ST 1820 ADVANCED DISTRIBUTION THEORY

Semester I **Category: MC**
Hours/Week : 6 **Credits : 5**

Objective : To enable students with the necessary knowledge towards constructing models applicable to real life situations

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. Characterisations 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. Generalised 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 and reference

Hogg, R.V.and Craig, A.T.(2002). Introduction to Mathematical Statistics . *Pearson Education, Asia.*

Johnson, N.L. and Kotz, S.(1972). Distributions in Statistics.Vol. 1 – 4. *John Wiley and Sons, New York* .

Johnson,N.L. Kotz,S and Balakrishnan,N(1997). Discrete Multivariate Distributions. *John Wiley and Sons, New York.*

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

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

Johnson, N.L., Kotz, S.and Kemp, A.W (1992). Univariate Discrete Distributions . *John Wiley and Sons, New York.*

Rohatgi, V.K. and Saleh, A.K.Md.E (2002). Introduction to Probability and Statistics, *Pearson Education, Asia.*

ST 1821 APPLIED REGRESSION ANALYSIS

Semester I **Category: MC**
Hours/Week : 6 **Credits : 5**

Objective: 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:

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

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).

* * * * * * *

ST 1822 STATISTICAL MATHEMATICS

Semester I **Category: MC**
Hours/Week : 6 **Credits : 4**

Objective: To give the necessary mathematical foundation for post-graduate studies in statistics.

Unit – I: Real Sequences – boundedness , monotonicity , convergence. Operations on sequences. Infinite series – convergence – tests for convergence

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

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

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

Unit – V: 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 (1971): *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. Burkill, J. C. (1962): *A First Course in Mathematical Analysis* – [Cambridge University Press]

2. Chakrabarti, A (2006): *A First Course in Linear Algebra* – [Vijay Nicole Imprints Pvt. Ltd]

3. Goldberg, R. R. (1970): *Methods of Real Analysis* – [Oxford & IBH Publishers]

4. Hadley, G. (1987) : *Linear Algebra* – [Narosa Publishing House]

* * * * * * *

ST 1823 FUNDAMENTALS OF DATA BASE ADMINISTRATION

Semester I **Category: MC**
Hours/Week : 6 **Credits : 3**

Objective : To introduce the concepts and practice of data base administration for purposes of analyzing large scale statistical data sets.

Unit 1: Client Server Architecture – Three Tier Architecture – Application Server Components – SQL Basics that covers creating database objects, stosing, retrieving and manipulating data in a relational database – Advanced SQL – SQL* Loader

Unit 2: PLSQL fundamentals, language structures, flow of execution and interface with SQL – Stored Procedures, Functions, Packages and Triggers.

Unit 3: Oracle Architectural Components, Creating a database & creating data dictionary – Views and Standard Packages.

Unit 4: Managing Table Spaces – Maintaining Daat files, Control files, Redo Log Files – Storage structure and relationships – Managing rollback segments, Tables and indexes, Data Integrity.

Unit 5: Managing PAsswaord Security and resources – Users, Roles and Privileges.

Books for study and reference

Christopher Allen. ORACLE Database 10g PL / SQL 101. *Tata McGrawHill Publishers*

Gavin Powell. Beginning Database Design. *Wiley Publication Inc.*

Ivan Bayross. SQL, PL?SQL The Programming Language of ORACLE. *BPB Publications*.

Kevin Loney. Oracle Databse 10g : The Complete Reference.

Lakshman Bulusu. ORACLE PL / SQL Programming. *CENGAGE Learning*.

Rick Greenwald, Robert Stackowiak, Gary Dodge, Daviv Klein, Ben Shapiro, Christopher G. Chelliah. Professional ORACLE Programming. *Wiley Publication Inc.*

ST 1824 STATISTICS LAB – I

Semester I **Category: MC (Practical)**
Hours/Week : 6 **Credits : 3**

Objective: To Provide Statistical Model Building Knowledge through R-Programming and Develop Model Building Skills by analysing Real life Problems.

Basics of R Programming

- Importing and Exporting Datasets
- Usage of Inbuilt Mathematical, Statistical and Special Functions in R
- Subset Observations , Variables from Datasets
- Creating new variables from existing variables
- Merging Datasets – Inner Joins, Outer Join, Left Outer Join, Left Inner Join
- Coding user defined functions in R Language

Linear and Matrix Algebra

- Linear Dependency and Linear Independency
- Determining Matrix Determinant and Matrix Inverse
- Determining Characteristic Roots and Characteristic Vectors
- Determining Rank, Index and Signature of a Quadratic form

Advanced Distribution Theory

- Fitting of Binomial and Poisson Distribution
- Fitting of Truncated Binomial and Truncated Poisson Distribution
- Fitting of Mixture of Geometric and Poisson
- Fitting of Mixture of Two Poisson
- Generating Samples from Discrete and Continuous Distributions

Applied Regression Analysis

- Building Linear Regression Model with Categorical Explanatory Variable
- Testing for Overall Model fit and Individual Regression Coefficients
- Determining R-Square, Adjusted R-Square, MAE and MAPE

- Study of Interaction Effects among Explanatory variable
- Detection of Outliers by Standardised and Studentized Residuals
- Testing for Multicollinearity using VIF and Conditional Index
- Transformation and Combining Variables to deal Multicollinearity
- Residual Analysis – Testing for Normality of Residuals, Transformation
- Residual Analysis – Testing Functional form of explanatory variables
- Testing for Homoscadasticity, Weighted Least Squares Regression
- Model Validation, Bootstrap Regression Methods

ST 2814 ESTIMATION THEORY

Semester II **Category: MC**
Hours/Week : 5 **Credits : 5**

Objective: To provide a strong theoretical foundation in Statistical Estimation Theory.

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 and reference:

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.and Casella,G.(1998).Theory of point estimation. Springer-Verlag**
4. **Rohatgi,V.K. and Saleh A.K.Md.E.,(2002).An Introduction to Probability and Statistics. John Wiley and Sons,N.Y.**
5. **Zacks, S. (1971). The Theory of Statistical inference. John Wiley and Sons,N.Y.**

ST 2815 : TESTING STATISTICAL HYPOTHESES

Semester II **Category: MC**
Hours/Week : 5 **Credits : 5**

Objective: To impart knowledge on statistical modeling for decision support based on sample characteristics.

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 and reference:

1. Kale, B.K.(2005) A first course on parametric inference, Narosa publishing house, New Delhi.
2. Kendall, M.G. and Stuart, A. (1967) The Advanced theory of Statistics, Vol-2, Hafner publishing co., New York.
3. Lehmann, E.L. (1986) Testing Statistical Hypotheses, John Wiley and sons, NY.
4. Rohatgi, V.K. and Saleh, E.A.K. Md.(2002)An Introduction to Probability and Statistics, John Wiley and sons, NY..

ST 2816 SAMPLING THEORY

Semester II **Category: MC**
Hours/Week: 5 **Credits: 5**

Objective: To impart knowledge on various sampling techniques useful in survey methodology.

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 and Reference:

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, New Delhi.
4. Murthy.M.N.(1967).Sampling Theory and Methods. Statistical Publishing Society, Calcutta.
5. Sampath,S.(2005). Sampling Theory and Methods. Narosa Publishing House.
6. **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**

ST 2817 CATEGORICAL DATA ANALYSIS

Semester II **Category: MC**

Hours/Week : 5 **Credits : 3**

Objective: (i) To introduce students to the exciting new area of analysis of categorical data

(ii) 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*

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

ST 2818 STATISTICS LAB – II

Semester II **Category: MC**
Hours/Week : 4 **Credits : 2**

Objective: To Provide hands on experience in implementation of concepts in Estimation Theory, Testing of Statistical Hypothesis and Categorical Data Analysis

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

- Basic Parametric and Non Parametric Tests
- Determining Test Function, Level and Power of Test Function, Power Curve
- MP Test, Uniformly Most Powerful Test, Uniformly Most Powerful Unbiased test

Sampling Theory

- Simple Random Sampling With and Without Replacement
- Stratified Random Sampling, Midzuno Sampling
- Horwitz – Thompsonn Estimator
- Hansen – Horwitz Estimator
- Desraj Ordered Estimator
- Calculation of Inclusion Probabilities in Fixed and Varying Sampling Designs

Categorical Data Analysis

- Building Binary Logistic Regression Model

- Obtaining Estimated Probability, Optimal Cut Point
- Deriving Classification Table, Sensitivity and Specificity, AUC Measures
- Model Validation through AUC and Gains Chart
- Probit Model
- Multinomial Logistic Regression Model
- Cumulative / Ordinal Logistic Regression Model
- Poisson Regression Model

ST 2959 MODERN PROBABILITY THEORY

Semester II **Category: Elective (ES)**
Hours/Week : 4 **Credits : 3**

Unit –I: 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 – II: 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 – III : Modes of Convergence of Random Variables : Limits of Random Variables, Convergence in Probability, Convergence Almost Surely, Convergence in Distribution, Convergence in r^{th} mean,

Unit –IV : Laws of Large Numbers: Weak and Strong law of large Numbers

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

Book for Study:

Bhat,B.R.(1985) Modern Probability Theory, 3 Ed. *New Age International Publishers.,*

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*

* * * * * * *

ST 2960 ACTUARIAL STATISTICS

Semester II **Category: Elective (ES)**

Hours/Week : 4 **Credits : 3**

Objective: To educate the students about the applications of statistics in insurance industry.

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 reference:

Bowers N.L.,.Gerber H.U,.Hickman, J.C and Nesbitt, C.J.(1997) Acturial Mathematics,

Society of Actuaries, Itasca, USA second edition.

Dixit. S.P..Modi .C.S and Joshi R.V. (2000) Mathematical basics of Life Assurance,

Insurance Institute of India, Bombay.

Donald, D.W.A.(1970) Compound Interest and annuities. Heinemann, London

Mccutcheon J.J.and Scot (1989).Mathematics of Finance, Heinemann, London

Neil, A (1977)..Life contingencies, Heinemann, London

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

ST 3815 MULTIVARIATE ANALYSIS

Semester III **Category: MC**

Hours/Week : 6 **Credits : 5**

Objective: To provide students the requisite knowledge and skills to handle multi-dimensional data and extract useful information from the data.

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. Dsitributions of Sample correlation coefficient, partial correlation coefficient & Multiple correlation coefficient. The Generalised T^2 Statistic – 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 – 3rd edn. *John Wiley & Sons*.
2. Johnson, R. A., & Wichern, D. W. (2007): Applied Multivariate Statistical Analysis – 2nd edn. *Prentice Hall International*

ST 3816 STOCHASTIC PROCESSES

Semester III **Category: MC**
Hours/Week: 6 **Credits: 5**

Objectives: i) To introduce basic concepts in Stochastic Processes.

ii) To expose several processes in disciplines like Biology, Engineering and Economics.

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 and Reference.

1. Karlin, S and Taylor, H.M.(1975). A first course in Stochastic Processes. Academic Press, New York.
2. Karlin, S and Taylor, H.M.(1981). A second course in Stochastic Processes. Academic Press, New York.
3. Medhi, J(1996). Stochastic Processes, Wiley Eastern Limited.
4. Ross, S.M.(1983). Stochastic Processes, John Wiley and Sons.

ST 3817 STATISTICAL QUALITY CONTROL

Semester III **Category: MC**

Hours/Week : 4 **Credits : 3**

Objective : To train students in modern statistical process control techniques.

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(S^4) measurements : converting defect retes(DPMO or PPM) to six sigma quality levels- six sigma relationships- six sigma assumptions- S^4 assessment- basic control charts and S^4 assessments- examples.

Books for study and reference

Duncan A.J. – Quality control and industrial statistics, 2nd edition, Homewood, I. Grant, E.L. and R.S. Leaven worth- statistical quality control, 2nd edition, Mc-Graw Hill Book Co.

Shilling, E.G. – Acceptance Sampling in Quality Control

Juran, J.M. and Gryana, F.M.- quality planning and analysis, Tata Mc-Graw Hill.

Wetherill, G.B.- sampling Inspection and Quality Control, Halsted Press, N.Y.

Ott, E.R.- Process QualityControl, Mc-Graw Hill.

Montogomery D.C.- Introduction to statistical quality control, John Wiley

Handbook of statistics, Indian statistical institute

Forrest W. Breyfogle III(1999)- implementing six sigma: smarter solutions using statistical methods, john Wiley and Sons, Inc.

ST 3818 STATISTICS LAB – III

Semester III **Category: MC (Practical)**
Hours/Week : 4 **Credits : 2**

Objective: To enable students to use statistical packages for analyzing Multivariate Data, Data on Stochastic Processes and on Quality Control.

Multivariate Analysis

1. Computation of Means, Variances, Covariances 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 T^2 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 – $\overline{X}$ -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

ST 3957 DATA WAREHOUSING AND DATA MINING

Semester III **Category: Elective (ES)**

Hours/Week : 4 **Credits : 3**

Objective: To recourse knowledge discovery through database which leads to Business Intelligence.

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, Constructionof 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 and Reference

Anahory S, Murray D(2001) Data Warehousing In The Real World: Practical Guide For Building Decision Support Systems, Addison Wesley.

Han J and Kamber M (2002) , Data Mining concepts and Techniques, Morgan Kaufmann Publishers (Only relevant sections)**,** SRI ESWAR ENTERPRISES

Pieter Adriaans and Dolf Zantinge (2000), Data Mining, Addison Wesley**.**

Pujari A..K (2001) Data Mining Techniques, University Press, Hyderabad.

ST 3958 NON-PARAMETRIC METHODS

Semester III **Category: Elective (ES)**

Hours/Week : 4 **Credits : 3**

Unit 1

Introduction, Run test for randomness. χ^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 sample problems – 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 references :

1. Gibbons (1971), Non parametric Statistical Inference,

   McGraw –Hill Kogakusha, Ltd.

2. Hollander Myles & Wolfe D.A.(1973) , Non parametric Statistical Methods, John Wiley & Sons.

3. Rohatgi.V.K. (1976), An introduction to probability theory and Mathematical Statistics, John Wiley & Sons.

ST 4813 APPLIED EXPERIMENTAL DESIGNS

Semester IV **Category: MC**
Hours/Week : 6 **Credits : 5**

Objectives: To provide both basic and advanced experimental designs applied in Agriculture, Pharmaceutical, Industrial and Biological sciences.

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 – 2^n; 3^n 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 and strip-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 and reference:

1. Das, M.N. and Giri,N. (2008). Design and Analysis of Experiments, Wiley Eastern.
2. Ferderer, W.T. (1993). Experimental Designs – Theory and Applications, McMillan.
3. Joshi,D.D (1987). Linear estimation and design of experiments. Wiley Eastern.
4. Kempthorne, O. (2000). Design and Analysis of Experiments, Wiley Eastern.
5. Montgomery, D.C.(2012) Design and Analysis of Experiments, John Wiley & Sons.

ST 4814 ADVANCED OPERATIONS RESEARCH

Semester IV **Category: MC**

Hours/Week : 6 **Credits : 5**

Objectives: i) To introduce students the statistical and mathematical formulations for handling a range of business based problems.

ii) 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.

Unit 3: 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 and reference

1. Hadley, G. (1997)- Non-Linear Programming and Dynamic Programming, *Addison- Wesley, New York.*
2. Hiller, S.F. and Lieberman J.G. (2000)- Operations Research, *CBS Publishers & Distributors, New Delhi.*
3. Nirmal Singh Kambo (1982)- Mathematical Programming Techniques, *East-West press.*
4. Philips, D.T. and Ravindra, A.& Solberg, J. (1976)- Operation Research, Principles & Practice, *John Wiley, New York.*
5. Taha, H.A. (1982) – Operations Research-An Introduction, *Macmillan Publishing, Company, New York.*
6. Wagner (1973)- Principles of Operations Research: with applications to managerial decisions, *Prentice Hall of India, New Delhi.*

ST 4815 BIOSTATISTICS AND SURVIVAL ANALYSIS

Semester IV **Category: MC**
Hours/Week : 6 **Credits : 5**

Objective: (i) To develop sound judgment about data applicable to clinical care
(ii) 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 for Study and Reference:

Daniel, Wayne W. (1995): Biostatistics: A Foundation For Analysis in the Health Sciences – 6th edn. *John Wiley & Sons*.

David G.Kleinbaum (1996): Survival Analysis, Springer

Dawson, Beth & Robert G. (2001): Basic & Clinical Biostatistics. *McGraw-Hill*

Elisa.T.Lee (1992) Statistical Methods for Survival Data Analysis , 2e, John Wiley and Sons

ST 4816 STATISTICS LAB – IV

Semester IV **Category: MC (Practical)**
Hours/Week : 4 **Credits : 2**
Objective: To Provide Practical Knowledge in Analysing problems in Design of Experiments, Operations Research, Biostatistics and Survival Analysis

Design of Experiments

- Complete Randomized Design, Randomized Block Design, Latin Square Design
- Balanced Incomplete Block Design, Split Plot Design
- Factorial Design - 2^2, 2^3, 2^4 , 3^2, 3^3, 3^4
- Construction of Contours, Response Surface Methods.
- ANOCOVA

Advanced Operations Research

- Graphical Method, Simplex Method - Linear Programming Problems
- Branch and Bound , Cutting Plan Method – I.P.P
- Queueing Theory – (M|M|1) : (GD|∞|∞), (M|M|1);(GD|N|∞),

  (M|M|C): (GD|∞|∞), (M|M|C): (GD|N|∞)
- Wolf's Algorithm – Beales Algorithm

Biostatistics and Survival Analysis

- Paired comparison of means; Kappa statistic; McNemar test. Sign test for median
- Wilcoxon Signed Rank test, Levene test,Wilcoxon Rank Sum test,Chi-Square test
- Survival Function, Hazard Function, Kappa Statistic, Relative Risk, Likelihood Ratio
- Odds Ratio, Sensitivity and Specificity, Mortality Rates, Adjusted Rates,
- Kaplan – Meier Survival Curves, Life Table Method
- Log Rank Test, Wilcoxon Test, Likelihood Ratio Test
- Cox Proportional Hazard Model, Cox Model with time dependent Covariates