Curriculum for Master's in Statistics Programme (2017 Onwards) Program Code : 74

Department of Statistics
Indira Gandhi National Tribal University
Amarkantak, India

Contents

1 The Curriculum

1.1 The Program

The four semester programme leading to degree of Masters in Statistics, shall, henceforth be termed as the *Program*.

1.2 Courses

Courses refer to various papers that are offered in the program.

1.3 Credit

A Credit is equivalent to an hour of teaching (Lecture (L) / Tutorial (T) per week or two hours of Practicals (P) / Field Work (F) per week. Each course in the program carries credits. The credits attributed to a course indicate the relative weight / importance of the respective course among the set of courses offered in the program.

1.4 Choice Based Credit System (CBCS)

Under the CBCS the learner may opt for courses and earn the required number of *credits* so as to qualify for the award of the degree.

The constituent courses of the program consist of the following types:

1. **Core Courses** (CC): Core Courses are a compulsory requirement for completion of the program.
2. **Elective Courses**: Elective Courses are mandatory courses that have to be chosen by the learner so as to enhance her / his generic proficiency while undergoing the program.

    (a) **Discipline Specific Electives** (DSE):These are courses from Statistics that help in enhancing generic proficiency in specialized areas.

(b) **Open Electives** (OE): These are courses from disciplines other than Statistics and are offered by other departments. Henceforth, OE courses shall be termed as **Generic Elective** (GE) **Courses**.

The learner has to choose the (DSE) courses in consultation with their mentors for semesters - III and IV. The learner is also advised to explore all GE courses available in various departments in Semesters - I and II. Based upon her / his choices she / he may enrol for the chosen GE course(s) for the respective semester.

The program shall be a four semester program with the following course and credit components

1. CC : 76 Credits
2. DSE Courses : 8 Credits
3. GE Courses : 6 Credits

In order to successfully complete the program the student shall earn at least 90 Credits from among CC, DSE and GE courses as mentioned above.

1.5 Admission to the Program

Eligibility Criteria A Bachelors degree from a recognized Indian or foreign university (as per the AIU foreign equivalence list) and secured a minimum of 50% aggregate of marks (45% for SC/ST/PWD/ Transgender and Kashmiri Migrant Students) and having studied Mathematics or Statistics as a major/optional subject at undergraduate level are eligible to apply. To pass the entrance examination conducted by the University the candidates belonging to open category have to secure a minimum of 40% marks [35% for SC/ST and Kashmiri Migrant students/ Transgender].

Number of Seats The number of students to be admitted to the Program are 20.

Reservation The reservation of seats shall be as per Government of India norms.

1.6 An Overview of the Syllabus

Introduction The programme for Masters in Statistics, henceforth termed as the Program, aims to impart domain specific knowledge to the learner. Additionally it attempts to inculcate skills into her/him that are beneficial not only for the individual's personal growth but also for the development of the society at large. Whence, the learner as a potential domain expert may grow into efficient appliers and creators of the domain knowledge. Further, it is expected that these experts should be able to work with experts of other domains as tackling the challenges of a growing complex society requires interdisciplinary approaches to problems.

Expected Learning Outcomes The learner at the post graduate level is expected to be well equipped with all necessary skills that make out of him a competent statistician. These include a firm base in mathematical underpinnings, the statistical inference and computational expertise to apply theory for extracting information from large data sets. A specialization in one of the applied fields like Biostatistics, Quantitative Social Sciences, Environmental Sciences, Actuarial Sciences, Industrial Statistics and Operations Management or Economic and Financial Statistics gives an additional advantage.

Computational Expertise As mentioned above computational skills are an added advantage for a practitioner. Therefore data centric programming skills with a knowledge of C/C++/R are essential tools for a statistician. Keeping this in view the curriculum emphasizes R - programming as acquaintance with *R* and its IDE *RStudio* have an edge over other statistical software due to the following reasons:

1. Open source virtually comes without any cost.
2. In tune to STATISTICAL THINKING must for learners of statistics.
3. Plenty of help available online. 5000+ packages.
4. Can be learned easily by self - learning.
5. Most powerful. Whatever is in Statistics is available in R.
6. Updated regularly.

7. Everyone can work with it after an initial training [even non-programmers].
8. Conducive to development of statistical-analytical abilities among learners.
9. Most of the proprietary software provide R interface.

Outline of the Syllabus

The core courses offered in *Semester-I* pertaining to *Mathematical Analysis*, *Probability Theory* and *Distribution Theory* aim at firming the foundation for the study of courses that follow in the subsequent semesters.

The courses mentioned above along with the core courses offered in *Semester-II* namely *Linear Algebra* and *Sampling Distributions* aim at laying firm foundations for taking up courses in *Statistical Inference* , *Design of Sample Surveys / Experiments*, Inferring from data arising out of *Sample Surveys / Designed Experiments*, courses based on *Linear Models*, *Multivariate Statistical Analysis* and *Time Series Analysis*.

The computational aspects of Statistics have been addressed through a series of *Statistical Computing and Data Analysis* courses numbered *I-IV* amounting to 16 credits. The *Statistical Computing and Data Analysis* courses also take care of the practical aspects of the theory courses offered in the respective semesters. In addition, the learner shall opt for minimum two courses in specialized areas in the form of *Discipline Specific Electives*.

The Dissertation to be submitted at the end of the Semester-IV shall be treated as an ambiance to further research.

1.7 Core Courses

S. No.	Title of the Course	Code	L-T-P-C
1	Analysis and Multi-variable Calculus	MSTAT-CC-01	3-1-0-4
2	Probability Theory	MSTAT-CC-02	3-1-0-4
3	Distribution Theory	MSTAT-CC-03	3-1-0-4
4	R Programming	MSTAT-CC-04	3-1-0-4
5	Statistical Computing and Data Analysis - I	MSTAT-CC-05	2-0-2-4
6	Linear Algebra	MSTAT-CC-06	3-1-0-4
7	Inference - I	MSTAT-CC-07	3-1-0-4
8	Sampling Distributions	MSTAT-CC-08	3-1-0-4
9	Design of Sample Surveys	MSTAT-CC-09	3-1-0-4
10	Statistical Computing and Data Analysis - II	MSTAT-CC-10	2-0-2-4
11	Inference - II	MSTAT-CC-11	3-1-0-4
12	Survival Analysis	MSTAT-CC-12	3-1-0-4
13	Regression Analysis	MSTAT-CC-13	3-1-0-4
14	Statistical Computing and Data Analysis - III	MSTAT-CC-14	2-1-0-4
15	Multivariate Statistical Analysis	MSTAT-CC-15	3-1-0-4
16	Design and Analysis of Experiments	MSTAT-CC-16	3-1-0-4
17	Time Series Analysis	MSTAT-CC-17	3-1-0-4
18	Statistical Computing and Data Analysis - IV	MSTAT-CC-18	2-1-0-4
19	Masters Dissertation	MSTAT-CC-19	0-0-0-4
20	Summer Internship (optional non-credit)	MSTAT-CC-20	0-0-2-2
21	Bridge Course (non - credit)	MSTAT-CC-00	2-0-2-0

1.8 Discipline Specific Elective Courses

S. No.	Title of the Course	Code	L-T-P-C
1	Statistical Methods in Demography	MSTAT-DE-01	3-1-0-4
2	Bayesian Inference	MSTAT-DE-02	3-1-0-4
3	Analysis of Directional Data	MSTAT-DE-03	3-1-0-4
4	Spatial Data Analysis	MSTAT-DE-04	3-1-0-4
5	Statistical Genetics	MSTAT-DE-05	3-1-0-4
6	Stochastic Processes	MSTAT-DE-06	3-1-0-4
7	Bioassay	MSTAT-DE-07	3-1-0-4
8	Clinical Trials	MSTAT-DE-08	3-1-0-4
9	Quantitative Epidemiology	MSTAT-DE-09	3-1-0-4
10	Advanced Statistical Computing	MSTAT-DE-10	3-1-0-4
11	Operations Research	MSTAT-DE-11	3-1-0-4
12	Statistical Quality Assurance	MSTAT-DE-12	3-1-0-4
13	Econometrics	MSTAT-DE-13	3-1-0-4
14	Categorical Data Analysis	MSTAT-DE-14	3-1-0-4
15	Fuzzy Sets and Systems	MSTAT-DE-15	3-1-0-4

1.9 Generic Elective Courses Offered to Students Pursuing Masters Programmes other than Statistics

S. No.	Semester	Code	Title	L-T-P-C
1	I	MSTAT-GE-01	Epidemiology and Vital Statistics	2-1-0-3
2	I	MSTAT-GE-02	Probability Theory and Models	2-1-0-3
3	II	MSTAT-GE-03	Designing Scientific Studies	2-1-0-3
4	II	MSTAT-GE-04	Scientific Computing	2-0-1-3
5	I	MSTAT-GE-05	Descriptive Data Analysis with R	2-0-1-3
6	II	MSTAT-GE-06	Inferential Data Analysis with R	2-0-1-3
7	II	MSTAT-GE-07	Big Data and Analytics	2-0-1-3

1.10 Scheme of Evaluation

Students undergoing a course shall be examined on a continuous basis by the course instructor during the semester in which the course is conducted. Broadly there shall be *Continuous Internal Evaluations (CIE)* during the semester and an *Semester-End Examination (SEE)* at the completion of the course. The distribution of marks over the CIE and SEE for CC / DSE Courses is as follows:

S. No.	Type of Assessment	Marks
1	CIE	40
2	SEE	60

Aiming to assess values, skills and knowledge imbibed by students, *CIE* and the designing of *SEE* is to be done by the respective course instructor. *CIE* may have different components for internal evaluation with marks distribution given as follows:

S. No.	Mode of Assessment	Marks
1	Term Paper / Sessional Tests	20
2	Other modes	20

The pattern and schedule of Continuous Internal Assessment and evaluation need to be decided by the concerned Faculty and made known to all students.

The distribution of marks for Summer Internship / Masters Dissertation is as follows:

S. No.	Type of Assessment	Marks
1	Periodical Presentation	60
2	Concise Dissertation	20
3	Viva - voce	20
	Total Marks	100

Dissertation / project report will be valued jointly by the supervisor and an examiner external to from a neighboring institution within the city/state/country.

1.11 Semester-wise Structure of the Programme

Semester	Title of the Course	Code	L-T-P-C
I	Analysis and Multi-variable Calculus	MSTAT-CC-01	3-1-0-4
	Probability Theory	MSTAT-CC-02	3-1-0-4
	Distribution Theory	MSTAT-CC-03	3-1-0-4
	R Programming	MSTAT-CC-04	3-1-0-4
	Statistical Computing and Data Analysis - I	MSTAT-CC-05	2-0-2-4
	Generic Elective - I		x-x-x-3
II	Linear Algebra	MSTAT-CC-06	3-1-0-4
	Inference - I	MSTAT-CC-07	3-1-0-4
	Sampling Distributions	MSTAT-CC-08	3-1-0-4
	Design of Sample Surveys	MSTAT-CC-09	3-1-0-4
	Statistical Computing and Data Analysis - II	MSTAT-CC-10	2-0-2-4
	Generic Elective - II		x-x-x-3
II	Summer Internship (optional non-credit)	MSTAT-CC-20	0-0-2-2
III	Inference - II	MSTAT-CC-11	3-1-0-4
	Survival Analysis	MSTAT-CC-12	3-1-0-4
	Regression Analysis	MSTAT-CC-13	3-1-0-4
	Statistical Computing and Data Analysis - III	MSTAT-CC-14	2-0-2-4
	Discipline Specific Elective - I		x-x-x-4
IV	Multivariate Statistical Analysis	MSTAT-CC-15	3-1-0-4
	Design and Analysis of Experiments	MSTAT-CC-16	3-1-0-4
	Time Series Analysis	MSTAT-CC-17	3-1-0-4
	Statistical Computing and Data Analysis - IV	MSTAT-CC-18	2-0-2-4
	Discipline Specific Elective - II		x-x-x-4
IV	Masters Dissertation	MSTAT-CC-19	0-0-0-4

2 Detailed Syllabus

2.1 Core Courses (CC)

2.1.1 MSTAT-CC-01 Analysis and Multi-variable Calculus

Sets and Metric Spaces Finite, countable and uncountable sets;the set of Real numbers; Metric Spaces: compactness, connectedness and completeness; Bolzano Weierstrass theorem, Heine Borel theorem. Normed linear Spaces. Spaces of continuous functions as examples.

Functions on Metric Spaces limits, continuity, uniform continuity, derivatives and mean value theorem; Monotonic functions, types of discontinuity, functions of bounded variation; Sequences as functions, convergence of sequences; series and sequence of partial sums, convergence of series;

Sequence of Functions Sequences and series of functions; Uniform and point-wise convergence, consequences of uniform convergence.

Integration Riemann - Stiltjes integral; fundamental theorem of calculus; improper integrals, beta function, gamma function; Differentiation under the sign of integral - Leibnitz rule.

Multi-variable Calculus Scalar fields and vector fields; Limit and continuity; directional derivative, partial derivative, derivative as a linear functions, inverse and implicit function theorems; maxima-minima using legrangian multiplier method; Evaluation of multiple integrals; transformation of variables and integration.

Texts

1. Apostol, T. M. (1985). Mathematical Analysis, Second Edition, Narosa Publishing House, New Delhi.
2. Apostol, T. M. (2007). Calculus Vol. 2, John Wiley & Sons.
3. Rudin, Walter. (1976). Principles of Mathematical Analysis, McGraw Hill.

References

1. Apostol, T. M. (2007). Calculus Vol. 1, John Wiley & Sons.

2. Bartle, R. G. and Sherbert, D. R. (2000). Introduction to Real Analysis, 3rd edition, John Wiley & Sons, Inc., New York.

3. Ghorpade, Sudhir R. and Limaye, Balmohan V. (2006). A Course in Calculus and Real Analysis, Undergraduate Textss in Mathematics, Springer (SIE), Indian reprint.

4. Tao, T. Analysis, Vol. I, Hindustan Book Agency, Delhi, India.

5. Tao, T. Analysis, Vol. II, Hindustan Book Agency, Delhi, India.

2.1.2 MSTAT-CC-02 Probability Theory

Probability Measure and Distribution Functions Probability space of a random experiment, probability measures, random variables as a measurable function. Field induced by a sequence of random variables, decomposition of distribution functions in purely discrete, absolutely continuous and singular components, Lebesgue measure, Lebesgue integration and Lebesgue - Steltjes measure.

Probability Inequalities CR-inequality, Cauchy-Schwartz inequality, Holder inequality, Minkowski inequality, Jensen inequality, Lyapunov inequality, Kolmogorov inequality, Hajck-Renyki inequality.

Convergence Convergence of Sequences of distribution functions, Helly - Bray theorem, Different types of convergence of sequence of random variables; Weak and strong law of large numbers, Khinchin. Borel and Kolmogorav theorems.

Characteristic Function and Central Limit Theorems Borel-Cantelli lemmas and zero-one law, Characteristic function, Inversion theorem, Continuity theorem, One dimensional central limit problem: lindeberg-levy, Lyapunov, Lindeberg-Feller theorems.

Texts

1. Ash, R. (1972). Real Analysis and Probability. Academic Press.
2. Burrill C.W. : Measure, Integration and Probability.
3. Chang K.L.: A Course in Probability Theory.
4. Loeve M. : Probability Theory.
5. Rohatgi, V.K. and Saleh, A.K. Md. E. (2001). An Introduction to Probability and Statistics. Wiley.

References

1. Billingsley, P. Probability and Measure.
2. Dudley, R. M. (1989). Real Analysis and Probability, Wadsworth and Brooks/Cole.
3. Kingman, J F C and Taylor, S. J. (1966). Introduction to Measure and Probability. Cambridge University Press.
4. Medhi, J. (2009). Stochastic Processes, New Age International Publishers.

2.1.3 MSTAT-CC-03 Distribution Theory

Discrete Distributions Uniform, Bernoulli, Binomial, Poisson, Negative Binomial, Multinomial, Hyper Geometric, Geometric, their characteristics and simple applications.

Continuous Distributions Continuous Distributions - Uniform, Normal, Exponential, Gamma, Chi-Square, Pareto, Lognormal, Logistic distributions, Laplace, Beta and Weibull - their characteristics and applications.

Bivariate Distributions Bivariate Binomial - Bivariate Poisson - Bivariate Normal - Bivariate Exponential of Marshall and Olkin, Multivariate Normal Distribution (Definition and concept only).

Compound Distributions - Compound Binomial, Compound Poisson and Compound Negative Exponential (Pareto) distributions, Concept of Truncation Zero Truncated Binomial and Poisson distributions their applications.

Concept of Convolution Mixture of Distributions, Extreme Value Distributions and Simulations

Texts

1. 1. Mood M., Graybill F.A. and Boes D.C. (2001) : Introduction to the Theory of Statistics, Tata McGraw-Hill, New Delhi.

2. 2. Johnson, N.L, Kotz, S. and Balakrishnan, N. (1994): Continuous Univariate Distributions, Vol. 1 & 2, Wiley Series in Probabilty and Statistics.

3. 3. Johnson, N.L , Kemp A.W. & Kotz, S. (1994): Univariate Discrete Distributions, Wiley Series in Probabilty and Statistics.

References

1. Rao C. R., (1973): Linear Statistical Inference and its Applications, Wiley Eastern Ltd, New Delhi.

2. Dudewicz, E.J and Mishra, S.N (1980): Mathematical Statistics, John Wiley, NY.

3. Kocherlakota S and Kocherlakota K (1992): Bivariate Discrete distributions, M. Dekker.

4. Balakrishnan N and Lai C.D. (2009): Continuous Bivariate Distributions, Springer.

5. Rohatgi, V.K. and Saleh (2002): An Introduction to Probability Theory and Mathematical Statistics, John Wiley.

6. Parimal Mukhopadhyay(2006): Mathematical Statistics, 3/e, Books and Allied (P) Ltd, Kolkata.Anderson, T.W. (1987). An Introduction to Multivariate Statistical Analysis, 2^{nd} edn., Wiley.

2.1.4 MSTAT-CC-04 R Programming

R language Essentials Expressions and objects, Assignments, creating vectors, vectorized arithmetic, creating matrices, operations on matrices, lists, data frames - creation, indexing, sorting and conditional selection ; examples.

R Programming: conditional statements if and if else; loops - for, while, do-while; functions - built-in and user defined; Data entry - reading from text file, data editor; examples.

Descriptive Statistics and Graphics Obtaining summary statistics; generating tables; Bar plots, Pie charts, Box plots, Histogram; exercises.

Probability and Distributions Random sampling and combinatory; obtaining density, cumulative density and quantile values for discrete and continuous distributions; generating samples from discrete and continuous distributions; Plotting density and cumulative density curves; Q-Q plot.

Correlation Pearson, Spearman and Kendalls correlation; Regression fitting, obtaining residuals and fitted values; one and two sample tests for mean and variance one way and two way ANOVA.

Text

1. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R, Springer.

2. Press, W. H., Teukolsky, S. A., Vellering, W. T. and Flannery, B. P. (1993). Numerical Recipes in C, II edition, Cambridge University Press.

3. Ross, S. M. Simulation.

References

1. Chambers, J. M. and Hastie, T. J (Editors).(1997). Statistical Models in S. Chapman and Hall.

2. Venables, W. N. and Ripley, B. D. (2000). S Programming, Springer.

3. www.r-project.org

2.1.5 MSTAT-CC-05 Statistical Computing and Data Analysis - I

Lab Exercises based on SPSS

1. Basics Import and Export of data files, Recoding, computing new variables Descriptive statistics.
2. Selection of cases, splitting and merging of files.
3. Bar and Pie charts.
4. Box plots for single and multiple groups.
5. Density and cumulative density plots for Binomial, Poisson, Normal and exponential distributions.
6. Checking Normality using Histogram and Q-Q plot.
7. Computation of simple, multiple, partial and rank correlation coefficients.
8. Computation of simple regression.
9. Fitting of curves Linear, parabola, cubic and exponential.
10. Testing of Hypothesis t, F, Chi square
11. One way and Two way ANOVA.

Lab Exercises based on R

1. Operations on vectors and matrices
2. Creating and manipulating data frames.
3. Writing user defined functions for finding arithmetic mean, median, factorial, matrix addition and multiplication.

4. Bar and Pie charts.
5. Box plots for single and multiple groups.
6. Density and cumulative density plots for Binomial, Poisson, Normal and exponential distributions.
7. Checking Normality using Histogram and Q-Q plot.
8. Correlation coefficient Pearsons, Spearman and Kendalls Tau.
9. Fitting simple linear and multiple linear regressions.
10. One sample and two sample t test.
11. One way and Two way ANOVA.

Texts

1. Crawley, J.(2007) *The R Book by Michael*, John Wiley and Sons.
2. Goon, A.M., Gupta, M.K. and Das Gupta, B. (2005). *Fundamentals of Statistics, Vol. I*, The World Press Pvt. Ltd., Calcutta.
3. Dalgaard, P. *Introductory Statistics with R* , Springer, 2nd edition, 2008.
4. Ross, S. M. *Simulation*.

References

1. Chambers, J. M. and Hastie, T. J (Editors).(1997). *Statistical Models in S*. Chapman and Hall.
2. Chambers, J. M.(1998). *Programming with Data: A Guide to S Language*. Springer.
3. Venables, W. N. and Ripley, B. D. (2000). *S Programming*, Springer.
4. *www.r-project.org*

2.1.6 MSTAT-CC-06 Linear Algebra

Vector Spaces Fields, vector spaces, subspaces, linear dependence and independence, basis and dimension of a vector space, finite dimensional vector spaces, completion theorem; Vector spaces with an inner product, Gram-Schmidt orthogonalization process, orthonormal basis and orthogonal projection of a vector.

Linear Transformations and Matrices Linear Transformations: Matrix representation of linear transformations. Change of basis, canonical forms, diagonal forms, triangular forms, Jordan forms; Algebra of matrices, row and column spaces of a matrix, elementary matrices, determinants, rank and inverse of a matrix, null space and nullity, partitioned matrices, Kronecker product; linear equations.

Characteristic Roots and Vectors Cayley - Hamilton theorem, minimal polynomial, similar matrices, algebraic and geometric multiplicity of a characteristic root, spectral decomposition of a real symmetric matrix, reduction of a pair of real symmetric matrices, Hermitian matrices.

Generalised Inverse of a Matrix Hermite canonical form, generalized inverse, Moore-Penrose generalized inverse, Idempotent matrices, Solutions of matrix equations.

Texts

1. Hoffman, K. and Kunze, R. (1971). Linear Algebra, 2nd ed., Prentice Hall, Inc.

2. Graybill, F.A. (1983): Matrices with Applications in Statistics, 2nd Ed. Wadsworth.

3. Rao, A. R.and Bhimsankaram,P. Linear Algebra, Hindustan Book Agency, Delhi, India.

4. Searle, S.R. (1982): Matrix Algebra for Statistical Applications, John Wiley and Sons inc.,

References

1. Bellman, R. (1970). Introduction to Matrix Analysis, 2nd ed. McGraw Hill.

2. Halmos, P.R. (1958): Finite Dimensional Vector Spaces, 2nd ed. van. Nortrand Company Inc.

3. Rao, C. R. (1973). Linear Statistical Inference and its Applications, Wiley.

4. Rao, C. R. and Mitra, S. K. (1971). Generalized Inverse of Matrices and its Applications, John Wiley and Sons, Inc.

5. Shanti Narain: A text book of matrices, S. Chand and Company (Pvt.) Ltd.

2.1.7 MSTAT-CC-07 Inference - I

Parametric Point Estimation properties of estimators - Consistency and its different forms , sufficient condition for consistency, Unbiasedness, sufficient statistics, factorization theorem, distributions admitting sufficient statistic, Exponential and Pitman families procedure for finding minimal sufficient statistic.

The Information Measure Cramer - Rao Bound, Chapman - Robbins Bound (single parameter case only), Bhattacharya Inequality, Minimum Variance Bound Estimator.

Uniformly Minimum Variance Unbiased Estimators (UMVUE) condition for the existence of UMVUE, completeness and bounded completeness, relation between complete statistic and minimal sufficient statistic, Rao - Blackwell Theorem, Lehmann - Scheffes theorem.

Methods of Estimation method of moments and its properties, method of maximum likelihood and its properties, Large sample properties of MLE, Methods of least squares - Optimum properties of least square estimates in linear model.

Interval Estimation Pivotal method of construction, shortest confidence intervals and their construction (minimum average width), construction of shortest confidence intervals in large samples; Decision Theory: Simple problems involving quadratic error loss function, Elementary notions of minimax estimation: simple illustrations.

Texts

1. Casella, G and Berger, R. L. (2002). *Statistical Inference*, 2nd edition. Duxbury Press

2. Kale, B.K. (1999). *A first course in Parametric Inference*, Narosa Publishing House.

3. Rohatgi, V.K. (1986): *Statistical Inference*, Wiley Eastern Ltd,.

4. Zacks, S. (1981). *Parametric Statistical Inference*, John Wiley, NY.

References

1. Goon, A. M, Gupta, M. K, and Das Gupta, B.C (1980). *An outline of Statistical Theory, Vol. II*,The World Press, Calcutta.

2. Lehmann, E .L. and Casella, G. (1998).*Theory of Point Estimation*, Springer International.

3. Mood, A.M., Graybill, F.A and Boers, D.C (1974). *Introduction to Theory of Statistics*, Mc Graw-Hill Book Company.

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.

2.1.8 MSTAT-CC-08 Sampling Distributions

Sampling distributions of sampling mean; chi-square, t and F distributions and their interrelations and properties and applications based on them; Non-central Sampling Distributions; Non-central chi-square, t and F distributions and their properties and applications.

Quadratic Forms Distributions of quadratic forms under normality-independence of quadratic forms and linear form- Cochran's Theorem.

Order Statistics Distributions and properties of order statistics - Joint and marginal distributions of order statistics - Distribution of range and mid range - Extreme values and their asymptotic distributions (concepts only).

Text

1. Rao, C.R. (1973). Linear Statistical Inference and Its Applications, 2/e, Wiley Eastern.

2. Rohatgi, V.K. and Saleh, A.K. Md. E. (2001). An Introduction to Probability and Statistics. Wiley.

3. Searle, S.R. Linear Models, John Wiley & Sons.

References

1. Anderson, T.W. (1987). An Introduction to Multivariate Statistical Analysis, 2^{nd} edn., Wiley.

2. Kotz, S., N. Balakrishnan and N. L. Johnson. (2005). Univariate Discrete Distributions 3rd Edition, John Wiley & Sons.

3. Kotz, S., N. Balakrishnan and N. L. Johnson. (2005). Continuous Univariate Distributions 2nd Edition, Volume 1, John Wiley & Sons.

4. Kotz, S., N. Balakrishnan and N. L. Johnson. (2005). Continuous Univariate Distributions 2nd Edition, Volume 2, John Wiley & Sons.

5. Kotz, S., N. Balakrishnan and N. L. Johnson. Discrete Multivariate Distributions, John Wiley & Sons.

6. Kotz, S., N. Balakrishnan and N. L. Johnson. (2005). Continuous Multivariate Distributions, Volume 1, Models and Applications, John Wiley & Sons.

7. McCullagh, P. and Nelder, J. A. Generalized Linear Models.

2.1.9 MSTAT-CC-09 Design of Sample Surveys

Preliminaries Sampling Designs: Simple random sampling; Stratified Random Sampling, Allocation problems; Systematic Sampling Schemes - Linear.

Ratio Estimators and Regression Estimators Ratio estimators and their properties in Simple Random Sampling - Ratio estimators in Stratified Random sampling - Regression Estimators

Double Sampling Double sampling on successive occasions, double sampling for stratification; cost and variance functions;

Two-Stage Sampling Concept of multistage sampling; Two-stage sampling with equal number of second stage units. Two-stage sampling with unequal number of second stage units.

Cluster Sampling Equal cluster sampling - Estimators of mean and variance, optimum cluster size, Unequal cluster sampling - Estimators of mean and variance; Idea of small area estimation.

Official Statistical System in India Indian Official Statistics System; Methods of collection of official statistics, their reliability and limitations; Designs of large sale sample surveys conducted by NSSO. Government of India's Principal publications containing data on the topics such as population, industry, agriculture, health and finance.

Texts

1. Cochran, W.G. (1999). *Sampling Techniques*, Third edition, John Wiley & Sons.

2. Des Raj and Chandhok (1998). *Sampling Theory*, Narosa.

3. Gupta, A. K. and Kabe D.G, (2011). *Theory of Sample Surveys*, World Scienific Publishing Co. Pte. Ltd., Singapore.

4. Mukhopadhay, P. (2009). *Theory and Methods of Survey Sampling*, Second edition, PHI Learning Pvt Ltd., New Delhi.

5. Singh, D. and Chaudhary, F. S. (1986). *Theory and Analysis of Sample Survey Designs*. New Age International Publishers.

6. Sukhatme, P. V. et al. (1984). *Sampling Theory of Surveys with Applications*. Iowa State Univ. Press.

7. Mukhopadhyay, P. (1998). *Small Area estimation in Survey Sampling*, Narosa.

References

1. Chaudhuri, A. and J.W.E. Vos (1988). *Unified Theory and Strategies of Survey Sampling*, North-Holland, Amsterdam.

2. A. and R. Mukerjee (1988). *Randomized Response : Theory and Techniques*, New York : Marcel Dekker Inc.

3. Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). *Sample Survey Methods and Theory*, Volume II, John Wiley.

4. Heyday, A. S. and Sinha, B. K. (1991). *Design and Inference Sampling in Finite Population*. Wiley.

5. Kish, L. (1995). *Survey Sampling*, John Wiley and Sons.

6. Mukhopadhyay, P. (1996). *Inferential Problems in Survey Sampling*, New Age International (P).

7. Murthy, M. N. (1977). *Sampling Theory & Methods*, Statistical Publishing Society, Calcutta.

8. Sarjinder Singh (2004). *Advanced Sampling - Theory with Applications*, Kluwer Publications.

9. Sampath, S. (2001). *Sampling Theory and Methods*, Alpha Science International Ltd., India.

10. Wolter, K. M. (1985). *Introduction to Variance Estimation*, Springer-Verlag.

2.1.10 MSTAT-CC-10 Statistical Computing and Data Analysis - II

Prerequisites *Statistical Computing and Data Analysis - I*

Theory of Estimation

1. MLE and Standard error of ML estimators.
2. MLE through the method of successive approximation.
3. Method of Moments.
4. Method of Least square
5. Interval Estimation: Confidence interval for mean, difference of means and ratio of variances.

Design of Sample Survey

1. Simple random sampling without replacement - Estimation of the population total and its variance.
2. Ratio, Regression and Difference estimation.
3. Stratified sampling SRS- Equal, Proportional allocations.
4. Linear and circular systematic sampling.
5. Cluster sampling of equal sizes

Texts

1. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R, Springer.
2. Press, W. H., Teukolsky, S. A., Vellering, W. T. and Flannery, B. P. (1993). Numerical Recipes in C, II edition, Cambridge University Press.
3. Ross, S. M. Simulation.

References

1. Chambers, J. M. and Hastie, T. J (Editors).(1997). Statistical Models in S. Chapman and Hall.

2. Venables, W. N. and Ripley, B. D. (2000). S Programming, Springer.

3. www.r-project.org

2.1.11 MSTAT-CC-11 Inference - II

Tests of Hypotheses Statistical hypothesis testing: Simple and Composite hypothesis, Null and Alternative hypothesis; Types of errors; Critical region; Level of significance; Power of a test; Most powerful test; Simple problems for calculating probability of Type I and Type II errors and power of the test; Randomized and non-randomized tests; Neyman - Pearson fundamental lemma; Most powerful tests; Uniformly most powerful test; Uniformly most powerful test for distributions with monotone likelihood ratio; Relation between Testing of Hypothesis and Interval Estimation - Uniformly Most Accurate interval.

Likelihood Ratio Test (LRT) Unbiasedness for hypothesis testing, Uniformly most powerful unbiased tests, Unbiased tests for one parameter exponential family; Similar test and complete sufficient statistics; Similar tests with Neyman structure, Locally most powerful tests. Likelihood ratio test, its properties and its asymptotic distribution, Applications of the LR method.

Non-Parametric Tests Goodness of fit test: Chi-square and Kolmogorov Smirnov test - Test for randomness, Wilcoxon Signed rank test - Two sample problem: Spearman and Kendall Tau Tests, Kolmogrov - Smirnov test, Wald - Wolfowitz run test, Mann - Whitney U test, Median test, K Samples - Kruskal Wallis test and Friedman test.

Sequential Tests Basic Structure of Sequential tests Sequential Probability Ratio Test (SPRT) and its applications Determination of the boundary constants; OC andASN Curves.

Texts

1. Casella, G., and Berger, R. L. (2002). Statistical Inference, 2^{nd} edition. Duxbury Press.

2. Gibbons, J.D. (2003). Nonparametric Statistical Inference, 4^{th} Edition, Marcel Dekker.

3. Kale, B.K. (1999). A first course in Parametric Inference, Narosa Publishing House.

4. Lehmann, E L. (1986). Testing Statistical Hypotheses, Springer.

5. Rohatgi, V.K. and Saleh, A.K. Md. E. (2001). An Introduction to Probability and Statistics. Wiley.

References

1. Ghosh, B.K (1970): Sequential Tests of Statistical Hypotheses, Addison Wesley.

2. Parimal Mukhopadhyay (2006): Mathematical Statistics, 3/e, Books and Allied (P) Ltd, Kolkata.

3. Rao, C. R. (1973). Linear Statistical Inference and its Applications, Wiley.

4. Randles, R.H. and Wolfe, D.A. (1979). Introduction to the Theory of Nonparametric Statistics, Wiley.

5. Wald, A (1949): Sequential Analysis, John Wiley, NY.

2.1.12 MSTAT-CC-12 Survival Analysis

Basic concepts Concepts of time, order and random censoring and likelihood in these cases - Life distributions Exponential, Gamma, Weibull, Lognormal, Pareto, Linear Failure rate Parametric inference (Point estimation, Scores, MLE).

Life tables failure rate, mean residual life and their elementary properties - Ageing classes and their properties - Bathtub Failure rate.

Estimation Estimation of survival function Actuarial Estimator - Kaplan-Meier Estimator - Estimation under the assumption of IFR / DFR - Tests of exponentiality against non- parametric classes Total time on test, Deshpande test.

Two Sample Problems Two sample problem: Gehan test, Log rank test. Mantel Haenszel test, Tarone Ware tests. Semi- parametric regression for failure rate Coxs proportional hazards model with one and several convariates - Rank test for the regression coefficients.

Texts

1. Miller, R.G. (1981) : Survival analysis, John Wiley
2. Cox, D.R. and Oakes, D. (1984): Analysis of Survival Data, Chapman and Hall, New York.
3. Elisa T.Lee, John Wenyu Wang and Timothy Wenyu Patt (2003): Statistical Methods for Survival Data Analysis, 3/e, Wiley Inter Science.
4. Klein P. John and Moeschberger(2003): Survival Analysis: Techniques for Censored and Truncated Data, 2/e, Springer.

References

1. Gross, A.J. and Clark, V.A. (1975): Survival distribution : Reliability applications in the Biomedical Sciences, John Wiley and Sons.
2. Elandt Johnson, R.E. Johnson N.L.(1999): Survival Models and Data Analysis, John Wiley and sons.
3. Kalbfleisch J.D. and Prentice R.L.(2003), The Statistical Analysis of Failure Time Data, John Wiley.
4. Lawless J.F. (2002): Statistical Models and Methods for Life Time Data, 2/e, John Wiley & Sons.
5. Xian Liu(2012): Survival Analysis Models and Applications- John Wiley & Sons.

2.1.13 MSTAT-CC-13 Regression Analysis

Prerequisites *Linear Inference, Statistical Computing and Data Analysis - I*

Simple Linear Regression Assumptions, least square (LS) estimators of parameters, standard error of estimators, testing of hypothesis for coefficient of regression, s.e. of prediction, testing of hypotheses about parallelism (Slopes), equality of intercepts, congruence, extrapolation, optimal choice of independent variables, diagnostic checks and correction: graphical technique, tests for normality , uncorrelatedness, homoscadasticity, lack of fit.

modifications like polynomial regression, transformations of dependent or independent variables , weighted LS, inverse regression.

Multiple Regression Standard Gauss-Markov setup, least square estimation, error and estimation spaces, variance and covariance of LS estimators, properties of LS estimators, estimation of error variance, case with correlated observation, LS estimation with restriction on parameters, simultaneous estimation of linear parametric functions, testing of hypothesis for one and more than one linear parametric functions, confidence intervals and regions. Mallows Cp, forward, backward selection method.

Regression Diagnostics Multicollinearity: consequences, detection and remedies, autocorrelation consequences, Durbin Watson test, estimation of parameters in autocorrelation. Multiple correlation, adjusted multiple correlation coefficient, null distribution of simple correlation and multiple correlation coefficient, partial correlation coefficient and its relation with multiple correlation coefficient, test for significance of simple , multiple and partial correlation coefficients, variable selection procedures. Residual and residual diagnostics, transformation of variables: Box- Cox power Transformation, generalized weighted least sequence.

Non-linear Regression Non-linear least squares transformation to a linear model, statistical inference in non-linear regression; Logistic regression: Logit transform, ML estimation, tests of hypothesis, Wald test, LR test, score test, test for overall regression, introduction to link functions such as binomial, inverse binomial, inverse Gaussian and Gamma.

Texts

1. Draper, N.R. and Smith H. (1998). Applied Regression Analysis, 3^{rd} Ed. Wiley.

2. Hosmer, D. W. and Lemeshow, S. (1989) Applied logistic regression, John Wiley

3. McCullagh, P. and Nelder, J. A.(1989) Generalized linear models, Chapman and Hall

4. Neter, J.; Wasserman, W. and Kutner, M.H.(1985) Applied linear statistical models

5. Ratkowsky, D. A.(1983) Nonlinear regression modeling (Marcel Dekker

6. Weisberg, S. (1985). Applied Linear Regression, Wiley.

References

1. Cook, R.D. and Weisberg, S. (1982). Residuals and Inference in Regression, Chapman and Hall.

2. Montogomery, D. C., Peck, E. A. and Vining, G. G. Introduction to Linear Regression Analysis.

3. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R, Springer.

4. Ryan, T. P. Modern Regression Methods.

5. Searle, S.R. Linear Models, John Wiley & Sons.

2.1.14 MSTAT-CC-14 Statistical Computing and Data Analysis - III

Prerequisites *Statistical Computing and Data Analysis - I, Statistical Computing and Data Analysis - II*

Testing of Statistical Hypothesis

1. Construction of randomized and nonrandomized MP, UMP and UMPU tests of hypotheses and drawing the power curves.

2. Construction of SPRT and its OC and ASN curves. Based on R

3. Non parametric tests: Runs Test, Kolmogorov - Smirnov test, Mann - Whitney U test, Wilcoxon Sign Test, Chi-Square Test for independence of Attributes, Median test, Kruskal Wallis test and Friedmans test.

Regression

1. Simple Regression
2. Multiple Regression
3. Regression Diagnostics
4. Logistic Regression
5. Poisson Regression

Texts

1. Johnson, R.A. and Wichern, D.W. (1990). Applied Multivariate Statistical Analysis, Prentice Hall.
2. Montogomery, D. C., Peck, E. A. and Vining, G. G. Introduction to Linear Regression Analysis.

References

1. Chambers, J. M. and Hastie, T. J (Editors).(1997). Statistical Models in S. Chapman and Hall.
2. Venables, W. N. and Ripley, B. D. (2000). S Programming, Springer.
3. www.r-project.org

2.1.15 MSTAT-CC-15 Multivariate Statistical Analysis

Prerequisites *Inference - I and Inference - II*

Multivariate normal distribution Marginal and conditional distributions - characteristic function. Maximum likelihood estimation of the parameters of Multivariate Normal and their sampling distributions - Inference concerning the mean vector when covariance matrix is known.

Total, Partial, Multiple Correlation in the Multivariate Setup MLEs of Total, Partial and Multiple correlation coefficients. Sampling distributions of Total and Multiple Correlation in the null case. Hotelling T2 statistic, derivation and its distribution - Uses of T2 statistic - relation between T2 and D2 - Mahalanobis D2 statistic and its distribution.

Generalized Variance Wishart distribution (statement only) Properties of Wishart distribution - Test for covariance matrix - Test for equality of covariance matrices.

Classification Problems Classification into one of two populations (known and unknown dispersion matrix) - Classification in to one of several populations - Fishers Linear discriminant function.

Principal Components Properties, Extraction of Principal components and their variances Canonical correlation - Estimation of canonical correlation and variates. Factor analysis - Mathematical model- Estimation of Factor Loadings - Concept of factor rotation - Varimax criterion.

Texts

1. Anderson, T.W. (1984) An Introduction to Multivariate Statistical Analysis, John Wiley.
2. Johnson, R. A and. Wichern D.W (2007): Applied Multivariate Statistical Analysis, 6 /e, Prentice-Hall of India Private Ltd., New Delhi.

References

1. Alvin C. Rencher(2002): Methods of Multivariate Analysis, 2/e, Wiley Interscience.
2. Giri, N. Multivariate Statistical Inference, Academic Publishers.
3. Jolliffe I.T.(2002): Principal Component Analysis, 2/e, Springer.
4. Ksheersagar, A. M. Multivariate Analysis. Marcell Dekkar.
5. Morrison, D.F. Multivariate Analysis.

6. Rao, C.R(1998): Linear Statistical Inference and its Applications, Wiley Eastern Ltd.

7. Seber, G.A.F. (1977) Multivariate Observations, Wiley.

8. Srivastava M.S. and Khatri C.G.(1979):Introduction to Multivariate Analysis, Elsevier.

2.1.16 MSTAT-CC-16 Design and Analysis of Experiments

Prerequisites *Inference - I, Inference - II, Linear Inference*

Designed Experiments and Block Designs One way classification - ANOVA; Two way classification with equal number of observations per cell (with and without interaction), BIBD intra block analysis, incidence matrix, connectedness balanced, orthogonality for two way classification with unequal number of observations per cell, random effect models for one factor, estimation of variance components.

2^k **Full Factorial Experiments** Diagrammatic presentation of main effects and first and second order interactions, model, analysis of single as well as more than one replicates using ANOVA, total confounding of 2^k design in 2p blocks p 2, partial confounding in 2p blocks; p = 2, 3, fractional factorial experiments, resolution of a design (III, IV & V), aberration of a design.

3^2 **designs** contrasts for linear and quadratic effects , statistical analysis of 3^2 design. Response surface methodology (RSM): linear and quadratic model, stationary point, canonical analysis, central composite designs (CCD), ridge 24 systems, multiple responses, concept of rotatable designs.

Taguchi Methods Concept of loss function, S/N ratio, orthogonal arrays, triangular tables, linear graphs, inner and outer arrays; nested and split plot designs.

Texts

1. Das, M.N. and Giri, N. (1979). Design and Analysis of Experiments, Wiley Eastern.

2. Dey, A. Theory of Block Designs.

3. Dean, A. and Voss, D. (1999). Design and Analysis of Experiments, Springer.

4. Ogawa J. (1974). Statistical Theory of the Analysis of Experimental Design, Marcel Dekker.

5. Joshi, D. D. Linear Estimation and Design of Experiments.

References

1. Atkinson, A. C. and Donev, A. N. (1992). Optimal Experimental Designs. Oxford University Press.

2. George E. P. Box, Draper N.R. (1987). Empirical Model-Building and Response Surfaces, Wiley.

3. Hicks, C.R., Kenneth V. and Turner, Jr. (1999). Fundamental Concepts in the Design of Experiments, Oxford University Press.

4. John P.W.M. (1971). Linear Models, Wiley.

5. Kshirsagar A.M. (1983). Linear Models, Marcel Dekker .

6. John, P. W. M. (1971). Statistical Design and Analysis of Experiments. MacMillan.

7. Montgomery, C.D. (1976). Design and Analysis of Experiments, Wiley, New York.

8. Pukelsheim, F. (1993). Optimal Design of Experiments. Wiley.

9. Shah, K. R. and Sinha, B. K. (1989). Theory of Optimal Designs. Springer-Verlag.

2.1.17 MSTAT-CC-17 Time Series Analysis

Time Series as Stochastic Process Time series as a discrete parameter stochastic process, Auto - Covariance, Auto-correlation and their properties. Exploratory time series analysis, Test for trend and seasonality, Exponential and moving average smoothing, Holt Winter smoothing and Forecasting.

Time Series Models Wold representation of linear stationary processes, Detailed study of the linear time series models: Autoregressive, Moving Average, Autoregressive Moving Average and Autoregressive Integrated Moving Average models; Model identification.

Estimation in Time Series Models Estimation of ARMA models: Yule-Walker estimation for AR Processes, Maximum likelihood and least squares estimation for ARMA Processes, Discussion (without proof) of estimation of mean, Auto-covariance and auto-correlation function under large samples theory, Residual analysis and diagnostic checking. Forecasting using ARIMA models.

Analysis of Seasonal Models parsimonious models for seasonal time series, General multiplicative seasonal models, forecasting, identification, estimation and diagnosis methods for seasonal time series. Spectral analysis of weakly stationary process. Herglotzic Theorem. Periodogram and correlogram analysis.

Text

1. Brockwell, P.J and Davis R.A. (1987). Time Series: Theory and Methods, Springer-Verlag.
2. Chatfield, C. (2004). The Analysis of Time Series - An Introduction, Sixth edition, Chapman and Hall.
3. Fuller, W. A. Introduction to statistical time series.
4. Montgomery, D.C., C. L. Jennings, Murat, Kulahci. Introduction to Time Series Analysis and Forecasting, Wiley Interscience.

References

1. Box, G.E.P and Jenkins G.M. (1970), Time Series Analysis, Forecasting and Control, Holden-Day.
2. Kendall, M.G. (1978). Time Series, Charler Graffin.

2.1.18 MSTAT-CC-18 Statistical Computing and Data Analysis - IV

Multivariate Statistical Analysis

1. Test for Mean vector when dispersion matrix in known (Single and Two sample).
2. Hotelling T^2 test (One and two sample).
3. Test for covariance matrices.
4. Discriminant Analysis.
5. One way MANOVA.
6. Principal Component Analysis.
7. Canonical Correlation Analysis.
8. Factor Analysis

Design and Analysis of Experiments

1. Design and analysis of a BIBD.
2. Design and analysis of a 2^k factorial experiment.
3. Design analysis of 3^2 factorial experiment.
4. Response surface designs.

Time Series Analysis

1. Exploratory time Series analysis.
2. Auto regressive integrated moving average (ARIMA) models
3. Seasonal ARIMA (SARIMA) models.
4. Forecasting with time series models.

Texts

1. Das, M.N. and Giri, N. (1979). Design and Analysis of Experiments, Wiley Eastern.

2. Johnson, R. A and. Wichern D.W (2007): Applied Multivariate Statistical Analysis, 6 /e, Prentice-Hall of India Private Ltd., New Delhi.

3. Kendall, M.G. (1978). Time Series, Charler Graffin.

References

1. Box, G.E.P and Jenkins G.M. (1970), Time Series Analysis, Forecasting and Control, Holden-Day.

2. Montgomery, C.D. (1976). Design and Analysis of Experiments, Wiley, New York.

2.1.19 MSTAT-CC-19 Masters Dissertation

The student pursuing the program shall work for a dissertation under the supervision of a faculty member of the Department of Statistics, IGNTU. The work may be started after the Semester -II and be submitted before the culmination of Semester - IV.

2.1.20 MSTAT-CC-20 Summer Internship

This is a non-credit and optional training to be taken up by the student under the supervision of experts / agencies that are involved in research / training in areas related to Statistics.

2.1.21 MSTAT-CC-00 Bridge Course

Note This is a non-credit course. The course instructor may recommended this course for students who have not studied statistics courses at the under graduate level.

Data and Random Samples Uncertainty, randomness and variation; Concepts of experiments: Deterministic and Probabilistic (concept only); Data generated in controlled conditions and that generated freely in nature; Data generated by random mechanisms (for ex sampling); Prospective studies and Retrospective Studies; Data Types (concepts only with real life examples):Linear Data: Cross sectional, Time series, Longitudinal and Panel Data. Circular Data. Merits of a sample surveys over census surveys. Various useful sampling designs and how these are practically applied in field works. Examples on Sampling Designs like Simple Random Sampling, Proportional Allocation in Stratified Random Sampling, Systematic Sampling and Probability Proportional to Size Sampling [When to use what], weighting the observation and using the weights in the analysis of sample based data.

Descriptive Statistics Presentation of Data and Graphical Tools: Univariate frequency distribution, Tabulation of data and Graphical representation of linear and circular data. Summarizing a frequency distribution in terms of summary measures for Central tendency, Variance; Utility of third and fourth central moments; The concept of change in origin and scale; Standardization of a variable; The concept of association in simultaneous study of two or more variables; Various measures of association depending on the nature of variables involved.

Probability Models Concept of a probability distribution, Normal distribution as a model for data arising in social sciences, probability curve for a standardized normal variable: its area properties and uses.

Statistical Inference Inferences about Population Parameters based on Sample Data. Concept of STATISTIC and Examples on how sampling generates a SAMPLING DISTRIBUTION of a statistic and STANDARD ERROR. Estimates of population mean, variance, proportion and coefficient of correlation (with confidence regions). Tests of hypothesis regarding population mean, variance, proportion and coefficient of correlation.

References

1. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. I, 8th Edn. The World Press, Kolkata.

2. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. II, 8th Edn. The World Press, Kolkata.

2.2 Discipline Specific Elective Courses (DE)

2.2.1 MSTAT-DE-01 Statistical Methods in Demography

Demographic Data and Measures Sources of demographic Statistics, Basic demographic measures: Ratios, Proportions and percentages, Population Pyramids, Sex ratio Crude rates, Specific rates, Labour force participation rates, Density of population, Probability of dying.

Life tables Construction of a life table, Graphs of l_x, q_x , d_x , Functions L_x , T_x and E_x. Abridged life tables Mortality: Rates and Ratios, Infant mortality, Maternal mortality, Expected number of deaths, Direct and Indirect Standardization, Compound analysis, Morbidity.

Fertility Measures of Fertility, Reproductivity formulae, Rates of natural increase, Fertility Schedules, Differential fertility, Stable populations, Calculation of the age distribution of a stable population, Model Stable Populations.

Population Projections Component method, Mortality basis for projections, Fertility basis for projections, Migration basis for projections, Ageing of the population, Estimation of demographic measures from incomplete data.

Texts

1. Ram, F. and Pathak, K. B. (1998): Techniques of Demographic Analysis, 2^{nd} Ed, Himalaya Publishing House, Bombay.

2. Alho, J., and Spencer, B. (2005). Statistical Demography and Forecasting. Springer-Verlag, New York.

References

1. Bhende, A. A. and Kanitkar,T. (2003). Principles of Population Studies, Sixteenth Revised Edition, Himalaya Publishing House, Mumbai.

2. Keyfilz, N. and Caswell, H. (2005) Applied Mathematical Demography, Third edition, Springer.

3. Pollard, A.H. Yusuf, F. and Pollard, G.N. (1990). Demographic Techniques, Pergamon Press.

4. Retherford, R. D. and Choe, M. K. (1993). Statistical Models for Causal Analysis, John Wiley & Sons, Inc.

5. Siegel, J. S. and Swanson, D. A. (2004). The Methods and Materials of Demography, Second Edition, Elsevier Science, USA.

6. United Nations Manuals on Demography.

7. Weeks, J. R. (2005). Population: An Introduction to Concepts and Issues, Nineth Edition, Wadsworth Publishing Company, Belmont, California.

2.2.2 MSTAT-DE-02 Bayesian Inference

Subjective Interpretation of probability in terms of fair odds - Evaluation of Subjective probability of an event using a subjectively unbiased coin - Subjective prior distribution of a parameter - Bayes theorem and computation of the posterior distribution.

Natural Conjugate family of priors for a model - Hyper parameters of a prior from conjugate family - Conjugate families for exponential family models - models admitting sufficient statistics of fixed dimension - Enlarging the natural conjugate family by enlarging hyper parameter space - mixtures from conjugate family - choosing an appropriate member of conjugate prior family - Non informative, improper and invariant priors - Jeffreys invariant prior.

Bayesian point estimation Prediction problem from posterior distribution - Bayes estimators for absolute error loss, squared error loss and 0 -1 loss - Generalization to convex loss functions - Evaluation of the estimate in terms of the posterior risk - Bayesian interval estimation : Credible intervals - Highest posterior density regions - Interpretation of the confidence coefficient of an interval.

Bayesian Testing of Hypothesis Prior and Posterior odds - Bayes factor for various types of testing hypothesis problems - Lindleys paradox for testing a point hypothesis for normal mean.

Bayesian prediction problem Prediction for Exponential family of Distributions - Predictive Distributions and Reliability Estimation - Predictive Interval - Ideas on Bayesian Robustness, Monte-Carlo Integration and Markov chain Monte Carlo techniques (without proof).

Texts

1. Bansal A.K.(2007): Bayesian Parametric Inference, Narosa Publications.

2. Sinha S K (1998): Bayesian Estimation, New Age International(P) Ltd, New Delhi.

References

1. Berger, J.O.(1985): Statistical Decision Theory and Bayesian Analysis, 2/e, Springer Verlag.

2. Robert C.P. and Casella, G.(2004): Monte Carlo Statistical Methods, 2/e,Springer Verlag.

3. DeGroot, M.H.(2004): Optimal Statistical Decisions, Wiley-InterScience.

4. Gamerman, D. and Lobes H.F. (2000): Stochastic Simulation for Bayesian Inference, Taylor and Francis.

5. Box, G.P. and Tiao, G.C.(1973): Bayesian Inference in Statistical Analysis, Addison Wesley.

2.2.3 MSTAT-DE-03 Analysis of Directional Data

Descriptive Analysis of Directional Data Graphical representation of data, Frequency distribution, Measures of location, circular variance and concentration, Correction for mean grouping, Measures of skewness and kurtosis.

Probability Models Circular models, distribution theory, independence, convolution, moments, distributions of an arc, mixtures, lattice distributions, wrapped normal, Cauchy, Poisson distributions, Von Mises, Fisher distribution characteristic functions, Polar distributions, isotropic random walk on the circle.

Estimation Point estimation, Cramer Rao type bound, sufficiency, Methods of estimation, testing hypothesis from parametric models. Neyman-Pearson and likelihood ratio principles.

Non-parametric methods Tests for randomness, goodness of fit, Rayleighs test. Durand and Greenwoods test, Range test, Kupers test, Watsons test, Uniform score tests, Runs test, Rank sum test, Tests for dispersion.

Texts

1. Mardia, K.V. (1972). Statistics of Directional data, Academic Press.

References

1. Batschelet, E. (1981). Circular Statistics in Biology, Academic Press.

2. Watson, G.S (1983). Statistics on Spheres, Wiley.

2.2.4 MSTAT-DE-04 Spatial Data Analysis

Geostatistical data, Lattice data, Point patterns. Ripleys & Cressies approaches to analysis of spatial data. Basic stochastic processes.

Spatial sampling, Autoregression and Autocorrelation. Point patterns Distance methods, Nearest-neighbor methods. Variogram & Correlogram. Variogram model fitting.

Spatial Prediction and Kriging, Spatial models on Lattices for discrete and continuous data.

GIS GIS and its applications.

Texts

1. Cressie, N.A. (1993). Statistics for Spatial Data. Wiley.

References

1. Ripley, B.D. (1981). Spatial Statistics. Wiley

2.2.5 MSTAT-DE-05 Statistical Genetics

Basic Concepts Basic biological concepts in genetics; Mendels law, Hardy Weinberg equilibrium. Mating tables, estimation of allele frequency (dominant / co-dominant cases). Approach to equilibrium for X-linked gone, Natural selection, mutation, genetic drift, equilibrium when both natural selection and mutation are operative. Non-random mating, inbreeding, phenotypic assortative mating. Analysis of family data (a) Relative pair data, I, T; 0 matrices, identity by descent, (b) family data - estimation of segretation ratio under ascertainment bias, (c) Pedigree data - Elston - Stewart algorithm for calculation of likelihoods. Linkage, Estimation of recombination fraction, inheritance of quantitative traits. Models and estimation of parameters. Sequence similarity, homology and alignment. Algorithms for (a) pairwise sequence alignment, (b) multiple sequence alignment, construction of phylogenetic trees, UPGMA, Neighbour joining, maximum parsimony and maximum likelihood algorithms.

Modelling in Genetics Brief review of genetics, including the Hardy-Wienberg laws, their ramifications including mutation and fitness coefficient, inbreeding and changes of coefficient of inbreeding over generations, Markovian models: sibmating,Wright-Fisher, Moran, Kimura models, Wright-Fisher model with varying generation sizes, hidden Markov models.

Texts

1. Ewens, W. J. Mathematical Population Genetics.

2. Kingman, J. F. C. Mathematics of Genetic Diversity.

References

1. Durbin,R., Eddy, S. R., Krogh,A., and Mitchison, G. (1998). Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.

2. Li, C. C. (1976). First Course on Population Genetics. Boxwood Press, California.

3. Nagylaki, T. (1992). Introduction to Theoretical Population Genetics. Springer Verlag.

2.2.6 MSTAT-DE-06 Stochastic Processes

Stohastic Processes Definition and Types.

Markov Chains Definition, Examples and classification, Discrete renewal equation and basic limit theorem, Absorption probabilities, Criteria for recurrence; Limiting Distribution, Stationary Distribution and Random walk.

Continuous Time Markov Chains Poisson process, General pure birth process, birth and death process, finite state continuous time Markov chains.

Branching Processes Review of discrete time branching process, extinction probabilities and asymptotic behaviour brief excursion to continuous time branching process, two-type branching process, branching process with general lifetime variable (Bellman-Harris process).

Renewal Processes Renewal equation, renewal theorem, applications, generalizations and variations of renewal processes, applications of renewal theory, Brownian motion.

Texts

1. Medhi, J. (1994) Stochastic Processes, Second edition, Wiley Eastern.

2. Ross, S. (1996) Stochastic Processes, Second edition, John Wiley.

References

1. Bhat, U. N. (1972). Elements of Applied Stochastic Processes, Wiley.
2. Harris, T. E. The theory of Branching Processes.
3. Hoel, P.G.., Port, S.C. and Stone, C.J. (1972). Introduction to Stochastic Processes, Houghton Miffin & Co.
4. Karlin, S. and Taylor, H.M. (1975). A First Course in Stochastic Processes, second edition, Academic Press.
5. Kulkarni,V.G. (1995). Modeling and Analysis of Stochastic Systems, Chapman and Hall, London.

2.2.7 MSTAT-DE-07 Bioassay

Bioassay Direct and indirect assays, quantal and quantitative assays, parallel line and slope ratio assays, design of bioassays, dose allocation schemes.

Methods of estimation of parameters estimation of extreme quantiles; Quantal responses; estimation of points on the quantal response function, sequential procedures, estimation of safe doses, polychotomous quantal response; Bayesian approach to bioassay.

Ratio Estimators Asymptotic distributions; Fiellers theorem.

Estimating dose-response relationships Regression approaches, Logit and probit approaches when dose-response curve for standard preparation is unknown.

Texts

1. Finney, D. J. (1971). Statistical Method in Bioassay, Griffin.
2. Govindarajulu, Z.(2000). Statistical Techniques in Bioassay, S. Kargar.

References

1. Finney, D. J. (1971). Probit Analysis (3^{rd} Ed.), Griffin.
2. Weatherile, G. B.(1966). Sequential Methods in Statistics, Methuen.

2.2.8 MSTAT-DE-08 Clinical Trials

Introduction to Clinical Trials The need and ethics of clinical trials, outcome measures, protocols, sample size determination, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I-IV trials, multi-center trials.

Data Management Data definitions, case report forms, database design, data collection systems for good clinical practice.

Design of clinical trials Random allocation, symmetric designs, adaptive designs, group sequential designs, Zelen's designs, comparative and controlled trials, parallel vs. cross-over designs, cross-sectional vs. Longitudinal designs, review of factorial designs; objectives 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 of bioequivalence trials, selection and design of trials with surrogate endpoints.

Reporting and Analysis

Analysis of categorical outcomes from Phase I - III trials, analysis of survival data from clinical trials, analysis of surrogate endpoint data, meta-analysis of clinical trials;

Baye's theorem, diagnostic tests sensitivity and specificity, applications of Bayes theorem, Receiver Operator Characteristics (ROC) curves.

Texts

1. Fleiss, J. L. (1989). The Design and Analysis of Clinical Experiments. Wiley and Sons.

2. Friedman, L. M., Furburg,C. and Demets, D. L. (1998). Fundamentals of Clinical Trials, Springer Verlag.

References

1. Jennison, C. and Turnbull, B. W. (1999). Group Sequential Methods with Applications to Clinical Trials, CRC Press.

2. E. Marubeni and M. G. Valsecchi (1994). Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley and Sons.

3. Piantadosi, S. (1997). Clinical Trials: A Methodological Perspective, John Wiley & Sons, New York.

2.2.9 MSTAT-DE-09 Quantitative Epidemiology

Introduction to Modern Epidemiology Principles of epidemiologic investigation, surveillance and disease monitoring in populations, epidemiologic resources on the Web; Organizing and presenting epidemiologic data.

Measuring Population Health Incidence and prevalence rates, disease attack rates, odds ratios, absolute and relative risk; standardization of rates direct and indirect standardization; application of life tables;causation and causal inference.

Design and analysis of Epidemiologic Studies

Types of Studies Case-control studies, cohort studies, cross-over designs.

Estimation estimation of prevalence and incidence; regression models for the estimation of relative risk; meta-analysis; quantitative methods in screening; age at onset distributions; assessing spatial and temporal patterns, disease mapping;

Probabilistic and Deterministic Models in Public Health Epidemic Modelling: Simple and general epidemics - both deterministic as well as stochastic; threshold theorems; Greenwood, Reed-Frost models, Neyman-Scott models of spatial spread of epidemics. Deterministic modelling of infectious diseases.

Special Topics Environmental epidemiology, molecular and genetic epidemiology, computational epidemiology; big data analytics in public health.

Texts

1. Rothman, K. J. and Geenland, S. (ed.) (1998). Modern Epidemiology, Lippincott-Raven.

2. Selvin, S. (1996). Statistical Analysis of Epidemiologic Data, Oxford University Press.

References

1. Bailey, N. T. J. The Mathematical Theory of Infectious Diseases and its applications.

2. Jekel, J. F., Elmore, J. G., and Katz, D.L. (1996). Epidemiology, Biostatistics and Preventive Medicine. WB Saunders Co.

3. McNeil, D. (1996). Epidemiological Research Methods. Wiley and Sons.

2.2.10 MSTAT-DE-10 Advanced Statistical Computing

Bootstrap Methods Resampling paradigms, bias and standard errors, confidence intervals; bootstrapping in regression; Jackknife in sample surveys, cross-validation for tuning parameters.

Simulation Based Testing Simulating test statistics and power functions.

Solution of Likelihood Equations Method of scoring, Newton - Raphson and other iterative procedures.

Other Computational Techniques EM Algorithm, Genetic Algorithm; Simulation based estimation : MCMC, Gibbs Sampler.

Text

1. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R, Springer.

2. Press, W. H., Teukolsky, S. A., Vellering, W. T. and Flannery, B. P. (1993). Numerical Recipes in C, II edition, Cambridge University Press.

3. Ross, S. M. Simulation.

References

1. Chambers, J. M. and Hastie, T. J (Editors).(1997). Statistical Models in S. Chapman and Hall.

2. Venables, W. N. and Ripley, B. D. (2000). S Programming, Springer.

3. www.r-project.org

2.2.11 MSTAT-DE-11 Operations Research

Linear programming Convex sets and associated theorems, graphical method, definition of linear programming problem, properties of a solution to the linear programming problem, generating extreme-point solutions, simplex computational procedure, artificial variables technique - big M method, two phase method; Revised simplex method; Duality problems of linear programming: unsymmetric primal-dual problems, symmetric primal-dual problems, Degeneracy and anticycling procedures: perturbation techniques.

Transportation Problems General transportation problem, Finding initial basic feasible solution, test for optimality, degeneracy in transportation problem, unbalanced transportation problem, maximization transportation problem, Assignment problem: mathematical formulation of the problem, the assignment method (Hungarian method), Non-linear programming problem (NLPP): general non-linear programming problem,Constrained optimization with equality constraints - necessary conditions for a generalized NLPP, sufficient conditions for a general NLPP with one constraint, sufficient conditions for a general problem with m(¡n) constraints, Constrained optimization with inequality constraints - Kuhn-Tucker conditions for general NLPP with m(¡n) constraints.

Inventory models Deterministic inventory models - general inventory model, Static economic-order quantity (EOQ) models - classic EOQ model, EOQ with price breaks, multi-item EOQ with storage limitation, Probabilistic inventory models:- Continuous review models - probabilitized EOQ model, probabilistic EOQ model, Single-period models - No setup model (Newsvendor model), setup model (s-S policy).

Queuing Systems Elements of a queuing model, roles of exponential distribution, pure birth and death models (relationship between the exponential and Poisson distributions) - pure birth model, pure death model, generalized Poisson queuing model, Specialized Poisson queues - steady-state measures of performance, single server models, multiple server models.

Texts

1. Gass S.I. (1985). Linear Programming - methods and applications, Fifth edition, Mc Graw Hill, USA.

2. Ravindran A, Philips D.T and Soleberg J.J. (1997) Operation Research - Principles and Practice, John Wiley & Sons.

References

1. Beale. (1988). Introduction to Optimization, John Wiley, 1988.

2. Taha H.A. (2007). Operations Research - An introduction, Eighth edition, Prentice-Hall of India Ltd.

3. Joshi, M. C. and Moudgalya, K. (2004). Optimization: Theory and Practice, Narosa, New Delhi, 2004.

2.2.12 MSTAT-DE-12 Statistical Quality Assurance

Quality and Quality Assurance Methods of quality assurance, Introduction to TQM and ISO 9000 standards, statistical quality control: Acceptance sampling for attributes, Single sampling, Double sampling, Multiple and sequential sampling plans, Measuring the performance of these plans.

Control charts Basic ideas, designing of control charts for the number of non-conformities and fraction non-conformities, mean charts, Median charts, Extreme value charts, R-charts, and S-charts, ARL, Economic design of Shewarts control charts.

Acceptance Sampling by Variables Sampling plans for a single specification limit with known and unknown variance, Sampling plans with double specification limits, Comparison of sampling plans by variable and attributes, Continuous sampling plans I, II and III.

Process Capability Studies Statistical aspect of six sigma philosophy, Control charts with memory - CUSUM charts, EWMA-mean charts, OC and ARL for control charts, The Taguchi Method: The Taguchi philosophy of Quality, Loss functions, SN ratios, Performance measures, Experimental design in Taguchi Methods: Orthogonal arrays and linear graph, Estimation of effects, Parameter Design.

Texts

1. Montgomery, R.C. (1985). Introduction to Statistical Quality Control, Fourth edition, Wiley.

2. The ISO 9000 book, Second Edition, Rabbit, J T and Bergle, PA Quality resources, Chapter-I

References

1. Schilling, E.G. (1982) Acceptance Sampling in Quality Control, Marcel Dekker.

2. Mitra, A.(2000) Fundamentals of Quality Control and Improvement Pearson Education Asia.

2.2.13 MSTAT-DE-13 Econometrics

Multicollinearity Effects of multicollinearity and detection, Remedial methods including the ridge regression. Specification error analysis, inclusion of irrelevant variables and deletion of dominant variables, their effects on the efficiency of optimization procedure.

Hetroscedasticity Consequences and tests for it, estimation procedures under hetroscedastic disturbances.

Auto Correlated Disturbances Effects on estimation of parameters, Cochran Orcutt and Prais-Winston transformation, Durbin-Watson test. Errors-in-varables model, Inconsistency of least squares procedures, Consistent estimation of Parameters by instrumental variables.

SURE Model Seemingly unrelated regression equation model, estimation of parameters under SURE model and their properties, generalized least squares methods and their asymptotic properties.

Simultaneous Equation Model Problem of indentification, a necessary and sufficient condition for the identifiability of parameters in a structural equation, Ordinary Least squares, indirect least squares, two stage least squares and limited information maximum likelihood method, K-class estimators, Asymptotic properties of estimators.

Suggested Readings

1. Gujarathi, D. (1979): Basic Econometrics, McGraw Hill.
2. Intrulligator, M.D. (1980): Econometric modelsTechniques and Applications, Prentice Hall of India.
3. Johnston, J. (1984): Econometric methods. Third edition, McGraw Hill.
4. Koutsoyiannis, A. (1979): Theory of Econometrics, Macmillan Press.

2.2.14 MSTAT-DE-14 Categorical Data Analysis

Prerequisites *Inference - I and Inference - II*

Categorical Data Measures of association, contingency tables,

Estimation Estimation in complete and incomplete tables; missing data and E-M algorithm for contingency tables and goodness-of- fit tests.

Generalized Linear Models Poisson and Logistic (Binary and Multinomial) Regression; Log-linear models;

Exact Tests Exact tests in Contingency tables.

Texts

1. Agresti, A. Introduction to Categorical Data Analysis. Wiley.
2. McCullagh, P and Nelder, J. A. Generalized Linear Models. Chapman and Hall.

References

1. Little, R.J. and Rubin, D. B. (1987). Statistical Analysis with Missing Data. Wiley.

2.2.15 MSTAT-DE-15 Fuzzy Sets and Systems

Fuzzy Sets and Uncertainty Uncertainty and information, fuzzy sets and membership functions, chance versus fuzziness, properties of fuzzy sets, fuzzy set operations.

Fuzzy Relations Cardinality, operations, properties, fuzzy Cartesian product and composition, fuzzy tolerance and equivalence relations, forms of composition operation.

Fuzzification and Defuzzification Various forms of membership functions, fuzzification, defuzzification to crisp sets and scalars.

Fuzzy Logic and Fuzzy Systems Classic and fuzzy logic, approximate reasoning, Natural language, linguistic hedges, fuzzy rule based systems, graphical technique of inference.

Development of membership functions Membership value assignments: intuition, inference, rank ordering, neural networks, genetic algorithms, inductive reasoning.

Fuzzy Arithmetic and Extension Principle Functions of fuzzy sets, extension principle, fuzzy mapping, interval analysis, vert ex method and DSW algorithm.

Fuzzy Optimization One dimensional fuzzy optimization, fuzzy concept variables and casual relations, fuzzy cognitive maps, agent based models.

Fuzzy Control Systems Fuzzy control system design problem, fuzzy engineering process control, fuzzy statistical process control, industrial applications.

Texts

1. T.J. Ross, Fuzzy Logic with Engineering Applications, 3rd Ed., Wiley India Pvt. Ltd., 2011.

2. H.J. Zimmerman, Fuzzy Set Theory and its Application, 3rd Ed., Springer India Pvt. Ltd., 2006.

3. G. Klir and B. Yuan, Fuzzy Set and Fuzzy Logic: Theory and Applications, Prentice Hall of India Pvt. Ltd., 2002.

4. G. Klir and T. Folger, Fuzzy Sets, Uncertainty and Information, Prentice Hall of India Pvt. Ltd., 2002.

2.3 Generic Electives (GE)

2.3.1 MSTAT-GE-01 Epidemiology and Vital Statistics

2-1-0-3

Prerequisite High School Algebra

Population Composition and Distribution Population density, percentage distribution, population potential, Lorenz curve, Gini concentration ratio; Age distribution: per cent distribution, per cent change in distribution, index of relative difference, index of dissimilarity; Population pyramid.

Epidemiology Definition and scope, health and disease; Population at risk, prevalence rate, incidence rate, risk, case-fatality; Morbidity; Disability; Comparing disease occurrence: risk difference, attributable fraction, population attributable risk, relative comparison.

Vital Statistics Rates of vital events: measurement of mortality, measurement of fertility, measurement of population growth.

Life Tables Life Tables: description and construction; Abridged life tables.

References

1. Bonita, R., Beaglehole, R., and Kjellström, T. (1993). Basic epidemiology. World Health Organization, Geneva.

2. Goon A.M., Gupta M.K. and Dasgupta B. (2002). Fundamentals of Statistics, Vol. II, 8th Edn. The World Press, Kolkata.

3. Ram, F. and Pathak, K. B. (1998). Techniques of Demographic Analysis, 2^{nd} Ed, Himalaya Publishing House, Bombay.

2.3.2 MSTAT-GE-02 Probability Theory and Models

Prerequisite High School Algebra

Definition of statistics Scope and limitations of statistics, Primary and Secondary data and its sources.

Measures of central tendency Mean, Median and Mode - Measures of dispersion Range, Variance and Standard deviation Coefficient of variation.

Elementary Probability Theory Sample spaces, events, probability; Axioms of Probability, conditional probability and independence, Bayes Theorem.

Random Variables Discrete and continuous random variables; distribution function; Random vectors, joint, conditional and marginal distributions.

Expectation Moments of a probability distribution, moment generating functions and characteristic functions, conditional expectation.

References

1. P. Billingsley: Probability and Measure (2 nd edition), John Wiley & Sons.

2. S M Ross: Introduction to Probability Models, Academic Press. H M Taylor and S Karlin: An Introduction to Stochastic Modelling, Academic Press.

3. P.G. Hoel, S.C. Port and C.J. Stone: Introduction to Probability, Universal Book Stall, New Delhi.

4. J.S. Rosenthal: A First Look at Rigourous Probability Theory, World Scientific.

5. M. Woodroofe: Probability with Applications, McGraw-Hill.

2.3.3 MSTAT-GE-03 Designing Scientific Studies

Prerequisite High School Algebra and Elementary Probability

Introduction to the Design of Experiments Basic terminology and principles; Construction and analyses of Completely Randomized Design, Randomized Block Design, Latin Square Design.

Design of Factorial Experiments Construction and analyses of 2^2 and 2^3 factorial design.

Design of Sample Surveys Sampling designs and sampling schemes; basic finite population sampling techniques [SRS WR/WOR, stratified, systematic] and related results on estimation of population mean/total. Allocation problem in stratified sampling.

Designing Epidemiological Studies Descriptive studies: ecological studies, cross-sectional studies, case-control studies, cohort studies; Experimental studies: randomized controlled trials, community trials, field trials; Potential errors in epidemiological studies: random error, systematic error, bias and confounding; Validity and reliability.

References

1. Bonita, R., Beaglehole, R., and Kjellström, T. (1993). Basic epidemiology. World Health Organization, Geneva.

2. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. II, 8th Edn. The World Press, Kolkata.

2.3.4 MSTAT-GE-04 Scientific Computing

Prerequisite High School Algebra

Programming in Matlab/Scilab Executing a function, Global variables, Loops, branches and control flow, Multidimensional matrices.

Applications in Linear Algebra Solving linear system of equations, Gaussian elimination, matrix factorization.

Matrices and Vectors Matrix and array operations, character strings, Finding the determinant, Eigen values and Eigen vectors of a matrix.

Other Applications Curve fitting and interpolation; Numerical integration; Finding the roots of polynomial equations by various numerical methods.

References

1. Rudra Pratap: Getting started with Matlab, Oxford University Press.
2. Brian R Hunt, Ronald L Lipsman, Jonathan M Rosenberg: A Guide to Matlab for beginners and experienced users, Cambridge University Press.
3. S R Otto and J P Denier: An Introduction to programming and Numerical Methods in Matlab, Springer Inc.
4. Won Young Yang, Wenwu Cao, Tae-Sang Chung, and John Morris: Applied Numerical Methods using Matlab, John Wiley & Sons Inc.

2.3.5 MSTAT-GE-05 Descriptive Data Analysis with R

Unit I Scientific Studies and variables as means to study a phenomenon, Classification of variables based on scale [nominal, ordinal, interval and ratio], based on discreteness and continuity

[With examples form Life Sciences, Environmental Sciences and Social Sciences]

Unit II Uncertainty, randomness and variation; Concepts of experiments: Deterministic and Probabilistic (concept only); Data generated in controlled conditions and that generated freely in nature; Data generated by random mechanisms (for ex sampling); Prospective studies and Retrospective Studies;

Data Types (concepts only with real life examples): Linear Data: Cross sectional, Time series, Longitudinal and Panel Data. Circular Data: (As it arises in Environmental Sciences).

[The student should be able to distinguish between various data types and variable types]

Unit III Presentation of Data and Graphical Tools: Univariate frequency distribution, Tabulation of data and Graphical representation of linear and circular data.

[The student should know appropriate use and interpretation of Bar Charts, Pie Charts, Histograms and Scatter Plots for Bivariate Data]

Unit IV Summarizing a frequency distribution in terms of summary measures for Central tendency, Variance; Utility of third and fourth central moments; The concept of change in origin and scale; Standardization of a variable.

[The student should be able to identify the best measure of central tendency and thus variance when dealing with a real life data set (Examples of Human Development Index, Consumer Price Index etc can be cited). He should also be able to appreciate the fact that these summary measures vary by the nature of the variable and the investigator has to be cautious when the summary measures are scale dependent]

Unit V Association: The concept of association in simultaneous study of two or more variables; Various measures of association depending on the nature of variables involved.

[The teacher is expected to emphasize that all the measures based on samples need further testing to be taken up in advance courses]

References

1. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. I, 8th Edn. The World Press, Kolkata.

2. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. II, 8th Edn. The World Press, Kolkata.

2.3.6 MSTAT-GE-06 Inferential Data Analysis with R

Unit I Probability Distributions:

Concept of a probability distribution [Draw similarities with the frequency distribution studied in ST: 551 Techniques of Data Analysis I], Normal distribution as a model for data arising in social sciences [the student may be cautioned by citing examples from daily life that it is not the best model always], probability curve for a standardized normal variable: its area properties and uses.

Unit II Sample Surveys:

Why do we conduct a survey, merits of a sample surveys over census surveys. Various useful sampling designs and how these are practically applied in field works. [The students may be asked to design a sample survey for their dissertation work]

Examples on Sampling Designs like Simple Random Sampling, Proportional Allocation in Stratified Random Sampling, Systematic Sampling and Probability Proportional to Size Sampling [When to use what], weighting the observation and using the weights in the analysis of sample based data.

[Students should be asked to practice entering data into spread sheets of Excel/SPSS]

Unit III Inferences about Population Parameters based on Sample Data

Concept of STATISTIC and Examples on how sampling generates a SAMPLING DISTRIBUTION of a statistic and STANDARD ERROR.

Estimates of population mean, variance, proportion and coefficient of correlation. [With confidence regions]

Tests of hypothesis regarding population mean, variance, proportion and coefficient of correlation. [The concepts of statistical hypotheses: null and alternative should be incorporated simultaneously using examples; when to use one-tailed test/ two-tailed test should be made clear to the students]

Unit IV Regression Analysis

Distinction between independent (explanatory) and dependent variables [the teacher may limit to keeping both the independent variables and dependent variables quantitative]; the linear regression model, parameters of the model, their estimates, interpretation of the estimated parameters, testing the significance of the parameters and fit of the model itself;

[One or two thorough examples are sufficient to make students understand the application of regression analysis]

Unit V Useful non parametric tests

References

1. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. I, 8th Edn. The World Press, Kolkata.

2. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. II, 8th Edn. The World Press, Kolkata.

2.3.7 MSTAT-GE-07 Big Data and Analytics

Introduction Big data overview, data pre-processing, concepts of supervised and unsupervised learning; An overview of descriptive statistics and regression; Ridge Regression, Lasso Regression, K Nearest Neighbours Regression.

Classification Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Regression and Classification Trees, Support Vector Machines.

Ensemble methods and Clustering Bagging, random forests, boosting. K-means, K-medoids, Hierarchical clustering, X-means.

Evaluation and Validation Cross-validation, assessing the statistical significance of data mining results.

Texts

1. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R, Springer.

2. Prakasa Rao, B.L.S. (2015). Brief Notes on BIG DATA: A Cursory Look. Lecture Notes No.: LN2015-01, C R RAO AIMSCS Lecture Notes Series.

References

1. Montogomery, D. C., Peck, E. A. and Vining, G. G. Introduction to Linear Regression Analysis.