MASTER OF SCIENCE

MSc STATISTICS with DATA SCIENCE

PROGRAM STRUCTURE AND SYLLABUS
2020-21 ADMISSIONS ONWARDS

(UNDER MAHATMA GANDHI UNIVERSITY PGCSS REGULATIONS 2019)

EXPERT COMMITTEE FOR M.Sc. STATISTICS
MAHATMA GANDHI UNIVERSITY

2020

ACKNOWLEDGEMENT

The Expert Committee for M.Sc. Statistics with specialization in Applied Algorithms, Statistical Techniques in data Mining, Analysis of Multi-type Data, Statistical Modelling and Non-parametric Statistics (later renamed as MSc Statistics with Data Science) is grateful to the members for their constructive and creative role in the preparation of the curriculum and syllabus - The Expert Committee also gratefully acknowledges the contribution of the members for the finalization of the syllabus.

I would like to place on record my sincere gratitude to Dr. K.K. Jose, Honourable Director, School of Mathematics and Statistics, Mahatma Gandhi University, Kottayam, for providing all academic support as subject expert. My sincere thanks are due to the expert committee members Dr. Gijo EV, Professor, Indian Statistical Institute, Bangalore, Dr.S.M.Sunoj, Professor, Department of Statistics, Cochin University of Science and Technology, Dr.E.I.Abdul Sathar, Professor, Department of Statistics, University of Kerala, Dr.Sudheesh Kumar Kattumannil, Associate Professor, Indian Statistical Institute, Chennai, Dr.Angel Mathew, Assistant Professor, Department of Statistics, Maharajas College, Ernakulam, Dr.Mini L., Associate Professor, Department of Statistics, Sree Sankara College, Kalady, Ligi George, Associate Professor of Statistics, St. Thomas College, Kozhencherry, Dr.Stephy Thomas, Assistant Professor of Statistics, B.C.M. College, Kottayam , Dr.Sooraj K Ambat, Scientist E, Naval Physical and Oceanographic Laboratory, Defence R & D Organization, Ministry of Defence, Thrikkakara, Kochi, Prasanth V P, Manager - Biostatistics, IQVIA, Infopark, Kochi, Dr.Biju Thomas, Associate Professor, Department of Statistics, Sree Sankara College, Kalady. and all the university officials for their endless support. I express my sincere thanks to **Dr. Varghese K Cherian**, Member, Syndicate for all the help he has rendered. Above all I express my wholehearted thanks to Abin John, Assistant

Professor of Statistics, St. Thomas College, Kozhencherry, secretary of the expert committee for his support.

We, the Expert Committee for M.Sc. Statistics with specialization in Applied Algorithms, Statistical Techniques in data Mining, Analysis of Multi-type Data, Statistical Modelling and Non-parametric Statistics (MSc Statistics with Data Science), express our sincere thanks to all who have been helping for the success to this endeavour academically and administratively.

The expert committee acknowledges the contribution of the academic section AcAIX M.G.University.

Benno Mathew,
Associate Professor, Sree Sankara College, Kalady,
Convenor, Expert Committee for M.Sc. Statistics with Data Science

30 December 2020

Expert Committee

1. Benno Mathew (Convenor)
   Associate Professor, Department of Statistics,
   Sree Sankara College, Kalady

2. Dr Gijo EV
   Professor, Statistics Quality Control & Operations Research Unit
   Indian Statistical Institute, Bangalore

3. Dr S. M. Sunoj,
   Professor, Department of Statistics, CUSAT

4. Dr. E.I. Abdul Sathar,
   Professor, Department of Statistics,
   University of Kerala, Thiruvananthapuram

5. Dr Sudheesh Kumar Kattumannil,
   Associate Professor, Indian Statistical Institute,
   Chennai Centre

6. Dr. Biju Thomas,
   Associate Professor, Department of Statistics,
   Sree Sankara College, Kalady

7. Dr Angel Mathew,
   Assistant Professor of Statistics,
   Maharajas College, Ernakulam

8. Dr Stephy Thomas,
   Assistant Professor of Statistics,
   BCM College, Kottayam

9. Abin John,
   Assistant Professor of Statistics,
   St Thomas College, Kozhencherry

10. Dr Sooraj K Ambat,
    Scientist E, Naval Physical and Oceanographic Laboratory,
    Defence R&D Organization, Ministry of Defence, Thrikkakara, Kochi

11. Prasanth V P (Expert from Industry)
    Manager - Biostatistics, IQVIA, Infopark, Kochi.

Post Graduate Programme in M. Sc. Statistics with Data Science
(From 2020-21 Academic Year onwards)

1. Objectives of the Programme

Statistical analysis, modelling and inference are required in almost all areas of the natural and social sciences, technology and industrial research and development. There has always been tremendous demand for trained post-graduates and specialists in different branches of Statistics.

Advent and expansion of Information Technology in almost all domains of life has increased the significance of *"Data-Science and Technology"* with a need for specialized academic courses at the interface of Statistics and Data analytics. Mahatma Gandhi University pioneers in this direction with the initiation of M.Sc. Statistics programme with specialization in state- of –the-art courses viz.,Applied Algorithms, Statistical Techniques in Data Mining, Analysis of Multi-type Data, Statistical Modelling and Non-parametric Statistics. The curriculum of this trend-setting academic programme offers a blend of foundational theoretical courses as well as applied core courses which will equip the students to clear Indian Statistical Service (ISS), UGC-NET and Data Science specific industrial jobs.

The list of elective courses includes — Machine Learning, Data Visualization, Analysis of Categorical Data, Epidemiology and Study Designs, Population Dynamics, and Statistical Techniques for Quality Control.

This academic programme also has an *Internship/Project* component in the second year under which the students may opt for industry oriented or applied/basic research-oriented project as per their inclination, background, and potential.

2. Eligibility for admissions:

B.Sc. Degree in mathematics or statistics main or B.Sc.(triple main) with Mathematics, Statistics and Computer science as main subject with at least 50% marks for the optional subjects taken together. **No private /distant students will be admitted for the programme.**

3. Faculty under which the degree is to be awarded: Faculty of Science

4. Programme Structure

Table of Courses and Credits

COURSE CODE	COURSE TITLE	TEACHING (LECTURE+ PRACTICAL)	TYPE	CREDITS
SEMESTER I		**TOTAL CREDITS: 19**		
ST500103	SAMPLING THEORY	5L	THEORY	4
ST040101	MEASURE AND PROBABILITY THEORY	5L	THEORY	4
ST040102	THEORY OF PROBABILITY DISTRIBUTIONS	5L	THEORY	4
ST040103	ANALYTICAL METHODS FOR STATISTICS	5L	THEORY	4
ST040104	STATISTICAL COMPUTING 1 USING PYTHON	3L + 2P	PRACTICAL *	3
SEMESTER 2		**TOTAL CREDITS: 19**		
ST500201	ESTIMATION THEORY	5L	THEORY	4
ST500202	STOCHASTIC PROCESSES	5L	THEORY	4
ST500203	MULTIVARIATE DISTRIBUTIONS	5L	THEORY	4
ST040201	TESTING STATISTICAL HYPOTHESES	5L	THEORY	4
ST040202	STATISTICAL COMPUTING 2 USING R	3L + 2P	PRACTICAL *	3

SEMESTER 3		TOTAL CREDITS: 19		
ST500302	DESIGN AND ANALYSIS OF EXPERIMENTS	5L	THEORY	4
ST500304	TIME SERIES ANALYSIS	5L	THEORY	4
ST040301	MULTIVARIATE ANALYSIS AND STATISTICAL TECHNIQUES FOR DATA MINING	5L	THEORY	4
ST040302	STATISTICAL MODELLING	5L	THEORY	4
ST040303	STATISTICAL COMPUTING 3 USING R AND PYTHON	2L + 3P	PRACTICAL *	3
SEMESTER 4		**TOTAL CREDITS: 23**		
ST040401	NON-PARAMETRIC STATISTICS	5L	THEORY	4
ST040402	APPLIED ALGORITHMS AND ANALYSIS OF MULTITYPE DATA	2L + 3P	PRACTICAL *	4
	ELECTIVE I			3
	ELECTIVE II			3
	ELECTIVE III			3
ST040403	PROJECT WORK AND INTERNSHIP (IN A REPUTED INDUSTRY/RESEARCH INSTITUTE)			4
ST040404	COMPREHENSIVE VIVA-VOCE			2
GRAND TOTAL OF CREDITS				80

COURSE CODE	LIST OF ELECTIVE COURSES	TEACHING (LECTURE+ PRACTICAL)	TYPE	CRED-ITS
ELECTIVES – BUNCH 1				
ST880401	MACHINE LEARNING	2L + 3P	THEORY	3
ST880402	DATA VISUALIZATION	2L + 3P	THEORY	3
ST880403	ANALYSIS OF CATEGORICAL DATA	5L	THEORY	3

ELECTIVES – BUNCH 2				
ST890401	EPIDEMIOLOGY AND STUDY DESIGNS	5L	THEORY	3
ST890402	POPULATION DYNAMICS	5L	THEORY	3
ST890403	STATISTICAL TECHNIQUES FOR QUALITY CONTROL	5L	THEORY	3

* *All practical question papers shall be generated from university*

First Semester Courses

COURSE CODE	COURSE TITLE	TEACHING (LECTURE+ PRACTICAL)	TYPE	CREDITS
ST500103	SAMPLING THEORY	5L	THEORY	4
ST040101	MEASURE AND PROBABILITY THEORY	5L	THEORY	4
ST040102	THEORY OF PROBABILITY DISTRIBUTIONS	5L	THEORY	4
ST040103	ANALYTICAL METHODS FOR STATISTICS	5L	THEORY	4
ST040104	STATISTICAL COMPUTING 1 USING PYTHON	3L + 2P	PRACTICAL	3

SAMPLING THEORY (ST500103)

Module 1

1.1 Official Statistical Systems in India – Role of NSSO and CSO and their activities – For general awareness of students (1 or 2 hours) **1.2**Census and Sampling methods, Advantages and disadvantages, Principles of sampling theory, Principal steps in a sample survey, **1.3**probability sampling and non-probability sampling, sampling and non-sampling errors, bias, variance and MSE, **1.4** simple random sampling with and without replacement – estimation of population mean, total and proportions, estimation of sample size -. **1.5** Properties of the estimators, variance and standard error of the estimators, confidence intervals, determination of the sample size.

Module 2

2.1 Stratified random sampling, estimation of the population mean, total and proportion, properties of estimators, various methods of allocation of a sample, comparison of the precisions of estimators under proportional allocation, optimum allocation and SRS. **2.2** Systematic sampling – Linear and Circular, estimation of the mean and its variance, intraclass correlation coefficient, **2.3** comparison of systematic sampling, SRS and stratified random sampling for a population with a linear trend.

Module 3

3.1 Ratio method of estimation, estimation of population ratio, mean and total,**3.2** Bias and relative bias of ratio estimator, comparison with SRS estimation. Unbiased ratio type estimators- Hartley- Ross estimator, Regression method of estimation. Comparison of ratio and regression estimators with mean per Modulemethod,**3.3** Cluster sampling, single stage cluster sampling with equal and unequal cluster sizes, estimation of the population mean and its standard error. **3.4** Two- stage cluster sampling with equal and unequal cluster sizes, **3.5**Multistage and Multiphase sampling (Basic Concepts), estimation of the population mean and its standard error.

Module 4

4.1Varying probability sampling, PPS sampling with and without replacement,**4.2** cumulative total method, Lahiris method, Midzuno-Zen method and its inclusion probabilities, estimation of the population total and its estimated variance under PPS wr sampling, **4.3**ordered and unordered estimators of the population total under PPS wor, Horwitz – Thomson estimator and its estimated S. E,**4.4** Des-Raj's ordered estimator, Murthy's unordered estimator (properties of these estimators for n=2 only). Inclusion probability proportional to size Sampling Procedures.

Textbooks

1. Cochran W.G (1992): Sampling Techniques, Wiley Eastern, New York.
2. Parimal Mukhopadhyay (2009) Theory and Methods of Survey Sampling, Second Edition, PHI Learning (P) Ltd.
3. Singh,D and Chowdhary,F.S. (1999): Theory and Analysis of Sample Survey Designs, Wiley Eastern (New Age International), New Delhi.

References

1. P.V. Sukhatmeet.al. (1984): Sampling Theory of Surveys with Applications. IOWA State University Press, USA.
2. M.N. Murthy (1977) Sampling Theory and Methods, Statistical Publishing Society
3. Sampath S. C. (2001) Sampling Theory and Methods, Alpha Science International Ltd., India.
4. Thomas Lumley (1969) Complex Surveys- A guide to analysis using R, Wiley eastern Ltd.
5. Desraj (1967) Sampling theory. Tata McGraw Hill , NewDelhi
6. MOSPI website.

MEASURE AND PROBABILITY THEORY (ST040101)

MODULE 1

Algebra of sets, limit of a sequence of sets, Fields, Sigma fields, General measure space, Lebesgue measure, Lebesgue-Stieltjes measure, Counting measure and their simple properties. Measurable functions and their properties. Integrals of indicator function, simple function and measurable functions, basic integration theorems. Monotone convergence theorem, Fatou's Lemma, Bounded convergence theorem and Lebesgue dominated convergence theorem. Lebesgue, Lebesgue-Stieltjes Integral and Riemann Integral (Basic concepts only).

MODULE 2

Discrete and Continuous probability spaces and their properties. Monotone, continuity and other properties. Conditional probability, multiplication theorem, total probability and Bayes theorem. Independence of events, Random variables, properties of random variables. Distribution function and its properties. Jordan decomposition theorem, Correspondence theorem (statement only). Independence of random variables. Mathematical expectation and its properties.

MODULE 3

Basic, Chebychev's, Markov's, Liaponov's, Jensen's, Cr, Cauchy-Swartz's, Holder's, Minkowski's inequalities. Convergence: Modes of convergence, Convergence in probability, in distribution, in r^{th} mean, almost sure convergence and their inter-relationships, Borel 0-1 criterion, Characteristic function of a random variable, properties, statement of Bochner's Theorem. Continuity and inversion theorems of characteristic functions (statements only).

MODULE 4

Law of large numbers: Weak law of large numbers – Chebychev's WLLN and Khinchine WLLN, Strong law of large numbers - Kolmogrov strong law of large numbers. Central limit theorem: Lindberg-Levy central limit theorem, Liaponov's central limit theorem, Lindberg-Feller central limit theorem (Without proof).

Textbooks

1) Ash R.B. and Doléans-Dade C.A. (2000) Probability and measure theory, Academic Press.
2) Bhat B.R (1999) Modern Probability theory, Third Edition, Wiley Eastern Ltd, New Delhi.

References

1) Basu A.K. (2012). Measure Theory and Probability, Second Edition, PHI Learning Pvt. Ltd, New Delhi.
2) Billingsley P. (2012) Probability and Measure, Anniversary edition, Wiley Eastern ltd.
3) Loeve M. (1977) Probability Theory, Fourth edition, Springer-Verlag.
4) Rohatgi V.K. and SalehM. (2015) An introduction to probability and statistics, Third edition, Wiley.
5) Robert G. Bartle (2001), A Modern Theory of Integration, American Mathematical Society (RI)
6) Laha R.G. and Rohatgi V.K. (1979) Probability theory, John Wiley.

THEORY OF PROBABILITY DISTRIBUTIONS (ST040102)

Module I

Probability Generating functions, Moment generating functions and their properties, Discrete Distributions - Bernoulli, Binomial, Geometric, Poisson,

Negative binomial and Hypergeometric. Power series and Ord family of distributions- Definition, Identification of members.

Module 2

Continuous Distributions -Rectangular, Exponential, Weibull, Beta, Gamma, Pareto, Normal, Lognormal, Cauchy, Laplace, Logistic, Inverse Gaussian. Pearson family and Exponential family of distributions – Definition and Identification of members.

Module 3

Functions of Random variables and their distributions. Probability integral transform, Distributions of sums, products and ratios of independent random variables, Truncated distributions, Compound distributions.

Module 4

Sampling distributions - Chi-square, t and F distributions (central and non-central forms), Order statistics and their distributions - joint and marginal distributions; Distributions of sample median, range and mid–range (Exponential and Uniform), Quantiles and QQ plot.

Textbooks

1. Gupta S.C. and Kapoor V.K. (2000) Fundamentals of Mathematical Statistics, S. Chand & Co, New Delhi.
2. Hogg R.V and Craig A.T. (2013) Introduction to Mathematical Statistics, Mac Millian publishing company.

Reference Books

1. Arnold B.C, Balakrishnan N. and Nagaraja H.N. (1992) A first Course in Order Statistics.
2. Biswas S. and Srivastava G.L (2008) Mathematical Statistics: A textbook, Alpha Science International Ltd
3. Johnson N.L, Kotz S. and Balakrishnan N. (1991) Continuous Univariate distributions I & II, Wiley.
4. Johnson N.L, Kotz S. and Kemp A.W. (1992) Univariate discrete distributions, Wiley.
5. Kotz S, Balakrishnan N. and Johnson N.L. (2000) Continuous Multivariate distributions, Wiley.
6. Rohatgi V.K. and Saleh M. (2015) An introduction to probability and statistics, Third edition, Wiley.

ANALYTICAL METHODS FOR STATISTICS (ST040103)

Module I

Sequence and series of real numbers, Convergence of sequence and series of real numbers - Definitions and problems, Continuity, Uniform continuity, Differentiability. Functions of several variables: maxima and minima, Method of Lagrangian multipliers, Laplace transform and its application to Differential equations, Fourier transform (Definition only)

Module 2

Vector spaces, Subspaces, Linear independence of vectors, Basis and dimension of a vector space, Inner product and orthogonal vectors, Gram-Schmidt orthogonalization process, Orthonormal basis, Matrix and its properties, Rank of a matrix, Partitioned matrices.

Module 3

Linear equations, Rank-Nullity theorem, Characteristic roots and vectors, Cayley-Hamilton theorem, Characteristic subspaces of a matrix, Nature of characteristic roots of some special types of matrices, Algebraic and geometric multiplicity of a characteristic root, Generalized inverse, Properties of g-inverse, Moore-Penrose inverse, and its computations.

Module 4

Quadratic forms, Congruent transformations, Congruence of symmetric matrices, Canonical reduction and orthogonal reduction of real quadratic forms, Nature of quadratic forms, Simultaneous reduction of quadratic forms, Similarity, and spectral decomposition of real symmetric matrices.

Textbooks

1. Malik S.C. and Arora S. (2014) Mathematical analysis, Fourth edition, New age international.
2. Shayle R. Searle, Andre I. Khuri - Matrix Algebra Useful for Statistics, Wiley (2017)
3. Rao A.R. and Bhimasankaram P. (2000) Linear Algebra, Second edition, Hindustan Book Agency.
4. Shanti Narayan and Mittal P.K (1962) A Course of Mathematical Analysis, S. Chand and Company.

Reference Books

1. Apostol T.M. (1996) Mathematical Analysis, Second edition, Narosa Publishing House, New Delhi.
2. Gilbert Strang (2014) Linear Algebra and its Applications, 15th Re-Printing edition, Cengage Learning.
3. Hoffman K. and Kunze R. (2014) Linear Algebra, Second edition, Phi Learning.

4. Bapat, Ravindra B., Linear Algebra and Linear Models, Springer 2012
5. Rao C.R. (2009) Linear Statistical Inference and its Applications, Second edition, Wiley Eastern.

STATISTICAL COMPUTING 1 USING PYTHON (ST040104)

MODULE 1

Introduction to Python, Types of numeric data, strings, Basic output statements, List, tuples, and Files- Introduction, Control flow structures – if else statements, while, for loops, defining functions in python, introduction to Program Design

Chapters 2,3,4 of textbook 1

MODULE 2

Processing of text data and csv files, dictionaries; exception handling; Turtle graphics; Classes and objects; Inheritance

Chapters 5,6,7 of textbook 1

MODULE 3

Statistical Testing using Python: (Implementation using Python), Plotting data in Python: Scatter plots, histogram, cumulative frequencies, error-bars, box plots, pie charts, grouped bar charts, Bivariate scatter plots, sensitivity and specificity, one sample t-test, paired t-test, independent sample t-test, ANOVA- one way. Practical of courses Analytic methods for Statistics and Theory Sampling using Python

Chapter 4, 7.1, 7.3, 8, 9 of textbook 2

MODULE 4

Random Number Generation- Uniform random Generation using numpy, Generation of non-uniform random variables - Simple discrete random variables like Bernoulli, Binomial, Uniform, Inversion methods, Simple acceptance rejection, Box Muller Method, p-p plots, qq plots, establishing limit theorems using random number generation and various relationships between distributions

Section 2.2 of textbook 3

Textbooks

1. Schneider, David I, An Introduction to Programming Using Python, Pearson Education Limited 2016
2. Haslwanter, Thomas, An Introduction to Statistics with Python: With Applications in the Life Sciences, Springer 2016

3. Asmussen, Søren, Glynn, Peter W. Stochastic Simulation: Algorithms and Analysis, Springer 2007.

Reference Books

1. Ceder, Vernon L, The Quick Python Book, Manning Publications Co., Greenwich
2. Saha, Amit Doing math with Python: use programming to explore algebra, statistics, calculus, and more!, No Starch Press, 2015
3. https://machinelearningmastery.com/how-to-generate-random-numbers-in-python/

Evaluation: 6 numerical questions each with weightage 10 (marks per question is 50) are to be asked. The student is expected to answer 3 full questions. Use of the packages (Python) are allowed for answering the questions in this paper. Examination of 3 hours duration must be conducted in the computer lab under the supervision of an external examiner appointed by the Controller of Examinations.

Question papers shall be generated from university.

Second Semester Courses

COURSE CODE	COURSE TITLE	TEACHING (LECTURE+ PRACTICAL)	TYPE	CREDITS
ST500201	ESTIMATION THEORY	5L	THEORY	4
ST500202	STOCHASTIC PROCESSES	5L	THEORY	4
ST500203	MULTIVARIATE DISTRIBUTIONS	5L	THEORY	4
ST040201	TESTING STATISTICAL HYPOTHESES	5L	THEORY	4
ST040202	STATISTICAL COMPUTING 2 USING R	3L + 2P	PRACTICAL	3

ESTIMATION THEORY (ST500201)

Module 1

1.1 Point estimation-properties of estimators – unbiasedness - consistency, sufficient condition for consistency – Sufficiency, minimal sufficiency,**1.2** completeness, bounded completeness, Fisher-Neymann factorization theorem, **1.3** exponential families, UMVUE estimators and their characterization, **1.4** Rao- Black well theorem, Lehmann – Scheffe theorem, **1.5** ancillary statistics, Basu's theorem.

Module 2

2.1 Fisher information measure and its properties, Fisher information matrix, **2.2** Lower bound to the variance of an unbiased estimator, Cramer - Rao inequality, Bhattacharyya's bounds, **2.3** Efficiency, minimum variance

Module 3

3.1 Methods of estimation: method of moments, method of maximum likelihood & their properties, Cramer- Huzurbazar theorem, Fisher's scoring method, **3.2** method of minimum chi-square and method of modified minimum chi-square- **3.3** Interval estimation – Pivotal method of construction - shortest confidence intervals and their construction (minimum average width) -**3.4** Construction of shortest confidence intervals in large samples.

Module 4

4.1 Basic elements of Bayesian inference, Loss function and risk functions, Standard forms of loss functions, **4.2** Prior distribution, Bayes Theorem, Posterior distribution, **4.3** Bayes risk, Bayes principle, Bayes estimators, Minimax estimators.

Textbooks

1. Rohatgi V.K. and Saleh A.K. (2015) An Introduction to Probability Theory and Mathematical Statistics, Wiley.
2. Berger J.O. (1993) Statistical Decision Theory and Bayesian Analysis, Third Edition, Springer.
3. Casella, G and Berger, R.L (2007) Statistical Inference, Second Edition, Cengage Learning.

Reference Books

1. Hogg R. V. and Craig A. T. (2013) Introduction to Mathematical Statistics, Pearson
2. Kale B. K. (2005) A First Course on Parametric Inference, Alpha Science International.

3. Lehmann E.L. (1983) Theory of point estimation – Wiley, New York.
4. Lindgren B.W (1976) Statistical Decision Theory (3rd Edition), Collier Mac Millian, New York.
5. Rao C.R (2009) Linear Statistical Inference and its Applications, John Wiley, New York.

STOCHASTIC PROCESSES (ST500202)

Module 1

1.1 Introduction to stochastic processes:- classification of stochastic processes according to state space and time space, wide sense and strict sense stationary processes, processes with stationary independent increments, **1.2** Markov process, Markov chains-transition probability matrices, Chapman-Kolmogorov equation, **1.3** first passage probabilities, generating functions, classification of states, criteria for recurrent and transient states, **1.4** mean recurrence time, mean ergodic theorem, the basic limit theorem of Markov chains (statement only),**1.5** reducible and irreducible Markov chains, stationary distributions, limiting probabilities and absorption probabilities.

Module 2

2.1 Random walk, gambler's ruin problem; **2.2** Galton-Watson branching process, generating function relations, **2.3** mean and variance functions, extinction probabilities, criteria for extinction.

Module 3

3.1 Continuous time Markov chains, Poisson processes,**3.2** pure birth processes and the Yule processes, birth and death processes, **3.3** Kolmogorov forward and backward differential equations, linear growth process with immigration, **3.4** steady-state solutions of Markovian queuing models--M/M/1, M/M/1 with limited waiting space,**3.5** M/M/s, M/M/s with limited waiting space.

Module 4

4.1 Renewal processes– concepts, examples, **4.2** Poisson process viewed as a renewal process, renewal equation, elementary renewal theorem,**4.3** asymptotic expansion of renewal function, central limit theorem for renewals,**4.4** key renewal theorem (statement only), delayed renewal processes.

Textbooks

1. Medhi J. (2017) Stochastic Processes, Second Edition, Wiley Eastern, New Delhi
2. Ross S.M. (2007) Stochastic Processes. Second Edition, Wiley Eastern, New Delhi
3. Ross S.M. (2014) Introduction to Probability Models. Eleventh Edition, Elsevier

Reference Books

1. Feller W. (1968) Introduction to Probability Theory and its Applications, Vols. I & II, John Wiley, New York.
2. Karlin S. and Taylor H.M. (1975) A First Course in Stochastic Processes, Second edition, Academic Press, New-York.
3. Cinlar E. (1975) Introduction to Stochastic Processes, Prentice Hall, New Jersey.
4. Basu A.K. (2003) Introduction to Stochastic Processes, Narosa, New-Delhi.
5. Bhat U.N. and Miller G. (2003) Elements of Applied Stochastic Processes. (Third edition), John Wiley, New York.

MULTIVARIATE DISTRIBUTIONS (ST500203)

Module 1

1.1 Notions of bivariate distributions, Gumbel's bivariate exponentials and basic properties, **1.2** Bivariate normal distribution- marginals and conditionals, independence of random vectors, **1.3** multinomial distribution and its basic properties.

Module 2

2.1 Multivariate normal (singular and non-singular), characteristic function, marginals, and conditionals–**2.2** properties, characterizations, **2.3** estimation of mean vector and dispersion matrix, independence of sample mean vector and sample dispersion matrix.

Module 3

3.1 Jacobian of matrix transformations of Y= AXB; Y= AXA'; X=TT',**3.2** matrix variate gamma and beta distributions. **3.3** Wishart distribution and its basic properties, characteristic function,**3.4** generalized variance and its distribution.

Module 4

4.1 Quadratic forms and their distributions (both scalar and vector forms), **4.2** Independence of quadratic forms, Cochran's theorem. **4.3** Simple, partial, and multiple correlation distributions, properties and their inter-relationships, tests. **4.4** Null and non-null distribution of simple and partial correlations, null distribution of multiple correlation.

Textbooks

1. Anderson T.W. (1984): An introduction to multivariate statistical analysis, Second edition, John Wiley.
2. Dean W. Wichern, Richard A. Johnson, Applied Multivariate Statistical Analysis, Sixth Edition, Pearson

Reference Books

1. Seber G.A.F. (1983): Multivariate Observations, John Wiley.
2. GiriN.(1984): Multivariate Statistical Inference, Academic publishers.
3. Kollo T and Rosen D.V. (2005): Advanced Multivariate Statistics with Matrices, Springer.
4. Kotz S, Balakrishnan N, and Johnson N.L (2000): Continuous Multivariate Distributions, Models and Applications,Volume 1,Second Edition, John Wiley.
5. Mathai A.M. (1996): Jacobins of Matrix Transformations and functions of Matrix Argument, World Scientific Pub CoPvt.Ltd
6. Rao.C.R(2009): Linear statistical inference and its applications,Second Edition, Wiley Eastern.

TESTING STATISTICAL HYPOTHESES (ST040201)

Module 1

1.1 Basic concepts in statistical hypotheses testing-simple and composite hypothesis, critical regions, Type-I and Type-II errors, Significance level, p-value, and power of a test. **1.2** Neyman-Pearson lemma and its applications; **1.3** Construction of tests using NP lemma- Most powerful test, uniformly most powerful test.**1.4** Monotone Likelihood ratio and testing with MLR property; Testing in one-parameter exponential families-one sided hypothesis, **1.5** Unbiased and Uniformly Most Powerful Unbiased tests for different two-sided hypothesis; Extension of these results to Pitman family when only upper or lower end depends on the parameters.

Module 2

2.1 Similar regions tests, Neymann structure tests, Likelihood ratio (LR) criterion and its properties, **2.2** LR tests for testing equality of means and variances of several normal populations. Testing in multi-parameter exponential families-tests with Neyman structure, **2.3** UMP and UMPU similar size-tests; **2.4** Confidence sets, UMA and UMAU confidence sets, Construction of UMA and UMAU confidence sets using UMP and UMPU tests respectively.

Module 3

3.1 Sequential probability ratio tests (SPRT), Properties of SPRT, Determination of the boundary constants **3.2** Construction of sequential probability ratio tests, Wald's fundamental identity, **3.3** Operating characteristic (OC) function and Average sample number (ASN) functions for Normal Binomial, Bernoulli's, Poisson and exponential distribution.

Module 4

4.1 Notion of likelihood ratio tests, Hotellings-T^2 and Mahalnobis-D^2 statistics-Their properties, interrelationships and uses, **4.2** Null distributions (one sample and two sample cases), Testing equality of mean vectors of two independent multivariate normal populations with same dispersion matrix,**4.3** Problem of symmetry, Multivariate Fisher- Behren problem.

Textbooks

1) Rohatgi V.K. (1976) An Introduction to Probability Theory and Mathematical Statistics, John Wiley & Sons, New York.
2) Anderson T.W. (1984): An introduction to multivariate statistical analysis, Second edition, John Wiley.

References Books

1. Casella G. and Berger R.L. (2002) Statistical Inference, Second Edition Duxbury, Australia.
2. Lehman E.L. (1998) Testing of Statistical Hypothesis. John Wiley, New York.
3. Wald A. (1947) Sequential Analysis, Wiley, Doves, New York.
4. Parimal Mukhopadhyay (2006): Mathematical Statistics, 3/e, Books and Allied (P) Ltd, Kolkata.
5. Rao C.R. (1973) Linear Statistical Inference and its Applications, Wiley.

STATISTICAL COMPUTING 2 USING R (ST040202)

Module 1

Introduction to statistical software R, Data objects in R, manipulating vectors, matrices, lists, importing of files, data frame, and computations of descriptive statistics measures. R-Graphics- Histogram, Box-plot, Stem and leaf plot, Scatter plot, Plot options; Multiple plots in a single graphic window, frequency table, Controlling Loops- For, repeat, while, if, if else etc, Functions in R.

Module 2

Plotting of probability distributions and sampling distributions, P-P plot, Q-Q Plot, Simulation of random numbers, Fitting of discrete and continuous distributions. Chi Square goodness of fit.

Module 3

One- and two-sample tests, paired sample t-tests and independent sample t-tests, Chi-square test for goodness of fit, Simple regression, Logistic regression and correlation, ANOVA and Kruskal–Wallis

Chapters 2,3,4,5,6 of textbook 2

Module 4

Random Number Generation- Uniform random Generation, Generation of non-uniform random variables - Simple discrete random variables like Bernoulli, Binomial, Uniform, Inversion methods, Simple acceptance rejection, Box Muller Method, p-p plots, qq plots, establishing limit theorems using random number generation and various relationships between distributions, Practical implementation of multivariate distribution, estimation and testing

Textbooks

1. An Introduction to R by W. N. Venables, D. M. Smith and the R Core Team
2. Dalgard, Peter, Introductory statistics with R, Springer,
3. Schumacker, Randall E., Using R with multivariate statistics, sage 2016
4. The R Book by Michael J. Crawley, John Wiley and Sons, Ltd., 2007.

Evaluation: 6 numerical questions each with weightage 10 (marks per question is 50) are to be asked. The student is expected to answer 3 full questions. Use of the packages (R) are allowed for answering the questions in this

paper. Examination of 3 hours duration must be conducted in the computer lab under the supervision of an external examiner appointed by the Controller of Examinations.

Question papers shall be generated from university.

Third Semester Courses

COURSE CODE	COURSE TITLE	TEACHING (LECTURE+ PRACTICAL)	TYPE	CREDITS
ST500302	DESIGN AND ANALYSIS OF EXPERIMENTS	5L	THEORY	4
ST500304	TIME SERIES ANALYSIS	5L	THEORY	4
ST040301	MULTIVARIATE ANALYSIS AND STATISTICAL TECHNIQUES FOR DATA MINING	5L	THEORY	4
ST040302	STATISTICAL MODELLING	5L	THEORY	4
ST040303	STATISTICAL COMPUTING 3 USING R AND PYTHON	2L + 3P	PRACTICAL	3

DESIGN AND ANALYSIS OF EXPERIMENTS (ST500302)

Module 1

1.1 Linear estimation: Gauss Markov set up, Estimability of parameters, **1.2** Method of least squares, best linear unbiased Estimators, Gauss-Markov Theorem, Tests of linear hypotheses, **1.3** Analysis of variance- one-way, two-way and three-way classification models.

Module 2

2.1 Planning of experiments: Basic principles of experimental design, Uniformity trails, **2.2** Completely randomized design (CRD), Randomized block design (RBD), **2.3** Latin square design (LSD) and Graeco-latin square designs, **2.4** Analysis of covariance (ANACOVA), ANACOVA with one concomitant variable in CRD and RBD

Module 3

3.1 Incomplete block design: Balanced incomplete block design (BIBD); Incidence Matrix, C- Matrix, Parametric relations;**3.2** Intra-block analysis of BIBD, Connectedness, Construction of BIBD by developing initial blocks,**3.3** Basic ideas of partially balanced incomplete block design (PBIBD).

Module 4

4.1 Factorial experiments, 2^n and 3^n factorial experiments, Analysis of 2^2, 2^3 and 3^2 factorial experiments,**4.2** Confounding in 2^n and 3^n factorial experiments, Construction of confounded scheme in 2^n factorial experiments,**4.3** Split plot experiments (RBD).

Textbooks

1. Das M.N. and Giri N.C. (1994) Design and analysis of experiments, Wiley Eastern Ltd
2. Joshi D.D. (1987) Linear estimation and Design of Experiments, Wiley Eastern Ltd
3. Montgomery, DC Design and Analysis of Experiments, Wiley Eastern Ltd

Reference Books

1. Agarwal B.L (2010) Theory and Analysis of Experimental Designs, CBS Publishers & Distributers
2. Dean A. and Voss D. (1999) Design and Analysis of Experiments, Springer Texts in Statistics
3. Dey A. (1986) Theory of Block Designs, Wiley Eastern, New Delhi.
4. Gomez K.A. and Gomez A.A. (1984) Statistical Procedures for Agricultural Research, Wiley Eastern Ltd

5. Kempthrone,O. (1952) Design and Analysis of Experiments,Wiley Eastern, New York
6. Montgomery, C.D. (2012) Design and Analysis of Experiments, John Wiley, New York.
7. Rangaswamy, R (2010) A textbook on Agricultural Statistics, New Age International Publishers.

TIME SERIES ANALYSIS (ST500304)

Module 1

1.1 Time series, Components of time series, Additive and multiplicative models, **1.2** Estimation and elimination of trend and seasonality, moving average, **1.3** Simple Exponential Smoothing, Holt's exponential smoothing, Holt-Winter's exponential smoothing, **1.4** Forecasting based on smoothing.

Module 2

2.1 Time series as a discrete parameter stochastic process, Auto-covariance and autocorrelation functions, Partial Auto-correlation function, and their properties, **2.2** Stationary processes, Wold representation of linear stationary processes, **2.3** Detailed study of the Box - Jenkins linear time series models: Autoregressive, Moving Average, Autoregressive Moving Average and Autoregressive Integrated Moving Average models.

Module 3

3.1 Estimation of ARMA models: Yule-Walker estimation for AR Processes, **3.2**Maximumlikelihood and least squares estimation for ARMA Processes. Choice of AR and MA periods, MMSE forecasting using ARIMA models**3.3** Residual analysis and diagnostic checking.

Module 4

4.1 Spectral density of a stationary time series and its elementary properties, Periodogram,**4.2** Spectral density of an ARMA process. Seasonal ARIMA models (Basic concepts only),**4.3** ARCH and GARCH models (Basic concepts only).

Textbooks

1. Abraham B. and Ledolter J.C. (2005) Statistical Methods for Forecasting, Second edition Wiley.
2. Box G.E.P, Jenkins G.M. and Reinsel G.C. (2008) Time Series Analysis: Forecasting and Control, Fourth Edition, Wiley.

3. Brockwell P.J and Davis R.A. (2002) Introduction to Time Series and Forecasting Second edition, Springer-Verlag.

Reference Books

1. Cryer, J. D. and Chan, K. (2008). Time Series Analysis with Applications in R, Second Edition, Springer-Verlag.
2. Shumway, R. H. and Stoffer, D. S. (2011) Time Series Analysis and Its Applications with R Examples, Third Edition, Springer-Verlag.

MULTIVARIATE ANALYSIS AND STATISTICAL TECHNIQUES FOR DATA MINING (ST040301)

Module 1

Introduction to data mining; data types for Data mining, Data mining functionalities -Concept/class description: characterization and discrimination, Association analysis, Classification and prediction, Clustering analysis, Evolution and deviation analysis, Data Pre-processing, Data cleaning, Data integration and transformation, Data reduction, Discretization and concept hierarchy generation

Module 2

2.1 Dimension Reduction methods: Profile Analysis and the associated tests,**2.2** Principal component Analysis-Method of extraction-properties, the associated tests, **2.3** Factor Analysis-Orthogonal Model-Estimation of factor loadings, **2.4** Canonical variates and canonical correlation, use, estimation and computation.

Module 3

3.1 Classification problems: Discriminant Analysis-Bayes' procedure, Classification into one of the two populations (Normal distribution only), Classification into several populations (Normal distribution only), **3.2** Fishers linear discriminant function and its associated tests, **3.3** Cluster Analysis: proximity measures, Hierarchical and non-hierarchical methods.

Module 4

4.1 Multivariate General linear models-MANOVA (one way and two way), **4.2** Wilk's λ, Rau's U, Pillai's trace, Hotelling-Lawley trace, Roy's Maximum Root Statistics (Concepts only),**4.3** Tests-Independence of sets of variables, Equality of dispersion matrices and Sphericity test.

Textbooks

1. Jiawei Han, Micheline Kamber, Jian Pei - Data Mining: Concepts and Techniques-Morgan Kaufmann (2011)
2. Johnson R.A. and Wichern D.W. (2008) Applied Multivariate Statistical Analysis. (6thed.) Pearson education.
3. Rencher,A. C. (2012) Methods of Multivariate Analysis.(3rd ed.) John Wiley.

Reference Books:

2. Rao C. R. (2009) Linear Statistical Inference and Its Applications (2nd Ed.), Wiley
3. Johnson, D. E. (1998): Applied Multivariate methods for Data Analysts, Duxbury Press, USA-An International Thomson Publishing Company.
4. Morrison, F (2003): Multivariate Statistical Methods, Brooks/Cole, 4thRevisededn. McGraw Hill Book Company.
5. Kshirsagar A.M. (1972): Multivariate Analysis, M.Dekker.
6. Srivastava M.S.and Khatri C.G. (2002): Methods of Multivariate Statistics,John Wiley & Sons, N.Y.
7. Anderson T. W. (2010) An Introduction to Multivariate Statistical Analysis (3rd ed.) John Wiley.
8. Seber G. F. (2004) Multivariate Observations, John Wiley.

STATISTICAL MODELLING (ST040302)

Module 1

Difference between Statical and Mathematical Modelling, Steps in Statistical Modelling, Principal of Least squares, Simple linear regression models, Multiple linear regression models, estimation of the model parameters, tests concerning the parameters, confidence intervals, prediction, use of Dummy variables in regression, polynomial regression models, stepwise regression.

Module 2

Multicollinearity- consequences, Detection, Farrar-Glauber test, remedial measures. Heteroscedasticity- consequences, Detection, tests, remedial measures Aitken's generalized least square method. Auto-correlation-tests for auto correlation, consequences, and estimation procedures, Errors in variables-consequences, detection, remedial measures, Stochastic regressors. Diagnostics, outlier, Influential observations, Leverage, Non-parametric regression basics.

Module 3

Introduction to nonlinear regression, least squares in the nonlinear case and estimation of parameters, 3.2 Models for binary response variables, estimation, and diagnosis methods for logistic and Poisson regressions.3.3 Prediction and residual analysis, 3.4 Generalized Linear Models – estimation and diagnostics.

Module 4

Transformations and weighting to correct model inadequacies, Analytical methods for selecting a transformation, The Box-Cox method, Transformation on the regress or variables, 4.2 Ridge regression, Basic form of ridge regression, Robust regression Least absolute deviation regression, Least median of squares regression,4.3 Inverse estimation- The calibration problem, Resampling procedures for regression models (Bootstrapping)

Textbooks

1. Montgomery D.C., Peck E.A. and Vining G.G. (2007) Introduction to Linear Regression Analysis, John Wiley, India.
2. Norman R. Draper, Harry Smith, Applied Regression Analysis, Third Edition, Wiley

Reference Books

1. Kutner M. H, Nachtsheim C.J, Neter J and Li W. (2005), Applied Linear Statistical Model, Fifth edition. McGraw Hill

STATISTICAL COMPUTING 3 USING R AND PYTHON (ST040303)

Applications of topics covered in

1. ST500302: Design and Analysis of Experiments
2. ST500304: Time Series Analysis
3. ST040301: Multivariate Analysis and Statistical Techniques for Data Mining

Evaluation: 6 numerical questions each with weightage 10 (marks per question is 50) are to be asked. The student is expected to answer 3 full questions. Use of the packages (R / Python) are allowed for answering the questions in this paper. Examination of 3 hours duration must be conducted in the computer lab under the supervision of an external examiner appointed by the Controller of Examinations.

Question papers shall be generated from university.

Fourth Semester Courses

COURSE CODE	COURSE TITLE	TEACHING (LECTURE+ PRACTICAL)	TYPE	CREDITS
ST040401	NON-PARAMETRIC STATISTICS	5L	THEORY	4
ST040402	APPLIED ALGORITHMS AND ANALY-SIS OF MULTITYPE DATA	2L + 3P	PRACTICAL	4
	ELECTIVE I		THEORY	3
	ELECTIVE II		THEORY	3
	ELECTIVE III		THEORY	3
ST040403	PROJECT WORK AND INTERNSHIP (IN A REPUTED INDUSTRY/RE-SEARCH INSTITUTE)			4
ST040404	COMPREHENSIVE VIVA-VOCE			2

NON-PARAMETRIC STATISTICS (ST040401)

Module 1

Order Statistics, Quantiles, QQ Plot, Empirical distribution function and its properties, Glivenko-Cantelli lemma (statement only), Kaplan Meier estimator, One sample and paired sample procedure; Sign test, Wilcoxon signed rank test, Test for Median and Quantiles.

Module 2

Goodness-of fit test; Kolmogorov-Smirnov test, Cramer-von Mises test, Anderson- Darling test, Chi-square test, Shipiro-Wilk test, Jarque-Bera test, Testing for Randomness: Run test, Wald Wolfowitz Run Test

Module 3

McNemar test, Cochran Q test, Two and k samples procedures; Two sample Kolmogrov- Smirnov Test, Mann-Whitney U test, Kruskal-Wallis test, Friedmans test, Measure of association; Kendall Tau, Spearman Rho, Re-sampling techniques; Jackknife and Bootstrap.

Module 4

Smoothing, Bias variance trade off, Non-parametric regression, Density estimation; Histogram and Kernel density estimation.

Practical using R or Python.

Textbook

1. J. D. Gibbons, S. Chakraborti, Nonparametric Statistical Inference, Springer, (2003).
2. L. A. Wasserman, All of Nonparametric Statistics, Springer, (2006).
3. M. Hollander, D. A. Wolfe and E. Chicken, Nonparametric Statistical Methods, John Wiley & Sons, (2013).
4. J. Kloke, J. W. McKean, Nonparametric Statistical Methods Using R, CRC Press, (2015).

APPLIED ALGORITHMS AND ANALYSIS OF MULTITYPE DATA (ST040402)

Module 1

Em Algorithm: Two-Component Mixture Model, Gaussian Models, The EM Algorithm in General, EM as a Maximization–Maximization Procedure

Section 8.5 of Textbook 1

Module 2

Support Vector Machines: Maximal Margin Classifier, Support Vector Classifiers, Support Vector Machines, SVMs with More than Two Class- One-Versus-One Classification and One-Versus-All Classification *(Section 9.1,9.2,9.3,9.4 of textbook 2)* Proximal SVM *(Research Paper 3)*

Module 3

Multi-dimensional scaling, Definition, Perceptual Map, Interpreting the axes, decision framework for perceptual mapping, Aggregate and disaggregate analysis, Decompositional and Compositional approaches, Interpreting the MDS results,

Chapter four of Textbook 2, Chapters 10 of textbook 4

Module 4

Structural Equation Modelling, importance of SEM, variable and construct, Various stages in SEM, Performing SEM and Interpreting them

Chapter seven of Textbook 2, Chapters 11 of Textbook 4

Practical using R or Python: Applications of topics covered in

1. ST040302: Statistical Modelling
2. ST040401: Non Parametric Statistics

Textbook

1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer
2. Trevor Hastie, Robert Tibshirani, Jerome Friedman The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, Second Edition
3. Glenn Fung, Olvi L. Mangasarian, Proximal support vector machine classifiers, Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data miningAugust 2001 Pages 77–86https://doi.org/10.1145/502512.502527
4. (Use R) Brian Everitt, Torsten Hothorn (auth.) - An Introduction to Applied Multivariate Analysis with R-Springer-Verlag New York (2011)
5. Joseph F. Hair, William C. Black, Barry J. Babin, Rolph E. Anderson - Multivariate Data Analysis (7th Edition)-Prentice Hall (2009)

Reference Texts

1. Rex B. Kline - Principles and Practice of Structural Equation Modeling (Methodology in the Social Sciences)-Guilford Press (2010)
2. Randall E. Schumacker - Using R With Multivariate Statistics-SAGE Publications (2015)

Evaluation: 6 numerical questions each with weightage 10 (marks per question is 50) are to be asked. The student is expected to answer 3 full questions. Use of the packages (R / Python) are allowed for answering the questions in this paper. Examination of 3 hours duration must be conducted in the computer lab under the supervision of an external examiner appointed by the Controller of Examinations.

Question papers shall be generated from university.

COURSE CODE	LIST OF ELECTIVE COURSES	TEACHING (LECTURE+ PRACTICAL)	TYPE	CREDITS
ELECTIVES – BUNCH 1				
ST880401	MACHINE LEARNING	2L + 3P	PRACTICAL	3
ST880402	DATA VISUALIZATION	2L + 3P	PRACTICAL	3
ST880403	ANALYSIS OF CATEGORICAL DATA	5L	THEORY	3
ELECTIVES – BUNCH 2				
ST890401	EPIDEMIOLOGY AND STUDY DESIGNS	5L	THEORY	3
ST890402	POPULATION DYNAMICS	5L	THEORY	3
ST890403	STATISTICAL TECHNIQUES FOR QUALITY CONTROL	5L	THEORY	3

MACHINE LEARNING (ST880401)

Module 1

Introduction to Machine learning, difference between machine learning and Statistics, Decision Tree Learning, Appropriate Problems for Decision tree learning, Basic decision tree algorithm, Hypothesis space in decision tree learning, inductive bias in decision tree learning, issues in decision tree learning, Supervised and Unsupervised learning.

Module 2

Artificial Neural Networks: Neural network representation, Appropriate problems for neural network learning, perceptron, multilayer networking, and Backpropagation algorithm

Module 3

Bayesian Learning: Bayes theorem and concept of learning, ML and least squared error hypothesis, ML hypothesis for predicting probabilities, minimum length description principle, Bayes optimal classifier, Gibbs Algorithm, Naive Bayes Classifier, Bayesian Belief Networks

Module 4

Ensemble Learning: Boosting Procedures, The AdaBoost Algorithm, Initial Analysis, Margin Explanation, Statistical View, Multiclass Extension, Noise Tolerance, Two Ensemble Paradigms, The Bagging Algorithm, Random Tree Ensembles- Random Forest.

Practical using R or Python.

Textbooks

1. Tom Mitchell, Machine Learning, McGraw Hill, 1997. (For Modules 1 to 3)
2. Zhi-Hua Zhou, (2012) Ensemble Methods Foundations and Algorithms, Chapman & Hall/CRC (For fourth Module)
3. Pratap Dangeti, Statistics for Machine Learning Techniques for exploring supervised, unsupervised, and reinforcement learning models with Python and R, Packt Publishing; 1st edition (2017)

Reference Books

1. Trevor Hastie, Robert Tibshirani, Jerome Friedman The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, Second Edition.
2. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer

Evaluation: 6 numerical questions each with weightage 10 (marks per question is 50) are to be asked. The student is expected to answer 3 full questions. Use of the packages (R / Python) are allowed for answering the questions in this paper. Examination of 3 hours duration must be conducted in the computer lab under the supervision of an external examiner appointed by the Controller of Examinations.

Question papers shall be generated from university.

DATA VISUALIZATION (ST880402)

Module 1

Purpose of visualization, Cognitive load and clutter, Gestalt principles of visual perception, Identifying and eliminating clutter and improving visual perception, Data Design Concepts

Ref Books:

1. Storytelling with Data: A Data Visualization Guide for Business Professionals, Cole Nussbaumer Knaflic (Ch:3, 4, 5)
2. Tufte, E., & Graves-Morris, P. (2014). The visual display of quantitative information.; 1983.

Module 2

Multidimensional visualization- Visualizing proportions (eg: Histograms, Bar Charts, Pie Charts) and relationships (eg: scatter plot, line chart, area chart, heat maps), tree visualization, graph visualization. Time series data visualization techniques.

Ref Book:

Data Visualization: a successful design process, Andy Kirk (Ch-5)

Module 3

Understanding analytics output and their usage, basic interaction techniques such as selection and distortion, evaluation, Examples of

information visualization applications and systems, user tasks and analysis-visualization packages

Module 4

Grammar of graphics using R-Construct/Deconstruct a graphic into a data-order of accuracy of perceptual tasks and its impact and Case study presentations and lab based on R package of Data Visualizations.

Data Visualization with Python – Matplotlib

References

1. Wickham, H. (2016). Ggplot2: Elegant Graphics for Data Analysis. Springer.2nd Edition
2. Keen, K. J. (2010). Graphics for Statistics andData Analysis with R. CRC Press.
3. Buja, A., Swayne, D. F. & Cook, D., (2007). Interactive and Dynamic Graphics for Data Analysis: with R and Ggobi. Springer Science & Business Media.
4. Dalgaard, P. (2008). Introductory statistics with R. Springer Science & Business Media.
5. Verzani, J. (2014). Using R for introductory statistics. CRC Press.
6. Murrell, P. (2016). R graphics. CRC Press.
7. Cleveland, W. S. (1993). Visualizing data. Hobart Press.
8. Tufte, E. R., Goeler, N. H., & Benson, R. (1990). Envisioning information (Vol. 126). Cheshire, CT: Graphics press.
10. Tufte, E., & Graves-Morris, P. (2014). The visual display of quantitative information.; 1983.

Evaluation: 6 numerical questions each with weightage 10 (marks per question is 50) are to be asked. The student is expected to answer 3 full questions. Use of the packages (R / Python) are allowed for answering the questions in this paper. Examination of 3 hours duration must be conducted in the computer lab under the supervision of an external examiner appointed by the Controller of Examinations.

Question papers shall be generated from university.

ANALYSIS OF CATEGORICAL DATA (ST880403)

Module 1

Categorical variables, Introduction to Binary data, The linear probability models, The logit model, The Probit model, the latent variable approach, the odds ratio, Relarive risks, Sensitivity and specificity, MNemar's test, Binomial response models, log-log models, Likelihood ration Chi-squared statistic, Log-rate models, Time Hazard models, Semi-parametric rate models.

Module 2

Logistic Regression Analysis: Logit Models with Categorical Predictors Logistic Regression models, regression diagnostics, Predictions, Interpreting parameters in logistic Regression. Inference for logistic Regression, Multiple logistic regression.

Module 3

Poisson regression: interpretations, regression diagnostics, Predictions, negative binomial regression, Proportional hazards regression.

Module 4

Principles of Bayesian statistics, Inference using simulations - Standard distributions, Understanding Markov Chain Monte Carlo, The Gibbs sampler and the WinBUGS [Necessary topics from Chapter 1-5 of Ioannis Ntzoufras (2009)]

Textbooks

1)Agresti, A. (1990) Categorical Data Analysis. New York: John Wiley
2)Carlin, B.P. and Louis, T.A. (2000) Bayes and Emperical Bayes Methods for Data Analysis, Second Edition

EPIDEMIOLOGY AND STUDY DESIGNS (ST890401)

Module 1

Basic concepts & Measures of exposure and outcome: What is epidemiology, History of Epidemiology, Emergence of modern epidemiology, Measures of Exposures, Types of exposures, Sources of exposures, Measures of outcome,

Disease registries, Classification of diseases, Measures of disease frequency, Prevalence, Incidence, Risk, Odds of disease, Incidence time, Incidence rate, Relationship between prevalence, rate and risk, Routine data to measure disease occurrence, Age standardization, Direct method of Standardization, Indirect method of standardization, Cumulative rate, Cumulative risk, Proportional incidence.

Module 2.

Overview of study designs: Type of study design, Intervention studies, Cohort studies, Case-control studies, Cross-sectional studies, Ecological studies, Measures of exposure effect, Relative and absolute measures of effect, Confidence intervals and significance tests for measures of occurrence and effect.

Module 3

Case-control studies: Definition of cases and controls, Methods of selecting cases and controls, matching, Sample size, Power calculations, Basic methods of analysis of grouped data, Basic methods of analysis of matched data, Logistic regression for case-control studies, Estimation and interpretation of logistic parameters, Matched analysis- estimation of logistic parameters, unmatched analysis of matched data, confounder score, Categorical data analysis.

Module 4.

Cohort studies: Prospective cohort studies: planning and execution, retrospective cohort, Nested case-control, Case-cohort studies: planning and execution, Matching and efficiency in cohort studies, Cohort studies – statistical analysis, Longitudinal studies: Design, execution and analysis of longitudinal studies, Repeated measurement analysis.

Textbooks

Isabel dos Santos Silva, (1999) Cancer Epidemiology: Principles and Methods, International Agency for Research on Cancer

Reference Books:

1. Ahrens W. and PigcotI. (2005). Handbook of Epidemiology, Springer. Penny Web ,
2. Chiris Bain & Sandi Pirozzo (2005). Essential Epidemiology-An introduction for students & Health Professionals,Cambridge University Press.
3. Rao, K.V. (2007). Biostatistics: A Manual of Statistical Methods for use in Health Nutrition and Anthropology, Raven publishers.
4. Rothman K.I and Greenland S (1998). Modem Epidemiology, Second edition,

5. Lippincott Pressat R. & Atherton A. (1972). Demographic Analysis.
6. Preston S.H.,Heuveline P. & Guillot M. Demography-Measuring and Modelling Population Processes.
7. Sundaram, K.R.(2010) Medical Statistics-Principles & Methods, BI Publications,New Delhi Penny Web ,
8. Chiris Bain & Sandi Pirozzo (2005).Essential Epidemiology-An introduction for students & Health Professionals,Cambridge University press.

POPULATION DYNAMICS (ST890402)

Module 1

Sources of mortality data-mortality measures-ratios and proportions, crude mortality rates, specific rates, standardization of mortality rates, direct and indirect methods, gradation of mortality data, fitting Gompertz and Makeham curves.

Module 2

Life tables-complete life table-relation between life table functions, abridged life table-relation between abridged life table functions, construction of life tables, Greville's formula, Reed and Merrell's formula- sampling distribution of life table functions, multivariate pgf –estimation of survival probability by method of MLE.

Module 3

Fertility models, fertility indices, relation between CBR, GFR, TFR and NRR, stochastic models on fertility and human reproductive process, Dandekar's modified binomial and Poisson models, Brass, Singh models, models for waiting time distributions, Sheps and Perrin model.

Module 4

Population growth indices, logistic model, fitting logistic, other growth models, Lotka's stable population, analysis, quasi stable population, effect of declining mortality and fertility on age structure, population projections, component method-Leslie matrix technique, properties of time independent Leslie matrix-models under random environment.

Textbooks

1.Biswas S (2007) Applied Stochastic Processes-A Biostatistical and Population Oriented Approach, Second Edition, New Central Book Agency.
2.Pollard J.H (1975) Mathematical Models for the growth of Human population, Cambridge University Press.

Reference Books

1)Biswas S (1988) Stochastics processes in Demography and applications, Wiley Eastern.
2)Keyfitz N (1977) Applied Mathematical Demography A Wiley Interscience publication.
3)Ramkumar R (1986) Technical Demography, Wiley Eastern. 4)Srinivasan K (1970) Basic Demographic Techniques and Applications

STATISTICAL TECHNIQUES FOR QUALITY CONTROL (ST890403)

Module 1

Meaning of quality, need for statistical quality control. Meaning and scope of statistical process control. Introduction to Shewhart control charts. Control charts for variables- Xbar – R charts, Xbar – S charts, X – MR charts. Attribute control charts - p, np, c, u charts. 1.4 Concept of OC and ARL for control charts.

Module 2

Modified control charts, Sloping control chart, CUSUM chart, EWMA charts. Economic design of mean charts. Process capability analysis, process capability indices – Cp and Cpk. Introduction to total quality management and six sigma.

Module 3

Concept of sampling plans. Acceptance sampling for attributes - single sampling, double sampling, multiple sampling, and sequential sampling plans. ASN curves. Measuring performance of sampling plans through OC curves. Rectifying inspection plans. AOQ and ATI curves, concept of AOQL.

Module 4

Acceptance sampling by variables. Sampling plan for a single specification limit with known and unknown variance. Performance evaluation through OC curves. Designing a variable sampling plan with a specified OC curve.

Textbooks

1. Montgomery, D.C. (2012). Introduction to Statistical Quality Control, Seventh edition, Wiley.
2. Duncan, A.J. (1986). Quality control and Industrial Statistics, Irwin, Homewood
3. Grant E.L. and Leaven Worth, R.S. (2000) Statistical Quality Control, Seventh edition, TATA McGraw Hill.

Reference Books

2. Mittag, H.J. and Rinne, H. (1993) Statistical Methods for Quality Assurance, Chapman & Hall, Chapters 1, 3 and 4.
3. Rabbit, J T and Bergle, P.A. The ISO 9000 book, Second Edition, Quality resources, Chapter-I
4. Schilling, E.G. (1982) Acceptance Sampling in Quality Control, Marcel Dekker.

PROJECT WORK AND INTERNSHIP (IN A REPUTED INDUSTRY / RESEARCH INSTITUTE) (ST040403)

Every student must undertake project work and internship in a reputed industry or research institute. Project work shall be executed by working outside the regular teaching hours under the supervision of a reputed researcher/ scientist/ an expert faculty in a reputed research institute//industry. At the end of the project work, the candidate must submit 3 copies of the Project Report consisting of the Title of the Study, Objectives, Review of Literature, Materials and Methods, Analysis of Data, Presentation of Results, Applications/ Conclusions, References etc. of about 40-50 pages. Online internships also may be permitted in compliance with the requirements. The guidelines for evaluating the Project Report will be issued by the University. There will be an internal assessment and external assessment for the project work. The external evaluation of the project work is followed by a presentation based on the work and comprehensive Viva

COMPREHENSIVE VIVA-VOCE (ST040404)

In order to assess the overall knowledge in theory and applications as well as general understanding of the different courses studied as part of the program, a Comprehensive Viva-voce shall be conducted at the end of the fourth semester of the programme and it shall cover oral questions from all courses in the program. The Viva-Voce board will consist of an outside expert, the Chairman of the Board of Examiners, and the Head of the concerned Department.