SHRIMATHI DEVKUNVAR NANALAL BHATT
VAISHNAV COLLEGE FOR WOMEN

DEPARTMENT OF STATISTICS

DEGREE OF MASTER OF PHILOSOPHY (M. Phil.)

(EFFECTIVE FROM THE ACADEMIC YEAR 2013-2014)

REGULATIONS

FULL-TIME/PART TIME

1. ELIGIBILITY FOR ADMISSION

Candidates who have secured a minimum of 55% in aggregate in M.Sc. statistics/Biostatistics (General or with any other specialization) degree of the University of Madras or its equivalent course recognized by the University of Madras are eligible to apply for M.Phil. in Statistics.

2. DURATION OF THE COURSE

The course shall extend over a period of ONE academic year consisting of TWO semesters for full time.

The course shall extend over a period of TWO academic year for part time.

3. COURSE OF STUDY

(i) For Full time:

The course of study for the degree shall consist of

Part I: Comprising of three written papers. The written examination will be conducted at the end of first semester.

Part II: Dissertation. To be submitted at the end of second semester.

Part I consist of:

Paper I: Research Methodology

Paper II: Advanced level paper in Statistics

Paper III: Elective (which will be a background relating to the proposed dissertation to be conducted internally by the college)

(ii) For Part time:

First year : course work only.

Part I : Theory examination (for all the three papers) will be held at the end of first semester in the **second year.**

Part II : Project / Dissertation evaluation at the end of II semester in the **second year.**

4. SCHEME OF EXAMINATION

The examination for papers I, II and III will be held in the month of April / May of the following year. The duration for each paper will be 3 hours carrying 75 marks each.

The written papers shall be evaluated by one internal and one external examiner.

Part – I credits

	Title of the Paper	Contact Hours	Library Hours	Total Hours	Credits
Core Paper I	Research Methodology	6	4	10	6
Core Paper II	Advanced Statistical Inference	6	4	10	6
Elective Paper	Applied Regression Analysis / Categorical Data Analysis / Survival Analysis/ Advanced Multivariate Analysis	6	4	10	6
Total				30	18

Part II Dissertation

The candidates should submit the dissertation to the Controller of Examinations through the Supervisor and the Head of Department at the end of one year from the commencement of the course and before August 31st. If the candidate is not able to submit for Dissertation within the period stated above she shall be given an extension of one month time with penalty. One external examiner and the Research Guide / Supervisor shall value the Dissertation. The external examiner should be from within the University

of Madras. The examiner who values the dissertation shall report on the candidate as "Highly Commended" or "Commended" or "Not Commended".

Part II Credits

	Marks	Credits
Part II Dissertation	**200**	**18**

5. QUESTION PAPER PATTERN

The pattern of question paper will be of essay type. The students should answer FIVE out of EIGHT questions and each question will carry 15 marks (5x 15 = 75 Marks).

6. PASSING MINIMUM

i) A candidate shall be declared to have passed Part I of the Examination if she secures not less than 50% marks in each paper, including paper III for which examinations are conducted internally.

ii) A candidate shall be declared to have passed Part II of the Examination if her dissertation is Commended/ Highly Commended.

iii) All other candidates shall be declared to have failed in the examinations.

Range of Marks for Dissertation

Not Commended – 49% and below; Commended – 50% to 74%;

Highly Commended – 75% and above.

7. ATTENDANCE

A candidate will be permitted to appear for the examinations only if she secures not less than 75% of attendance in each subject during the semester.

8. QUALIFICATION OF GUIDES FOR CONDUCTING THE M.Phil COURSE

A teacher is eligible to be a guide if he/she possess a Ph.D degree or two years of post graduate teaching experience after qualifying for M.Phil degree. She should have obtained recognition from the University.

COURSE STRUCTURE

SEMESTER I	TITLE OF THE PAPER	CONTACT HOURS	MARKS		CREDITS
			Int.	Ext.	
Part I Core Paper I	Research Methodology	10	25	75	6
Core Paper II	Advanced Statistical Inference	10	25	75	6
Paper III Elective	Applied Regression Analysis / Categorical Data Analysis / Survival Analysis/Advanced Multivariate Analysis	10	25	75	6
SEMESTER II Part II			Dissrtatn:150 Viva-voce:50		18
	TOTAL		500		36

RESEARCH METHODOLOGY

CORE PAPER - I

Credits: 6 **6 Hours/Week**

UNIT 1: Matrices

Introduction-some basics of Matrix - Positive definite matrices-square root matrix- Random vectors and matrices - mean vector and covariance matrices - partitioning the covariance matrix -Matrix inequalities and Maximization - Quadratic forms in random variables - canonical reduction - generalized inverse and its properties - Moore Penrose inverse - Spectral decomposition - Cholesky decomposition – SVD.

UNIT 2: Probability Theory

Convergence in sequence - Almost uniform convergence - Convergence in Probability - Convergence in measure - Convergence in mean. Central limit theorem, statement of CLT, Lindeberg, Levy and Liapounov forms with proof and Lindeberg Feller's form examples. Khintchine weak law of large numbers, Kolmogorov inequality, strong law of large numbers.

UNIT 3: Functions of Several Variables

Optimization - Techniques with and without constraints - non linear optimization, soft computing-genetic algorithm-ant colony optimization

UNIT 4: Bivariate distributions

Some special bivariate distributions - Bivariate binomial distribution -Bivariate poisson distribution- Bivariate exponential distribution -simulation from discrete, continuous, bivariate and multivariate distributions- bootstrapping technique.

UNIT 5: Research Methodology

Definition of Research- stages in Research-Types of Research-Research design and planning – Review of literature- Thesis writing- Review of journals. Writing a project proposal to a funding agency.

BOOKS FOR STUDY

1. Bhat, B.R. (1985) : Modern probability theory, 2nd ed. Wiley Eastern.
2. Graybill. F.A. (1983): Matrices with Applications in Statistics, second edition, Wadsworth.
3. Kothari. C.R. (2004): Research Methodology, New Age International (p) Ltd.
4. Taha, H (2007): Operations Research, Prentice Hall of India, 7th edition.
5. Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000): Continuous Multivariate distributions, John Wiley Publications.

BOOKS FOR REFERENCE

1. Bansal, A. K. (2007): Bayesian Parametric inference, Narosa Publishing House Pvt. Ltd.
2. Billingstey, (1984): Probability and Measure, Wiley Interscience
3. Biswas, S. (1984): Topics in Algebra of Matrices, Academic Publication

ADVANCED STATISTICAL INFERENCE

CORE PAPER II

Credits: 6 **6 Hours/Week**

UNIT 1: Estimation

Sufficient statistics, Neyman Fisher Factorisation theorem, the existence and construction of minimal sufficient statistics, Minimal sufficient statistics and exponential family, sufficiency and completeness, sufficiency and invariance.

UNIT 2:

Unbiased estimation : Minimum variance unbiased estimation, locally minimum variance unbiased estimators, Rao Blackwell – theorem. Completeness- Lehmann Scheffe theorems, Necessary and sufficient condition for unbiased estimators.

UNIT 3: Testing

Generalization of the fundamental Neyman Pearson lemma, two sided hypotheses, testing the mean and variance of a normal distribution.

UNIT 4:

Unbiased ness for hypotheses testing, similarly and completeness, UMP unbiased tests for multi parameter exponential families, comparing two Poisson or Binomial populations, testing the parameters of a normal distribution (unbiased tests), comparing the mean and variance of two normal distributions.

UNIT 5: Non-parametric tests:

Rank tests for comparing two treatments: Linear rank Statistics – Definition, Distribution Properties and usefulness; Linear rank tests for location problem- Wilcoxon test; Linear rank tests for scale problem – Siegel-Tukey test.

Comparison of more than two treatments - Kruskal, Wallis test - Testing against trend using Kendall's statistic; randomised complete blocks- Friedman's two way analysis of variance by ranks, test for independence, z x t contingency tables.

BOOKS FOR STUDY & REFERENCE:

1. Lehmann, E.L. (1986) : Testing of statistical hypothesis, John Wiley.
2. Lehmann, E.L. (1983): Theory of point estimation, John Wiley.
3. Lehmann, E.L.(1975) : Non parameteric: Statistical methods based on Ranks, McGraw Hill.
4. Gibbons, J.D. (1971): Nonparametric Statistical Inference, McGraw Hill.
5. Rohatgi, V.K. and Saleh, A.K. (2002): An Introduction to Probability and Statistics, John Wiley & Sons.

ADVANCED MULTIVARIATE ANALYSIS

ELECTIVE PAPER

Credits: 6 **6 Hours/Week**

UNIT 1:

Cluster Analysis

Distance and similarity measures - hierarchical clustering techniques - agglomerative techniques - single linkage, complete linkage, average linkage methods

UNIT 2:

Nonhierarchical clustering methods-theory of k means and k medoids, fuzzy c means-derivations, Introduction to theory of rough k means; validity measures based on clustering methods

UNIT 3:

Support Vector Machine

Linear Support Vector Machines – The Linearly Separable Case – The Linearly Nonseparable Case – Nonlinear Transformations – Kernels and Their Properties – Examples of Kernels – Binary Classification – Multiclass Support Vector Machines – Multiclass SVM as a Series of Binary Problems – Multiclass SVM – Optimization Algorthms for ANM's – Software Packages.

UNIT 4:

Correspondence Analysis

Introduction, Attribute independency and contingency table, Graphic representation: file profiles or column profiles, Relationship with main components: double weighting, Weight of the profiles, Profiles standardization, Distributional distance between profiles, Concept and interpretation.

UNIT 5:

Principal component analysis

Extraction of components - properties and characteristics of components - total variation, relative importance, standardization of variables and components, special covariance structures - interpretation of principal components

Canonical correlation analysis

Extraction of canonical correlations and their variable - testing the significance of canonical correlation - interpretation of canonical variables

Note: Implementation of the above techniques using R language

BOOKS FOR STUDY:

1. Johnson, R.A. and Wichern, D.W. (2002): Applied Multivariate Statistical Analysis, 5th edition, Pearson Education, Asia.
2. Tan, P.N., Steinbach, M. and Kumar,V.(2008): Introduction to Data Mining, Pearson Education.

BOOKS FOR REFERENCE:

1. Anderson, T. W. (2003): An Introduction to Multivariate Statistical Analysis, 3rd edition, John Wiley & Sons.
2. Hair, J.F., Anderson, R.E., Tatham. R.L. and Black, W.C. (2006): Multivariate Data Analysis, 5th edition, Pearson Education, Asia.
3. Han, J. and Kamber, M. (2002): Data Mining: Concepts and Techniques, Morgan Kaufman Publishers.
4. Soman, K. P., Diwakar, S. and Ajay, V. (2008): Insight into Data Mining: Theory and Practice, Prentice Hall of India Pvt, Ltd.
5. Morrison, D.F. (1990): Multivariate Statistical Methods, 3rd edition, McGraw Hill.

APPLIED REGRESSION ANALYSIS

ELECTIVE PAPER

Credits: 6 **6 Hours/Week**

UNIT 1:

Simple linear regression-Assumptions, estimation of model parameters, standard error of estimators, testing of hypotheses on slope and intercept (β's), interval estimation of model parameters, prediction interval of a new observation, coefficient of determination, regression through origin, Tests of hypotheses about parallelism and equality of intercepts.

UNIT 2:

Multiple linear regressions : Standard Gauss Markov setup, least square estimation of model parameters, variance covariance of least squares estimators, estimation of error variance

Tests of hypotheses – significance of regression (ANOVA, $R^2$and adjusted R^2), individual regression coefficients, subset of regressor variables, general linear hypotheses- Confidence intervals and regions, prediction intervals, detecting hidden interpolation.

UNIT 3:

Model adequacy checking - residual plots for checking normality homoscedasticity and autocorrelation, detection of outliers, Test for Lack of fit of the model. Durbin – Watson test for autocorrelation. Analytical methods for selecting a transformation generalized and weighted least squares- Detection of influential observations – Cooks statistic, DFFITS, DFBETAS.

Variance stabilizing transforms and transforms to linearize the model, analytical methods for selecting a transform, generalized and weighted least squares. Dummy (or indicator variables) – general concepts and their use.

UNIT 4:

Multicollinearity – sources, effects, diagnostics, Methods of dealing with multicollinearity (collection of additional data, model respecification, Ridge regression).

Selection of Variables – forward selection, backward elimination and stepwise regression (algorithms only)

UNIT 5:

Nonlinear regression – transformation to a linear model, their use and limitations, initial estimates (starting values), parameter estimation using iterative procedures – Gauss-Newton, steepest Descent, Marquardt's compromise.

BOOKS FOR STUDY:

1. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003): Introduction to Linear regression analysis, third edition, John Wiley and Sons, Inc.
2. Zar, J.H. (2006): Biostatistical Analysis, fourth edition, Pearson education.

BOOKS FOR REFERENCE:

1. Draper, N.R. and Smith, H. (2003): Applied Regression Analysis, third edition, John Wiley and Sons, Inc.
2. Johnston, J. (1984): Econometric methods, third edition, McGraw Hill International.

CATEGORICAL DATA ANALYSIS

ELECTIVE PAPER

Credits: 6 **6 Hours/Week**

UNIT 1:

Nature of Categorical data - Contingency tables and their distribution: Binomial and Multinomial sampling. Comparing proportions in two-by-two tables: Difference of proportions - Relative risk - Odds Ratio - Properties of Odds Ratio - relationship between Odds Ratio and Relative Risk - Odds ratio in I x J tables -Nominal and Ordinal Measures of Association.

Inference for Contingency tables: Interval estimation for difference of proportions, odds ratio, log odds ratio and relative risk.

Testing Independence in Two-Way tables: Pearson and Likelihood-ratio chi-square tests - Yate's correction for continuity-Residuals for cells in a contingency table-Partitioning chi-square

UNIT 2:

Testing Independence for Ordinal Data-Trend tests for 2 x J tables-Exact Inference for small samples - Fisher Exact Test for 2 x 2 tables

Association in Three-Way Tables: Partial Tables - Conditional and Marginal Odds Ratios - Homogeneous Association - Cochran-Mantel-Haenszel methods.

UNIT 3:

Components of Generalized Linear Models-Logit models for Binary data-Binomial GLM for 2 x 2 contingency tables

Logistic regression: parameter interpretation - maximum likelihood estimate - test of overall regression and goodness of fit - Wald test, deviance statistic, LR test, score test-Logistic regression diagnostics.

UNIT 4:

Logit model with Categorical predictors: Linear logit model for I x 2 tables - Cochran-Armitage Trend test. Logit models for multiway contingency tables.

Multinomial Logit models: Logit models for Nominal responses- Base-line category logits-estimating response probability

Logit models for ordinal responses: Cumulative logits - proportional odds model.

Probit models – complementary log-log link model (description only).

UNIT 5:

Poisson log-linear model for count data- ML estimation, goodness of fit test, Loglinear Models for Two-way Tables: Independence Model-Interpretation of parameters- Poisson regression for rates.

Models for matched pairs: Comparing dependent proportions-McNemar's test, small sample test for comparing matched proportions.

Measures of agreement-Kappa measure - Weighted Kappa - Bradley-Terry model for paired preferences, generalized estimating equation methodology (basic idea).

BOOKS FOR STUDY:

1. Agresti, A. (2002): Categorical data analysis, John Wiley & Sons.
2. McCullagh, P. and Nelder, J.A. (1991): Generalized Linear Models, Second Edition, Chapman and hall, London.

BOOKS FOR REFERENCE:

1. Agresti, A. (1991): An Introduction to Categorical data analysis, John Wiley & Sons.
2. Armitage, P. and Berry, G. (1987): Statistical methods in Medical Research, Blackwell Scientific Publications, USA.
3. Deshpande, J.V., Gore, A.P. and Shanubhogue, A. (1995): Statistical Analysis of Non Normal Data, New Age International Publishers Ltd., New Delhi.
4. Hardin, J.W., and Hilbe, J.M. (1994): Generalized Estimating Equation, Chapman and Hall, London.

5. Hosmer, D.W. and Lemeshow, S. (1989): Applied Logistic Regression,
John Wiley & Sons Inc.

SURVIVAL ANALYSIS

ELECTIVE PAPER

Credits: 6 **6 Hours/Week**

UNIT 1:

Concept of Time and event, Order and Random Censoring, Censoring mechanism and truncations, Survival, hazard and density functions. Mean and median residual life and their elementary properties, Ageing classes - IFR, IFRA, NBU, NBUE, HNBUE, DMRL and their duals, Bathtub Failure rate

UNIT 2:

Life distributions - Exponential Gamma, Weibull, Lognormal, Pareto, Linear Failure rate-Parametric inference: Point estimation, Confidence Intervals, Scores, tests based on LR and MLE, Partial likelihood estimation-log logistic distribution.

UNIT 3:

Estimation of survival function: reduced sample method, actuarial estimator, Kaplan-Meier Estimator-life table estimation.

UNIT 4:

Nonparametric methods: Gehan Test, Log rank test. Mantel - Haentzel Test, Tarone - Ware tests, Efron Tests.

UNIT 5:

Semi-parametric regression for failure rate - PH assumptions - Cox's proportional hazards model with one and several covariates-goodness of fit. Extended Cox's model.

BOOKS FOR STUDY:

1. Deshpande, J.V., Gore, A.P. and Shanubhogue, A. (1995): Statistical Analysis of Non Normal Data, New Age International Publishers Ltd., New Delhi
2. Klein, J.P. and Moeschberger, M.L. (2003): Survival Analysis-Techniques for Censored and Truncated data
3. Miller, R.G. (1981): Survival Analysis, John Wiley and Sons.

BOOKS FOR REFERENCE:

1. Barlow, R. E. and Proschan, F. (1975): Statistical Theory of Reliability and Life testing, Holt, Rinehart and Winston, New York.
2. Johnson, E.R.E. and Johnson, N.L. (1980): Survival models and Data Analysis, John Wiley and Sons.
3. Lee, C.T. (1997): Applied survival analysis, John Wiley.