Discipline Specific Elective
Category V

DISCIPLINE SPECIFIC ELECTIVE COURSE –4A: BIOSTATISTICS

CREDIT DISTRIBUTION, ELIGIBILITY, AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Biostatistics	**4**	**3**	0	**1**	Class XII pass with Mathematics .	knowledge of Statistical Inference and stochastic processes

Learning objectives:

- Parametric Models for Survival data.
- Different types of censoring and its application in public health.
- Estimation of death probabilities by using the theory of competing risks.
- Non-parametric methods for incomplete survival data.
- Computation of the probability of gametes in different generations under random mating.

Learning Outcomes:

After completing this course, students will develop a clear understanding of:

- The fundamental concepts of survival functions and their interrelationship.
- Survival models and their applications.
- Handling censored data and estimating mean survival time of the patients.
- Actuarial and Kaplan-Meier methods.
- Competing Risk Theory.
- Basic concept of Statistical genetics.

SYLLABUS OF DSE-4A

Theory

UNIT I **(11 Hours)**

Survival Analysis

Survival Analysis: Functions of survival times, survival distributions and their applications exponential, gamma, Weibull, Rayleigh, lognormal, death density function for a distribution having bath-tub shaped hazard function.

UNIT II **(13 Hours)**

Censoring Schemes

Censoring Schemes: Type I, Type II and progressive or random censoring with biological examples. Estimation of mean survival time and variance of the estimator for Type I and Type II censored data with numerical examples. Non-parametric methods: Actuarial and Kaplan-Meier methods for estimating survival function and variance of the Estimator.

UNIT III **(10 Hours)**

Competing Risk Theory:

Indices for measurement of the probability of death under competing risks and their inter-relations. Estimation of probabilities of death using maximum likelihood method and modified chi squire method.

UNIT IV **(11 Hours)**

Statistical Genetics:

Statistical Genetics: Introduction, concepts-Genotype, Phenotype, Dominance, Recessiveness, Linkage and Recombination, Coupling, and Repulsion. Mendelian laws of Heredity, Random mating, Gametic array, Genotypic array, Relation between genotypic array and gametic array under random mating. Distribution of genotypes under random mating. Hardy-Weinberg law. Concept of gene frequencies.

PRACTICAL/ LAB. WORK (30 HOURS)

List of Practical:

1. Estimation of survival function, death density function and hazard function.
2. Estimation of mean survival time using various parametric survival models.
3. To Identify and analyse type-I censored data.
4. To Identify and analyse type-II censored data.
5. To Identify and analyse progressively type I censored data.
6. Estimation of mean survival time and variance of the estimator for type I censored data.
7. Estimation of mean survival time and variance of the estimator for type II censored data.
8. Estimation of mean survival time and variance of the estimator for progressively type I censored data.
9. To estimate the survival function and variance of the estimator using Actuarial methods.
10. To estimate the survival function and variance of the estimator using Kaplan-Meier method.
11. To estimate Crude probability of death.
12. To estimate Net-type I probability of death.
13. To estimate Net-type II probability of death.
14. To estimate partially crude probability of death.
15. To estimate gene frequencies.

Practical work to be conducted using electronic spreadsheet / EXCEL/ Statistical Software Package/ SPSS/ calculators.

ESSENTIAL READINGS:

- Biswas, S. (2007). Applied Stochastic Processes: A Biostatistical and Population Oriented Approach, Reprinted 2nd Ed., New Central Book Agency.
- Lee, E.T. and Wang, J.W. (2003). Statistical Methods for Survival data Analysis, 3rd Ed., John Wiley & Sons.
- Indrayan, A. (2008). Medical Biostatistics, 2nd Ed., Chapman and Hall/CRC.

SUGGESTIVE READINGS:

- Narayan P. (1999). Statistical Genetics, New Age International Pvt. Ltd.
- Miller, R. G. (2011). Survival Analysis. John Wiley & Sons.
- Elandt-Johnson R.C (1971).Probability models and Statistical Methods in Genetics, John Wiley & Sons.

Note: Examination scheme and mode shall be as prescribed by the Examination Branch University of Delhi, from time to time.

DISCIPLINE SPECIFIC ELECTIVE COURSE – 4b: ORDER STATISTICS AND ITS APPLICATIONS

CREDIT DISTRIBUTION, ELIGIBILITY, AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Order Statistics and its Applications	4	3	0	1	Class XII pass with Mathematics	knowledge of statistical distributions and stochastic processes

Learning Objectives

The learning objective of this course is to make the students aware of the properties and applications of order statistics.

Learning Outcomes:

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

- Find joint, marginal distributions and conditional distributions of order statistics in the continuous and discrete case.
- Find the distribution of sample range and other systematic statistics in case of sampling from an arbitrary continuous population and from some specific continuous distributions such as uniform and exponential.
- Understand the Markov Chain property of order statistics in the continuous case.
- Learn how to obtain distribution-free confidence intervals for population quantile for population distributions based on order statistics.
- Understand the distribution-free bounds for moments of order statistics and of the range.
- Derive the recurrence relations and identities for moments of order statistics drawn from an arbitrary population (discrete or continuous), as well as from some specific distributions.
- Understand the concept of L-moments and L-moments estimation of parameters.
- Derive the Linear estimation of location and scale parameters based on the moments of order statistics.

SYLLABUS OF DSE-4b

Theory

UNIT I **(15 hours)**
Introduction to Order Statistics

Definition and applications of order statistics. Basic distribution theory. Joint and marginal distributions of order statistics in the continuous case. Distribution of the median, range and other systematic statistics. Order statistics for a discrete parent. Examples based on discrete and continuous distributions.

UNIT II **(10 hours)**

Conditional distribution of order statistics

Conditional distribution of order statistics. Order statistics as a Markov Chain. Distribution-free confidence intervals for population quantiles. Distribution-free bounds for moments of order statistics and of the range.

UNIT III **(10 hours)**

Moments of order statistics

Moments of order statistics. Recurrence relations and identities for moments of order statistics from an arbitrary distribution. Recurrence relations for moments of order statistics from some specific distributions.

UNIT IV **(10 hours)**

Order statistics in statistical inference

Order statistics in statistical inference. L-moments and L-moments estimation. Linear estimation based on order statistics. Examples based on some specific continuous distributions.

PRACTICAL/LAB WORK – (30 hours)

List of Practical:

1. Problem-solving using joint, marginal, and conditional distributions of order statistics for some specific continuous distributions.
2. Distribution-free confidence intervals for population quantiles for various distributions.
3. Calculating Means, variances, and covariances by using exact expressions for the moment of order statistics for some specific continuous distribution.
4. Calculating Means, variances, and covariances by using recurrence relations for some specific continuous distributions.
5. Calculation of L-moments for some specific continuous distributions.
6. L-moments estimation of parameters for some specific continuous distributions.
7. Calculation of linear unbiased estimation for location and scale parameters for some specific continuous distributions.

Practical work to be conducted using electronic spreadsheet / EXCEL/ Statistical Software Package/ SPSS/ calculators.

ESSENTIAL READINGS

- David, H. A. and Nagaraja, H. N. (2003). Order Statistics, 3rd ed., John Wiley & Sons.

SUGGESTIVE READINGS:

- Arnold, B. C., Balakrishnan, N. and Nagaraja H. N. (2008). A First Course in Order Statistics, SIAM Publishers.
- Arnold, B.C. and Balakrishnan, N. (1989). Relations, Bounds and Approximations for Order Statistics, Vol. 53, Springer-Verlag.
- Ahsanullah, M., Nevzorav, V.B. and Shakil, M. (2013). An Introduction to Order Statistics, Atlantis Studies in Probability and Statistics, Vol. III. Atlantis Press.
- Gibbons, J.D. and Chakraborti, S. (1992). Nonparametric Statistical Inference, 3rd ed., Marcel Dekker.

- Shahbaz, M. Q., Ahsanullah, M., Shahbaz, S. H. and Al-Zahrani, B. M. (2016). Ordered Random variables: Theory and Applications. Springer.

Note: Examination scheme and mode shall be as prescribed by the Examination Branch University of Delhi, from time to time.

DISCIPLINE SPECIFIC ELECTIVE COURSE – 4c: STATISTICAL COMPUTING AND BASIC DATA MINING

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Statistical Computing and Basic Data Mining	4	3	0	1	Class XII pass with Mathematics	Knowledge of MATLAB / OCTAVE / R / Python / C

Learning Objectives

learning objectives include:

- Understand the theoretical foundations and practical aspects of statistical computing and data mining.
- Develop skills in the use of statistical computing and data mining software to solve problems and analyze data. The programming implementations will be completed using MATLAB/OCTAVE/R/Python/C.

Learning Outcomes:

After completion of this course, students will develop a clear understanding of:

- Apply knowledge of statistical computing and data mining techniques to solve problems and analyze data.
- Communicate effectively about statistical computing and data mining concepts and techniques both orally and in writing.
- Develop ability for programming implementation using MATLAB/OCTAVE/R/Python/C.

SYLLABUS OF DSE-4C

Theory

UNIT I **(15 hours)**

Simulation techniques
Random number generation: Review; Simulating multivariate distributions; Simulating stochastic processes. Variance reduction methods.

UNIT II **(12 hours)**
Markov Chain Monte Carlo methods
Markov Chain Monte Carlo methods: The Metropolis–Hastings Algorithm; Gibbs sampling.

UNIT III **(18 hours)**
Data Mining and its applications
Introduction to Data Mining and its Applications. Data Pre-processing Techniques: Data Cleaning, Data Integration, Data Transformation, and Data Reduction. Exploratory Data Analysis. Classification Techniques: Decision Trees, Naive Bayes, k-Nearest Neighbors (k-NN). Clustering Techniques: K-Means, Hierarchical Clustering. Association rule mining. Evaluation of Data Mining Models.

PRACTICAL/LAB WORK – (30 hours)
List of Practical:

1. Practical based on random number generation: univariate and multivariate distributions.
2. Practical on simulating stochastic processes; variance reduction.
3. Simple practical problems on MCMC.
4. Practical based on Data pre-processing, transformation, reduction.
5. Practical based on classification and clustering.

Practical work to be conducted using electronic spreadsheet / EXCEL/ Statistical Software Package/ SPSS/ calculators.

ESSENTIAL READINGS

- Rubinstein, R.Y. (2017). Simulation and the Monte Carlo Methods, Wiley.
- Voss, J. (2014). An introduction to statistical computing: a simulation-based approach,Wiley series in computational statistics.
- Tan, P. N., Steinbach, M., & Kumar, V. (2016). Introduction to data mining. Pearson Education India.
- Han, J., Kamber, M., & Pei, J. (2012). Data mining concepts and techniques third edition. University of Illinois at Urbana-Champaign Micheline Kamber Jian Pei Simon Fraser University.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). Data Mining: Practical machine learning tools and techniques, Elsevier Inc.

SUGGESTIVE READINGS:

- Vetterling, William T., Saul A. Teukolsky, William H. Press, and Brian P. Flannery. Numerical recipes in C: the art of scientific computing. Cambridge university press, 1999.
- Christian, P. R., & George, C. (1999). Monte Carlo statistical methods. Springer Texts in Statistics.
- Hancock, M. F. (2012). Practical data mining. CRC Press.
- Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). Data mining for business analytics: concepts, techniques, and applications in R. John Wiley & Sons.
- Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2019). Data mining for business analytics: concepts, techniques and applications in Python. John Wiley & Sons.
- Hastie, T., Tibshirani, R., Friedman, J. (2008). The Elements of Statistical Learning: Data Mining, Inference and Prediction, 2nd ed., Springer.
- Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. United States: MIT Press.

Note: Examination scheme and mode shall be as prescribed by the Examination Branch University of Delhi, from time to time.

DISCIPLINE SPECIFIC ELECTIVE COURSE – 4d : RESEARCH METHODOLOGY

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Research Methodology	4	3	0	1	Class XII pass with Mathematics	Nil

Learning Objectives:

The learning objectives include

- To provide scientific approaches to develop the domain of human knowledge through empirical studies.
- To enable the student researchers to understand basic concepts and aspects related to research, data collection, analyses, interpretation and report writing.

Learning Outcomes:

After completion of this course, students will develop a clear understanding of

- Research Methods.

- Research Problems.
- Research Designs.
- Comparative study of different methods of data collection.
- Guidelines for construction of questionnaires.
- Processing and Analysis of data.
- Interpretation and Report writing.

SYLLABUS OF DSE – 4D

Theory

UNIT I **(09 hours)**

Introduction to Research:

Importance and need for research ethics, Objectives of research, Types of research, Research approaches, Review of literature, Mode of literature survey: Books and Monographs, Journals, Conference proceedings, Abstracting and Indexing Journals, E-Journals/Books, Formulation of a research problem, Identifying variables, Constructing hypothesis, Conceptualization of a research design.

UNIT II **(09 hours)**

Methods & Techniques of Data Collection:

Survey methodology and Data collection, Source of data collection-Use of secondary data, Methods of collecting primary data, Develop a questionnaire, Questions and answers in surveys, Non-response, Errors in surveys, Sample size, sampling frames and coverage error.

UNIT III **(15 hours)**

Data Processing & Analysis:

Data processing, Exploratory data analysis, Review of various techniques (Parametric and Nonparametric tests, Correlation and Regression analysis, ANOVA, Multivariate Techniques) for data analysis covered in core statistics papers, Techniques of interpretation, Precaution in interpretation.

Report writing:

Discussions, Conclusions, Referencing and various formats for reference writing, Bibliography, Thesis writing, Formats of publications in research journals including subject classification, Impact factor, Citation index.

UNIT IV **(12 hours)**

Computer Application:

Data Communication and networks, Website, Webpage, Search Engines, Scientific search engines. Scientific Word Processing with LaTeX and MS-Word, MS Equation editor, Slides making-Power Point Features, Slide preparation, SPSS, Statistical Programming with R, Simulation.

PRACTICAL/LAB WORK – (30 hours)

PROJECT WORK (using a spreadsheet, Scientific Word Processing with LaTeX and MS-Word, MS Equation editor, Slides making-Power Point Features, Slide preparation, SPSS, Statistical Programming with R, Simulation.)

ESSENTIAL READINGS

- Kothari, C.R., Garg, Gaurav (2015): Research Methodology: Methods and Techniques, 3rd Edition (Reprint), New Age International Publishers.
- Kumar, R. (2011): Research Methodology: A Step-by-Step Guide for Beginners, SAGE publications.

- Anderson, J., Durston, B.H., Pooole, M. (1970): Thesis and Assignment Writing, Wiley Eastern. Ltd., New Delhi.
- Braun, J., Duncan, W. and Murdock, J. (2008): A First Course in Statistical Programming with R, Cambridge University Press, London.
- Lamport, L. (1999): LATEX: A Document Preparation System, Addison, Wesley, 2nd Edition, New York.
- Cunningham, B.J. (2012): Using SPSS: An Interactive Hands-On Approach, SAGE South Asia Edition.
- Voss, J. (2014): An Introduction to Statistical Computing: A Simulation-based Approach, Wiley series in computational statistics.

SUGGESTIVE READINGS

- Pannerselvan, R. (2006): Research Methodology, Prentice-Hall of India Pvt., New Delhi.
- Landau, Sabine and Everitt, Brian S. (2004): A Handbook of Statistical Analyses using SPSS, Chapman & Hall/CRC.
- Dalgaard, P. (2008): Introductory Statistics with R, Springer Science, New York.
- Gardener, M. (2012): Beginning R: The Statistical Programming Language, Wiley Publications.
- Robert, C.P. and Casella, G. (2004): Monte Carlo Statistical Methods, Springer Science, New York.
- Rubinstein, R.Y. (1981): Simulation and the Monte Carlo Methods, Wiley.
- Venkataraman, M.K. (1998): Numerical Methods in Science and Engineering, The National Publishing Company, Chennai.

Note: Examination scheme and mode shall be as prescribed by the Examination Branch University of Delhi, from time to time.