MAHATMA GANDHI UNIVERSITY

PRIYADARSHINI HILLS, KOTTAYAM
KOTTAYAM – 686560
www.mgu.ac.in

RESTRUCTURED REGULATIONS, SCHEME AND SYLLABUS

UNDER CREDIT AND SEMESTER SYSTEM

FOR

POST-GRADUATE PROGRAM

IN

M.Sc. BIOSTATISTICS

2012

MAHATMA GANDHI UNIVERSITY, KOTTAYAM

REGULATIONS FOR CREDIT AND SEMESTER SYSTEM

Duration

The duration of PG program shall be 4 semesters. The duration of each semester shall be 90 working days. Odd semesters extend from June to October and even semesters from December to April. There will be one month semester breaks each in November and May. A student may be permitted to complete the program, on valid reasons, with in a period of 8 continuous semesters from the date of commencement of the first semester of the programs.

Program Structure

The program includes two types of courses namely Program Core courses and Program Elective Courses. In the fourth semester the colleges can choose 4 electives that will suit the needs of students there, from the electives specified in the syllabus. There shall also be a Program Project with dissertation to be undertaken by all students. Every Program conducted under Credit Semester System shall be monitored by the College Council.

Viva Voce

Comprehensive Viva-voce shall be conducted at the end of the fourth semester of the program and it shall cover questions from all courses in the program.

Project work

Project work shall be completed by working outside the regular teaching hours under the supervision of a teacher in the concerned department. There should be an internal assessment and external assessment for the project work. The external evaluation of the Project work is followed by presentation of work including dissertation and Viva-Voce.

Examinations

There shall be University examination at the end of each semester. Project evaluation and Viva -Voce shall be conducted at the end of the program only. Project evaluation and Viva-Voce shall be conducted by two external examiners and one internal examiner.

There shall be one end-semester examination of 3 hours duration in each lecture based course and practical course. The examinations for which computers are essential should be conducted in the computer lab supervised by an external examiner appointed by the university.

Evaluation and Grading

Evaluation: The evaluation scheme for each course shall contain two parts; (a) internal evaluation and (b) external evaluation. 25% weightage shall be given to internal evaluation and the remaining 75% to external evaluation and the ratio and weightage between internal and external is 1:3. Both internal and external evaluation shall be carried out using direct grading system.

Internal evaluation: The internal evaluation shall be based on predetermined transparent system involving periodic written tests, assignments, seminars and attendance in respect of theory courses and based on written tests, lab skill/records/viva and attendance in respect of practical courses. The weightages assigned to various components for internal evaluation are as follows.

Components of Internal Evaluation

Component	**Weightage**
i) Assignment	1
ii) Seminar	1
iii) Attendance	1
iv) Two Test papers	2

Letter Grade	Performance	Grade Point(G)	Grade Range
A	Excellent	4	3.5 to 4.00
B	Very Good	3	2.5 to 3.49
C	Good	2	1.5 to 2.49
D	Average	1	0.5 to 1.49
E	Poor	0	0.0 to 0.49

Grades for Attendance

% of attendance	Grade
>90%	A
Between 85 and 90	B
Between 80 and below 85	C
Between 75 and below 80	D
< 75	E

To ensure transparency of the evaluation process, the internal assessment grade awarded to the students in each course in a semester shall be published on the notice board at least one week before the commencement of external examination. There shall not be any chance for improvement for internal grade.

A separate minimum of C Grade for internal and external are required for a pass for a course. For a pass in a program a separate minimum grade C is required for all the courses and must score a minimum CGPA of 1.50 or an overall grade of C and above. Each course is evaluated by assigning a letter grade (A, B, C, D or E) to that course by the method of direct grading. The internal (weightage =1) and external (weightage =3)

components of a course are separately graded and then combined to get the grade of the course after taking into account their weightages.

A student who fails to secure a minimum grade for a pass in a course will be permitted to write the examination along with the next batch. There will be no supplementary examination.

Assignments:

Every student shall submit at least one assignment as an internal component for every course. The Topic for the assignment shall be allotted within the 6th week of instruction.

Seminar Lectures

Every PG student shall deliver at least one seminar lecture as an internal component for every course. The seminar lecture is expected to train the students in self-study, collection of relevant matter from the books and Internet resources, editing, document writing, typing and presentation.

Class Tests

Every student shall undergo at least two class tests as an internal component for every course. The weighted average shall be taken for awarding the grade for class tests.

Attendance

The attendance of students for each course shall be another component of internal assessment. The minimum requirement of aggregate attendance during a semester for appearing the end semester examination shall be 75%. Condonation of shortage of attendance to a maximum of 10 days in a semester subject to a maximum of two times during the whole period of post graduate program may be granted by the University.

If a student represents his/her institution, University, State or Nation in Sports, NCC, NSS or Cultural or any other officially sponsored activities such as college union / university union activities, he/she shall be eligible to claim the attendance for the actual number of days participated subject to a maximum of 10 days in a Semester based on the specific recommendations of the Head of the Department and Principal of the College concerned.

A student who does not satisfy the requirements of attendance shall not be permitted to take the end Semester examinations.

Mahatma Gandhi University, Kottayam

M.Sc. BIOSTATISTICS: SCHEME AND SYLLABUS 2012
(UNDER CREDIT AND SEMESTER SYSTEM)

Semester I (Total credits-20)

Course Code	Course Title	Credits	Teaching Hours
BSTA 101	Descriptive Statistics, Probability and Distributions	4	5
BSTA 102	Sample Survey Methods	4	5
BSTA 103	Computer Programming in C++ and SAS	4	5
BSTA 104	Statistical Genetics and Ecology	4	5
BSTA 105	Statistical Data Analysis Using Microsoft Excel and BASE SAS	4	5

Semester II (Total credits-20)

Course Code	Course Title	Credits	Teaching Hours
BSTA 201	Linear Algebra, Regression Techniques and Bioassays	4	5
BSTA 202	Statistical Estimation: Theory and Practice	4	5
BSTA 203	Basic Epidemiology and Vital Statistics	4	5
BSTA 204	Statistical Testing of Hypotheses	4	5
BSTA 205	Statistical Data Analysis using SPSS and BASE SAS.	4	5

Each student has to carry out a mini project work during the second semester vocation under the guidance of a reputed doctor or scientist in a hospital or a research institute and submit the project report to the course coordinator.

Semester III (Total credits-20)

Course Code	Course Title	Credits	Teaching Hours
BSTA 301	Design of Experiments and Quality Control	4	5
BSTA 302	Stochastic Modeling and Time Series Analysis	4	5
BSTA 303	Multivariate Statistical Methods	4	5
BSTA 304	Advanced Epidemiology	4	5
BSTA 305	Statistical Computing and Data Analysis using ADVANCED SAS and SPSS	4	5

Semester IV (Total credits-20)

Course Code	Course Title	Credits	Teaching Hours
BSTA 401	R Programming, Bayesian Inference and MCMC Methods	3	5
BSTA 402	Elective 1- Survival Analysis and Demography	3	5
BSTA 403	Elective 2- Controlled Clinical Trials and Operations Research	3	5
BSTA 404	Elective 3- Bioinformatics and Computational Biology	3	5
BSTA 405	Elective 4- Statistical Computing and Data Analysis using ADVANCED SAS and R	3	5
BSTA 406	Project Work and on the job training (in a reputed industry/Research Institute)	3	
BSTA 407	Viva-voce	2	

SYLLABI OF COURSES OFFERED IN SEMESTER - 1

BSTA 101 DESCRIPTIVE STATISTICS, PROBABILITY AND DISTRIBUTIONS

UNIT 1. Elementary concepts in Statistics: Concepts of statistical population and sample from a population; qualitative and quantitative data; nominal, ordinal, ratio, interval data; cross sectional and time series data; discrete and continuous data. Collection and scrutiny of data**:** Primary data; designing a questionnaire and a schedule; secondary data and sources of secondary data. Presentation of data: Diagrammatic and graphical representation of data; frequency distributions and cumulative frequency distributions; histogram, frequency polygon, stem and leaf chart and ogives. Descriptive statistics: Concepts of central tendency or location, Absolute and relative measures of dispersion; Box plot, Lorenz curve; skewness and kurtosis.

UNIT 2. Probability: Random Experiment; sample point; sample space; events; mutually exclusive and exhaustive events; frequency and classical definitions of probability. Axiomatic definition of probability; addition and multiplication theorems; conditional probability and independence; Bayes' theorem. (The main thrust is on numerical problems and applications), Discrete and continuous random variables; probability density functions and distribution functions; expectation of a random variable.

UNIT 3. Standard Univariate Distributions: Standard univariate discrete and continuous distributions- uniform; binomial; Poisson; geometric; negative binomial and hyper-geometric distributions. Uniform; exponential; normal; Laplace, gamma, beta, log-normal, logistic and Weibull distributions.(elementary properties and applications only)

UNIT 4. Sampling Distributions, Law of large numbers and Central Limit Theorem: Concepts of random sample and statistic; distribution of sample mean from a normal population; chi-square distribution; F and t statistics, distributions (no derivations) and their applications. Chi-square test for goodness of fit, Central Limit Theorem for i.i.d case (statement and examples only). Evaluation of probabilities from the binomial and Poisson distributions using central limit theorem. Chebychev's inequality and weak law of large numbers (statement and applications only).

References

Dutta, N. K. (2004). Fundamentals of Biostatistics, Kanishka Publishers.

Gurumani N. (2005) . An Introduction to Biostatistics, MJP Publishers.

Daniel, W. W. (2007). Biostatistics- A Foundation for Analysis in the Health Sciences,Wiley.

Rao, K. V. (2007). Biostatistics - A Manual of Statistical Methods for use in Health Nutrition and Anthropology.

Pagano, M.& Gauvreau, K. (2007). Principles of Biostatistics.

Rohatgi, V.K.& Saleh, A.K.Md. (2001). An Introduction to Probability and Statistics, John Wiley & Sons.

Sundaram, K.R.(2010) Medical Statistics-Principles & Methods, BI Publications,New Delhi

BSTA 102 SAMPLE SURVEY METHODS

UNIT 1. Concepts of population and sample, need for sampling, census and sample surveys, sampling and non-sampling errors, sample size determination, finite population sampling techniques-SRSWR, SRSWOR, estimation of mean or total in each case and their variances.

.

UNIT 2. Systematic sampling, stratified sampling, allocation problems in stratified sampling, estimation of mean or total in each case and their variances. Examples from health sciences.

UNIT 3. Ratio and regression estimators based on SRSWOR method of sampling, Auxiliary information in sample surveys; Randomized response technique: Warner's model-related and unrelated questionnaire methods. Cluster and Multistage Sampling: double sampling, two stage and multi-stage sampling, cluster sampling. Examples based on biostatistical experiments.

UNIT 4. Unequal probability sampling and Non-probability sampling: PPS WR/ WOR methods and related estimators of a finite population mean, Horvitz- Thompson estimator of a finite population total/mean and expressions for variance and its unbiased estimator, convenient sampling, non-random sampling.

References

Cochran, W.G. (2002). Sampling Techniques. Wiley
Des Raj and Chandhok (1998). Sampling Theory, Narosa.
Murthy, M.N. (1967). Sampling Theory and Methods. Statistical Publishing Company, Calcutta.
Sampath S.(2005). Sampling Theory and Methods.

BSTA 103 COMPUTER PROGRAMMING IN C++ AND SAS

Unit 1: Fundamentals, Operating Systems- Windows, Linux, Internet, algorithms, Flow charts, data types and variables, Operators, Input Output statements, Control statements:- if, if-else, nested if-else, goto and switch statements.

Unit 2: Loops: for, while, do…while loops. Break, continue, exit ( ), Library functions. One dimensional, two-dimensional and multi-dimensional arrays. Functions, definition and declaration, Illustrative examples from statistics. Pointers and references.

Unit 3: Introduction to SAS: SAS variables, Libraries ,Windows, Parts of a SAS program, Data sets-Creation, Reading data from an external file, Data step statements like CARDS, DATA, Assignment, Do-loops, DROP, KEEP, INPUT, OUTPUT, SET, STOP, IF-THEN-ELSE, SAS Operators , Functions , Arrays.

Unit 4: Procedures in SAS, Proc statements, CLASS, BY, DROP, FREQ, KEEP, OUTPUT, LABEL etc. Procedures-PRINT, FREQ, MEANS, UNIVARIATE, SORT, CONTENTS, CORR, PLOT, REG, ANOVA, LOGISTIC, IMPORT, TABULATE.

References

Balaguruswamy E. (1997). Object-Orinted Programming with C++, Tata McGraw-Hill Publishing Company Ltd.

Der, G. and Everitt, B.S.(2006). A Handbook of Statistical Analysis Using SAS, CRC Press.

Der, G. and Everitt, B.S.(2006). Statistical Analysis of Medical Data Using SAS, CRC Press.

Littell R.C., Stroup W.W. & Freud R.J. (2002). SAS For Linear Models, SAS Institute Inc.

Lora, D. and Susan, S.(2009)The Little SAS, support.sas.com

BSTA 104 STATISTICAL GENETICS AND ECOLOGY

Unit 1: Basic biological concepts in genetics, Mendel's law, Hardy Weinberg equilibrium, estimation of allele frequency (dominant/co-dominant cases), Approach to equilibrium for X-linked gene. The law of natural selection, mutation, genetic drift.

Unit 2: Non-random mating ,inbreeding, phenotypic assortative mating. I ,T,O matrices, identity by descent. Family data-estimation of segregation ratio under ascertainment bias, pedigree data : Elston - Stewart algorithm for calculation of likelihoods. Linkage, estimation of re-combination fraction, inheritance of quantitative traits.

Unit 3: Introduction to ecology and evolution, population dynamics: single species-Exponenetial, Logistic and Gompertz models, Leslie matrix model for age and stage Structured population, survivorship curves-Constant, monotone and bath tub shaped hazard rates.

Unit 4: Two species: Lotka-Volterra equations, isoclines. Abundance estimation: Capture –recapture, Nearest Neighbor, line transect sampling, indirect methods. Ecological Diversity: Species abundance curve, indices of diversity (Simpson's index, Shannon-Wiener index). Game theory in ecology - Evolutionarily stable strategy, its properties, simple games such as Hawk-Dove game, Prisoner's dilemma, etc. Preservation of ecology and biodiversity.

References

Anil Gore & Sharayu Paranjpe (2001). A Course in Mathematical And Statistical Ecology,Kluwer academic Publishers.
Gardner E.J. & Snustad D.P. Principles of Genetics, John Wiley & Sons Inc.
Lange, K (2002). Mathematical and Statistical Methods for Genetic Analysis, Springer.

BSTA 105 BASIC STATISTICAL COMPUTING

This paper includes practical problems using data from Biostatistical contexts based on papers BSTA101-104. There will be 8 questions (2 from each course) of which 5 are to be answered. Data Analysis using Microsoft Excel & SAS is expected.

SYLLABI OF COURSES OFFERED IN SEMESTER II

BSTA 201 LINEAR ALGEBRA, REGRESSION TECHNIQUES AND BIOASSAYS

UNIT 1. Linear Algebra: Set operations, vectors and matrices, matrix operations, determinants, inverse of a square matrix; linear independence, rank of a matrix, generalized inverse and applications, linear equations, characteristic roots and vectors, quadratic forms and nature of definiteness.

UNIT 2. Analysis of Bivariate data: Scatter diagram, Principle of least squares; Karl Pearson's correlation coefficient; coefficient of determination; correlation ratio; rank correlation; partial and multiple correlations, Linear regression, Simple linear regression, multiple regression, fit of polynomials and use of orthogonal polynomials. Residuals and their plots, tests for departure from assumptions such as fitness of the model, normality, homogeneity of variances, detection of outliers and remedies, influential observations, power transformations for dependent and independent variables.

UNIT 3. **Generalized linear models**, analysis of binary and grouped data by using logistic models, large sample tests about parameters, goodness of fit, analysis of deviance, variable selection, introduction to Poisson regression, log-linear models, Random and mixed effect models, Nonparametric regression and generalized linear models.

UNIT 4. Bioassays: Types of biological assays, direct assays, ratio estimators, asymptotic distributions, regression approaches for estimating dose response relationships. Quantal responses, methods of estimation of parameters, dose allocation schemes, median dose, estimation of points on the quantal response function, Estimation of safe doses.

References

Draper, N.R. and Smith, H (2003). Applied Regression Analysis, John Wiley & Sons.
Rossi R.J.(2010).Applied Biostatistics for Health Sciences, Wiley.

BSTA 202 STATISTICAL ESTIMATION: THEORY AND PRACTICE

UNIT 1. Basic concepts and properties of estimators: Parametric models, parameters, random sample and its likelihood, statistic and its sampling distribution, problem of inference, estimator and estimate, mean square error (MSE), properties of estimators-unbiasedness, consistency, efficiency.

UNIT 2. **Sufficiency,** Cramer-Rao lower bound, Minimum variance unbiased estimator, relative efficiency of an estimator. Methods of estimation**:** Fisher information, complete and sufficient statistic, Rao-Blackwell theorem, UMVUE.

UNIT 3. **Methods of Estimation**: Method of moments, method of MLE, properties of MLE (statements only), method of minimum chi-square, linkage estimation (Examples from Genetics).

UNIT 4. Interval estimation: concepts of confidence interval, confidence coefficient, confidence interval for the parameters of univariate normal, proportion, mean, difference of means. Small sample and large sample confidence intervals. Large sample confidence intervals for binomial and Poisson parameters, bootstrap methods.

References

Hogg R.V. and Tanis E.A.(2001). Probability and Statistical Inference, Prentice Hall International Inc.
Kale, B.K. (1999). A first Course on Parametric Inference, Narosa Publishing House.
Manly, B. F. (2007). Randomization, Bootstrap and Monte Carlo methods in Biology, Chapman & Hall / CRC.
Rohatgi, V.K. and Saleh, A.K.Md.(2001). An Introduction to Probability and Statistics, John Wiley & Sons.

BSTA 203 BASIC EPIDEMIOLOGY AND VITAL STATISTICS

UNIT 1. Basic concepts & Measures of exposure and outcome: What is epidemiology? History of Epidemiology, Emergence of modern epidemiology, Measures of Exposures, Types of exposures, Sources of exposures, Measures of outcome, disease registries, Classification of diseases. Measures of disease frequency: Prevalence, Incidence, Risk, Odds of disease, Incidence time, Incidence rate, Relationship between prevalence, rate and risk, Routine data to measure disease occurrence, age standardization, direct method of Standardization, indirect method of standardization, cumulative rate, cumulative risk, proportional incidence.

UNIT 2. Overview of study designs: Type of study design, Intervention studies, Cohort studies, case-control studies, cross-sectional studies, ecological studies, Measures of exposure effect, relative and absolute measures of effect, Confidence intervals and significance tests for measures of occurrence and effect.

UNIT 3. Validity and reliability of measures of exposure and outcome: Sensitivity, Specificity, predictive value method for selecting a positivity criterion, receiver operator characteristic (ROC) curve, Intra and Inter-observer reliability, Kappa measure of agreement.

UNIT 4. Coverage and content errors in demographic data, use of balancing equations and Chandrasekharan-Deming formula to check completeness of registration data, adjustment of age data- use of Whipple, Meyer and UN indices. Population composition, dependency ratio. Stable and quasi-stable populations: Population-stable, stationary, 5 stages, inter census, post census; population projection, migration; factors affecting population-internal and international; age pyramid, age structure intrinsic growth rate. Models for population growth and their fitting to population data. Stochastic models for population growth.

References

Ahrens W. and Pigcot I.(2005). Handbook of Epidemiology, Springer.

Penny Web ,Chiris Bain & Sandi Pirozzo (2005).Essential Epidemiology-An introduction for students & Health Professionals,Cambridge University Press.

Rao, K.V.(2007). Biostatistics: A Manual of Statistical Methods for use in Health Nutrition and Anthropology, Raven publishers.

Rothman K.I and Greenland S (1998). Modem Epidemiology, Second edition, Lippincott

Pressat R. & Atherton A. (1972). Demographic Analysis.

Preston S.H.,Heuveline P. & Guillot M. Demography-Measuring and Modelling Population Processes.

Sundaram, K.R.(2010) Medical Statistics-Principles & Methods, BI Publications,New Delhi

BSTA 204 STATISTICAL TESTING OF HYPOTHESES

UNIT 1. Testing of hypothesis: Basic concepts, simple and composite hypotheses, two types of errors, critical region, significance level, size and power of the test, p-value and its interpretation. Neymann-Pearson Lemma (Statement only) and its application in testing of hypothesis.

UNIT 2. Testing Composite Hypotheses, Uniformly most powerful test, Likelihood ratio test, Test for mean, variance, equality of means, equality of variances, large sample and small sample tests.,

UNIT 3. Basics of sequential testing, Wald's SPRT with illustrations, OC and ASN functions for tests regarding binomial and normal populations. Linkage estimation and testing. Partitioning of Chi-square.

UNIT 4. Non parametric tests: Sign test, Quantile Test, Wilcoxon signed rank, Mann Whitney U test, runs test, median test, Chi-square test for independence of attributes, homogeneity, goodness of fit, Kolmogorov-Smirnov one sample and two sample tests, Freidman's test, Kruskal Wallis test.

References

Hogg R.V. and Tanis E.A.(2001). Probability and Statistical Inference, Prentice Hall International Inc.
Kale, B.K. (1999). A first Course on Parametric Inference, Narosa Publishing House.
Rohatgi, V.K. and Saleh, A.K.Md.(2001). An Introduction to Probability and Statistics, John Wiley & Sons.
Sundaram, K.R.(2010) Medical Statistics - Principles & Methods, BI Publications, New Delhi

BSTA 205 BASIC STATISTICAL COMPUTING

This paper includes practical problems from papers BSTA 201-204. There will be 8 questions (2 from each course) of which 5 are to be answered. Data Analysis using Excel SPSS, MINITAB, or SAS is expected.

SYLLABI OF COURSES OFFERED IN SEMESTER III

BSTA 301 DESIGN OF EXPERIMENTS AND QUALITY CONTROL

UNIT 1. Introduction to design of experiments: estimable linear parametric functions and their estimation, Gauss-Markov Theorem (meaning and statement only), testing of linear hypotheses, Basic principles of experimental design, uniformity trails, analysis of variance, CRD, RBD, LSD (equal and unequal number of observations, missing observations).

UNIT 2. Incomplete block designs, Balanced incomplete block designs (BIBD), group testing, PBIBD, hierarchical and nested designs. Split plot experiments, Analysis of Covariance.

UNIT 3. General factorial experiments, factorial effects, 2^n and 3^n factorial experiments in randomized block, Yate's method, complete and partial confounding, simple problems

UNIT 4. Quality and related concepts, ISO Certification, six-sigma, Statistical process control, theory of control charts, Shewhart control charts for variables - $\bar{x}$, R, sigma charts, attribute control charts - p, np, c charts, modified control charts.

References

Angela Dean & Daniel Voss (2006). Design and Analysis of Experiments, Springer Verlag

Campbell M.J, Machin D. & Walters S.J (2007). Medical Statistics – A Text Book for the Health Sciences, Wiley.

Cochran & Cox (2000). Experimental Designs, Wiley Asia

Das M.N. & Giri N.C. (2006). Design and Analysis of Experiments, New Age Publications

Montgomery, D.C. (2001). Design and Analysis of Experiments, Wiley.

Montgomery D. C. (2005) Introduction to Statistical Quality control, 5th edition, Wiley.

BSTA 302 STOCHASTIC MODELLING AND TIME SERIES ANALYSIS

UNIT 1. Introduction to stochastic processes (sp's); classification of sp's according to state space and time domain. Countable state Markov chains (MC's), Chapman-Kolmogorov equations; calculation of n-step transition probability and its limit. Stationary distribution, classification of states; first –passage time problems.

UNIT 2. Stationary process; weakly stationary and strongly stationary processes; Discrete state space continuous time MC: Kolmogorov differential equations; Poisson processes and properties, birth and death process, Yule process.

UNIT 3. Renewal theory: Elementary renewal theorem and applications. Statement and uses of key renewal theorem; Branching process: Galton –Watson branching process, pgf relations, probability of ultimate extinction, distribution of population size.

UNIT 4. Box Jenkins Models, Moving average processes, auto regressive processes, ARIMA models, Auto correlation function and correlogram, diagnostic checks, modeling and prediction, non-Gaussian models, applications in biostatistical contexts.

References

Basu A.K. (2003).Introduction to Stochastic Processes, Narosa Publishing House.
Feller, W. (1968): Introduction to Probability and its Applications, Vol.1, Wiley Eastern.
Medhi, J, (1982): Stochastic Processes, Wiley Eastern.
Suddhendu Biswas (1995). Applied Stochastic Processes: A Biostatistical and Population oriented Approach, Wiley Eastern.
Bhat B.R. (2008) Stochastic Models: Analysis and Applications, New Age Publshers
Karlin, S. and Taylor,H.M.(1998) An Introduction to Stochastic Modelling, Edition 3, Academic Press

BSTA 303 MULTIVARIATE STATISTICAL METHODS

UNIT 1. Multivariate data, multivariate normal distribution, random sampling from a multivariate normal distribution. Maximum likelihood estimators of parameters. Distribution of sample mean vector.

UNIT 2. Hotelling's T^2 and Mahalanobis D^2 statistics, applications in tests on mean vector for one and more multivariate normal populations and also on equality of the components of a mean vector in a multivariate normal population. Wishart distribution and applications.

UNIT 3. Likelihood ratio test criterion. Multivariate Analysis of Variance [MANOVA] of one-and two-way classified data. Dimension reduction, Principal components– estimation and computation, canonical correlations and applications.

UNIT 4. Classification and discrimination procedures for discrimination between two multivariate normal populations – sample discriminant function, tests associated with discriminant functions, classification into more than two multivariate normal populations, Cluster analysis, Hierarchical and agglomerative methods

References

Anderson, T.W. (1983): An Introduction to multivariate statistical analysis. 2nd Ed. Wiley.

Martin Bilodeau, David Brenner (1999).Theory of multivariate statistics, Springer.

Bhuyan K.C. (2005). Multivariate Analysis and its Applications,

Johnson, R. and Wychern (1992): Applied multivariate Statistical analysis, Prentice – Hall, 3rd Ed.

Morrison, D.F. (1976): Multivariate statistical methods. 2nd.Ed. McGraw Hill.

Rencher, A.C.(1998). Multivariate Statistical Inference with Applications, Springer.

Seber, G.A. F. (2001): Multivariate observations. Wiley.

BSTA 304 ADVANCED EPIDEMIOLOGY

UNIT 1. Confounding: Confounding, Assessment of confounding, Mantel-Haenszel summary measures of effect, Interaction, Mantel-Haenszel method to adjust for several confounders, Confidence intervals and statistical tests for adjusted relative measures of effect.

UNIT 2. Case-control studies: Definition of cases, and controls, methods of selecting cases and controls, matching, sample size, power calculations, basic methods of analysis of grouped data, basic methods of analysis of matched data. Logistic regression for case-control studies, estimation and interpretation of logistic parameters, matched analysis- estimation of logistic parameters, unmatched analysis of matched data, confounder score, Categorical data analysis.

UNIT 3. Cohort studies: Prospective cohort studies: planning and execution, retrospective cohort, nested case-control, case-cohort studies: planning and execution, matching and efficiency in cohort studies, cohort studies –statistical analysis.**Longitudinal studies:** Design, execution and analysis of longitudinal studies, repeated measurement analysis**.**

UNIT 4. Bias, attributable risk and causation in Epidemiology: Sources of bias**,** Selection bias, measurement bias, misclassification of exposure and outcome, Differential and non-differential exposure and outcome classification, Excess risk and Attributable risk, Causation and Hill's criteria.

References

Ahrens W. and Pigcot I.(2005). Handbook of Epidemiology, Springer.
Penny Web ,Chiris Bain & Sandi Pirozzo (2005).Essential Epidemiology-An introduction for students & Health Professionals, Cambridge University Press.
Rao, K.V.(2007). Biostatistics: A Manual of Statistical Methods for use in Health Nutrition and Anthropology. Raven publishers.
Rothman K.I and Greenland S (1998). Modem Epidemiology, Second edition, Lippincott

BSTA 305 ADVANCED STATISTICAL COMPUTING

This paper includes practical problems from papers BSTA 301-304. There will be 8 questions (2 from each course) of which 5 are to be answered. Data Analysis using SAS and SPSS is expected.

SYLLABI OF COURSES OFFERED IN SEMESTER IV

BSTA 401 R PROGRAMMING, BAYESIAN INFERENCE AND MCMC METHODS

UNIT 1. Elements of R Programming, Introduction to statistical software R, Data objects in R, Manipulating vectors, matrices, lists, importing of files, data frame, and computations of descriptive statistics measures. R-Graphics- Histogram, Box-plot, Stem and leaf plot, Scatter plot, Plot options; Multiple plots in a single graphic window, frequency table, Plotting of probability distributions and sampling distributions.

UNIT 2. Elements of Bayesian approach, Subjective interpretation of probability. Evaluation of subjective probability of an event, prior distribution of a parameter. Bayes' theorem and computation of the posterior distribution. Natural Conjugate family of priors for a model. Conjugate families for (i) exponential family models, (ii) models admitting sufficient statistics of fixed dimension.

UNIT 3. Bayesian Inference: Bayes estimators for (i) absolute error loss (ii) squared error loss (iii) 0-1 loss. Generalization to convex loss functions. Evaluation of the estimate in terms of the posterior risk. Bayesian interval estimation : Credible intervals. Highest posterior density regions. Bayesian testing of hypothesis problem. Prior odds, Posterior odds, Bayes factor for various types of testing of hypothesis problems.

UNIT 4. Simulation Techniques, Gibbs sampling, Monte-Carlo methods, Markov Chain Monte Carlo (MCMC) methods, bootstrap methods and other computer simulation methods.

References

Bansal A.K. (2007). Bayesian Parametric Inference, Narosa.
Bolstad W.M.(2007). Introduction to Bayesian Statistics, Wiley.
Lee, P.M. (2004). Bayesian Statistics, Arnold publishers.
Purohit S.G., Gore,S.D. and Deshmukh,S.R.(2008) Statistics Using R, Alpha Science

BSTA 402 SURVIVAL ANALYSIS AND DEMOGRAPHY

UNIT 1. Concepts of time, order and random censoring- right and left, likelihood in these cases, Survival function-Actuarial estimator, Kaplan-Meier (K-M) estimator, Graphical Display for Survival, Median survival time and confidence interval for median survival, Hazard Ratio, Relation between Hazard Ratio, Relative Risk and Odds ratio, Relative survival estimation.Nonparametric Methods for Comparing Survival Distributions - log rank test, Confidence Interval for Hazard Ratio, Stratified Log-rank test, Peto's test, Gehan test, Mantel-Haenzel test.

UNIT 2. Life Time distributions - Parametric (exponential, gamma, Weibull, loglogistic), Linear failure rate, Parametric inference (point estimation, confidence intervals, scores, Likelihood ratio test. Accelerated failure time model, Cox-Snell residuals.

UNIT 3. Identification of Prognostic Factors Related to Survival Time, Cox's proportional hazards regression model with one and several covariates, Rank test for the regression coefficients, Adequacy Assessment of the Proportional Hazards Model, Time dependent Extension of the Cox model, Tests with non-proportional hazards, parametric and nonparametric inference for this model.

UNIT 4. Measures of fertility and mortality, stochastic models for reproduction, distributions of time to first birth, inter-live birth intervals and of number of births for both homogeneous and non homogeneous groups of women, estimation of parameters, estimation of parity progression ratios from open birth interval data. Measures of mortality, construction of abridged life tables. Distributions of life table functions and their estimation, model life tables-Coul and Deming, U.N. model life tables. Estimation of measures of mobility. Methods for population projection. Use of Leslie matrix.

References

Klein, J.P. and Moeschberger, M.L.(2003). Survival Analysis, Springer.

Elandt, Johnson and Johnson (1998). Survival Models and Data Analysis, John Wiley & Sons.

Miller, R.G. (2000). Survival Analysis, Second Edition, John Wiley & Sons.

Machin D. ,Cheung Y.B. & Parmar M.KB.(2006).Survival Analysis-A Practical Approach, John Wiley & Sons.

Fisher L.D. & Belle G.V.(1993). Biostatistics-A Methodology for the Health Sciences, John Wiley & Sons.

Pressat R. & Atherton A. (1972). Demographic Analysis.

Preston S.H.,Heuveline P. & Guillot M. Demography-Measuring and Modelling Population Processes.

Deshpande, J.V. and Purohit,S.G.(2005) Life Time Data: Statistical Models And Methods,World Scientific

BSTA 403 CONTROLLED CLINICAL TRIALS AND OPERATIONS RESEARCH

UNIT 1. Introduction to clinical trials: the need and ethics of clinical trials, Drug Development Process, ICH GCP, Relevant FDA and EMEA guidelines (Industry, TA-, phase-specific), data management, objectives and end points of clinical trials, bias and random errors in clinical studies, conduct of clinical trials, overview of phase I-IV trials, multi-center trials. Design of clinical trials: parallel vs cross-over designs, cross-sectional vs: longitudinal designs. Design and analysis of Phase - I, Phase -II and Phase-III trials.

UNIT 2. Statistical Methods (Industry-, TA-, phase-specific),Defining objectives and end-points, Various study designs, Analysis data sets, Handling missing data, Handling multiplicity, Baseline and covariates, Sub-group analysis, Modeling treatment effects , Design of bio-equivalence trials, Understanding Protocol, Sample Size Determination, Inputs to Data Management Documents, Understanding Clinical Study Report, Randomization Methods, Statistical Analysis Plan, TLG Shells

UNIT 3: Analysis methods / models for continuous, categorical, binary, survival data, Non-parametric methods, Repeated measures analysis, Quality of life data analysis, Interim analysis, Data Comprehension, Data Interpretation, Adaptive Trials, Meta Analysis, SAS Programming – Efficacy reporting, Validation Plans, Blind Data Review, Data Visualizations Methods, Sensitivity Analysis, Unblinding, DSMB Reviews.

UNIT 4: Introduction to Operations Research, linear programming problems (LPP), framing an LPP problem, graphical solution, feasible, basic feasible and optimal basic feasible solutions to an LPP, simplex method, dual of linear programming, transportation problems, assignment problems, simple numerical problems as illustration.

Reference books

1. Friedman L.M.,Furberg C.D. & Demets D.L.(1998). Fundamentals of clinical trials, Springer
2. Shein-Chung Chow and Jen-Pei Liu(2004). Design and Analysis of Clinical Trials: Concepts and Methodologies (2nd edition) Wiley-Interscience
3. Stuart J. Pocock (2010)Clinical Trials – A practical approach (Reprint), John Wiley & Sons
4. Stephen Senn (2009) Statistical Issues in Drug Development (2nd edition), John Wiley & Sons Ltd.
5. Alex Dmitrienko, Geert Molenberghs, Christy Chuang-Stein, Walter Offen(2005). Analysis of Clinical Trials Using SAS – A Practical Guide ,SAS Publishing
6. David Collett (2003) Modeling Binary Data (2nd edition), Chapman & Hall/CRC
7. Alan Agresti (2002) Categorical Data Analysis (2nd edition), Wiley-Interscience

BSTA 404 BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

UNIT 1. Introduction to Bioinformatics: Bioinformatics Overview, Bioinformatics Concepts:- Functional Genomics, Comparative genomics, Structural biology, classification of protein structure, Medical information, Objectives of Bioinformatics. Applications, Challenges in Molecular biology, Careers in Bioinformatics, Major databases & tools, Bioinformatics in India.

UNIT 2. Genomics: Data Mining –ORF, Pubmed, Phylogenetic Analysis, MSA, Gen BANK, COG Cluster, OMIM, Gene Mapping, Sequence Assembly & Expression, Alignment of MS. **Proteomics***:* Visualization & prediction of Protein Structure, Methods used in protein structure prediction, PROSITE, DNA Micro array (DNA chip).

UNIT 3. Tools in Bioinformatics: Web based Bioinformatics Applications, Desktop based softwares, Online Analysis Tools & Servers, PDB, SWISS-PROT, CATH, Annotation Systems-DAS, Homology Tools –BLAST, FASTA, Multiple Alignment-CLUSTALW, Molecular visualization software-Swisspdb viewer, Rasmol Gene Prediction Softwares- Genescan, Protein, Modelling software-SWISSMODEL.

UNIT 4. Computational Biology: Genetic Algorithms, HMMR, Dynamic Programming Algorithm. Local & Global Alignment Algorithm, Needleman- Wunsch Algorithm, Heuristic Algorithm like BLAST, FASTA-Multiple Segment Alignment Algorithm, Protein secondary structure prediction Algorithm.

NB: As this paper requires computational techniques, hand-on practical sessions are important and should be held in conjunction with lectures.

References

Bergeron, B.(2003). Bioinformatics Computing, Prentice Hall of India.
Bozdogan, H (2003). Statistical Data Mining & Knowledge Discovery, CRC Press
Chen, Z (2001). Intelligent Data Warehousing, CRC Press
Ewens, W.J. and Grant, G.R. (2002). Statistical Methods in Bioinformatics, Springer.
Mount D.W. (2003). Bioinformatics – Sequence and Genome Analysis, CBS Publishers.
Rajan S.S. and Balaji R. (2002). Introduction to Bioinformatics, Himalaya Publishing House.
Shanmughavel P. (2005). Principles of Bioinformatics, Pointer Publishers.
Waterman, M.S.(2000). Introduction to Computational Biology, CRC Press.
Xiong, J.(2006). Essential Bioinformatics, Cambridge University Press.
Deshmukh, S.R. and Purohit,S.G.(2007) Microarray Data: Statistical Analysis Using R, Alpha Science.

BSTA 405 ADVANCED STATISTICAL COMPUTING

This paper includes practical problems from papers BSTA 401- 404. There will be 8 questions (2 from each course) of which 5 are to be answered. Data Analysis using SAS, SPSS and R software packages is expected.