Curriculum for Doctor of Philosophy Programme in Statistics Program Code: 177

Department of Statistics
Indira Gandhi National Tribal University
Amarkantak, India

Contents

1 PhD(Statistics) Program

1.1 Qualifying Examination Requirements

Subject to the University regulations, the following shall be applicable to all students joining the PhD program in the Department of Statistics:

1. Each student in the Ph.D. programme in the Department of Statistics is required to apply for PhD Qualifying Examination.

2. The qualifying examination, if required (in case of non - NET candidates) shall consist of two sections namely **Part - A (Research Aptitude Test)** and **Part - B (Statistics)**. Part- A and Part - B shall consist of 30 and 70 objective type questions respectively.

3. The qualifying examination shall be conducted as per the notification issued by the office of the Controller of Examinations, IGNTU every year.

4. Notwithstanding the qualifying examination requirement, the course requirement stipulated by the University regulations for the Ph.D. programmes will have to be satisfied. It may be noted that currently the course requirement for Ph.D. students consist of completing at least 12 course credits.

5. **PhD Regulations - 2016** of the IGNTU may be referred to for all details.

1.2 Syllabus for Qualifying Examination

The entrance examination shall be designed to test the understanding of the fields mentioned below. The questions shall be designed to test how theory can be applied to solve problems / frame solutions to practical problems. The level of questions shall be as expected of a post-graduate student.

1. Probability Theory
2. Distribution Theory
3. Statistical Inference
4. Design of Sample Surveys
5. Design and Analysis of Experiments
6. Multivariate Analysis
7. Exploratory Data Analysis
8. Programming in **R** / **C**
9. Real Analysis
10. Linear Algebra

2 Courses Offered in PhD(Statistics) Program

2.1 Mandatory Core Courses

S. No.	Title of the Course	Code	Credits
1	Research Methodologies	SCC-01	4
2	Lab	SCC-02	1

2.2 Discipline Specific Core Courses

The student may select at least two from among the following set of courses.

S. No.	Title of the Course	Code	Credits
1	Biostatistics	RSTAT-01	5
2	Analysis of Biostatistical Data	RSTAT-02	5
3	Advanced Design of Sample Surveys	RSTAT-03	5
4	Advanced Design and Analysis of Experiments	RSTAT-04	5
5	Testing of Hypothesis	RSTAT-05	5
6	Generalised Linear Models and Inference	RSTAT-06	5
7	Bayesian Inference	RSTAT-07	5
8	Theory of Estimation	RSTAT-08	5
9	OR: Stochastic and Deterministic Models and Optimization	RSTAT-09	5

2.3 Semester-wise Structure of the Programme

Semester	Code	Title of the Course	L-P-T-C	Marks
I	SCC-01	Research Methodologies	3-0-1-4	80
	SCC-02	Lab	0-1-0-1	20
	RSTAT-***	Discipline Specific Core Course - I	3-1-1-5	100
	RSTAT-***	Discipline Specific Core Course - II	3-1-1-5	100
II - IX		Research Work and Dissertation		

3 Detailed Syllabus

3.1 SCC-01 Research Methodologies

An overview of research methodology Research concept, steps involved, identification, selection and formulation of research problem, justification, hypothesis; literature collection- textual and digital resources (internet).

Research Design, Data Collection and Interpretation Research design; sampling techniques, collection and documentation, presentation, analysis and interpretation of data

Scientific Writing Forms of scientific writing- Article, notes, reports, review article, monographs, dissertations, popular science articles, bibliographies.

Formulation of Scientific Communication Outline preparation, drafting title, sub titles, tables, illustrations; Formatting tables- title, body footnotes; figures & graphs- structure, title and legends, Impact factor, citation indices, plagiarism.

Computer application MS office, excel, power point, graphics (Sigma plot), statistical software (SPSS).

Elementary Biostatistics Standard deviation/error; Correlation coefficient, types of correlation, regression equation, biological significance of correlation and regression; Test of significance, chi-square test, analysis of variance.

References

1. Research Methodology - Methods & Techniques, CR Kothri CR (1990), Vishva Prakashan, New Delhi.

2. Research Methodology & Statistical Techniques, S Gupta (1999) Deep & Deep Publications, New Delhi.

3. Research Methodology for Biological Sciences, N Gurumani (2007), MJP Publishers, Chennai.

4. Introduction to Biostatistics, L Forthofer (1995), Academic Press, New York.

5. Biostatistical Analysis, JH Zar (2006), Prentice-Hall.

6. Research Design: Qualitative, Quantitative & Mixed Method Approaches, John W. Creswell (2009), Sage Publication, USA.

7. Experimental Design & Data Analysis for Biologists. PQ Gerry & JK Michael (2002), Cambridge University Press.

8. Choosing & Using Statistics: A Biologists Guide, D Calvin (2003), Blackwell Publisher.

3.2 SCC-02 Lab

Lab work based on **SCC-01**.

3.3 RSTAT-01 Biostatistics

Rates for Measuring Health and Disease Incidence and prevalence rates, disease attack rates, odds ratios, absolute and relative risk;

Standardization of Rates Direct and indirect standardization.

Life Tables Life tables and their applications.

Diagnostic Tests Bayes theorem, diagnostic tests sensitivity and specificity, applications of Bayes theorem, Receiver Operator Characteristics (ROC) curves, calculation of prevalence. The relative risk and the odds ratio.

Probabilistic Models in Public Health Modelling of infectious disease process: Infectious diseases of human malaria, tuberculosis, hepatitis, HIV/AIDS. Deterministic modelling of infectious diseases.

References

1. Daniel, Wayne W. (1978) Biostatistics: A Foundation for Analysis in the Health Sciences, 3rd Edition, John Wiley & Sons, New York.

2. Finney, D. J. (1997) Probit Analysis, Cambridge University Press, London.

3. Forthofer, Ronald N., and EunSul Lee (1995) Introduction to Biostatistics A Guide to Design, Analysis, and Discovery, Academic Press, San Diego, California.

4. Pagano Marcello and KimberleeGauvereau (2000) Principles of Biostatistics, Duxbury, Thomson Learning Inc., United States. (First Indian Reprint, 2007).

5. Rosner, Bernard (2000) Fundamentals of Biostatistics, Duxbury, Thomson Learning Inc., United States.

3.4 RSTAT-02 Analysis of Biostatistical Data

Cohort and Case Control Studies Cohort studies, case control studies, estimation of prevalence and incidence, Age at onset distributions, Assessing Spatial and Temporal patterns, etc.

Clinical Trials The need of clinical trials, bias and random error in clinical studies, conduct of clinical trials. Overview of Phase I IV trials. Randomized control trial. Ethical issues.

Probit Analysis Concept and definition, dose response analysis.

Contingency Tables The Chi-Square test, McNemars test, the odds ratio, Berksons fallacy. Multiple 2 x 2 tables.Simpsons paradox, the Mantel Haenszel method.

Research Synthesis and Meta-Analysis in Public Health and Medicine Introduction to Meta-Analysis and application of these to data sets of clinical trials and observational studies. Bias in Meta-Analysis.

References

1. Petitti, D. B. (2000) Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine, 2ndedn. Oxford University Press, New York.

2. Piantadosi, S. (1997) Clinical Trials: A Methodological Perspective, John Wiley & Sons, New York.

3. Wolf, F.M. (1986) Meta-Analysis: Quantitative Methods for Research Synthesis, Beverly Hills, C.A.: Sage.

4. Zhou, Xuyu and Ying Lu (2003) Meta-Analysis, In Ying Lu and Ji-Qian Fang (Eds.) Advanced Medical Statistics, World Scientific Publishing Co. Pte. Ltd. Bangalore, 233-317.

3.5 RSTAT-03 Advanced Design of Sample Surveys

Non-sampling errors Incomplete surveys, Hansen and Hurwitz technique, Demings model of the effects of call-backs; modeling observational errors, estimation of variance components, application to longitudinal studies (repetitive surveys); Non-response; Hunsen-Hurwitz estimator. Politz-Simmons technique for Not At Home's, RRT: Warner's model, related and unrelated questions, non-response stratum and double sampling.

Unequal Probability Sampling pps wr/wor methods [including Lahiris scheme] and related estimators of a finite population mean [Hansen-Hurwitz and Desraj estimators for a general sample size and Murthys estimator for a sample of size 2]; PPSWOR sampling method - Horvitz- Thompson estimator; Midzuno scheme; Lahiri-Midzuno-Sen scheme, nonnegativity of Sen-Yates-Grundy estimator: PPS sampling, Hunsen-Hurwitz estimator, Desraj's ordered estimators for WOR selection, Murthy's symmetrized Desraj estimator, Var. estimator.

Variance Estimation method of random groups, balanced half samples (IPNSS), Jack- knife method.

Small Area Estimation Issues in small area estimation - synthetic and generalized regression estimators.

Unified Theory of Sampling non-existence theorems relating to labelled populations. Traditional model-based and Bayesian theories of inference in finite population sampling. Sufficiency, Bayesian sufficiency, completeness. Optimal and various other useful sampling strategies.

Texts

1. Chaudhuri, A. and J.W.E. Vos (1988) Unified Theory and Strategies of Survey Sampling. North-Holland, Amsterdam.

2. Chaudhuri, A. and R. Mukerjee (1988) Randomized Response : Theory and Techniques, New York : Marcel Dekker Inc.

References

1. Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). Sample Survey Methods and Theory, Volume II, John Wiley.

2. Heyday, A. S. and Sinha, B. K. (1991). Design and inference sampling in finite population. Wiley.

3. Mukhopadhyay, P. (1996). Inferential problems in survey sampling. New Age International (P).

4. Mukhopadhyay, P. (1998). Small area estimation in survey sampling. Narosa.

5. Singh, D. and Chaudhary, F. S. (1986). Theory and Analysis of Sample Survey Designs. New Age International Publishers.

6. Wolter, K. M. (1985). Introduction to variance estimation. Springer-Verlag.

3.6 RSTAT-04 Advanced Design and Analysis of Experiments

Partially Balanced Incomplete Block Designs Two-associate PBIB designs - association scheme and intra-block analysis, group divisible designs, dual and linked block designs, resolvable and affine-resolvable designs, general row-column designs - connectedness and intra-block analysis.

Fractional Factorial Designs Fractional replication for symmetric factorials, orthogonal and balanced arrays and their connections with confounded and fractional factorials, optimality of fractional factorials based on orthogonal arrays and construction of orthogonal arrays.

Response Surface Designs Orthogonality, rotatability and blocking; construction and analysis, method of steepest ascent; canonical analysis and ridge analysis of fitted surface.

Regression Designs Optimal regression designs for multiple linear regression and quadratic regression with one explanatory variable; introduction to D-optimal design measure; equivalence theorem.

Texts

1. Dean, A. and Voss,D. Design and Analysis of Experiments.

2. Khuri, A. and Cornell. M. (1991). Response Surface Methodology. Marcel Dekker.

References

1. Bose, R. C. and Shimamoto, T. (1952, 1973). Classification and analysis of PBIB Designs with two associate classes. Jour. Amer. Stat. Assoc. Vol. 47, pp 151-184.

2. Chakravarty, M. C. Mathematics of Design and Analysis of Experiments.

3. Raghavarao, D. Constructions and Combinatorial Problems in Design of Experiments.

4. Wu, C.F.J., and Hamada, M. (2002). Experiments: Planning Analysis, and Parameter Design Optimization, John Wiley & Sons.

3.7 RSTAT-05 Testing of Hypothesis

Neymann-Pearson fundamental lemma, Distributions with ML ratio, Confidence bounds, Generalization of the fundamental lemma.

Least favourable distributions, applications to normal distribution.

UMP unbiased two-sided tests, Applications to exponential families, Fisher-Beherns problem,.

Unbiased confidence sets. Most powerful permutation and invariant tests, Admissibility of tests, Chi-tests and invariance, The Hunt-Stein theorem and its applications.

References

1. T. S. Ferguson, Mathematical Statistics: A Decision Theoretic Approach, Academic Press, 1967.
2. L. Le. Cam, Asymptotic in Statistics, Springer-Verlag, 1990.
3. Jun Shao, Mathematical Statistics, $2^n d$ Ed., Springer, 2003.
4. E. L. Lehmann, Testing Statistical Hypotheses, Wiley, 1986.

3.8 RSTAT-06 Generalised Linear Models and Inference

Data and models Components of a generalized linear model, estimation and fit of the model. Model selection, estimation and prediction; the algorithm for fitting generalized linear models.

Models for Binary Responses Link function, parameter interpretation, retrospective sampling. Likelihood functions for binary data: parameter estimation, deviance function, bias and precision of estimates. Over-dispersion.

Models for Polytomous Responses The multinomial distribution, likelihood functions: log-likelihood for multinomial responses, parameter estimation, deviance function. Over-dispersion.

Log-linear models The Poisson log-likelihood function, over dispersion, asymptotic theory, multiple responses: canonical correlation models, multivariate regression models, log-linear regression models, Likelihood equations.

Model Diagnostics for Generalised Linear Models Score tests for extra parameters, checks for systematic departure from the model: residuals, variance function, link function, scale of covariates, checks for compound discrepancies. Checks for isolated departure from the model: measure of leverage, consistency.

Texts

1. McCullagh, P and Nelder, J. A. Generalized Linear Models. Chapman and Hall.

References

1. Agresti, A. Introduction to Categorical Data Analysis. Wiley.

3.9 RSTAT-07: Bayesian Inference

Natural Conjugate family of priors for a model Hyper parameters of a prior from conjugate family. Conjugate families for (i) exponential family models, (ii) models admitting sufficient statistics of fixed dimension. Enlarging the natural conjugate family by (i) enlarging hyper parameter space (ii) mixtures from conjugate family, choosing an appropriate member of conjugate prior family. Non informative, improper and invariant priors. Jeffreys invariant prior.

Point Estimation Bayesian Point Estimation as a prediction problem from posterior distribution. Bayes estimators for (i) absolute error loss (ii) squared error loss (iii) 0 - 1 loss. Generalization to convex loss functions. Evaluation of the estimate in terms of the posterior risk.

Bayesian Interval Estimation Credible intervals. Highest posterior density regions. Interpretation of the confidence coefficient of an interval and its comparison with the interpretation of the confidence coefficient for a classical confidence interval.

Bayesian Testing of Hypothesis Specification of the appropriate form of the prior distribution for a Bayesian testing of hypothesis problem. Prior odds, Posterior odds, Bayes factor for various types of testing hypothesis problems depending upon whether the null hypothesis and the alternative hypothesis are simple or composite. Specification of the Bayes tests in the above cases. Discussion of Lindleys paradox for testing a point hypothesis for normal mean against the two sided alternative hypothesis.

Bayesian Calculations for Non Conjugate Priors (i) Importance sampling, (ii) Obtaining a large sample of parameter values from the posterior distribution using Acceptance - Rejection methods, Markov Chain Monte Carlo methods and other computer simulation methods.

Texts

1. Berger, J. O. Statistical Decision Theory and Bayesian Analysis, Springer Verlag.

References

1. Robert C. P. and Casella, G. Monte Carlo Statistical Methods, Springer Verlag.
2. Leonard T. and Hsu, J. S. J. Bayesian Methods. Cambridge University Press.
3. Gemerman, D. Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference, Chapman Hall.
4. Press, S. J. (2002). Subjective and Objective Bayesian Statistics: Principles, Models, and Applications, 2nd Edition, Wiley.

3.10 RSTAT-08: Theory of Estimation

Elements of Decision Theory Complete class theorem, admissibility of Bayes rule, Minmax Theorem.

Estimators Review of sufficiency, consistency and efficiency. UMVU estimators and their properties. Application to normal and exponential one and two sample problems.

Information Inequality Multiple parameter case; Equivariance, Invariance. Application to location and scale families. Minimum Risk Equivariant (MRE) estimation.

Bayes and Minimax Estimation Estimation for Exponential families. Admissibility of estimators, Blyth's ratio method, Karlin's sufficient conditions.

Pitman's estimator Properties, Simultaneous estimation. Stein's phenomenon, Shrinkage estimation.

References

1. J. Berger, Statistical decision theory, Springer-Verlag, 1980.
2. T. S. Ferguson, Mathematical Statistics: A Decision Theoretic Approach, Academic Press, 1967.
3. E. L. Lehmann, Theory of Statistical Inference, Wiley, 1983.
4. S. Zacks, The Theory of Statistical Inference, Wiley, 1971.
5. Jun Shao, Mathematical Statistics, 2^{nd} Ed., Springer, 2003.

3.11 RSTAT-09: Operations Research : Stochastic and Deterministic Models and Optimization

Operations Research Models Overview of operations research models; Overview of Linear Programming; Integer programming, Multi Objective programming problems, Dynamic programming problems .

Queuing Models Overview of Queuing Models and their applications, Replacement models.

Nonlinear Optimization Techniques Overview of Results in Nonlinear optimization techniques. Unconstrained and constrained optimization techniques: Gradient method, Kuhn-Tucker conditions. Monte Carlos techniques.

Inventory Models Stochastic and Deterministic Inventory Models. Models for Deteriorating Items, Multi- echelon models.

References

1. Optimization Theory and Applications: S.S. Rao, Wiley Eastern.
2. Nonlinear Programming Theory and Algorithms: Bazaraa, M.S. and Shetty, C.M., John Wiley Sons.
3. Operations research: An Introduction: Taha, H. A., Pearson.
4. Elements of Applied Stochastic Processes: U.N. Bhat, John Wiley.