# Department of Statistics, Siksha-Bhavana
# MSc in Statistics

(To be effective from 2020-2021 Academic Session)

| Structure of Course Contents | |
|---|---|
| Existing | Proposed |
| **SEMESTER- I** | **SEMESTER- I** |
| MSC-11: Linear Algebra and Linear Models | MSC-11: Linear Models and Distribution Theory |
| MSC-12: Regression Techniques | MSC-12: Real Analysis and Measure Theory |
| MSC-13: Stochastic Process and Distribution Theory | MSC-13: Statistical Inference I |
| MSC-14: Statistical Inference I | MSC-14: Sample Survey |
| MSC-15: Practical on Linear Algebra, Regression Techniques and Statistical Inference I | MSC-15: Practical on Linear Models, Distribution Theory and Statistical Inference I |
| MSC-16: Practical on Linear Models, Distribution Theory and Stochastic Process | MSC-16: Practical on Sample Survey |
| **SEMESTER- II** | **SEMESTER- II** |
| MSC-21: Statistical Inference II | MSC-21: Statistical Inference II |
| MSC-22: Applied Multivariate Analysis | MSC-22: Applied Multivariate Analysis |
| MSC-23: Sample Survey | MSC-23: Regression Techniques |
| MSC-24: Design of Experiments | MSC-24: Design of Experiments |
| MSC-25: Practical on MSC-21 and MSC-22 | MSC-25: Practical on MSC-21 and MSC-22 |
| MSC-26: Practical on MSC-23 and MSC-24 | MSC-26: Practical on MSC-23 and MSC-24 |
| **SEMESTER- III** | **SEMESTER- III** |
| MSC-31: Real Analysis and Measure Theory | MSC-31: Stochastic Process |
| MSC-32: Categorical Data Analysis and Advanced Data Analysis Techniques | MSC-32: Advanced Data Analysis Techniques |
| MSC-33:Course from the **Elective** module | MSC-33: Course from the **Elective** module |
| MSC-34: Course from the **Special** module | MSC-34: Course from the **Special** module |
| MSC-35: Practical on MSC-31 and MSC-32 | MSC-35: Practical on MSC-31 and MSC-32 |
| MSC-36: Practical on MSC-33 and MSC-34 | MSC-36: Practical on MSC-33 and MSC-34 |
| **SEMESTER- IV** | **SEMESTER- IV** |
| MSC-41: Courses from the **Special** module | MSC-41: Reliability Analysis |

| | |
|---|---|
| | |
| MSC-42: Courses from the **Special** module | MSC-42:Courses from the **Special** module |
| MSC-43: Practical on MSC-41 and MSC-42 | MSC-43: Courses from the **Special** module |
| MSC-44: Grand Viva-Voce | MSC-44: Practical on MSC-41, MSC-42 and MSC-43 |
| MSC-45: Project Work | MSC-45: Project Work |

| Existing | Proposed |
|---|---|
| **Structure of Elective Modules** | |
| MSE-1: Operations Research and Optimization Technique | MSE-1: Operations Research and Optimization Technique |
| MSE-2: Statistical Genetics | MSE-2:Statistical Genetics |
| | MSE-3: Statistical Ecology |
| 20% marks is for Internal Examination and 80% marks is for Semester-end Examination of each Module except Project Module. | 20% marks is for Internal Examination and 80% marks is for Semester-end Examination of each Module except Project Module. |
| **Department of Statistics, Siksha-Bhavana**<br>**Structure of Special Modules** | |
| Structure of Course Contents (Existing) | Structure of Course Contents (Proposed) |
| MSS-1: Actuarial Statistics | MSS-1: Actuarial Statistics |
| MSS-2:Reliability Analysis | MSS-2:Time Series Analysis |
| MSS-3:Time Series Analysis | MSS-3:Demography |
| MSS-4:Demography | MSS-4:Survival Analysis |
| MSS-5:Advance Design of Experiments | MSS-5:Clinical Trials and Bioassays |
| MSS-6:Survival Analysis | MSS-6:Advanced Mathematics for Statistics |
| MSS-7:Clinical Trials and Bioassays | MSS-7: Bayesian Inference |
| MSS-8:Statistical Ecology | MSS-8: Econometrics |
| MSS-9:Bayesian Inference | MSS-9: Introductory Data Science and Statistical Machine Learning |
| MSS-10:Advanced Mathematics for Statistics | |
| | |

**D E P A R T M E N T   O F   S T A T I S T I C S**

**S I K S H A  B H A V A N A,  V I S V A – B H A R A T I**

**Proposed Syllabus for the 4-Semester M.Sc. course in Statistics**

## Semester-I

**MSC – 11**
**Linear Models and Distribution Theory**

**Course Objective:**
*This is a basic and prerequisite course for the students before studying the courses like multivariate analysis.The first part isto make them understand what a linear model is and how various real-life problems can beexpressed and analyzed using linear models.*
*The second part covers a vast area of advanced mathematical statistics- stretching to different multivariate inferential problems.*

**Learning Outcomes:**
- *Understand and be proficient at theoretical developments in the analysis of linear models, including linear and quadratic forms, least squares, linear hypothesis testing, analysis of variance, etc.*
- *Apply the results from linear model theory in further advanced topics, such as nonparametric models, multivariate analysis, high-dimensional inference, etc.*
- *Sampling distributions under multivariate set-up—distribution of sample mean vector and sample variance covariance matrix—Wishart distribution, Hotelling's $T_2$*
- *Multivariate analysis of variance (MANOVA),*
- *Mahalanobis Distance*

### Linear Models

Gauss-Markov set-up and related concepts. (3L)
Tests of hypotheses for one and more than one linear parametric functions, confidence intervals and regions, Analysis of Variance, Power of F-test, Multiple comparison tests due to Tukey and Scheffe. (12L)

### Distribution Theory:

Non-central $\chi^2$, t and F distributions. (7L)
Distributions of quadratic forms under normality, Fisher-CochranTheorem and related results. (10L)
Random Sampling from $N_p(\mu, \Sigma)$, MLE's of $\mu$ *and* $\Sigma$ and their Stochastic independence. (8L)
Central Wishart distribution (without derivation) and its properties (with proof). (8L)
Distribution of Hotelling's $T^2$ and Mahalanobis's $D^2$ with applications.(4L)
Distribution of sampling multiple correlation coefficients, partial correlation coefficient and regression coefficient vector. (8L)

**References:**

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

Bapat, R.B: Linear Algebra and Linear Models

Johnson, S. and Kotz, N. L. (1972): Distributions in Statistics, Vol. I, II and III, Houghton and Miffin.

Anderson, T. W. (1983): An Introduction to Multivariate Statistical Analysis, 2$^{nd}$ Ed., Wiley.

Johnson, R. and Wichern (1992): Applied Multivariate Statistical Analysis, prentice-Hall, 3$^{rd}$ Ed.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979):Multivariate Analysis. Academic Press, London.

**MSC-12**

## Real Analysis and Measure Theory

***Course Objective:***

*The course is spitted into two parts: Real analysis and Measure theory. The objective of this course is to provide some basic and advanced ideas of real analysis and measure theory. The real analysis part mainly develops the student's analytical thinking and some prerequisite ideas of measure theory. The measure theory part helps the student to gain the knowledge about various measures. This part also develops the concept of probability and its related terms from the view of measure theory.*

***Learning Outcomes:***

*After completing the course the students will be able to:*

*· Understand various advanced ideas of real analysis like compact sets, Heine – Borel theorem etc.*
*· Understand sequence, series of functions and their convergence.*
*· Analyze the power series and related terms about its convergence.*
*· Know the terms related to different measures and their importance.*
*· Study the concept of random variables and its convergence.*
*· Grasp the idea of characteristic functions, law of large numbers and central limit theorem.*

**Real Analysis:**

Introduction to real numbers, Introduction to n-dimensional Euclidian space, open and closed intervals (rectangles), compact sets, Bolzano-Weirstrass theorem, Heine–Borel theorem. (15L)

Real valued functions, continuous functions, uniform continuity, sequences of functions, uniform convergence; Power series and radius of convergence. (10L)

**Measure Theory:**

Classes of sets, fields, sigma-fields, Borel sigma-field in $R_k$, sequence of sets, limsup and liminf of a sequence of sets. Measure, Probability measure, properties of a measure, Lebesgue and Lebesgue-Steljes measures on $R_k$. (10L)

Measurable functions, Random variables, sequence of random variables, almost sure convergence, convergence in probability and convergence in distribution. Monotone convergence theorem (Statement only). Fatou's lemma (Statement only), Dominated convergence theorem (Statement only). (10L)

Borel-Cantelli Lemma (Statement only), Weak law and strong law of large numbers for iid sequences.CLT for a sequence of independent random variables under Lindeberg's condition, CLT for iid random variables. (10L)

Characteristic function, Levy's continuity theorem (statement only) (5L)

**References:**

Adke, S. R. and Manjunath, S. M. (1984): An Introduction to Finite Markov Processes, Wiley Eastern.
Bhat, B. R. (2000): Stochastic Models: Analysis and Applications, New Age International, India.
Jagers, P. (1974): Branching Processes with Biological applications, Wiley.
Karlin, S. and Taylor, H. M. (1975): A First Course in Stochastic Processes, Vol.1, Academic Press.
Medhi, J. (1982): Stochastic Processes, Wiley Eastern.
Cramer, H. (1946): Mathematical Methods of Statistics, Princeton.
Pitman, J. (1993): Probability, Narosa publishing House.
Ash, R. (1972): Real analysis and probability, Academic Press.
Billingsley, P. (1986): Probability and Measure, Wiley.
Dubley, R. M. (1989): Real Analysis and Probability, Wordsworth and Brooks/Cole.
Kingman, J. F. C. and Taylor, S. J. (1966); Introduction to Measure and Probability.

**MSC-13**

## Statistical Inference-I

**Course Objectives:**

*The main objective of the course is to draw statistically valid conclusions about a population on the basis of a sample in a scientific manner. This course deals with fundamental concepts and techniques of statistical inference including point and interval estimation. Parametric, Nonparametric and Bayesian Estimation methods are to be explained. Students will be accustomed with theory as well as methods of estimation in this course.*

**Learning Outcomes:** *On completion of the course, students will be able to:*

- *Estimate unknown parameters of a given probability distribution using standard and nonstandard estimation techniques.*
- *Understand how to perform point and interval estimation.*
- *Familiar with the fundamental properties of estimators.*
- *Familiar with the different methods of finding out estimators of parameters.*
- *Familiar with loss functions, Bayes risks, prior distributions, derivation of Bayes estimates.*
- *Non-parametric estimates on abstract space and their properties.*

Introduction: Parametric models, Point estimation, Tests of hypothesis and interval estimation. (5L)

Information in data about the parameters as variation in Likelihood function, concept of no information, Sufficiency, Neyman-Fisher Factorization theorem, likelihood equivalence, Minimal sufficient statistic, Exponential families and Pitman families, Invariance property of sufficiency under one-one transformation of sample space and parameter space. Fisher information for one and several parameters models. (15L)

Methods of estimation: maximum likelihood method, methods of moments and percentiles, solution of likelihood equations, iterative procedures, choice of estimators based on unbiasedness, minimum variance, mean squared error, minimum variance unbiased estimators, Rao - Blackwell theorem, completeness, Lehmann – Scheffe theorem, necessary and sufficient conditions for MVUE, Cramer – Rao lower bound approach. (15L)

One sample U-statistics, Kernel and symmetric kernel, Two sample U-statistics, Asymptotic distribution of U-statistics. UMVUE property of U-statistics, Asymptotic distribution of linear function of order statistics. (15L)

Loss functions, expected loss, decision rules, optimal decision rules, Bayes Estimation. (10L)

**References:**


Ferguson, T. S. (1967): Mathematical Statistics, Academic Press.
Kale, B. K. (1999): A First Course on Parametric Inference, Narosa Publishing House.
Lehmann, E. L. (1986): Theory of Point Estimation (Student Edition).
Rao, C. R. (1973): Linear Statistical Inference.
Rohatgi, V. (1988): An Introduction to probability and Mathematical Statistics, Wiley Eastern Ltd., New Delhi (Student Edition).

**MSC-14**

## Sample Survey

***Course Objective:***
*After the introduction of the sampling schemes like Simple random Sampling, Stratified Random Sampling and Systematic Sampling in UG, this course consists of the details of the some advanced sampling schemes like Cluster Sampling, Two-stage sampling, Double Sampling, PPS sampling etc. The randomized response techniques for the surveys regarding sensitive topics are also introduced. The objective of the course is to prepare the students for all real life survey situations.*

***Learning Outcomes:*** *After completion of the course, the students will be able to*
*(1) Conduct a randomized response based survey regarding some sensitive characteristic.*
*(2) Estimate the proportion of individuals having some sensitive characteristic.*
*(3) Understand some basics of survey design.*
*(4) Estimate the population mean/total as well as the variance of the estimators under different sampling schemes, viz Cluster Sampling, Two-stage Sampling, Double Sampling and PPS Sampling.*

Two-stage and Double Sampling and related results in estimation of population mean and total. Allocation problem in two-stage and double sampling. (15L)

Concepts of Sampling Design and Design based Inference. Rao-Blackwellization of estimators. (10L)

Unequal probability sampling: ppswr/wor methods [including Lahiri's Scheme] and related estimators of a finite population [Hansen-Hurvitz and Desraj Estimators for a general sample size and Murthy's Estimator for a sample of size 2]. Rao-Hartley and Cochran's sampling strategy.
(25L)

Randomized Response Technique. (10L)

**References:**

Chaudhuri, A. and Vos, J. W. E. (1988): Unified Theory and Strategies of Survey Sampling, North-Holland, Amsterdam.
Chaudhuri, A. and Mukerjee, R. (1988): Randomized Response – Theory and Techniques, New York, Marcel Dekker Inc.
Cochran, W. G. (1984): Sampling Techniques, 3rd Ed. Willey.
Des Raj and Chandhok (1998): Sampling Theory, Narosa.
Mukhopadhyay, P. (1998): Small area estimation in survey sampling, Narosa.

**MSC-15:**

Practical on Linear Models, Distribution Theory and Statistical Inference I

**MSC-16:**

Practical onSample Survey

**Semester-II**

**MSC-21**

**Statistical Inference-II**

***Course Objective:***
*This course catches the flavor of advanced level testing of hypothesis. At the end detailed discussion on nonparametric tests is furnished.*

***Learning Outcome:***
*After completion of the course, the students will be able to understand*
- *Neyman-Pearson fundamental lemma*
- *MP test, Uniformly most powerful test, Uniformly most powerful unbiased test*
- *Alpha-similar tests constructions and its applications*
- *Invariance tests*
- *Test for composite null vs composite alternatives*
- *Construction of nonparametric tests statistic and their exact and large sample distributions.*
- *Wilcoxon signed rank test, Mann Whitney test*
- *Goodness of fit test-Kolmogorov Smirnov test, Anderson Darling tests*
- *Nonparametric one way ANOVA test , Nonparametric two way ANOVA test.*

Tests of Hypothesis: Concepts of critical regions, test functions, two kinds of errors, size function, power function, level, MP, UMP tests, Neyman – Pearson Lemma, MP test for simple null against simple alternative hypothesis, UMP tests for simple null hypothesis against one sided alternatives and for one sided null against one sided alternatives in one parameter exponential family. Extension of these results to Pitman family when only upper and lower end depends on the parameter and to distributions with MLR property, non-existence of UMP test for simple null against two sided alternatives in one parameter exponential family. (25L)

Rank tests, Locally most powerful rank tests, Linear rank statistics and theirdistributional properties under null hypothesis, Pitman's asymptotic relative efficiency. (15L)

One sample location problem, sign test and signed rank test, two sample Kolmogorov-Smirnov tests. Two sample location and scale problems. Wilcoxon-Mann-Whitney test, normal score test, ARE of various tests based on linear rank statistics. Kruskal-Wallis K sample test. (20L)

**References:**

Rao, C. R. (1973): Linear Statistical Inference.
Lehmann, E. L. (1986): Testing Statistical Hypothesis (Student Edition).
Gibbons, J.D. (1985): Nonparametric statistical inference, 2nd ed., Marcel Dekker, Inc.

**MSC– 22**

## Applied Multivariate Analysis:

***Course Objectives:*** *To impart the concepts and applications of various multivariate statistical techniques. To make the students understand how to analyze multivariate data using statistical theories.*

***Learning Outcomes:*** *After this course students should be able to*
*1. formulate analysis of real-life multivariate data using statistical principles along with softwares.*
*2. model and forecast various continuous and or discrete dependent variables depending on more than one independent variables.*
*3. apply various supervised and unsupervised learning methods for real-life applications.*

Principal Component Analysis and Canonical Correlation Analysis: Concepts, Computation and Large sample inferences. (12L)

Factor Analysis: Concept, Factor model, estimation of factor loadings,Factor rotation, estimation of factor scores, Model fit. (12L)

Multivariate Analysis of Variance (MANOVA): one way and two-way classified data with one observation per cell. (12L)

Developing discriminant function considering cost factor. Fisher's method for discriminating two multivariate populations with common dispersion matrix, sample discriminate function, extension for several populations. (12L)

Cluster Analysis: Proximity measures, Hierarchical and Non-hierarchical clustering techniques (12L)

**References:**

Giri, N. C. (1977): Multivariate Statistical Inference, Academic Press.
Seber, G. A. F. (1984): Multivariate Observations, Wiley.
Anderson, T. W. (1983): An Introduction to Multivariate Statistical Analysis, 2$^{nd}$ Ed., Wiley.
Johnson, R. and Wichern (1992): Applied Multivariate Statistical Analysis, prentice-Hall, 3$^{rd}$ Ed.
Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979):Multivariate Analysis. Academic Press, London.

**MSC – 23**

## Regression Analysis

**Course Objective:**

*Regression Analysis is the most common statistical modelling approach used in data analysis. In this course students will learn various statistical methods for investigating functional relationships among variables. Regression analysis is an applied topic that is used in various sectors like academic, company, forecasting etc. The objective is to provide the basic and advanced idea of regression analysis, so that students can be applied this modelling to solve various real life problems and draw inferences from the data.*

**Learning Outcomes:**

*After completing the course the students will be able to:*

- *Analyze and fit linear, polynomial and multiple linear regression models using data.*
- *Detect and overcome the issues like model adequacy, multicollinearity and influential points.*
- *Perform various statistical inferences related to regression analysis.*
- *Fit the nonlinear, logistic, poisson regression model and their inferences.*
- *Perform all the above computation using R/ SAS.*

Multiple linear regression. (5L)
Model Adequacy Checking: Residual Analysis. (5L)
Diagnostics of Leverage and influence, Variable selection and Model Building.(5L)
Multicollinearity, Robust Regression. (10L)
Introduction to Nonlinear Regression: Parameter estimation, Statistical Inference.(10L)
Introduction to Generalized Linear Models: Logistic and Poisson Regression. (10L)
Nonparametric Regression. (10L)
Use of Statistical Packages: R/ SAS (5L)

**References:**

Cook, R. D. and Weisberg, S. (1982). Residual and Influence in Regression. Chapman and Hall.
Draper, N. R. and Smith, H. (1998). Applied Regression Analysis. 3rd Ed. Wiley.
Gunst, R. F. and Mason, R. L. (1980). Regression Analysis and Its Applications - A Data Oriented Approach. Marcel and Dekker.
Rao, C. R. (1973). Linear Statistical Inference and Its Applications. Wiley Eastern.
Weisberg, S. (1985). Applied Linear Regression. Wiley.

**MSC– 24:**

## Design of Experiments

***Course Objective:***
*The objective of the course is to develop a systematic method to determine the relationship between factors affecting a process and the output of the process. It is used to find out cause and effect relationship. This information is needed to manage process inputs in order to optimize the output.*

***Learning Outcomes:*** *After completion of the course, the students will be able to*
*(1) Analyze generalized block designs.*
*(2) Analyze a block design with missing observations.*
*(3) Understand the purpose of confounding and analyze a factorial design with some confounded effects.*
*(4) Construct Balanced Incomplete Block designs by different methods.*
*(5) Construct fractional factorials*

Factorial experiments: $2^n$, $3^n$factorial experiments, illustrations, main effects and interactions, confounding and illustrations. (12L)

Split and Strip plot Analysis. (8L)

General block design and its information matrix (C). (10L)

Criteria for connectedness, balance and orthogonality. (10L)

Intra-block analysis including estimability, best point estimates,interval estimates of estimable linear parametric functions and testing of linear hypotheses). (10L)

BIBD: Description, construction and Analysis. (10L)

**References:**


Dey, Aloke (1986): Theory of Block Designs, Wiley Eastern.
Dean, A. and Voss, D. (1999): Design and Analysis of Experiments, Springer.
Das, M. N. and Giri, N. (1979): Design and Analysis of Experiments, Wiley Eastern.
Joshi, D. D. (1987): Linear Estimation and Design of Experiments, Wiley Eastern.
Montgomery, C. D. (1976): Design and Analysis of Experiments, Wiley, New York.
Searle, S. R., Casella, G. and McCulloch, C. E. (1992): Variance Components, Wiley.


**MSC-25:**

Practical on Applied Multivariate Analysis and Statistical Inference II

**MSC-26:**

Practical on Regression Techniques and Design of Experiments

## <u>Semester-III</u>

**<u>MSC – 31</u>**

## <u>Stochastic Process:</u>

**Course Objective:**
*This course covers a vast area of advanced mathematical statistics- stretching to stochastic process.*

**Learning Outcomes:**
- *Stability of stochastic process—stationarity and limiting distribution of a process*
- *Markovian model, classifications of state under discrete time Markov model*
- *Gambler's ruin, Random walk model*
- *Poisson process, death and birth process, queueing theory*

Introduction to Stochastic Processes (SP): classification of SP's according to state space and time domain. Countable state Markov chains (MC's), Chapman-Kolmogorov equations, calculation of n-step transition probability and its limit. Stationary distribution, classification of states, transient Mc, random walk and gambler's ruin problem. Applications from social, biological and physical sciences. (25L)

Discrete state space continuous MC: Kolomogorov-Feller differential equations. Poisson process, birth and death process. Applications to queues and storage problems. First-passage time and other problems. Martingales and related results. (25L)

Renewal theory: Elementary renewal theorem and applications. Statement and applications of key renewal theorem. (10L)

**References:**

Adke, S. R. and Manjunath, S. M. (1984): An Introduction to Finite Markov Processes, Wiley Eastern.
Bhat, B. R. (2000): Stochastic Models: Analysis and Applications, New Age International, India.
Jagers, P. (1974): Branching Processes with Biological applications, Wiley.
Karlin, S. and Taylor, H. M. (1975): A First Course in Stochastic Processes, Vol.1, Academic Press.
Medhi, J. (1982): Stochastic Processes, Wiley Eastern.

**MSC – 32:**

## Advanced Data Analysis Technique

***Course Objectives:***
*In most of the applied research problems, it is a common practice to deal with categorical variables. Besides, some of the analyses are to be made on the basis of simulated data because of the lack of proper real-life data set. This particular course is designed to give some idea about the inferences on categorical data as well as about some popular classical and Bayesian computing techniques which are appropriate in absence of proper data support.*

***Learning Outcomes:*** *After completion of the course, the students will be able to*
*1. Perform the analysis of contingency tables and fitting of generalized linear models.*
*2. Perform Gibbs sampling technique to simulate data from a high-dimensional posterior distribution.*
*3. Apply Markov Chain Monte Carlo Technique for simulation.*
*4. Apply Bootstrap and jackknife resampling techniques.*

Generalized linear models and its properties. (10L)

Logisitic and Poisson regression: logit model for dichotomous data with single and multiple explanatory variables, ML estimation, large sample tests about parameters,
goodness of fit, analysis of deviance, variable selection, extension to polytomous data,
Introduction to Poisson regression. (10L)

Introduction to Hierarchical log-linear models. Fitting of log-linear models. (5L)

EM algorithm: applications to missing and incomplete data problems, mixture models, Application in Cluster analysis (10L)

Markov Chain Monte Carlo methods: Gibbs sampling for multivariate simulation, Metropolis-Hastings Algorithm. (10L)

Bootstrap methods: resampling paradigms, bias and standard errors, confidenceintervals, bootstrapping in regression. (10L)

Jackknife and cross-validation: jackknife in sample surveys, cross-validation for tuning parameters.(5L)

**References:**

Agresti (1990). Categorical Data Analysis. Wiley, New York.
P. McCullagh and J.A. Nelder (1999). Generalized Linear Models, Second edition. Chapman and Hall, New York.
Robert and Casella (2004): Monte Carlo Statistical Methods, Springer

Efron, B. and Tibshirani, R.J. (1993) An Introduction to the Bootstrap. Chapman & Hall, New York, London.
J.K. Ghosh, M. Delampady& T. Samanta: Bayesian Inference

**MSC – 33**

Course from the **Elective**module

**MSC – 34**

Course from the **Special**module

**MSC – 35**

**Practical** on MSC-31 and MSC-32

**MSC – 36**

**Practical** on MSC-33 and MSC-34

## Semester-IV

**MSC – 41:**

## Reliability Analysis

***Course Objectives:*** *To impart the concept of reliability and how statistical and probabilistic theories are applied to model and explain life of a mechanical component along with prediction of the same.*

***Learning Outcomes:*** *After this course students will be able to*
*1. model and explain the operation time of a mechanical component.*
*2. to predict the reliability of a component, system and of a finished product.*
*3. explain the nature of the lifetime of an item as well.*

Reliability concepts and measures, Components & systems: coherent systems, reliability of the coherent systems. Cuts and paths, modular decomposition, bounds on system reliability; structural and reliability importance of components. (12L)

Life distributions, reliability function, hazard rate, common life distributions: exponential, weibull, gamma etc. (6L)

Reliability estimation based on failure time in various censored life tests. Stress-strength reliability and its estimation. (8L)

Notions of ageing, IFR, IFRA, NBU, DMRL and NBUE and their duals, loss of memory property of the exponential distribution. Closures of these classes under formation of Coherent systems, Convolution and Mixtures. (14L)

Univariate shock models and life distribution arising out of them. Bivariate shock models, common bivariate exponential distribution and their properties. (6L)

Maintenance and replacement policies; availability of repairable systems; modeling of a repairable system by a non-homogeneous Poisson process. (8L)

Basic ideas of accelerated life testing and basic ideas of software reliability. (6L)

**References:**


Barlow R.E. and Proschan F. (1985): Statistical Theory of Reliability and Life testing: Holt, Rinehart and Winston.
Lawless J.F. (1982): Statistical model and Methods of Life time data, John Willey.
Bain L.J. and Engelhardt (1991): Statistical Analysis of Reliability and Life testing Models, Marcel Dekker.
Nelson, W. (1982): Applied Life Data Analysis, John Willey.

**MSC – 42**

Courses from the **Special**module

**MSC – 43**

Courses from the **Special** module

**MSC – 44**

**Practical** on MSC-41, MSC-42and MSC-43

**MSC – 45**

**Project(Maximum Marks 100)**

## Module: Elective

## MSE-1: Operations Research and Optimization Technique

**Course Objectives:**
*The objective of the course is to provide basic idea about operations research and utilizing optimization techniques as its basic tools. Use of statistical and mathematical tools in operations research and their applications to decision making process is primary concern. Role of operations research under different constraint conditions are to be studied.*

***Learning Outcomes:*** *On completion of the course, students will be able to:*
- *Formulate the problem in operations research.*
- *Establish the relationship between the variables and constraints by constructing the model.*
- *Identify the possible alternative solutions and select the optimal one.*
- *Install, test and establish the optimal solution.*
- *Learn the tools like Linear Programming Problems, Transportation, assignment, replacement and operational gamming.*
- *Familiar with the queuing and different inventory models.*

Definition and scope of Operational research; phases in Operations research; models and their solutions; decision-making under uncertainty and risk, use of different criteria; sensitivity analysis.(12L)

Review of LP problems; transportation and assignment problems. Introduction to game theory, 2-person zero sum game. (12L)

Queueing models-specifications and effectiveness measures. Steady-state solutions of M/M/1 and M/M/c models with associated distributions of queue-length and waiting time. M/G/1 queue and Polazcek-Khinchine result. Steady-state solutions of $M/E_k/1$ and $E_k/M/1$ queues. (12L)

Branch and bound method for solving traveling salesman problem. 0-1 Programing.
Replacement problems; block and age replacement policies; dynamic programming approach for maintenance problems; replacement of items with long life. (12L)

s-S policy for inventory and its derivation in the case of exponential demand, multi echelon inventory models, models with variable supply and models perishable items, estimation of EOQ in some simple cases. (12L)

**References:**

Taha, H. A. (1982): Operational Research: An Introduction, Macmillan.
Kanti, S., Gupta, P. K. and Singh, M. M. (1985): Operations Research, Sultan Chand & Sons.

## MSE-2: Statistical Genetics

**Course Objective:** *Statistical genetics has played a pivotal role in the discovery of genes that cause disease in humans. This module introduces the basic concepts and terms in genetics and demonstrates the use of statistical models to identify disease genes in humans. This course will provide an introduction to statistical methods for genetic studies. The basic material in statistical genetics is covered, focusing on association analysis. The emphasis of this course is on understanding basic concepts and methods and how they are applied in the health sciences.*

**Learning Outcomes:**
*Having successfully completed this module, students will be able to:*
*(1) Understand the basic concepts of genetics .*
*(2) Know how to analyze the most usual forms of genetic linkage and allelic association data.*

Fundamentals of Genetics (10L)

Mendel's laws, Estimation of allele frequencies, Hardy-Weinberg law, Mating tables, Snyder's ratios, Models of natural selection and mutation, Detection and estimation of linkage (recombination), Inheritance of quantitative traits. (15L)

Evolution of DNA sequences - Kimura's two-parameter and Jukes-Cantor models, DNA sequence alignment - Needleman-Wunsch and Smith-Waterman algorithms.
Gene trees and species trees.Estimation of evolutionary and population genetic parameters from DNA sequence data. (20L)

Gene mapping methodologies: (a) transmission-disequilibrium test, (b) linkage disequilibrium mapping, (c) quantitative trait locus (QTL) mapping , (d) expression QTL mapping. (15L)

**References:**

Kenneth Lange: Mathematical and statistical methods for genetic analysis, Springer
**Wu**, Rongling, **Ma**, Changxing, **Casella**, George: Statistical Genetics for Quantitative Traits, Springer, 2007
Robert V. **Hogg** and T. **Craig:** Introduction to Mathematical Statistics, Pearson, 2005.

## MSE-3: Statistical Ecology

***Course Objectives:***
*Statistical ecology is a recently introduced topic that creates a bridge between statistics and ecology. Here various statical methods are used to study plenty of ecological questions throughout the world. The main objective of this course is to provide the theory as well as application of statistical techniques in various ecological contexts. Some of the topics are closely associated with real-life issues and using statistics we can draw the inferences.*

***Learning Outcomes:***
*After completing the course the students will be able to:*
• *Understand population dynamics and related single species and interactive models.*
• *Understand various ideas related to abundance estimation like capture-recapture, nearest neighbor method.*
• *Recognize the concept of ecological diversity through statistical approach.*

Introduction to Ecology and evolution. (5L)
Population dynamics: single species – Exponential, logistic and Gompertz models, Leslie matrix model for age and stage structured population, Survivorship curves – Constant, monotone and bathtub shaped hazard rates (20L)
Two species: Lotka – Volterra equations, isoclines, competition and coexistence, predator-prey oscillations (15L)
Abundance estimation: Capture – recapture, nearest neighbour, line transect sampling, indirect methods (14L)
Ecology Diversity: Species abundance curve, Indices of diversity (Simpson's index, Shannon – Wiener index), Diversity as average rarity (6L)

**References:**

Anil Gore and S. A. Paranjpe (2000): A course on Mathematical and Statistical Ecology (Kluwer)
Clark, C. W. (1976): Mathematical Bioeconomics: Optimal Management of Renewable Resources (Wiley)
Maynard Smith, J. (1982): Evolution and the Theory of Games (CambridgeUniversity Press)
Pielou, E. C. (1977): An Introduction to Mathematical Ecology (Wiley)

## Module: Special

## MSS-1: Actuarial Statistics

**Course Objectives:**
*The aim of Actuarial Statistics is to provide grounding in mathematical and statistical methods that are of relevance for actuarial work. It is a discipline that assesses financial risks in the insurance and finance fields. It applies the mathematics of probability and statistics to define, analyze and solve the financial implications of uncertain future events.*

**Learning Outcomes:** *On completion of the course, students will be able to:*
- *Equipped with knowledge of statistical distributions, methods to summarize data, the principles of statistical inference, regression models (including generalized linear models) .*
- *Accustomed with individual and aggregate claims and their applications.*
- *Utility functions and their uses in insurance.*
- *Life table functions and their applications.*
- *Life insurance, Life annuities, Net premiums, Net reserves.*
- *Multiple life functions and Multiple decrement functions.*

General insurance: Claim amount distributions, approximating the individual model, stop-loss insurance. (5L)

Models for individual claims and their sums, Distribution of aggregate claims, compound Poisson distributions and its applications. (5L)

Utility theory, insurance and utility theory (2L)

Probability Models Life tables: Survival function, curtate future life time, force of mortality. (3L)

Life table and its relation with survival function, examples, assumptions for fractional ages, some analytical laws of mortality, select and ultimate tables. (5L)

Principles of compound interests: Nominal and effective rates of interest and discount, force of interest and discount; compound interest, accumulation factor, continuous compounding. (5L)

Life insurance: insurance payable at the moment of death and at the end of the year of death level benefit insurance, endowment insurance, differed insurance and varying benefit insurance, recursion, commutation functions. (5L)

Life annuities : Single payment, continuous life annuities, discrete life annuities, life annuities with monthly payments, commutation functions, varying annuities, recursions, complete annuities – immediate and apportionable annuities- due. (6L)

Net premiums: continuous and discrete premiums, true monthly payment premiums, apportionable premium, commutation functions, accumulation type benefits

(4L)

Net premium reserves: Continuous and discrete net premium reserve, reserves based on true monthly premiums, reserves on an apportionable or discounted continuous basis, reserves at fractional durations, allocations of loss to policy years, recursive formulas and differential equations for reserves, commutation functions.(8L)

Multiple life functions, joint life and last survivor status, insurance and annuity benefits through multiple life functions evaluation for special mortality laws. (4L)

Multiple decrement models, deterministic and random survivor groups, associated single decrement tables, central rate of multiple decrement, net single premiums and their numerical evaluations.

(6L)

Some practical considerations: Premiums that include expenses – general expenses types of expenses, per policy expenses. (2L)

**References:**

Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. (1997): Actuarial Mathematics, Society of Actuaries, Ithaca Illiois, U.S.A.

Spurgeon, E. T. (1972): Life Contingencies, Cambridge University Press.

Neill, A. (1977): Life Contingencies, Heineman.

## MSS-2: Time Series Analysis

***Course Objective:*** *To impart the concept of time series and how to develop statistical models to forecast a time series for practical use/planning.*

***Learning Outcome:*** *After completion of the course, students should be able to*
*1. Distinguish time series and cross-sectional data.*
*2. Model various real-life time series data and forecast them along with the forecast errors.*

Smoothing Time series using filters. (6L)

Representation of time series as a stochastic process. Weakly and strongly Stationary process and their examples. (6L)

Ergodicity, sufficient condition for ergodicity. (6L)

Autocorrelation and partial autocorrelation functions and their properties (10L)

AR, MA, ARMA models and their properties. (10L)

Estimation of ARIMA model parameters, (6L)

Forecasting: minimum MSE forecast using ARIMA models (6L)

Spectral Analysis of weakly Stationary process. Periodogram and correlogram analysis. (10L)

**References:**

Anderson, T.W.: The Statistical Analysis of Time series
Brockwell and Davis (2002): Introduction to Time series and forecasting
Gujarati, D.: Basic Econometrics.

## MSS-3: Demography

**Course Objectives:**
*The scientific nature of demography proves the following four objectives of demography. These are to achieve knowledge about the size, composition, organization and distribution of the population. To describe the past evolution, present distribution and future changes in the population of an area.*

**Learning Outcomes:** *On completion of the course, students will be able to:*
- *Coverage and content errors in demographic data*
- *Measure of fertility, stochastic model for reproduction*
- *Measures of mortality.*
- *Life table functions and their applications.*
- *Population growth and population projection.*
- *Stochastic models for social and occupational mobility.*

Coverage and content errors in demographic data, use of balancing equation and Chandrasekharan – Deming formula to check completeness of regression data. Adjustment of age data – use of Whipple, Myer and UN indices. Population composition, dependency ratio. (15L)

Measure of fertility, stochastic model for reproduction, distributions of time to first birth, inter – live birth intervals and of number of births, estimation of parameters, estimation of parity progression ratios from open birth interval data. (15L)

Measures of mortality, construction of abridged life tables, Distributions of life table functions and their estimation. (10L)

Stable and quasi stable populations, intrinsic growth rate. Models for population growth, and their fitting to population data. Stochastic models for population growth. (8L)

Stochastic models for migration and for social and occupational mobility based on Markov chains. Estimation of measures of mobility. (7L)

Methods for population projection. Use of Leslie matrix. (5L)

**References:**


Bartholomew, D.J. (1982): Stochastic Models for Social Processes, John Willey.
Benjamin B., (1969): Demographic Analysis, George Allen and Unwin.
Chiang,C.L. (1968): Introduction to Stochastic Process in Bio-statistics; John Willey.
Cox, P.R. (1970): Demography, Cambridge University Press.
Keyfitz, N. (1977): Applied Mathematical Demography, Springer Verlag.
Spiegelman, M. (1969): Introduction to Demographic Analysis, Harvard University Press.
Wolfenden, H.H. (1954): Population Statistics and their Compilation; American Actuarial Society.

## MSS-4: Survival Analysis

***Course Objectives:***
*The aim of this course is to enable students to analyse data from studies in which individuals are followed up until a particular event occurs - e.g. death, cure, relapse - making use of follow-up data for those who do not experience the event, with proper attention to underlying assumptions and a major emphasis on the practical interpretation and communication of results.*

***Learning Outcomes:***
*After the successful completion of the course, students should be able to:*
*1. Collaborate with Health scientists.*
*2. Apply basic methods for estimation and statistical inference when working with censored data.*

Concepts of time, Order and random Censoring, likelihood in these cases. Life tables, hazard rate, mean residual life and their elementary properties. Lifedistributions-Exponential Gamma, Weibull, Lognormal, Pareto, Linear Failure rate. Parametric inference (Point estimation, Confidence Intervals, Scores, LR, MLE tests (Rao-Willks-Wald)) for these distributions. (25L)

Estimation of survival function - Actuarial Estimator, Kaplan -Meier Estimator, Nelson-Aalen Estimator, Greenwood's formula (15L)

Two sample problem- Weighted Log rank test. Mantel-Haenszel test (6L)

Semi-parametric regression for failure rate - Cox's proportional hazards model with one and several covariates. Inference based on partial likelihood. (8L)

Competing risks model, parametric and non-parametric inference for this model. (6L)

**References:**

Cox, D.R. and Oakes, D. (1984) : Analysis of Survival Data, Chapman and Hall, New York.
Gross A.J. and Clark, V.A. (1975) : Survival Distribution : Reliability applications in the Biomedical Sciences, John Wiley and Sons.
Elandt - Johnson, R.E. Johnson N.L. : Survival Models and Data Analysis, John Wiley and Sons.
Miller, R.G. (1981) : Survival Analysis (John Wiley).
Kalbfleisch J.D. and Prentice R.L. (1980), The Statistical Analysis of Failure Time Data, John Wiley.

## MSS-5: Clinical Trials and Bioassay

***Course Objective:***
*Bioassays are essential tools for pre-clinical research. By revealing whether a compound or biologic has the desired effect on your biological target, bioassays can drive decision-making throughout the drug discovery process, to ultimately bring new drugs to patients. This course will specially help the students who are interested to get involved in clinical research.*

***Learning Outcome:***
*After the completion of the course, the students should be able to:*
*(1) Understand the concept of basic study designs, group sequential design and adaptive designs.*
*(2) Understand the notion of clinical trials.*
*(3) Understand different types of biological assays.*
*(4) Get involved in medical research.*

### Clinical Trials:

Introduction, Basic Study Design, Randomization and Blinding (3L)

Ethical Issues in Clinical Trials. (3L)

Sample Size, Trial Conduct and Monitoring (4L)

Group sequential designs and interim stopping rules. (8L)

Adaptive Designs. (7L)

Issues in Data Analysis, Closeout, Reporting and Interpreting of Results (5L)

### Bioassay

Logic of biological assay; dosage response curves; quantitative and quantal responses; parallel line and slope-ratio assays; simplified estimators; sequential assays; problem of design. (10L)

PK/PD Models : Compartment models and their identifiability, Numerical solutions of coupled differential equations, Mixed effects models and population PK/PD
(20L)

### References:

Friedman LM, Furberg CD, DeMets DL. Fundamentals of ClinicalTrials. 3rd ed. Mosby-Year Book, Inc., St. Louis, 1996.
S.Piantadosi: Clinical Trials - A Methodologic Perspective

B.S.Everitt and A.Pickles: Statistical Aspects of Design & Analysis of Clinical Trials
J.Whitehead: The Design and Analysis of Sequential Clinical Trials
Z.Govindarajulu: Statistical Techniques in Bioassay
D.J.Finney: Statistical Methods in Bioassay

## MSS-6: Advanced Mathematics for Statistics

***Course Objectives:***
*This is an advanced mathematics course that is designed for the interdisciplinary research purpose. This course will also be helpful for various competitive exams like NET, SET, RET, GATE etc. The main objective of this course is to provide the theory as well as application of some applied mathematics topic. Some of the topics are very much associated with Statistics and its application.*

***Learning Outcomes:***
*After completing the course the students will be able to:*
- *Understand various advanced techniques related to ordinary differential equations like: Existence and Uniqueness of solutions, Green's function etc.*
- *Understand various ideas related to partial differential equations.*
- *Know some basic and advanced terms related to linear algebra and numerical analysis.*
- *Recognize the concept of calculus of variation and linear integral equation.*

Ordinary Differential Equations (ODEs):

Existence and Uniqueness of solutions of initial value problems for first order ordinary differential equations, singular solutions of first order ODEs, system of first order ODEs. General theory of homogenous and non-homogeneous linear ODEs, variation of parameters, Sturm-Liouville boundary value problem, Green's function. [10L]

Partial Differential Equations (PDEs):

Lagrange and Charpit methods for solving first order PDEs, Cauchy problem for first order PDEs. Classification of second order PDEs, General solution of higher order PDEs with constant coefficients, Method of separation of variables for Laplace, Heat and Wave equations. [10L]

Linear Algebra:

Vector spaces, subspaces, linear dependence, basis, dimension, algebra of linear transformations. Algebra of matrices, rank and determinant of matrices, linear equations. Eigenvalues and eigenvectors, Cayley-Hamilton theorem. Matrix representation of linear transformations. Change of basis, canonical forms, diagonal forms, triangular forms, Jordan forms. Inner product spaces, orthonormal basis. Quadratic forms, reduction and classification of quadratic forms. [10L]

Numerical Analysis :

Numerical solutions of algebraic equations, Method of iteration and Newton-Raphson method, Rate of convergence, Solution of systems of linear algebraic equations using Gauss elimination and Gauss-Seidel methods, Finite differences, Lagrange, Hermite and spline interpolation, Numerical differentiation and integration, Numerical solutions of ODEs using Picard, Euler, modified Euler and Runge-Kutta methods. [10L]

Calculus of Variations:

Variation of a functional, Euler-Lagrange equation, Necessary and sufficient conditions for extrema. Variational methods for boundary value problems in ordinary and partial differential equations. [10L]

Linear Integral Equations:

Linear integral equation of the first and second kind of Fredholm and Volterra type, Solutions with separable kernels. Characteristic numbers and eigenfunctions, resolvent kernel. [10L]

## MSS-7: Bayesian Inference

**Course Objective:**

*Bayesian methods have some advantages over the classical methods, viz. these provide a natural and principled way of combining prior information with data, within a solid decision theoretical framework. These also provide inferences that are conditional on the data and are exact, without reliance on asymptotic approximation. Small sample inference proceeds in the same manner as if one had a large sample. These advantages make these methods widely applicable to the data scientists. This course is designed to enlighten the students about the basics of Bayesian inferences.*

**Learning Outcomes:**

*1. Explain in detail the Bayesian framework for data analysis and its flexibility and be able to demonstrate when the Bayesian approach can be beneficial*
*2. Develop, analytically describe, and implement both single and multiparameter probability models in the Bayesian framework.*
*3. Demonstrate the role of the prior distribution in Bayesian inference and be able to articulate the usage of non-informative priors and conjugate priors.*
*4. Show high level Interpretation of Bayesian Analysis Results and be able to readily perform Bayesian model evaluation and assessment.*
*5. Demonstrate the necessary skills to: fit hierarchical models, provide thorough technical specifications for these models.*
*6. Perform Bayesian computation using Markov chain Monte Carlo methods using R*
*7. Demonstrate how Bayesian Methods can be used to solve real world problems.*
*8. Communicate complex statistical ideas to a diverse audience.*
*9. Demonstrate the necessary research skills to form a hypothesis, collect and analyse data, and reach appropriate conclusions.*

Concepts of admissibility, Bayes rules, admissibility of Bayes rules. Minimax Estimation. (12L)
Bayesian Linear Regression with Conjugate priors, Bayesian Model Selection, Bayesian Information Criterion.(13L)
Bayesian Hypothesis Testing (One-sided and Two-sided Example), The Bayes Factor, A Test for Comparing Two Population Means.(10L)
Hierarchical Bayes Examples, Exchangeability, Hierarchical Bayesian Analysis Shrinkage and Bayesian Estimation, Empirical Bayes Estimation (with examples), Comparison of Hierarchical vs. Empirical Bayes.(10L)
Bayesian Probit and Logistic Regression (Multi-category Ordinal Response).
Multinomial-Dirichlet model.(10L)
Nonparametric Bayes.(5L)

**References:**

J.O. Berger: Statistical Decision Theory and Bayesian Analysis
C.P. Robert: The Bayesian Choice
J.K. Ghosh, M. Delampady& T. Samanta: Bayesian Inference
P. Lee: Bayesian Statistics – An Introduction
Berger, J. O. (1985): Statistical Decision Theory and Bayesian Analysis, 2nd Ed., Springer.

## MSS-8 (Econometrics)

***Course Objectives:*** *To impart the knowledge of problems with violation of different assumptions while modelling real-life economic problems to solve using statistical theories and how to find solutions to those.*

***Learning Outcomes:*** *After finishing this paper the students should be able to identify the assumption(s) being violated in a real-life (economic) dataset and to remove the issue from it and model the data for practical use.*

Introduction: Objective behind building econometric models, nature of econometrics, model building, role of econometrics, residual diagnostics, review of Gauss Markov model (15L)

Multicollinearity: Introduction and concepts, detection of multicollinearity, consequences, tests and solutions of multicollinearity, specification error. (15L)

Generalized least squares estimation, Aitken estimators. Autocorrelation: concept, consequences of autocorrelated disturbances, detection and solution of autocorrelation.Heteroscedastic disturbances: Concepts and efficiency of Aitken estimator with OLS estimator under heteroscedasticity. Consequences of heteroscedasticity. Tests and solutions of heteroscedasticity. (20L)

Autoregressive and Lag models, Dummy variables, Qualitative data for structural and reduced model, logit, probit model. (10L)

## MSS-9 : Introductory Data Science and Statistical Machine Learning

***Course Objectives:*** *The aim of this module is to apply quantitative modeling and data analysis techniques to the solution of real world business problems, communicate findings and effectively present results using data presentation and visualization techniques.*

***Learning Outcomes:*** *After finishing this paper the students should be able to*

- *(i)* *recognize and analyze ethical issues in business related to intellectual property, data security, integrity, and privacy.*
- *(ii)* *apply ethical practices in everyday business activities and make well-reasoned ethical business and data management decisions.*
- *(iii)* *demonstrate knowledge of statistical data analysis techniques utilized in business decision making.*
- *(iv)* *apply principles of Data Science to the analysis of business problems.*
- *(v)* *use data mining software to solve real-world problems.*
- *(vi)* *apply algorithms to build machine intelligence.*

Definitions, Data wrangling, Descriptive Analytics and visualization, Feature extraction techniques, Supervised learning: Regression and Classification, Unsupervised learning, Ensemble learning (averaging, Decision tree, Logistic regression, Support Vector Machines, Bagging, Random Forest, Boosting and gradient boosting), association rules, text mining. (40L)

All the above techniques should be demonstrated using either of R or Python or both. (20L)

**References:**

An Introduction to Statistical Learning- with Applications in R: Gareth James, Daniela Witten, Trevor Hastie, Springer, 2013.
The Elements of Statistical Learning- Data mining, Inference and Prediction: Trevor Hastie, Robert Tibshirani, Jerome Friedman, Springer, 2003
Machine Learning for Hackers: Drew Conway and John Myles White, O'Reilly, 2012.