PONDICHERRY UNIVERSITY

DEPARTMENT OF STATISTICS

SYLLABUS FOR M.Sc STATISTICS
(CBCS Pattern)
Effective from the Academic Year 2009-2010
M.Sc (STATISTICS) – COURSE STRUCTURE  
(With effect from 2009-10 onwards)

Objectives

The present course is intended to provide a platform for talented students to undergo higher studies in the subject as well as to train them to suit for the needs of the society. Apart from teaching core Statistics subjects, the students are also trained to handle real life problems through the practical classes. As a part of the course, the students are taught some programming languages and also exposed to various statistical softwares such as SPSS, SYSTAT, MINITAB.

Eligibility

B.Sc degree in Statistics or Mathematics with Statistics as a minor subject with a minimum 55% of marks.

Duration of the Course

The course duration shall normally be of two years duration spread over four semesters.

Medium

The medium of instruction shall be English.

Choice Based Credit System (CBCS)

The M.Sc Statistics program is offered through a unique CBCS. The salient features of the CBCS is that the program is offered through credit based courses. Subjects are divided into Hard Core and Soft Core. Hard Core subjects are compulsory. The students have the choice to select from among the list of soft core subjects. Soft core subjects are similar to elective subjects.

A student is expected to complete a minimum of 72 credits within four semesters. Students are assessed and awarded letter grades based on their performances in the given class.

Weightage of marks

The weightage of marks for continuous internal assessment (CAA) and end semester examinations shall be 40 and 60 respectively. A student is declared passed in a given subject when he/she secures a minimum of 40% in the end semester examination in that subject.

Continuous Internal Assessment

The weightage of 40 marks for internal continuous assessment component shall consist of the following:

a) Written test (best 2 of 3 class tests) = 30 marks
b) Written assignments/ Seminar presentations = 10 marks
   TOTAL = 40 marks
Aim of the Course:

The Degree of Master of Science in Statistics aims to train the students in the development and applications of Statistical techniques for analyzing data arising in the scientific investigation of problems in the various disciplines. It is also proposed to provide first hand practical experience in handling modern statistical softwares in the analysis of such data.

Eligibility for admission:

Candidates for admission to the first year of the M.Sc (Statistics) degree shall be required to have passed the B.Sc degree examination of any Indian University recognized by the University Grants Commission with Statistics as the main subject or Mathematics as the main subject with Statistics as one of the minor subject with a minimum of 55% marks in the main and allied subjects.

Duration of the Course:

The course shall be of two years duration spread over four semesters. The maximum duration to complete the course shall not be more than 8 semesters.

Eligibility for admission to Examination:

A candidate shall be permitted to appear for the examination in a subject of study only if he/she secures not less than 70% in the subject concerned.

Medium:

The medium of instructions shall be English.

Passing Minimum:

As per the Choice Based Credit System regulations of the Pondicherry University.
PONDICHERRY UNIVERSITY
CHOICE BASED CREDIT SYSTEM
M.Sc. STATISTICS SYLLABUS

Effective from the Academic Year 2009 – 2010

<table>
<thead>
<tr>
<th>COURSE CODE</th>
<th>TITLE OF THE COURSE</th>
<th>NATURE OF THE COURSE</th>
<th>NO. OF CREDITS</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 411</td>
<td>Mathematical Methods for Statistics</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 412</td>
<td>Probability Theory – I</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 413</td>
<td>Sampling Theory</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 414</td>
<td>Distribution Theory</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 415</td>
<td>Statistical Laboratory – I (Based on STAT 413)</td>
<td>Hard Core</td>
<td>2</td>
</tr>
<tr>
<td>STAT 421</td>
<td>Probability Theory - II</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 422</td>
<td>Statistical Quality Control and Operations Research</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 423</td>
<td>Stochastic Processes</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 424</td>
<td>Theory of Estimation</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 425</td>
<td>Statistical Laboratory - II (Based on STAT 422, 424)</td>
<td>Hard Core</td>
<td>2</td>
</tr>
<tr>
<td>STAT 531</td>
<td>Multivariate Statistical Analysis</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 532</td>
<td>Testing of Statistical Hypotheses</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 533</td>
<td>Linear Models and Regression Analysis</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Elective – I (Group A)</td>
<td>Soft Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 534</td>
<td>Statistical Laboratory – III (Based on STAT 531, 532, 533)</td>
<td>Hard Core</td>
<td>2</td>
</tr>
<tr>
<td>STAT 541</td>
<td>Design and Analysis of Experiments</td>
<td>Hard Core</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Elective – II (Group B)</td>
<td>Soft Core</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Elective – III (Group B)</td>
<td>Soft Core</td>
<td>4</td>
</tr>
<tr>
<td>STAT 542</td>
<td>Statistical Laboratory – IV (Based on STAT 541)</td>
<td>Hard Core</td>
<td>2</td>
</tr>
<tr>
<td>STAT 543</td>
<td>Project and Viva-Voce</td>
<td>Hard Core</td>
<td>4</td>
</tr>
</tbody>
</table>

**Electives:**

<table>
<thead>
<tr>
<th>Group A</th>
<th>Group B</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT 535 Time Series Analysis</td>
<td>STAT 544 Data Base Management Systems</td>
</tr>
<tr>
<td>STAT 536 Survival Analysis</td>
<td>STAT 545 Bayesian Inference</td>
</tr>
<tr>
<td>STAT 537 Data Mining Methods</td>
<td>STAT 546 Reliability</td>
</tr>
<tr>
<td>STAT 538 Econometrics</td>
<td>STAT 547 Biostatistics</td>
</tr>
<tr>
<td>STAT 539 Actuarial Statistics</td>
<td>STAT 548 Advanced Operations Research</td>
</tr>
</tbody>
</table>
STAT 411 - MATHEMATICAL METHODS FOR STATISTICS

Credits: 4

Unit I
Convergence of infinite numerical sequences and series (review only) – Absolute and conditional convergence – Sequences and series of functions – Pointwise and Uniform convergence – Tests for Uniform convergence – Properties of Uniform convergence

Unit II
Riemann - Stieltjes integral: Definition and properties – Integrals with step function and monotonic functions as integrators and their properties – Mean value theorem, Taylors theorem – Evaluation of Riemann - Stieltjes integral – Fundamental theorem

Unit III
Functions of several variables : Limits and continuity – Partial derivatives and Differentiability - Properties of differentiable functions – Higher order derivatives and differentials – Young and Schwartz theorems – Taylors theorem - Maxima and Minima – Extrema under constraints

Unit IV
Vector space and sub-space – Linear independence and orthogonality – Dimension and basis of a vector space – Orthonormal basis – Gram-Schmidt orthogonalization – Inner product space – Simultaneous linear equations (homogeneous and non-homogeneous)

Unit V
Matrices: Rank, inverse, trace and their properties – Characteristic roots and vectors – Idempotent and partitioned matrices – G-inverse and Moore Penrose inverse - their properties – Reduction of a matrix into diagonal, echelon, canonical and triangular forms – Quadratic forms – reductions of different types – Definite quadratic forms – Cochran’s theorem

Books for Study:


Books for Reference:

Unit I

Unit II

Unit III
Expectation and moments – Definition and properties – Moment Generating Function – Moment Inequalities – \( C_r, \) Holder, Jensen and basic inequalities.

Unit IV
Convergence: Modes of convergence – Convergence in probability – Convergence in distribution – Convergence in \( r^{th}\) mean – Almost sure convergence and their interrelationships.

Unit V

Books for Study:

Books for Reference:
Unit I

Unit II

Unit III
Varying probability Sampling: Cumulative total method and Lahiri’s method - Estimation in pps sampling with replacement, pps sampling without replacement; General selection procedures, Narian’s Scheme of sample selection and Sen-Midzona method - Ordered estimator; Des Raj Unordered estimators: Horvitz – Thompson estimator and Murthy’s estimator.

Unit IV
Cluster Sampling: Equal cluster sampling – Estimators of mean and variance, optimum cluster size, Unequal cluster sampling – Estimators of mean and variance, varying probability cluster sampling - Two stage sampling – variance of the estimated mean - Three stage sampling – variance of the estimated mean.

Unit V

Books for Study:
1. Cochran, W.G(1977): Sampling Techniques, Wiley Eastern Ltd., (Chapter 6 for Unit I, Chapter 7 for Unit II and Chapter 13 for Unit V)

Books for Reference:
STAT 414 – DISTRIBUTION THEORY CREDITS: 4

Unit I
Brief review of distribution theory, functions of random variables and their distributions - Laplace, Cauchy, lognormal, logarithmic series and power series distributions - Multinomial distribution

Unit II
Bivariate Binomial – Bivariate Poisson - Bivariate Exponential - Compound, truncated and mixture of distributions, concepts of convolution

Unit III
Multivariate normal distribution – marginal and conditional distributions – characteristic function and other properties - Sampling distributions: Non-central chi-square, t and F distributions and their properties

Unit IV
Distributions of quadratic forms under normality-independence of quadratic forms and linear form- Cochran’s theorem - Order statistics, their distributions and properties- Joint and marginal distributions of order statistics - Distribution of range and mid range - Extreme values and their asymptotic distributions (concepts only)

Unit V
Empirical distribution function and its properties, Kolmogorov Smirmov distributions, life distributions, exponential, Weibull and extreme value distributions, Mills ratio, distributions classified by hazard rate.

Books for Study:

Books for Reference:
I. Sampling Theory (20 marks)

2. PPSWR – Hurwitz Thompson estimator - Des Raj ordered estimator – Murthy’s unordered estimator – Midzuno scheme.
3. Linear and circular systematic sampling.
4. Stratified sampling – SRS, PPSWR, PPSWOR
5. Cluster sampling – of equal sizes.
6. Ratio, Regression and Difference estimation.

II. Computations based on SYSTAT software (40 marks)

1. Random number generation. (i) Binomial, (ii) Poisson, (iii) Normal
2. Descriptive Statistics
3. Computation of
   (i) Simple and multiple correlation coefficients
   (ii) Simple and multiple regression coefficients
4. Statistical Tests
   (i) Z test and Confidence Interval
   (ii) One sample t-test and Confidence Interval
   (iii) Two sample t-test and Confidence Interval
   (iv) Chi-square test for independence of attributes
   (v) Tests for homogeneity of proportions, variances and correlation coefficients
5. Non – Parametric Tests
   (i) Run test
   (ii) Sign test
   (iii) Wilcoxon one sample test
6. Design of Experiments
   (i) One-way ANOVA
   (ii) Two-way ANOVA
   (iii) Analysis of Covariance (ANCOVA)
Unit I
Characteristic function - Definition and properties – Inversion formula and its application – Characteristic Function and Moments – Bochner’s theorem (statement only) – Simple problems

Unit II
Weak and complete convergence of distribution functions – Helly’s First and Second limit theorems

Unit III
Law of large numbers: Kolmogrov Inequality – Weak law of large numbers (Khinchin’s and Kolmogrov) - Kolmogrov Strong law of large numbers – Glinvenko-Cantelli Theorem (statement only)

Unit IV
Central Limit Theorem : iid case – Lindeberg-Levy and Liapounov’s form - Lindeberg - Feller form – Infinitely Divisible distributions – definition, elementary properties and examples – canonical representation (without proof)

Unit V
Conditioning: Radon Nikodym theorem and derivative (without proof) - Conditional expectation – definition – properties (probability and expectation properties) - conditional probability and its applications – Definition and properties of Martingales and Sub-martingales – Martingale convergence theorem

Books for Study:

Books for Reference:
Unit I
Quality improvement: Meaning of quality and quality improvement – Different types of Quality costs and their management
Control charts: Review of $\bar{X}$, R, p, c, d charts - Modified control charts for mean – CUSUM chart – technique of V-mask – Weighted Moving average charts – Slopping control charts and group control charts

Unit II
Process Capability analysis: Meaning, Estimation technique for capability of a process – Capability Indices: $C_p$, capability ratio and $C_{pk}$ index – Estimation of natural tolerance limit of a process
Acceptance Sampling plans for attributes: Single, double, multiple and continuous sampling plans for attributes (Dodge type)

Unit III
Acceptance Sampling plans for variables: one sided and two sided specification – Standardized plans (ANSI/ANSQ Z1.9) and MIL-STD-414
Reliability: Concept, Definition and need - Concepts of Hazard rate, IFR and DFR - Relevance of exponential distribution in Reliability – Failure models – Taguchi’s approach in Quality and reliability - Six sigma approach

Unit IV
Review of LPP – Simplex and revised simplex methods - Duality in LPP – Dual Simplex method – Some important theorems on duality - Sensitivity Analysis – Variation in cost vector ‘c’ – Variation in the requirement vector ‘b’ – Addition and deletion of single variable – Addition and deletion of single constraint

Unit V
Replacement problem – Replacement of policy when value of money change does not change with time – Replacement of equipment that fails suddenly – Group replacement – Inventory – Various types of inventory – Costs associated with inventory models – Deterministic inventory models with and without shortages

Books for Study:

Books for Reference:
Unit I

Unit II
Markov chains continuous in time – General pure birth processes and Poisson process, birth and death processes, finite state continuous time Markov chains.

Unit III
Branching processes discrete in time – Generating functions relations – Mean and variance – Extinction probabilities – Concept of Age dependent Branching process

Unit IV
Renewal processes – Definition and examples – key renewal theorem – Study of residual life time process –

Unit V
Stationary process – weakly and strongly stationary process – Moving average and Autoregressive process and its covariance functions - Brownian Motion process – Joint probabilities for Brownian motion process – Brownian motion as a limit of random walk

Books for Study:

Books for Reference:
Unit I
Parametric point estimation – properties of estimates – Consistency and its different forms
Sufficient condition for consistency- Unbiasedness – sufficient statistics – Factorization theorem
- Distributions admitting sufficient statistic, procedure for finding minimal sufficient statistic.

Unit II
The information measure – Cramer - Rao (CR) inequality - Kiefer – Chapman - Robbins (KCR)
inequality - Bhattacharya inequality - minimum variance bound estimator- Invariant
(equivariant) estimators (concepts only)

Unit III
Uniformly minimum variance unbiased estimators (UMVUE)- condition for the existence of
UMVUE- Completeness and Bounded completeness- Relation between complete statistic and

Unit IV
Methods of estimation – method of moments- method of maximum likelihood and its
properties-Large sample properties of MLE - Method of minimum chi-square and its properties
– Methods of least squares – Optimum properties of least square estimates in linear model.

Unit V
Interval estimation – Pivotal method of construction - shortest confidence intervals (minimum
average width) - Constructions of shortest confidence intervals.
Notion of Bayes estimation – Concepts of prior, posterior and conjugate priors. Simple
problems involving quadratic error loss function - Elementary notions of minimax estimation -
Simple illustrations.

Books for Study:

Books for Reference:
   Graw-Hill Book Company.
I    Estimation

1. MLE and Standard error of ML estimators.
2. MLE through the method of successive approximation.
3. MLE for truncated distribution.
4. Method of Moments
5. Method of Minimum Chi-square
6. Method of Least square
7. Interval estimation: Confidence interval for mean, difference of means, variance and ratio of variances.

II    Statistical Quality Control

Control charts:

   a. CUSUM chart
   b. Modified Control chart
   c. Moving Average Control chart
   d. Exponentially Weighted Moving Average chart
   e. Sloping Control Chart

Acceptance sampling:

   1. Single sampling plans and double sampling plans (for attributes)
   2. Variable Sampling plans (Single and double specifications)
   3. Standard plans
Unit I
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 and their sampling distributions in the null case

Unit II
Hotelling $T^2$ distribution and its applications - derivation of generalized $T^2$ statistic and its distribution - Uses of $T^2$ statistic - optimum properties of $T^2$ statistic - Mahalanobis $D^2$ statistic and its distribution - relation between $T^2$ and $D^2$ – Test based on $T^2$ statistic

Unit III
Generalized variance - Wishart distribution (statement only) – Properties of Wishart distribution - Test for covariance matrix – Test for equality of covariance matrices – Test for independence of sets of variables

Unit IV
Classification problems - Classification into one of two populations (known and unknown dispersion matrix) - Classification into one of several populations – Linear discriminant function – Multivariate analysis of variance (MANOVA) – One-Way classification.

Unit V

Books for Study:


Books for Reference:

STAT 532 – TESTING OF STATISTICAL HYPOTHESES

Unit I
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, Generalization of fundamental lemma and its applications

Unit II
Unbiasedness for hypothesis testing, Uniformly most powerful unbiased tests, Unbiased tests for one parameter exponential family, Similar regions and complete sufficient statistics, Tests with Neyman structure, Uniformly most powerful similar tests, Locally most powerful tests.

Unit III
Invariant tests, maximal invariants, Uniformly most powerful invariant tests, Consistent tests, Likelihood ratio test, its properties and its asymptotic distribution, Applications of the LR method.

Unit IV
Non-parametric tests: Goodness of fit test : Chi-square and Kolmogorov Smirnov test - k-sample problem: Extension of Median test, Kruskal Wallis test, Friedman test – Notion of ARE.

Unit V
Sequential methods: Sequential unbiased estimation – Application to Normal distribution - Sequential test - Basic Structure of Sequential tests – Sequential Probability Ratio Test (SPRT) and its applications – Determination of the boundary constants – Operating Characteristic and expected sample size of SPRT - Optimum properties of SPRT.

Books for Study:

Books for Reference:
Unit I
Full rank linear model – least square estimators of the parameters and their properties – model in centered form – Estimators under normality assumption and their properties – Coefficient of determination – Generalized least squares – misspecification of the error structure and the model.

Unit II
Test for overall regression and for a subset of the parameters – test in terms of $R^2$ – General Linear Hypothesis testing – special cases – confidence region for the parameters and the mean – prediction intervals – likelihood ratio tests for the parameters – study of the residual outliers and influential observations

Unit III
Selection of input variables and model selection – Methods of obtaining the best fit - Stepwise regression – Forward selection and backward elimination

Unit IV

Unit V
Robust regression – Linear absolute deviation regression – M estimators – Robust regression with rank residuals – Resampling procedures for regression models – methods and its properties (without proof) - Jackknife techniques and least squares approach based on M-estimators.

Books for Study:

Books for Reference:
I  Multivariate analysis

1. Maximum likelihood estimators – Mean vector and dispersion matrix
2. Test for Mean Vectors (\( \Sigma \) is known)
3. Hotelling’s \( T^2 \) statistic
4. MANOVA (one way)
5. Test for covariance matrix
6. Discriminant analysis
7. Principal component analysis
8. Canonical correlation and canonical variables

II  Testing of 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.
3. Non parametric tests: Kolmogorov Smirnov test, Median test for k-sample problem, Kruskal Wallis test and Friedman’s test

III  Linear Models and Regression Analysis

1. Multiple linear regression
2. Linear Models
Unit I
Notion of design matrix- general analysis of design models (Inter and Intra Block analysis) – C Matrix and its properties – EMS and its uses, Algorithm for calculating EMS - Two way elimination of heterogeneity – Orthogonality – Connectedness and resolvability

Unit II
Principles of scientific experimentation – Pen and Plot techniques - Basic Design: CRD, RBD and LSD, Analysis of RBD (with one observation per cell, More then one but equal number of observations per cell) – Multiple comparison test: LSD, SNK, DMR, Tukey tests.

Unit III
Factorial experiments: 2^n and 3^n experiments and their analysis – Complete and Partial Confounding - Fractional Replication in Factorial Experiments – Split plot and strip plot design and their analysis.

Unit IV
BIBD - Types of BIBD - Simple construction methods - Concept of connectedness and balancing – Intra Block analysis of BIBD – Recovery of InterBlock information – Partially Balanced Incomplete Block Design with two associate classes – intra block analysis only.

Unit V
Youden square and lattice design and their analysis – Mutually Orthogonal Latin Squares – Analysis of Covariance with one concomitant variable – Analysis for CRD and RBD only –Response Surface Designs – Method of Steepest Ascent

Books for Study:
2. Douglas C. Montgomery (1984) : Design and Analysis of Experiments, John Wiley and Sons, (Chapter 16 for Parts of Unit IV and Unit V)

Books for Reference:
I. Design of Experiments (20 marks)

1. Multiple Comparison tests
2. $2^4, 3^2, 3^3$ factorial experiment
3. Complete and partial confounding in $2^4, 3^2, 3^3$ factorial experiments
4. Fractional factorial
5. Split plot design
6. Strip plot design
7. BIBD
8. PBIBD with two associate classes
9. Youden Square Design
10. Analysis of Covariance – CRD and RBD

II. Computations based on SPSS software (40 marks)

1. Statistical Quality Control and Regression Analysis
   (i) Simple Linear Regression
   (ii) Multiple regression
   (iii) $\bar{X}$ and R charts
   (iv) p and np chart
   (v) c chart
2. Multivariate Analysis
   (i) Tests based on Hotelling $T^2$
   (ii) MANOVA
   (iii) Discriminant Analysis
   (iv) Principal Component Analysis
   (v) Factor Analysis
3. Parametric and Non-parametric tests
   (i) Tests based on mean
   (ii) Chi-square test for independence of attributes
   (iii) Run test
   (iv) Sign test and Wilcoxon signed rank test
   (v) Kruskal Wallis and Friedman test
   (vi) Kolmogrov-Smirnov test
   (vii) Median test
4. Design of Experiments
   (i) One-way and two-way ANOVA
   (ii) Factorial experiments
   (iii) ANOCOVA
A project work is compulsory and shall be offered in semester X. It will have 4 credits.
2. A project work may be taken individually or by a group of two students.
3. Project work shall be supervised by a faculty member assigned by the Head of the Department in the beginning of the semester.
4. The project work should be selected in such a way that there is enough scope to apply and demonstrate the statistical techniques learnt in the course.
5. At the end of the semester, before the last working day, a report on the work done should be submitted (two copies). If a team of two students jointly do a project work then they must submit individual report separately (not copy of the same report).
6. The project report shall clearly state the selected problem, the statistical methodologies employed for data collection and analysis and the conclusions arrived at. Details of previous studies in the area and related references should also be given.
7. The project work will be assessed for a maximum of 100 marks. Each student will give a seminar before the end of the semester on their project work which will be evaluated internally for a maximum of 30 marks. There will be viva-voce examination for a maximum of 10 marks by an internal and an external examiner. The project report will be valued by the same external and internal examiner for a maximum of 60 marks.
STAT 535 - TIME SERIES ANALYSIS

CREDITS: 4

Unit I

Unit II
General linear stationary models – stationarity and invertability – Autoregressive and moving average processes and their autocorrelation functions – mixed autoregressive moving average processes
Linear nonstationary models – ARIMA processes and their explicit forms – Integrated MA processes

Unit III
Forecasting: MMSE forecasts and their properties – Forecasts and their updating – Forecast of functions and forecast weights – examples

Unit IV
Model Identification – Identification techniques – Initial estimates for different processes – MA, AR, ARMA – choice between stationary and nonstationary models – model multiplicity

Unit V

Books for Study:

Books for Reference:
Unit I
Concepts of time, Order and random Censoring, likelihood in these cases. Life distributions – Exponential, Gamma, Weibull, Lognormal, Pareto, Linear Failure rate. Parametric inference (Point estimation, Scores, MLE)

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

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

Unit IV
Two sample problem- Gehan test, Log rank test. Mantel–Haenszel test, Tarone–Ware tests. Semi-parametric regression for failure rate – Cox’s proportional hazards model with one and several convariates. Rank test for the regression coefficients.

Unit V
Competing risks model, parametric and non-parametric inference for this model. Multiple decrement life table.

Books for Study:

Books for Reference:
STAT 537 – DATA MINING METHODS  CREDITS: 4

Unit I
Introduction to Data Mining – Different data mining tools – Different data types – Distance metrics – Information measures- Entropy, Gini Index

Unit II
Basic Clustering: Hierarchical clustering methods – partitioning clustering methods – k-means, k-medoids, fuzzy k-means – clustering validity measures – Purity, precision, recall, silhouette

Unit III
Advanced Clustering: Density Based Clustering – DBSCAN – DENCLUE, Genetic clustering – Model based clustering

Unit IV
Decision trees- rule deduction using decision trees – k-nn classification, naïve Bayesian classification – vector support machines and simple applications

Unit V
Association analysis – Market Basket Analysis – A priori algorithm – Pruning and Candidate generation – Rule Mining

Books for Study:
Tan, Steinbach and Kumar :Introduction to Data Mining , Pearson Education
(Portions of Chapters 2,4,5,6,8,9)

Books for Reference:
Rajan Chattamvelli (2009): Data Mining Methods, Narosa Publishing House, New Delhi
Unit I
Aitken’s generalized least squares (GLS) estimator – Heteroscedasticity – Auto-correlation – Test of auto-correlation.

Unit II

Unit III
Linear regression with stochastic regressors – Errors in variable models and instrumental variable estimation – Independent stochastic linear regression – Auto regressive linear regression – Contemporaneously uncorrelated linear regression – Distributed lag models.

Unit IV

Unit V

Books for Study:

Books for Reference:
Unit-I
Life Tables and Survival functions: Probability for the age at death and time until death for a person, curtate future life time and force of mortality- Relation between Life Table function and Survival function

Unit-II
Life Insurance: Principals of Compound Interest- Nominal and Effective rates of interest and Discount-Force of Interest and Discount- Accumulation factor and continuous compounding-Insurance payable at the time of death-Level-Benefit Insurance-Endowment insurance-Deferred Insurance- Varying benefit Insurance-Insurance payable at the end of the year of death-recursions and commutation functions.
Life-Annuities: Single payment –continuous and discrete Life annuities. Life annuities with monthly payment-varying annuities-apportionable annuities- Due and complete annuities immediate.

Unit-III
Benefit premiums: Fully continuous and fully discrete premiums, true monthly payment premiums, apportionable premiums and accumulation benefits
Benefit reserves: Fully continuous benefit reserves, benefit reserves on a semi-continuous basis, benefit reserves based on monthly benefit premiums, benefit reserves on an apportionable or discounted continuous basis.
Analysis of Benefit Reserves: benefit reserves for general insurance, reserves at fractional durations, allocation of loss to policy years, recursive formulae and differential equations for reserves and commutation functions.

Unit-IV
Multiple Life Functions: Joint distribution of future life times, the joint life status, the last-survivor status, dependent lifetime models, common shock, capulas, insurance and annuity benefits, survival status, special two life annuities, reversionary annuities.
Gompertz and Macham laws, uniform distribution, simple contingent functions and their evaluation.
Multiple decrement Models: Deterministic and random survivor groups, associated single decrement tables, central rates of multiple decrement, net single premiums and their evaluation. Construction of multiple decrement table.
Applications of Multiple Decrement Theory: Actuarial present value and their numerical evaluation, withdrawal benefit patterns that can be ignored in evaluating premiums and reserves, valuation of pension plans, demographic assumptions, projecting payment and contribution rates, defined-benefit plans, defined-contribution plans, disability benefits with individual life insurance, disability income benefits waiver premium benefits.

UNIT-V
Collective risk Model for a Single Period: Distribution of aggregated plans, selection of basic distributions, distribution of individual claim amount, properties of certain compound distributions and approximations to the distribution of aggregate claims.
Collective Risk models over an extended period: Discrete and Continuous models, ruin probabilities and the claim amount distribution, first surplus below the initial level, maximum aggregate loss.

Books for Study:

Books for Reference:
3. Ammann,M(2002): Credit Risk Valuation, Springer Verlag
UNIT-I
Introduction to DB Systems: Overview-historical perspective-file system versus DBMS-Advantages of DBMS-Storing Data in DBMS-Queries in DBMS-Structure of a DBMS-Use of DBMS.
ER Model: Overview-Entities -Attributes and Entity Sets-Relationships and Relationship Sets-Features of ER Model-

UNIT-II
Relational Model: Introduction-Integrity Constraints over relations -. Enforcing Integrity Constraints-Querying Relational Data-Logical Database Design: ER to Relational - View: Introduction- Altering/Destroying tables and views.

UNIT-III

UNIT-IV

UNIT-V
Design of ER Model, Design of Database, Normalization and Queries for the Case Studies: Student Information System-Library Information System-Railway Reservation System.

Books for study:

Books for Reference:
STAT 545 - BAYESIAN INFERENCE  
CREDITS: 4

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

Unit II
Natural Conjugate family of priors for a model. Hyper parameters of a prior from conjugate family. Conjugate families for (i) exponential family models. (ii) models admitting sufficient statistics of fixed dimension. Enlarging the natural conjugate family by (i) enlarging hyper parameter space (ii) mixtures from conjugate family, choosing an appropriate member of conjugate prior family. Non informative, improper and invariant priors. Jeffrey’s invariant prior.

Unit III
Bayesian point estimation: as a prediction problem from posterior distribution. Bayes estimators for (i) absolute error loss (ii) squared error loss (iii) 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 and its comparison with the interpretation of the confidence coefficient for a classical confidence interval.

Unit IV
Bayesian Testing of Hypothesis: Specification of the appropriate form of the prior distribution for a Bayesian testing of hypothesis problem - Prior odds, Posterior odds, Bayes factor for various types of testing hypothesis problems depending upon whether the null hypothesis and the alternative hypothesis are simple or composite. Specification of the Bayes tests in the above cases. Discussion of Lindley’s paradox for testing a point hypothesis for normal mean against the two sided alternative hypothesis.

Unit V
Bayesian prediction problem - Large sample approximations for the posterior distribution - Bayesian calculations for non conjugate priors: (i) Importance sampling, (ii) Obtaining a large sample of parameter values from the posterior distribution using Acceptance – Rejection methods, Markov Chain Monte Carlo methods and other computer simulation methods.

Books for Study:

3. Leonard T. and Hsu, J.S.J. Bayesian Methods, Cambridge University Press.
4. Bansal, Bayesian Inference, Narosa Publications

Books for Reference:

Unit I
Reliability concepts and measures; components and systems; coherent systems; reliability of coherent systems; cuts and paths; modular decomposition; bounds on system reliability; structural and reliability importance of components.

Unit II
Life distributions; reliability function; hazard rate; common life distributions – exponential, Weibull, Gamma, etc. Estimation of parameters and tests in these models.

Unit III
Notions of ageing: IFR, IFRA, NBU, DMRL, and NBUE Classes and their duals and implications.

Unit IV
Common bivariate exponential distributions and their properties.

Unit V
Reliability growth models; probability plotting techniques; Hollander-Prochan and Deshpande tests for exponentiality. Basic ideas of accelerated life testing.

Books for Study:

Books for Reference:
Unit I

Unit II

Unit III

Unit IV

Unit V

Books for Study:

Books for Reference:
Unit I

Unit II
Inventory models with one or two price breaks - Multi item deterministic problem – Constraints on storage and investment – Probabilistic Inventory models – Periodic Review systems – Fixed order quantity system

Unit III
Non-linear programming problem – Kuhn Tucker conditions – Quadratic programming problem (QPP) - Wolfe’s and Beale’s algorithms for solving QPP – Geometric programming

Unit IV
Dynamic programming problem (DPP) - Bellman’s principle of optimality - General formulation - computation methods and application of DP - Solving LPP through DP approach - Convex programming

Unit V
Queuing theory – Basic characteristics of queuing models – Arrival and service distribution – steady state solution of M/M/1 and M/M/C models with associated distribution of queue length and waiting time - M/G/1 queue-steady results using embedded Markov chain Methods - Pollaczek Khinchin result.

Books for Study:

Books for Reference: