# SRI DHARMASTHALA MANJUNATHESHWARA COLLEGE, UJIRE–574240

## (AUTONOMOUS)

(Re–Accredited by NAAC at 'A' Grade with CGPA 3.61 out of 4)



# DEPARTMENT OF PG STUDIES IN STATISTICS

## SYLLABUS

(With effect from 2020–2021)

# SRI DHARMASTHALA MANJUNATHESHWARA COLLEGE, UJIRE–574240

## (AUTONOMOUS)

(Re-Accredited by NAAC at 'A' Grade with CGPA 3.61 out of 4)



# DEPARTMENT OF PG STUDIES IN STATISTICS

*Syllabus of*
*Master's Degree in Statistics*

(CREDIT BASED SEMESTER SCHEME)
2020–2021 onwards

Approved by the BOS meeting held on _____
Academic Council meeting held on _____

## Course Objective:

- To prepare the younger generation to tackle the emerging problems through applications of statistics.
- To develop the department as a hub for teaching, research and extension activities in the field of modern statistical theory and methods.

## List of Hardcore papers:

i. Real Analysis
ii. Probability and Distributions – I
iii. Theory of Sampling
iv. Probability and Distributions – II
v. Design and Analysis of Experiments
vi. Theory of Estimation
vii. Theory of Testing of Hypothesis
viii. Regression Analysis
ix. Stochastic Processes
x. Time Series Analysis
xi. Reliability and Survival Analysis

## List of Softcore papers:

i. Linear Algebra
ii. Data Management and Statistical Computing with Python
iii. Multivariate Analysis
iv. Statistical Modelling
v. Big Data Analytics
vi. Artificial Intelligence
vii. Elements of Statistical Computing
viii. Survival Analysis
ix. Stochastic Finance

x.     Data Mining

xi.    Bayesian Inference

xii.   Statistical Methods for Reliability

xiii.  Nonparametric Inference

xiv.   Actuarial Methods

xv.    Pattern Recognition and Image Processing

xvi.   Operations Research

# Course Pattern for M.Sc. Statistics Programme

## I Semester:

| Course Code | Title of the Paper | Hrs/Week | Credits |
|---|---|---|---|
| STH 411 | Real Analysis | 4 | 4 |
| STS 412 | Linear Algebra | 3 | 3 |
| STH 413 | Probability and Distributions – I | 4 | 4 |
| STH 414 | Theory of Sampling | 4 | 4 |
| STP 415 | Practical–I (Based on R programming and STH 414) | 8 | 4 |
| STP 416 | Practical–II (Based on Programming in Python) | 8 | 4 |
| | Mini project** | | |
| | | Total | 23 |

**Mini project will be incorporated as 'Skill Component'.

## II Semester:

| Course Code | Title of the Paper | Hrs/Week | Credits |
|---|---|---|---|
| STE 421 | Introductory Statistics and Data Analysis (Open Elective Course) | 3 | 3 |
| STH 422 | Probability and Distributions – II | 4 | 4 |
| STH 423 | Design and Analysis of Experiments | 4 | 4 |
| STH 424 | Theory of Estimation | 4 | 4 |
| STS 425 | Data Management and Statistical Computing with Python | 3 | 3 |
| STP 426 | Practical–III (Based on STH 423 and STH 424 using R) | 6 | 3 |
| STP 427 | Practical–IV (Based on STS 425) | 6 | 3 |
| | Mini project** | | |
| | | Total | 21+3* |

*This credit is not included for CGPA.

**Mini project will be incorporated as 'Skill Component'.

SDM College (Autonomous) Ujire

## III Semester:

| Course Code | Title of the Paper | Hrs/Week | Credits |
|---|---|---|---|
| STE 531 | Inferential Statistics and Data Analysis (Open Elective Course) | 3 | 3 |
| STH 532 | Theory of Testing of Hypothesis | 4 | 4 |
| STH 533 | Regression Analysis | 4 | 4 |
| STH 534 | Stochastic Processes | 4 | 4 |
| STS 535 | Multivariate Analysis | 3 | 3 |
| STP 536 | Practical-V (Based on STH 532 and STH 533 using R) | 6 | 3 |
| STP 537 | Practical-VI (Based on Machine Learning with Python) | 6 | 3 |
| | Mini project** | | |
| | | Total | 21+3* |

*This credit is not included for CGPA.

**Mini project will be incorporated as 'Skill Component'.

## IV Semester:

| Course Code | Title of the Paper | Hrs/Week | Credits |
|---|---|---|---|
| STH 541 | Time Series Analysis | 4 | 4 |
| STH 542 | Reliability and Survival Analysis | 4 | 4 |
| STS 543 | Statistical Modelling | 3 | 3 |
| STS 544 | Big Data Analytics | 3 | 3 |
| STP 545 | Practical-VII (Based on STH 541, STH 542 and STS 543 using R) | 6 | 3 |
| STP 546 | Project Work** | 8 | 4 |
| | | Total | 21 |

**Project Work will involve 'Skill Component'.

# Scheme of Internal Assessment Evaluation

The scheme of evaluation for internal assessment marks shall be as follows:

| | | |
|---|---|---|
| i. | Two Internal Tests | 20 marks |
| ii. | Seminar/Assignments/Classroom Activities etc. | 10 marks |
| | **Total:** | **30 marks** |

# Question Paper Pattern

The pattern of question paper in theory examinations shall be as follows:

  i.   There shall be totally 8 questions in which Q. No. 1 is compulsory. Students have to answer any 4 questions from the remaining 7 questions.

  ii.  Q. No. 1 will contain 8 questions of short answer type, each question carrying 3 marks. Students will have to answer any 6 questions. Thus Q. No. 1 carries 18 marks.

  iii. Q. No. 2 to Q. No. 8 will be of long answer type, each question carrying 13 marks.

The distribution of marks will be as follows:

Q. No. 1            $3 \times 6 = 18$

Any 4 questions out of
remaining 7 questions     $13 \times 4 = 52$

Total = 70

(Prof. Shanthiprakash)
Chairman
PG B. O. S. in Statistics

SDM College (Autonomous) Ujire

# Proposed Syllabus

## *I Semester*

## STH 411 - Real Analysis (4 Credits)

**Rationale/Learning Objectives:**

This course provides necessary mathematical foundations required for understanding different theoretical aspects in statistics.

- <u>Unit 1</u>: Elements of set theory, sets in Euclidean space of $k$-dimensional $R^k$ rectangles. Metric spaces, neighbourhood, interior point and limit point, open and closed sets, Bolzano-Weierstrass theorem in $R^2$, compact set, real-valued functions, Heine-Borel theorem (Statement only), continuity and uniform continuity. (13hrs)

- <u>Unit 2</u>: Sequences and Series of real numbers - Cauchy sequence, convergence of bounded monotone sequence. Limit superior, limit inferior and limit properties. Series of positive terms - tests for convergence, divergence. Series of arbitrary terms - absolute and conditional convergence. (13hrs)

- <u>Unit 3</u>: Sequences of functions - uniform convergence and point wise convergence, series of functions - uniform convergence, Weierstrass' $M$ test. Power series and radius of convergence. Reimann-Stieltjes integration continuous integrand and monotonic/differentiable integrator. (13hrs)

- <u>Unit 4</u>: Functions of two variables - partial and directional derivatives. Maxima and minima of functions, maxima-minima under constraints (Lagrange's multipliers). Improper integrals. (13hrs)

## Books for Reference:

1. Apostol, T. M. (1985). *Mathematical Analysis*. Narosa India Ltd.

2. Bartle, R. G. (1975). *The Elements of Real Analysis* (2nd ed.). Wiley.

3. Courant, R. and John, F. (1965). *Introduction to Calculus and Analysis*. Wiley.

4. Goldberg, R. R. (1970). *Methods of Real Analysis*. Oxford Publishing Co.

5. Khuri, A. T. (1993). *Advanced Calculus with Applications in Statistics*. John Wiley.

6. Rudin, W. (1976). *Principles of Mathematical Analysis*. McGraw Hill.

## STS 412 - Linear Algebra (3 Credits)

**Rationale/Learning Objectives:**

This course provides necessary mathematical foundations on matrix algebra and vector geometry for better understanding of linear models and multivariate analysis.

- <u>Unit 1</u>: Fields, vector spaces, subspaces, linear dependence and independence, basis and dimension of a vector space, finite dimensional vector spaces, completion theorem. Examples of vector spaces over real fields. Vector spaces with an inner product, Gram-Schmidt orthogonalization process, orthonormal basis. (10hrs)

- <u>Unit 2</u>: Row and column spaces of a matrix. Rank and inverse of a matrix, properties of inverse. Rank of a product of matrices, partitioned submatrices, rank factorization of a matrix, rank of a sum, inverse of a partitioned matrix. General linear system of equations, generalized inverse, Moore-Penrose inverse, idempotent matrices. Solutions of matrix equations. (10hrs)

- <u>Unit 3</u>: Characteristic roots and vectors, Cayley-Hamilton theorem, minimal polynomial, similar matrices. Algebraic and geometric multiplicity of characteristic roots, spectral decomposition of a real symmetric matrix, reduction of a pair of real symmetric matrices, singular value decomposition. (10hrs)

- <u>Unit 4</u>: Real quadratic forms, reduction and classification of quadratic forms, index and signature. Extrema of quadratic forms. Vector and matrix differentiation. (10hrs)

## Books for Reference:

1. Biswas, S. (1984). *Topics in Algebra of Matrices*. Academic Publications.

2. Hadley, G. (1987). *Linear Algebra*. Narosa.

3. Graybill, F. A. (1983). *Matrices with Applications in Statistics*.

4. Rao, A. R. and Bhimasankaran, P. (1992). *Linear Algebra*. Tata McGraw Hill.

5. Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). Wiley.

6. Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications*. Wiley.

7. Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. Wiley.

## Practicals based on R Programming

1. Introduction to R programming and basics of R.

2. Algebra of matrices.

3. Vector space and linear independence of vectors.

4. Diagonalization of a matrix.

5. Solution to system of linear equations.

6. Transformation and Gram-Schmidt orthogonalization.

7. Eigen values and eigen vectors.

8. Spectral decompositions.

9. Sketching of p.m.fs, p.d.fs and d.f.

10. Model sampling from univariate and bivariate distributions.

## STH 413 - Probability and Distributions - I (4 Credits)

**Rationale/Learning Objectives:**

This course provides necessary theoretical foundations on the developments of statistical concepts and develops problem solving skills.

- <u>Unit 1</u>: Algebra of sets, sequence of sets and limits, fields and sigma-fields, minimal sigma-field. Events, sample space. Probability measure, probability space, property of probability measure, properties related to sequences of events, independent events, conditional Probability. (13hrs)

- <u>Unit 2</u>: Measurable functions, random variables, probability induced by a random variable. Definition of simple random variables. Integration of measurable functions with respect to measures. Expectation, properties of expectation, moments, inequalities. (16hrs)

- <u>Unit 3</u>: Standard discrete and continuous univariate distributions and their properties, probability generating function and moment generating function. Bivariate normal and multinomial distributions. Transformation techniques. Distribution of functions of random variables. (13hrs)

- <u>Unit 4</u>: Order Statistics - their distributions and properties, joint and marginal distributions. Distribution of range and median. Truncated and mixture distributions. (10hrs)

## Books for Reference:

1. Ash, R. B. and Catherine Doleans-Dade (2000). *Probability and Measure Theory*. Academic Press.

2. Bhat, B. R. (1999). *Modern Probability Theory* (3rd ed.). New Age International Publishers.

3. Johnson, S. and Kotz (1972). *Distributions in Statistics*. Vols. I, II and III, Houghton and Miffin.

4. Mukhopadhyaya, P. (1996). *Mathematical Statistics*. Calcutta Publishing House.

5. Pitman, J. (1993). *Probability*. Narosa.

6. Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). Wiley Eastern.

7. Rohatgi, V. K. and Saleh, A. K. Md. E. (2015). *An Introduction to Probability Theory and Mathematical Statistics*. Wiley Eastern.

8. Laha, R. G. and Rohatgi, V. K. (1979). *Probability Theory*. Wiley Eastern.

9. Ross, S. M. (1993). *First Course in Probability*. Academic Press.

10. Billingsley, P. (1986). *Probability and Measure*. John Wiley and Sons.

## STH 414 - Theory of Sampling (4 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical knowledge on various sampling techniques used for designing and selecting a sample from a population.

- Unit 1: Basic Concepts - sampling design, sampling scheme, sampling strategy, interpenetrating subsampling, concept of non-random sampling. Probability proportional to size with replacement (PPSWR) sampling - selection of PPSWR sample, estimation of population mean, total and their sampling variances. Hansen-Hurwitz strategy, estimation of sampling variance. Comparison with SRSWR, estimation of gain due to PPSWR sampling.          (13hrs)

- Unit 2: Varying probability without replacement (PPSWOR) sampling - some properties of sampling design. Horwitz-Thomson estimator, sampling variance of population total and its unbiased estimator. Sen-Midzuno sampling scheme, Des-Raj's ordered estimator (general case), Murthy's unordering principle (sample of size two).          (13hrs)

- Unit 3: Single stage cluster sampling - concepts, estimation of efficiency of cluster sampling, clusters of varying sizes. Two-stage sampling - notions, estimation of population total and its variance, efficiency of two-stage sampling relative to cluster and uni-stage sampling. (13hrs)

- Unit 4: Ratio and regression estimators based on SRSWOR, method of sampling, bias and mean square errors, comparison with mean per unit estimator. Two-phase sampling - notion, double sampling for ratio estimation, double sampling for regression estimation. Randomized response techniques - Warner's model, related and unrelated questionnaire methods, non-sampling errors. Official Statistics for National Development - NSO, CSO, MOSPI, Human Development Index. Measuring inequality in income - Lorenz Curve, Gini coefficient. (13hrs)

## Books for Reference:

1. Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). Wiley.

2. Des Raj and Chandok (1998). *Sampling Theory*. Narosa Publication.

3. Mukhopadhyay, P. (1998). *Theory and Methods of survey Sampling*. Prentice Hall of India.

4. Murthy, M. N. (1977). *Sampling Theory and Methods*. Calcutta: Statistical Publishing Society.

5. Sampath, S. (2001). *Sampling Theory and Methods*. Narosa Publishers.

6. Sen, A. (1997). *Poverty and Inequality*.

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

8. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Ashok (1984). *Sampling Theory of Surveys with Applications*. ICAR Publication.

9. Banett, V. (2002). *Sample Survey: Methods and Principles*. Arnold Publishers.

## Practicals based on Theory of Sampling

1. Determination of sample size.

2. Probability proportional to size with replacement (PPSWR) - I

3. Probability proportional to size with replacement (PPSWR) - II

4. Probability proportional to size without replacement (PPSWOR) - I

5. Probability proportional to size without replacement (PPSWOR) - II

6. Des Raj's ordered estimator.

7. Des Raj's ordered estimator and Murthy's unordered estimator.

8. Stratified and systematic sampling.

9. Cluster sampling with clusters of equal size.

10. Cluster sampling with clusters of unequal size.

11. Two stage sampling.

12. Multistage sampling.

13. Ratio method of estimation.

14. Ratio and regression method of estimation.

## Practicals based on Programming in Python

1. Write a program to perform addition, subtraction, multiplication, division and modulo operations on two integer numbers (read the input from keyboard).

2. (a) Write a program to read three integer numbers from keyboard and find the largest among three numbers.

   (b) Write a program to determine whether the entered character is a vowel or not.

3. Write a python program to perform addition, subtraction, multiplication, division on two matrices (read the input from keyboard).

4. Write a python program to perform arithmetic operation in excel.

5. Write a program to convert a square matrix into a lower and upper triangular matrix.

6. (a) Write a program to identify whether an input matrix is symmetric or not.

   (b) Write a program to identify whether an input matrix is binary or not.

7. Write a program to count the number of digits, uppercase characters, lowercase characters and special characters in a given string.

8. Write a python program to find mean, median, mode of a entered list of numbers.

9. Write a program that has dictionary of names of students and a list of their marks in four subjects. Create another dictionary from this dictionary that has name of the student and their total marks. Find out topper and his/her score.

10. Write a program that has a class student that stores roll number, name and marks (in three subjects) of the students. Display the information (roll number, name, and total marks) stored about the student.

*II Semester*

# STE 421 - Introductory Statistics and Data Analysis (3 Credits)

**Rationale/Learning Objectives:**

This course provides knowledge on fundamentals involved in descriptive statistics, probability and sampling, which are widely used in data analysis. (**Open Elective Course**)

- Unit 1: Statistics - introduction, meaning, definition and scope of the subject as a science of decision making against uncertainty. Data types, methods of collection, presentation in the form of tables and graphs. Descriptive Statistics - measures of central tendency, positional averages, measures of dispersion, skewness and kurtosis. Methods of summarizing categorical data - univariate and bivariate contingency tables. Box plots - construction and interpretations. Exploratory data analysis using descriptive measures and graphical tools. (13hrs)

- Unit 2: The concept of random experiment, simple events, sample space, types of events, probability of an event, rules of probability, conditional probability, Baye's rule, exercises on computation of probabilities using these rules to fix the ideas. The concept of random variables - discrete and continuous type, Binomial, Poisson and Normal distributions - their use in practical applications, computing probabilities using these distributions. (13hrs)

- Unit 3: Sampling methods - population and sample, parameter and statistic, concept of a random sample, simple random sampling, stratified sampling, systematic, sampling, sample size determination. The concept of sampling distribution of a statistic and standard error. (14hrs)

## Books for Reference:

1. Campbell, R. C. (1974). *Statistics for Biologists.* Cambridge University Press.

2. Chatfield, C. (1981). *Statistics for Technology.* Chapman and Hall.

3. Frank, H. and Athoen, S. C. (1997). *Statistics: Concepts & Applications.* Cambridge University Press.

4. Medhi, J. (1992). *Statistical Methods: An Introductory Text.* Wiley Eastern Limited.

5. Ross, S. M. (2017). *Introductory Statistics.* Academic Press.

6. Rice, J. A. (2006). *Mathematical Statistics and Data Analysis.* Singapore: Thomson-Duxbury.

# STH 422 - Probability and Distributions - II (4 Credits)

**Rationale/Learning Objectives:**

This course provides adequate conceptual basis for the asymptotic theory and characterization properties of the distributions.

- Unit 1: Measure, probability measure, properties of a measure and probability, Carathedory extension theorem (statement only). Lebesgue and Lebesgue-Stieltjes measure on the real line. Absolute continuity, definition of Radon-Nikodym derivative and illustrations. (13hrs)

- Unit 2: Monotone convergence theorem, Fatou's lemma and dominated convergence theorem. Borel-Cantelli lemma, convergence in probability, convergence almost surely, convergence in distribution, convergence in $r^{th}$ mean, convergence theorem for expectations. Slutsky's theorem. (13hrs)

- Unit 3: Weak law of large numbers - Kolmogorov's generalized WLLN (proof of sufficient condition only), Khintchine's WLLN as special case, Chebyshev's WLLN. Kolmogorov's strong law of large number sequence of independent and iid random variables. Kolmogorov's inequality. (13hrs)

- Unit 4: Characteristic function - properties, inversion theorem (statement only and proof for density version), uniqueness theorem, continuity theorem (statement only). Central limit theorem, Lindeberg-Levy and Liapounov central limit theorems. Statement of Lindeberg-Feller form (statement only). Application of these theorems. Sampling distributions, chi-square, $t$, $F$ and non-central chi-square, mgf of non-central chi-square distribution, reproductive property. Non-central $t$ and non-central $F$. (13hrs)

## Books for Reference:

1. Bhat, B. R. (1999). *Modern Probability theory* (3rd ed.). New Age International Publishers. (To be used as Text.)

2. Mukhopadhyaya, P. (1996). *Mathematical Statistics.* Calcutta Publishing House.

3. Pitman, J. (1993). *Probability.* Narosa.

4. Billingsley, P. (1986). *Probability and Measure.* John Wiley and Sons.

5. Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley.

6. Ash, R. B. and Catherine Doleans-Dade (2000). *Probability and Measure Theory.* Academic Press.

7. Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory.*

8. Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). Wiley Eastern.

9. Rohatgi, V. K. and Saleh, A. K. Md. E. (2015). *An Introduction to Probability Theory and Mathematical Statistics.* Wiley Eastern.

## STH 423 - Design and Analysis of Experiments (4 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical foundations on the fundamentals and principles involved in designed experiments.

- Unit 1: Gauss-Markov setup, normal equations and least squares estimates, estimable function and estimation space, variance and covariance of least squares estimates, estimation of error variance, estimation with correlated observations, simultaneous estimates of linear parametric functions. Tests of hypothesis for one and more than one linear parametric functions. Confidence intervals and regions, analysis of variance, power of F-test, multiple comparison tests - Tukey and Bonferroni, simultaneous confidence interval.        (13hrs)

- Unit 2: Introduction to designed experiments, general block design and its information matrix, criteria for connectedness, balance and orthogonality. Intrablock analysis - estimability, best point estimates/interval estimates of estimable linear parametric functions and testing of linear hypotheses, estimation of parameters.        (13hrs)

- Unit 3: BIBD - definition, concept of connectedness, balancing, properties, estimability, recovery of inter-block information. Analysis of covariance in a general Gauss-Markov model, applications to CRD and RCBD. Fixed, mixed and random effects models, variance components estimation, study of various methods. (13hrs)

- Unit 4: General factorial experiments, factorial effects - best estimates and testing the significance of factorial effects, study of $2^n$ and $3^n$ factorial experiments in randomized blocks. Complete and partial confounding. Nested designs. Split-plot, strip plot designs. (13hrs)

## Books for Reference:

1. Bapat, R. B. (2012). *Linear Algebra and Linear Models.* Hindustan Book Agency.

2. Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* Wiley Eastern.

3. Aloke Dey (1986). *Theory of Block Designs.* Wiley Eastern.

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

5. Chakrabarti, M. C. (1962). *Mathematics of Design and Analysis of Experiments.* Asia.

6. Cochran and Cox, D. R. (1957). *Experimental Designs.* John Wiley.

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

8. Giri, N. (1986). *Analysis of Variance.* South Asian Publishers.

9. John, P. W. M. (1911). *Statistical Design and Analysis of Experiments.* Macmillan.

10. Joshi, D. D. (1987). *Linear Estimation and Design of Experiments.* Wiley Eastern.

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

12. Mukhopadhyay, P. (1998). *Applied Statistics.* Books and Allied (P) Ltd.

13. Pearce, S. C. (1984). *Design of Experiments.* New York: Wiley.

14. Rao, C. R. and Kleffu, J. (1988). *Estimation of Variance Components and Applications.* North Holland.

15. Searle, S. R., Casella, G. and McCullugh, C. E. (1992). *Variance Components.* Wiley.

## Practicals based on Design and Analysis of Experiments

1. One way classification and multiple comparison tests.

2. Two way classification with equal number of observations per cell (model with interaction).

3. Two way classification with unequal number of observations per cell (model without interaction).

4. Estimability and completeness.

5. Analysis of general block design.

6. Analysis of LSD and BIBD.

7. Analysis of covariance in one way and two way model.

8. $2^k$ factorial experiments and analysis of single replicate of $2^k$.

9. Total and partial confounding in $2^k$ factorial experiments.

10. Analysis of $2^k$ fractional factorial experiments.

11. Analysis of $3^2$ factorial experiments.

12. Analysis of one way classification random effects data.

## STH 424 - Theory of Estimation (4 Credits)

**Rationale/Learning Objectives:**

This course provides necessary theoretical foundations on the developments and applications of various estimation techniques.

- Unit 1: Parametric models, likelihood function, example from standard discrete and continuous models. Plotting likelihood functions. Sufficiency, Neyman factorization criterion, Fisher information for single and several parameters. Minimal sufficient statistic, likelihood equivalence. Exponential families and Pitman families. Completeness, Ancillary Statistics, Basu's theorem and applications. (13hrs)

- Unit 2: Minimum variance unbiased estimation, unbiasedness, locally unbiased estimators, minimum variance, locally minimum various, mean squared error, Cramer-Rao lower bound approach. Minimum variance unbiased estimators (MVUE), Rao-Blackwell theorem, completeness, Lehman-Scheffe's theorem, necessary and sufficient condition for MVUE. Bhattacharya bounds (without proof). Introduction to interval estimation, construction of confidence intervals using pivot. (13hrs)

- Unit 3: Consistency, estimation of real and vector valued parameters, invariance properties. Consistency of estimators by method of moments and method of percentiles, mean squared error criterion, asymptotic relative efficiency, consistent asymptotic normal (CAN). (13hrs)

- Unit 4: Method of maximum likelihood - notion, MLE in location and scale family, exponential family, Cramer family (statement only). Cramer-Huzurbazar theorem. Solutions to likelihood equations method of scoring, Newton-Raphson and other iterative procedures. Fisher lower bound to asymptotic variance, extension to multi-parameter case (without proof). (13hrs)

## Books for Reference:

1. Casella, G. and Berge, R. L. (2002). *Statistical Inference* (2nd ed.). Singapore: Thomson-Duxbury.

2. Kale, B. K. (1999). *A First Course on Parametric Inference*. Narosa Publishing House.

3. Lehman, E. L. (1986). *Theory of Point Estimation*. John Wiley.

4. Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley Eastern.

5. Rohatgi, V. K. and Saleh, A. K. L. (2001). *An Introduction to Probability and Mathematical Statistics*. Wiley Eastern.

6. Rajagopalan, M. and Dhanavanthan, P. (2012). *Statistical Inference*. Phi Learning Pvt. Ltd.

7. Zacks, S. (1981). *Parametric Statistical Inference*. Pergamon Press.

## Practicals based on Theory of Estimation

1. Estimation by the methods of moments and percentile.

2. Estimation by the methods of MLE for non-regular and multi-parameters.

3. Estimation by the methods of MLE using iterative method.

4. Construction of UMVUE.

5. Asymptotic behavior of estimation.

## STS 425 - Data Management and Statistical Computing with Python (3 Credits)

### Rationale/Learning Objectives:

This course provides comprehensive knowledge of python programming paradigms required for data management and statistical computing.

- <u>Unit 1</u>: Using Numpy - Basics of NumPy-Computation on NumPy-Aggregations-Computation on Arrays Comparisons, Masks and Boolean Arrays-Fancy Indexing-Sorting Arrays-Structured Data: NumPy's Structured Array.        (14hrs)

- <u>Unit 2</u>: Data Manipulation with Pandas - Introduction to Pandas Objects-Data indexing and Selection-Operating on Data in Pandas-Handling Missing Data-Hierarchical Indexing - Combining Data Sets.        (14hrs)

- <u>Unit 3</u>: Visualization and Matplotlib - Basic functions of matplotlib-Simple Line Plot, Scatter Plot-Density and Contour Plots- Histograms, Binnings and Density-Customizing Plot Legends, Colour Bars-Three- Dimensional Plotting in Matplotlib.        (12hrs)

## Books for Reference:

1. VanderPlas, J. (2016). *Python Data Science Handbook - Essential Tools for Working with Data.* O'Reily Media, Inc.

2. Zhang, Y. (2016). *An Introduction to Python and Computer Programming.* Springer Publications.

3. Thareja, R. *Python Programming using Problem Solving Approach.* Oxford University Press.

4. Grus, J. (2016). *Data Science from Scratch First Principles with Python,.* O'Reilly Media.

5. Padmanabhan, T. R. (2016). *Programming with Python.* Springer Publications.

## Practicals based on Data Management and Statistical Computing with Python

1. Write a NumPy program to compute sum of all elements, sum of each column and sum of each row of a given array.

2. Write a NumPy program to find rank, determinant, trace and eigen value of an array.

3. Write a Numpy program to sort array content in row and column wise.

4. Write a program to preprocess the data using Numpy and sklearn.preprocessing packages.

5. Write a python program to import pandas library to perform joining, merging and concatenating different dataframe.

6. Write program to do the following:

   (a) Create data frame df consisting 10 rows and 4 columns of randomly generated numbers between 1 to 100.

   (b) Create a new column such that, each row contains the row number of nearest row-record by Euclidean distance.

7. Use Iris data set, write program to answer the following questions:

   (a) Find the mean, median, standard deviation of iris's sepal-length (1st column).

   (b) Create a normalized form of iris's sepal-length whose values range exactly between 0 and 1 so that the minimum has value 0 and maximum has value 1.

   (c) Find the number and position of missing values in iris's sepal-length (1st column).

8. Use Automobile dataset, write program to answer the following questions:

   (a) Clean the data and update the CSV file.

   (b) Find the most epesensive car.

   (c) Find each company's highest price car.

9. Use Companies sales dataset, write program to answer the following questions:

   (a) Read total profit of all months and show it using a line plot.

   (b) Read all product sales data and show it using a multiline plot.

(c) Calculate total sale data for last year for each product and show it using a pie chart.

10. Use SAHeart dataset, write program to answer the below questions:

(a) Draw a bar plot to show the number of person having CHD or not in comparison to they having family history of the disease or not.

(b) Find out the number of CHD cases in different age categories. Do a Bar Plot and sort them in the order of age groups.

*III Semester*

# STE 531 - Inferential Statistics and Data Analysis (3 Credits)

**Rationale/Learning Objectives:**

This course provides fundamentals for developing various tests for the validity of different hypotheses, which are widely used in data analysis. (**Open Elective Course**)

- Unit 1: The concept of hypothesis and tests of hypothesis: null hypothesis, alternate hypothesis, test statistic, level of significance, $p$-value, testing hypothesis about population means, and population proportions, confidence intervals. Nonparametric tests - sign test, Wilcoxon-Mann-Whitney test, Wilcoxon signed rank test. Contingency tables, chi-square test for independence of attributes. (16hrs)

- Unit 2: Testing for the equality of several population means. The concept of analysis of variance, one way analysis of variance, its utility in the analysis of survey data and data obtained from designed experiments. (10hrs)

- Unit 3: Regression and correlation - bivariate data, correlation, scatter plot, correlation coefficient and its properties, testing for correlation coefficient, rank correlation. Regression - use of simple linear regression model to study the linear relationship between two variables, fitting the simple linear regression model, testing significance of regression coefficient, coefficient of determination. (14hrs)

## Books for Reference:

1. Campbell, R. C. (1974). *Statistics for Biologists*. Cambridge University Press.

2. Chatfield, C. (1981). *Statistics for Technology*. Chapman and Hall.

3. Frank, H. and Athoen, S. C. (1997). *Statistics: Concepts & Applications.* Cambridge University Press.

4. Medhi, J. (1992). *Statistical Methods: An Introductory Text.* Wiley Eastern Limited.

5. Ross, S. M. (2017). *Introductory Statistics.* Academic Press.

6. Rice, J. A. (2006). *Mathematical Statistics and Data Analysis.* Singapore: Thomson-Duxbury.

# STH 532 - Theory of Testing of Hypothesis (4 Credits)

**Rationale/Learning Objectives:**

This course provides necessary theoretical foundations on developments and applications of various tests for the validity of different hypotheses.

- Unit 1: Framing of null hypothesis, critical region, level of a test, randomized and non-randomized tests, two kinds of error, size of a test, p-value, power function. Most powerful tests in class of size a test, Neyman-Pearson lemma, MP test for simple null against simple alternative hypothesis. UMP tests for one sided null against one sided alternatives, monotone likelihood ratio property. Extension of these results in Pitman family when only upper or lower end points depend on the parameter. (13hrs)

- Unit 2: Non-existence of UMP test. Neyman-Pearson generalized lemma (statement only), concept of UMP for simple null against two sided alternatives in one parameter exponential family and UMPU tests with application to one parameter exponential family, UMP for two sided null (statement only). Likelihood ratio test (LRT), asymptotic distribution of LRT statistic, Pearson's chi-square test for goodness of fit, Bartlett's test for homogeneity of variances, large sample tests. (16hrs)

- Unit 3: Interval estimation, confidence level, construction of confidence intervals by inverting acceptance region. Shortest expected length confidence interval, evaluating interval estimators using size and coverage probability and test related optimality, uniformly most accurate one-sided confidence interval and its relations to UMP test for one sided null against one sided alternative hypothesis. (10hrs)

- Unit 4: U-statistics, properties and asymptotic distributions (in one and two sample case). Nonparametric tests: One sample test - test based on total number of runs, the ordinary sign test, the Wilcoxon signed rank test, the

Kolmogorov-Smirnov one sample goodness of fit test. Two sample tests - the median test, the Wilcoxon-Mann-Whitney test, Kolmogorov-Smirnov two sample test. (13hrs)

## Books for Reference:

1. Casella, G. and Berger, R. L. (2002). *Statistical Inference.* Wadsworth Group.

2. Gibbons, J. D. (1971). *Nonparametic Inference.* McGraw Hill.

3. Kale, B. K. (1999). *A First Course on Parametric Inference.* Narosa Publishing House.

4. Lehmann, E. L. and Romano, J. (2008). *Testing Statistical Hypotheses.* John Wiley.

5. Pratt, T. W. and Gibbons, J. D. (1981). *Concepts of Nonparametric Theory.* Springer.

6. Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* Wiley Eastern.

7. Rohatgi, V. K. and Saleh, A. K. L. (2001). *An Introduction to Probability and Mathematical Statistics.* Wiley Eastern.

8. Rajagopalan, M. and Dhanavanthan, P. (2012). *Statistical Inference.* Phi Learning Pvt. Ltd.

## Practicals based on Theory of Testing of Hypothesis

1. MP tests-I

2. MP tests-II

3. UMP tests-I

4. UMP tests-II

5. UMPU tests.

6. Likelihood ratio tests.

7. Confidence intervals.

8. Large sample tests using variance stabilizing transformation.

9. Non-parametric tests.

10. Goodness of fit tests.

11. Kolmogorov-Smirnov tests.

## STH 533 - Regression Analysis (4 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical foundations on regression techniques, which are extensively used in data analysis.

- <u>Unit 1</u>: Simple linear regression, multiple linear regression, basic assumptions, ordinary least squares (OLS) - estimation and their properties, tests of hypothesis about regression coefficients, likelihood ratio criterion. Dummy variables. Prediction - best linear unbiased predictor. (10hrs)

- <u>Unit 2</u>: Regression diagnostics and specification tests - residual analysis for identifying influential observations, recursive residuals and their applications, specification tests, subset selection of explanatory variables, Mallows $C_p$ statistic. Use of prior information. Restricted least squares estimators and mixed regression estimator. (10hrs)

- <u>Unit 3</u>: Violation of basic ideal conditions - disturbance with non-zero mean, asymptotically uncooperative regressors. Multicollinearity - its consequences and testing. Ridge estimator and its properties, ridge regression. Stochastic regressors, autoregressive models, instrumental variables, errors in variables. Distributed lag models. (10hrs)

- <u>Unit 4</u>: Heteroscedasticity - tests for heteroscedasticity. Generalized least squares (GLS) estimators and its properties, feasible generalized least squares estimators. Grouping of observations. Sets of Regression Equations. Auto correlation - its consequences and testing for autocorrelation, estimation and prediction. Autoregressive conditional heteroscedasity (ARCH) models. (10hrs)

- <u>Unit 5</u>: Simultaneous equation models. Identification problem, identification using linear homogeneous restrictions on structural parameters, rank and order conditions, estimation in simultaneous equation models. Indirect least squares, two stage least squares, structural equation modelling. (12hrs)

## Books for Reference:

1. Cook, R. D. and Weisberg, S. (1982). *Residual and Influence in Regression.* London: Chapman and Hall.

2. Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis* (3rd ed.). New York: Wiley.

3. Gunst, R. F. and Mason, R. L. (1980). *Regression Analysis and its Application - A Data Oriented Approach.* Marcel Dekker.

4. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). *Introduction to Linear Regression Analysis.* John Wiley.

5. Ryan, T. P. (1997). *Modern Regression Methods.* New York: John Wiley.

6. Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis* (2nd ed.). New York: John Wiley.

7. Fomby, T. B., Hill, C. R. and Johnson, S. R. (1988). *Advanced Econometric Methods.* Springer.

8. Greene, W. H. (2002). *Econometric Analysis* (5th ed.). New York: Prentice Hall.

9. Johnston, J. and Dinardo, J. (1996). *Econometric Methods* (4th ed.). McGraw-Hill.

10. Maddala, G. S. (1992). *Introduction to Econometrics* (2nd ed.). New York: Macmillan.

11. Gujarati, D. N. (2004). *Basic Econometrics* (4th ed.). McGraw-Hill.

# STH 534 - Stochastic Processes (4 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical knowledge on modelling for sequence of non-independent random variables, which are extensively used in the analysis of time dependent data.

- <u>Unit 1</u>: Introduction to stochastic processes - classification according to state space and time domain. Stationary process - weakly stationary and strongly stationary processes. Countable state Markov chains (MCs), Chapman Kolmogorov equations, calculation of $n$-step transition probability and its limit. Stationary distribution, classification of states, random walk and gamblers ruin problem, estimation of TPM for finite states of MC.  (15hrs)

- <u>Unit 2</u>: Discrete state space continuous time MC, Kolmogorov-Feller differential equations, Poisson process, birth and death process, applications to queues. Wiener process as a limit of random walk, first passage time and other problems.  (13hrs)

- <u>Unit 3</u>: Renewal theory - elementary renewal theorem and applications. Statement and uses of key renewal theorem, study of residual life time process.  (11hrs)

- Unit 4: Branching process - Galton-Watson branching process, probability of ultimate extinction, distribution of population size. Martingale in discrete time - definition and elementary properties, convergence theorem, applications. (13hrs)

## Books for Reference:

1. Basu, A. K. (2003). *Introduction to Stochastic Processes.* Narosa Publications.

2. Bhat, B. R. (2000). *Stochastic Models: Analysis and Applications.* New Age International.

3. Karlin, S. and Taylor, H. M. (1975). *A First Course in Stochastic Processes.* Academic Press.

4. Medhi, J. (1982). *Stochastic Processes.* Wiley Eastern.

5. Ross, S. M. (1983). *Stochastic Processes.* John Wiley & Sons.

6. Lawler, G. F. (2006). *Introduction to Stochastic Processes* (2nd ed.). Chapman and Hall.

## Practicals based on Stochastic Processes

1. Realization of stochastic processes.

2. Calculation of $n$-step transition probabilities.

3. Classification of states and mean recurrence time of state.

4. Simulation of Markov chain and estimating the stationary distribution of ergodic Markov chain.

5. Simulation of Poisson processes.

6. Realization of queues and computations of typical events limiting.

7. Simulation of branching process and estimating its mean and variance.

## STS 535 - Multivariate Analysis (3 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical foundations on multivariate techniques, which are extensively used in data analysis.

- Unit 1: Nature of a multivariate problem, main types of multivariate problems, objectives of multivariate analysis. Organization of multivariate data, descriptive statistics, visualization techniques. Multivariate normal distribution properties, maximum likelihood estimators of the parameters. Independence of sample mean vector and sample covariance matrix. Assessing the assumptions of normality Q-Q plot, chi-square plot, transformations to near normality. (10hrs)

- Unit 2: Inference problems in multivariate normal distribution, Hotellings $T^2$, Mahalanobis $D^2$ statistics, likelihood ratio tests for collinearity, $q$-sample problem. Roy's union and intersection test. Test for symmetry. Confidence regions, simultaneous confidence statements. Independence of subvectors, sphericity test. Wishart matrix, statement of Wishart distribution, its properties and applications. (10hrs)

- Unit 3: Principal component analysis (PCA) - definition and properties, graphing the principal components, sample principal components, interpretation of zero, small and repeated eigenvalues, component loadings and component correlations, the problem of scaling, tests of hypotheses. Cluster Analysis - distances and similarity measures, hierarchical clustering methods, $k$-means method. (10hrs)

- Unit 4: Classification and discrimination problems - concepts of separation and classification, Bayes and Fisher's criteria, classification rules based on expected cost of misclassification (ECM) and total probability of misclassification (TPM), classification with two multivariate normal populations (equal and unequal covariance matrices), evaluating classification rules, classification with several populations, Fisher's linear discriminant function, tests associated with discriminant functions. (10hrs)

## Books for Reference:

1. Anderson, T. W. (1984). *An Introduction to Multivariate Analysis* (2nd ed.). John Wiley.

2. Flury, B. (1997). *A First Course in Multivariate Statistics*. Springer Texts in Statistics.

3. Kshirasagar, A. M. (1972). *Multivariate Analysis*. Marcel Dekker.

4. Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivatiate Analysis.* Academic Press.

5. Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* Wiley Eastern.

6. Johnson, R. A. and Wichern, D. W. (1986). *Applied Multivariate Statistical Analysis* (6th ed.). Prentice Hall of India.

7. Rencher, A. C. (2003). *Methods of Multivariate Analysis.* Wiley.

## Practicals based on Multivariate Analysis

1. Graphical representation of multivariate data.

2. Model sampling from multivariate normal distribution.

3. Applications of Hotelling's $T^2$.

4. Testing equality of covariance matrices.

5. Principal component analysis.

6. Cluster analysis.

7. Classification.

8. Discriminant analysis.

## Practicals based on Machine Learning with Python

1. Fashion trends online (FTO) is an e-commerce company that sells women apparel. It is observed that 10% of their customers return the items purchased by them for many reasons (such as size, color and material mismatch). On a specific day, 20 customers purchased items from FTO. Write program to answer the following:

    (a) Probability that exactly five customers will return the items.
    (b) Probability that a maximum of five customers will return the items.
    (c) Probability that more than five customers will return the items purchased by them.
    (d) Average number of customers who are likely to return the items and the variance and the standard deviation of the number of returns.

2. The number of calls arriving at a call center follows a Poisson distribution at 10 calls per hour. Write program to answer the following:

(a) Calculate the probability that the number of calls will be maximum five.

(b) Calculate the probability that the number of calls over a 3 hour period will exceed 30.

3. As per survey of pesticides among 1000 farmers in grape farming for around 10 acres of grape farmland, it was found that the grape farmers spray 38 liters of pesticides in a week on an average with the corresponding standard deviation of 5 liters. Assume that the pesticide spray per week follows a normal distribution. Write program to answer the following questions:

(a) What proportion of the farmers is spraying more than 50 liters of pesticide in a week?

(b) What proportion of the farmers is spraying less than 10 liters?

(c) What proportion of the farmers is spraying between 30 liters and 60 liters?

4. Design a python program to perform principal component analysis for a sample training data set.

5. Design a python program to perform discriminant analysis for a sample training data set.

6. Design a program to implement the simple linear regression model for a sample training data set stored as a CSV file.

7. Design a program to implement the multiple linear regression model for a sample training data set stored as a CSV file.

8. Design a program to implement the Bayesian classifier for a sample training data set stored as a CSV file.

9. Write a program to implement the gradient descent algorithm for predicting future sales using the dataset Advertising.csv.

10. Design a program to implement the k-means clustering for a sample training data set stored as a CSV file.

## *IV Semester*

# STH 541 - Time Series Analysis (4 Credits)

## Rationale/Learning Objectives:

This course provides theoretical knowledge on the developments and applications of various techniques used in analyzing time series data and also forecasting.

- <u>Unit 1</u>: Simple descriptive techniques - time series plots, trend, seasonal effect. Tests for trend and seasonality - estimation and elimination of trend and seasonal components. Exponential and moving average smoothing. Time series as discrete parameter stochastic process. Stationarity, autocovariance and autocorrelation function and their properties. Partial autocorrelation function. (13hrs)

- <u>Unit 2</u>: Probability models - White noise model, random walk, linear processes, moving average (MA), autoregressive (AR), ARMA and ARIMA models, invertibility, ACF and PACF of these processes. Spectral properties of stationary models - periodogram, spectrum. (13hrs)

- <u>Unit 3</u>: Spectral density function - estimation of spectral densities of AR, MA and ARMA models. Sample ACF and PACF for model identification. Model building - estimation of mean, autocovariance function and autocorrelation function. Estimation in AR models, Yule-Walker equations, estimation in MA model and ARMA models. Order selection in AR and MA models. (13hrs)

- <u>Unit 4</u>: Forecasting - forecast mean square error (FMSE), least squares prediction, BLUP, innovation algorithm. Box-Jenkins forecasting for ARMA Models. Forecasting through exponential smoothing and Holt Winters smoothing. Residual analysis and diagnostic checking. Non-stationary time series models and their identification. Introduction to ARCH and GARCH models. (13hrs)

## Books for Reference:

1. Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden Day.

2. Brockwell, P. J. and Davis, R. S. (2002). *Introduction to Time series and Forecasting* (2nd ed.). Springer.

3. Chatfield, C. (1996). *The Analysis of Time Series: An Introduction.* Chapman Hall.

4. Kendall, M. G. and Ord, J. K. (1990). *Time Series* (3rd ed.). Edward Arnold.

5. Montgomory, D. C. and Johnson, D. A. (1977). *Forecasting and Time Series Analysis.* McGraw Hill.

6. Tanaka, K. (1996). *Time Series Analysis.* Wiley Series.

7. Tsay, R. S. (2005). *Analysis of Time series.* John Wiley & Sons.

## Practicals based on Time Series Analysis

1. Time series plots and elimination of trend and seasonality.

2. Estimation of ACF and PACF.

3. Model identification and estimation of ARMA model.

4. Model identification and estimation of ARIMA model.

## STH 542 - Reliability and Survival Analysis (4 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical knowledge on reliability techniques and statistical lifetime models, which are being used in medical sciences and industries.

- Unit 1: Coherent structures, representation of coherent systems in terms of paths and cuts, duals systems, modules of coherent systems. Reliability of system of independent of components, association of random variables, bounds on system reliability, improved bounds on system reliability using modular decompositions, lifetime distribution of $k$ out of $n$ system. (13hrs)

- Unit 2: Measures of reliability, survival/failure rate, hazard function, cumulative hazard function, lack of memory property, graphs of the system reliability functions. Notion of aging, life distributions of coherent systems, classes of life distributions - parametric and nonparametric models, mean residual lifetime with survival function. NBU, NBUE, NWU, NWUE classes of life distributions and their implications. (13hrs)

- Unit 3: Complete and censored samples, type I, II and random censoring, life distributions - Exponential, Gamma, Weibull, Lognormal, Pareto family. Estimation of parameter for exponential and gamma distribution under various censoring situations. Confidence interval for parameters of Exponential, Weibull, and Lognormal distributions. Wald, Score and LR tests for Exponential against Gamma and Weibull. (13hrs)

- Unit 4: Estimation of survival function - Kaplan-Meier estimator, Nelson-Aalen estimator, Greenwoods formula. Other life table estimators. Actuarial method of estimation of survival function. Semi-parametric regression for failure rate, Cox's proportional hazards model with one and more number of covariates, log likelihood function, log linear hazards, test for regression coefficients with and without ties. (13hrs)

## Books for Reference:

1. Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models.* Holt, Rinehart and Winston Inc.

2. Barlow, R. E. and Proschan, F. (1996). *Mathematical Theory of Reliability.* John Wiley.

3. Tobias, P. A. and Trindane, D. C. (1995). *Applied Reliability* (2nd ed.). CRC Press.

4. Lawless, J. R. (1982). *Statistical Models and Methods for Lifetime Data.*

5. Bain, L. J. and Engelhardt (1991). *Statistical Analysis of Reliability and Life Testing Data.*

6. Zacks, S. (1992). *Introduction to Reliability Analysis: Probability Models and Statistical Methods.* Springer.

7. Cox, D. R. Oakes, D. (1984). *Analysis of Survival Data.* New York: Chapman and Hall.

8. Kalbfeish, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data* (2nd ed.). John Wiley & Sons, Inc.

9. Deshpande, J. V. and Purohit, S. G. (2005). *Lifetime Data: Statistical Models and Methods.* World Scientific.

## STS 543 - Statistical Modelling (3 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical knowledge on Bayesian statistics, nonparametric techniques and generalized linear models for data analysis and inference.

- Unit 1: Introduction to Bayesian theory and philosophy - loss function and risk, foundations of optimal decision making, Bayes rule, minimax rule, admissibility. Prior and posterior distributions, conjugate families, non-informative priors - uniform and Jeffrey's prior. Bayesian estimation. Introduction to credible sets, Bayesian hypothesis testing, Bayesian prediction. (15hrs)

- Unit 2: Nonparametric density estimation (kernel based), nonparametric regression techniques - kernel, nearest neighbour, local polynomial (LOESS regression) and spline based methods. Concept of Resampling techniques - bootstrap and jackknife methods. Bootstrap procedure - hypothesis testing and bootstrap interval estimation. (12hrs)

- <u>Unit 3</u>: Introduction to generalized linear model (GLM), logistic regression, multi-logit regression, count data regression, log linear regression.     (13hrs)

## Books for Reference:

1. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer-Verlag.

2. Ghosh, J. K., Delampady, M. and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer Texts in Statistics.

3. Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Massachusetts: Addison-Wesley, Reading.

4. Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.

5. Hardle, W. (1991). *Smoothing Techniques*. Springer Science & Business Media.

6. Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer Science & Business Media.

7. Wasserman, L. (2005). *All of Nonparametric Statistics*. Springer Science & Business Media.

8. Dobson, A. J. (1983). *Introduction to Statistical Modelling*. Chapman and Hall.

9. Agresti, A. (1990). *Categorical Data Analysis* (3rd ed.). Wiley.

10. Myers, R. H., Montgomery, D. C., Vining, G. G. and Robinson, T. J. (2010). *Generalized Linear Models: with Applications in Engineering and the Sciences* (2nd ed.). John Wiley & Sons.

11. Davison, A. C. and Hinkley, D. V. (1991). *Bootstrap Methods and their Application*. Cambridge University Press.

12. Higgins, J. J. (2004). *An Introduction to Modern Nonparametric Statistics*. Brooks/Cole.

## STS 544 - Big Data Analytics (3 Credits)

### Rationale/Learning Objectives:

This course enables the students to understand about big data and Hadoop ecosystem tools for large data analytics.

- <u>Unit 1</u>: Introduction to Big Data - Classification of Digital Data, Characteristics of Data, Evolution of Big Data, Challenges with Big Data, Business Intelligence Vs Big Data, A Typical Data Warehouse Environment. Big Data Analytics - Introduction to Big Data Analytics, Classification of Analytics, Importance of Big Data Analytics, Data Science, Terminology used in Big Data Environment. (10hrs)

- <u>Unit 2</u>: Hadoop - Hadoop Distributed File System Basics, Hadoop MapReduce Framework, MapReduce Programming. Hadoop Essential Tools - Apache HIVE, Apache PIG, Sqoop, Apache Flume. (10hrs)

- <u>Unit 3</u>: Business Intelligence Concepts and Application, Data Warehousing, Data Mining, Data Visualization. (10hrs)

- <u>Unit 4</u>: Artificial Neural Network, Association Rule Mining, Text Mining, Web Mining, Social Network Analysis. (10hrs)

## Books for Reference:

1. Acharya, S. and Chellappan, S. (2015). *Big Data and Analytics*. Wiley Publications.

2. Eadline, D. (2016). *Hadoop 2 Quick-Start Guide: Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem* (1st ed.). Pearson Education. ISBN-13: 978-9332570351.

3. Maheshwari, A. (2017). *Data Analytics* (1st ed.). McGraw Hill Education. ISBN-13: 978-9352604180.

4. White, T. (2015). *Hadoop: The Definitive Guide*. (4th ed.). O'Reilly Media. ISBN-13: 978-9352130672.

5. Lam, C. *Hadoop in Action*. Manning. ISBN 9781935182191.

## Artificial Intelligence (3 Credits)

### Rationale/Learning Objectives:

This course enables the students to identify the problems where artificial intelligence is required and the different methods available.

- <u>Unit 1</u>: What is artificial intelligence? Problems, Problem Spaces and search, Heuristic search technique. (10hrs)

- <u>Unit 2</u>: Knowledge Representation Issues, Using Predicate Logic, representing knowledge using Rules, Symbolic Reasoning under Uncertainty. (10hrs)

- <u>Unit 3</u>: Statistical reasoning, Weak Slot and Filter Structures, Strong slot-and-filler structures. (10hrs)

- <u>Unit 4</u>: Game Playing, Natural Language Processing, Learning. (10hrs)

## Books for Reference:

1. Rich, E., Knight, K. and Nair, S. B. *Artificial Intelligence* (3rd ed.). McGraw Hill.

2. Rusell, S. and Norving, P. *Artificial Intelligence: A Modern Approach* (2nd ed.). Pearson Education.

3. Patterson, D. W. *Introduction to Artificial Intelligence and Expert Systems.* Prentice Hall of India.

4. Luger, G. (2002). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving* (4th ed.). Pearson Education.

5. Rolston, D. W. *Artificial Intelligence and Expert Systems Development.* McGraw Hill.

6. Padhy, N. P. (2015). *Artificial Intelligence and Intelligent Systems.* Oxford University Press.

## Elements of Statistical Computing (3 Credits)

### Rationale/Learning Objectives:

This course provides foundations for statistical simulation and validation of models.

- <u>Unit 1</u>: Random number generation, requisites of a good random number generator, methods of random number generation such as linear congruential, mixed congruential and multiplicative congruential. Testing of random number generator, run test, Kolmogrov-Smirnov test, sign test, rank test, gap test, digit frequency test and serial correlation, selection of a random number generator. Methods of generating random observations such as inverse transforms, composition, convolution and acceptance-rejection. (10hrs)

- <u>Unit 2</u>: Simple optimization method, direct search, grid search, interpolatory search, gradient search. Newton-Raphson method, Muller's method, Aitken's extrapolation, simple problems and applications. (10hrs)

- <u>Unit 3</u>: Methods to compute integrals - quadrature formula, double integration, singularity, Gaussian integration. Monte Carlo Methods - Monte Carlo integration and simple case studies, applications of Monte Carlo methods to compute expected values of functions of random variables such as Laplace transform, fourier transform etc., some case studies. (10hrs)

- <u>Unit 4</u>: Approximating probabilities and percentage points in selected probability distribution, verification of WLLN and CLT using random number generator, simulating null distribution of various test statistics, simple applications and case studies. (10hrs)

## Books for Reference:

1. Kennedy, W. J. Gentle, J. E. (1980). *Statistical Computing*. Marcel Dekker.

2. Sen, K. V. (1993). *Numerical Algorithm Computation in Science and Engineering* (2nd ed.). Affiliated East West Press.

3. Law, A. M. and Kelton, W. D. (2000). *Simulation, Modeling and Analysis* (3rd ed.). Tata McGraw Hill.

4. Rajaraman, V. (1993). *Computer Oriented Numerical Methods* (4th ed.). Prentice Hall.

5. Ripley, B. D. (1987). *Stochastic Simulation*. John Wiley.

6. Ross, S. M. (2000). *Introduction to Probability Models*. Academic Press.

7. Ross, S. M. (2013). *Simulation*. Academic Press.

8. Thisted, R. A.(1988). *Elements of Statistical Computing*. Chapman and Hall.

# Practicals based on Elements of Statistical Computing

1. Generation of random numbers by acceptance rejection method.

2. Generation of random numbers by linear congruential method.

3. Solution of the equations using iterative methods.

4. Numerical integration.

5. Monte-Carlo integration.

6. Empirical distributions of the test statistics.

7. Applications of CLT.

# Survival Analysis (3 Credits)

### Rationale/Learning Objectives:

This course provides theoretical foundations on statistical lifetime models being used in medical sciences and industries.

- <u>Unit 1</u>: Complete and censored samples, type I, II and random censoring, life distributions - Exponential, Gamma, Weibull, Lognormal, Pareto family. Estimation of parameter for exponential and gamma distribution under various censoring situations. Confidence interval for parameters of Exponential, Weibull, and Lognormal distributions. Wald, Score and LR tests for Exponential against Gamma and Weibull. (10hrs)

- <u>Unit 2</u>: Life tables - standard methods for uncensored and censored data, asymptotic properties of estimates under a random censorship model. Failure rate, mean residual life and their elementary properties. Estimation of survival function - Kaplan-Meier estimator, Greenwoods formula. Other life table estimators. Actuarial method of estimation of survival function. (10hrs)

- <u>Unit 3</u>: Fully parametric analysis of dependency accelerated life model - simple form, log logistic accelerated life model, proportional hazards model in relation with accelerated life model. Semi-parametric regression for failure rate, Cox's proportional hazards model with one and more number of covariates, log likelihood function, log linear hazards, test for regression coefficients with and without ties. (10hrs)

- <u>Unit 4</u>: Two sample problem - Gehan test, log rank test, Mantel-Haenszel test. Competing risks model - parametric and nonparametric inference for these model. (10hrs)

## Books for Reference:

1. Cox, D. R. Oakes, D. (1984). *Analysis of Survival Data.* New York: Chapman and Hall.

2. Kalbfeish, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data* (2nd ed.). John Wiley & Sons, Inc.

3. Lawless, J. F. (2002). *Statistical Models and Methods for Lifetime Data.* John Wiley & Sons, Inc.

4. Miller, R. G. (1981). *Survival Analysis.* John Wiley & Sons, Inc.

5. Hosmer, D. W., Lemeshow, S. and May, S. (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data* (2nd ed.). John Wiley & Sons, Inc.

6. Deshpande, J. V. and Purohit, S. G. (2005). *Lifetime Data: Statistical Models and Methods.* World Scientific.

## Practicals based on Survival Analysis

1. Estimation of parameters under censoring.

2. Construction of life tables.

3. Test for class properties.

4. Kaplan-Meier estimators.

5. Accelerated failure time model.

6. Cox proportional hazard model.

7. Two-sample tests under censoring.

## Stochastic Finance (3 Credits)

**Rationale/Learning Objectives:**

This course provides foundations on fundamentals of financial markets and stocks and to analyze the data on finance.

- <u>Unit 1</u>: Basic concepts of financial markets and stocks, types of traders, forward contracts and futures, call and put options, European option and American options. Interest rates, continuous compounding, present value analysis, bond pricing, risk free interest rates. Returns, gross returns, log returns. (10hrs)

- <u>Unit 2</u>: Portfolio theory, mean variance portfolio theory. One risky asset and one risk free asset, two risky assets. Sharpes ratio, tangency portfolio, optional mix of portfolio. Market portfolio, beta, security market line, and capital asset pricing model (CAPM) and their assumption. Value at risk (VAR), nonparametric and parametric estimation of VAR, VAR for a derivative and for a portfolio of assets, delta normal method, simulation of VAR models. (10hrs)

- <u>Unit 3</u>: Financial derivatives, options, pricing via arbitrage, law of one price. Risk neutral valuation, arbitrage theorem. Convexity of cost of call option, binomial model single and multiperiod binomial model. Modeling returns - lognormal model, random walk model, modeling through geometric Brownian motion process. Ito lemma (without proof). Arbitrage theorem. The Black Scholes formula and assumptions, properties of the Black Scholes option cost. (10hrs)

- <u>Unit 4</u>: Black Scholes Merton differential equations and assumptions, the delta hedging arbitrage strategy, volatility and estimating the volatility parameter, implied volatility. Pricing American options, pricing of an European option using Monte Carlo and pricing an American option using finite difference methods. Call options on dividend paying securities. (10hrs)

## Books for Reference:

1. Ross, S. M. (2003). *An Elementary Introduction to Mathematical Finance.* Cambridge University Press.

2. Ruppert, D. (2004). *Statistics and Finance: An Introduction.* Springer International Edition.

3. Hull, J. C. (2008). *Options, Futures and Other Derivatives.* India: Pearson Education.

4. Cuthbertson, K. and Nitzsche, D. (2001). *Financial Engineering: Derivatives and Risk Management.* John Wiley & Sons Ltd.

5. Leuenberger, D. G. (1998). *Investment Science.* Oxford University Press.

6. Wilmott, P. (2000). *Quantitative Finance.* John Wiley & Sons.

7. Tsay, R. S. (2005). *Analysis of Time series.* John Wiley & Sons.

# Data Mining (3 Credits)

**Rationale/Learning Objectives:**

This course provides foundations on various statistical methods used in data analysis, including artificial intelligence and machine learning techniques.

- <u>Unit 1</u>: Data mining - motivations and importance, knowledge discovery in databases (KDD) process search - introduction, querying, approximation and compression. Kinds of data considered for data mining, basic data mining tasks, data mining issues. Data mining models - predictive and descriptive, inter connections between statistics, data mining, artificial intelligence and machine learning, applications of data mining. (10hrs)

- <u>Unit 2</u>: Data marts, databases and data warehouses, OLTP systems, multidimensional models - data cubes, OLAP operations on data cubes, multidimensional schemes. Data Pre-processing - data cleaning, data integration, data transformation and data reduction. Visualization techniques for multidimensional data - scatter plot matrix, star plots, Andrews plots, Chernoff faces, parallel axis plots. (10hrs)

- <u>Unit 3</u>: Supervised learning - classification and prediction, statistical classification, linear discriminants, Mahalanobis linear discriminant, Fisher's linear discriminant, Bayesian classifier, regression based classification, $k$-NN (nearest neighbour) classifier. Tree classifiers - decision trees, ID3 algorithm, CART. Artificial neural networks (ANN) - the learning problem, perceptron, the delta rule, multilayer feed forward neural network, back propagation learning algorithm. Support vector machines - Lagrangian formulation and solution, measuring classifier accuracy. (10hrs)

- <u>Unit 4</u>: Unsupervised learning - clustering problem, similarity and distance measures, partitioning algorithms - $k$-means, $k$-medoids (PAM) algorithms. Density based clustering algorithms (DBSCAN). Association rule mining - market basket analysis, frequent item sets, support and confidence of an association rule, a priori algorithm, partition algorithm. (10hrs)

## Books for Reference:

1. Han, J. and Kamber, M. (2002). *Data Mining: Concepts and Techniques.* USA: Morgan Kaufman Publishers.

2. Dunham, M. H. (2005). *Data Mining: Introductory and Advanced Topics.* Pearson Education.

3. Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* New York: Springer.

4. Berthold, M. R. and Hand, D. J. (2003). *Intelligent Data Analysis: An Introduction* (2nd ed.). Springer.

5. J. P. Marques de Sa (2001). *Pattern Recognition: Concepts, Methods and Applications*. Springer.

6. Chattamvelli, R. (2009). *Data Mining Methods*. Narosa Publishing House.

## Bayesian Inference (3 Credits)

### Rationale/Learning Objectives:

This course provides theoretical knowledge on Bayesian techniques for data analysis and inference.

- <u>Unit 1</u>: Introduction and philosophy - loss function and risk, foundations of optimal decision making, Bayes rule, minimax rule, admissibility, sufficiency and Rao-Blackwellization. (10hrs)

- <u>Unit 2</u>: Utility theory, utility and loss, personal utility function, prior and posterior, conjugate families, non-informative priors - uniform prior, Jeffrey's prior, left and right invariant prior. (10hrs)

- <u>Unit 3</u>: Bayesian analysis - the posterior distribution, Bayesian estimation, credible sets, Bayesian hypothesis testing, Bayesian prediction, empirical Bayes analysis, hierarchical Bayes analysis, Bayesian robustness. (10hrs)

- <u>Unit 4</u>: Bayesian computation - analytic approximation, the EM algorithm, Monte Carlo sampling, Markov Chain Monte Carlo methods. (10hrs)

## Books for Reference:

1. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis* (2nd ed.). New York: Springer-Verlag.

2. Ghosh, J. K., Delampady, M. and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. New York: Springer Texts in Statistics.

3. Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Massachusetts: Addison-Wesley, Reading.

## Practicals based on Bayesian Inference

1. Computation of risk functions.

2. Computation of Bayes and minimax rules.

3. EM algorithm.

## Statistical Methods for Reliability (3 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical knowledge on reliability techniques for data analysis and inference.

- <u>Unit 1</u>: Coherent structures, representation of coherent systems in terms of paths and cuts, duals systems, modules of coherent systems. Reliability of system of independent of components, association of random variables, bounds on system reliability, improved bounds on system reliability using modular decompositions. (10hrs)

- <u>Unit 2</u>: Measures of reliability, graphs of the system reliability functions. Notion of aging, life distributions of coherent systems, distributions with increasing failure rate average arising from shock models, preservation of life distribution classes under reliability operations. Reliability bounds, lifetime distribution of $k$ out of $n$ system. (10hrs)

- <u>Unit 3</u>: Classes of life distributions - parametric and nonparametric models, mean residual lifetime with survival function. Applicable in replacement models, NBU, NBUE, NWU, NWUE classes of life distributions and their implications. Shock models leading to NBU. Age replacement and block replacement policies. Renewal theory useful in replacement models. (10hrs)

- <u>Unit 4</u>: Replacement policy comparisons, preservation of life distribution classes under reliability operations. Reversed hazard rate, cumulative reversed hazard function, relation between hazard function and reversed hazard function. Lack of memory property. (10hrs)

## Books for Reference:

1. Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models.* Holt, Rinehart and Winston Inc.

2. Barlow, R. E. and Proschan, F. (1996). *Mathematical Theory of Reliability.* John Wiley.

3. Tobias, P. A. and Trindane, D. C. (1995). *Applied Reliability* (2nd ed.). CRC Press.

4. Lawless, J. R. (1982). *Statistical Models and Methods for Lifetime Data.*

5. Bain, L. J. and Engelhardt (1991). *Statistical Analysis of Reliability and Life Testing Data.*

6. Zacks, S. (1992). *Introduction to Reliability Analysis: Probability Models and Statistical Methods.* Springer.

## Nonparametric Inference (3 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical knowledge on various nonparametric testing procedures for data analysis.

- Unit 1: Empirical distribution function, Glivenko-Cantelli theorem, Kolmogorov goodness of fit test. 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. (10hrs)

- Unit 2: Rank tests, locally most powerful rank test, linear rank statistics and their distributional properties under null hypothesis, Pitman's asymptotic relative efficiency. (10hrs)

- Unit 3: 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 test based linear rank statistics. Kruskal-Wallis $k$ sample test. (10hrs)

- Unit 4: Cox's proportional hazards model, rank test (partial likelihood) for regression coefficients, concepts of Jackknifing method of quenouille for reducing bias, bootstrap methods, confidence intervals. (10hrs)

## Books for Reference:

1. Cox, D. R. and Oakes, D. (1983). *Survival Analysis.* Chapman and Hall.

2. Davison, A. C. and Hinkley, D. V. (1991). *Bootstrap Methods and their Application.* Cambridge University Press.

3. Fraser, D. A. S. (1957). *Nonparametric Methods in Statistics.* John Wiley.

4. Gibbons, J. D. (1985). *Nonparametric Statistical Inference* (2nd ed.). Marcel Dekker.

5. Hajek, J. and Sidak, Z. (1961). *Theory of Rank Tests*. Academic Press.

6. Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley.

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

## Actuarial Methods (3 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical knowledge on various models used in insurance, risk analysis and theory of credibility.

- Unit 1: Review of decision theory and actuarial applications. Loss distributions - modeling of individual and aggregate losses, moments, fitting distributions to claims data, deductibles and retention limits, proportional and excess of loss reinsurance, share of claim amounts, parametric estimation with incomplete information. (10hrs)

- Unit 2: Risk models - models for claim number and claim amount in short term contracts, moments, compound distributions, moments of insurer's and reinsurer's share of aggregate claims. (10hrs)

- Unit 3: Review of Bayesian statistics - estimation and application. Experience rating - rating methods in insurance and banking, claim probability calculation, stationary distribution of proportion of policyholders in various levels of discount. (10hrs)

- Unit 4: Delay/run-off triangle - development factor, basic and inflation-adjusted chain-ladder method, alternative methods, average cost per claim and Bornhuetter Ferguson methods for outstanding claim amounts, statistical models. (10hrs)

## Books for Reference:

1. Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A. and Nesbitt, C. J. (1997). *Actuarial Mathematics* (2nd ed.). Society of Actuaries.

2. Klugman, S. A., Panjer, H. H., Willmotand, G. E. and Venter, G. G. (1998). *Loss Models: From Data to Decisions*. John Wiley & Sons.

3. Daykin, C. D., Pentikainen, T. and Pesonen, M. (1994). *Practical Risk Theory for Actuaries*. Chapman Hall.

# Pattern Recognition and Image Processing (3 Credits)

### Rationale/Learning Objectives:

This course provides theoretical knowledge on various statistical techniques involved in pattern recognition and image processing for data analysis and inference.

Pattern Recognition

- Unit 1: Review of Bayes classification - error probability, error bounds, Bhattacharya bounds, error rates and their estimation. Parametric and nonparametric learning, density estimation. Classification trees. $k$-NN rule and its error rate. (8hrs)

- Unit 2: Neural network models for pattern recognition - learning, supervised and unsupervised classification. Unsupervised classification - split/merge techniques, hierarchical clustering algorithms, cluster validity, estimation of mixture distributions. (8hrs)

- Unit 3: Feature selection - optimal and suboptimal algorithms. Some of the other approaches like the syntactic, the fuzzy set theoretic, the neurofuzzy, the evolutionary (based on genetic algorithms), and applications. Some recent topics like data mining, support vector machines, etc. (8hrs)

## Books for Reference:

1. Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition* (2nd ed.). New York: Academic Press.

2. Devijver, P. A. and Kittler, J. (1982). *Pattern Recognition: A Statistical Approach*. Prentice Hall.

3. Jain, A. K. and Dube, R. C. (1988). *Algorithms for Clustering Data*. Prentice Hall.

4. Everitt, B. S. (1993). *Cluster Analysis*. Halsted Press.

5. Fu, K. S. (1982). *Syntactic Pattern Recognition and Applications*. Prentice Hall.

6. Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press.

7. Hastie, T., Tibshirani, R. and Friedman, J. H. (2001). *Elements of Statistical Learning.* Springer-Verlag.

8. Ripley, B. D. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press.

9. Theodoridis, S. and Koutroumbas, K. (1999). *Pattern Recognition.* Academic Press.

Image Processing

- Unit 4: Introduction, image definition and its representation. Typical IP operations like enhancement, contrast stretching, smoothing and sharpening, greylevel thresholding, edge detection, medial axis transform, skeletonization/thinning, warping. (8hrs)

- Unit 5: Segmentation and pixel classification, object recognition, some statistical (including Bayesian) approaches for the above, like Besag's ICM algorithm, deformable templates approach of Grenander. (8hrs)

## Books for Reference:

1. Young, T. Y. and Fu, K. S. (1986). *Handbook of Pattern Recognition and Image Processing.* Vols. 1 & 2, Academic Press.

2. Jain, A. (1989). *Fundamentals of Digital Image Processing.* Prentice Hall.

3. Castleman, K. R. (1996). *Digital Image Processing.* Prentice Hall.

4. Mardia, K. V. and Kanji, G. K. (1993). *Statistics and Images.* Carfax.

## Operations Research (3 Credits)

**Rationale/Learning Objectives:**

This course provides theoretical foundations on optimization techniques for managerial decision making process.

- Unit 1: Linear programming problem (LPP) - definition, formulation. Simplex method - canonical form, improving non-optimal basic feasible solution (B.F.S), conditions for optimality, conditions for unboundedness. Convex sets, geometry of simplex method extreme point and B.F.S, existence of B.F.S, existence of optimal B.F.S. Two phase method, big $M$ method. (10hrs)

- Unit 2: Duality theory of LPP - weak duality theorem and its properties, the fundamental duality theorem, complementary slackness theorem. Dual simplex method. Sensitivity analysis. Integer programming cutting plane technique, Gomory's algorithm for pure integer program. Dynamic Programming, multistage decision making problems, Bellman's principle of optimality, recursive nature of computation, application, applications of dynamic programming. (10hrs)

- Unit 3: Inventory theory - nature of inventory problem, motives for carrying inventory, deterministic inventory model with decay. Probabilistic inventory models, continuous review and periodic review systems, $(s, S)$ policy, heuristic solution of lot size reorder point model $((Q, r)$ policy). (10hrs)

- Unit 4: Queuing theory - characteristics of queues, $M/M/1$ system, steady state solution, measures of effectiveness, waiting time distributions, Little's formula, $M/M/1/K$ system, $M/M/C$ system, machine interference problem, $M/G/1$ system, Pollaczek Khintchine formula. (10hrs)

## Books for Reference:

1. Gross, D. and Harris, C. M. (1985). *Fundamentals of Queuing Theory* (2nd ed.). John Wiley.

2. Hadley, G. (1975). *Linear and Combinatorial Programming*. John Wiley & Sons.

3. Murty, K. G. (1976). *Linear and Combinatorial Programming*. John Wiley & Sons.

4. Kambo, N. S. (1991). *Mathematical Programming Techniques*. Affiliated East West Press.

5. Taha, H. A. (2001). *Operations Research: An Introduction* (6th ed.). India: Prentice Hall.

6. Sivazlian, B. D. and Stanfel, L. E. (1975). *Analysis of Systems in Operations Research*. Prentice Hall.

7. Daellenbach, H. G. and George, J. A. (1978). *Introduction to Operations Research Techniques*. Allyn and Bacon Inc.