UNIVERSITY OF RAJSHAHI

Rajshahi – 6205, Bangladesh

Syllabus
for
M.Phil./Ph.D. in Statistics

Session: 2023 – 2024

DEPARTMENT OF STATISTICS

www.ru.ac.bd/stat/

University of Rajshahi
Faculty of Science

Department of Statistics
Syllabus for M. Phil. /Ph. D. in Statistics
Session: 2023 - 2024

As per the ordinance of the University of Rajshahi, the requirements for the degree of M.Phil./Ph.D. are (a) course work of 200 marks, (b) at least a seminar presentation, (c) submission of a thesis on an approved topic and (d) an oral or both oral and practical examination.

The M.Phil./Ph.D. course in Statistics shall consist of Two theory papers with 200 marks, each carrying 100 marks. The duration of the examination is four hours for each paper. The pass mark of each paper is 45 percent. The coursework must be completed in the First/Second year of admission.

The titles of the papers are as follows:

Stat. 601: Advanced Statistical Inference
Stat. 602: Advanced Order Statistics
Stat. 603**:** Advanced Stochastic Processes
Stat. 604: Advanced Regression Diagnostics
Stat. 605: Advanced Survival Analysis
Stat. 606**:** Contemporary Survey Sampling
Stat. 607: Population Studies
Stat. 608: Advanced Time Series Analysis
Stat. 609: Advanced Environmental Statistics
Stat. 610: Advanced Statistical Methods and Development Studies
Stat. 611: Advanced Growth Modeling and Anthropology
Stat. 612: Statistical Genomics and Bioinformatics
Stat. 613: Statistical Proteomics and Bioinformatics
Stat. 614 : Machine Learning and Deep Learning
Stat. 615: Research Methodology and Scientific Writing

Fellow(s) shall take any Two of the above courses subject to the recommendation of the Supervisor(s) and approval of the Academic Committee of the Department.

The M.Phil./ Ph.D. degree in Statistics shall be awarded in accordance with the Ordinances of the University of Rajshahi.

Stat. 601: Advanced Statistical Inference

Full Marks: 100 **Examination: 4 Hours**

Review of Theory of Estimation and Hypothesis Testing.

Bayes estimation and Minimax Estimation: Prior and posterior densities; Theorem on conjugate family of prior density; Structure of Bayes estimates for quadratic and convex loss functions. Theorem of Girshick and Savage. Theorem of De Groot and Rao relating to the above loss functions. Definition of minimax estimation. Important theorems on minimax estimation.

Admissibility of Estimation: Basic theory of admissibility and complete class. Admissibility under quadratic loss. Concept of equivalent estimation. Pitman Estimation.

Interval Estimation: Bayesian procedure of Interval Estimation. Shortest confidence interval. Expected lengths of a confidence interval.

Test of Hypothesis: Statement of the problem, Generalized Neyman Pearson Lemma. Locally most powerful unbiased (LMPU) test. Testing of a composite hypothesis. Fisher-Behren's problem. Asymptotic efficiency of a test. Similar Region and Neyman structure. Multiple hypothesis testing, monotone procedure of testing. Empirical Bayes testing of multiple hypothesis. Bayes optimality of the Wald sequential probability ratio test.

Decision Theory: Statement of the problem. Formulation of the problem of inference in the light of decision theory and its relationship with two persons Zero-Sum game. Problem of point estimation as a decision function. Randomised and Non-Randomised decision rules. Bayes and minimax decision rules. Complete class of decision rule. Admissibility of Bayes rules. Separating Hyperplane theorem, Minimax theorem. Complete class theorems.

Non Paramatric test: Statement of a Non-Paramatric test. ARE and Robustness of a non-paramatric test. McNemar test in a 2X2 contingency table. Cox and Stuart test for trend. Cremer's contingency coefficient. Cochran's test for related observations. ARE of Mann-Whitney test w.r. to t-test. ARE of sign test w.r. to one-sample students t-test. Kruskal-Walls test and C.R.S. Design. Square Ranks Test for variances. Wilcoxon signed rank test. Quade test, Firedman test, BIB Kolmogorov test.

References:

Bkacjwekkm D. Girshick, M.A. (1954): Theory of Games & Statistical Decision, Wiley, NY.

Chernoff, H. & Moses, L.E. (1959): Elementary Decision Theory, Wiley, N.Y.

Conover, W.J.: Practical Nonparametric Statistics, 2nd Ed., Wiley, N.Y.

Ferguson, T.S. (1967): Mathematical Statistics, A Decision Theoretical Approach, Academic Press, New York.

Gibbons J.D.: Nonparametric statistical inference.

John, P.W.M.: Statistical Design and Analysis of Experiments.

Lehmann, E.L.(1959): Testing Statistical Hypothesis, Wiley, N.Y.

Mood, O. Graybill, F. & Boes.: An introduction to Mathematical Statistics, N.Y.

Rao, C.R. (1965): Linear Statistical Inference and Its Applications, Wiley, N.Y.

Scheffe, H.: Analysis of Variance.

Weiss, L. (1961): Statistical Decision Theory, McGraw Hill, N.Y.

Zacks, S. (1971): Statistical Inference, Wiley, N.Y.

Stat. 602: Advanced Order Statistics

Full Marks: 100 **Examination: 4 Hours**

Review of Preliminaries of order statistics: Distribution-free continence intervals for quantities, Distribution–free tolerance intervals, Conditional distributions, Order statistics as a Markov chain, Order statistics for independent non-identically distributed variates. Concomitants of order statistics. Expected values and moments: Basic formulae, recurrence ratios and identities, and results for Uniform, Exponential, Logistic, Gamma, Weibull, Normal and Haley Logistic distributions.

Bounds and approximations for moments: Distribution – free bounds for the moments of order statistics and of the range, bounds for the expected values of order statistics in terms of quantities of the parent distribution, Universal bounds and series approximations.

Order statistics in estimation:- Least squares estimation of location and scale parameters by order statistics. Estimation of location and scale parameters for censored data.

Asymptotic Theory: Asymptotic joint distribution of quantiles, asymptotic distribution of the extreme, asymptotic distribution of linear functions of order statistics, asymptotic theory of concomitants of order statistics.

References:

Ali, M. A. (1994). Moments of Order Statistics and related results, Ph.D. Thesis AMU, India.

Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1992). A First course in order statistics, John Wiley, New York.

Balakrishnan, N. and Cohxn, A.C. (1991). Statistics and Inference: Estimation Methods, Academic, Boston.

Begum, A.A. (1998). Moments of Concomitants of order Statistics, Ph.D. Thests, AMU, India.

David, H.A. (1981): Order Statistics, 2nd ed, Wiley, New York.

Mood, A.M. Graybill, F.A. and Boes, B.C. (1974). Introduction to the Theory of Statistics, 3rd ed. McGraw Hill, N.Y.

Stat. 603: Advanced Stochastic Processes

Full Marks: 100 **Examination: 4 Hours**

Probability and Stochastic Processes: Introduction, Probability Space, Random Variables, Expectation, Independence, Conditional Expectation, Conditional Probability, Limit Theorems, Basic Properties of Stochastic Processes, Stationary Processes with Independent Increments, Markov Processes, Diffusions Processes, Wiener Process and White Noise.

Basic Theory of Stochastic Differential Equations: Ordinary, Random and Stochastic Differential Equations, Existence and Uniqueness of Solutions; Dependence on Parameters and Initial Conditions, Ito's Formula and Moments of Solutions of Stochastic Differential Equations, Solutions of Stochastic Differential Equations as Diffusion Processes.

Stochastic Modeling: Models and Modeling, The Markov Model for Social Mobility, Some Generalizations of the Markov Model, Interactive Markov Chain Models, Maintainability in a Deterministic Environment, Attainability in a Deterministic Environment, Maintainability in a Stochastic Environment, Attainability in a Stochastic Environment. Growth and Aging

Forestry Management: The Faustmann Model, A Model of Optimal Forest Thinning, The Beverton-Holt Fisheries Model, The Logistic Growth Model, Generalized Logistic Models: Depensation, Non-linear Fishery Model, Non-linear Forestry Model, Economic Models of Renewable-Resource Harvesting.

References:

Bartholomew, D.J : Stochastic Models for Social Processes John Wiley & Sons, 3rd Edition.

Colin W. Clark: Mathematical Bio-economics: The Optimal Management of Renewable Resources John Wiley & Sons.

Narendra. s. Goel and Nira Richter. Dyn : Stochastic Models in Biology Academic Press, New York.

Robert Gilmore: Catastrophe Theory for Scientists and Engineers John Wiley & Sons, New York.

Tracy R. Lewis: *Stochastic Modeling of Ocean Fisheries Resource Management.*

Stat. 604: Advanced Regression Diagnostics

Full Marks: 100 **Examination: 4 Hours**

Introduction: Preliminaries. Departure from classical assumptions. Shortcomings and modifications of least squares residuals. The Hat matrix. Influential observations, high leverage points and outliers.

Diagnostic Quantities: Deletion of observations. Measures of influence. Cook's distance. DFFITS. DFBETA. MDFIT. Outlier diagnostics. Measures of high leverage points.

Tests for Normality: Basic ideas. Moments, skewness and kurtosis of OLS residuals. Tests on residual plots. Shapiro-Wilk test. Anderson-Darling test. Bowman-Shenton test. Rescaled moments test, Tests on other residuals.

Diagnostic Plots: Residual plots. Normal probability plot. Leverage-Residual plots. Added variable plots. Partial regression leverage plots. Partial residual plots.

Robust Regression: Group deletion. Masking and swamping. The Breakdown point and Robust estimators. The Least Median of Squares (LMS) technique. The Reweighted Least Squares residuals. Detection of multiple outliers.

Further Diagnostic Quantities: Transformation diagnostics. Logistic regression diagnostics, Diagnostics in nonlinear regression, Diagnostics for generalized linear models.

References:

Atkinson, A.C.: Plots, Transformations, and Regression. Clarendon Press.

Barnett, V. & Lewis, T.: Outliers in Statistical Data, 3rd Ed., Wiley, N.Y.

Belsley, D.A., Kuh, E. & Welsch, R.E.: Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, Wiley, New York.

Chatterjee, S. & Hadi, A.S.: Sensitivity Analysis in Linear Regression, Wiley, N.Y.

Chatterjee, S. and Hadi, A.S.: Regression Analysis By Example, Wiley, N.Y.

Cook, R.D. & Weisberg, S.: Residuals and Influence in Regression, Chapman & Hall, London

D'Agostino, R.B. & Stephens, M.A.: Goodness-of-fit Techniques, Marcell-Dekker, New York.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. & Stahel, P.J.: Robust Statistics, The Approach Based on Influence Function, Wiley, N.Y.

Hawkins, D.M.: Identification of Outliers, Chapman & Hall, London.

Huber, P.J: Robust Statistics, Wiley, New York.

McCullagh, P. & Nelder, J.A.: Generalized Linear Models, Chapman & Hall.

Montgomery, D.C., Peck, E.A. and Vining, G.G. (2003): Introduction to Linear Regression Analysis, 3rd ed., Wiley, N.Y.

Mosteller, F. & Tukey, J.W.: Data Analysis and Regression: A Second Course in Statistics, Addison-Wesley, Massachusetts.

Rousseeuw, P.J. & Leroy, A.: Robust Regression and Outlier Detection, Wiley, New York.

Ryan, T.P. (1997): Modern Regression Methods, Wiley, N.Y.

Rosseeuw, P.J. and Leroy, A. (1987): Robust Regression and Outlier Detection, Wiley, N.Y.

Seber, G.A.F and Wild, C.J. (1989): Nonlinear Regression, Wiley, N.Y.

Stat. 605: Advanced Survival Analysis

Full Marks: 100 **Examination: 4 Hours**

Elements of Biometrics: The Sample size, Study Designs- Cross sectional, Retrospective and Prospective study Designs, The Odds, Ratio- properties, estimation & test. Attributable Risk & Relative Risk- estimation, test and large sample confidence interval. The Screening Test, Misclassification error, Predictive value, Sensitivity & Specificity.

Linear Models: Classical linear models, Probit models, Logit models. Multinomial models for multivariate binary responses, Loglinear models, Generalized linear models, Likelihood and Quaxi likelihood methods of estimation, Generalized estimating equations, Hypothesis Testing.

Logistic Regression Model: Development of the model, Interpretation of parameters in terms of odds ratio, Models with single indicator and dummy explanatory variables, Models with multiple explanatory variables, Models with mixed explanatory variables. Inference procedures. Application of the model in cohort and case control studies.

Statistical Methods For Continuous PH Regression: Models Conditional, Marginal and Partial Likelihood, Estimation and test of regression vector, Comparison of two or more survival distributions, Estimation of survivor function, Model cheek and data analysis.

Regression Methods For Grouped Data: Life table models with regressor variables. Grouped PH model, The Logistic model, Estimation and test of regression vector, Testing equality of survival distributions, Time dependent regressor variables, Model cheek and data analysis.

Nonparametric and Distribution Free Methods: Estimation of survivor functions and quantiles, Standard error of estimates and confidence interval construction for censored and uncensored data. Linear Rank Test for comparing two or more survival distributions with uncensored & censored data. Exponential ordered score (Log Rank) test with censored data, Generalized Wilcoxon (Wilcoxon and Kruskal Wallis) test with censored data.

Markov Chain Models: Bernoulli trials with Markov dependence, Correlated Bernoulli trials, Stationary Markov chain. Markov chain model with serially dependent observations, Markov models for covariate dependence of binary sequences, Estimation and test of parameters, Test for Markovity.

Multivariate Survivor Models: Survivor function, hazard function and probability density function, Marginal and conditional density functions, Study of bivariate models, Nonparametric inference procedures, Special case with PH models Study of Competing Risk as a special case of multivariate survivor models.

References:

Bhat, U.N.(1988): Elements of Stochastic Process, Wiley

Cox, D.R.(1970): Analysis of Binary Data, Methuen, London

Cox,D.R. and Hinkley, D.V(1974): Theoretical Statistics, Chapman & Hall.

Fleiss, J.L (1981): Statistical Method for Rates and Proportions, Wiley. N.Y.

Hosmer, D.W & Lemeshow, S (1989): Applied Logistic Regression, Wiley, N.Y

Karim, M. R. and Islam, A. (2019): *Reliability and Survival Analysis*, Springer, Singapore.

Lee, E.T. (1980): Statistical Methods For Survival Data Analysis, Belmont: Wadsworth, Inc.

Lawless, J.F.(1982): Statistical Models & Methods For Lifetime Data, Wiley, N.Y.

McCullagh, P and Nedler, J.A.(1983): Generalized Linear Models, Chapman & Hall, London.

Mehedi, J. (1981): Stochastic Process, Wiley Eastern Ltd. New Delhi.

Stat. 606: Contemporary Survey Sampling

Full Marks: 100 **Examination: 4 Hours**

Sampling: Evaluation of various sampling procedures. Sampling Policy. Desiderata of Statistical Survey design. Determination of Sample size relevant to Sampling Schemes. The Replacement and Nonreplacement Sampling Lattice Sampling. Problems of Randomization. Building Sampling Frame. Two- Way Frames. Multiple Frame, Imperfection of Sampling Frames, Ideal Conditions for Probability Sampling, Multipurpose descriptive surveys. Cost oriented sampling schemes. Sampling though time. Statistical Techniques for Analysis of Survey Data.

Survey Methodology: Survey Method vs. Scientific Method. Survey planning and General planning. Survey Design and Structure. Pilot Survey. Multirecord Survey (MRS) and Dual Record System (DRS) and their comparison with SRS. Coverage Errors Models- M_g (the General Model), M_o (the Equal Catchability Model) M_b (the behavioral Response Model), M_t (the Patterson Model). M_t Models for Survey Design, Model for Response Errors, Response Bias and Reliability in sensitive topic surveys. Assessment of the quality of survey data. Strategies for the analysis of imputed data from a sample survey. Types of problems in investigations based on surveys. The Missing and Hard- to- Get Data- Effect and Estimation,. Significant and Interpretation of Non-Sampling Errors

Methodology and Questionnaire Related to Census and National Samples Surveys (Like WFS, BFS, BDHS, BNS, BAS and so on) of Bangladesh

References:

Cochran, W.G: Sampling techniques, 3rd ed. Wiley, N.Y.

Carl-Erik Sarndal, Bengt Swensson and Jan Wretman: Model Assisted Survey Sampling, Springer, NY

Kish, L.: Survey Sampling, Wiley.

Raymond J. Jessen: Statistical Survey Techniques, Wiley, NY.

D. J. Bogue, E.E. Ariaga and D. L. Alderton (Eds): Reading in Population Research Methodology. Vol. I, Basic Tools, UNFPA

Stat. 607: Population Studies

Full Marks: 100 **Examination: 4 Hours**

Population Study: Field of Population Study. Study of Population Size, Composition, Distribution and Structure. Population Trends (Time Series) and Growth. Socio-Economic Characteristics of Population. Population Study and Demography.

Demographic Analysis: Some Basic Demographic Methods -(Balancing equation, Rates, Ratios and Probabilities, Life Tables, Demographic Models, Standardization and Decomposition). Analysis of Age - Sex Data, Mortality and Morbidity, Fertility and Reproduction, Nuptiality and Marriage , Migration and Urbanization Statistics. Direct and Indirect Measurement (from Stock and Flow type Data) Estimation of Fertility and Mortality using various Indirect Techniques.

Demographic Models: Mortality Model -- UN Model Life Tables. Coale-Demeny Model Life Tables. Lederman's System of Model Life Tables. Brass Logit Life Table System. UN Life Tables for Developing Countries. Nuptiality Model -- Coale's Parameters of Nuptiality. Coale-McNicol Model. Fertility Model Coale- Trussel Model. Brass Relational Gompertz Model. Age-Period Cohort (APC) Model, Davis-Blake Framework for Fertility Analysis. Bongaart's Framework.

Demographic Data Screening: Evaluation and Adjustment of Age, Sex, Fertility, Mortility and other Demographic Data by various Screening Procedures. Graduation Methods. Evaluation of Sources. Factors Important in Demographic Analysis.

Population Problems and Issues: The Growth of World Population. Demographic Trends. Demographic Transition. Population Models- Exponential, Mulltusion and Logistic Models. Stationary, Stable and Quasi-stable Models. Problems and Issues related to population growth.

Understanding Population Problems: Population Trend, Problems related to Population Growth, Reproductive Health, Ageing, etc. Some contemporary population problems in Bangladesh. Population Policies and Programmes. Family Planning and other programmes.

References:

UNFPA (1993). Population Research Methodology Vols. 1-8. Chicago, Illinois, and other UNFPA publications.

UN Publications. Manual IV and Manual X, Population Bulletins, Population Debate and UN others.

ESCAP Publications.

Journals/Reports: Population Studies, Demography, Population and Development Review, Genus, Social Biology, Biosocial Science, ESCAP Population Journals, Development (SID), JASA, JRSS, Biometrica, Demography-India, Applied Statistics and some other journals, Reports relevant to study & research.

Stat. 608: Advanced Time Series Analysis

Full Marks: 100 **Examination: 4 Hours**

Lag structures: Distributed Lags, Disaggregation. Autoregressive Process (AR), Moving Average (MA) Process. ARMA Process.

Unit roots: Time series with Unit roots, non-Stationarity, Co-integration. Autoregressive Integrated Moving Average (ARIMA) modeling. Forecasting and testing in co-integrated systems. Maximum likelihood estimation and inferences on co-integration. Distribution of estimators for AR time series with a unit root. Likelihood ratio statistic for AR time series with unit root.

Error Correction: Co-integrated and error correction representation, estimation and testing. Error correction model and agricultural supply response.

Causality: Inference and causality in economic time series models. Causality and the Granger test. Forecasting with ARIMA and vector Autoregressive (VAR) models.

Seasonality: Seasonality and Unit roots. Seasonal unit root in aggregate data. Seasonality non-Stationarity and forecasting. Seasonality and Stochastic trend. Seasonal integration and co-integration.

Neural network: Neural network analysis and time series analysis. Chaotic time series and it's predictivity by neural network analysis.

References:

Fuller, W.A. (1976): Introduction to Statistical time series, Wiley, N.Y.

Fuller, W.A. Non-stationary Auto-regressive Time Series, North Holland.

Harvey, A.C. The Econometric Analysis of Time Series, Philip Allan.

Kosko B. (1994) Neural Network and Fuzzy System. Prentice Hall of India, New Delhi.

Rao V. and Rao H. (1996) C++ Neural Network and Fuzzy Logic. BPB Publications, New Delhi.

Wasserman P.D. (1989): Neural Computing. Van Nostrand Reinhold, NY.

Stat. 609: Advanced Environmental Statistics

Full Marks: 100 **Examination: 4 Hours**

Fundamentals of Environment: Definition, basic term, components and functioning of environment.

Major Issues of Environment: Environmental pollution (land, water and air), Green house effect.

Environmental Extreme Events: Natural hazard and disaster, flood, tropical cyclone, drought, earthquake, deforestration and agricultural practices.

Agroclimatic Classification: Rainfall, temperature, solar radiation evapotranspiration, humidity, moisture availability index (MAI), drought index (DI), soil moisture deficit and its impact on agricultural crops. Application of these two important statistical tools for analysis of crop-climate relationships.

Probability Distributions: Gamma, Exponential, Mixed Gamma, Normal, Poison, Binomial, Eggenberger-Polya distribution and its application to environmental factors.

Markov Chain Model: Review of Markov Process, Higher Transition Probabilities, Classification of States and Chains, Determination of Higher Transition Probabilities, stability of a Markov System: Limiting Behaviour, Statistical Inference for Marhov Chains, Markov Chains with continuous State Space, Non-Stationary or Non-homogeneous Chains. Applications of Markov chain model for environmental events.

GLS and Simulation Modeling: Geographical Information Systems (GIS) Simulation Modeling for environmental research.

References:

Abrol, Y.P., 1999, Rice-in a variable climate. APC Publications Pvt. Ltd., New Delhi.

Bachelet, D. and Kropff, M.J., 1995, The impact of climate change on agroclimatic zones in Asia. CAB International, U.K.

Bratley, P.,B.L. Fox and E.L. Schrage, 1983, A guide to simulation. Springer Verlag, New York.

Commission on the 29th IGC Climatology, 2000, Proceedings of International Symposium of Climate Change Variability and their Impacts, August 9-13, 2000, Konkuk University, Seoul, Korea.

Ghairs, D.D., Environmental Science: A Framework for Decision Making.

Lamb, H.H. 1972, Climate: Present, past and future. Methnen, London.

Law, A.M., and W.D. Kelton, 1991, Simulation and modelling. McGraw-Hill, Inc.

Lockwook, World Climatology: An Environmental Approach, O,U.P.

Papadakis, J., 1975, Climate of the World and Their Agricultural Potentialities, Buenon Aires., Agrentina.

Unwin, D., Introductory Spatial Analysis, Methuen, New York.

Vose, D., 1996, Quatitative risk analysis: A Guide to Monte Carlo Simulation Modelling., John Wiley & Sons, New York.

Stat. 610: Advanced Statistical Methods and Development Studies

Full Marks: 100 **Examination: 4 Hours**

Method of Multivariate Analysis: Principal Component Analysis, Factor Analysis, Canonical Correlation Analysis, Discriminate Analysis, Cluster Analysis, Multivariate Analysis of Variance. Screening of Data for Multivariate Analysis.

Input Output Analysis: Assumptions for the General Solution to the Input-Output Models. Backward and Forward Linkages. Input Output Linkages Analysis and Development Strategy.

Obstacles to Development: Dualism, The Process of Cumulative Causation, Regional Growth Rate Differences. The Prebisch Model, The Seers Model. Theories of Dependence and Unequal Exchange.

The Environment and Development: Models of the Environment and Economic activity. The market based approach to Environmental analysis. Measures of Environmental values. Economic growth and Environment. Rural development and Environment. Urban development and Environment. Framework for Analysis of Population, Environment and Development Interrelationships. Sustainable Development and Environment. Brundtland Report on Environment. Environment-Development Paradigm.

Education and Development: Public educational Expenditure. Economics of Education and Employment. The relation between Employment opportunities and Educational demand. Education and Economic Growth. Education, Inequality and Poverty. Education, Internal Migration and the Mrain Drain. Education and Rural Development. Education Modelling. Education and Development Planning. Role of UNESCO.

References:

Anderson, T.W. An Introduction to Multivariate Statistical Analysis.

Srivastava, M.S. & Carter M. An Introduction to Applied Multivariate Statistics.

Thirlwall A.P. Growth and Development 5th Edition ELBS. with Macmillan 1994.

Trodo M.P. Economic Development. 5th edition, Longman 1995.

UN Population Debate, 1978.

Stat. 611: Advanced Growth Modeling and Anthropology

Full Marks: 100 **Examination: 4 Hours**

Growth modeling: Introductions: prenatal, neonatal and postnatal growth. Early childhood, mid-childhood and adolescent growth. Different parametric and nonparametric growth models: Allometry Model, the Gompertz and logistic growth models, Jenss model, Count model, double logistic model, PB models, ICP model, Reed models, SSC model, JPPS model, JPA-2 models, modified ICP model, BTT model and Kernel's (non-parametric) model, Wavelet, polynomial model, simulated growth models. Prediction of future growth. Analysis of predicting errors.

Anthropology: Introduction, Social and physical Anthropology. Their impact on Human Growth. Measurement of Craniometry and Osteometry. Eliptical Faurier Analysis and Weblet Analysis.

References:

Alex F. Roche (1992): Growth Maturation and Body Composition: The Longitudinal Study 1029.1991, Cambridge University press

Beck,R.D.dueToit SHC and Thissen,D.(1994): AUXAL: Auxological Analysis of Longitudinal Measurements of Human Stature, Chicago: SSI

Carterer, J.E.L and Heath B.H.(1990): Somatotyping Development and Applications, Cambridge University Press, New York.

Johnston, F.E, Roche, A.F. and Susanne, C. (1980): Human Physical Growth and Maturation Methodologies and Factors, Plenum Press, New York.

Juan Comas(1960): Manual of Physical Anthropology, Charies C. Thomas Publisher, Springfield, Illinois, USA .

Hastic, T.,Tibshirani ,T., and Friedman,J.(2000): The Element of Statistical Learning: Data Mining Inference and Prediction, Springer, New York

Falkner, F. and Tanner, J.M.(1978): Human Growth, vol. I: Principles and Prenatal Growth, Plenum Press, New York and London.

Falkner, F. and Tanner, J.M.(1978): Human Growth, vol.3: Neurobiology and Nutrition, Plenum Press, New York and London.

Falkner, F. and Tanner J.M.(1978): Human Growth, vol.3: Neurobiology and Nutrition, Plenum Press, New York.

Rao, V.V.(1992): Biostatistics.

Spiegelman,M.(1968): Introduction to Demography, North Holland.

Shephard, R.J.(1991): Body Composition in Biological Anthropology, Cambridge University Press, New York.

Stevens, J.(1996): Applied Multivariate Statistics for Social Science, third edition. Lawrence Erlbaum Associates, Inc, publishers. Mahwah, New Jersey.

Tanner, J.M.(1962): Growth at Adolescence, Blackwell Scientific Publication, second edition. Alden Press, Oxford.

Stat. 612: Statistical Genomics and Bioinformatics

Full Marks: 100 **Examination: 4 hours**

Basic Genomics: Introduction, Classification of Genomics. DNA, RNA, Gene, Genome and Chromosomes. Cell division, Linkage and mapping, Molecular genomics. Genotype and Genotyping technology. Central dogma and bioinformatics. Overview of Gene-Expression and Microarray.

DNA Sequence Analysis: DNA sequencing. Classification of sequencers including NGS. Analysis of DNA patterns, overlaps counted, overlaps not counted and motifs, sequence accuracy, sequence formats. Conversions of one sequence format to another. Single and multiple sequence alignments approaches. Phylogenetic analysis approaches. Some bioinformatics databases including GenBank, DDBJ and NCBI, EMBL. BLAST and FASTA analysis.

Genetic Linkage Analysis: Introduction, Mendelian segregation, Segregation patterns in a full-sib family. Genetic distance, Recombination fraction and map functions. Two-point analysis for backcross and F2-intercross, three-point analysis, Multilocus likelihood and locus ordering, Estimation with many loci, Mixture likelihoods and order probabilities. Linkage Analysis with Controlled Crosses and Recombinant Inbred Lines (RILs).

Quantitative Trait Loci (QTL) Analysis: .Introduction. Marker Analysis of Phenotypes. Whole-Genome Marker Analysis. The Structure of QTL Mapping. Interval Mapping Approaches for QTL Analysis (Linear regression and maximum likelihood approaches with backcross and F2 populations). Composite and multiple interval mapping approaches for QTL analysis.

Gene-Expression Analysis: Basic concept. Preprocessing (Transformation, normalization, Image analysis and filtering). Identification of differential expressed (DE) genes in two or more groups using statistical test. Clustering and Classification for Gene-Expression Data Analysis. Modeling genetic regulatory networks using gene expression profile. Gene-set enrichment analysis. MicroRNA- concepts, application and prediction..

Genome-wide Association Studies (GWAS): Introduction. QTL and Single Nucleotide Polymorphism (SNP) analysis with gene-expression data. SNP analysis using contingency table. Examples of disease related SNPs. GWAS using linear mixed models and GeneABLE. Haplotype estimation. Regional multilocus association models. Linkage disequilibrium and tagging. Practical guide to linkage disequilibrium analysis and tagging using Haploview.

Metagenomics : Definition, Evolutionary Significance, Gene conservation between species, Difference of adaptive traits, Species diversity of microflora, Shotgun meta genomics,.

Toxicogenomics: Use of transcriptomics in toxicogenomics and environmental toxicology, Analysis, Modeling and interpretation of toxicogenomic data, Principles of data mining in toxicogenomics, *Escherichia coli* stress response as a tool for detection of toxicity.

Main Books:

Ben Hui Liu and Leming M Shi. (2013). Statistical Genomics and Bioinformatics, Chapman and Hall/CRC press, 2nd edition, New York.

David W. Mount. (2004): Bioinformatics: Sequence and Genome Analysis, Second Edition, Publisher: Cold Spring Harbor Laboratory Press.

Recommended Books:

Shizhong Xu (2013). Principles of Statistical Genomics. Publisher: Springer; 1 edition.

Ben Hui Liu (2008). Statistical Genomics: linkage, Mapping and QTL analysis. Chapman and Hall/CRC press, New York.

Rongling W., Changxing M., George C. (2007). Statistical Genetics of Quantitative Traits: Linkage, Maps and QTL (Statistics for Biology and Health). Springer; 1.

Benjamin M. N., Manuel A.R. Ferreira, Sarah E. and Medland, Danielle P. (2008). Statistical Genetics: Gene Mapping Through Linkage and Association. Taylor & Francis; 1 edition.

Sorin I. , Michael W., and Andrew C., (2004). Computational Methods for SNPs and Haplotype Inference. Springer; 1 edition (April 28, 2004)

Cedric G., Julius van der W., and Hayes B. (2013). Genome-Wide Association Studies and Genomic Prediction. Springer; 1 edition..

Robert G., Vincent C., Huber W , Rafael I., and Sandrine Dudoit (2005). Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Statistics for Biology and Health). Springer; 1 edition

Robert Gentleman (2008). R programming for bioinformatics (Chapman & Hall/Crc Computer Science & Data Analysis). Chapman & Hall/CRC; 1 edition.

Alberts B., Johnson A., Lewis J., RaffM. , Roberts K., and Walter P., (2007). Molecular Biology of the Cell: Reference Edition, , Garland Science; 5 edition.

Sahu, Saura C., (2008). Toxicogenomics : a powerful tool for toxicity assessment. Chichester, West Sussex : John Wiley & Sons.

Burczynski, Michael E., Boca Raton, FL, (2003), An introduction to toxicogenomics, CRC Press.

Sahu, Saura C., (2008). Toxicogenomics : a powerful tool for toxicity assessment. Chichester, West Sussex : John Wiley & Sons.

Diana Marco (2010). Metagenomics: Theory, Methods and Applications. Publisher: Caister Academic Press.

Wolfgang Streit (Editor), Rolf Daniel (2010): Metagenomics: Methods and Protocols (Methods in Molecular Biology) **.** Publisher: Humana Press.

Michael E. Burczynski (2013). An Introduction to Toxicogenomics. Publisher: CRC Press.

Stat. 613: Statistical Proteomics and Bioinformatics

Full Marks: 100 Examination: 4 hours

Basic Proteomics and Bioinformatics: Concepts, historical background, importance, scopes and applications of proteomics and bioinformatics, Cell Structure and Function, Cell components. Chromosome, Chromosome structure and organisation. Modern concept of central dogma.

Amino Acids and Protein Structure: Amino acids, Types and Structure of Amino Acids, Properties of Amino Acids, Codons for amino acids. Protein folding and factors for Protein Folding, primary, secondary, tertiary and quaternary structure, protein motifs and domain.

Biological Databases: Overview of the use and maintenance of different databases in common use in biology. Databases: GenBank, EMBL NCBI, Swiss-Prot, PDB, ExPASy, EBI and KEGG.

Sequencing and Sequence Alignments: Sequencing techniques, Long repeats, r-scan, Definition of sequence alignment, Significance of sequence alignment, Overview of methods of sequence alignments. Pairwise and Multiple Sequence alignments. Comparison of local and global alignments. BLAST, FASTA and ClastalOmega.

Phylogenetic Analysis of Protein Sequences: Motivation and background on phylogenetics**,** Distance and clustering approach, Maximum Likelihood methods**,** Maximum Parsimony, Phylogenetic Tree Estimation, Evolutionary perspectives of phylogenetic tree.

Protein Structure-function relationship: Methods for predicting 3D structure of proteins; simulation of protein functions from structural analysis. Protein identification by MS/MS – software tools for MS/MS identification: SEQUEST, MASCOT, OMSSA, X!HUNTER. Protein-protein interaction analysis.

Protein design and Engineering: Basic concept, application of protein engineering, peptide synthesis, recombinant protein production in *E. coli* and Yeast, Peptide synthesis, PCR based mutagenesis for engineering of recombinant protein. Concept and application of RE in protein engineering.

Metabolomics and Drug Discovery: Definition and necessity of metabolomics, Metabolites and Metabolome. Biochemical pathways e.g. Glycolysis. Different types of pathways in ***E.coli.*** Separation and Detection techniques of Metabolomics. Drug Discovery Protocol. Concept of Systems Biology**.** Approaches to drug and vaccine design using bioinformatics tools, Molecular docking using Autodock and/or other computer aided programs.

Main Books:

Warren J. Ewens, Gregory R. Grant (2004): Statistical Methods in Bioinformatics: An Introduction (Statistics for Biology and Health). 2nd edition. Publisher: Springer.

R.M.Twyman (2010): Principles of Proteomics, BIOS Scientific Publishers, 2004.

Books Recommended:

Josip Lovric (2011): Introducing Proteomics; edition 1, Publisher Wiley- Blackwell.

David W. Mount. (2004): Bioinformatics: Sequence and Genome Analysis, Second Edition, Publisher: Cold Spring Harbor Laboratory Press.

Robert G., Vincent C, Wolfgang H., Rafael I., and Sandrine D.,(2005). Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Statistics for Biology and Health), Springer; 1st edition.

Robert Gentleman (2008). R programming for bioinformatics (Chapman & Hall/Crc Computer Science & Data Analysis), Chapman & Hall/CRC; 1st edition.

Andreas D. B., Francis O., (2004). Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. Wiley- Interscience; 3rd edition.

Richard S. Larson (2005). Bioinformatics and Drug Discovery (Methods in Molecular Biology) (Editor), Humana Press.

Jones, Oliver A H; (2014). Metabolomics and systems biology in human health and medicine. CAB International.Wallingford, Oxfordshire : CAB International.

Lämmerhofer, Michael; Weckwerth, Wolfram(2013). Metabolomics in practice: successful strategies to generate and analyze metabolic data. Wiley-VCH.

Fan, Teresa Whei-Mei; Higashi, Richard M, Lane, Andrew N. (2012). The handbook of metabolomics, New York : Humana Press : Springer.

Stat.614: Machine Learning and Deep Learning

Full Marks: 100 **Examination: 4 hours**

Machine Learning:

The Fundamentals of Machine Learning, Dataset Loading, Exploratory Data Analysis.

Feature Selection and Extraction (image, text), Standardization and Normalization.

Regression: Linear regression, Ridge regression, Kernel Ridge regression, Polynomial regression, Decision tree regression.

Classification: SVM, Logistic regression, Stochastic Gradient Descent, KNN, Naive Bayes Classification, Decision Trees.

Ensemble Methods: Random Forest Classifier, Voting Classifier, AdaBoost, Gradient Boosting, XGBoost.

Neural Networks.

Model selection and evaluation: Cross-Validation, Hyperparameter Tuning, Pipeline.

Deep Learning:

Introduction: Deep Learning Fundamentals.

Algorithms: Deep Learning and Neural Networks (Classification and Regression), Convolutional Neural Network (CNN), Transfer Learning (TL), CNN Architectures and Autoencoder, Recurrent Neural Networks (RNN) or Sequence Models, and Time-series Forecasting.

Natural Language Processing (NLP), Generative Adversarial Networks (GAN), Reinforcement Learning Basics, Audio Classification.

Main Books:

Dasgupta, N. (2018). *Practical Big Data Analytics: Hands-on techniques to implement enterprise analytics and machine learning using Hadoop*, Spark, NoSQL and R. Packt Publishing.

Favero, L.P., Belfiore, P., Souza, R.D.F. (2023). *Data Science, Analytics and Machine Learning with R*, Elsevier Inc.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.

Ghatak, A. (2019). Deep Learning with R. Springer.

Al-Malah, K.I.M. (2024). Machine and Deep Learning Using MATLAB: Algorithms and Tools for Scientists and Engineers, Wiley.

References:

Shalev-Shwartz, S. and Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press.

Bertrand Clarke · Ernest Fokou´e Hao Helen Zhang (2009). *Principles and Theory for Data Mining and Machine Learning*, Springer Science+Business Media, LLC, Dordrecht Heidelberg, Germany.

Hastie, T., Tibshirani, R. and Friedman, J. (2008). *The Elements of Statistical Learning: Data mining, Inference and Prediction*, Springer Series in Statistics.

Sumathi, S. and Sivanandam, S.N. (2006). *Introduction to Data Mining and its Applications*, Springer-Verlag, Berlin, Heidelberg.

Stat. 615: Research Methodology and Scientific Writing

Full Marks: 100 **Examination: 4 hours**

Introduction to Research: Meaning, objectives and motivations in research, Characteristics and limitations of research, Components of research work, Application of Research, Meaning and sources of research problem, Characteristics of good research problem, Research process, Outcomes, Meaning and types of research hypothesis, Importance of review of literature, Organizing the review of literature, Research protocol, Research methodology.

Types of Research: Types of research, Fundamental, Pure or Theoretical research, Applied research, Descriptive research, Evaluation research, Experimental research, Survey research, Scientific research, Non-scientific research, Action research, Research and development (R & D), Qualitative research, Quantitative research, Historical research.

Research Design: Meaning, need, types of research design, Exploratory, Descriptive, Casual research design, Components of research design, and Features of good research design. Experiments, surveys and case study research design.

Questionnaire Design and Construction: Types of questions, framing of questions, sequencing questions, construction of a model questionnaire, question wording, guidelines for avoiding poor question wording, data processing and analysis, Pre and post-tested questionnaire.

Reliability and Validity in Measurements: Measurement error, test for sound measurement, reliability and its measurements, validity and its types, measurements of validity, construction of measurements scales.

Sampling, Data Collection and Analysis: Contribution of statistics to complete a scientific research, Types and sources of data, Methods of collecting data, classification and tabulation of data, graphical representation of data, Concept of sampling, Probability and non-probability sampling methods, Sample size determination.

Statistical Tools for Data Analysis: Selection of Suitable Statistical tools/methods/models for data analysis.

Ethics in Research: Plagiarism: Definition, different forms, consequences, unintentional plagiarism, copyright infringement, collaborative work. Qualities of good researcher, Ethical clearance letter.

ICT Tools for Research: Role of computers in research, maintenance of data using software such as Mendeley, Endnote, Tabulation and graphical presentation of research data and software tools. Web search: Introduction to Internet, use of Internet and WWW, using search engines and advanced search tools.

Scientific Report: Title selection, Objectives, Research questions and Research Hypothesis, Research report and its structure, Journal articles, Components of Journal article, Explanation of various components, Structure of an abstract and keywords, Thesis and dissertations, Components of thesis and dissertations, Referencing styles and bibliography.

Main Books:

1. Kothari, C. R. (2009). R*esearch Methodology- Methods and Techniques*, New Age International Publishers, New Delhi.
2. Lohr, S. L. (2010). *Sampling: Design and Analysis*, Books Cengage Learning, Australia.
3. Pandey, P. and Pandey, M. M. (2015). *Research Methodology: Tools and Techniques*. Bridge Center, Romania.
4. Ted Palys (1997). *Research Decisions – Quantitative and Qualitative Perspectives*, International Thomson Publishing.