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: Statistics

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

(Syllabus to be implemented from June 2024)

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

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

M.Sc. Statistics Semester-I			
Course Type	**Course code**	**Course Title**	**No. of Credits**
Major Mandatory	DSC-1	Distribution Theory	4
	DSC-2	Estimation Theory	4
	LAB-1	Statistical Practical-I	2
	LAB-2	Statistics Practical-II	2
Major Elective	DSE-1(A)	Statistical Mathematics	4
	DSE-1(B)	Demography	
	DSE-1(C)	Official Statistics	
	LAB-3	Statistics Practical-III	2
Research Methodology	RM	Research Methodology in Statistics	4
M.Sc. Statistics Semester-II			
Course Type	**Course code**	**Course Title**	**No. of Credits**
Major Mandatory	DSC-3	Stochastic Processes	4
	DSC-4	Theory of Testing of Hypotheses	4
	LAB-4	Statistics Practical-IV	2
	LAB-5	Statistics Practical-V	2
Major Elective	DSE-2 (A)	Probability Theory	4
	DSE-2 (B)	Nonparametric Methods	
	DSE-2 (C)	Modeling and Simulation	
	LAB-6	Statistics Practical-VI	2
On Job Training/ Field Project	OJT/FP	-----	4
M.Sc. Statistics Semester-III			
Course Type	**Course code**	**Course Title**	**No. of Credits**
Major Mandatory	DSC-5	Multivariate Analysis	4
	DSC-6	Regression Analysis	4
	LAB-7	Statistics Practical-VII	2
	LAB-8	Statistics Practical-VIII	2
Major Elective	DSE-3 (A)	Design and Analysis of Experiments	4
	DSE-3 (B)	Time Series Analysis	
	DSE-3 (C)	Actuarial Statistics	
	LAB-9	Statistics Practical-IX	2

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

Punyashlok Ahilyadevi Holkar Solapur University Solapur

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

M.Sc. (Statistics) 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	Multivariate Analysis	4	4	---	4	40	---	60	---	100	---
DSC-6	Regression Analysis	4	4	---	4	40	---	60	---	100	---
DSE-3	(A) Planning and Analysis of Industrial Experiments (B) Time Series Analysis (C) Actuarial Statistics	4	4	---	4	40	---	60	---	100	---
LAB-7	Statistics Practical-VII	2	---	4	4	---	20	---	30	---	50
LAB-8	Statistics Practical-VIII	2	---	4	4	---	20	---	30	---	50
LAB-9	Statistics Practical-IX	2	---	4	4	---	20	---	30	---	50
RP-1	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. (Statistics) 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	Industrial Statistics	4	4	---	4	40	---	60	---	100	---
DSC-8	Reliability and Survival Analysis	4	4	---	4	40	---	60	---	100	---
DSE-4	(A) Data Mining (B) Clinical Trials (C) Artificial Intelligence	4	4	---	4	40	---	60	---	100	---
LAB-10	Statistics Practical-X	2	---	4	4	--	20	---	30	---	50
LAB-11	Statistics Practical-XI	2	---	4	4	--	20	---	30	---	50
RP-2	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	Asymptotic Inference	---	No equivalence
	HCT 3.2	Multivariate Analysis	DSC-5	Multivariate Analysis
	HCT 3.3	Planning and Analysis of Industrial Experiments	DSE-3 (A)	Planning and Analysis of Industrial Experiments
	SCT 3.1	Regression Analysis	DSC-6	Regression Analysis
	SCT 3.2	Official Statistics	---	No equivalence
	OET 3.1	Applied Statistics	---	No equivalence
	OET 3.2	Modeling and Simulation	---	No equivalence
IV	HCT 4.1	Data Mining	DSE-4 (A)	Data Mining
	HCT 4.2	Industrial Statistics	DSC-7	Industrial Statistics
	HCT 4.3	Reliability and Survival Analysis	DSC-8	Reliability and Survival Analysis
	HCT4.4	Optimization Techniques	---	No equivalence
	SCT 4.1	Time Series Analysis	DSE-3(B)	Time Series Analysis
	SCT4.2	Clinical Trials	DSE-4(B)	Clinical Trials

M.Sc. (Statistics) Semester -III

Major Mandatory	DSC-5: MULTIVARIATE ANALYSIS	Credits: 04

Course Outcomes:

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

1. Use multivariate distributions and analysis techniques appropriately.
2. Undertake multivariate hypothesis tests, and draw appropriate conclusions.
3. Analyze multivariate data, dimension reduction and the dependence structure of variables to extract the useful information from multivariate dataset
4. Apply suitable tools for classification to formulate and solve real –life problems.
5. Apply suitable tools for exploratory data analysis and clustering techniques to extract the useful information from real-life problems.

Unit-1: Multivariate Normal Distribution: Sample mean vector, sample dispersion matrix, correlation matrix, Two definitions and their equivalence, singular and non-singular normal distribution, characteristic function, moments, marginal and conditional distributions. Maximum likelihood estimator of the parameters of the multivariate normal distribution and their sampling distributions. Partial and multiple correlation coefficients. Null distribution of sample correlation coefficient and sample multiple correlation coefficient. (15L)

Unit-2: Wishart matrix and its distribution, properties of Wishart distribution, Distribution of generalized variance. Hotelling's T^2 statistic and its null distribution. Applications of T^2 statistic and its relationship with Mahalanobis D^2 statistic, Confidence region for the mean vector. (15L)

Unit-3: Discrimination and Classification, Fisher's discriminant function and likelihood ratio procedure, Minimum Expected Cost of Misclassification (ECM) rule, Rao's U statistics and its use in tests associated with discriminant function. Introduction to principal component analysis as dimension reduction technique, canonical correlation, canonical variables. (15L)

Unit-4: Cluster and Factor Analysis: Hierarchical and Non-hierarchical clustering, single, complete, average linkage method and k-means clustering. Introduction to factor analysis, orthogonal factor model, estimation of factor loading, MLE and principle 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 Mandatory	**DSC-6: 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.

Major Elective	DSE-3(A): PLANNING AND ANALYSIS OF INDUSTRIAL EXPERIMENTS	Credits: 04

Course Outcomes:

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

1. Design and analyze two-level and three-level full factorial experiments.
2. Analyze un-replicated factorial experiments.
3. Design and analyze two-level and three-level confounded factorial experiments.
4. Design and analyze two-level and three-level fractional factorial experiments.
5. Design, analyze and interpret first and second order response surface 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, their graphical representation, Analysis of full 2^n replicated and unreplicated factorial design. Concept of confounding: Total and partial confounding, construction and analysis of confounded design. (15L)

Unit-3: 3^n Factorial Experiments: Concepts of main effects, interaction, their graphical representation, linear and quadratic components. Analysis of Full 3^n replicated and unreplicated factorial design. Confounding: Construction and analysis of 3^n confounded design. (15L)

Unit-4: Fractional Replication for Symmetric Factorial: Concept of generator, defining contrast, aliasing, resolution and minimum aberration, construction and analysis of 2^{n-k} and 3^{n-k} fractional designs. Introduction to Response Surface Methodology (RSM). (15L)

Reference Books:

1) Jeff Wu C. F., Hamada M. (2000): Experiments: Planning, Analysis and Parameter design optimization, John Wiley & Sons.
2) Phadke, M.S. (1989): Quality Engineering using Robust Design, Prentice-Hall.
3) Montgomery D.C. (2001): Design and Analysis of Experiments, 5^{th} edition, Wiley New York.
4) Angela Dean and Daneil Voss (1999): Design and Analysis of Experiments, Wiley.
5) Das, M. N. and Giri, N. (1979), Design and Analysis of Experiments, Wiley Eastern.

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, inevitability, 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.

Major Elective	DSE-3(C): ACTUARIAL STATISTICS	Credits: 04

- **Course Outcomes:**

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

1. To understand how actuarial science is used in finance, investments, banking and insurance.
2. Explain the concept of survival models
3. Describe estimation procedures for lifetime distributions.
4. To understand the statistical behaviour of actuarial indicators.
5. To solve the problems related to the benefit amounts in insurance, annuities, premiums and reserves.

Unit-1: Future life time random variable, its distribution function and density function, concept of force of mortality, curtate future life time random variable its probability mass function, deferred probabilities, all these functions in terms of international actuarial notation. Analytical laws of mortality such as Gompertz' law and Makeham's law, single decrement life table, select and ultimate life table. (15 L)

Unit-2: Concept of compound interest rate, discount factor, present value of the money, nominal rate of interest, force of interest. Assurance contracts with level and varying benefits, such as whole life insurance, term insurance endowment insurance. Means and variances of the present value random variables of the payments under these contracts under the assumption of constant force of interest, when the benefit payments are made at the end of year (discrete set up) or when it is paid at the epoch of death(continuous set up). Actuarial present value of the benefit. Net single premiums. (15 L)

Unit-3: Annuity contracts, annuity certain, discrete annuity, m-thly annuity, continuous annuity, deferred annuity, present values and accumulated values of these annuities. Continuous and discrete life annuity, such as whole life annuity, temporary life annuity, n-year certain and life annuity, life annuities with mthly payments. Present value random variables for these annuity payments, means and variances. Actuarial present value of the annuity. (15 L)

Unit-4: Loss at issue random variable, various principle1s to decide net premiums for insurance products and annuity schemes defined in unit II and III, fully continuous premiums and fully discrete premiums, True m-thly payment premiums. Extended equivalence principle to decide gross premiums. Concept of reserve, prospective & retrospective approach. Fully continuous reserve. Fully discrete reserve. (15 L)

Reference Books:

1) Bowers, JR. N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. (1997). Actuarial Mathematics, 2nd Edn., The Society of Actuaries.
2) Deshmukh S.R. (2009). Actuarial Statistics: An Introduction Using R, Universities Press.

3) Harriett, E.J. and Dani, L. L.(1999). Principles of Insurance: Life,Health, and Annuities, 2nd Edn., Life Office Management Association.
4) Neill, Alistair (1977). Life Contingencies, The Institute of Actuaries.
5) Palande, P. S., Shah, R. S. and Lunawat, M. L. (2003). Insurance in India - Changing Policies and Emerging Opportunities, Response Books.

LAB-7	STATISTICS 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. Apply multivariate data analysis tools such as Multiple linear regression, Principle component analysis, Canonical correlations, Discriminant analysis, Cluster and Factor analysis.
2. Summarize and interpret multivariate data using appropriate multivariate methods to analyze data with statistical software.
3. Apply multivariate techniques appropriately, undertake multivariate hypothesis tests and draw appropriate conclusions.
4. Perform statistical tests and construct statistical intervals in a multiple linear regression set up.
5. Apply various statistical tests to determine the acceptability of a fitted model.

List of Practicals	
1	Multivariate Analysis
2	Maximum Likelihood Estimation of μ and Σ
3	Hotelling's T^2 Statistic
4	Discriminant Analysis
5	Dimension Reduction Techniques
6	Multiple Linear Regression
7	Residual Analysis
8	Testing the significance of regressors and ANOVA.
9	Best subset selection based on MSE, R^2 and Mallows Cp-criterion.
10	Testing for autocorrelation and fitting auto-correlated model.

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:
Students know about the solving of numerical problems related to planning and analysis of industrial experiments, time series analysis and actuarial statistics.

LAB-9	STATISTICS PRACTICAL-IX (Optimization Techniques-II)	Credits: 02

Course outcomes:

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

1. Formulate a problem as an appropriate optimization problem (IPP, QPP)
2. Apply various methods to obtain optimum solution of IPP and QPP.
3. Solve two person zero sum games with pure and mixed strategies using various methods.
4. Explain, formulate and solve dynamic Programming problem.

Unit 1: Integer Linear Programming Problem (ILPP): The concept of cutting plane, cutting plane method for all ILPP and mixed ILLP, Branch and Bound method. Quadratic programming: Kuhn-Tucker conditions, methods due to Beale, Wolfe.

Unit 2: Theory of Games: two person zero sum games, minimax and maximin principles, Saddle point, mixed strategies; rules of dominance, solution of 2×2 game by algebraic method, Graphical method, Reduction of the game problem as LPP. Dynamic Programming: The Recursion Equation Approach, Computational Procedure, Characteristics of Dynamic Programming, Solution of L.P.P. by Dynamic Programming.

Practical List:

- Game Theory.
- Quadratic programming
- Integer programming.
- Dynamic Programming

References:

1) Hadley G. (1969): Linear Programming, Addison Wesley
2) Taha H. A. (1971): Operation Research: An Introduction, Macmillan N.Y.
3) Kanti Swaroop and Gupta M. M.(1985): Operations Research, Sultan Chand & Co. ltd.
4) P. Gupta and D. S. Hira (2010): Operation Research, Sultan Chand & Co. ltd.
5) J. K. Sharma. (2003): Operation Research: Theory and Applications. Macmillan.

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: INDUSTRIAL STATISTICS	Credits: 04

Course outcomes:

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

1. Can identify the cause of defects using statistical quality management techniques.
2. Able to apply statistical quality control techniques to minimize the variability in manufacturing and business process.
3. Perform process capability analysis.
4. Acquainted with Six Sigma and lean thinking in industrial experimentation.
5. Design and implement sampling inspection plans.
6. Expertise in the most import field of applied statistics that contributes to quality control in all most all industries.

Unit-1: Concept of quality and quality control, Basic concept of process monitoring and control, Process control and product control, Seven SPC tools, Deming's PDCA cycle for continuous improvements and its applications, General theory of Control charts, Different types of limits: Natural tolerance limits, Specification limits, Control limits, Warning limits, Performance measures of a control chart: ARL, ATS, OC. Control charts for measurement and attributes: $\overline{X}$, R, S, S^2, X-MR charts, p, np, c, u, charts, Demerit chart. (15L)

Unit-2: Memory type control charts: CUSUM chart, tabular form and V-mask. Moving and exponentially weighted moving average charts, use of these charts for process control. Multivariate control charts for measurements data, Hotelling's T^2 control charts. (15L)

Unit-3: Process Capability Analysis: Process capability indices (PCI)C_p, C_{pk} and C_{pm} under normal distribution of quality characteristics, connection between proportion of nonconforming and C_p, C_{pk}. Estimation, confidence interval and testing hypotheses about C_p. Process capability analysis for non-normal data. Introduction to Six-sigma methodology, DMAIC Cycle and case studies. (15L)

Unit-4: Acceptance Sampling Plans: Consumer's risk, Producer's risk, LTPD, AOQ. Sampling plans by attributes, single, double and sequential plans, Sampling by variables for one-sided and two-sided specifications. (15L)

Reference Books:

1) Guenther W. C. (1981): Sampling Inspection in Statistical Quality Control, Charter Grifits.
2) Montgomery D. C. (1996): Introduction to Statistical Quality Control, John Wiley & Sons
3) Kotz S. (1993): Process Capability Indices, Chapman and Hall.
4) Abraham Boyas (1998): Quality Improvements through Statistical Methods, Birkhauser.
5) Mittag H. J. and Rinne H. (1993): Statistical methods Quality Assurance, Chapmam and Hall.

Major Mandatory	**DSC-8: RELIABILITY AND 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) Miller R. C. (1981): Survival Analysis, John Wiley.
3) Bain L.J. and Engelhardt (1991): Statistical Analysis of Reliability and Life testing Models, Marcel Dekker.
4) Deshpande, J. V. and Purohit, S. G. (2005): Life Time Data: Statistical Models and Methods, Word Scientific.
5) Lawless J. F. (1982): Statistical models and methods for failure time data, John Wiley.
6) 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): 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 bio-statistical methods and statistical software and how to apply them.

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.

Major Elective	DSE-4(C): ARTIFICIAL INTELLIGENCE	Credits: 04

- **Course Outcomes:**

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

1. Understand application of statistics in the field of Artificial Intelligence.
2. Understand different algorithms of machine learning such as ANN, Genetic Algorithm.
3. Understand difference between database and knowledge base.
4. Apply ANN to real life data datasets to extract knowledge out of it.

Unit 1: AI Problem Solving: Historical development of AI, Role of heuristic in problem solving. Knowledge representation and inference. Comparison of database knowledgebase. Expert Systems: Expert Problems. Predicate logic. Fact- Table. Rule base. Fuzzy Logic. Case based reasoning. Design of fuzzy rule base. Construction and implementation of knowledgebase systems. (15L)

Unit 2: Artificial Neural Network: Signal processing in biological and artificial neurons. ANN architectures. Perceptron learning. Multilayer Perceptron: Back Propagation Algorithm, XOR Problem, Heuristics, Output Representation and Decision Rule, Training and implementation of a neural network. (15)

Unit 3: Genetic Algorithm: History and evolution of G.A. Modeling a problem for the application of G.A. Representation of data in chromosomes. GA operators: Encoding, Crossover, Selection, Mutation, etc. Fitness function. Reproduction and convergence. Comparison of ANN and GA. Application of G.A. (15)

Unit 4: Natural Language Processing: Text categorization. Text summarization and text elaboration. Vision and perception. Image analysis and pattern matching. Robotics. (15)

References:

1. S. Rajsekaran ,G. A. Vijaylaxmi Pai: Neural Networks. Fuzzy Logic and Generic Algorithms. Synthesis and Applications. (EEE)
2. David Goldberg: Genetic Algorithms (Addison and Wesley)
3. David Rolston: Principles of AI and Expert System Development (MGII)
4. E. Ritch and K.Knight: Artificial Intelligence (MGII)
5. Mehrika, K., Mohan, C., and Ranka (1997) Elements of Artificial neural networks. Penram International.
6. Chattamvelli, R. (2015). Data mining methods. Alpha Science.

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

- **Course Outcomes:**

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

Students can solve numerical and real life problems related to DSC-7 and DSC-8 of semester –IV theory papers.

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

- **Course Outcomes:**

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

Students can solve numerical and real life problems related to DES-4 of semester–IV theory papers.

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.

*****