BEGUM ROKEYA UNIVERSITY, RANGPUR
Department of Statistics
Faculty of Science

Syllabus for
M.Sc. (Master of Science) program in Statistics

The M.Sc. (Master of Science) program in Statistics promote at specializing and training in statistical methodology in its theoretical, practical/Applied and in scientific research aspects of modern age particularly with computer intensiveness.

The M.Sc. (Master of Science) program in Statistics shall extend over one academic year which is divided into two semesters: **1st Semester** (July to December) and **2nd semester** (January to June).Students enrolled in the second semester under this program comprise of two groups namely General Group and Thesis Group .Thesis shall be offered subject to the approval of the departmental Academic Committee. The courses offered by the department are compulsory and optional in nature.

The Examination in 1st and 2nd semester shall be of total 1200 Marks. The class attendance carries 5%, assignments & presentation carry 10%, quizzes carry 10%, mid-term carries 25% and the semester final carries 50% of the total marks both in theoretical and practical courses. The general group students shall have to submit a report at the end of the 2nd semester carrying 50 marks, out of which 20 for oral presentation and 30 for report evaluation and the thesis group students shall have to submit a thesis/dissertation at the end of the 2nd semester carrying 150 marks, out of which 50 for Thesis defense and 100 for thesis evaluation. At the end of all course examinations all the students from two groups shall have to face an interview board for viva-voce in the second semester, which carries 50 marks.

The group-wise structure for the M.Sc. program in 2nd semester is given below:
General Group: The Examination shall consist of six courses in 2nd semester of 500 marks (15 Credits), Report of 50 Marks (1.5 Credits) out of which 20 for oral presentation and 30 for report evaluation and Viva Voce Examination of 50 Marks (1.5 Credits).

Thesis Group: The Examination shall consist of four courses in 2nd semester of 400 marks (12 Credits), thesis/dissertation of 150 Marks (4.5 Credits) out of which 50 for Thesis defense and 100 for thesis evaluation and Viva Voce Examination of 50 Marks (1.5 Credits).

The breakdown of marks, Units and Credits are as Follows:

Level		Course Code	Course Title	Marks	Credit Points
1st Semester	Compulsory Courses	STAT5101	Advanced Multivariate Techniques	100	3
		STAT 5102	Statistical Signal Processing	100	3
		STAT 5103	Statistical Genomics	100	3
	One from the (selected Optional Courses)	STAT 5104	Population Studies	100	3
		STAT 5105	Data Base Management System	100	3
		STAT 5106	Applied Bayesian Statistics	100	3
	First three are Compulsory Courses & **One from the** aforesaid related Optional(theoretical) Course	STAT 5107	Statistical Data Analysis-I (Lab)	50	1.5
		STAT 5108	Statistical Data Analysis-II (Lab)	50	1.5
		STAT 5109	Statistical Data Analysis-III (Lab)	50	1.5
		STAT 5110	Statistical Data Analysis-IV (Lab)	50	1.5
		STAT 5111	Statistical Data Analysis-V (Lab)	50	1.5
		STAT 5112	Statistical Data Analysis-VI (Lab)	50	1.5
2nd Semester	Compulsory Courses	STAT 5201	Advanced Statistical Inference	100	3
		STAT 5202	Advanced Demography and Survival Analysis	100	3
		STAT 5203	Advanced Bioinformatics	100	3
	One from the (Optional Courses)	STAT 5204	Computer Intensive Statistics	100	3
		STAT 5205	Stochastic Modeling	100	3
		STAT 5206	Design of Scientific Experiments and Generalized Linear Models	100	3
	Compulsory for General group	STAT 5207	Statistical Data Analysis-VII (Lab)	50	1.5
		STAT 5208	Statistical Data Analysis-VIII(Lab)	50	1.5
		STAT 5209	Thesis	150	4.5
		STAT 5210	Report	50	1.5
		STAT 5211	Viva-Voce	50	1.5

STAT 5101
Advanced Multivariate Techniques
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Advanced Multivariate Techniques

Principal Components: Population Principal Components, Summarizing Sample Variations by Principal, Graphing the principal Components, Large Sample Inference.

Factor Analysis: The Orthogonal factor models, methods of Estimation (maximum Likelihood Estimates and Principal factor Analysis), Selection of Loading and factor (Factor Rotation, Varimax Rotation, Quartimax Rotation, Oblimin Rotations), Structural Equations Models, Graphical representation of the results of factor analysis , Robust versus classical factor analysis.

Canonical Correlation Analysis: Canonical Vitiates and Correlation, Interpreting the population canonical variable, Sample Canonical Verities and Sample Canonical Correlations, Large Sample Inference.

Discrimination and Classification: Separation and Classification two populations, Classification of two Multivariate Normal populations, Evaluating Classification Functions, Fisher's Discriminate Function, Classification With Several populations, Fisher's Method for Discriminating Several populations.

Clustering: Similarity and Disimilarity measures, possible data problems in the context of cluster analysis, Types of Clustering, Hierarchical Clustering methods, Nonhierarchical Clustering Methods, Evaluation of ClusterValidity [2].

References:

1. Jonson, R.A. Wichern, D. W. (2002): *Applied Multivariate Statistical Analysis,* 5th edition, Pearson Education, Asia.
2. *StatisticalDataAnalysisExplained: Applied Environmental Statistics with R. C. Reimann, P. Filzmoser, R. G. Garrett,R. Dutter © 2008 John Wiley & Sons, Ltd. ISBN: 978-0-470-98581-6*
3. Srivastava, K.S. (2002):*Methods of Multivariate Statistics. Wiley.*
4. Anderson, T. W. (1984): *Introduction to Multivariate Analysis,* 2nd Edition, John Wiley, New York.
5. *Gan, G., Ma, C., and Wu, J.,(2007): Data Clustering: Theory, Algorithms, and Applications*

STAT 5102
Statistical Signal Processing
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Statistical Signal Processing

Signal and their Classification Real World Analog Signal: Audio, Video, Biomedical (EEG, ECG, PET, CT, US), SAR, Microarray, etc. Digital Representation of Analog Signal, Role of Transformation in Signal Processing. Signal Estimation Theory. Estimation of Signal parameters Using ML, EM algorithm, Minimum Variance Unbiased Estimators (Rao-Blackwell Theorem, CRLB,BLUE),Bayesian Estimators (MAP, MMSE, MAE), Liner Bayesian Estimators (Winner Filter, Kalman Filter). Signal Detection theory, Hypothesis, Detection Criteria (Bayes risk, Probability of error, Neyman-Pearson).LRT, Detection With Unknown Signal Parameters: UMP tests, Karlin-Rubin Theorem, GLRT, Bayes Factor. Application of Signal Estimation and Detection Theory to Signal Communication, Signal Recovery from Various Types Various Types of Linear and non-linear degradations, Feature Extraction, Compression, Pattern Recognition, Copyright Protection, Enhancement, Etc.

References:

1. Kay, S.M. (2013): *Fundamentals of Statistical Signal Processing: Estimation Theory. Prentice-Hall.*
2. Kay, S. M. (1998)*: Fundamentals of Statistical Signal Processing: Detection Theory. Prentice-Hall*.
3. Chonavel, T. (2002):*Statistical Signal Processing, Modeling and Estimation*, Springer.

STAT 5103
Statistical Genomics
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Statistical Genomics

Basic Concepts: Introduction, Genes and Chromosomes, Meiosis, Mandel's Laws, Linkage and Mapping, Quantitative genetics, Molecular Genetics, SNP.

Linkage Analysis and Map Construction: Introduction, Mendel an segregation, Segregation Patterns in a full-sib family , 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, Map functions. Linkage Analysis in Controlled Crosses, Linkage Analysis with Recombinant inbred Lines: Linkage Analysis for Distorted and Misclassified Markers.

Marker Analysis of phenotypes: Introduction, QTL regression model; Analysis at the marker, moving away from the marker, Power calculation, Marker interaction analysis, Whole-genome marker analysis.

The Structure of QTL mapping: Introduction, The mixture model, Population genetic Structure of the mixture model, Quantitative genetic structure of the mixture model, Experimental setting of the mixture model, Estimation in the mixture model, Computational Algorithms for the mixture model.

Interval Mapping: Introduction, Interval mapping for both backcross and intercross (F2) populations for QTL analysis.

Main Books:

1. *Benjamin M. Neale, Manuel A.R. Ferrera,Sarah E.(2008): Statistical Genetics: Gene Mapping Through Linkage and Association, Medland, Danielle posthuma. Publisher: Tayor and Francis; 1 edition.*
2. *Rongling Wu, Changxing Ma. (2007): statistical Genetics of Quantitative Traits: Linkage,Maps and QTL (statistics for Biology and Health), George Casella, 1st edition. Publisher: springer.*

References:

3. *Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, peter Walter (2007): Molecular Biology of the Cell: Rerference Edition, publisher: Garland Science; 5 edition*

4. *Dr.M. Kearsey (2004): Genetical Analysis of Quantitative Traits, 1st edition publisher: Garland Science.*
5. *Freddy Bugge Christiansen (2000): population Genetics of Multiple Loci (Wiley Series Mathematical and Computational Biology). 1s tedition .Publisher: Wiley.*
6. *M.C. Yang (2000): Introduction to Statistical Methods in Modern Genetics (Asian Mathematics Series) Publisher: CRC.*
7. *Nicola J. Camp, Angela Cox (2002): Quantitative Trait Loci: Methods and protocols (Methods in Molecular Biology), 1st edition, publisher: Humana press.*
8. *Robert Gentleman, Vincent Carey, Wolfgang Huber, and Rafael Irizarry, Sandrine Dudoit (2005): Bioinformatics and Computational Biology solutions Using R and Bioconductor (statistics for Biology and Health), 1st edition, publisher: Springer.*
9. *Robert Gentleman (2008): R programming for bioinformatics (Chapman and Hall/Crc computer science and Data Analysis) 1st edition, publisher: Chapman and Hall/CRC.*
10. *Robert Gentleman, Vincent Carey, Wolfgang Huber, Rafael Irizarry (Editor), Sandrine Dudoit (2005): Bioinformatics and Computational Biology solutions Using R and Bioconductor (statistics for Biology and Health), publisher: springer;1st edition.*
11. *Sorin Istrail (Editor), Michael Waterman and Andrew Clark (2004): Computational Methods for SNPs and Haplotype Inference. Publisher: Springer, 1 edition.*

STAT 5104
Population Studies
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Population Studies

Gender Based Domestic Violence (GBDV): Reasons of Domestic Violence, Types of Physical Violence, Socio- economic and Reproductive Health Implication of GBDV, Important Steps in Reducing Gender Based Domestic Violence.

Social Development Indicator: Human Capital Indicator, Population Policy, Social Capital Development/Social Capital, Gender and Development, Social Protection/Social Network.

Couple Year Protection: Definition, Sterilization (Conversion Factor, Achievement Index, Prevalence Index), Tubectomy (Conversion Factor, Achievement Index, Prevalence Index), Intrauterine Device (IUD) (Conversion Factor, Achievement Index, Prevalence Index), Vasectomy (Conversion Factor, Achievement Index, Prevalence Index).

Effectiveness of Contraceptive Use: Fecundability and Fecundity, Life Table Analysis of Contraceptive Failure, Construction of Single and Multiple Decrement Life Table.

Decomposition of Change in TFR between Two Time Periods: Bongaart's Model, Target setting by Bongaarts Model, Relationship between Target Fertility and Contraceptive use.

Millenium Development Goal (MDG): Rational of MDG, Motivation Behind MDG, Goal, Target, Indicator of MDG, Current Situation of Bangladesh Considering Different Indicators.

Population Aging: Elderly Situation, Aging Index, Support Ratio Index, Care Index, Elderly Situation in Bangladesh, Components (Elements) of Aging Policy in Bangladesh, Goals and Objectives of Aging Policy in Bangladesh.

Gender Preference: Family Size, Ideal Family Size, Sex Preference of Family Size, Factors Affecting Sex Preference in Bangladesh, Relationship between Actual Fertility and Ideal Fertility, Fertility of Spacers and Limiters and their Effect, Effect of under Five Mortality or Infant Mortality on Desired Family Size.

Advocacy: Advocacy, Components of Advocacy, Importance of Advocacy.

Monitoring and Evaluation: Monitoring and Evaluation, Steps in Monitoring and Evaluation, Indicators of Monitoring and Evaluation.

Management Information System (MIS): Purpose of Information, Purpose of Management Information System.

Urbanization: Magacity, Urban Projection, Estimating Megacity Population and Implications on Basic Needs, Social, Economic and Demographic Implications.

Gompertz Model: Assumption, Estimation of Process Advantages and Disadvantages, Derivation of Model Parameters.

Population Stabilization: Population Stabilization, Tempo Effect, Quantum Effect, Implication of Population Stabilization If Replacement Fertility is Not Achieved, Population Momentum, Reduction of Population Momentum, Factors to be Considered in Reduction Population Momentum.

Demographic Benefits: Achieved Replacement Fertility in Time, Its Benefit in Falling Fertility, Demographic Window/Bonus: Implication of Macro Economic Growth.

Health Policy of Bangladesh: Definition, Reasons of Health Policy, Health Policy of Bangladesh, Health Infrastructure Information, Selected Health and Family Planning Indicators, National Health Policy (NHP), Objectives of National Health Policy, Principles of National Health Policy.

Truncation Estimator of Age at First Marriage: Truncation or Censoring, Estimate Mean Age and Marriage of Truncated Distribution.

Disability Adjusted Life Years (DALY): Necessity of Measuring DALY, Concepts, Principle and Philosophy of DALY, Measurement of DALY, Construction of Life Table of Disability Prevalence, Problems in DALY.

Influence of Age Structure on Fertility: Estimation of Fertility When It is Affected by Age Structure.

References

1. *Chiang, C. L. (1984): The Life Table and Its Applications, Krueger Pule, John Wiley, New York.*
2. *Bongaarts, J. and, Potter, R. G. (1983): Potter Fertility, Biology and Behaviour: An Analysis of the Proximate Determinants of Fertility, Academic Press, Sandiego, California.*
3. *Colin, N. (1988): Methods and Models in Demography, Belhaven Press, London.*
4. *Selected articles from Population Studies, Demography, Population and Development Studies in Family Planning etc.*

STAT 5105
Data Base Management System
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Data Base Management System

Introduction to Database system: Overview, File system VS database system, Advantage of a DBMS, Describing and storing in a DBMS, Applications.

The Entity-Relationship model: Basic concept, Design issue, Mapping constraints, Keys, E-R diagram, Weak entity sets, Extended E-R features ,Design of an E-R database schema, Reduction of a E-R schema to tables.

Relational model: Structure of relational databases, The relational algebra, The tuple relational calculus, the domain relational calculus, relational algebra operations, modification of the database, introduction to views.

Structured Query Language: The form of a basic SQL query, UNION, INTERSECTION and EXCEPT, nested queries, aggregate operations, null Values, embedded SQL, cursors, dynamic SQL, ODBC and JDBC, triggers and active database.

Relational Database Design: Pitfalls in relational database design, Decomposition, normalization using functional Dependencies, normalization using multivalve Dependencies, Normalization using join dependencies, domain-key normal form.

Object Oriented and Object Relational Databases: The Object oriented data model, Nested relations, Complex types and object orientation, Querying with complex types, creation of complex values and objects.

Indexing and Hashing: Index definition, Benefits of using index, Various type Index, Index Structure, B-tree Structure of Index, Clustered Index, Non Clustered Index, Index design, Index in SQL, Hash file organization, Hash function, Hash Indices, Dynamic Hashing, Hashing queries.

Storage and file Structure: Overview physical Storage media, file organization, organization of records in files, Data-dictionary storage. **Query Processing:** Catalog Information for cost estimation, measuring of query Cost, different operations, Evaluation of Expressions, Choice of Evaluation Plans.

Transactions: Transaction Concept, Transaction atomicity, consistency, isolation, durability, Transaction state, Implementation of atomicity and durability, Concurrent Execution, Serializability, Recoverability, Recoverable Schedules, Implementation of isolation, Transaction Definition in SQL.

Concurrency Control: Lock based protocols, Timestamp based protocols, Lock based protocols, timestamp based protocols, validation based protocols, multiple granularity, multiversion schemes, deadlock handling, insert and delete operations.

Recovery system: Failure classification, Storage structure, recovery and atomicity, log-based recovery, shadow paging, recovery with concurrent transactions, buffer management, advanced recovery techniques.

Database system architecture: Centralized systems, client-server systems, parallel systems, distributed systems, network types.

An introduction to parallel and distributed database: Oracle: introduction to SQL plus, PL/SQL, triggers, forms, reports, query, procedures, and project builder. Case study: MS SQL server, My SQL server.

Reference:

1. *A. Silberschatz : Database System Concepts.*
2. *R. Ramakrishnan : Database Management System.*
3. *James Martin : Principles of Database Management.*
4. *Uillman : Database Management systems.*

STAT 5106
Applied Bayesian Statistics
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Applied Bayesian Statistics

Bayesian Thinking Background, Benefits and Implementation; Bayes Theorem, Components of Bayes Theorem-Likelihood, Prior and posterior; Proper and Improper Priors; Conjugate Prior, Method of finding Conjugate Prior, Informative and Non-Informative Priors, bayes Factor.

Bayesian Inference and Prediction: Single Parameter Models- Binomial, Poisson, Normal With Known Variance, Normal With Known Mean; Multi-Parameter Models-Concept of Nuisance Parameters, Normal Model With a Non Informative, Conjugate, and Semi-Conjugate Priors, Multinomial Model, Multivariate Normal Model, Inference Based on bayes factor.

Bayesian Computations: Summarizing Posterior Distributions-Approximation Methods, Monte Carlo Method- Direct Sampling and Acceptance- Rejection Sampling; Importance Sampling; Markov Chain Monte Carlo (MCMC) Methods-Gibbs Sampling, Metropolis- Hastings (MH) Sampling; relationship between Gibbs and MH Sampling Evaluating MCMC and Model fit – assessing Convergence, Acceptance rates of the MH Algorithm, Autocorrelation; Evaluating Model fit; EM Algorithm-EM Theory for Regular Exponential Family, Generalized EM Algorithm; Introduction to Win BUGS.

Regression Models; Normal Linear Regression-Model Development, Sampling Form the Posterior and Predictive Distribution using Gibbs and MH sampling Algorithms, Generalized Liner Models Model Development, Sampling from the Posterior Distribution and Predictive Distribution Gibbs and MH Sampling Algorithm; Hierarchical Liner Regression Models-Random Intercept Model and Random Coefficient Model, Fixed Versus Random Effects Models.

References:

1. *Gelman, A. Carlin J.B. and Stern H.S. (20004). Bayesian Data Analysis. Chapman and Hall.*
2. *Lynch, S.M. (2007). Introduction to Applied Bayesian Statistics and Estimation for Social Scientists Springer.*
3. *McLachlan, G. L. and Krishnan, T. (2008) The EM Algorithm and Extensions. Wiley.*

STAT 5107
Statistical Data Analysis-I (Lab)
Full Marks-50
No. of Lectures: Minimum-15.
Duration of each lecture: 1.5 hours
Duration of Examination: 2 Hours

Problems Related to STAT 5101.

STAT 5108
Statistical Data Analysis-II (Lab)
Full Marks-50
No. of Lectures: Minimum-15.
Duration of each lecture: 1.5 hours
Duration of Examination: 2 Hours

Problems Related to STAT 5102.

STAT 5109
Statistical Data Analysis-III (Lab)
Full Marks-50
No. of Lectures: Minimum-15.
Duration of each lecture: 1.5 hours
Duration of Examination: 2 Hours

Problems Related to STAT 5103.

1. Linkage Analysis and Map Construction for both backcross (B1 & B2) and inter-cross (F2) population. Analysis of recombination fraction.

2. QTL analysis by ANOVA, Regression, Interval mapping for both backcross (B1 & B2) and inter-cross (F2) population.

3. SNP data analysis for both backcross (B1 & B2) and inter-cross (F2) population

STAT 5110
Statistical Data Analysis-IV (Lab)
Full Marks-50
No. of Lectures: Minimum-15.
Duration of each lecture: 1.5 hours
Duration of Examination: 2 Hours

Problems Related to STAT 5104.

STAT 5111
Statistical Data Analysis-V (Lab)
Full Marks-50
No. of Lectures: Minimum-15.
Duration of each lecture: 1.5 hours
Duration of Examination: 2 Hours

Problems Related to STAT 5105.

STAT 5112
Statistical Data Analysis-VI (Lab)
Full Marks-50
No. of Lectures: Minimum-15.
Duration of each lecture: 1.5 hours
Duration of Examination: 2 Hours

Problems Related to STAT 5106.

STAT 5201
Advanced Statistical Inference
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Advanced Statistical Inference

Theory of Estimation

Points Estimation: Principles of Data Reduction, Empirical bayes (EB) Method, Bayes Conventional and Empirical Bayes Techniques, Approximation of Bayes and EB Method of Estimation for Prior Distribution, Application of EB Methods, Minimaxity and Admissibility in Exponential Families and their Properties, Bayesian Estimation in Linear Model, predictive Inference with Reference to Bayesian Analysis, James Stein Estimator, Shrinkage Estimators, EM Algorithm.

Robust Statistics: Fundamental Concept and Example, One-Dimensional Estimation and Test, Influence Functions, Classes of M-Estimators, L-Estimators, R-Estimators, Multidimensional Estimators, Application of Robust Estimators.

Method of Evaluating Estimators: U- Statistics and Its Properties, Best Unbiased Estimators and their Properties.

Confidence Sets: Length of Confidence Intervals, Shortest Length Confidence Intervals, UMA and UMAU Confidence Sets, Randomized Confidence Sets.

Bootstrap Confidence Sets: Contraction of Bootstrap Confidence Intervals, Asymptotic and Accurate Bootstrap Confidence Sets.

Simultaneous Confidence Intervals: Bonferronis Method, Schaffer's Method in Liner Models, Tukey's Method in One-Way ANOVA Models, Confidence Bands for Cumulative Distribution Functions.

Theory of Hypothesis Testing

Generalized Neyman Pearson Lemma, Uniformly Most Powerful (UMP) Test, Locally Most Powerful (LMP) Test, Locally Uniformly Most Powerful Unbiased (LUMPU) Test, Optimal Test, Locally Best Test, Tests Under Restricted Alternatives, Similar Region and Structure, Most Powerful Similar Region (MPSR) Test, Uniformly most Powerful Similar Region (UMPSR) Test, Asymptotic Efficiency of Test, Sequential Probability Ratio Test (SPRT) for Three Hypotheses, Sobel and Wald Test, Lagrange multiplier (LM) Test, Test in Presence if Nuisance Parameters, Union-Intersection and Intersection-Union Test, Armitage Method for Composite Hypotheses, Sequential T, X2 and T2 Test.

Empirical Bayes Tests Testing if multiple hypothesis, Lindley's Procedure for Test of Significance, Lindley's Paradox, p-value and Bayesian Significance Probability, Bays test in Liner Model.

Texts:

1. *Heorge, C. and Barger, R.L. (2003): Statistical Inference, 2nd edition, Thompson-Duxbury, USA.*
2. *Hogg, R.H. Mckean, J.W. and Craig, A.T. (2007): Introduction to Mathematical Statistics, 6th edition, Pearson Education (Singapore) Pte Ltd.*

References:

1. *Lehman, E. and Cassela, G. (1998): Theory of Point Estimation, Springer Verlag, New York.*
2. *Lehman, E. L. (1997): Testing Statistical Hypothesis, 2nd edition, Springer-Verlag, New York.*
3. *Bansal, A.K. (2007): Bayesian Parametric Inference, Norosa Publishing House India.*
4. *Rohatgi, U. K. and Saleh, A. K. Md. E. (2005) : An Introduction to the Probability and Statistics, 2nd edition, John Wiley and Sons Ins. New York.*
5. *O' Hagan, A and Forster, J. (2004): Advanced Theory of Statistics, Bayesian Inferences, Vol. 2B Arnold.*
6. *Carlin, B.P. and Louis, T.A. (2002): Bayes and Empirical Bayes Method for Data Analysis, 2nd edition, CRP Press.*
7. *Stuart, A. and Keith, J.O. (1986): Advanced Theory of Statistics, Vol. II, Charles Griffin and Company Ltd. London.*
8. *Kalbfleisch, J.G. (1985) : Probability and Statistical Inference, Vol. I and Vol. II. 2nd edition, Springer-Verlag, New York.*
9. *Wasan, M. T. (1970) : Parametric Estimation, McGraw-Hill Inc, New York.*
10. *Knight, K. (2000): Mathematical Statistics, Chapman and Hall, London.*
11. *Shao, J. (1999): Mathematical Statistics, Springer-Verlag, New York.*
12. *Mukhopadhyaya, P. (1996): Mathematical Statistics, New General Book Agency (P)Ltd, India.*

STAT 5202
Advanced Demography and Survival Analysis
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Advanced Demography and Survival Analysis

Group A: Advanced Demography

Life Time Analysis: General Idea, Ordinary life Table. Properties and Interrelationship. Probability Distributions of life Table Functions. Methods of Constriction of Double and Multiple tables and Increment Decrement Life tables. Joint life functions, multi-life functions. last survivor status. General multi-life status. Application-construction of school life table, Working life table. Life table analysis of birth intervals, breastfeeding.

Demographic Estimation: Concept and Applicability of the indirect techniques involve in the estimation of infant, child, adult and maternal mortality. Estimates of fertility. Estimation of migration. Dual record system. Chandra-Sekar and Deming Method. Coal's indices, Coal's nuptiality model.

Stable Population Theory and Models: Concept of Stationary, Stable and Quasistable population. Annual growth rate and intrinsic growth rate, Lotka's integral Equation. Met maternity function. Graduation of normal, Wicksell and Hadwiger. Effects of change of birth and death rates on stable population. Study of some growth models-Exponential, Malthusian, Logistic and Quasi-Stable models.

Population Projection: Development of Leslie projection matrix. Properties of Leslie matrix. Forward and backward operation of population projection. Stable vector, dominant root. Freak's method for population projection. projection of fertility and mortality.

Micro Demography: Reproductively, fecundity, fecund ability and sterility. Effective fecund ability. Residual fecund ability. Estimation of fecund ability. pearl index. Effectiveness and efficiency of FP methods. Davis-Blake framework of intermediate variables. proximate determinants of fertility.

Population Trends and Population planning: Demographic transition, world population situation, recent trends and future prospects. Population planning, planning for education, health, housing, population control and family planning program. Frame work for population and development research and planning.

Group B : Survival Analysis

Overview of Biostatistics: lifetime data. lifetime distributions and their characteristics. basic concept of censoring and methods. Survival curve estimates, Kaplan Meier estimator. nelson and –Harrington estimators. Variance estimation. Mean and survival. Comparison of survival curves.

Proportional Hazards Models: Introduction, Hypothesis test. Stratification, Residuals, Penalized Cox models. Partial likelihood. Applications of proportional hazards model.

Parametric Regression models: Introduction, Inclusion of strata. Specifying a Distributions, Residual analysis and other model checks. Predicted values. Fitting the model. Exponential, Weibull, Normal, Log-normal and Gamma regression models. Accelerated life models.

Goodness of fit test: Some general methods of testing fit. Chi-square tests, kolmogorov-Smirnov test, Anderson-Darling test. Tests of fit for specific models. Model selection critetia.

Main Books:

1. *Keyfitz, N. (1977). Applied Mathematical Demography, Wiley & sons.*
2. *Lawless, J.F. (1982). Statistical models and methods for Lifetime Data, Wiley, N.Y.*
3. *J.H. Pollard (1980). Mathematical models for the growth of human popualtions.*

References:

1. *Bain, L.J. and Engelhardt, M. (1991). Statistical Analysis of Reliability and Life Testing Models, Theory and Methods, 2nd Edition. Marcel Dekker, New York.*
2. *Balakrishnan, N.(ED.) (1995). Recent Advances in Life- Testing and Reliability, CRC press, Boca Raton,FL.*
3. *Biswas, S. (1988). Stochastic Processes in Demography and Applications. Wiley Estern ltd. India.*
4. *Islam, N. (1996). Levels and Correlates of Marriage and Fertility in Bangladesh. Unpublished Ph.D. deseration, R.U.*
5. *Jonson RCE & Jonson (1980). Survival Models and Data Analysis, Wiley NL & Sons, NY.*
6. *Kalbflesch, J.D. and prentice, R.L. (1980). the Statistical Analysis of Failure Time Data, Wiley, new York.*
7. *Nelson, W. (1990). Accelerated Testing. Statistical Models, Test Plans, and Analysis, Wiley, New York.*
8. *Shryock, H.J.S. Siegel and Associates (1976). The method and materials of Demography, Cond .ed. New York, Academic Press.*
9. *UNFPA (1993). population Research Methodology Vols. 1-8. Chicago, IIlinois, and other UNFPA publications.*
10. *UNO (1983). Indirect Technique Demographic Estimation, Population Studies No. 81.*

STAT 5203
Advanced Bioinformatics
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Advanced Bioinformatics

Introduction to Molecular Biology of the Cell: DNA, RNA, Chromosome, Gene and Central dogma.

Analysis of DNA Sequence: Shotgun sequencing, Modeling DNA , Modeling signals in DNA, Long repeats, r-scene, Analysis of DNA patterns, Overlaps counted, Overlaps not counted and motifs, Genomic sequencing, Sequence accuracy, Sequence formats, Conversions of one sequence format to another. The analysis of multiple DNA or protein sequences.

Statistical phylogenetic: Motivation and background on phylogenetic, Distance and clustering approach, Likelihood methods, parsimony, RNA-based phylogenetic methods, phylogenetic Tree Estimation.

Protein Classification and Structure prediction: Introduction to protein structure prediction, Review of protein structure and terminology, protein classification. Methods of protein classification: Viewing protein structures, protein structure classification databases, Alignment of protein structures, structural prediction, structural modeling.

Statistical Methods for Gene Expression Data Analysis: Introduction to microarrays and microarray data, image analysis, preprocessing: Transformation and normalization. Identification of differential expressed genes in two or more groups using statistical test. The family-wise error rate (FWER) and the false discovery rate (FDR) in the algorithms.Geneclustering and classification. Inferring genetic regulatory networks from microarray experiment with Bayesian networks. Modeling regulatory networks using gene expression profile.

Main Books:

1. *David W. Mount. (2004): Bioinformatics: Sequence and Genome Analysis, second Edition, publisher: Cold spring Harbor Laboratory press.*
2. *Husmeier, D., Dybowski, R., Roberts, S. (2005): probabilistic Modiling in Bioinformatics and Medical Informatics, 2nd edition, publisher: springer.*
3. *Warren J. Ewens, Gresgory R. Grant (2004): statistical Methods in Bioinformatics: An Introduction (statistics for Biology and Health), 2nd edition. Publisher: springer.*

References:

4. *Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, peter Walter (2007): Molecular Biology of the Cell: Rerference Edition, publisher: Garland Science; 5 edition*
5. *D. Baxevanis, B.F. Francis Ouellette (2004): Bioinformatics: A Practical Guide to the Analysis of Genes and proteins Andreas, publisher: Wiley-Interscience; 3 edition*
6. *Michael R. Barnes (2007): Bioinformatics for Geneticists: A Bioinformatics primer for the analysis of genetic data, publisher: Wiley; 2nd edition.*
7. *Robert Gentleman, Vincent Carey, Wolfgang Huber, and Rafael Irizarry, Sandrine Dudoit (2005): Bioinformatics and Computational Biology solutions Using R and Bioconductor (statistics for Biology and Health), 1st edition, publisher: Springer.*
8. *Robert Gentleman (2008): R programming for bioinformatics (Chapman and Hall/Crc computer science and Data Analysis) 1st edition, publisher: Chapman and Hall/CRC.*
9. *Robert Gentleman, Vincent Carey, Wolfgang Huber, Rafael Irizarry (Editor), Sandrine Dudoit (2005): Bioinformatics and Computational Biology solutions Using R and Bioconductor (statistics for Biology and Health), publisher: springer;1st edition.*
10. *Sunil K. Mathur (2010): statistical Bioinformatics: with R, publisher: Academic press.*

STAT 5204
Computer Intensive Statistics
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Computer Intensive Statistics

Jackknifing : Bias Correction , Pseudo-Value, Approximate Confidence Intervals, Extension to –or-More-Sample Problems.

Bootstrapping: Bootstrap Strategy, Sampling Distributions, Empirical Distributions, Bootstrap Distributions, Percentile Bootstrap: Definition and use of Confidence Limits, Relation to Jackknife, Application to Hypothesis Testing, Number of Simulation Samples Required, Variants: Balanced Re-sampling.

Simulation Testing: Randomization Test, Approximate Randomization Test, Monte-Carlo Tests, Examples, Unbiasedness, Power, Number of Simulated Samples Needed.

Density Estimation: Definition, Examples, Bias, MSE and IMSE, Choice of Kernel and Smoothing Parameter, Computation via Fast Fourier Transform, Kernel Estimates for Non-

Negative and Circular Data, Variable and Adaptive kernel Estimators, Kernel Estimations in Nonparametric Regression.

References:

1. *Eforn, B. and Tibshirani, r.j. (1993): An Introduction to Bootstrap, Chapman and Hall, new York.*
2. *Eforn, B. (1984): The Jackknife, The Bootstrap and other Re-sampling Plans, Society for industrial mathematics.*
3. *Noreen, E. W. (1982): Computer- Intensive Methods for Testing Hypothesis, Wiley, New York.*
4. *Shao, J. and Tu, D. (2000): Jackknife and Bootstraps, Springer-Verlag, New York.*
5. *Silverman, B.W. (1986) : Density Estimation for Statistics and Data Analysis, Chapman and hall, London.*

STAT 5205
Stochastic Modeling
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Stochastic Modeling

Reviewing the concepts of stochastic process: Finite Markov process, higher order Markovchains, tests of stationarity and order of Markov chain, lumpability and reversed Markov chain.

Stochastic models of population growth and other related processes: A simple illness deathprocess; illness transition probability and death transition probability; expected duration of stayin illness and death states; multiple exit transition probabilities; multiple transition probabilitiesleading to death; Kolmogorov differential equations and finite Markov processes; A general model for illness-death process;

Renewal Processes: renewal process when time is discrete; renewal process when time iscontinuous; renewal function and renewal density; backward and forward recurrence times; thestationary process; altering renewal process; cyclically altering renewal process; Markov renewalprocess;

Queuing Processes: a generalized queuing model, the queue M/M/1, the queue M/M/s,the queue M/M/1 with balking, the M/M/1 queue with state-dependent service, the queueM/M/1 with additional server for longer queues, the loss system M/M/s; The queue M/G/1and G/M/1; the machine interference *problem; Queuing Networks: the tandem queue, openand closed queuing networks;*

Text Books:

1. *Chiang, C. L. (1968). Introductory Stochastic Processes to Biostatistics. Wiley.*
2. *Bhat, U. N. (1972). Elements of Applied Stochastic Process. Wiley.*

.

References:

1. *Anderson, T.W. (1971): The statistical Analysis of Time series, Wiley, N.Y*
2. *Cryer, J. D. and K. Chan (2008): Time series Analysis: with applications in R, 2nd Ed., spinger, N.Y.*
3. *Elandt Johnson, R.C.(1971): Porbability Models and statistical Methods in Genetics, John Wiley, N.Y*
4. *Findley, D.F.(1981): Applied Time series Vol. I & II,Academic press, N. Y.*
5. *Fuller, W.A(1976): Introduction to Statistical Time series, Willey N.Y*
6. *Hamilton, J.D. (1994):Time series Analysis, Princeton University Press, N.J.*
7. *Lutkepohl, H. (2005): New Introduction to multiple Time series Analysis, Springer, N.Y.*
8. *Reinsel, G.C. (2003): Elements of Multivariate Time series Analysis, Springer, N.Y.*
9. *Tan Wai-Yuan(1991): Stochastic Process of Carcionogenesis,Marcel Dkker, N.Y.*
10. *Tsay, R.S. (2010): Analysis of Financial Time series, Wily & Sons, N.Y.*

STAT 5206
Design of Scientific Experiments and Generalized Linear Models
Full Marks-100
Number of Lectures: Minimum 30
Duration of each lecture: 1.5 hours
Duration of Examination: 3 hours

Design of Scientific Experiments and Generalized Linear Models

Group A : Experimental Design

Introduction: Review of two way classification with equal and unequal number of observation per cell. Concept of non-orthogonal design. Balanced incomplete Block design, Youden design and related design and Split-plot design.

Fractional Factorials and Main Effect Plans: Review of experimental design. Method of constructions of plans with factor at 2 levels. Orthogonal arrays of strength 3 with factors at 2 levels. Orthogonal main effect plans factor at 3 and other levels. Orthogonal main effect plans of size 2 * Sr . Analysis of orthogonal plans. Analysis of covariance: Analysis of covariance of non-orthogonal data in two way classification. Analysis of covariance with two ancillary variables. Covariance and analysis of experiments with missing observation transformation.

Weighing Design: Complete black design as weighing designs. Two pan weighing designs from BIB design. Two associate PBIB designs as one pan weighing design. Weighing designs from truncated BIB design. Efficiency.

Lattice Designs: balanced lattices. Partially balance lattices. Rectangular lattices. Cubic lattice squares-description. statistical analysis with different replications. multivariate analysis of variance: Introduction. Omnibus MANOVA tests. Analysis and interpreting MANOVAS. Causal models underlying MANOVA analysis. Complex designs.

Group B : Generalized Liner models

Introduction: the origins of generalized linear models, processes in model fitting. The P-components of a GLM. Measuring the goodness of fit. Residuals. Continuous data with constant variance. Error structure. Liner predictor. model formula for liner predictors. Estimation, Algorithms for least for squares. Selection of covariates.

Models for binary and polytomous data: Introduction, Models for binary responses. likelihood function for binary data. The multinomial distribution. Likelihood functions. OKVRR- dispersion for both binary polytomous data.

Log-Linear Models: Likelihood function. Multinomial response Models. Multiple responses. The variance function. The deviance. The canonical link. Multiplicative models. Log link. Liner models.

Identity link. Estimation of the dispersion parameters. Techniques in model checking. Score tests for extra parameters. Smoothing. Checks for systematic and isolated departures from the model.

Main Books:

1. *Choran and Cox (2000) Experimental Design, 2nd edition. Jone Wiley, N.Y.*
2. *Montegomery, D.C. (2005). Design and Analysis of Experiment. John Wiley, N.Y.*
3. *Christensen, R. (1997) b. Log-Linear Models and Logistic Regression, 2nd edition. Springer-Verlag, N.Y.*
4. *Dobson, A.J. (2001). An Introduction to Generalized Linear Models, Chapman & Hall, N.Y.*

References:

1. *Fisher, R.A. (1995). Design of Experiment, 8th edition. Hafner, N.Y.*
2. *Hosmer, D.W. and Lemeshow, S. (2000). Applied Logistic Regression, Wiley,N.Y.*
3. *John and Quenouille (1977) Experiments: Design and Analysis, 2nd Edition. Charles Griffin, London.*
4. *Kleibum, D.G. (1993). Logistic Regression, Springer-Verlag. N.Y.*
5. *McCullagh, P.and Nelder, J.A. (1989). Generalized Liner models, Chapman & Hall, London.*
6. *McCullagh, E. and Searle, S. (2001) Generalized Liner and Mixed Models, Wiley,N.Y.*
7. *Montgomery, D.C. Peck, E.A. and Vining, G.G. (2003). Introduction to linear Regression Analysis, 3rd edition. Wiley, N.Y.*
8. *Rosseeuw, P.J. and Leroy, A. (1987). Robust Regression and Outlier Detection, Wiley, N.Y.*
9. *Ryan, T.P. (1997). Modern Regression Methods, Wiley,N.Y.*
10. *Seber, G.A.F. and Wild, C.J. (1989). Nonlinear Regression, Wiley,N.Y.*

STAT 5207
Statistical Data Analysis-VII (Lab)
Full Marks-50
No. of Lectures: Minimum-15.
Duration of each lecture: 1.5 hours
Duration of Examination: 2 Hours

Problems Related to courses STAT 5201 & STAT 5202.

Group-A : Estimation

1. Box plots and its various in interpretations and outlier detection.
2. Cdf estimation and quantile estimation.
3. Density estimation.
4. Robust estimation of univariate location and scale parameter.

5. Huber`s M-estimation of location and parameter.
6. Variance estimation by Jackknife, bootstrap and influence function.
7. Robust estimation of multivariate location and scatter matrix.
8. Introduction to S-plus/R.
9. Data manipulation and various graphs.
10. Estimation of SE of r and b1 by non-parametric and parametric bootstrap and jackknifing.
11. Different bootstrap CI of r and b1.

Group- B: Hypothesis Testing

1. MPT
2. UMPT
3. SRT
4. MPSRT
5. UMPSRT
6. Asymptotic efficiency of the above tests.
7. Sequential test
8. OC and ASN function
9. Construction of decoction regions.
10. Test of 2×2 contingency table.
11. ARE of Mann-Whitney test and sign test.
12. Kruskal-Wallis test.
13. Square rank test for variances.
14. Friedman test
15. Permutation test and bootstrap test of equality of two means.
16. Fitting regression models by bootstrapping.
17. Different density estimations of univariate data.

Group- A: Advanced Demography

1. Construction of multiple decrement life tables
2. Estimation of fertility and mortality by indirect techniques
3. Estimation of completeness of registration and survey data by Chandra Secker method
4. Estimation of nuptiality parameters of nuptiality model
5. Estimation of Cole's indices
6. Graduation of net maternity function by Normal, Wicksell and Hadwihger Model.
7. Problems related to population projection
8. Problems related to fecund ability.
9. Estimation of proximate determinants of fertility and inhibiting effects of marriage and divorce.

Group- B: Survival Analysis

1. Non-parametric estimation of survival and hazard functions with standard errors and confidence intervals.
2. Fitting of parametric survival distributions under different types of censored data.
3. Comparison of two and/or more than two survival curves.
4. Check of proportional hazard assumptions. Cox proportional hazard model analysis.
5. Fitting of parametric regression models and tests of fit.

STAT 5208
Statistical Data Analysis-VIII (Lab)
Full Marks-50
No. of Lectures: Minimum-15.
Duration of each lecture: 1.5 hours
Duration of Examination: 2 Hours

Problems Related to STAT 5203 & one from the selected theoretical courses.

1. DNA sequence data Analysis by statistical phylogenetics approaches.
2. Detection of differentially expressed (DE) genes by classical, Bayesian and non-parametric approaches.
3. Detection of differentially co-expressed (DCE) genes by classical, Bayesian and non-parametric approaches.
4. Gene and individual classification by supervised and unsupervised statistical algorithms.
5. Gene-set enrichment analysis.

And

Problems Related to course STAT5204

OR

Problems Related to course STAT5205

1. Study of simple epidemic models.
2. Analysis of Kermack and McKendreck's epidemic model.
3. Study of Daley and Kendall's epidemic model.
4. Study of Optimum allocation model in clinical drug trials.
5. Study of Robbins model.
6. Study of Hardy Weinberg law of heredity.
7. Analysis of M/M/1 Queuing models.
8. Analysis of multi cannel Queuing models.
9. Study of Network models.
10. Study of models of Waiting times of conceptions-Sheps and perrion model of reproductive process.

OR

Problems Related to course STAT5206

Group-A: Experimental Design.

1. Analysis of data using two way and three way classification with equal and unequal number of observations per cell.
2. Analysis of data using non-orthogonal model.
3. Arrange the data in BIBD, PBIBD and You den design. Also analyze the data using such design.
4. Estimate and test the data using intra-block and inter-block BIBD model.
5. Analyze the data using Split-plot and design.
6. Analyze factorial experiment of levels 2 and 3 with different factors. Construct $\frac{1}{2}$ and $\frac{1}{4}$ replicate of such design and analyze the data using RBD and LSD model. Also study confounding. Defining contrast and aliases.
7. Covariance analysis of one way and two way classified data with two concomitant variables.
8. Analyze lattice design/Balanced lattice design/Partially balanced lattice design with different replications.
9. Construct the methods of estimation and analyzing procedure of one pan and two pan weighing design.
10. Develop the method of estimating variance using one/two pan weighing design from BIBD. Check properties.
11. Develop the estimating and analyzing procedure using MANOVA. Develop test procedure Omnibus MANOVA tests and other important methods. Interpretation the results of MANOVA.

Group-B: Generalized Linear Models

1. Checking the fitness of different probability distributions (Binomial, Poission, Normal, Gamma,……etc.) to the exponential family.
2. Computation of deferent residuals , deviance of estimate and measuring of goodness of fit.
3. Estimation of parameters and computation of Generalized variance by different methods of estimation.
4. Maximum likelihood estimation of parameters by iterative weighted least squares procedures.
5. Retrospective Sampling of binary data.