SARDAR PATEL UNIVERSITY
Vallabh Vidyanagar, Gujarat
(Reaccredited with 'A' Grade by NAAC (CGPA 3.25)
Syllabus with effect from the Academic Year 2022-2023

PROGRAMME STRUCTURE
Master of Science in Statistics
MSc (Applied Statistics) Semester: III

Programme Outcome (PO) - For MSc Applied **Statistics** Programme	Master of Science program provides extended theoretical and practical knowledge of different science subjects. Master of Science programme at Sardar Patel University is designed keeping the overall back ground preparation in mind for the student to either seek a job or to become an entrepreneur. The students, after completion of Bachelor of Science can select the master's programme in the subject they have had at the final year or in a related discipline (depending upon eligibility criteria prescribed by university). **Programme outcomes: At the end of the program, the students will be able to** 1. Have a deep understanding of both the theoretical and practical concepts in the respective subject. 2. Understand laboratory processes and use scientific equipments and work independently. 3. Develop research temperament as a consequence of their theory and practical learning. 4. Communicate scientific information in oral and written form. 5. Understand the issues related to nature and environmental contexts and think rationally for sustainable development. 6. The students are able to handle unexpected situations by critically analyzing the problem.
Programme Specific Outcome (PSO) - For MSc Applied **Statistics** Semester - III	1. 2. …

To Pass	At least 40% Marks in the University Examination in each paper and 40% Marks in the aggregate of University and Internal examination in each course of Theory, Practical & 40% Marks in Viva-voce.

Course Type	Course Code	Name Of Course	Theory/ Practical	Credit	Exam Duration in hrs	Component of Marks		
						Internal Total	External Total	Total
Core Course	PS03CAST51	Knowledge Discovery and Data Mining	T	4	3	30	70	100
	PS03CAST52	Operations Research - II	T	4	3	30	70	100
	PS03CAST53	Practicals (Based on PS03CAST51 and Based on PS03CAST51)	P	4	3	30	70	100
	PS03CAST54	Practicals	P	4	3	30	70	100
	PS03CAST55	Comprehensive Viva-Voce	P	4	3	30	70	100
Elective Course (Any Two)	PS03ESTA51	Planning and Analysis of Industrial Experiment	T	4	3	30	70	100
	PS03ESTA52	Generallized Linear Models	T	4	3	30	70	100
	PS03ESTA53	Survival Analysis	T	4	3	30	70	100

MASTER OF SCIENCE IN APPLIED STATISTICS
M.Sc. Applied Statistics, Semester – III

Course Code	PS03CAST51	Title of the Course	KNOWLEDGE DISCOVERY AND DATA MINING
Total Credits of the Course	4	Hours per Week	4

Course Objectives:	The objective of this course is to provide introduction to the principles and design of data mining and statistical machine learning algorithms. The course is aimed at providing foundations for conceptual aspects of data mining and machine learning algorithms along with their applications to solve real-world problems.

Course Content		
Unit	Description	Weightage* (%)
1.	Review of classification methods from multivariate analysis, classification and decision trees. Clustering methods from both statistical and data mining viewpoints.	25
2.	Unsupervised learning from univariate and multivariate data, dimension reduction and feature selection	25
3.	Supervised learning from moderate to high dimensional input spaces, artificial neural networks and extensions of regression models, regression trees	25
4.	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	25

Teaching-Learning Methodology	Interactive Class Lectures, ICT tools, Live Demonstrations & Algorithm Building.

Evaluation Pattern		
Sr. No.	Details of the Evaluation	Weightage
1.	Internal Written / Practical Examination (As per CBCS R.6.8.3)	15%

2.	Internal Continuous Assessment in the form of Practical, Viva-voce, Quizzes, Seminars, Assignments, Attendance (As per CBCS R.6.8.3)	15%
3.	University Examination	70%

Course Outcomes: Having completed this course, the learner will be able to	
1.	understand data mining principles and techniques: Introduce DM as a cutting edge business intelligence method and acquaint the students with the DM techniques for building competitive advantage through proactive analysis, predictive modelling, and identifying new trends and behaviours.
2.	have a good understanding of the fundamental issues and challenges of machine learning: data, model selection, model complexity, etc.
3.	have an understanding of the strengths and weaknesses of many popular machine learning approaches.
4.	appreciate the underlying mathematical relationships within and across Machine Learning algorithms and the paradigms of supervised and un-supervised learning.
5.	design and implement various machine learning algorithms in a range of real-world applications.
6.	construct basic feed-forward artificial neural networks with backpropagation. This develops a good foundation for studying deep learning.

Suggested References:

Sr. No.	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.

On-line resources to be used if available as reference material
On-line Resources

*****

MASTER OF SCIENCE IN APPLIED STATISTICS
M.Sc. Applied Statistics, Semester – III

Course Code	PS03CAST52	Title of the Course	OPERATIONS RESEARCH – II
Total Credits of the Course	4	Hours per Week	4

Course Objectives:	The main objectives of the course are as follows: 1. Ability to understand and analyze managerial problems in industry so that they are able to use resources (capitals, materials, staffing, and machines) more effectively. 2. Knowledge of formulating mathematical models for quantitative analysis of managerial problems in industry. 3. Skills in the use of Operations Research approaches and computer tools in solving real problems in industry. 4. Mathematical models for analysis of real problems in Operations Research.

Course Content		
Unit	Description	Weightage* (%)
1.	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)	25
2.	Assignment Problem (AP): Introduction, Area of application, mathematical model of AP – maximization and minimization problems, Hungarian Method, Multiple Optimal Solutions.	25
3.	Network Analysis: Introduction, Minimal spanning tree problem, Maximal flow problem. PERT and CPM- Terminologies used, Similarity and Differences, steps in PERT and CPM.	25
4.	Inventory Control-Deterministic and Probabilistic models, Non-linear Programming Problem- Kuhn-Tucker Conditions, Introduction to Simulation Techniques and Sequencing Problems.	25

Teaching-Learning Methodology	Interactive Class Lectures, ICT tools used

Evaluation Pattern		
Sr. No.	Details of the Evaluation	Weightage
1.	Internal Written / Practical Examination (As per CBCS R.6.8.3)	15%
2.	Internal Continuous Assessment in the form of Practical, Viva-voce, Quizzes, Seminars, Assignments, Attendance (As per CBCS R.6.8.3)	15%
3.	University Examination	70%

Course Outcomes: Having completed this course, the learner will be able to	
1.	formulate a real-world problem as a mathematical programming model
2.	solve network models like the shortest path, minimum spanning tree, and maximum flow problems
3.	understand the applications of, basic methods for, and challenges in non-linear programming
4.	identify, mathematically express and solve transportation problems.
5.	learn optimality conditions for single- and multiple-variable unconstrained and constrained non-linear optimization problems, and corresponding solution methodologies
6.	be able to design and solve simple models of CPM and queuing to improve decision making and develop critical thinking and objective analysis of decision problems.
7.	be able to solve simple problems of replacement and implement practical cases of decision making under different business environments.
8.	identify, mathematically express and solve assignment problems.

Suggested References:	
Sr. No.	References
1.	Kambo, N.S.(1991) Mathematical Programming Techniques Affiliated East-West Press Pvt. Ltd.
2.	Operations Research, Paneerselvan, PHI.

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

On-line resources to be used if available as reference material
On-line Resources

*****

MASTER OF SCIENCE IN APPLIED STATISTICS
M.Sc. Applied Statistics, Semester – III

Course Code	PS03CAST53	Title of the Course	PRACTICALS
Total Credits of the Course	04	Hours per Week	04

Course Objectives:	1. 2. …

Course Content		
Unit	Description	Weightage* (%)
1.		25
2.		25
3.		25
4.		25
...		

Teaching-Learning Methodology	

Evaluation Pattern		
Sr. No.	Details of the Evaluation	Weightage
1.	Internal Written / Practical Examination (As per CBCS R.6.8.3)	15%
2.	Internal Continuous Assessment in the form of Practical, Viva-voce, Quizzes, Seminars, Assignments, Attendance (As per CBCS R.6.8.3)	15%
3.	University Examination	70%

Course Outcomes: Having completed this course, the learner will be able to	
1.	
2.	
3.	
…	

Suggested References:	
Sr. No.	References
1.	
2.	
…	

On-line resources to be used if available as reference material
On-line Resources

*****

MASTER OF SCIENCE IN APPLIED STATISTICS
M.Sc. Applied Statistics, Semester – III

Course Code	PS03CASTA54	Title of the Course	PRACTICALS
Total Credits of the Course	04	Hours per Week	04

Course Objectives:	1. 2. …

Course Content		
Unit	Description	Weightage* (%)
1.		25
2.		25
3.		25
4.		25
...		

Teaching-Learning Methodology	

Evaluation Pattern		
Sr. No.	Details of the Evaluation	Weightage
1.	Internal Written / Practical Examination (As per CBCS R.6.8.3)	15%
2.	Internal Continuous Assessment in the form of Practical, Viva-voce, Quizzes, Seminars, Assignments, Attendance (As per CBCS R.6.8.3)	15%
3.	University Examination	70%

Course Outcomes: Having completed this course, the learner will be able to	
1.	
2.	
3.	
…	

Suggested References:	
Sr. No.	References
1.	
2.	
…	

On-line resources to be used if available as reference material
On-line Resources

*****

MASTER OF SCIENCE IN APPLIED STATISTICS
M.Sc. Applied Statistics, Semester – III

Course Code	PS03CAST55	Title of the Course	COMPREHENSIVE VIVA-VOCE
Total Credits of the Course	01	Hours per Week	04

Course Objectives:	1. 2. …

Course Content		
Unit	Description	Weightage* (%)
1.		
2.		
3.		
4.		
...		

Teaching-Learning Methodology	

Evaluation Pattern		
Sr. No.	Details of the Evaluation	Weightage
1.	Internal Written / Practical Examination (As per CBCS R.6.8.3)	15%
2.	Internal Continuous Assessment in the form of Practical, Viva-voce, Quizzes, Seminars, Assignments, Attendance (As per CBCS R.6.8.3)	15%
3.	University Examination	70%

Course Outcomes: Having completed this course, the learner will be able to	
1.	
2.	
3.	
…	

Suggested References:	
Sr. No.	References
1.	
2.	
…	

On-line resources to be used if available as reference material
On-line Resources

*****

MASTER OF SCIENCE IN APPLIED STATISTICS
M.Sc. Applied Statistics, Semester – III

Course Code	PS03EAST51	Title of the Course	Planning and Analysis of Industrial Experiments
Total Credits of the Course	04	Hours per Week	04

Course Objectives:	1. To learn how to plan, design and conduct experiments efficiently and effectively. 2. To analyze the resulting data to obtain objective conclusions. 3. To equip the student with modern statistical designs, and their applications in the industry.

Course Content		
Unit	Description	Weightage* (%)
1.	A review of basic concepts of design of experiment. Factorial Experiments: Concepts of main effects, interaction, Analysis of full 2n and 32 factorial designs, Confounding: Total and partial confounding. Analysis of 2n and 3n confounded design	25
2.	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.	25
3.	3 (n-p) Fractional Factorial Designs: Overview, Designing 3(n-p) Experiments, Box-Behnken Designs, Analyzing the 3(n-p) Design, ANOVA, Parameter Estimates.	25
4.	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	25
...		

Teaching-Learning Methodology	

Evaluation Pattern		
Sr. No.	Details of the Evaluation	Weightage
1.	Internal Written / Practical Examination (As per CBCS R.6.8.3)	15%
2.	Internal Continuous Assessment in the form of Practical, Viva-voce, Quizzes, Seminars, Assignments, Attendance (As per CBCS R.6.8.3)	15%
3.	University Examination	70%

Course Outcomes: Having completed this course, the learner will be able to	
1.	Appreciate the advantages and disadvantages of a design for a particular experiment.
2.	Understand the potential practical problems in its implementation.
3.	Construct optimal or good designs for a range of practical experiments.
4.	Describe how the analysis of the data from the experiment should be carried out

Suggested References:	
Sr. No.	References
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.

On-line resources to be used if available as reference material
On-line Resources

*****

MASTER OF SCIENCE IN APPLIED STATISTICS
M.Sc. Applied Statistics, Semester – III

Course Code	PS03EAST52	Title of the Course	GENERALIZED LINEAR MODELS
Total Credits of the Course	04	Hours per Week	04

Course Objectives:	To introduce the theory of generalized linear models (GLIM), estimation and testing procedure and explain procedure for fitting of Genralized Linear Models for real data.

Course Content		
Unit	Description	Weightage* (%)
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	25
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	25

	Models with Qualitative or/and Quantitative Covariates; Parallel Line Regression	
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	25
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	25

Teaching-Learning Methodology	Discussion and question answers based learning Black board/Multimedia projector using ICT Tools Learning through Problem solving approach Assignments and seminars are given for development of confidence among students Fitting of various models to data using software (demonstration of software R language for data handling)

Evaluation Pattern		
Sr. No.	Details of the Evaluation	Weightage

1.	Internal Written / Practical Examination (As per CBCS R.6.8.3)	15%
2.	Internal Continuous Assessment in the form of Practical, Viva-voce, Quizzes, Seminars, Assignments, Attendance (As per CBCS R.6.8.3)	15%
3.	University Examination	70%

Course Outcomes: Having completed this course, the learner will be able to	
1.	Understand the differences between Linear model and generalized model and their underlying assumptions.
2.	Various estimation procedures for estimation of parameters of different generalized models.
3.	know Logit, Probit for binary data and their applications
4.	

Suggested References:	
Sr. No.	References
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.
8.	Agresti, A. (2002). Categorical Data Analysis, ED.II, Wiley InterScience
9.	Fahrmier,L .and Tutz,G.(2001). Multivariate Statistical Modeling Based on Generalizes Linear Models, Springer
10.	Gill, J.(2001).Generalized Linear Models: A Unified Approach, Sage Publication

On-line resources to be used if available as reference material
On-line Resources

*****

MASTER OF SCIENCE IN APPLIED STATISTICS
M.Sc. Applied Statistics, Semester – III

Course Code	PS03EAST53	Title of the Course	SURVIVAL ANALYSIS
Total Credits of the Course	04	Hours per Week	04

Course Objectives:	1. Survival analysis involves the analysis of time-to-event data and is widely used in health and medicine. The objective of this course is introducing the fundamental concepts of survival analysis and basic principles such as censoring, survival and hazard functions, Cox proportional hazards regression, model diagnostics and model extensions to incorporate recurrent events and competing risks. 2. To introduce various frailty models.

Course Content		
Unit	Description	Weightage* (%)
1.	Survival data, Concepts of time, order and random and hybrid censoring, Life distributions - exponential, gamma, Lognormal, Pareto, linear failure rate, Ageing classes - IFR, IFRA, NBU, NBUE, HNBUE and their duals, Bathtub failure rate. Parametric inference, point estimation, confidence Intervals, scores, tests based on LR, MLE	25
2.	Life tables, failure rate, mean residual life and their elementary properties. Estimation of survival function - Actuarial estimator, Kaplan - Meier estimator, Estimation under the assumption of IFR/DFR	25
3.	Semi-parametric regression for failure rate - Cox's proportional hazards model, partial likelihood, estimation and inference methods for the Cox models, time-dependent covariates, residuals and model diagnosis, functional forms of the Cox models, goodness-of-fit tests for the Cox models, Competing risk models, Repair models, Probabilistic models, Joint distribution of failure times Unconditional tests for the time truncated case, Tests for exponentiality, two sample nonparametric problem.	25
4.	Nelson-Aalen estimators, counting processes and martingales, modeling counting processes, Regression models for modeling multiple events, Frailty models, Shared frailty models, Identifiability of frailty models, Frailty regression models, Bivariate and correlated frailty models, Additive frailty models.	25

...		

Teaching-Learning Methodology	

Evaluation Pattern		
Sr. No.	Details of the Evaluation	Weightage
1.	Internal Written / Practical Examination (As per CBCS R.6.8.3)	15%
2.	Internal Continuous Assessment in the form of Practical, Viva-voce, Quizzes, Seminars, Assignments, Attendance (As per CBCS R.6.8.3)	15%
3.	University Examination	70%

Course Outcomes: Having completed this course, the learner will be able to	
1.	decide the type of censoring and truncation that is the basis for given survival data
2.	estimate survival functions using parametric and non-parametric methods
3.	construct a life table using the Kaplan-Meier approach
4.	interpret coefficients in Cox proportional hazards regression analysis
5.	compare survival functions of two or more populations; use software for survival analysis
6.	explain theoretical concepts underlying parametric survival models; and
7.	analyse survival data using basic parametric models and interpret the results

Suggested References:	
Sr. No.	References
1.	Klein, J. P. and Moeschberger, M. L. (1997). Survival Analysis: Techniques for Censored and Truncated Data , Springer, New York

2.	Collett, D. (2003). Modelling Survivaldata in Medical Research, Second Edition, Chapman & Hall/CRC
3.	Therneau, T. M. and Grambsch, P. M. (2000). Modeling Survival Data, Extending the Cox Model, Springer, New York.
4.	Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data, Chapman and Hall.
5.	Deshpande, J.V. and Purohit, S.G. (2005). Life Time Data: Statistical Models and Methods, Word Scientific.
6.	Duchateau, L. and Johnson, P. (2008). The Frailty Model. Springer: New York.
7.	Hanagal, D. D. (2011). Modeling Survival Data Using Frailty Models. CRC Press.
8.	Kalbfleish, JD. and Prentice, RL. (2002). The Statistical Analysis of Failure Time Data. New York: Wiley.
9.	Hougaard, P. (2000). Analysis of Multivariate Survival Data. Springer: New York.
10.	Wienke, A. (2011). Frailty Models in Survival Analysis, CRC Press: New York.

On-line resources to be used if available as reference material
On-line Resources

*****