DEPARTMENT OF STATISTICS
SARDAR PATEL UNIVERSITY
VALLABH VIDYANAGAR

SYLLABUS EFFECTIVE FROM: 2017-18

MASTER OF SCIENCE (APPLIED STATISTICS)
Course Syllabus and Structure of for M.Sc. (Applied Statistics)
Semester – III

PS03CAST21: KNOWLEDGE DISCOVERY AND DATA MINING

Unit-I	Review of classification methods from multivariate analysis, classification and decision trees. Clustering methods from both statistical and data mining view points.	[10]
Unit-II	Unsupervised learning from univariate and multivariate data, dimension reduction and feature selection.	[10]
Unit-III	Supervised learning from moderate to high dimensional input spaces, artificial neural networks and extensions of regression models, regression trees	[10]
Unit-IV	Introduction to databases, including simple relational databases, data warehouses and introduction to online analytical data processing. (Revision) Association rules and prediction, data attributes, applications to electronic commerce	[10]

References

1. Berson, A. and Smith, S.J. (1997). Data Warehousing, Data Mining, and OLAP. McGraw-Hill.
2. Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). Classification and Regression Trees. Wadsworth and Brooks/Cole.
3. Han,J. and Kamber.M. (2000). Data Mining; Concepts and Techniques. Morgan Kaufmann.
4. Mitchell, T.M. (1997). Machine Learning. McGraw-Hill.
5. Ripley, B.D. (1996). Pattern Recognition and Neural Networks. Cambridge University Press.

PS03CAST22: OPERATIONS RESEARCH II

Unit-I	Transportation Problem(TP): Introduction, Area of application, mathematical model of TP- maximization and minimization problems, Terminologies used in LPP. Degeneracy,Balanced and unbalanced TP. North-West Corner Method(NWCM), Least Cost Method(LCM), Vogel's Approximation Method(VAM), Modified Distribution Method(MODI)	[12]
Unit-II	Assignment Problem(AP): Introduction, Area of application, mathematical model of AP- maximization and minimization problems ,Hungarian Method, Multiple Optimal Solutions.	[12]
Unit-III	Network Analysis: Introduction, Minimal spanning tree problem, Maximal flow problem. PERT and CPM- Terminologies used, Similarity and Differences, steps in PERT and CPM.	[12]
Unit-IV	Inventory Control-Deterministic and Probabilistic models, Non-linear Programming Problem- Kuhn-Tucker Conditions, Introduction to Simulation Techniques and Sequencing Problems,	[12]

References

1. Kambo, N.S.(1991) Mathematical Programming Techniques Affiliated East-West Press Pvt.Ltd.
2. Hadley,G. (1987) Linear Programming.
3. Taha, H.A.(1992) Operations Research 5th Ed., Macmillan
4. Kanti Swarup, Gupta P. K. and Man Mohan Singh (1977) Operations Research, Sultan Chand & Sons.
5. N. D. Vohra (2011) Quantitative Techniques in Management, 4th Ed., Mc Graw Hill.
6. V. K. Kapoor(1998) Problems & Solutions in Operations Research, 2nd Ed., Sultan Chand & Sons.
7. S. D. Sharma (2001) Operations Research, 13th Ed., Kedar Nath Ram Nath & Co.
8. J. K. Sharma(2009) Quantitative Techniques For Managerial Decisions, 1st Ed. , Macmillan

PS03EAST21: PLANNING AND ANALYSIS OF INDUSTRIAL EXPERIMENTS

Unit-I	A review of basic concepts of design of experiment. Factorial Experiments: Concepts of main effects, interaction, Analysis of full 2^n and 3^2 factorial designs, Confounding: Total and partial confounding. Analysis of 2^n and 3^n confounded design.	10
Unit-II	$2^{(n-p)}$ Fractional Factorial Designs: Basic Idea, Generating the Design,The Concept of Design Resolution, Plackett-Burman Designs for Screening, Enhancing Design Resolution via Foldover, Aliases of Interactions: Design Generators, Blocking, Replicating the Design, Adding Center Points,Analyzing the Results of a $2^{(n-p)}$ Experiment.	10
Unit-III	$3^{(n-p)}$ Fractional Factorial Designs: Overview, Designing $3^{(n-p)}$ Experiments, Box-Behnken Designs, Analyzing the $3^{(n-p)}$ Design, ANOVA, Parameter Estimates.	10
Unit-IV	Central Composite and Non-Factorial Response Surface Designs: Overview, Design Considerations, Alpha for Rotatability and Orthogonality, Available Standard Designs, Analyzing Central Composite Designs, The Fitted Response Surface, Categorized Response Surfaces. Taguchi Methods: Robust Design Experiments: Overview, Quality and Loss Functions, Signal-to-Noise (S/N) Ratios, Orthogonal Arrays, Analyzing Designs, Accumulation Analysis	10

1. Kshirsagar A.M. (1983) Linear Models (Marcel Dekker)
2. John P.W.M.(1971) Linear Models (John Wiley Ltd.)
3. Jeff Wu C. F., Hamada M. (2000): Experiments: Planning, Analysis and parameter design optimization, John Wiley & Sons.
4. Montgomery D.C. (2001): Design and Analysis of Experiments, 5th edition, Wiley New York.
5. Angela Dean and Daneil Voss (1999): Design and Analysis of Experiments, Wiley.
6. Phadke, M.S. (1989): Quality Engineering using Robust Design, Prentice-Hall.

PS03EAST22: GENERALIZED LINEAR MODELS

Unit 1	Review of Linear Statistical Models, Discrete Response Data, Introduction of Generalized Linear Models (GLMs), Components: Linear Predictor, Link Function, Natural Parameters, Scale Parameters; Exponential Family of Distributions (EFD): Members of EFD: Normal, Lognormal, Exponential, Gamma, Binomial, Poisson, Negative Binomial; Steps for Model Fitting, Mean and Variance of EFD; Frequent Inference: Estimation of Parameters through Iteratively Reweighted Least Square (IRLS) and Algorithms, Form of Adjacent Dependent Variable and Weights, Analysis of Deviance, Nested Model and Non-Nested Model; Goodness of Fit Criteria: RSquare, Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), Bayesian Information Criterion (BIC); Step Wise Selection; Testing of Parameters through Wald Test; Confidence Intervals; GLMs Residuals: Residual Analysis, Pearson Residual, Anscombe Residual, Deviance Residual. Model Checking: Hat Matrix, Outlier, Leverage, Influence	12 L
Unit 2	Binary Data: Models for Binary Data: Group Data and Ungroup Data, Linear Predictor, Link Function: Logit, Probit and Complementary Log Log; Prospective Study and Retrospective Study, Likelihood Function, Estimation of Parameters through IRLS Method; Deviance; Probit Model, Residual Analysis Polytomous Data: Introduction of Multinomial Logistic Regression and Ordinal Regression, Examples and their inference. Ordinal Regression Models with Qualitative or/and Quantitative Covariates; Parallel Line Regression	12L
Unit 3	Count Data: Introduction Poisson Regression, Likelihood, Estimation and Testing of Parameters: Log Linear Model for Contingency Table and their Analysis Generalized Linear Models with distribution having Constant Coefficient of Variation; Gamma Family; Canonical Link Function; Inference and Residual Analysis of GLMs with Gamma Distribution; Comparison between Response having Gamma distribution and lognormal distribution; Examples and Applications	12L
Unit 4	Models for Survival Data: Estimation with Censored Survival Data and Survival Distribution: Exponential Distribution; Weibull Distribution and their Examples. Under and Over Dispersion Problem of Data, Quasi Likelihood for Estimation of Parameters, Properties of Quasi Likelihood, Analysis of Deviance; Quasi Likelihood: Binomial, Poisson, Normal, Gamma, Lognormal, Exponential; Comparison of Quasi Likelihood with Likelihood; Concept of Marginal Likelihood, Conditional Likelihood; Models with Nonlinear Parameters in Covariates; Model Checking: Checking Link Function, Checking Covariance Scale, Checking the Variance Function, Score Test for Extra Parameters, Checking Form of Covariates, Detection of Influential Observations	12L

Books Recommended

1 Agresti, A. (2002). Categorical Data Analysis, ED.II, Wiley InterScience

2 Fahrmier,L .and Tutz,G.(2001). Multivariate Statistical Modeling Based on Generalizes Linear Models, Springer

3 Gill, J.(2001).Generalized Linear Models: A Unified Approach, Sage Publication

4 Lindsey, J.K. (1997). Applying Generalized Linear Models , Springer

5 Maindonald,J. And Braun,J. (2007). Data Analysis and Graphics using R: An example based approach Ed.II, Cambridge University Press

6 McCullagh, P. And Nelder,J.A. (1983). Generalized Linear Models- Monographs on Statistics

and Applied Probability, Chapman and Hall
7 Myers, R.H, Montegomery, D.C., Vinning, G.G and Robinson, T.J.(2010). Generalized Linear Models with Applications in Engineering and the Sciences, Ed.II , Wiley Series in Probability and Statistics, A John Wiley & Sons.

PS03EAST23: SURVIVAL ANALYSIS

Unit 1	Concepts of time, order and random censoring. Life distributions – exponential,l gamma, Lognormal, pareto, linear failure rate. Parametric inference, point estimation, confidence Intervals, scores, tests based on LR, MLE.	12L
Unit 2	Life tables, failure rate, mean residual life and their elementary properties. Ageing classes - IFR, IFRA, NBU, NBUE, HNBUE and their duals, Bathtub failure rate. Estimation of survival function - Actuarial estimator, Kaplan - Meier estimator. Estimation under the assumption of IFR/DFR	12L
Unit 3	Semi-parametric regression for failure rate - Cox's proportional hazards model. Competing risk models. Repair models. Probabilistic models. Joint distribution of failure times. Unconditional tests for the time truncated case. Tests for exponentiality, two sample non-parametric problem.	12L
Unit 4	Concept of frailty. Shared frailty models. Identifiability of frailty models. Various frailty models. Gamma, positive stable, inverse Gaussian, power variance function, compound Poisson and compound negative binomial shared frailty models. Frailty regression models. Bivariate and correlated frailty models. Additive frailty models. Reversed hazard rates, Cox's proportional reversed hazards model	12L

Books Recommended

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

2 Deshpande, J.V. and Purohit S.G. (2005). Life Time Data: Statistical Models and Methods, Word Scientific.

3 Duchateau, L. and Johnson, P. (2008). The Frailty Model. Springer: New York

4 Gross A.J. and Clark, V. A. (1975) Survival Distributions: Reliability Applications in the Biomedical Sciences, John Wiley and Sons

5 Hanagal, D. D. (2011). Modeling Survival Data Using Frailty Models. CRC Press: New York.

6 Hougaard, P. (2000). Analysis of Multivariate Survival Data. Springer: New York

7 Wienke, A. (2011). Frailty Models in Survival Analysis, CRC Press: New York