DEPARTMENT OF STATISTICS

M.Sc. - STATISTICS

SYLLABUS

Effective from the Academic Year 2016-2017

LOYOLA COLLEGE (Autonomous)

Ranked 2 in INDIA RANKING 2017 - NIRF
'College of Excellence' Status Conferred by UGC in 2014
Re-accredited with 'A' Grade (3.70 CGPA) by NAAC in 2013
Chennai - 600 034

RESTRUCTURING-2016 (2016-17 batch ONWARDS)
PG - Arts / Science / Commerce / Social Work

Part	Semester 1	Semester 2	Summer Vacation	Summer 3	Summer 4	Total Hours
Major Core (MC)	30(20 C)	24(20 C)	– –	20(15 C)	30(24 C)	104(79 C)
Elective Subject (ES)	– –	4(3 C)	– –	4(3 C)	– –	8(6 C)
Inter - Disciplinary (ID)	– –	– –	– –	6(5 C)	– –	6(5 C)
Self study Paper (SSP)				Outside class hours(2C)		(2 C)
Summer Training Program (STP)	– –	– –	3 to 4 weeks (1 C)	– –	– –	(1 C)
Life Skills Training (LST)	– –	2h + 2h# (2 C)	– –	– –	– –	2+2# (2 C)
Extension Activities	LEAP	LEAP(3 C)	– –	– –	– –	(3 C)
Total Hours (Total Credits)	30 (20 C)	30+2# (23+5 C)	–(1 C)	30 (23+2 C)	30 (24 C)	120+2# (90+6+2*)C

Note: A theory paper shall have 5 to 6 contact hours and a practical session shall have 3 to 5 contact hours.

New format of the subject codes from the 2016 regulation

Subject codes are 10 characters long:

1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th

- 1st & 2nd digits – last two digits of regulation year in YY format (If 2016, it will be 16).
- 3rdalphabet: U – UG / P – PG / M – M.Phil. / D – Ph.D.
- 4th & 5thalphabets: department wise program code(example – MT / CO / HT…..)
- 6th digit: Semester for UG/ PG / M.Phil. and year for Ph.D.
- 7th & 8th alphabet: Category of paper or group of category of papers (GE/RL/OL/HE/OR/AL /ES/SK/MS/CM/CC/ ……)
- 8th & 9th digits: subject number range (01 to 99).

For example,

Example 1:16UCH1MC01

16 – Admitted in 2016
U – UG student
CH – Chemistry Student
1 – 1st Semester subject
MC01 – Major paper

Example 2:16PCO2ID01

16 – Admitted in 2016
P – PG student
CO – Commerce Student
2 – 2nd Semester subject
ID01 – Inter disciplinary paper

- For subjects which are carried forward from one regulation to the next, the first two digits representing the regulation alone will change.
- Subjects which are not carried forward from one regulation to the next, will not appear in the new regulation.
- For new subjects which need to be added to a regulation, a new subject code must be created in continuation of the last created code under that type/category.
- Subject codes which are identical (except for the first two digits which represent the regulation year) are treated as equivalent for the purpose of syllabus / question paper setting / conducting examination / etc.

DEPARTMENT OF STATISTICS (PG)

Sl. No	Sub. Code	Title
1	16PST1MC01	ADVANCED DISTRIBUTION THEORY
2	16PST1MC02	APPLIED REGRESSION ANALYSIS
3	16PST1MC03	STATISTICAL MATHEMATICS
4	16PST1MC04	STATISTICAL DATA ANALYSIS USING SAS
5	16PST1MC05	STATISTICS LAB – I
6	16PST2MC01	ESTIMATION THEORY
7	16PST2MC02	TESTING STATISTICAL HYPOTHESES
8	16PST2MC03	SAMPLING THEORY
9	16PST2MC04	CATEGORICAL DATA ANALYSIS
10	16PST2MC05	STATISTICS LAB – II
11	16PST2ES01	ACTUARIAL STATISTICS
12	16PST2ES02	MODERN PROBABILITY THEORY
13	16PST3MC01	MULTIVARIATE ANALYSIS
14	16PST3MC02	STOCHASTIC PROCESSES
15	16PST3MC03	STATISTICAL QUALITY CONTROL
16	16PST3MC04	STATISTICS LAB – III
17	16PST3ES01	ADVANCED OPERATIONS RESEARCH
18	16PST3ES02	NON-PARAMETRIC

		METHODS
19	16PST3ID01	MATHEMATICAL AND STATISTICAL COMPUTING
20	16PST4MC01	APPLIED EXPERIMENTAL DESIGNS
21	16PST4MC02	DATA WAREHOUSING AND DATA MINING
22	16PST4MC03	BIOSTATISTICS AND SURVIVAL ANALYSIS
23	16PST4MC04	STATISTICS LAB – IV
24	16PST4PJ01	PROJECTS

16PST1MC01 ADVANCED DISTRIBUTION THEORY

SEMESTER	**I**	**CREDITS**	**5**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**6**

Objectives:

1) To impart knowledge on the construction of statistical models
2) To apply statistical models which are relevant to real life problems

Unit 1: Discrete distributions : Uniform. Binomial, Poisson, Geometric, Negative Binomial, Hypergeometric, Power series. Continuous distributions : Uniform, Normal, Exponential, Gamma, Chi-square, t, F, Lognormal, Weibull, Cauchy, Beta, Inverse Gaussian. Characterizations of distributions : Geometric, Normal, Exponential. Truncated distributions : Binomial, Poisson, Normal.

Unit 2: Multivariate discrete distributions : Trinomial and Bivariate Poisson distribution, their properties, Multinomial and Multivariate Poisson distributions.

Unit 3: Multivariate continuous distributions : Bivariate normal and Bivariate exponential (Marshall and Olkin) distributions, properties, Multivariate extensions.

Unit 4: Non-central distributions : Non-central Chi-square, Non-central t and Non-central F distributions and their properties. Compound distributions and Mixtures of distributions, order statistics, their distributions and properties.

Unit 5: Quadratic forms in Normal variates, properties of idempotent matrices. Quadratic forms, definiteness of a quadratic form. Generalized inverse (elementary ideas only). Necessary and Sufficient condition for a Quadratic form to be distributed as a Chi-square, Cochran's theorem.

Books for study

1. Hogg, R.V.and Craig, A.T.(2002). Introduction to Mathematical Statistics . Pearson Education, Asia.
2. Johnson, N.L. and Kotz, S.(2004). Distributions in Statistics. Vol. 1 – 4. John Wiley and Sons, New York .

Books for reference:

1. Johnson, N.L., Kotz, S. and Balakrishnan, N. (2004). Continuous Univariate Distribution. Vol. 1 John Wiley and Sons,(Asia) Pte.Ltd. Singapore.
2. Johnson, N.L., Kotz, S. and Balakrishnan ,N(2004). Continuous Univariate Distributions Vol. 2. John Wiley and Sons,(Asia) Pte.Ltd. Singapore

16PST1MC02 APPLIED REGRESSION ANALYSIS

SEMESTER	**I**	**CREDITS**	**5**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**6**

Objectives:

1. To study linear and non linear relationships between variables.
2. To train students in applications of regression models in real life situations.

Unit 1: Multiple Linear Regression. Estimation of model parameters, Hypothesis testing, Confidence intervals, Prediction, Residual analysis, PRESS statistic, Lack of fit.

Unit 2: Correcting Model Inadequacies. Variance stabilizing transformations, Linearizing a model, Selection of transformation, Generalized and weighted least squares, Indicator variables.

Unit 3: Model Building. Model building problem, Variable selection – Stepwise regression methods, Multicollinearity problem, Diagnostic, Methods for dealing with multicollinearity – Lift curve – KS statistic – Cross validation.

Unit 4: Polynomial and Non Linear Regression. Polynomial regression model in one variable – Piecewise polynomial fitting (Splines), Non-parametric regression, IVM-Linear regression models, Non-linear least squares, Transformation to a linear model, Inference problem in non-linear regression.

Unit 5 : Auto correlation, Partial Auto Correlation, Stationarity, Unit Root Test, Non Stationarity in Variance, Random Walk,

Random Walk with Drift, Auto Regressive Model, Moving Average Process, ARIMA – Determining Model, Estimation and Forecasting.

Books for Study:

1. Montgomery,D.C., Peck E.A, Vining G.G.(2003).Introduction to Linear Regression Analysis. John Wiley and Sons, Inc.NY
2. Ngai Hang Chan(2002), Time Series Applications to Finance, Wiley Series

Books for reference:

Draper, N. R. & Smith, H(1998) Applied Regression Analysis, 3rd Ed. (John Wiley).

16PST1MC03 STATISTICAL MATHEMATICS

SEMESTER	**I**	**CREDITS**	**4**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**6**

Objectives:

1. To train students in the mathematical foundation for post-graduate studies in Statistics.
2. To expose students the mathematical pre-requisites of statistics.

Unit 1: Real Sequences – boundedness, monotonicity, convergence. Operations on sequences.

Infinite series – convergence – tests for convergence

Unit 2: Real functions – Boundedness, monotonicity , continuity. Differentiable functions – Extreme Values.

Unit 3: Riemann Integrals – Properties – Fundamental theorem. Improper integrals – Tests for convergence.

Unit 4: Euclidean Spaces – Linear Independence / Dependence – Basis – Dimension. Inner products – Orthogonality.

Unit 5: Eigen values and vectors – Quadratic forms – Diagonal forms – Matrix square root. Gram-Schmidt Orthogonalization (Notion only)

Books for Study:

1. Somasundaram,D and Choudhry,B (1999): A First Course in Mathematical Analysis –[Narosa Publishing house]

Unit – I : Sections 2.1, 2.3 to 2.8, 3.1 to 3.4, 3.6

Unit – II : Sections 4.2, 4.3, 4.4, 4.5, 7.1, 7.3, 9.2

Unit – III : Sections 8.1, 8.3, 8.4, 8.5

2. Franz, E. Hohn (1973): Elementary Matrix Algebra – [Amerind Publishing Co. Pvt. Ltd] Unit IV : Sections 5.9 to 5.13, 5.17, 5.19, 5.22, 6.4, 7.2, 7.3, 7.4

Unit V : Sections 8.1 to 8.4

3. Bellman, R (1974): Introduction to Matrix Analysis – [Tata-McGraw-Hill Publishing Co. Ltd] Unit – V : Sections 3.4 to 3.7, 6.5

Books for Reference:

1. Chakrabarti, A (2006): A First Course in Linear Algebra – [Vijay Nicole Imprints Pvt. Ltd]
2. Burkill, J. C. (1962): A First Course in Mathematical Analysis – [Cambridge University Press]
3. Goldberg, R. R. (1970): Methods of Real Analysis – [Oxford& IBH Publishers]
4. Hadley, G. (1987) : Linear Algebra – [Narosa Publishing House]

16PST1MC04 STATISTICAL DATA ANALYSIS USING SAS

SEMESTER	**I**	**CREDITS**	**3**
CATEGORY	**MC(L)**	**NO.OF HOURS/ WEEK**	**6**

Objective

1. To enable students to use SAS for Data processing
2. To expose students on the applications of Statistical Analysis using SAS

Unit 1: Data step and Proc Step, SAS Data Libraries, Creating dataset using data lines, Importing data using INFILE statement, Importing data using Proc Import, Creating HTML Output, Sub setting observations using conditional statements, Sub setting variables using Keep/Drop, Creating variables using IF-THEN else statements, Retain statement, FIRST. , LAST. , Date functions, Character functions.

Unit 2: SAS procedures, Sub setting in Procedures with the WHERE Statement, Sorting Data with PROC SORT, Printing Data with PROC PRINT, Summarizing Your Data Using PROC MEANS, Writing Summary Statistics to a SAS Data Set, Counting Data with PROC FREQ, Producing Tabular Reports with PROC TABULATE, PROC SORT, PROC SUMMARY

Unit 3: Modifying a Data Set Using the SET Statement, Stacking Data Sets Using the SET Statement, Interleaving Data Sets Using the SET Statement, Combining Data Sets Using a One-to-One Match Merge, Combining Data Sets Using a One-to-Many Match Merge, Merging Summary Statistics with the Original Data, Writing Multiple Data Sets Using the OUTPUT Statement, Changing Observations to Variables Using PROC TRANSPOSE

Unit 4: SAS Macro Concepts, Substituting Text with Macro Variables, Creating Modular Code with Macros, Adding Parameters to Macros, Writing Macros with Conditional Logic, Writing Data-Driven Programs with CALL SYMPUT. Proc SQL, Using Proc SQL to create tables, Modifying tables, Aggregating tables, Stacking and Merging tables

Unit 5: PROC UNIVARIATE, PROC MEANS, PROC CORR, PROC PLOT, PROC FREQ, PROC TTEST , PROC NPAR , PROC ANOVA, PROC REG, PROC ARIMA.

Books for Reference:

1. The Little SAS Book: A Primer, Fourth Edition, Lora D. Delwiche, Susan J. Slaughter
2. Learning SAS by Example: A Programmer's Guide, Ron Cody, SAS Institute

16PST1MC05 STATISTICS LAB – I

SEMESTER	**I**	**CREDITS**	**3**
CATEGORY	**MC(L)**	**NO.OF HOURS/ WEEK**	**6**

Objectives:

1) To expose students on the applications of Statistical Model Building using R-Programming

2) To develop statistical model building skills through analyzing real life problems

Basics of R Programming

1. Importing and Exporting Datasets
2. Usage of Inbuilt Mathematical, Statistical and Special Functions in R
3. Subset Observations , Variables from Datasets
4. Creating new variables from existing variables
5. Merging Datasets – Inner Joins, Outer Join, Left Outer Join, Left Inner Join
6. Coding user defined functions in R Language
7. Linear and Matrix Algebra
8. Linear Dependency and Linear Independency
9. Determining Matrix Determinant and Matrix Inverse
10. Determining Characteristic Roots and Characteristic Vectors
11. Determining Rank, Index and Signature of a Quadratic form

Advanced Distribution Theory

1. Fitting of Binomial and Poisson Distribution
2. Fitting of Truncated Binomial and Truncated Poisson Distribution
3. Fitting of Mixture of Geometric and Poisson
4. Fitting of Mixture of Two Poisson
5. Generating Samples from Discrete and Continuous Distributions

Applied Regression Analysis

1. Building Linear Regression Model with Categorical Explanatory Variable
2. Testing for Overall Model fit and Individual Regression Coefficients
3. Determining R-Square, Adjusted R-Square, MAE and MAPE
4. Study of Interaction Effects and outlier detection
5. Testing for Multicollinearity using VIF and Conditional Index
6. Transformation and Combining Variables to deal Multicollinearity
7. Residual Analysis – Normality of Residuals, Transformation, Functional form
8. Autocorrelation, Partial Autocorrelation , Stationarity
9. Constructing ARIMA model and measuring model performance
10. Model Validation, Bootstrap Regression Methods

16PST2MC01 ESTIMATION THEORY

SEMESTER	**II**	**CREDITS**	**5**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	5

Objectives:

1) To provide a strong theoretical foundation to aid in optimal inference strategies.
2) To illustrate the real life applications of estimation procedures.

Unit 1: Problem of point estimation -unbiasedness-uniformly minimum variance unbiased estimator. Necessary and sufficient condition for UMVUE. Properties of UMVUE, Examples. Cramer-Rao inequality.

Unit 2: Sufficiency, Fisher-Neyman factorization theorem, examples. Rao-Blackwell theorem,

Unit 3: Completeness and bounded completeness. Basu's theorem. Lehmann-Scheffe theorem, examples.

Unit 4: Method of maximum likelihood, consistent asymptotic normal (CAN) estimators, examples. Invariance property of CAN estimators.

Unit 5: Baye's and minimax estimation, examples. M-estimation, Jack knife and Bootstrap methods.

Books for study:

1. Kale, B.K.(2005): A first course on parametric inference, Narosa Publishing House.
2. Kendall, M.G and Stuart,A.(1967).The Advanced Theory of Statistics. Vol.2. Inference and Relationship. Hafner Publishing Co., New York.
3. Lehmann,E.L.andCasella,G.(1998).Theory of point estimation. Springer-Verlag

Books for reference:

1. Rohatgi,V.K. and SalehA.K.Md.E.,(2002).An Introduction to Probability and Statistics. John Wiley and Sons,N.Y.
2. Zacks, S. (1971). The Theory of Statistical inference. John Wiley and Sons,N.Y.

16PST2MC02 TESTING STATISTICAL HYPOTHESES

SEMESTER	**II**	**CREDITS**	**5**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**5**

Objectives:

1. To impart knowledge on techniques for testing of hypotheses towards decision support based on sample characteristics.
2. To illustrate the real life applications of testing problems and procedures

Unit 1: Statistical hypotheses, Neyman – Pearson fundamental lemma, distributions with monotone likelihood ratio, Generalization of the fundamental lemma (without proof).

Unit 2: Two-sided hypotheses, Unbiasedness for hypothesis testing. Applications to one-parameter exponential family. Similarity and completeness.

Unit 3: UMP unbiased tests for multiparameter exponential family and applications. Confidence Intervals. Unbiased confidence sets.

Unit 4: Symmetry and invariance, maximal invariants, most powerful invariant tests, unbiasedness and invariance.

Unit 5: Likelihood ratio tests, large sample properties, asymptotic distribution of LRT statistic for simple null hypothesis.

Books for study:

1. Kale, B.K.(2005) A first course on parametric inference, Narosa publishing house, New Delhi.
2. Lehmann, E.L. (1986) Testing Statistical Hypotheses, John Wiley and sons, NY.

Books for reference

1. Rohatgi, V.K. and Saleh, E.A.K. Md.(2002)An Introduction to Probability and Statistics, John Wiley and sons, NY
2. Kendall, M.G. and Stuart, A. (1967) The Advanced theory of Statistics, Vol-2, Hafner publishing co., New York.

16PST2MC03 SAMPLING THEORY

SEMESTER	**II**	**CREDITS**	**4**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**5**

Objectives:

1. To impart knowledge on various sampling techniques useful in survey methodology.
2. To demonstrate sampling inspection procedures using sampling techniques.

Unit 1: Introduction to the theory of sampling – sampling designs – estimation procedures – properties of estimators – SRSWOR – properties of SRSWOR –optimal properties of the sample mean.

Unit 2: Sampling with varying probability – procedures for pps selection –Desraj ordered and Murthy's unordered estimators – HT estimator – optimal properties of HT estimators – estimation of variance of HTE

Unit 3: Stratified sampling – estimation of the population mean – allocation problems. Systematic sampling – Methods for populations with linear trend – Comparison with SRSWOR and stratified sampling for standard populations.

Unit 4: Multi stage sampling- Multi phase sampling - Ratio estimation - approximation to bias and MSE – regression estimation - approximation to bias and MSE - Double sampling for ratio and regression. Cluster sampling and multistage sampling under SRS methods.

Unit 5: Non-response – effects of non-response, Warner' model, Simmons randomized response technique. Planning and organization of large scale surveys

Books for Study:

1. Cochran W.G .(2000). Sampling Techniques. John Wiley and Sons, New York.
2. Deming, W.E.(2000). Some Theory of Sampling. John Wiley and Sons, New York.
3. Desraj and Chandok,P.(1998). Sampling theory.Narosa Publishing House, NewDelhi.
4. Sampath,S.(2005). Sampling Theory and Methods. Narosa Publishing House.

Books for Reference:

1. Murthy.M.N.(1967).Sampling Theory and Methods. Statistical Publishing Society, Calcutta.
2. Sukhatme,P.V., Sukhatme,B.V., Sukhatme,S.and Asok,C. (2000).Sampling theory of Surveys with Applications. Indian Society of Agricultural Statistics, New Delhi

16PST2MC04 CATEGORICAL DATA ANALYSIS

SEMESTER	**II**	**CREDITS**	**4**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**5**

Objectives:

1. To introduce students to the exciting new area of analysis of categorical data
2. To equip students with knowledge and techniques required to handle data- modeling situations involving categorical data.

Unit 1: Categorical Response data – Inference procedures. Contingency tables – Comparison of proportions, partial association in 2 x 2 and I x J tables. Testing independence in two-way contingency tables.

Unit 2: Generalized Linear Model – For binary data & count data. Inference for & Fitting of GLMs.

Unit 3: Logistic Regression Model – Fitting & diagnostics. Conditional associations in 2 x 2 x K tables. Multinomial logit models – Baseline logit models for nominal responses & Cumulative logit model for ordinal responses.

Unit 4: Loglinear models for two-way tables; Loglinear models for Independence & Interaction in three-way tables. Loglinear - Logit model connection. Diagnostics for checking models. Ordinal Association Models. Probit Models.

Unit 5: Comparison of dependent proportions. Conditional logistic regression for Binary Matched pairs. Marginal models for square contingency tables. Symmetry, Quasi-Symmetry & Quasi-independence.

Book for Study:

Alan Agresti (2002): Categorical Data Analysis. John Wiley & Sons

Book for Reference:

1. Hosmer, D.W. & Lemeshow, S. (1989)Applied Logistic Regression (John Wiley).

16PST2MC05 STATISTICS LAB – II

SEMESTER	**II**	**CREDITS**	**2**
CATEGORY	**MC(L)**	**NO.OF HOURS/ WEEK**	**4**

Objectives:

2. To Provide hands on experience in implementation of concepts in Estimation theory, Testing of Statistical Hypothesis and Categorical Data Analysis
3. To apply advanced statistical software for relevant applications Estimation Theory
    - Estimation of unknown parameter through MLE Procedure
    - Deriving Confidence Intervals for the unknown parameter
    - Bayesian Estimation Procedure
    - Bootstrap and Jackknife Procedures

Testing of Statistical Hypothesis

1. Basic Parametric and Non Parametric Tests
2. Determining Test Function, Level and Power of Test Function, Power Curve
3. MP Test, Uniformly Most Powerful Test, Uniformly Most Powerful Unbiased test

Sampling Theory

1. Simple Random Sampling With and Without Replacement
2. Stratified Random Sampling, Midzuno Sampling
3. Horwitz – Thompsonn Estimator
4. Hansen – Horwitz Estimator
5. Desraj Ordered Estimator
6. Calculation of Inclusion Probabilities in Fixed and Varying Sampling Designs

Categorical Data Analysis

1. Building Binary Logistic Regression Model
2. Obtaining Estimated Probability, Optimal Cut Point

3. Deriving Classification Table, Sensitivity and Specificity, AUC Measures
4. Model Validation through AUC and Gains Chart
5. Probit Model
6. Multinomial Logistic Regression Model
7. Cumulative / Ordinal Logistic Regression Model
8. PoissonRegression Model

16PST2ES01 ACTUARIAL STATISTICS

SEMESTER	**II**	**CREDITS**	**3**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**4**

Objectives:

1. To educate interested students about the applications of statistics in insurance industry.
2. To compute several measures which are relevant to actuarial statistics.

Unit 1: Compound Interest-Accumulated value and present value annuities certain, present values, amounts, annuities, perpetuities, Redemption of loans.

Unit 2: Further compound interest and Annuities certain, Nominal and effective rates of discount – capital redemption of policies

Unit 3: Mortality tables – construction of mortality tables comparison of different mortality tables.

Unit 4: Life Assurance premiums – Assurance benefits – Life annuities and temporary annuities – Net premiums for assurance plans – Net premiums for Annuity plans-premium conversion table.

Unit 5: Office premiums – policy values – Further life contingencies - methods of valuation – Data for valuation – special reserves and adjustments

Books for Study:

1. Bowers N.L.,.Gerber H.U,.Hickman, J.C and Nesbitt, C.J.(2006) Actuarial Mathematics, Society of Actuaries, Itasca, USA second edition.

2. Dixit. S.P.Modi .C.S and Joshi R.V. (2000) Mathematical basics of Life Assurance, Insurance Institute of India, Bombay.
3. Donald, D.W.A.(1970) Compound Interest and annuities. Heinemann, London

Books for reference:

1. Mccutcheon J.J. and Scot (1989).Mathematics of Finance, Heinemann, London
2. Neil, A (1977)..Life contingencies, Heinemann, London
3. Spurgeon, E.T(1972) Life Contingencies Cambridge University Press

16PST2ES02 MODERN PROBABILITY THEORY

SEMESTER	**II**	**CREDITS**	**5**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**6**

Objectives:

1. To impart knowledge in the advanced probability theory
2. To illustrate probabilistic pre-requesites which are required for building statistical models.

Unit 1: Classes of events: Classes - Fields and σ-Fields; Definition of Probability; Independence of Events, Properties, Discrete, General, and Induced Probability Spaces, Counting, Lebesgue-Stieltjes measures. Distribution Function of a random variable and random vector, Decomposition of Distribution Functions , Independence of Random variables.

Unit 2: Expectation and Moments: Definition and Properties of Expectation-Moments Inequalities, Characteristic Functions – Properties of Characteristic Functions – Inversion Formula. Convergence theorems for Expectations. Conditional Expectation.

Unit 3 : Modes of Convergence of Random Variables : Limits of Random Variables, Convergence in Probability, Convergence Almost Surely, Convergence in Distribution, Convergence in rth mean,

Unit 4 : Laws of Large Numbers: Weak and Strong law of large Numbers

Unit 5: Central Limit Theorem :Central Limit Theorems for Independent Random Variables – Lindeberg – Levy, Liapunov and Lindeberg – Feller Theorems.

Book for Study:

1. Bhat,B.R.(2007) Modern Probability Theory, 3 Ed. New Age International Publishers.,
2. Rohatgi, V.K. and Saleh, A.K.Md.E (2002). Introduction to Probability and Statistics, Pearson Education, Asia.

Books for Reference:

1. Ash,R.B.(1972). Real Analysis and Probability, Academic Press.
2. Billingsley,P.(1991). Probability and Measure. John Wiley & Sons, New York.
3. Loeve,M.(2000). Probability Theory. Van Nostrand, Princeton
4. Parthasarathy,K.R.(1977). Introduction to Probability and Measure. Thomson wadsworth

16PHE2FC01 LIFE SKILLS TRAINING

SEMESTER	**II**	**CREDITS**	**2**
CATEGORY	**FC(T)**	**NO.OF HOURS/ WEEK**	**2+2**

OBJECTIVES OF PG SYLLUBUS

1. To improve and sustain the primal level of competence and performance of PG students through an advanced training of holistic development of oneself.
2. To empower through various skills and strengthen them to face the future life issues and challenges.
3. To equip them with practical and value based learning of soft skills for a better life in future.

INSIDE CLASS HOURS (2 hrs)

Unit – I: Constructing Identity

Self Image – Understanding self image – shadows down the lane – self acceptance - Self Knowledge – Knowing oneself - Self confidence – Guilt and grudges - Power of belief – positive thinking– optimizing confidence - Self development – perception, attitude and Behavioural change, developing a healthy and balance personality - Self esteem – signs - indicators

Unit – II: Capacity Building

Motivation – Definition, types (Intrinsic and Extrinsic), Theories (Maslow's hierarchical needs, etc), Factors that affect motivation, Challenges to motivation, Strategies to keep motivated, motivational plan. Time Management Skills– steps to improve time management, overcoming procrastination, assessing and planning weekly schedule, challenges, goal settings, components of goal settings, consequences of poor time management, control of interruption and distractions. Communication, public speaking, talents, creativity, learning,

Unit – III: Professional Skills

Leadership development skills – difference between leader and manager, different styles and their utilities, functions of leadership, application of knowledge, overcoming from obstacles, influential skills and Leadership qualities. Application skills – Managing Career and self-direction, Visionary thinking, formulating strategies, shaping strategies, building organizations relationships, change management. Project Management Skills, Independent working skills, Writing skills, Public Speaking, analytical Skills, Neo Research and Development. Problem solving skills – Process, approaches and its components, creative problem solving, Tools and techniques, application of SMART analysis and barriers to problem solving.

Unit – IV: Life Coping Skills

Life skills – Personal and reproductive Health, love, sex, marriage and family – family life education – Gender Equity - child bearing and Childrearing practices, Geriatric Care - adjustability Human Relationship – formal and informal - peer group – friends – same

and other gender - family – Colleagues – community – emotional intelligence - Stress Coping skills – Definition of stress, strategies to alleviate stress, problem and emotion focused coping, techniques to reduce stress, stress reaction phases, crisis intervention steps, creating positive affirmations, Signs, Symptoms and Reactions of Stress.

Unit – V: Social Skills

Human Rights Education, Understanding Human Rights, International and national mechanisms, protection and preservation of HRs, Human Rights in the context of new, technological and electronic society, Peace Education, Social Harmony in the context of religious fundamentalism and fanaticism, Understanding Peace and Justice, Conflict Resolution Strategies

Reference books

1. Healing Your Emotional Self: A Powerful Program to Help You Raise Your Self-Esteem, Quiet Your Inner Critic, and Overcome Your Shame by Beverly Engel
2. Self-knowledge and self-discipline by B. W. Maturin
3. Motivation: Biological, Psychological, and Environmental (3rd Edition) by Lambert Deckers
4. Getting Things Done: The Art of Stress-Free Productivity by David Allen
5. Managerial Skills in Organizations by Chad T. Lewis
6. Social Intelligence: The New Science of Human Relationships by Daniel Goleman

Competence building	**Career Preparatory Training**
Power talk	Interview Guidance
Emotional Intelligence	Group Dynamics
Stress management	Leadership skills
Decision Making	Negotiation Skills
Positive image building	Creative writing

OUTSIDE THE CLASS HOURS (2 hrs)

- Each student will choose either of the above-mentioned modules and is expected to undergo a training/workshop in that module.
- She/he will have to accomplish ten hrs outside the class hours to fulfill the 2 credits.

Methodology

Inputs, ABL model, Documentaries, group activities and Interaction, Special workshop by professionals.

Evaluation

There will be end test and a Project for 'inside the class hours'. Viva Voce will be conducted for the 'Outside the class hours'.

16PST3MC01 MULTIVARIATE ANALYSIS

SEMESTER	**III**	**CREDITS**	**5**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**6**

Objectives:

1. To provide students the requisite knowledge and skills to handle multi-dimensional data and extract useful information from the data.
2. To derive statistical inference based on multivariate statistical analysis

Unit 1: (Basic Preparation): Multivariate Normal distribution – Properties, Distributions of linear combinations, independence, marginal distributions, conditional distributions, Partial and Multiple correlation coefficient. Moment generating function.

Unit 2: (Inference Means of MV normal populations)): Estimation of Mean and Var-Cov matrix. Distributions of Sample correlation coefficient, partial correlation coefficient &Multiple correlation coefficient. The Generalized T2Statistic – Distribution & Applications.

Unit 3: (Inference for several populations): Paired comparisons & Repeated Measures design, Multivariate ANOVA – One way &Two way. Profile analysis.

Unit 4: (Analysis of Covariance structure): Principal Components. Factor Analysis – Orthogonal Factor model, Factor rotation, Factor scores. Canonical Correlation Analysis.

Unit 5: (Grouping Techniques): Discrimination & Classification – Fisher's method. Optimality of classification rules. Discrimination & classification for several populations. Cluster Analysis – Similarity measures, Hierarchical & Non-Hierarchical methods.

Books for study:

1. Anderson, T. W. (2003): An Introduction to Multivariate Statistical Analysis – 3rdedn. John Wiley & Sons.
2. Johnson, R. A., &Wichern, D. W. (2007): Applied Multivariate Statistical Analysis – 2ndedn. Prentice Hall International

Books for Reference:

1. Everitt, B.S & Dunn, G (2001): Applied multivariate Data analysis, second edition, Arnold publishers, London.
2. Morrison, D.F (1990): A multivariate statistical methods, Third edition, Mc graw hall, New delhi.

16PST3MC02 STOCHASTIC PROCESSES

SEMESTER III CREDITS 5
CATEGORY MC(T) NO.OF HOURS/ WEEK 6

Objectives:

1. To introduce the variety of situations that can be handled by the theory of Stochastic Processes
2. To expose students to several processes in disciplines like Biology, Engineering and Economics that are stochastic in nature.

Unit 1: Elements of Stochastic processes – simple examples, Classification of general stochastic processes. Stationary independent increment process. Properties.

Unit 2: Markov Chains – discrete in time. Examples. Classification of states of a Markov Chain. Recurrence. Basic limit theorem of Markov Chains. Absorption probabilities. Criteria for recurrence.

Unit 3: Markov Chains continuous in time. Examples. General Pure birth process, Poisson process, Birth – Death process. Finite state continuous time Markov Chains. Bivariate Poisson process.

Unit 4: Renewal process – Definition and examples, Elementary Renewal Theorem, Martingales – Examples. Super and Sub - martingales.

Unit 5: Branching process – generating function relations, estimation probabilities, two – type branching process – Description of continuous time branching process. Stationary process – mean square distance, prediction and covariance stationary process.

Books for study:

1. Medhi, J(1996). Stochastic Processes, Wiley Eastern Limited.
2. Karlin, S and Taylor, H.M.(1978). A first course in Stochastic Processes. Academic Press, New York.

Books for Reference

1. Karlin, S and Taylor, H.M.(1981). A second course in Stochastic Processes. Academic Press, New York.
2. Ross, S.M.(1983). Stochastic Processes, John Wiley and Sons.

16PST3MC03 STATISTICAL QUALITY CONTROL

SEMESTER	**III**	**CREDITS**	**3**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**4**

Objectives:

1. To train students in modern statistical process control techniques.
2. To detect variation in production process and suggests inspection plans

Unit 1: Meaning and Scope of Statistical process control (SPC), Shewart Control Charts for X-bar, R, np, p, c charts. and their uses. OC and ARL of control charts. Uses of runs and related patterns of points.

Unit 2: control charts based on C.V. extreme values, moving averages, geometric moving averages, modified control charts CUSUM procedures, use of V mask, derivation of ARL. Multivariate control charts.

Unit 3: Process capability, tolerance limits, beta content and beta expectation, Normal theory and non-parametric approaches.

Unit 4: Sampling inspection plans. Classification and general properties Sampling plans by variables, estimation of lot defective and plan parameter determination in known and unknown cases. Continuous sampling plans – CSP- 1 and its modifications. Derivation of AOQL for CSP-1, operations of MLP's and Wald-Wolfowitz plans.

Unit 5: Implementing six sigma. Six sigma overview and implementations- smarter six sigma solutions(S4) measurements : converting defect rates (DPMO or PPM) to six sigma quality levels- six sigma relationships- six sigma assumptions- S4 assessment- basic control charts and S4 assessments- examples.

Books for study :

1. Duncan A.J. – Quality Control and Industrial Statistics (2010), 2nd edition, Homewood,
2. I. Grant, E.L. and R.S. Leaven worth (2004)- Statistical Quality Control, 2nd edition, Mc-Graw Hill Book Co.
3. Montogomery D.C (2007).- Introduction to Statistical Quality Control, John Wiley
4. Shilling, E.G.(2010) – Acceptance Sampling in Quality Control, second edition, A Chapman & Hall book.
5. Juran, J.M. and Gryana, F.M (2008).-Quality planning and analysis, Tata Mc-Graw Hill.

Books for reference:

1. Wetherill, G.B. (1977) - Sampling Inspection and Quality Control, Halsted Press, N.Y.Ott, E.R.-Process Quality Control, Mc-Graw Hill.

2. Forrest W. Breyfogle III (1999)- Implementing Six Sigma: Smarter solutions using statistical methods, John Wiley and Sons, Inc.

16PST3MC04 STATISTICS LAB – III

SEMESTER	**III**	**CREDITS**	**2**
CATEGORY	**MC(L)**	**NO.OF HOURS/ WEEK**	**4**

Objectives:

1. To enable students to use statistical packages for analyzing Multivariate Data, Stochastic process and Quality Control.
2. To expose students to interpret the output and provide recommendations

Multivariate Analysis

1. Computation of Means, Variances, Covariance and Correlations from a Multivariate dataset.
2. Computation of Partial correlation coefficients from the Var-Cov matrix of a multivariate normal population.
3. Computation of Multiple Correlation coefficients from the Var-Cov matrix of a multivariate normal population.
4. Tests for significance of correlation coefficient using samples from multivariate normal populations– Simple Correlation, Partial correlation and Multiple correlation coefficients.
5. Applications of T2 Statistics to different situations – Test for mean of a single MV normal population, Test for equality of mean vectors of two MV normal populations with equal var-cov matrices & unequal var-cov matrices, Special Applications.
6. MANOVA – One-way & Two-way models.
7. Principal component analysis.
8. Factor Analysis
9. Canonical Correlation Analysis
10. Fishers Discriminant Analysis – Two populations, several populations. Classification with Prior Probabilities.
11. Cluster Analysis – Hierarchical method with different linkages, K-Means method.

Statistical Process Control

1. Control Chart for Attributes – p-chart, np- chart, c-chart, u-chart.
2. Control charts for variables – X -chart, R-chart, S-chart.
3. Process capability computations.
4. Special charts – Moving range chart, CUSUM charts, Exponentially Weighted Moving Average charts.
5. OC Curves for various charts.
6. Single Sampling Plan – OC curve, AOQ curve, ATI curve.
7. Double Sampling Plan – OC curve, ASN curve, AOQ curve, ATI curve

Stochastic Process:

1. Detemination of Pn for a transition probability matrix.
2. Determination of stationary distribution.
3. Generating a Poisson process; waiting time distribution.
4. Extinction probability in a branching process.

16PST3ES01 ADVANCED OPERATIONS RESEARCH

SEMESTER	**III**	**CREDITS**	**3**
CATEGORY	**ES(T)**	**NO.OF HOURS/ WEEK**	**4**

Objectives:

1. To introduce students the statistical and mathematical formulations for handling a range of business based problems.
2. To develop a broad appreciation of different types of decision-making environments.

Unit 1: General Linear programming problem-Formulation-Solution through Graphical, Simplex, Big-M and Two phase Methods – Duality in Linear programming – Goal programming problem.

Unit 2: Non-linear programming-Kuhn Tucker theorem-Lagrangian multipliers method-Wolfe's and Beale's algorithm for solving Quadratic programming problems.

Unit3: Inventory control: Deterministic Models – Economic Order Quantity – Problems with no shortages – The fundamental EOQ Problems, EOQ problems with several production runs of unequal length – Problems with price breaks – One price break, More than one price break - Probabilistic models – Single Period Problem without set-up cost – I and II.

Unit 4: Queuing theory- (M/M/1): (GD/∞/∞), (M/M/1): (GD/N/∞), (M/M/C): (GD/∞/∞), (M/M/C): (GD/N/∞).

Unit 5: Integer Programming-Branch and Bound and Cutting plane methods- Dynamic Programming - Solution of LPP by DPP. Simulation- Formulating and Implementing a Simulation model.

Books for study:

1. Hiller, S.F. and Lieberman J.G. (2000)- Operations Research, CBS Publishers & Distributors, New Delhi.
2. Hadley, G. (1997)- Non-Linear Programming and Dynamic Programming, Addison- Wesley, New York.
3. Nirmal Singh Kambo (1982)- Mathematical Programming Techniques, East-West press.

Books for reference:

1. Philips, D.T. and Ravindra,A.& Solberg, J. (1976)- Operation Research, Principles &Practice, John Wiley, New York.
2. Taha, H.A. (1999) – Operations Research-An Introduction, Macmillan Publishing, Company, New York.
3. Wagner (1973)- Principles of Operations Research: with applications to managerial decisions, Prentice Hall of India, New Delhi.

16PST3ES02 NON-PARAMETRIC METHODS

SEMESTER III CREDITS 3
CATEGORY ES(T) NO.OF HOURS/ WEEK 4

Objectives:

1. To explore knowledge in the advanced methods of non-parametric inference.

1. To derive inference for samples drawn from distribution free population

Unit 1: Introduction, Run test for randomness. $\chi 2$ goodness of fit test .Kolmogrov – Smirnov one sample test, .Kolmogrov – Smirnov two sample test, Binomial test , Point estimator and confidence interval for probability of success.

Unit 2: One sample location problems – Wilcoxon signed rank test. Fishers sign test. Asymptotic test of symmetry – Estimators and confidence interval.

Unit 3: Two sampleproblems – Wilcoxon rank sum test for location parameter (Mann – Whitney).

Test for dispersion parameter – Rank test , Rank like test (Moses), Millers asymptotic test based on Jackknife .

Unit 4: One way layout – Kruskal Wallis test. Test for ordered alternatives , Multiple comparison based on Kruskal Wallis rank sums .Two way layout - Friedman's rank sums test. Test for ordered alternatives , multiple comparisons.

Unit 5: Kendals test for independence. Theil's test for regression coefficients .Hollander's test for parallelism of two regression lines.

Books for Study :

1. Gibbons (2003), Non parametric Statistical Inference,McGraw –Hill Kogakusha, Ltd.
2. Hollander Myles & Wolfe D.A.(1973) , Non parametric Statistical Methods, John Wiley & Sons.

Book for Reference:

1. Rohatgi.V.K. (2011), An introduction to probability theory and Mathematical Statistics, John Wiley & Sons.

16PST3ID01 STATISTICAL AND MATHEMATICAL COMPUTING

SEMESTER	**III**	**CREDITS**	**5**
CATEGORY	**ID(T)**	**NO.OF HOURS/ WEEK**	**4**

Objective :

To introduce the software skills to apply the statistical and mathematical tools.

Unit 1: Python Statements and Comments, Keywords and Identifier, Python Data types, Python I/O and Import , Python operators , Precedence and Associativity, If else statement, For loop, While loop, Break and Continue, Pass, Looping techniques

Unit 2: Function, Argument, Recursion, Modules, Python package, Data types, Numbers , List, Tuple, String, Set, Dictionary , File Operation , Directory , Exception , Exception Handling , User-defined exception , Namespace , Class, Inheritance, Multiple Inheritance, Operator overloading

Unit 3: Data visualization in Python, Fitting of Distributions – Graphical and Statistical Procedures , Statistical Measures, User defined functions for Parametric tests and Non parametric tests , Fitting Statistical Models - Multiple Linear regression, Binary logistic regression.

Unit 4: The MATLAB environment , basic commands, data types, variables, Assignment statement, mathematical operators, Managing workspace, Handling of Arrays, Matrix operations and analysis, program structures - if statement, for loop, while loop, break statement , Creating-saving and running m-files, Functions based on arguments, Nested functions, File I/O handling, Debugging techniques

Unit 5: Elementary Mathematics - Trigonometry, exponentials ,logarithms, rounding, remainders, descriptive statistics, Polynomials, Data visualization - 2D and 3D plotting, Random number generation, Integration , Double Integration, Differentiation, Partial Differentiation, Matrix Operations, Linear Equations, Eigen values and Eigen vectors, Matrix Analysis Books for Reference

1. Python programming for Absolute Beginner, Third Edition By Michael Dawson – Cengage Learning
2. Matlab A practical introduction to programming and problem solving, Third edition By Stormy Attaway, Elsiver, Butterworth Heinemann Publication

Books for Study

1. Python Data Analytics, Fabio Nelli – Apress
2. Python for Data Analysis , Wes McKinney – O'Reilly
3. Modeling Techniques in Predictive Analytics By Thomas W.Miller, Pearson Education
4. Matlab with applications to engineering, physics and finance By David Baez-Lopez, CRC Press
5. Solving Applied Mathematical Problems with MATLAB By Dingyu Xue Yangquan chen, CRC Press

16PST3ID02 STATISTICAL AND MATHEMATICAL COMPUTING

SEMESTER	**III**	**CREDITS**	**5**
CATEGORY	**ID(L)**	**NO.OF HOURS/ WEEK**	**2**

Objective:

To introduce the software skills to apply the statistical and mathematical tools.

Data Analysis using Python

1. Importing and Exporting Datasets
2. Sub setting Dataset
3. Aggregating dataset
4. Stacking and Merging dataset
5. For and While loop
6. Diagrammatic representation
7. Parametric Tests
8. Non-parametric tests
9. Multiple Linear Regression
10. Binary Logistic Regression

MATLAB environment

1. MATLAB commands
2. Matrix operations
3. Eigen values and Eigen vectors
4. Evaluation of Integral
5. Differentiation
6. Convergence of Sequence
7. Convergence of Infinite Series
8. 2D and 3D Plots

Books for Reference

1. Python programming for Absolute Beginner, Third Edition By Michael Dawson – Cengage Learning
2. Matlab A practical introduction to programming and problem solving , Third edition By Stormy Attaway, Elsiver, Butterworth Heinemann Publication

Books for Study

1. Python Data Analytics, Fabio Nelli – Apress
2. Python for Data Analysis , Wes McKinney – O'Reilly

16PST4MC01 APPLIED EXPERIMENTAL DESIGNS

SEMESTER	**IV**	**CREDITS**	**5**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**6**

Objectives:

1. To provide both basic and advanced experimental designs applied in Agriculture, Pharmaceutical, Industrial and Biological sciences.
2. To derive solutions for statistical inference problems

Unit 1: Review of Linear models – Block Design, C-matrix and its properties- Analysis of block design – (CRD) completely Randomized design – (RBD)- Randomized Block Design – (LSD)- Latin Square Design – (RLSD) Repeated Latin Square Design – Missing plot techniques – ANOCOVA.

Unit 2: Factorial Design – 2n; 3n factorial designs. Finite fields and design of experiments. Partial confounding and complete confounding – confounding in more than two blocks. Fractional factorials – construction and analysis-concept of resolution plans.

Unit 3: Asymmetrical factorial designs (AFD)- AFD- confounded asymmetrical factorial design construction of balanced confounded asymmetrical factorials-split andstrip-plot experiment.

Unit 4: Incomplete block designs – varietal Trials – incomplete block design balanced incomplete block designs (BIBD) construction of BIBD – analysis of BIBD, Youden square design – Lattice designs. Partially balanced incomplete block design (PBIBD) – analysis and construction of PBIBD - Group divisible-simple-triangular- Latin square type and cyclic PBIBD.

Unit 5: Orthogonal Latin square – maximum number of orthogonal Latin squares – construction of orthogonal Latin squares – construction of BIBD using orthogonal Latin squares. Response surface designs- definition of response surface design – first order and second order response surface design.

Books for study:

1. Das, M.N. and Giri,N. (2008). Design and Analysis of Experiments, Wiley Eastern.
2. Federer, W.T. (1993). Experimental Designs – Theory and Applications, McMillan.

Books for reference:

1. Joshi,D.D (1987). Linear estimation and design of experiments. Wiley Eastern.
2. Kempthorne, O. (2000). Design and Analysis of Experiments, Wiley Eastern.
3. Montgomery, D.C.(2012) Design and Analysis of Experiments, John Wiley & Sons.

16PST4MC02 DATA WAREHOUSING AND DATA MINING

SEMESTER	**IV**	**CREDITS**	**5**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**6**

Objectives:

1. To recourse knowledge discovery through database which leads to Business Intelligence.
2. To illustrate the applications of data mining and statistical tools

Unit 1 : Data Warehouse Basics: Definition of a Data Warehouse - Basic Elements of the Data Warehouse - Data Warehouse and OLTP Database Design, Differences - Data Warehouse Features - Manage Data, Decision Support System (DSS) - Data Warehousing Process - Comparing Warehouses and Data Marts - Dependent Data Mart Model - Independent Data Mart Model - Enterprise Model Architecture. Defining the Business and Logical Models: Documenting Business Measures and Documenting Business Dimensions Creating the Metadata -Designing the Dimensional Model: Data Warehouse Database Design Objectives - Data Warehouse Data Types - Star Dimensional Modeling - Fact Tables - Dimension Tables

Unit 2 : Translating Business Dimensions into Dimension- Star Dimensional Model Characteristics - Snowflake Model - Designing the Physical Model - Translating the Dimensional Model into a Physical Model.-Storage Considerations for the Physical Model- Database Sizing - Estimating the Database Size - Indexing Types - B*tree Index, Bitmap Indexes - Partitioning Tables and Indexes, Strategies for Extracting, Transforming, and Transporting: Extraction – Transformation - and Transportation Process (ETT) - Data Staging Area - Extracting Data , -Examining Source Systems – Mapping - Designing Extraction Processes - Designing Transformation Processes and ETT Tools

Unit 3 : Naive Bayes Classification Method, Bayesian Networks, Path Analysis, Back Popagation Algorithm, Building Predictive Model using Artificial Neural Network, Support Vector Machine, Decision Tree Methods - Classification Tree, Regression Tree, Decision tree based on Statistical Significance - Chi Square

Automated Interaction Detector(CHAID). Comparing Classifier Accuracy.

Unit 4 : Construction of Gains Chart, ROC Curve, Leave one out validation and N fold validation, Construction of Logit Model Tree, K th Nearest Neighbourhood Classification, Bagging and Boosting Principles, Adaptive Boosting Algorithm, Decision Stumps and Random Forest, Apriori Algorithm and Association Rule Mining

Unit 5 : Additive Regression, Logit Boost, Multi Class Classifier, Ordinal Class Classifier, Expectation Maximization Algorithm, Genetic Algorithm, Combining Classifiers, Cost Sensitive Classifier, Text Mining– Methods and Models

Books for Study:

1. Anahory S, Murray D(2001) Data Warehousing In The Real World: Practical Guide For Building Decision Support Systems, Addison Wesley.
2. Han J and Kamber M (2002) , Data Mining concepts and Techniques, Morgan Kaufmann Publishers (Only relevant sections), SRI ESWAR ENTERPRISES

Books for Reference:

1. Pieter Adriaans and DolfZantinge (2000), Data Mining, Addison Wesley.
2. PujariA..K (2001) Data Mining Techniques, University Press, Hyderabad.

16PST4MC03 BIOSTATISTICS AND SURVIVAL ANALYSIS

SEMESTER	**IV**	**CREDITS**	**5**
CATEGORY	**MC(T)**	**NO.OF HOURS/ WEEK**	**6**

Objectives:

1. To develop sound judgment about data applicable to clinical care
2. To emphasize study design and interpretation of results of medical research

Unit 1: Introduction to Medical research – Study Designs – Observational studies; Experimental Studies & Clinical trials; Meta Analysis, Research questions about mean of a group & proportions in a group. Repeated measures design – Paired comparison of means; Kappa statistic; McNemar test. Sign test for median; Wilcoxon Signed Rank test.

Unit 2: Research questions about means and variances of two groups; Levene test; Wilcoxon Rank Sum test. Decisions on proportions in two groups – z-test; Chi-Square test. Diagnostic Procedures with Threshold model. Measuring the accuracy of diagnosis – Sensitivity, Specificity; ROC curve.

Unit 3: Analytical Estimation procedures for survival distributions- The Exponential Distribution, Weibull Distribution, Lognormal Distribution, Gamma Distribution. Hazard Plotting. Relative, Corrected Survival Rates, Standardised Rates and Ratios.

Unit 4: Kaplan Meier Survival Curve, Life Table Analysis, Comparison of Survival Distribution – Log Rank Test for comparing two groups, Log rank test for comparing n-groups.

Unit 5: Cox Proportional Hazard Model, Meaning of PH Assumption, ML Estimation of Cox Proportional PH Model, Adjusted Survival Curves using Cox PH Models, Evaluating the Proportional Hazard Assumption – Graphical Approach, Goodness of fit test approach,

Book forStudy and Reference:

1. Dawson, Beth & Robert G. (2001): Basic & Clinical Biostatistics. McGraw-Hill
2. Daniel, Wayne W. (1995): Biostatistics: A Foundation For Analysis in the Health Sciences – 6th edn. John Wiley & Sons.

Book for Reference:

1. David G.Kleinbaum (1996): Survival Analysis, Springer
2. Elisa.T.Lee (1992) Statistical Methods for Survival Data Analysis , 2e, John Wiley and Sons

16PST4MC04 STATISTICS LAB – IV

SEMESTER	**IV**	**CREDITS**	**2**
CATEGORY	**MC(L)**	**NO.OF HOURS/ WEEK**	**4**

PG Statistics Category: MC(Practical)

Credits : Semester III Hours/Week : 4

Objectives:

1. To Provide Practical Knowledge in Analysing problems in Design of Experiments, Operations Research, Biostatistics and Survival Analysis
2. To demonstrate hands on experience for problems using statistical software

Design of Experiments

1. Complete Randomized Design, Randomized Block Design, Latin Square Design
2. Balanced Incomplete Block Design, Split Plot Design
3. Factorial Design - 22, 23, 24 , 32, 33, 34
4. Construction of Contours, Response Surface Methods.
5. ANOCOVA

Data mining Techniques

1. Segmentation using CART and CHAID
2. Classification and prediction using Random Forest
3. Predictive model building using ANN, SVM and KNN
4. Market Basket Analysis
5. Prediction using Adaboost and Bayesian Network

Biostatistics and Survival Analysis

1. Paired comparison of means; Kappa statistic; McNemar test. Sign test for median
2. Wilcoxon Signed Rank test, Levenetest, Wilcoxon Rank Sum test, Chi-Square test
3. Survival Function, Hazard Function, Kappa Statistic, Relative Risk, Likelihood Ratio

4. Odds Ratio, Sensitivity and Specificity, Mortality Rates, Adjusted Rates,
5. Kaplan – Meier Survival Curves, Life Table Method
6. Log Rank Test, Wilcoxon Test, Likelihood Ratio Test
7. Cox Proportional Hazard Model, Cox Model with time dependent Covariates