UNIVERSITY OF RAJSHAHI
Faculty of Science
Department of Statistics

Department of Statistics: At a glance

The Department of Statistics is one of the oldest Departments in the University of Rajshahi, which is the second largest public university of Bangladesh. Since its inception in 1961, the department has evolved steadily. The department offers good teaching and research facilities. At present there are 27 full-time teachers and 21 office staffs. The present enrolments of the students are about 100 and 400 at the graduate and undergraduate programs respectively. The department has its own seminar library with a large number of Textbooks, Journals, and Scientific Literature. A Computer Unit was established in the department in 1990. It is equipped with more than 180 modern personal computers of various models. Both the teachers and students of this department have been using these computers. Occasionally teachers of other departments of the University of Rajshahi are also allowed computer facilities for their research works. Presently the department is offering four courses on computer at the undergraduate and graduate levels. It also offers a certificate course on computer for researchers, professionals and students.

Apart from offering B.Sc. Honours and M.Sc. degrees in Statistics, the department also conducts research for higher studies leading to M. Phil. and Ph. D. degrees. The research works have been concentrated mainly in the areas of Anthropometric Statistics, Bayesian Inference, Biostatistics, Bioinformatics, Biochemical Engineering and Statistical Signal Processing, Bootstrap Techniques, Computer Programming & Simulation, Demography, Earthquake Prediction Modeling, Econometrics, Environmental Statistics, Geo-Statistics, Human Growth, Human Morphology, Industrial Statistics, Linear programming, Order Statistics, Probability, Regression Diagnostics, Reliability Theory, Robust Statistics, Sample re-use Techniques, Social & Educational Statistics, Statistical Data Mining, Statistical Inference, Stochastic Modeling etc.

The department organizes Seminar/Conferences regularly. In the last two decades five international conferences were organized. Occasionally distinguished foreign scholars visit the department and give seminar talks. The department publishes an annual journal named International Journal of Statistical Sciences (IJSS). This is a referred journal and abstracted in the Statistical Theory and Method abstract published by the International Statistical Institute, Netherlands and BANSDOC, Dhaka. For over last 20 years this department has a collaboration program with Indian Statistical Institute, Kolkata. In recent years, the department has launched link programs with Kyoto Gakuen University, Japan and The University of Electro-Communications, Tokyo, Japan.

The department has an association of its own, named “Parishankan Samity”. The aims of this “Samity” are to uphold the academic interests of the department, to organize indoor games, reception party for the new students, farewell party for the outgoing students and bring out a magazine annually. Recently, the alumni of this department have formed Rajshahi University Statistics Alumni (RUSA) in Dhaka to

incorporate all ex-statistics students of Rajshahi University from home and abroad to promote academic, research and all other activities of Department of Statistics.

Name and Research Interest of the Academic Staff

SL	Name	Specialization/Field of Interest
	Professor	
1	Dr. S. K. Bhattacharjee	Bayesian Inference
2	Dr. Md. Asaduzzaman Shah	Stochastic Modeling, Probability
3	Dr. Md. Nurul Islam	Demography, Survey Sampling, Applied Statistics, Anthropometric Study, Health Statistics
4	Dr. M. Sayedur Rahman	Environmental Studies, Statistical Data Mining, Bootstrap, Simulation Modeling
5	Dr. Anjuman Ara Begum	Demography, Statistical Inference, Order Statistics, Computer Programming
6	Dr. Md. Ripter Hossain	Demography, Sample Survey, Econometrics, Industrial Management and Health Statistics
7	Dr. Md. Ayub Ali	Time Series Analysis and Forecasting, Multivariate Analysis, Anthropometric Study, Physical Health and Human Growth, Statistical Modeling, Econometrics, Environmental Statistics, Descriptive Statistics and Applied Statistics
8	Dr. Md. Golam Hossain	Health Statistics and Physical Anthropology
9	Dr. Md. Rezaul Karim	Reliability, Bio-Statistics, Computer Programming, Data Science
10	Dr. Md. Monsur Rahman	Reliability, Bio-Statistics, Statistical Inference
11	Dr. Dulal Chandra Roy	Sample Survey, Estimation, Operation Research, Quality Control, Applied Statistics
12	Dr. Md. Nurul Haque Mollah	Robust Statistical Inference, Multivariate Statistics, Optimization, Data Mining, Statistical Signal Processing, Biostatistics, Statistical Genomics and Bioinformatics
13	Dr. Saroje Kumar Sarkar	Multivariate Analysis, Reliability
14	Dr. Md. Aminul Hoque	Demography, Bio-informatics, Forecasting, Biochemical Engineering
15	Dr. Md. Mahmudul Alam (Lien)	Distribution Fitting, Geo-Statistics, Environmental Statistics, Earthquake Prediction Modeling, Computer Programming & Simulation, Data Mining
16	Dr. Provash Kumar Karmokar	Econometrics, Time Series Analysis and Forecasting, Statistical Modeling of Climatic Variables, Computational Statistics, Simulation and Modeling, Agriculture Statistics
17	Dr. Md. Jahanur Rahman	Econometrics, Applied Macroeconomics, Financial Statistics and Climate Change

SL	Name	Specialization/Field of Interest
18	Dr. Papia Sultana	Semiparametric and Nonparametric Modelling, Medical Diagnostics, Epidemiology
19	Md. Mesbahul Alam, Ph. D.	Reliability Data Analysis, Warranty, Bio-statistics Optimization, Mathematical Statistics, Robust Statistics, Regression Diagnostics, Computer Programming and Data Mining
Associate Professor		
20	Md. Monimul Huq	Time Series Analysis and Forecasting, Econometrics, Robust Regression and Diagnostics, Multivariate Analysis
21	Dr. Md. Sabiruzzaman	Time Series Analysis, Econometrics, Robust Statistics, Estimation
22	Dr. Abu Sayed Md. Al Mamun	Regression Diagnostics and Statistical Modeling
23	Md. Abdul Khalek	Groundwater Modeling, Environmental Statistics, Computer Programming & Simulation
24	Dr. Md. Kamruzzaman	Demography and Health Statistics
25	Farhana Hasan	Demography, Mathematical Economics, Probability Theory, Reliability
Assistant Professor		
26	Md. Hadiul Kabir	Biomedical Statistics and Time Series Analysis
27	Dr. Md. Mostafizur Rahman	Financial Time Series, Data Mining and Econometrics

Vision of the Department:

- To establish the Department of Statistics, University of Rajshahi as a center of excellence for quality education and research;
- To produce skilled and competent human resources to serve the needs of national and international communities;
- To be a leading statistical institute by addressing “better data, better life”.

Mission of the Department:

- Providing high quality education through adoption of comprehensive programs with a regular updating curriculum for the requirements of contemporary job markets;
- Improving teaching learning environment and infrastructure of the entity to produce competent work force with high level of professionalisms equipped with modern technology and services of dedicated faculties under strategic leadership to contribute to the socio-economic development of the country and make a prosperous nation in the globalized environment.
- Organizing training, research collaboration, workshop, seminar, conference for improving teaching quality and research;

- Promoting scholarly research in statistical theory and application through publications in leading professional journals for the theoretical development of the multidisciplinary field;
- Contributing excellence in research and education in the statistical science through services to the university communities, statistical professions, and societies at large;
- Producing skilled graduates to lead quality research, training and innovation in the field of statistics and multidisciplinary areas to meet the needs of a dynamic world.

Objectives of this Department:

- Provide professional graduates in the field of statistics and the multidisciplinary areas according to the requirements of contemporary job markets.
- Publish in leading professional journals to contribute the theoretical development and application of statistics addressing substantive problems through scholarly research;
- Disseminate statistical knowledge to ensure effective applications of statistics in real life practices.
- Deliver adequate, relevant and timely statistics to facilitate research, planning and decision making process for the government and the community for achieving sustainable development goals (SDG) of Bangladesh.

Curriculum

M.Sc. in Statistics
Session: 2017-2018
Examination: 2018

Ordinance

The M.Sc. (Master of Science) Final Course in Statistics shall spread over One Academic year. The Course is divided into Two Groups - The General Group and the Thesis Group. The Examinees shall take Examination either in the General Group or in the Thesis Group. The Thesis shall be offered subject to the approval of the Departmental Academic Committee. The examination (General or Thesis Group) shall be of 900 Marks (9 Course Units: 36 Credits).

M.Sc. Degree: A candidate should be awarded M.Sc. degree in Statistics, if he obtains 32 credit points out of 36.

Course Improvement: A candidate obtaining a GPA of less than 3.00 shall be allowed to improve the result by reappearing in the exam up to 12 credits with F grade within consecutive next three years. Also, a promoted student earning of a GPA less than 3.00 individual course(s) shall be allowed to improve the grade(s).

General Group: The Examination shall consist of Eight Theory Courses of 600 marks (6 Units: 24 Credits), Eight Practical Sessions of 150 Marks (1.5 units, 6 Credits; Spread over Eight days, 30% of the total Practical marks shall be allotted for continuous laboratory assessment), Viva-voce Examination of 100 marks (1.0 unit, 4 Credits), research project 50 marks (0.5 unit, 2 credit). Each course contains 60 of the theory marks, 15 in-course marks (Tutorial/Terminal 11.25, and Attendance 3.75).

Thesis Group: The Examination shall consist of Eight Theory Courses of 600 Marks (6 units: 24 Credits) and Viva Voce Examination of 100 Marks (1.0 unit: 4 Credits), a Thesis/ Dissertation carrying 100 marks (1.0 units: 4 Credits) and Thesis Defense (Seminar and Viva voce on Thesis) of 50 marks (0.5 unit: 2 Credits) and in-plant training of 50 marks (0.5 unit: 2 Credits). The breakdown of marks, units and credits are as follows:

General Group				
Category	Theory	Viva	Practical	Total
Course / Session	8 Courses	-	8 Sessions	-
Total Marks	600	100	200	900
Units	6.0	1.0	2.0	9
Credits	24	4	8	36

Thesis Group						
	Theory	Viva	Thesis	Thesis Defense	In plant Training	Total
Courses / Sessions	8 Courses	-	-	-	-	-
Total Marks	600	100	100	50	50	900
Units	6.0	1.0	1.0	0.5	0.5	9
Credits	24	4	4	2	2	36

The Paper-wise Title of Courses (Both for General and Thesis Groups), Marks, Course Unit, Credits and Duration of Examination are as follows:

Courses M-Stat	**Title of Paper**	**Exam.**	**In Course**		**Full Marks**	**Unit**	**Credit**	**Exam. Hours**
			Tutorial	**CA**				
Compulsory								
501	Advanced Statistical Inference	60	11.25	3.75	75	0.75	3	4
502	Advanced Experimental Design	60	11.25	3.75	75	0.75	3	4
503	Advanced Multivariate Analysis	60	11.25	3.75	75	0.75	3	4
504	Time Series Analysis and Forecasting	60	11.25	3.75	75	0.75	3	4
505	Data Mining	60	11.25	3.75	75	0.75	3	4
	Total				**375**	**3.75**	**15**	**20**
Optional: ***Any one course from each group should be chosen subject to the Approval of the Academic Committee***								
Group 1								
506	Genomics and Bioinformatics	60	11.25	3.75	75	0.75	3	4
507	Advanced Demography	60	11.25	3.75	75	0.75	3	4
508	Environmental Statistics	60	11.25	3.75	75	0.75	3	4
509	Advanced Actuarial Statistics	60	11.25	3.75	75	0.75	3	4
Group 2								

Courses M-Stat	Title of Paper	Exam.	In Course		Full Marks	Unit	Credit	Exam. Hours
			Tutorial	CA				
510	Proteomics and Biomedical Informatics	60	11.25	3.75	75	0.75	3	4
511	Health and Epidemiology	60	11.25	3.75	75	0.75	3	4
512	Statistical Methods in Industrial Management	60	11.25	3.75	75	0.75	3	4
513	Statistical Methods for Reliability Data	60	11.25	3.75	75	0.75	3	4
	Group 3							
514	Physical Health and Human Growth Modeling	60	11.25	3.75	75	0.75	3	4
515	Advanced Stochastic Modeling	60	11.25	3.75	75	0.75	3	4
516	Advanced Biostatistics	60	11.25	3.75	75	0.75	3	4
517	Research, Planning, Monitoring and Evaluation	60	11.25	3.75	75	0.75	3	4
	Optional Paper Total				**225**	**2.25**	**9**	**12**
518	Viva-voce				100	1.0	4	-
General Group								
519	Practical				200	2.0	8	40
Thesis Group								
520	Written Thesis				100	1.0	4	-
521	In plant Training				50	0.5	2	-
522	Thesis Defense				50	0.5	2	-
Grand Total					**900**	**9**	**36**	**72**

COURSE: M-STAT 501
Advanced Statistical Inference
Full Marks: 75
(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)
Number of Lectures: 45
Examination hours: 4

Aim of the course
Aim of the course is to equip students in statistical inference covering the theory and the applications at an advanced level.

Objectives of the Course
- This course is designed to aid the interpretation of data that are subject to appreciable haphazard variability and to give a comprehensive Statistical basis for the analysis of such data, excluding considerations specific to particular subject matter.
- This course will give students a view of the nature of advanced statistical methods and to nurture advanced statistical thinking.

Learning Outcomes
After completing this course students will be able to
- know how attain the complex inferential targets using general statistical inference,
- handle any challenging statistical inference.

Course content

Sufficiency and Unbiasedness: Different types of statistical models, parametric, semi-parametric and non-parametric models, Group and exponential families of distribution, Sufficiency – minimal sufficient, Completeness and their relations – applications. Lehman – Scheffe Theorem, UMVU estimates, LMVU estimates, Necessary and sufficient condition of UMVU.

Asymptotic Estimation: Different types of consistency and their relations, Asymptotic normality, Asymptotic efficiency, properties of U and V statistics. Different examples and applications of U and V statistics. Bootstrap bias and standard error.

Parametric Estimation: Bayes decision and estimators. Invariance, equivariance and Pitman estimators. Minimaxity and admissibility. Loss function optimality. Maximum, Quasi-maximum and conditional likelihood methods of estimation. Asymptotically efficient estimators. Confidence region. Fiducial & Tolerance limits. Bayesian and Bootstrap intervals. EM algorithm.

Non-parametric and Robust Estimation: Distribution estimators. Density estimators. Different concepts of robustness. Statistical functional differentiability & asymptotic normality L-M-R-estimators. Robustness Vs. efficiency. Variance estimation. Robust estimation of multivariate location functional and scatter matrix.

Composite Hypotheses: Review of simple hypothesis and test criteria, generalized Neyman Pearson lemma. Locally UMPU test. Similar region & Neyman structure. Sufficient statistics and SR test. MPSR test. UMPSR test. Asymptotic efficiency of a test.

Sequential Test: Review of SPRT, OC and ASN functions. SPRT for three hypotheses. Sobel and Wald test. Armitage method for composite hypothesis. Wald theory of weight function. Cox's theorem. Sequential t-test, Sequential χ^2 test, Asymptotic Sequential t-test, Sequential analysis of variance, Sequential Multivariate Analysis.

Non-Parametric Test: Introduction. ARE and Robustness of a non-parametric test. McNemar test in 2x2 contingency analyses. Cox & Stuart test for trend. Cramer's contingency coefficient. Cochran test for related observations. ARE of Mann-Whitney test and Sign test. Kruskal-Wallis test & CRS design. Square rank test for variances. Quantile test. Friedman test. Kolmogorov one sample & two samples test.

Main Books:

Efron, B. & Tibshirani, R.J. (1993): An Introduction to Bootstrap. Chapman and Hall, N.Y.

George Casella and Roger L. Berger (2002): Statistical Inference, 2nd Ed., Thomson Learning Asia and China Machine Press.

Lehman, E.L. (1986): Testing Statistical Hypotheses, 2nd Ed., Wiley, N.Y.

Lehman, E.L. (1989): Theory of Point Estimation, 2nd Ed., Wiley, N.Y.

Rohatgi, V.K. & Ehsanes Saleh, A.K.M. (2001): An Introduction to Probability and Statistics. John Wiley and Sons, N.Y.

Silverman, B.W. (1986): Density Estimation for Statistics and Data Analysis, Chapman & Hall, London.

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

Books Recommended:

Ashraf Ali, M. (1974): Theory of Statistics, Vol.2, Angel Library, Nilkhet, Dhaka.

Barnet, V. (1982): Comparative Statistical Inference, 2nd Ed., Wiley, N.Y.

Efron, B. (1984): The Jackknife, the Bootstrap and Other Re-sampling Plans, SIAM.

Fraser, DHS (1985): Structure of Inference, Chapman and Hall, N.Y

Gibbons, J.D., and Chakraborti, S. (1992): Non-Parametric Statistical Inference, Marcell-Dekkar, N.Y.

Kalbfleisch, J.: Probability & Statistical Inference Vol.2, Springer-Verlag, N.Y.

Noreen, E.W. (1982): Computer-Intensive Methods for Testing Hypothesis, Wiley, N.Y.

Schervish (1995): Theory of Statistics, Springer Verlag, N.Y.

Shao, J. & Tu, D. (2000): Jackknife and Bootstrap, Springer-Verlag, N.Y.

Wetherill, G.B. & Glazebrook (1975): Sequential Methods in Statistics, 3rd ed., Chapman and Hall, London

COURSE: M-STAT 502
Advanced Experimental Design
Full Marks: 75
(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)
Number of Lectures: 45
Examination hours: 4

Aim of the Course
The main aim of experimental design is the observed variance in a particular variable is partitioned into components attributable to different sources of variation. It provides a statistical test of whether or not the means of several groups are equal. Comparisons of mean squares, along with an F-test allow testing of a nested sequence of models.

Objective of the Course
After completing this Course, student should

- Describe or explain the variation of information under condition that are hypothesized to reflect the variation,
- Include the establishment of validity, reliability and reliability and replicability
- Achieving appropriate levels of statistical power and sensitivity
- Understand comparative design are Important to choose between alternatives with narrow scope, suitable for initial comparison

Learning Outcomes

- Experimental design has long enjoyed the status of being the most used statistical technique in statistical, psychological as well as other field of research. It is probably the most useful technique in the field of statistical inference.
- Mathematical relationship which relates changes in a given response to changes in one or more factors may learn from statistical model.
- Design of experiment is computationally elegant and relatively robust against violations of its assumptions

Course content

Fractional Factorial and Main Effect Plan: Review of factorial experiment, confounding, fractional replication and related plan. Construction of plan with factors at 2 levels. Orthogonal arrays of strength 3 with factors at 2 levels. Orthogonal main effect plans factors at 3 and other levels. Mixed factorial experiment, Orthogonal main effect plans of size $2 \times S^r$. Analysis of orthogonal plans.

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

Lattice Design: Balanced lattices. Partially balance lattices. Rectangular lattices. Cubic lattices. Lattice squares- description. Statistical analysis with different replications.
Multivariate analysis of variance (MANOVA): Introduction. Omnibus MANOVA tests. Analysis and interpreting MANOVA. Causal models underlying MANOVA. Complex design.

Nested Design: Introduction, two stage nested design, three stage nested design.

Response Surface Design: Introduction, first order design, second order design, method of steepest ascent, difference between response surface design and usual design.

Main Books:

Cohran and Cox (2000): Experimental Design, 2nd ed., John Wiley, N.Y.
Federer, W.T. 1955. Experimental design, Mcmillan, New York.
Fisher, R.A.(1995): Design of Experiment, 8th ed., Hafner, N.Y.
John and Quenouille(1977): Experiments: Design and Analysis, 2nd ed., Charles Griffin, London.
Kutner, M. H., Nachtsheim, C. J., Neter, J. and Li, W. (2005). Applied Linear Statistical Models. 5th Edn., McGraw-Hill, Irwin.
Montegomery, D.C. (2005): Design and Analysis of Experiment, John Wiley, N.Y.
Steel, R.G.D and J.H. Torrie. 1980. Principles and procedures of statistics, 2nd edition, McGraw-Hill Book Co. New York.

COURSE: M-STAT 503
Advanced Multivariate Analysis
Full Marks: 75
(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)
Number of Lectures: 45
Examination hours: 4

Aims of the Course
The aim of this course is to explore the analyzing methods under Multivariate techniques such as Multivariate multiple regressions, Corresponding Analysis, Multivariate Mixture Distribution, Clustering, Independent Component Analysis and Bayesian Multivariate Regression and Factor Analysis.

Objectives of the Course
After completing this course, the students should

- understand all the features of Multivariate multiple regression analysis;
- apply and fit appropriate Multivariate multiple regression model according to the nature of data;
- understand all the features of Multivariate Mixture Distribution and appropriate clustering;
- understand how to separate the mixing signals through ICA; and

- understand how to apply Bayesian method in Regression and Factor Analysis.

Learning Outcomes

At the end of the course, the students will be able to

- know how to apply appropriate multivariate techniques to the data from a wide range of application area e.g., Neurology, Electroencephalographic recordings, Study of human brain, image analysis, Signal separation, Wireless communications, speech recognition, gene expression profile analysis, radiometric sky map, Volcanic eruption and climate change, Financial Econometrics, Business statistics, classifying and grouping sectors, Categorical data Management, Industrial Management and statistics, Bayesian prediction, noise analysis and so on.

Course content

Multivariate Regression Analysis: simple, multiple and multivariate multiple linear regression models. Assumptions. Parameter estimations and multivariate prediction. The distribution of likelihood ratio for the multivariate multiple regression model. Likelihood ratio test (LRT) including other multivariate test procedures. Relationship with canonical correlation analysis (CCA). Interpretation and conclusion.

Correspondence Analysis: Concept of correspondence analysis (CA). Algebraic development of correspondence analysis. Multiple correspondence analysis (MCA). Validation techniques in MCA. Similarities of CA and MCA with Categorical PCA and non-linear PCA. Application of CA for multiple factor analysis with contingency tables. Multiple factor analysis of mixed tables of metric and categorical data. Multi-block canonical correlation analysis for categorical variables. MCA and classification.

Multivariate Mixture Model: Definition and properties of finite mixture model. Maximum likelihood (ML) fitting of finite mixture models via EM algorithm. Multivariate Normal Mixture models: definition, properties and ML estimation via EM algorithm. Multivariate t Mixture models: definition, properties and ML estimation via EM algorithm. Mixture of principal component analyzers and mixture of factor analyzers.

Clustering and Data Mining: Classification, clustering and data mining. Some clustering methods: Fuzzy clustering, regression-wise clustering and clustering by finite mixture models. Data recovery models by averaging, linear regression, PCA, factor analysis and K-mean clustering.

Bayesian Fundamentals: Statistical distributions (scalar, vector and matrix distributions). Prior distributions (vague, conjugate, generalized and correlation priors). Hyper-parameter Assessment (binomial, scalar normal, multivariate normal and matrix normal likelihoods). Bayesian Estimation Methods (marginal posterior mean, maximum a posteriori).

Bayesian Multivariate Regression: Bayesian regression model, likelihood, conjugate priors and posterior, conjugate estimation and inference, generalized priors and posterior, generalized estimation and inference, interpretation and discussion.

Bayesian Factor Analysis: Bayesian factor analysis model, likelihood, conjugate priors and posterior, conjugate estimation and inference, generalized priors and posterior, generalized estimation and inference, interpretation and discussion.

Introduction ICA/BSS: Independent component analysis (ICA) or blind source separation (BSS) model, PCA verses ICA. Estimation and inference, interpretation and discussion, Application.

Main Books:

Greenacre, Michael (2007). *Correspondence Analysis in Practice, Second Edition*. London: Chapman & Hall/CRC.

Johnson, R. A & Wichern, D. W. (2002): *Applied Multivariate Statistical Analysis*, 5th ed. Prentice- Hall, N.Y.

Rowe, D. B. (2003): *Multivariate Bayesian Statistics*, Chapman & Hall /CRC

Books Recommended:

Anderson, T. W. (1984) : *An Introduction to Multivariate Statistical Analysis*, 5th ed. Wiley,N. Y.

Hyvarinen, A., Karhunen, J. & Oja, E. (2001): Independent Component Analysis, Wiley, N.Y.

Izenman, A.J. (2008); Modern Multivariate Statistical Techniques: Regression, Classification and Manifold learning, Springer, N.Y.

Jajuga, K., Sokolowski, A. & Bock, H.-H. (2002): *Classification, Clustering and Data Analysis*: Recent Advances and Applications, Springer. N.Y.

Jolliffe, I. T. (2002): *Principal Component Analysis*, 2nd ed. Springer, N.Y.

Kaufman, L. & Rousseeuw, P. J., (2005): *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, N.Y.

McLachlan, G. & Peel, D. (2000): *Finite Mixture Models*, Wiley, N.Y.

Mirkin, B. (2005): *Clustering for Data Mining: A Data Recovery Approach*, Chapman & Hall, N. Y.

COURSE: M-STAT 504
Time Series Analysis and Forecasting
Full Marks: 75

(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75
Number of Lectures: 45
Examination hours: 4

Aim of the Course

The aim of this course is to get acquainted with important concepts of time series analysis and its applications.

Objectives of the Course

Students should know how to:

- Solve difference equations of a system with time series operator
- Model and forecast time series data properly
- Analyze data and signals in frequency domain and compute spectral density
- Model volatility of financial time series
- Perform multivariate time series analysis and discover interdependence

Learning outcomes

Having completing this course, students will able to do

- Apply statistical theory and methods of time series regression applicable to in economic business, environmental, geological and astrophysical problems
- Explore trends of social and economic indicators
- Estimate models for time-series data
- Interpret the results of an implemented time series analysis
- Aware of limitations and possible sources of errors in the analysis
- Present results in oral and written form

Course content

Introduction: Components of time series, Stationarity, Ergodicity, White noise, Autocorrelation function, Partial autocorrelation function, Difference equations and their solution, Basic ARMA models and their extension. Box-Jenkins modeling philosophy and forecasting.

Spectral Analysis: Introduction, Fourier transformation. Periodogram, Spectral representation, Spectral density, Spectral densities for ARMA processes.

Non-stationary Time Series: Trend stationary and difference stationary time series, Integrated process, Unit roots, Unit root tests, Structural changes and their consequences, Filtering, ARIMA modeling, SARIMA modeling.

Multivariate Time Series: Structural, recursive and reduced form vector autoregressive (VAR) models, Granger causality, Impulse response functions, Forecast error variance decomposition. Spurious regression and cointegration, Tests for cointegration: Engle-Granger methodology and Johansen's methodology, Error correction models.

Time Series Model of Heteroskedasticity: Stylized facts of financial time series. Volatility clustering, Detection of autoregressive conditional heteroskedasticity (ARCH) effects, Modeling volatility, ARCH model, Extension of ARCH model: GARCH, TARCH, GJR-GARCH, FIGARCH, EGARCH, IGARCH, PARCH, NARCH models.

Main Books:

Andersen, T. G. and A. R. Davis (2009): Handbook of Financial Time Series, Jens-Peter Kreifs and Thomas Mikosch edition, Springer-Verlag.

Hamilton, J.D. (1994): Time Series Analysis, Princeton University Press, N.J.

Shumway, R. H. and D. S. Stoffer (2006): Time Series analysis and its Applications with R Examples.

Books Recommended:

Anderson, T.W. (1971): The Statistical Analysis of Time Series, Wiley, N.Y

Box, G.E.P. and Jenkins, G.M. (1976): Time Series Analysis: Forecasting and Control, Holden-Day, Sun Francisco.

Cryer, J. D. and K. Chan (2008): Time Series Analysis: with applications in R, 2nd Ed., Spinger, N.Y.

Findley, D.F. (1981): Applied Time Series Vol. I & II, Academic press, N.Y.

Fuller, W.A (1976): Introduction to Statistical Time Series, Willey N.Y

Lűtkepohl, H. (2005): New Introduction to multiple Time Series Analysis, Springer, N.Y.

Reinsel, G. C. (2003): Elements of Multivariate Time Series Analysis, Springer, N.Y.

Tsay, R. S. (2010): Analysis of Financial Time Series, Wily & Sons, N.J.

COURSE: M-STAT 505
Data Mining
Full Marks: 75

(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)

Number of Lectures: 45

Examination hours: 4

Aim of the Course

Data that has relevance for managerial decisions is accumulating at an incredible rate due to a host of technological advances. From this flood of digital data we have to extract meaningful information and knowledge for the development of business, government and scientific community. Data mining is a class of analytical techniques that examine a large amount of data to discover new and valuable information. This course is design is to introduce the core concepts of data mining.

Objectives of the Course

After completing this course, the students should

- understand the basic concept about data mining;
- able to explore categorical and numerical data and also apply proper technique for preprocessing the data;
- enlighten fundamental concepts and algorithms for supervised learning, unsupervised learning and semi-supervised learning, to provide the students with the necessary background for the application of data mining to real problems.
- develop and apply critical thinking, problem-solving, and decision-making skills

Learning Outcomes

At the end of the course, the student will be able to

- know what is data mining and how it is used;
- know the different technique for pre-processing the data;
- know different supervised, semi-supervised and unsupervised learning method for classification, prediction and clustering the data;
- know about association rules and model evaluation technique; and
- know how to use data mining technique for real data analysis and its interpretation.

Course content

Introduction: What and why is Data Mining? Need for Human Direction of Data Mining**,** Fallacies of Data Mining**,** Data Mining Tasks, Data Mining Process, Data Preprocessing, Data Cleaning, Handling Missing Data**,** Identifying Misclassifications**,** Graphical Methods for Identifying Outliers, Data Transformation and Numerical Methods for Identifying Outliers.

Exploratory Data Analysis (EDA): Introduction, Hypothesis Testing versus Exploratory Data Analysis**,** Getting to Know the Data Set, Dealing with Correlated Variables, Exploring Categorical Variables**,** Using EDA to Uncover Anomalous Fields Exploring Numerical Variables, Exploring Multivariate Relationships, Selecting Interesting Subsets of the Data**,** Binning.

Unsupervised Learning: Association rules- Affinity Analysis and Market Basket Analysis, Data Representation for Market Basket Analysis, Support, Confidence, Frequent Item sets, and the A Priori Property**,** Genetic Algorithm.

Supervised Learning: Introduction, Decision tree, Random Forest, Ensemble Learning, Neural Network, K-nearest-Neighbor methods, Support Vector Machines, Deep Learning.

Semi-Supervised Learning: Overview of Semi-Supervised Learning, Learning from Both Labeled and Unlabeled data, Inductive and Transductive Semi-Supervised Learning, Mixture Model for Semi-Supervised Learning.

Main Books:

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

Daniel T. Larose (2005)**.** Discovering Knowledge In Data: An Introduction to Data Mining, Wiley Interscience, N.J., USA.

References:

Patricia B. Cerrito (2006). Introduction to Data Mining *Using SAS ® Enterprise Miner,* SAS Institute Inc., Cary, NC, USA. (Practical book 1)

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.

B. D. Ripley (2002). Statistical Data Mining, Springer-Verlag, New York.

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

Zhu, X. J. and Goldberg, A.B. (2009). Introduction to Semi-Supervised Learning, Morgan & Claypool Publishers.

John Shawee-Taylor and Nello Cristianini (2004). Kernel Methods for Pattern Analysis, Cambridge University Press, New York. USA.

Danial T. Larose (2006). Data Mining Methods and Models, John Wiley & Sons, Inc.

COURSE: M-STAT 506
Genomics and Bioinformatics
Full Marks: 75
(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75**)**
Number of Lectures: 60
Examination hours: 4

Aim of the course:

- The aim of this course is twofold: to provide an overview of the most common statistical methods for molecular genomics and transcriptomics data analysis, and to provide the necessary information for solving the complex biological problems and achieving the satisfactory score of sustainable development goal (SDG) index from the agriculture and health sectors.

Objective of the Course:
The main objective of this course are

- to understand statistical modeling on molecular genomics and transcriptomics datasets
- to learn most common statistical methods for genomics and transcriptomics data analysis
- to develop the capability of statistical model building strategies for genomics and transcriptomics data analysis
- Hands-on training on genomics and transcriptomics data analysis to understand how to provide the necessary information to solve the complex biological problems

Learning Outcomes:
After completion of this course successfully, the learners/students would be able

- to analyze genomics and transcriptomics datasets to provide the necessary information to solve the complex biological problems that are associated with the genetic factors
- to select appropriate statistical algorithms for analyzing genomics and transcriptomics datasets
- to contribute to the development of high yielding varieties and to achieve the satisfactory score of SDG index from the agricultural sector.
- to contribute to the discovery of new drugs/vaccines for the complex diseases and to achieve the satisfactory score of SDG index from the health sector.

Course content

Basic Genomics: Introduction. Classification of Genomics. Genes and Chromosomes. Cell division, Nucleic acid, Molecular Genetics (Marker, DNA sequence, SNPs). Genotype and Genotyping technology. Central dogma (DNA/Gene transcription, RNA translation and Protein synthesis).

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 DNA Sequence. Some bioinformatics databases including GeneBank, NCBI, PDB, BLAST and FASTA.

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

Gene-Expression Data Analysis: Introduction to different types of microarray gene expression data. 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. Inferring genetic regulatory networks from microarray experiment with Bayesian networks. Modeling genetic regulatory networks using gene expression profile. Gene-set enrichment analysis.

Genome-wide Association Studies (GWAS): Introduction . QTL and SNP analysis with Gene-Expression data. SNP analysis using contingency table. 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.

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, Cold Spring Harbor Laboratory Press.

Recommended Books:

Xu, S. (2013). *Principles of statistical genomics*. Springer.

Liu, B. H. (1997). *Statistical genomics: linkage, mapping, and QTL analysis*. CRC press.

Ferreira, M. A. R., Medland, S. E., & Posthuma, D. (2008). *Statistical genetics: gene mapping through linkage and association*. New York: Taylor & Francis.

Gondro, C., Van der Werf, J., & Hayes, B. J. (Eds.). (2013). *Genome-wide Association Studies and Genomic Prediction*. Humana Press.

COURSE: M-STAT 507
Advanced Demography
Full Marks: 75
(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)
Number of Lectures: 45
Examination hours: 4

Aim: Firstly, observed the values of the demographic parameters using indirect techniques and secondly, take necessary future plan of action on the basis of the findings for the nation of country.

Objectives:

- to understand the concepts of single, double and multiple decrement life tables.
- to observed eliminating effect of specific disease on expectation of life.
- to estimate different parameters of vital events using indirect techniques in absence of direct information.
- to understand the concept of population growth models and their graduation.
- to understand the application of projection matrix for fertility and mortality.
- to understand the concept of fecundability, its distribution and effect the use of contraceptives.

Application:
The fields of application include public health, reproductive health, population policy, such as fertility, mortality, migration etc.

Course content

Life table analysis: General idea, ordinary life table. Properties and interrelationships. Probability distributions of life table functions. Methods of construction of double and multiple decrement tables and increment decrement life tables. Joint life functions, Multi-life functions. Last survivor status. General Multi-life status.

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, Coale's nuptiality model. Davis-Blake framework of intermediate variables. Proximate determinants of fertility.

Stable population theory and models: Concept of stationary, stable and quasi-stable population. Natural growth rate and intrinsic growth rate, Lotka's Intregal Equation, Net maternity function. Graduation of NMF by Normal, Wicksell and Hadwiger curve. 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. Frejka's component method for population projection. Projection of fertility and mortality.

Micro demography: Reproductivity, fecundity, fecundability and sterility. Effective fecundability. Residual fecundability. Estimation of fecundability. Pearl index. Effectiveness and efficiency of FP methods.

Main Books:

Pollard, J.H. (1980): Mathematical models for the growth of human populations.

Johnson RCE & Johnson (1980): Survival Models and Data Analysis, Wiley NL & Sons, NY.

Keyfitz, N. (1977): Applied Mathematical Demography, Wiley & Sons.

Books Recommended:

Biswas, S. (1988): Stochastic Processes in Demography and Applications. Wiley Eastern Ltd., India

Islam, N. (1996): Levels and Correlates of Marriage and Fertility in Bangladesh. Unpublished Ph.D desperation, R.U.

Shryock, H. J. S. Siegel and Associates (1976): The Methods and Materials of Demography; Cond., ed., New York, Academic Press.

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

UNO (1983): Indirect Technique Demographic Estimation, Population Studies No. 81.

COURSE: M-STAT 508
Environmental Statistics
Full Marks: 75

(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)

Number of Lectures: 45

Examination hours: 4

Aim of the Course:

The aim of this course is a broad discipline stretching from how and what to sample, through to modeling impacts on human and ecosystem health and ultimately to providing predictions of what changes might occur in the future. It acts as a bridge between the fundamental methods of the subject and important applications in a wide variety of environmental issues.

Objectives of the Course:

This course will develop the students' ability to

- Understand the complex relationships between natural and human systems.
- Improve knowledge of the environment.
- Provide quantitative information about the environment's state and its most important changes over time across territories.
- Achieve by setting up, strengthening and sustaining environmental issues which already operating in economic and social statistics.

Learning Outcomes:

On successfully completion of this course, the student will be able to:

- Apply mathematical concepts, including statistical methods, to field and laboratory data to study scientific phenomena.
- Develop an understanding of current environmental monitoring systems,
- Apply knowledge acquired to the process of environmental impact modelling and prediction as a design tool with application to a number of case studies.
- Design and execute a scientific project.
- Adapt skills in GIS to environmental management systems.
- Develop skill for achieving sustainable development goals.

Course content

Introduction: Environmental Variables - Discrete and continuous; Data collection - primary and secondary; Presentation of data - spatial and non-spatial data.

Design and Analysis of Environmental Data: Conceptual Foundations Methods, Environmental data, Data Exploration, screening and adjustment, Confidence Intervals and More, Deterministic functions, Bestiary of probability distributions, Continuous probability distributions, Discrete probability distributions, Statistical Models – putting it all together, Frameworks for statistical Inference, Bayesian Inference, Hypothesis testing concepts, Nonparametric Inference: Ordinary least squares and more, Maximum Likelihood inference.

Landscape of Statistical Methods: General linear models. Nonlinearity, Nonlinearity and non-normal errors (generalized linear models)., Heterogeneous errors, Correlated errors, Multi-level models (mixed effects models), Multivariate statistics, Finding groups - cluster analysis, Testing/describing group differences, Un-constrained ordination, Constrained ordination

Hazard in the environment: Concept of risk, vulnerability, hazard, and disaster; Types of Natural Hazards and their Global and National perspectives, Role of Global climatic changes and Global warming. Causes and consequences of Global Warming, Sea level rise in climate.

Study of Agro-meteorological Features: Fundamentals Concept of Meteorology and Climatology. Desertification, Drought and Flood management and Modeling Analysis: Flood hazard and its management: Definition, Causes, nature, frequency of flooding and its impacts.

Desertification and Drought – Causes of desertification; Evaluation of desertification hazard – potential and zoning: Drought - causes, types, distribution and management.

Food Security and Environmental Impact on Health and Agriculture: Pollution and Soil degradation, Deforestation, Land use pattern and regional pattern of productivity.

Case Study of Environmental Data Analysis:

1. Applications of probability distributions and Markov chain model.
2. Drought Identification and Characterization at Local, National and Global level:
3. Drought indices by Standardized Precipitation Index (SPI),
4. Drought Prediction Using Markov chains modeling,
5. Drought indicators : A Stochastic approach to evaluation.
6. Applications of non-linear and non-stochastic Time series analysis, Wavelets analysis, Spectral analysis.
7. Study of validity and uncertainty in environmental modeling.

Geographical Information System (GIS): Basic principles, Raster and vector data, Map Projection, Overlay analysis, Data structure and Digital cartography.
Global Positioning System (GPS): Basic principles, Applications to environmental studies.

Main Books:

G.P. Patil & C.R. Rao (1999): Environmental statistics: analyzing data for environmental policy, John Wiley and Sons.

H. R. Byers (1974): General Meteorology, McGraw-Hill

Isaacson D.L., Madsen R., John (1976): Markov Chains: Theory and Applications. Wiley, New York.

Vic Barnett (2004): Environmental Statistics: Methods and Applications (Wiley Series in Probability and Statistics), John Wiley & Sons.

Books Recommended:

Box, G.E.P. and G.M. Jenkins, (1976): Time Series Analysis Forecasting and Control. San Francisco: Holden-Day.

Bryan F.J. Manly, Statistics for Environmental Science and Management, Second Edition (Chapman & Hall/CRC Applied Environmental Statistics).

G. F. White (ed) (1974): Natural Hazards – Local, National, Global, Oxford University Press

G. F. White (ed): Natural Hazards – Local, National, Global: Oxford University Press.

G. T. Trewartha (1968): An Introduction to Climate; McGraw-Hill

Linda Courtenay Botterill and Geoff Cockfield (2013): Drought, Risk Management, and Policy: Decision-Making Under Uncertainty. Australia

P. Reining (1978): Handbook of Desertification Indicators (Washington D.C.: American Association for the Advancement of Science.

V.T. Chow (1964): Handbook of Applied Hydrology, McGraw-Hill, New York.

Velma I. (2012): Impact of Climate Change on Water and Health, Grover, Publisher: CRC Press, John Wiley and Sons.

Wayne R. Ott (1979): Environmental Indices: Theory and Practice, Publisher: CRC Press, John Wiley and Sons.

COURSE: M-STAT 509
Advanced Actuarial Statistics
Full Marks: 75
(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)
Number of Lectures: 45
Examination hours: 4

Aims and Objectives
The aim of the Actuarial Statistics subject is to provide a grounding in mathematical and statistical modelling techniques that are of particular relevance to actuarial work, including stochastic processes and survival models and their application.

Learning Outcomes
On successful completion of this subject, a student will be able to:

- describe and use statistical distributions for risk modelling.
- describe and apply the main concepts underlying the analysis of time series models.
- describe and apply Markov chains and processes.
- describe and apply techniques of survival analysis.
- describe and apply basic principles of machine learning.

Course content

Insurance: Nature and functions of insurance ; Benefits and costs of insurance system to the society; Economic theories of insurance; The mathematical basis for insurance; Insurable interest; Principle of indemnity; Doctrine of subrogation; Warranties; Proximate causes; Assignment of transfer of interest; Return of premium.

Life Insurance: Essential features of life insurance contract; Risk selection for life insurance; Sources of risk information; Classification of annuities.
Health insurance: Types of health insurance coverage; Exclusion in health insurance policies. Payment of claim. Fire insurance, Marine insurance, and other insurance: Concepts, features, policies and coverage; payment of claim;

Risk Management: Loss, peril, hazard and risk; Types of risk; Development and implementation of risk management programs; Methods of dealing with risk; personal risk management.

The economics of insurance, utility theory, Application of probability to problems of life and death, the determination of single premiums for insurances and annuities in both the discrete and continuous case.

Theory and practice of pension funding, Assumptions, Basic actuarial functions and population theory applied to private pensions.

Survival distributions and life tables, Life insurance, Life annuities, Net premium, Premium series, multiple life functions, multiple decrement models, Valuation theory for pension plans, the expense function and dividends.

Measurement of risk and Mortality Table: Mortality tables and its classification; Construction of mortality tables; Premium calculation of various life policies.

Exposure formulas: Assumed and using implications, Techniques of calculating exposures form individual records including consideration involving selection of studies, various observation periods and various methods of tabulating deaths. Techniques of calculating exposures from variation schedules including the general concepts of fiscal year. The use of interim schedules and variations in observation period or method of grouping deaths and practical aspects of construction of actuarial tables.

Books Recommended:

Dorfman, Mark.S (1991). Introduction to Risk Management and Insurance, Prentice Hall.

Mishra, M.N. (1989). Principles and practice, S.Chand and Company.

COURSE: M-STAT 510
Proteomics and Biomedical Informatics
Full Marks: 75

(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)

Number of Lectures - Minimum 60

Examination hours: 4

Aim of the course:

The aim of this course is twofold: to provide an overview of the most common statistical methods for molecular proteomics data analysis, and to provide the necessary information for solving the complex biological problems and achieving the satisfactory score of sustainable development goals (SDGs) index from the health sectors.

Objective of the Course:

The main objective of this course are

- to understand statistical modeling on molecular proteomics datasets
- to learn most common statistical methods for proteomics data analysis
- to develop the capability of statistical model building strategies for proteomics data analysis
- Hands-on training on proteomics data analysis to understand how to provide the necessary information to solve the complex biological problems

Learning Outcomes:

After completion of this course successfully, the learners/students would be able

- to analyze proteomics datasets to provide the necessary information to solve the complex biological problems that are associated with the genetic factors
- to select appropriate statistical algorithms for analyzing proteomics datasets
- to contribute to the development of high yielding varieties and to achieve the satisfactory score of SDG index from the agricultural sector.

- to contribute to the discovery of new drugs/vaccines for the complex diseases and to achieve the satisfactory score of SDG index from the health sector.

Course content

Introduction: Cell Structure and Function, Cell components. Chromosome, Chromosome structure and organization. DNA, RNA, Gene and Central dogma and bioinformatics. Introduction to Bioinformatics. Importance/scope/Applications of Bioinformatics. DNA sequencing. Shotgun sequencing, Long repeats, r-scane.

Protein Sequencing and Amino Acids: Amino acids and Amino Acids structure and functions. Codons. Metabolic and Biochemical pathway analysis e.g. *E.coli*. pathways. Sequence alignment, Overview of methods of sequence alignment. Dynamic programming algorithm for sequence alignment, Multiple Sequence alignments. statistical methods for aiding alignment.

Protein Databases: Overview of the use and maintenance of different databases in common use in biology. Databases: GenBank, DDBJ, EMBL NCBI, EFI, UniGene, UniProt, Swiss-Prot, PDB. BLAST and FASTA analysis.

Phylogenetic Analysis of Protein Sequence: Motivation and background on phylogenetics**,** Distance and clustering approach, Likelihood methods**,** Parsimony, RNA-based phylogenetics methods, Phylogenetic Tree Estimation.

Protein Classification, Structure and Prediction: Protein Structure Prediction: Methods for predicting the secondary and tertiary structure of proteins. Techniques: neural networks, SVMs, genetic algorithms and stochastic global optimization.

Medical Informatics: Introduction to Medical Informatics. Perspectives and goals of Medical Informatics. History, Taxonomy and standards of Medical Informatics, Organization of Medicine and Health Information, Paper-based Medical Report and Electronic Medical Report (EMR), Pervasive Healthcare.

Drug Discovery Informatics: Metabolome and Metabolomics. Systems biology, Approaches to drug and vaccine design using bioinformatics tools, Molecular docking using Autodock and/or other computer aided programs.

Network Analysis and Disease Prediction: Scope and applications of Network analysis in medical informatics. Bayesian Network (BN) Analysis and application, Artificial Neural Network (ANN) Analysis and application. Other relevant network. Disease Surveillance, Disease prediction models. Survival analysis. Risk classifications. CAPRA and D'Amico risk classifications. Nomogram development for disease prediction.

Main Books:

Shortliffe E.Hand J.J. Cimino (2006**).** *Biomedical Informatics: Computer Applications in Health Care and Biomedicine (Health Informatics)*. Springer-Verlag. .

Husmeier, D., Dybowski, R., Roberts, S. (2005). *Probabilistic Modeling in Bioinformatics and Medical Informatics*, 2nd edition, Publisher: Springer.

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

Books Recommended:

Carey, V. J., Huber, W., Irizarry, R. A., & Dudoit, S. (2005). *Bioinformatics and computational biology solutions using R and Bioconductor* (Vol. 746718470). R. Gentleman (Ed.). New York: Springer.

Pevsner, J. (2009). *Bioinformatics and functional genomics*. John Wiley & Sons.

Carey, V. J., Huber, W., Irizarry, R. A., & Dudoit, S. (2005). *Bioinformatics and computational biology solutions using R and Bioconductor* (Vol. 746718470). R. Gentleman (Ed.). New York: Springer.

COURSE: M-STAT 511
Health and Epidemiology
Full Marks: 75

(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)

Number of Lectures: 45

Examination hours: 4

Aim of the Course

The aim of this course is to provide students with the necessary knowledge and skills to be able to critique, design and conduct and human population based research, including observational studies and randomized controlled trials. More specifically, to provide advanced knowledge of statistical analysis for epidemiological data, study design and protocol development.

Objectives of the Course

- To define epidemiology and public health
- To list the functions of public health
- To describe the assessment of evidence of causation
- To outline the process of population-based health management
- To describe different public health programs
- To discuss about the study design, sample size determination, sampling procedure, and data analysis for health science and epidemiology

Learning outcomes

At the end of the course, students should have the knowledge and skills to be able to:

- Critically appraise and evaluate the design, analysis and interpretation of health science and epidemiological studies

- Describe and discuss the role and contribution of epidemiology to health
- Select, devise and develop appropriate study designs for health science and epidemiological research
- Conduct appropriate statistical analyses for health and epidemiological data

Course content

Introduction: Some examples of health study; Selection of health problem for research; Framing different type of study design: prospective study, retrospective study, longitudinal study, experimental study, observational study, intervention study, single blind study, double blind study; Selection of study population and cases for the study; causal variables, confounder variables; Ethical issue related to health study.

Measuring Health and Disease: Sources of health statistics. Meaning and Concept. Biological variations. Health indicators classification. Morbidity concepts and measures. Illness, diseases and their classifications. Multiple causation of diseases The conquest and resurgence of infectious diseases. The biomedical basis of chronic diseases. Measuring diseases frequency and errors in measurement. Mortality. Nutrition in Bangladesh; special emphasis on infant and child nutrition.

Statistics in Epidemiology: Categorical response data in epidemiology and diagnosis-prognosis; OR, RR, NNT, PPR, NPR, sensitivity, specificity, true positive rate, true negative rate; ROC analysis; power, precision, sample size calculation.

Statistical Models in Epidemiology: Data Analysis and interpretation using statistical models. Linear regression model; log linear models, building and applying log linear models, log linear and logit models for ordinal variables; multinomial response models. Models for mixed health hazards; multilevel sources of variation. Analysis of repeated epidemiological outcomes. Parametric and semiparametric models for complete and incomplete data in epidemiology, Cox proportional hazards model.

Mathematical Models in Epidemiology: Basic Concepts, SI Model Formulation, Solution and Interpretation, SIS Model with Constant Coefficient Formulation, Solution and Interpretation, SIS Model with Constant Coefficient is a function of time *t,* SIS Model with Constant Number of Carriers Formulation, SIS Model When the Carrier is a Function of Time t, General Deterministic Model with Removal (SIR Model) Formulation, Solution of Model Equation Interpretation, Epidemic Model with Vaccination Solution of Model Equation

Main Books:

Bonita R, Beaglehohe R and Kjellstrom T (2006): Basic Epidemiology, 2nd edition.
Josepn L. Fleiss (1973): Statistical Methods for Rates and Proportions. Wiley, NY.
Klien JP and Moeschberger, ML (2003): Survival Analysis: Techniques for Censored and Truncated Data. 2nd edi, Springer.

Books Recommended:

Armitage, P (1971). Statistical Methods in Medical Research. Blackwell, NY.
Lawless, JF (1982). Statistical Models and Methods for Lifetime Data. Wiley, NY.

Schneider, Mary Jane (2006): Introduction to Public Health, Gaithersburg, Maryland.
Schoenbach VJ and Rosamond WD (2000): Understanding the Fundamentals of Epidemiology, an evolving text.
Singh,B and Agarwal, N(2005), Bio-Mathematics, First Edition, Krishna Prokashan Media (p) Ltd., Meerat, India.
Spiegelman, M. (1968): Introduction to Demography, North Holland.

COURSE: M-STAT 512
Statistical Methods in Industrial Management
Full Marks: 75

(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)
Number of Lectures: 45
Examination hours: 4

Aim of the Course

In Industrial management prepares students to take on management responsibility is the middle management in private industry, the service sector or public administration.

Objective of the Course:

- To deepen and broaden the student's knowledge.
- To cover the theory and practice of modern econometric.
- To teach the students the habits of thought knowledge are understanding.
- The course is application oriented.

Learning Outcomes:

Student will be able to:

- Perform the management functions.
- Compare selected theories of management.
- Perform the functions is the marketing mix.
- Use basic nosiness application.
- Assessor ethical issue in business situation

Course Content

Industrial management: Meaning, Principles, Characteristics and importance. Management by objectives. Advantages and disadvantages of management by objectives. Industrial organization chart. Decision making techniques. Industrial accidents and safety. Causes of accidents. Methods of reducing accidents. Job analysis and evaluation. Methods of job analysis and evaluation. Production and productivity. Factors of production. Tools of productivity. Factors affecting productivity.

Industrial Psychology: Quality, Scope and Aims of industrial psychology, Problem of industrial psychology, Pattern of human behavior, Human needs, Difficulties and suggestion for improvement of human and social relationships.

Quality control and inspection: Objectives of quality control. Quality conformance. Quality principles. Advantages of quality control. Quality certification, Objectives of inspection. Various kinds of inspection. Advantages and disadvantages of inspection. Quality philosophies. Deming philosophy. Tools and techniques. Total quality management (TQM). Implementation of TQM, Philosophies of TQM. Pareto analysis, Cause-and-effect-diagrams. Failure modes and effect analysis (FMEA).

Corrective maintenance (CM) and preventive maintenance (PM): Preventive maintenance models for complex systems. Maintenance of repairable systems. Models for complex repairable systems. Parameter estimation. Model selection. Preventive maintenance scheduling. Reliability centered maintenance (RCM). Steps of the RCM process. Inventory. Managing and controlling inventory. Forecasting for inventory management of service parts. Supply-chain management. Case studies in industrial management and maintenance.

Network Analysis: Network Techniques. Project Management, Key decisions in project management, Project life cycles, planning and scheduling with Gantt Charts, PERT and CPM strategies, Steps in PERT and CPM project Plan. Deterministic time estimate, Probabilistic time estimate.

Main Books:

Mukhi, H. R. (2001): Industrial Management, Satya Prakashan, New Delhi.

William J. Stevenson (2005): Operation Management, 8th edition, McGraw Hill Company, N.Y.

Books Recommended:

Davis Aquilano and Chase (2004): Fundamentals of operation Management, 4th ed., McGraw Hill Company, N.Y.

Hill, T. (2000): Operations Management: Strategic, Context and Managerial Analysis, University of Oxford.

Hiller and Hiller (2005): Introduction to Management Science, 2nd ed., McGraw Hill Company, N.Y.

Kelly A. and Harris, M.J. (1978): Management of Industrial Maintenance, Butterworth Heinemann Ltd.

Kobbacy, K.A.H and Murthy, D.N.P. (2008): Complex System Maintenance Handbook, Springer-Verlag.

Zipkin (2005): Fundamental of inventory Management, 1st ed., McGraw Hill Company, N.Y.

COURSE: M-STAT 513
Statistical Methods for Reliability Data
Full Marks: 75
(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)
Number of Lectures: 45
Examination hours: 4

Aims of the course
The aim of this course is to provide an analytical introduction to the core concepts of reliability and maintenance with emphasis on more advanced topics in statistical methods for reliability data and analysis of field reliability data, accelerated failure time data, software reliability modeling.

Objectives of the Course
After completing this course, the students should

- understand a general and advanced strategies that can be used for data analysis, modeling, and inference from reliability data;
- critically analyze field-failure data, repairable system and recurrence data;
- develop the capacity of modelling accelerating life tests data, degradation data;
- comprehend maintenance and software reliability.

Learning Outcomes
At the end of this course, the students will be able to understand

- basic ideas behind product reliability, reasons for collecting reliability data and distinguishing features of reliability data;
- general strategy that can be used for data analysis, modeling, and inference from reliability data;
- system reliability modeling, the distribution of system failure time as a function of individual component failure-time distributions, estimate system reliability, analysis of data with more than one failure mode;
- typical data from repairable systems and other applications that have recurrence data, the combined use of simple parametric and nonparametric graphical methods for drawing conclusions from recurrence data;
- applications of accelerated life testing, nonparametric and graphical methods for presenting and analyzing accelerated life test data;
- degradation models, connection between degradation models and failure-time models, differences between degradation data analysis and traditional failure-time data analysis;
- failure time regression models and prediction of reliability;
- concept of maintenance, preventive and corrective maintenance.
- some basic ideas of software reliability modeling;

Course content

Reliability concepts: Basic concept of reliability. Examples and features of reliability data. Strategy for collection, modeling, and analysis of reliability data. Models for continuous failure-time processes. Models for discrete data from a continuous process.

Component and system reliability concepts and methods: Location-scale-based distributions - concept and applications in reliability. Probability plots. Reliability block diagram, component reliability, system reliability, reliability of series and parallel systems. Failure mode. Competing risk model. Mixture model.

Analysis of repairable system and recurrence data: Intensity function, mean cumulative function, tests for recurrence rate trend. Models for perfect repair, minimal repair and imperfect repair – derivation and estimation.

Accelerated failure time models: Accelerating variables, life-stress relationships and acceleration models. Guideline for the use of accelerating models. Non-parametric and graphical methods for presenting and analyzing accelerated life test (ALT) data. Likelihood methods for analyzing censored data from an ALT. Suggestions for drawing conclusions from ALT data. Potential pitfalls of accelerated life testing.

Degradation data, models, and data analysis: Degradation data. Models for degradation data. Estimation of model parameters. Comparison with traditional failure-time analysis. Approximate degradation analysis.

Failure time regression models: Models fitting and applications in reliability.

Software reliability modeling: Concept of software reliability. Software reliability modeling and estimation. Software testing procedures. Prediction and management of software reliability.

Prediction of reliability: Motivation and prediction problems. Naive method for computing a prediction interval. Prediction of future failures from a single group of units and from multiple groups of units with staggered entry into the field.

Maintenance: Maintenance, preventive and corrective maintenance, optimum preventive maintenance – concept and applications.

Case studies: Analysis of field reliability data.

Main Books:

Meeker, W. Q. and Escobar, L. A. (1998): Statistical Methods for Reliability Data, Wiley, N.Y.

Nelson, W. (1990): Accelerated Testing: Statistical Models, Test Plans, and Data Analyses, Wiley, New York.

Books Recommended:

Ansell, J. I. and Phillips, M. J. (1994): Practical Methods for Reliability Data Analysis, Clarendon Press, Oxford.

Balakrishnan, N. and Rao, C. R. (Eds.) (2001): Handbook of Statistics, Vol. 20, Advances in Reliability, Elsevier, The Netherlands.

Bain, L. J. and Engelhardt, M. (1991): Statistical Analysis of Reliability and Life Testing Models, Theory and Methods, 2nd ed., Marcel Dekker, New York.

Blischke, W.R., Karim, M.R., and Murthy, D.N.P. (2011). Warranty Data Collection and Analysis, Springer-Verlag London Limited.

Blischke, W. R. and Murthy, D. N. P. (2000) Reliability. Wiley, New York.

Gámiz, M. L., Nikolaos Limnios, K. B. Kulasekera and Bo Henry Lindqvist (2011): Applied nonparametric methods in reliability, Springer-Verlag.

Hamada, M. S., Alyson G. Wilson, C. Shane Reese and Harry F. Martz (2008): Bayesian Reliability, Springer-Verlag.

Kalbfleisch, J. D. and Prentice, R. L. (1980): The Statistical Analysis of Failure Time Data, Wiley, New York.

Kobbacy, K.A.H and Murthy, D.N.P. (2008). Complex System Maintenance Handbook, Springer-Verlag.

Lawless, J. F. (2003): Statistical Models and Methods for Lifetime Data, 2nd ed., Wiley, N.Y.

Sinha, S. K. (1986). Reliability and Life Testing, Wiley Eastern Ltd., India.

Tobias, P. A. and Trindade, D. C. (1995): Applied Reliability, 2nd ed., Van Nostrand Reinhold, New York.

COURSE: M-STAT 514
Physical Health and Human Growth Modeling
Full Marks: 75

(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)

Number of Lectures: 45

Examination hours: 4

Aim of the Course

The aim of this course is to cover out the in depth knowledge on human growth process, body composition, body shape and size, different types of growth maturations, analysis of growth events, lung capacity, lung related other variables, pre-caution of growth failure, heart failure and its causes, finding risk factor of heart failure, different parametric and non-parametric growth model and higher dimensional growth model to understand and predict biological variables and adult stature.

Objectives of the Course

After completing this course, the students should

- understand all the features of human growth, and morphometric;
- apply longitudinal data to fit human growth model in two or higher dimensions; and
- understand the diseases and disorder of growth process;

Learning Outcomes

At the end of the course, the students will be able to

- know how to develop growth chart;
- know how to use statistics in Medical Science;
- know how to assess growth status;
- diagnostic growth failure, heart failure and lung problems; and
- predict final stature of a child, and so on.

Course content

Morphometry: Basic concepts, stature, weight, BMI, vital capacity, strength, sitting-height, chest circumference, Bouchard's index. Livi's weight-height index. Body Mass Index (BMI), Rohrer's body build index, index of morphological equilibrium. Grid method and auxogram. Manouverier's indexes of body build. Cormic index. Pirguet's index of body build, Demeny's vital index, Speh's index, Bruugsch's chest-stature index, Pignet's coefficient of robusticity,

Craniometry and Osteometry: Planes of orientation, cranial landmarks, craniometric indices and cranial capacity. Osteometry, index of the body, foramen and Baudoin's sexual index vertebral index and Cunningham's index. Measurement of sacrum and long bones. Elliptical Fourier Analysis.

Growth and Maturation: Basic concepts, stages in child growth, early childhood, mid-childhood, late childhood, adolescence, prenatal and postnatal growth, Skeletal maturation, sexual maturation, age at menarche, body composition and nutritional status, sequence of adolescent events.

Lung Capacity: Introduction, Vital Capacity and its measurement, Tidal Volume, Inspiratory Vital Capacity, Expiratory Vital Capacity, Total Lung Capacity, Forced Expiratory Vital Capacity, Forced Expiratory Volume, and their relationship

Growth Failure: Introduction, synonyms and keywords, Symptoms, Causes, Exam and test, Treatment.

Heart Failure: Concept, Relation between Heart Rate and Age, Classes of Heart Failure, Warning sign of Heart Failure, Risk of Heart Failure, Symptom & Diagnosis, Prevention & Treatment, Advanced Heart Failure,

Growth Model: Biological variables and its secular trends, Algometry 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-1 and JPA-2 models, modified ICP model, BTT model and Kernel's (non-parametric) model, Wavelet model, polynomial model, growth variations due to genetics and nutrition. Twin growth, heritability of growth.

Higher Dimensional Growth Model: Extension of BTT model, Extension of other growth models.

Main Books:

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: Axiological Analysis of Longitudinal Measurements of Human Stature, Chicago: SSI

Books/Scientific Paper Recommended:

Shahin, M A, M. A Ali, ABM S. Ali (2013). An Extension of Generalized Triphasic Logistic Human Growth Model. J Biomet Biostat 4: 162. doi:10.4172/2155-6180.1000162

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

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

Fleiss, J.L.(1981) : Statistical Methods for Rates and Proportions, Second Edition, John Wiley and Sons. N.Y.

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

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

COURSE: M-STAT 515
Advanced Stochastic Modeling
Full Marks: 75

(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)

Number of Lectures: 45

Examination hours: 4

The aim of this course

The course is designed on advanced probability to cater to the needs of students and to the non-specialists. It would be suitable for research level courses in statistics. Uncertainties arises in several ways in different aspects of growth processes and also due to the involvement of unpredictable human behavior, a brief description of stochastic processes and stochastic differential equation in terms of which different models have been incorporated.

Objectives of the Course

After completion of the course, the students should

- Understand the self-contained modules of concepts and notations.
- Able to differentiate the deterministic and non-deterministic models.
- Understand all features of stochastic epidemic processes and queue processes.
- Develop maturity on stationary processes and time series.

Learning Outcomes

At the end of the course, the students will be able to

- Know when it is appropriate to use the properties of generating functions in different processes.
- Know how the renewal theory and arguments have often been advanced in a variety of situations, such as demography, manpower studies, reliability, replacement and maintenance.
- Create interest in the application of probability theory, concerned with objects or individuals that can generate objects of similar kinds, such as human beings, animals, genes, bacteria, and also neutrons which yield new neutrons under a nuclear chain.

Course content

Review of Stochastic Process and Markov Models: Definition, state space and parameters of stochastic process. Markov process, Markov Chain, Poisson Process, Birth and death process, Illness-Death Process, Branching process, Renewal Process and Queuing Process, Application of Markov Model and MCMC.

Stochastic Epidemic Process: The random variable technique and its application. Simple epidemic model, General epidemic model, Carrier borne epidemic model. Kermack and McKendrick's model, Daley and Kendall's Model. Stochastic Process of Clinical Drug Trials. Two armed bandit model, the Winner sampling model, the Optimum allocation model.

Models for Social and Occupational Mobility: Introduction, Models for social mobility, Models for occupational mobility.

Markov Models for Educational and Manpower Systems: A model for system with given Input. A model for an expanding system with given size.

Continuous Time Models for Stratified Social Systems: Some basic theory of Markov Processes. A manpower system with given Input. A Manpower system with given growth rate. Systems with given Input and loss rate depending on length of service. Hierarchical systems with given input and promotion rates depending on seniority. Fix and Neyman Model.

Stochastic Models of Reproductive Process: Dandekar's Modified binomial and Poisson model, Brass model, Singh's modified model, Model of waiting times of conception Sheps and Perrin model of reproductive process.

Stochastic Process in Genetics: Introduction, Physical basis of heredity. Genotypes under random mating, Herdy Weinberg law, Mating under various types of selection. Autosomal inheritance, Sex-linked inheritance, Change of gene frequencies, Homozygosity under random mating.

Stochastic Process in Queueing and Reliability: General concept, steady state and transient behavior of M/M/I models, Birth and death process, Multichannel models, Network of Markovian queueing system, GI/M/I and M/G (a,b)/I Models.

Main Books:

Bartholomew, D.J.(1973): Stochastic Models for Social Processes, 2nd ed. John Wiley and Sons.

Biswas, S.(2004): Applied Stochastic Process, New Central Book Agency (P) Ltd., Kolkata, India.

Medhi, J. (1994): Stochastic Process, Wiley Eastern Ltd., New Delhi, India.

Books Recommended:

Anderson, T.W. (1971): The Statistical Analysis of Time Series, Wiley, N.Y

Cryer, J. D. and K. Chan (2008): Time Series Analysis: with applications in R, 2nd Ed., Spinger, N.Y.

Elandt Johnson, R.C.(1971): Probability Models and Statistical Methods in Genetics, John Wiley, N.Y.

Lűtkepohl, H. (2005): New Introduction to multiple Time Series Analysis, Springer, N.Y.

Tan Wai-Yuan(1991: Stochastic Process of Carcionogenesis, Marcel Dekker, N.Y.

COURSE: M-STAT 516
Advanced Biostatistics
Full Marks: 75

(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)

Number of Lectures: 45

Examination hours: 4

Aims objectives and Applications:

This course mainly contains survival analysis, clinical trials and accelerated life testing (ALT) model.

Objective of survival analysis is (i) to estimate survival time for a group of patients, such as time until second heart-attack for a group of patients. (ii) to compare survival time between two or more groups, such as treated vs. placebo patients in a randomized controlled trail. (iii) to access the relationship of co-variates to survival time, such as: does weight, insulin resistance, or cholesterol influence survival time of patients?

Clinical trials (also called medical research and research studies) are used (i) to determine whether new drugs or treatments are both safe and effective. (ii) to compare a new treatment to a treatment that is already available.

Now a days, modern technology produce high reliability products. For such products, under operating stress(Temperature, Voltage, Load, Cycle, etc.) level, it takes a lot of time to get a sufficient number of failures to be used to estimate lifetime distributions which is useful to determine the products quality. Therefore, a sufficient number of failures obtained under high stress levels with small duration of times. These accelerated data are analyzed under accelerated life testing(ALT) environment (i) to estimate reliability for a specified time under x_0 (ii) to estimate median or other quantiles under x_0.

Therefore from the above aim and objectives it is clear that, the applications of this course are to analyze the medical research data and industrial products lifetimes data.

Course content

Basic Concept: Survival data. Different distributions of survival data and their characteristics. Survivor function, hazard function, parametric and nonparametric estimation of survival function.

Parametric Regression Models: Introduction. Inclusion of strata. Specifying a Distributions. Residuals. Residual analysis and other model checks. Predicted values. Fitting the Model. Exponential, Weibull, Normal, Lognormal and Gamma regression models. Interval estimation of the parameters and quantiles. Applications of parametric regression models.

Proportional Hazards Models: Introduction. Hypothesis test. Stratified and Penalized Cox models. Residual analysis. Partial likelihood. Applications of proportional hazards model. Estimation of hazard ratio. Comparison of two or more survival functions.

Extension of the Cox PH Model: Definition and examples of time-dependent variables, extended Cox model for time-dependent variables, hazard ratio for extended cox model, assessing time-dependent variables that do not satisfy PH assumption, extended cox likelihood. Application of the extended Cox model.

Accelerated life testing models: Accelerating variables, different types of life-stress relationships. Constant-stress and step-stress accelerated test models. Different methods for representing and analyzing accelerated life test (ALT) data based on different sampling schemes. Application of accelerated life testing models.

Multivariate lifetime models: Multivariate lifetime distributions and their characteristics, parametric and nonparametric estimation of multivariate lifetime distribution, models with multiple failure modes.

Clinical Trials: Basic concepts of clinical trials. Controlled and uncontrolled clinical trials, historical controls, protocol, placebo, randomization, blind and double blind trials, ethical issues, protocol deviations, volunteer bias. Simple comparative trials, Cross-over trials, size of trials, meta analysis, interim analysis, multi-centre trials, combining trials.

Main Books:

Bland, J.M.(1995): An Introduction to Medical Statistics, Second Edition, Oxford University Press.

Fleiss, J.L, Levin, B and Paik, M.C.(2003): Statistical Methods for Rates and Proportions, Third edition, John Wiley & Sons.

Kleinbum, D.G and Klein, M(2012): Survival Analysis; A self-learning Text, Third edition, Springer .

Lawless, J.F.(2003): Statistical Models and Methods for Lifetime Data, John Wiley and Sons, N.Y.
Nelson, W.(1990): Accelerated Testing: Statistical Models, Test Plans and Data Analyses, John Wiley and Sons, N.Y.

Books Recommended:

Bain, L.J. and Engelhardt, M. (1991): Statistical Analysis of Reliability and Life Testing Models, Theory and Methods, 2nd ed., Marcel Dekker, New York.
Balakrishnan, N.(Ed.)(1995): Recent Advances in Life – Testing and Reliability, CRC Press, Boca Raton, FL.
Johnson RCE & Johnson (1980): Survival Models and Data Analysis, Wiley NL & Sons, NY.

COURSE: M-STAT 517
Research Planning, Monitoring and Evaluation

Number of Lectures: 45

Full Marks: 75

(Examination 60, Tutorial/Terminal 11.25, and Attendance 3.75)

Examination hours: 4

Aim of the course

The aim of this course is to develop student's understanding of the nature of studies to monitor and evaluate intervention programs, using examples from Government and other related areas. Also this is to equip practitioners, project partners and external evaluators with the knowledge and expertise necessary to conduct project planning, monitoring, evaluation and reporting. There is a particular focus on the contribution of statistical methods in both the design and analysis of such studies.

Objectives of the course

This course will develop the students' ability to

- design of a broad strategy for the monitoring and evaluation of a specific government programs or policy;
- Aware of the appropriate analysis techniques to complement a chosen design strategy.
- Understand and be able to apply broad principles to guide the choice between alternative statistical designs that can accommodate the evaluation of an intervention.
- Understand and be able to articulate the important role monitoring plays, both as a policy tool in its own right and as an aid to evaluation.

Learning Outcomes

- Having successfully completed this course student will be able to:
- Produce a stakeholder analysis, a problem analysis and an objective analysis, using the Logical Framework Approach;
- Identify and develop "SMART" objectives and indicators, project risk factors and assumptions;
- develop a monitoring and evaluation plan;

- identify appropriate qualitative and quantitative data collection techniques;
- identify the elements of a monitoring and evaluation framework and system
- construct an evaluation report; and
- design a Terms of Reference for an evaluation.

Course Content

Introduction: Designs for program evaluation and principle. Research, data analysis techniques and reporting, Developing a M & E system, Developing a terms of reference and responding to proposals. Data sources and research methods, Performance monitoring, Theoretical framework, Statistical methods used in the analysis of data devised from such designs, in particular for the estimation and test of hypothesis.

Monitoring and Evaluation: Basic concept, Opportunities and Barriers, Purpose and benefits of planning, monitoring and evaluation. Introduction to the Logical Framework Approach (LFA), Steps of the LFA: Stakeholder analysis, Problem analysis, Objectives analysis, Alternatives analysis, Identification of indicators and means of verification, Identification of assumptions and risks, Setting baselines and targets, Theory of Change, How to produce a plan of action. Key concepts and approaches in evaluations (including effectiveness, efficiency, impact, relevance and sustainability). Development practitioners, project leaders and decision makers responsible for designing, implementing, monitoring or evaluating development projects. Collecting, Analyzing, and Using Monitoring Data, Key Elements of M & E Work Plan, Comprehensive Monitoring and Evaluation Framework,

Evaluation of Research Question: Outcome, Impact, Output, Input, Study Designs, Types and Objectives of Monitoring and Evaluation, Steps in Developing a Monitoring Plan, Country Monitoring and Evaluation Matrix. Planning, Monitoring and Evaluation for Development Projects.

Main Books

Planning, Monitoring and Evaluation Framework for Research Capacity Strengthening, 2016, ESSENCE Good practice document series, WHO.

M.Q. Patton, 2017, Managing for Sustainable Development Goals- An Intergrated approach to planning, Monitoring and Evaluation, Priciples-Focussed Evaluation: The Guide.

References

A step by step guide to Monitoring and Evaluation, 2014, Monitoring and Evaluation for Sustainable Communities by http://www.geog.ox.ac.uk/research/technologies/projects/ monitoringandevaluation.html

World Bank. Institutionalizing Impact Evaluation Within the Framework of a Monitoring and Evaluation System. 2009.

Rossi, P., M. Lipsey, and H. Freeman (RLF). A Systematic Approach. 7th edition 2004: Thousand Oaks, CA; Sage.

Handbook on Monitoring and Evaluating for Results, Evaluation Office, UNDP, 2002

Irwin Epstein and Tony Tripodi, 1977, Research Techniques for Program Planning, Monitoring, and Evaluation, Columbia University Press

COURSE: M-STAT 518

Viva-voce
Full Marks: 100

COURSE M-STAT 519

Practical: 8 Sessions (*For General Group)*

Full Marks: 200
(Practical examination140 marks and continuous assessment 60 marks)

Session-I : Advanced Statistical Inference

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 Scale parameter
6. Variance estimation by Jackknife, bootstrap and influence function.
7. Robust estimation of multivariate location and scatter matrix.
8. Data manipulation and various graphs
9. Estimation of SE of r and b_1 by non-parametric and parametric bootstrap and jackknifing
10. Different bootstrap CI of r and b_1
11. MPT
12. UMPT
13. SRT
14. MPSRT
15. UMPSRT
16. Asymptotic efficiency of the above tests
17. Sequential test
18. OC and ASN function
19. Construction of decoction regions
20. Tests of 2x2 contingency table.
21. ARE of Mann- Whitney test and sign test.
22. Kruskal-Wallis test
23. Square rank test for variances.
24. Friedman test
25. Permutation test and bootstrap test of equality of two means
26. Fitting regression models by bootstrapping
27. Different density estimators of univariate data

Session-II: Advanced 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 Youden design. Also analyze the data using such designs.
4. Estimate and test the data using intra-block and inter-block BIBD model.
5. Analyze the data using Split plot and Split-split plot design.

6. Analyze factorial experiment of levels 2, 3 and 5 with different factors. Construct plans of $\frac{1}{2^n}$, $\frac{1}{3^n}$ and $\frac{1}{5^n}$ replicate of such experiments 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 the 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.

Session-III: Advanced Multivariate Analysis

1. Fit the multivariate straight-line regression model and test the goodness of fit.
2. Apply CA and MCA for categorical data analysis.
3. Apply CA and MCA for mixed tables of metric and categorical data analysis.
4. Fit the multivariate normal and t mixture models by estimating the parameters using EM algorithm.
5. Find the clusters from a given datasets.
6. Compute Bayesian prediction using multivariate regression model.
7. Perform factor analysis of multivariate data using Bayesian approach. Recover source signals (original signals) from the mixed signals using Bayesian approach. Mixture of artificial signals, natural image and audio signals with Gaussian noise can be considered as the mixture dataset

Session – IV Time Series Analysis and Forecasting

1. **Single time series:** Checking Time series properties, Tests for order of integration. Spectral representation, Fitting an appropriate ARIMA model. Diagnostic and Stability, Checking and Forecasting.
2. **Multivariate time series:** Tests for cointegration and spurious correlation. Fitting different VAR and VEC models. Tests for Granger causality, Estimation of impulse responses functions. Decompositionof forecast error variance.
3. **Stylized facts of time series:** Test for ARCH effect, Modeling and forecasting volatility. Detection of structural breaks.

Session – V: Data Mining

1. Exploring Categorical and Numerical Variables.
2. k-nearest neighbor algorithm for estimation and prediction.
3. Classifying Credit risk using decision trees.
4. Neural network model for classification.
5. Clustering by Kohonen Networks.

Session – VI: Genomics and Bioinformatics

1. DNA sequence data Analysis using BLAST / PYTHON / PERL.
2. Phylogenetic Analysis of DNA sequence data using R/BLAST /PYTHON/ PERL..
3. QTL data analysis using R
4. Gene-expression data analysis using R.
5. GWAS/SNPs data analysis using R.

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-Sekar and Deming method
4. Estimation of nuptiality parameters of nuptiality model
5. Estimation of Cole's indices
6. Gradution of net maternity function by Normal, Wicksell and Hadwiger curve
7. Problems related to population projection
8. Problems related to fecundability
9. Estimation of proximate determinants of fertility and inhibiting effects of marriage and divorce

Environmental Statistics

Experiments based on Statistical methods and computer applications

1. Use of computer program R, Mathlab for analysis of environmental data.
2. Regression analysis, trend analysis, error analysis and application of statistical tests in environmental problems.
3. Fitting of polynomials to environmental data.
4. Use of Chi-square, F-test and t- test.
5. Principle component analysis of environmental variable.
6. Determine drought index and SPI by using Markov chain model using R.
7. Study of non-linear and non-stochastic time series analysis of environment data by using Matlab.
8. Study of Wavelets analysis and Spectral analysis of environment data by using Matlab.
9. Study of geo-statistical modeling analysis, spatio-temporal modeling, Extreme value modeling by using GIS.

Advanced Actuarial Statistics

1. Survival distributions and life tables.
2. Life annuities, Net premium, Premium series.
3. Multiple life functions.
4. Multiple decrement models.
5. Pension plans, the expense function and dividends.
6. Measurement of risk and Mortality Table: Construction of mortality tables.
7. Premium calculation of various life policies.
8. Techniques of calculating exposures form individual records including consideration involving selection of studies, various observation periods and various methods of tabulating deaths.
9. Techniques of calculating exposures from variation schedules including the general concepts of fiscal year.

10. The use of interim schedules and variations in observation period or method of grouping deaths and construction of actuarial tables.

Session – VII
Proteomics and Biomedical Informatics

1. Protein sequencing analysis using BLAST / PYTHON / PERL.
2. Protein classification using supervised and unsupervised statistical algorithms.
3. Phylogenetic Analysis of Protein Sequenceing using R.
4. Metabolomics data analysis using R/MATLAB.
5. Bayesian network (BN) analysis, Nomogram construction and disease prediction.

Health and Epidemiology

1. To fiend the growth over time by using algometric growth and other suitable models.
2. To detect the factors which directly and indirectly influence on health by using statistical tools.
3. Estimation of relative risk and odds ratio.
4. To solve the care management problem using by linear programming methods (graphical and simplex)
5. Statistical models to control extraneous factors : Linear regression, Log-linear regression, Logistic regression, Poisson regression and Negative Binomial regression
6. Epidemiological and health hazards models.
7. Using output of any appropriate statistical analysis (Using any statistical package/software) interpretation of causal variables and confounder variables, and cases of the study.
8. Calculation of OR, RR, NNT, PPR, NPR, sensitivity, specificity, true positive rate, true negative rate, ROC with 95% CI using raw data and their interpretation
9. Calculation of sample size with respect to various problem (regarding power and precision)
10. Fitting and interpretation of linear model, logistic model, multinomial model.
11. Analysis of repeated measure.
12. Analyzing parametric models and cox proportional hazards model.

Statistical Methods in Industrial Management

1. Economic order quantity
2. Order Cycle and order size
3. Optimal quantity
4. Optimal order quantity
5. Total annual cost
6. Quantity discounts
7. Constructing Networks
8. Draw a precedence diagram using AOA
9. Critical path
10. Determine an optimum crashing plan
11. Deterministic time estimates
12. Probabilistic time estimates
13. Compute slak times

14. Gantt chart

Statistical Methods for Reliability Data

1. Distribution ID plot and overview plot.
2. Parametric and nonparametric distribution analysis. Life data event plot, probability plots and parametric ML fit.
3. Selection of a suitable model from a set of competitive models for reliability data.
4. Multiple failure mode data analysis: individual and combined modes.
5. Accelerated life test models analysis.
6. Parametric regression model fit with censored product reliability data.
7. Recurrence (point process) data analysis.
8. Warranty claims data and warranty cost modeling.
9. Degradation data modeling and analysis.
10. Case studies with real product reliability data.

Session – VIII
Physical Health and Human Growth Modeling

1. Morphological indices
2. Differential indices of body building
3. Craniometric indices, detection of Cranial landmarks.
4. Extraction of growth parameters
5. Fitting of growth models: Models described in M-Stat 508

Advanced Stochastic Modeling

1. Study of simple epidemic models.
2. Analysis of Kermack and McKendrick'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.

Advanced Biostatistics

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.

Research, Planning, Monitoring and Evaluation

Assigned by course tutor

COURSE: M-STAT 520

Written Thesis

Full Marks: 100

COURSE: M-STAT 521

In-plant Training

Full Marks: 50

To complete in-plant training, students must have to submit a report on a specific topic designed by the examination committee.

COURSE: M-STAT 522

Thesis Defense

Full Marks: 50