SHIVAJI UNIVERSITY, KOLHAPUR.



Accredited By NAAC with 'A' Grade

Revised Syllabus For

M. Phil./ Ph. D. Course Work

Statistics

Syllabus to be implemented from

June, 2020 onwards.

Department of Statistics Shivaji University, Kolhapur The syllabi for the Pre-Ph.D./M.Phil. Programming Statistics (To be effective from Academic Year2020-2021)

• There shall be three courses for Pre-Ph.D./M.Phil. Programme:

Paper – I: Research Methodology. Marks:100

Paper – II: Recent Trends in Statistics. Marks:100

Paper - III: Related to the area of research topic from the following list of courses.

Marks:100 (80 marks for Theory Paper examination + 20 marks for seminar*)

- i. Advanced Distribution Theory
- ii. Applied Regression Analysis
- iii. Directional Data Analysis
- iv. Topics in Statistical Quality Control.
- v. Advanced Multivariate Analysis.
- vi. Optimization Techniques
- vii. Medical Statistics
- viii. Advanced Survival Analysis (Syllabus not available)
- ix. Advanced Inference (Syllabus not available)

* 20 Marks are reserved for a seminar. A student is expected to review a research paper in Statistics, published during last five years in national or International Journal of reputelisted in the UGC CARE list. The research paper preferably should be related to the topic of research and to be chosen in consultation with the research guide. The candidate should give a seminar on the review of the selected paper.

Paper – IResearch Methodology

Unit1:Introduction to Research Methodology:

Definition, Characteristics, Objectives, Research and Scientific method, Literature review, types of research, process of scientific research, Research design, Tools for data collection and statistical analysis of data, Use of open source software (Excel, R and Python etc.), Intellectual property rights (IPR) and Plagiarism; patents/startups and its e-filing procedures, norms and policies. (15L)

Unit2: R-software:

Introduction to R, data types and structures, data input, operators, R-packages, data import and export, built in functions for basic statistical computations and analysis, saving work in R, matrix algebra, apply family of functions, graphical procedures, graphs using ggplot2 package, control statements, programming, user defined functions, numerical methods for root finding; solving system of linear equations and numerical integration using R, parallel computing, performance evaluation of statistical procedures. (15L)

Unit3: Simulation Techniques, Monte Carlo Methods and permutation tests:

Simulation Techniques:Quantile Transformation and Accept–Reject, Importance Sampling and Its Asymptotic Properties, Optimal Importance Sampling Distribution, Algorithms for Simulating from Common Distributions

Monte Carlo Methods: Markov Chain Monte Carlo Methods and applications, EM algorithm, Metropolis Hasting Algorithm, Gibbs Sampling.

Permutation tests: General Permutation Tests, Exact Similarity of Permutation Tests, Power of Permutation Tests.

Unit-4:Resampling:

The Bootstrap: Bootstrap Distribution and the Meaning of Consistency, Consistency in the Kolmogorov and WassersteinMetrics. Delta Theorem for the Bootstrap, second—Order accuracy of the Bootstrap, m out of n Bootstrap, Bootstrap Confidence Intervals, Bootstrap Confidence intervals for Quantiles, Bootstrap in Regression, Residual Bootstrap, Block Bootstrap Methods: Non-overlapping block bootstrap, Moving block bootstrap (MBB), Optimal Block Length.

The Jackknife: Motivating Examples, Bias Correction by the Jackknife, Variance Estimation, Delete-d Jackknife and Von Mises Functionals.

(15L)

(15L)

References:

1. Wickham, H. (2019) Advanced R. CRC press.

- 2. Peng, R. D. (2016) R programming for data science. Leanpublishing.
- 3. Gardener, M. (2012) Beginning R: The statistical programming language. John Wiley & Sons.
- 4. Bloomfield, V. A. (2018) Using R for numerical analysis in science and engineering. CRC Press.
- 5. Anirban DasGupta (2008) Asymptotic Theory of Statistics and Probability. Springer.

- 7. Robert C P, Casella G (1999) Monte Carlo Statistical Methods (Berlin: Springer-Verlag)
- 8. Christian Robert and George Cassella (2010) Introducing Monte Carlo Methods with R-Springer.

Paper – II: Recent Trends in Statistics

Unit1: Robust Estimation: Types of Robustness, Basic Tools for Judging Robustness, Robust estimation of Location and Scale, Robust estimation in one and two sample problems.

Unit2: NonparametricRegression:

Basic idea of smoothing, spline smoothing, kernel regression. Nonparametric designing estimation. The naïve estimator, The Kernel estimator, The nearest neighbor method, The variable Kernel Method, Orthogonal series estimators, Maximum penalized likelihood estimators, General weight function estimators. (15L)

Unit3:Artificial Neural network and Support vector Machines:

Fundamental concept of Artificial Neural network, feed forward and feedback networks, perception learning rule. Single layer feed-forward neural network. Multilayer feed- forward neural network and Backpropagation training method.

Support vector Machines (SVM): concept of SVM, Problem formulation, Lagrangian theory, Duality, KKT conditions, support vector classification, support vector regression and their implementations. (15L)

Unit4:Big Data Analysis:

Defining big data: volume, variety, velocity, veracity, variability and value; big data analytics: text analytics, audio analytics video analytics, social media analytics and predictive analytics; large scale hypotheses testing: large scale testing, false discovery rates, empirical Bayes large scale testing, local false discovery rates and choice of null distribution; Sparse modeling: LASSO, least angle regression, post selection inference for LASSO.

(15L)

(15 L)

References

- 1. Critianini N and Shawe-Taylar John (2000) An Introduction to Support vector machines Cambridge University Press.
- 2. Zurada J.M. Introduction to Artificial Neural systems
- 3. Efron, B., & Hastie, T. (2016). *Computer age statistical inference* (Vol. 5). Cambridge University Press.
- 4. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144.
- 5. Härdle, W. (1990). Applied nonparametric regression. Cambridge university press.
- 6. Wilcox, R. R. (2017) Introduction to Robust Estimation and Hypothesis Testing, 4th Ed Academic Press.

7. Huber, P. J., & Ronchetti, E. M. (2009). Robust Statistics. 2nd John Wiley & Sons. Hoboken, NJ.

Paper -III i) Advanced Distribution Theory

Unit 1: Logarithmic Distributions: The Lognormal Distribution , Log-Logistic Distributions, Log-Extreme Value Distributions, The Log Cauchy Distribution, The Log Student's t Distribution, Alternatives for the Logarithm transformation with examples.

(15L)

Unit 2: Distributions with specific types of parameters (Semiparametric Families): Location, Scale Parameters, Power Parameters, Frailty and Resilience Parameters: Proportional Hazards and Reverse Hazards, Tilt Parameters: Proportional Odds Ratios, Extreme Stable Families, Hazard Power Parameters, Laplace Transform Parameters, Convolution Parameters, Age Parameters: Residual Life Families, Successive Additions of Parameters, Mixing Semiparametric Families. (15L)

Unit 3: Bivariate models: Bivariate continuous models, bivariate Pearson system, Farlie Morgenstern distribution; distributions with specified conditionals, bivariate Pareto of I, II, III and IV kind distributions. (15L)

Unit 4: Record values - definition, properties, distribution of nth record, record values from exponential, Weibull and logistic; Moments relationships, characterizations. (15L)

References:

1. Johnson, N.L., Kotz, S. and Kemp, A.W. (1992) Univariate discrete distributions, second edition, Wiley.

2. Kocherlakota, S. and Kocharlakota, K. (1992) Bivariate Discrete Distributions, Marcel-Dekker

3. Marshall, Albert W., and Ingram Olkin. (2007) *Life distributions: Structure of Nonparametric, Semiparametric, and Parametric Families*. Vol. 13. Springer, New York,

4. Kotz, S. , Balakrishnan, N. and Johnson, N.L. (2000) Continuous multivariate distributions, Volume I, John Wiley and Sons.

5. Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1998) Records, John Wiley and Sons.

Paper- III

(ii) Applied Regression Analysis

Unit-1: Introduction and review of basic results on multiple linear regression. Hypothesis testing, confidence and prediction intervals, General linear hypothesis testing. Residuals residual analysis including various residual plots,VST and Box-Cox transformation.(15L)

Unit-2: Concept of multicollinearity, sources and measures of multi-collinarity, Tools for handing Multicollinearity: methods based on singular value decomposition and ridge regression. Properties of ridge estimator. Concept of LASSO regression and Elastic net regression. (15L)

Unit-3:Robust Regression: Influential observations, leverage, outliers, outliers and influential observations detection methods. Estimation in the presence ofvertical outliers: M-estimator, robust criterion functions (Huber, Hampel, Ramsey etc.) breakdown point, influence function, efficiency, Asymptotic distribution of M-estimator, rank estimation procedures when error follows non normal distribution. (15L)

Unit-4: Introduction to general non-linear regression-least squares in non-linear case, estimating the parameters of a non-linear system. Parameterization of the model.Generalized Linear Models: The exponential family, Likelihood Theory and moments, Linear structure and the link functions, estimation procedures Newton Raphson, WLS, IWLS, Residuals and Model fit. (15L)

References:

1. Draper N and Smith H (1981): Applied Regression Analysis.

2. Gunst R. F. and Mason R. L. (1980): Regression Analysis and its applications a data oriented approach.

- 3. Mongomery, Peack and Vining (2006): Introduction to linear regression.
- 4. Rawlings, Pantula and Dickey (2001): Applied regression analysis, A research tools
- 5. Chatterjee and Hadi (2006): Regression Analysis by example.
- 6. D. Birks and Y. Dodge (1993) AlternativeMethod in Regression.
- 7. Gill, J.: Generalized linear models: A unified a approach.

PaperIII (iii) Directional Data Analysis

Unit 1: Graphical representation of data, Trigonometric moments, Measures of location, circular variance, Circular skewness and Circular kurtosis. Circular models: wrapped Cauchy, Von Mises distribution, mixtures of von Mises distributions; Point estimation for these distributions. (15L)

Unit 2: One, two and multi-sample tests for the von Mises distribution. Non-parametric methods: Tests for randomness, goodness of fit, Rayleigh's test. Range test, Kuiper's test, Watson's test. (15L)

Unit 3: Circular Correlation and Regression: Circular Measure for Correlation ρ_c , its sample estimate, Rank Correlation, Circular-Linear Correlation, Circular-Circular Regression, Estimation of Regression Coefficients, Circular-Linear Regression, Linear-Circular Regression.(15L)

Unit 4: Spherical Data, spherical polar coordinates, descriptive statistics, Spherical distributions: Uniform, von Mises Fisher, Fisher-Bingham Distributions, Kent Distributions. One, two and multi-sample tests for the von Mises Fisher distribution. (15L)

References:

1. Mardia, K. V. (2014). Statistics of directional data. Academic press.

- 2. Jammalamadaka, S. R., & Sengupta, A. (2001). Topics in circular statistics (Vol. 5). world scientific.
- 3. Fisher, N. I., Lewis, T., & Embleton, B. J. (1993). *Statistical analysis of spherical data*. Cambridge university press.

Paper III (iv)Topics in Statistical Quality Control

- Unit 1:Shewhart Control charts: basic principles and assumptions, phase I and phase II applications of control charts, performance measures of a control chart; \overline{X} , R, S, S^2 , p, c and D charts; warning limits, sensitizing rules; adaptive control charts; economic design of a control chart. (15L)
- Unit 2:CUSUM charts, EWMA charts, combined Shewhart-CUSUM charts, combined Shewhart-EWMA Charts; Hotelling's T^2 Chart, Multivariate CUSUM Charts, Multivariate EWMA Charts; GLR Charts, SPRT charts, nonparametric control charts, Bayesian control charts. (15L)
- **Unit 3:**Process capability, process capability indices (C_p, C_{pk}, C_{pm}, C_{pmk}), point and interval estimation of C_p and C_{pk}; Nonparametric capability indices: robust capability indices, capability indices based on fitted distributions, data transformation, capability indices computed using resampling methods; multivariate process capability indices. (15L)
- **Unit 4:**Acceptance sampling plans for attributes: single sampling plan, double and multiple sampling plans, sequential sampling, performance measures of sampling plans; Acceptance sampling plans for variables: sampling inspection plans by variables for one or two sided specifications, sequential sampling by variables. (15L)

References:

- 1. Guenther, W. C. (1977). Sampling Inspection in statistical quality control. Macmillan.
- 2. Kenett, R. S. and Zacks, S. (2014). *Modern Industrial Statistics with applications in R, MINITAB and JMP*. John Wiley & Sons.
- Montgomery, D. C. (2010). Statistical Quality Control: A Modern Introduction, 6th Edition. Wiley India Pvt Ltd.
- 4. Ryan, T. P. (2011). Statistical Methods for Quality Improvement. John Wiley & Sons.

Paper: III (v) Advanced Multivariate Analysis

Unit -1: Exploring Multivariate Data: Types of data, scatterplot marix, Coplots and trellis graphics. Multivariate Normal distribution. Checking distributional assumptions using probability plots. (15L)

Unit-2: Correspondence Analysis: Concept, two-dimensional contingency tables, applications, multiple correspondence analysis. Multidimensional Scaling: Proximity matrices, metric least squares multidimensional scaling, Non-metric multidimensional

scaling, Non-Euclidean metrics, Three-way multidimensional scaling, Inference. (15L)

- **Unit-3:** Models for Multivariate response variables: Multivariate linear regression model, repeated quantitative measure, logistic model for multivariate binary responses, marginal models for repeated binary responses. Conjoint Measurement Analysis: Design of data generation, estimation of preference ordering. Applications.(15L)
- Unit 4: Non-parametric tests: Tests for multivariate location, scale and location-scale set up. Mardia's test, bivariate Wilcoxon sign-rank test, Peters and Randles test. Depth based ordering of vectors. Depth function, tests based on data depth, data depth plot. Applications.(15L)

References:

1.Dodge, Y. (Ed.). (2012). Statistical data analysis based on the L1-norm and related methods. Birkhäuser.

2. Everitt, B. S., & Dunn, G. (1991). Applied multivariate data analysis (No. 519.5076 E9).

3. Giri, N. C. (2014). Multivariate statistical inference. Academic Press.

4.Härdle, W., & Simar, L. (2007). *Applied multivariate statistical analysis* (Vol. 22007, pp. 1051-8215). Berlin: Springer.

5. Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (Vol. 5, No. 8). Upper Saddle River, NJ: Prentice hall.

6. Liu, R., & McKean, J. W. (2015). Robust rank-based and nonparametric methods. Springer.

7. Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate Analysis*, Academic Press Inc. *London*) *LTD*, 15, 518.

8. Parelius, J. (1997). *Multivariate analysis based on data depth*. Ph.D. dissertation. Dept. Statistics, Rutgers Univ., New Jersey. Z. Z.

PaperIII (vi) Optimization Techniques

Unit 1: Deterministic inventory models: The meaning of inventory control, factors involved in inventory problem, inventory model building, Concept of EOQ.Single item inventory control models without shortages: EOQ model with constant rate of demand, EOQ model with different rates of demand in different cycles and Economic lot size with finite Rate of replenishment. (EOQ production model). Single item inventory control models with shortages: EOQ model with constant demand and variable order cycle time, EOQ model with constant demand and fixed reorder cycle time and The production lot size model with shortages (15L)

Unit 2: Replacement Problems: Types of failure: Gradual failure, Sudden failure, Replacement policy for items whose efficiency deteriorates with time: when value of money remains constant and when value of money changes.Replacement of items that completely fail: Mortality tables, mortality theorem, individual and group replacement policy.(15L)

Unit 3:Goal Programming: Meaning & Concept of GP, Difference between LP and GP Approach, Model Formulation, Graphical Method to Solve GP, Modified Simplex Method.(15L)

Unit 4: Integer Linear Programming Problem (ILPP): The concept of cutting plane, mixed ILLP, Branch and Bound method.Non- Linear programming: Kuhn-Tucker conditions, methods due to Beale, Wolfe. (15L)

References:

- 1. Hadley G. (1969): Linear Programming, Addison Wesley.
- 2. Taha H. A. (1971): Operation Research An Introduction- Macmillan
- 3. Kanti Swaroop& Gupta M. M. (1985): Operations Research, Sultan Chand & P. Gupta
- 4. D. S. Hira (2010): Operation Research, Sultan Chand & Co.ltd.
- 5. J. K. Sharma. (2003): Operation Research Theory and Applications. Macmillan.

Paper-III (vii) Medical Statistics

Unit-1. Basic Measures of Fertility and Reproduction:

Definition of natural fertility, fertility, fecundity, fecundability; Basic measures of fertility and reproduction; Cohort and period measures of fertility; Sources of fertility data; Nuptiality, Nuptiality Table and Measurements; Timings of Events-Age at first marriage, first birth, last birth, birth intervals, menarche, menopause, sterilization, Mean age childbearing; Tempo and quantum effects in fertility; Parity and birth order-distributions with Indian examples, Parity progression ratios (PPRs). (15L)

Unit-2. Estimation of Fertility and Analysis:

Coale's fertility indices; Coale's-Trussel model of natural fertility; SMAM, P/F Method to estimate fertility; Decomposition of fertility; Age-pattern of Fertility, Estimating fertility through PPRs, Calculation of Bongaarts' Indices, Rele's method of estimating fertility, Reverse survival method of estimating fertility; Estimating fertility from Own-children data; Coale's (1981) Robust Procedure to Estimate fertility from single census; Estimating of fertility from CEB data using Gompertz relational model; Estimating fertility from historical data; Estimating sex ratio at birth, birth intervals, Probit analysis to estimate age of menopause; Long term fertility projections: Intergenerational Rationale andtimeseriesmodels.(15L)

Unit-3. Public health and Mortality(i)

Concepts and definitions: Health, morbidity, disease burden, disability, prevalence and incidence, etc. Sources of health data/information: Civil Registration, Sample Registration System (SRS), Census and other large scale surveys, completeness and quality of data. Applications of health measures in planning, monitoring and evaluation; CDR, IMR or ASDR for estimating immunization needs, clustering, patterning of death, etc. Advanced methods of estimating/assessing mortality, and Construction and applications of life tables (multiple decrement). Age pattern of mortality: focus on adult mortality and morbidity/disease pattern. Avoidable mortality. (15L)

Unit 4: Public health and Mortality(ii)

Measures of health and burden of disease: Concepts of health expectancy, DALY, survivorship curve; epidemiological estimates for diseases (Years of Lost due to Disability-

YLD). Culture, community and disease (anthropological epidemiology): Traditional health providers (primitive/tribal/ancient) and practices, and Cultural and socio-religious interpretation/meaning associated with diseases, and health-seeking behaviors.(15L)

List of Reference

1. Preston, Samuel H., Heuveline, Patrick, and Guillot, Michel (2001). *Demography: Measuring and Modeling Population Processes*. Oxford: Blackwell Publishers.

2. Siegel, Jacob S., and David A. Swanson (eds.), (2004). *The Methods and Materials of Demography* (Second edition). San Diego: Elsevier Academic Press.

3. United Nations(1999). *Below Replacement Fertility*, Population Bulletin of the UN, Special Issue Nos. 40/41, Department of Economic and Social Affairs, UN, New York.

4. Rowland, Donald T.(2006), *Demographic Methods and* Concepts. New York: Oxford University Press.

5. Das Gupta, M and M. Rani(2004), *India's Public Health System: How well does it function at the National level*, Policy Research Working Paper No. 3447, World Bank, Washington, D.C.

6. Government of India (2002), *National Health Policy*, Ministry of Health and Family Welfare, New Delhi.

7. Government of India(2005), National Rural Health Mission – Framework for Implementation 2005-2012, Ministry of Health and Family Welfare, New Delhi.

8. Mathers, CD, T. Vos, AD Lopez, J. Salomon, and M. Ezzati(eds) (2001). *National Burden of Disease Studies: A Practical Guide*, Global Program on Evidence for Health Policy. Geneva: World Health Organization.

9. Mills, A, JP Vaughan, DL Smith and I Tabibzadeh(eds.) (1993). *Health System Decentralization: Concepts, issues and country experience,* WHO, Geneva.

10. Murray, CJL, and AD Lopez(Eds: 1996). The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries and Risk Factors in 1990 and Projected to 2020. Global Burden of Disease and Injury Series, Vol. 1. Cambridge: Harvard University Press.

11. Peter, D.H., AS Yazbeck, RR Sharma, GNV Ramana, LH Pritchett and A Wagstaff (2000). Better Health System for India's Poor: Findings, analysis and options, World Bank, Washington, D.C.

VHAI (1997). Report of the Independent Commission on Health in India, VHAI, New Delhi.
World Bank(2003). World Development Report: Making services work for poor people, Oxford University Press, Washington, D.C.

14. World Health Organization(2000). World Health Report 2000. Health systems: Improving Performance. Geneva, World Health Organization.

15. John P. Klein, Melvin L, Moeschberger, J.P. Klein, Alwyn B. Scott: Survival analysis: Techniques for Censored and Truncated data/Edition2, Springer, New York.

16. Devid Collett, Chapman and Hall/CRC, Boca Raton: Modeling Survival Data in Medical Research, Second Edition

17. Terry M. Therneau and Patricia M. Grambsch: Modeling Survival Data, Extending the Cox model Springer, New York.