU and NBUE. Two-sample problem: Gehen's test, Log rank test, Mantel-Haenszel test. (15 L)

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

Hard Core	SCT 4.1: TIME SERIES ANALYSIS	No. of credits: 04

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.

Soft Core	SCT 4.2: DATA MINING	No. of credits: 04

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.

M.Sc. Part-II (BIOSTATISTICS) SEMESTER-IV (Practical Papers)

Hard Core	HCP 4.1: STATISTICAL DATA ANALYSIS-VI	No. of credits: 02

This paper includes practical problems from paper HCT 4.1. There will be at least eight practicals. Each practical should consist of problems to be solved using each of the following software EXCEL/ R/ MINITAB/ MATLAB/ SYSTAT wherever applicable.

Hard Core	HCP 4.2: STATISTICAL DATA ANALYSIS-VII	No. of credits: 02

This paper includes practical problems from paper HCT 4.2. There will be at least eight practicals. Each practical should consist of problems to be solved using each of the following software EXCEL/ R/ MINITAB/ MATLAB/ SYSTAT wherever applicable.

Hard Core	HCP 4.3: STATISTICAL DATA ANALYSIS-VIII	No. of credits: 02

This paper includes practical problems from papers SCT 4.1/SCT4.2. There will be at least eight practicals. Each practical should consist of problems to be solved using each of the following software EXCEL/ R/ MINITAB/ MATLAB/ SYSTAT wherever applicable.

Hard Core	HCP 4.4: PROJECT WORK-II	No. of credits: 02

Project work consists of understanding the domain of the problem, formulation of the problem, collection of the relevant data, analysis of the data and report writing. Students are expected to design computer programs or use statistical software for computational purpose.

*****