MASTER OF SCIENCE

BIOSTATISTICS

PROGRAM STRUCTURE AND SYLLABUS
2019-20 ADMISSIONS ONWARDS

(UNDER MAHATMA GANDHI UNIVERSITY
PGCSS REGULATIONS 2019)

BOARD OF STUDIES IN BIOSTATISTICS (PG)
MAHATMA GANDHI UNIVERSITY

March 2019

SYLLABII RESTRUCTIONG WORKSHOP ORGANIZED BY THE PG BOARD OF STUDIES (STATISTICS)

The workshop on syllabi restructuring of Statistics started with an inaugural session on 17/12/2018 at 9.30 am at K E College Mannanam . The Board of studies , Chairperson Dr.Hitha.N welcomed the gathering. The workshop was inaugurated by the Hon'ble Vice Chancellor of Mahatma Gandhi university Dr. Sabu Thomas. In the inaugural address , the vice chancellor mentioned about the need of syllabus revision and added that foreign universities do revise their syllabi every year by incorporating the recent developments in the fields. Dr. Krishnadas, Member, Syndicate , Dr. Antony Thomas , Principal, K.E College, Dr. K.K Jose, Director, Department of Bio Statistics , St.Thomas college ,Pala felicitated the function . The inaugural session came to an end by 10.30 am. The workshop coordinator, Dr. Priya.P.Menon delivered the vote of thanks.

The technical session started at 10.45 am which was led by Dr. K.K Jose. The courses in each semester and their corresponding codes each for M.Sc Statistics (Pure), M.Sc Statistics (Applied) and M.Sc Biostatistics were decided in this session. Necessary corrections were made in the syllabus of each course. The general suggestion was to include text books and reference books of the latest edition. The technical session II started at 2.30 pm by Dr. Nibu A.George , Assistant Professor, Baselious College on the introduction of question bank system. The speaker gave a clear idea on how to include questions of varied difficulty level in the software. After this session, course wise discussion on the syllabus for Applied Statistics were made in detail.

Dr. K.R Sundaram , Prof. of Biostatistics , Amrita Institute of Medical Sciences led the first session on 18/12/2018. A lecture on the recent trends in Biostatistics was given by him. Both the morning and afternoon sessions were totally dedicated for the Biostatistics syllabus. Dr. Sundaram and Dr. K.K.Jose gave appropriate suggestions on the latest developments in the field and modifications in the syllabus were made accordingly.

The first session on 19/12/2018 started at 9.30 am. Dr Sebastian ,Associate Professor, St.Thomas College , Pala gave appropriate suggestions regarding the inclusions of certain topics in the courses Probability and Measure theory, Multivariate analysis, Advanced probability theory and Bayesian inference. The discussion on the syllabi of the courses Stochastic process, Estimation theory and Testing of hypotheses was led by Dr. Seemon Thomas , Associate Professor, St.Thomas college, Pala . Certain topics were included in the

existing syllabus of Design of Experiments and Sampling theory. In the afternoon session , discussion on the course ,Statistical computational techniques was initiated by Prof. K.A Rajeevan Pillai. He stressed the importance of R software in the field.

On 20/12/2018, discussion on the draft syllabi of all the courses, semester wise was done in detail. The afternoon session of 20/12/2018 and the morning session of 21/12/2018 were completely devoted for the model question papers on the draft syllabi.

The valedictory function started at 3 pm on 21/12/2018 .The duty certificates were distributed to the participants and the workshop came to a close by 4.00 pm .

A meeting of the BOS was organized at Maharajas College , Ernakulam on 01/03/2019 to finalise the syllabi and model question papers before submitting it to the University. Majority of the BoS members and the subject expert Prof. K.A.Rajeevan Pillai were present. The gathering revised the model question papers according to the new pattern proposed by the University.

Dr. Hitha.N,

Chairperson,

BoS(PG), Statistics.

Dr.Priya P. Menon,

Workshop Coordinator,

BoS(PG), Statistics.

ACKNOWLEDGEMENT

The PG Board of Studies is grateful to the members who have contributed in the curriculum restructuring of MGU-PG-CSS-2019- The Board of Studies also gratefully acknowledges the contribution of the participating members in the curriculum workshop and for the finalization of the syllabus.

Acknowledgements are due to Dr. K.K. Jose, Professor and Coordinator, Department of Biostatistics, St. Thomas College Pala and Prof. Rajeevan Pillai K.A, Associate Professor (Retd.) Maharaja's College, Ernakulam, by providing all academic support as subject experts. Thanks are due to Dr. K.R Sundaram , Prof. of Biostatistics, Amrita Institute of Medical Sciences for extending his help as a resource person. I express my gratitude to Dr. Nibu. A.George Assistant Professor, Baselious college, Kottayam for rendering a session on question bank preparation. My sincere gratitude to Dr. K.M. Kurian, Dr. Sebastian George, Dr. Benny Kurian, Dr.SeemonThomas, Dr. Deemat K.Mathew for their selfless effort throughout the preparation of the syllabi. Thanks to Dr.Smitha S , Dr..Jobin Varghese P,. Dr. Sindhu E.S, Dr. Dhannya P.Joseph and Sri. Tijo Mathew of K.E.College, Mannanam for their whole hearted support. I extend my thanks to the students of K.E College, Mannanam for their timely interactions in the workshop. My sincere gratitude to Dr. Maya. T.Nair , SVR ,N.S.S College, Vazhoor, Ms Rose Maria Jos and Ms. Meenu Tom and Mr. Noel George of St.Thomas college, Pala .

My sincere gratitude to Dr. Sabu Thomas, Hon'ble Vice Chancellor, M G University , Kottayam and all the university officials for their endless support. I express my sincere thanks to Dr. Praveenkumar V.S , member , Syndicate and convenor , Syllabus revision committee , Dr. K.Krishnadas, Member, Syndicate, M.G University , Kottayam and Dr. Antony Thomas, Principal K.E College, Mannanam for all the help they have rendered . Above all I express my wholehearted thanks to my dearest colleague Dr. Priya. P.Menon , co-ordinator , workshop on curriculum restructuring for her endless help without which I may not be able to finish this work.

We, the PG Board of Studies, Statistics , express our sincere thanks to all who have been helping for the success to this endeavor academically and administratively. The Board of Studies in Statistics acknowledges the contribution of the academic section

AcAIX M.G.University. Also we would like to place on records my appreciation and thanks to the faculty members of Nirmala College , Muvattupuzha who have been associated with this noble work in one or the other form.

Dr.Hitha . N

Ernakulam

Chairperson , PG BoS(Statistics)

April 02, 2019.

THE BOARD OF STUDIES IN STATISTICS&BIOSTATISTICS (PG)

Chairperson : Dr. Hitha N, Principal (Retd.),Govt. College Kodencherry, Kozhikode

Members

1	Dr. E.S Jeevanand	Associate Professor, UC College, Aluva.
2.	Dr. Benny Kurian	Associate Professor,St. Thomas College, Pala.
3.	Dr. Smitha S	Assistant Professor, K.E College, Mannanam
4.	Dr. Sindhu E.S	Assistant Professor,K.E College, Mannanam.
5.	Dr. Priya P.Menon	Assistant Professor,Maharajas College, Ernakulam.
6.	Dr. Anilkumar P	HoD(Retd),Dept of Statistics, Farook College.
7.	Dr. Rani Sebastian	Assistant Professor, St.Thomas College, Thrissur
8.	Shri Z.A.Ahamed Ashraf	Assistant Professor, Govt College, Kozhokode.
9.	Dr.Maya T. Nair	Assistant Professor,SVR NSS College, Vazhoor.
10.	Dr. Joby K.Jose	Associate Professor,Kannur University.
11.	Dr Jose.K.K (Subject Expert);	Professor and Coordinator, Dept. of Biostatistics Principal(Retd), St.Thomas College, Pala.
12.	Prof. Rajeevan Pillai.K.A(Subject Expert) :	Associate Professor(Retd),Maharajas College, Ernakulam.

Expert Committee Members:

1. Dr Jose.K.K(Subject Expert); Professor and Coordinator, Dept. of Biostatistics

2. Ms. Rose Maria Jose, Asst. Professor of Biostatistics, St. Thomas College Palai

Co-opted Members:

1.Ms. Meenu Tom, Asst. Professor of Biostatistics, St. Thomas College Palai

2.Mr. Noel George, Asst. Professor of Biostatistics, St. Thomas College Palai

.

.

.

.

.

.

Table of Contents

M. Sc Biostatistics Degree Program

(Mahatma Gandhi University Regulations PGCSS2019

From 2019-20 Academic Years)

1. **Aim of the Program**

M.Sc. Biostatistics Program is intended to develop statisticians with expertise to tackle Biostatistics and Epidemiology and health related issues. Biostatistics is the science of obtaining, analyzing, and interpreting data using statistical theory and methods to address problems in the biological and health sciences. Biostatisticians should be able to plan and organize health surveys, clinical trials, and analyze the data and draw meaningful conclusions. Interpretation of results and presentation in the form of a reportare essential components. They shall be experts in biological and epidemic modeling as well as apply statistical techniques in survival analysis, population studies and demography, ecology, genetics, gene sequence micro-array analysis etc. They should be familiar with statistical computing softwares like SPSS, SAS as well as open source programs like R and Python. The aim of the program is to generate good experts in statistical theory as well as practice. This will enable them with excellent career opportunities and research potential in emerging areas of interdisciplinary naturein academics as well as industry.

2. **Eligibility for Admissions:**Any candidate who has passed B.Sc., BDS, B.Tech. B.Pharm or MBBS with at least 50% marks(CGPA 2 out of 4) provided he/she has studied Mathematics or Statistics or Probability as a course at Plus Two level.
3. **Medium of Instruction and Assessment:** The medium of instruction and assessment shall be English.
4. **Faculty under which the Degree is Awarded** : Faculty of Science
5. Specializations offered, if any: Nil
6. Note on compliance with the UGC Minimum Standards for the conduct and award of Post Graduate Degrees: Ensured

7. THE PROGRAM STRUCTURE

Course Code	Title of the Course	Type of the Course	Hours per week	Credits
FIRST SEMESTER				
ST02 0101	Statistical Methods and Probability Distributions	Theory	5	4
ST02 0102	Theory and Methods of Sample Surveys	Theory	5	4
ST02 0103	Statistical Programming in R and Python	Theory	5	4
ST02 0104	Statistical Genetics and Ecology	Theory	5	4
ST02 0105	Statistical Data Analysis Using Microsoft Excel,R and Python	Theory & Practical	5	3
SECOND SEMESTER				
ST02 0201	Matrix Algebra and Regression Analysis	Theory	5	4
ST02 0202	Sampling Distributions and Statistical Estimation Methods	Theory	5	4
ST02 0203	Parametric and Nonparametric Tests	Theory	5	4
ST02 0204	Epidemiology and Study Designs	Theory	5	4
ST02 0205	Statistical Data Analysis Using SPSS, R and Python	Theory& Practical	5	3

Each student has to carry out a mini project work during the second semester vacation

THIRD SEMESTER				
ST02 0301	Designs and Analysis of Experiments	Theory	5	4
ST02 0302	Stochastic Models and Time Series Analysis	Theory	5	4
ST02 0303	Applied Multivariate Analysis	Theory	5	4
ST02 0304	Advanced Epidemiology and Bioassays	Theory	5	4
ST02 0305	Statistical Data Management using Python, ADVANCED R and SPSS	Theory & Practical	5	3
ST02 0306	Mini Project	2	Nil	2

FOURTH SEMESTER				
ST02 04 01	SAS Programming, Bayesian Inference and MCMC Methods	Theory	5	3
ST02 04 02	Survival Analysis and Lifetime Modeling	Theory	5	3
Electives Bunch I				
ST86 04 01	Elective 1- Clinical Trials and Bioinformatics	Theory	5	3

ST86 04 02	Elective 2-Demography & Vital Statistics	Theory	5	3
ST86 0403	Elective 3-- Statistical Computing and Data Analysis using R and SAS	Theory & Practical	5	3
Electives Bunch II				
ST87 0401	Elective 1- Advanced Clinical Trials and Operations Research	Theory	5	3
ST87 0402	Elective 2– Quality Control and Research Methodology	Theory	5	3
ST87 0403	Elective 3 Statistical Computing and Data Management using R and SAS	Theory & Practical	5	3
ST 02 04 03	Project Work and on the Job Training (in a reputed industry / Research Institute / Medical College)	Project Report & Presentation	Nil	3
ST02 04 04	Comprehensive Viva-voce		Nil	3
	Total Credits			**80**

FIRST SEMESTER COURSES

ST 02 0101 STATISTICAL METHODS & PROBABILITY DISTRIBUTIONS

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:**To impart basic knowledge & skills in the fundamentals of Statistics & Probability Theory & their applications in Biostatistics.

2. **UNIT 1:Elementary concepts in Statistics:1.1**Concepts of statistical population and sample from a population**1.2**Qualitative and Quantitative data**1.3**Nominal, Ordinal, Ratio, Interval data**1.4** Cross sectional and Time series data**1.5**Discrete and Continuous data **1.6** Collection and scrutiny of data**:** Primary data**,** Designing a questionnaire and a schedule**,**Secondary data and sources of secondary data **1.7** Presentation of data: Diagrammatic and Graphical representation of data; Frequency distributions and Cumulative frequency distributions; Histogram, Frequency polygon, Stem and Leaf Diagram and Ogives.**1.8**Descriptivestatistics: Concepts of central tendency or location, Absolute and relative measures of dispersion**1.9** Box plot, Lorenz curve**2.0**Skewness and kurtosis**2.1** Bivariate Data Analysis, **2.2**Scatter Diagram, **2.3**Pearson's Correlation Coefficient & Rank Correlation, Interpretation

3. **UNIT 2: Probability: 2**.1Random Experiment **2.2** Sample point **2.3** Sample space**2.4** Events**2.5** Mutually exclusive and exhaustive events**2.6** Frequency and Classical definitions of probability **2.7** Axiomatic definition of probability **2.8**Addition and Multiplication theorems **2.9** Conditional probability and independence **3.0**Bayes' theorem. (The main thrust is on numerical problems and applications) **3.1**Discrete and continuous random variables **3.2** Probability density functions and distribution functions **3.3**Expectation of a random variable.

4. **UNIT 3: Standard Univariate Distributions: 3.1** Standard univariate discrete distributions – Uniform; Binomial; Poisson; Geometric; Negative Binomial and Hyper-

geometric distributions. **3.2**Standard univariate continuous distributions –Uniform; Exponential; Normal; Laplace, Gamma, Beta, Log-normal, Logistic and Weibull distributions.(elementary properties and applications only) **3.3** Chi-square distribution **3.4** F and t statistics, distributions (no derivations) and their applications .**3.5** Chi-square test for goodness of fit.

5. **UNIT 4: 4.1** Central Limit Theorem for i.i.d case (statement and examples only) **4.2** Evaluation of probabilities from the binomial and Poisson distributions using central limit theorem **4.3**Tchebychev's inequality and weak law of large numbers (statement and applications only)

1. **Recommended Text Books:**

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

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

Pagano, M.& Gauvreau, K. (2007) –Principles of Biostatistics, Thomson India Edition.

P Mariappan (2013). Biostatistics :An Introduction, Pearson Publishers, New Delhi

2. **Recommended Reference Books:**

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

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

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

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

S.C Gupta & V.K. Kapoor (2002). Fundamentals of Mathematical Statistics, Sultan Chand & Sons, New Delhi

ST 02 01 02 THEORY& METHODS OF SAMPLE SURVEYS

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Theory & Methods of Sample Surveys& their applications in Biostatistics, Clinical Trials, Public Health Surveys etc.

2. **UNIT 1. 1.1 Concepts of population and sample1.2** Need for sampling **1.3** Census and sample surveys **1.4** Principle steps of sample surveys **1.5** Sampling and non-sampling errors **1.6** Probability & Non-probability sampling **1.7** Sample size determination **1.8** Finite population sampling techniques-SRSWR, SRSWOR estimation of mean, total & proportion in each case and their variances.

3. **UNIT 2. Stratified sampling & Systematic sampling,: 2.1** Allocation problems in stratified sampling –Proportional & Optimum allocations (estimation of mean or total in each case and their variances) **2.2** Post Stratification **2.3** Systematic Sampling – Linear & Circular , estimation of mean & variance (derivation not expected). Simple examples from health sciences.

4. **UNIT 3. 3.1 Ratio and regression estimators** based on SRSWOR method of sampling **3.2**Auxiliary information in sample surveys **3.3** Randomized response techniques: Warner's model-related and unrelated questionnaire methods.

5. **UNIT 4. Cluster Sampling & Unequal Probability Sampling**: **4.1**Cluster Sampling **4.2**Double sampling **4.3** Two stage and multi-stage sampling **4.4** Estimation of Population Mean &variances(no derivation expected) **4.5** Simple illustrative examples. **4.6**PPS WR / WOR methods and related estimators of a finite population

mean **4.7** Horvitz- Thompson estimator of a finite population total & mean and expressions for variance and its unbiased estimator.

1. **Recommended Text Books:**

Singh D and Choudhary F.S (1986) Theory and Analysis of Sample Survey Designs, Wiley Eastern Ltd.

Sampath S.(2005). Sampling Theory and Methods, Narosa Publishers

Murthy, M.N. (1967). Sampling Theory and Methods. Statistical Publishing Company, Calcutta.

2. **Recommended References:**

Cochran, W.G. (2002). Sampling Techniques. Wiley,

Des Raj and Chandhok (1998). Sampling Theory, Narosa.

Gupta.S.C and Kapoor V.K (2007) Fundamentals of Applied Statistics. Sultan Chand & Sons Educational Publishers, New Delhi.

Parimal Mukhopadhyay (2009) Theory and Methods of Survey Sampling, PHI Learning Private Limited,NewDelhi

ST 02 01 03 STATISTICAL PROGRAMMING IN R AND PYTHON

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To provide basic knowledge & skills inStatistical Programming usingR & PYTHON and their applications in Biostatistics and Data Science as well as Data Analytics field.

2. **Unit 1**: **1.1** Introduction to R programming **1.2** Features of R **1.3** Variables **1.4** R operators **1.5** Data objects in R (data types)- Strings, Vectors, Lists, Matrices, Arrays and Data Frames **1.6** Decision Making Statement in R-if statement, if-else statement

1.7 Switch statement 1.8 R Loops- While loops & For loops **1.9** R-functions – built-in functions & user –defined functions.

3. **Unit 2**: **2.1** R charts and graphs-Pie chart ,Bar chart ,Box plot, Histograms, Line graphs and Scatter plot **2.2** Computation of descriptive statistics like mean, median, mode, standard deviation, variance, coefficient of variation, coefficient of skewness and kurtosis, moments **2.3** Generating samples from different distributions**2.4** Plotting pdf, cdf.etc of standard distributions like binomial, Poisson, exponential, normal, etc.

4. **Unit 3**: **3.1** Introduction to Python **3.2** Variables **3.3** Operators **3.4** Data types **3.5** Decision Making Statements **3.6** Loops**3.7** Functions: Casting Functions, Mathematical Functions, Random Functions, **3.8**String Functions, The capitalize (), center (), and count () functions, The find (), isalpha (), and isdigit () functions, The join (), len (), and split () functions

5. **Unit 4**:**4.1**Charts and graphs-Pie chart,Bar chart ,Box plot, Histograms, Line graphs and Scatter plot **4.2** Computation of descriptive statistics **4.3** Generating samples from different distributions **4.4** Computing averages, measures of dispersion, skewness, kurtosis, correlation etc

References Textbook

Purohit S.G, Gore S.D and Deshmuk S.R(2008) Statistics Using R. Narosa Publishing House, New Delhi.

David M. Beazley (2009) Python Essential Reference

Recommended References:

Mark Lassoff and Julius Hernandez (2018) Introduction to Python LearnToProgram, LLC

ST 02 01 04 STATISTICALGENETICS AND ECOLOGY

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To provide basic knowledge & skills in Statistical Genetics and Ecology for solving the emerging issues in biological modeling, ecological studies, bio-diversity assessment etc.

2. **Unit 1: 1.1** Basic biological concepts in genetics **1.2** Mendel's law **1.3** The law of natural selection, mutation and genetic drift **1.4** Hardy-Weinberg equilibrium, estimation of allele frequency (dominant/co-dominant cases) **1.5** Approach to equilibrium for X-linked gene. **1.6** Non-random mating and inbreeding **1.7** Phenotypic assortative mating,

3. **Unit 2: 2.1**.Pedigree data: Elston-Stewart algorithm for calculation of likelihood **2.2**. Linkage, **2.3** Genetic mapping **2.4** Linkage equilibrium **2.5**Partitioning of Chi-square **2.6**Detection of linkage and estimation of re-combination fraction **2.6** Inheritance of quantitative traits.

4. **Unit 3: 3.1** Introduction to ecology and evolution **3.2** Population dynamics: single species-Exponential, Logistic and Gompertz models 3**.3** Leslie matrix model for age and stage Structured population 3**.4** Survivorship curves-Constant, monotone and bath tub shaped hazard rates 3**.5** Two species: Lotka-Volterra equations 3**.6** Isoclines

5. **Unit 4: 4.1**Abundance estimation: Capture –recapture **4.2**Nearest Neighbor**4.3** Line transect sampling **4.4** Indirect methods**. 4.5** Ecological Diversity: Species

abundance curve, indices of diversity (Simpson's index, Shannon-Wiener index) **5.6** Game theory in ecology – **4.7** Evolutionarily stable strategy, its properties**, 4.8** simple games such as Hawk-Dove game, Prisoner's dilemma.

1. Recommended Text Books:

Anil Gore &SharayuParanjpe (2001). A Course in Mathematical and Statistical Ecology,Kluwer academic Publishers.

Lange, K (2002). A Course in Mathematical and Statistical Methods for Genetic Analysis, Springer.

Falconer D.S.(1991) Introduction to quantitative Genetics, ELBS Logman group.

2. Recommended Reference Books:

Gardner E.J. &Simmons D. P.(2007)Principles of Genetics, John Wiley & Sons Inc.

Lange, K (2002). Mathematical and Statistical Methods for Genetic Analysis, Springer.

Robert J Booker (2009) Genetics: Analysis & Principles, McGraw-Hill.

Robert H Tamarin, (2001) Principles of Genetics, McGraw-Hill.

ST 02 01 05 STATISTICAL DATA ANALYSIS USING MICROSOFT EXCEL, R & PYTHON

Total Credits: 3

Total Hours: 25

Weightage: 30

1. **Objective of the Course:**This course is intended to impart the skills in data analysis and biostatistical computation. It also enables to draw valid conclusions and interpretation of the findings.

2. **Syllabus:**

   This paper includes practical problems using data from Biostatistical contexts based on all courses during Semester I namely, ST02 0101-0104. There will be 4 questions (1 from each course) of which 3 are to be answered. Questions shall be based on moderately sizedreal data sets to illustrate the applications of theory. Data Analysis using Microsoft Excel, R& Python is expected. The duration of examination is 3 hours.

   Practical training should be imparted to the students in the computer lab for analyzing data sets using Microsoft Excel, R. &Python and Practical Record should be prepared by each student.

SECOND SEMESTER COURSES

ST 02 02 01 MATRIX ALGEBRA ANDREGRESSION ANALYSIS

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Matrix Algebra, Correlation &Regression analysis& their applications in Biostatistics

2. **UNIT 1. Linear Algebra:1.1** Set operations **1.2** Vectors and matrices **1.3** Different types of matrices and their operations **1.4** Determinants **1.5** Inverse of a square matrix **1.6** Linear independence **1.7** Rank of a matrix **1.8** Generalized inverses and applications **1.9** Linear equations **1.10** Characteristic roots and vectors **1.11** Quadratic forms and nature of definiteness.

3. **UNIT 2. Linear regression: 2.1** Simple linear regression **2.2** Partial and multiple correlations **2.3** Multiple regression **2.4** Model adequacy checking **2.5** Residuals and their plots **2.6** Tests for departure from assumptions such as fitness of the model **2.7** normality **2.8** Homogeneity of variances **2.9** Detection of outliers and remedies **2.10** Influential observations **2.11** Variance stabilizing transformations **2.12** Power transformations for dependent and independent variables **2.13** Generalized and Weighted least squares.

4. **UNIT 3**. **Generalized linear models**, **3.1**Analysis of binary and grouped data by using logistic models **3.2** Large sample tests about parameters **3.3** Goodness of fit **3.4** Analysis of deviance **3.5** Variable selection **3.6** Introduction to Poisson regression **3.7** Logistic Regression **3.8** Log-linear models,

5. **UNIT 4.Nonparametric regression4.1** Polynomial regression models **4.2** Orthogonal polynomials **4.3** Random and mixed effect models **4.4** Multi-collinearity **4.5** Robust regression **4.6**Non-Linear Regression **4.7** Parameter estimation **4.8** Linearization **4.9** Diagnostics for Leverage and influence.

3. Recommended Text Books:

Bapat R.B. (1999) Linear Algebra and Linear Models; Hindustan Book Agency

Montgomery, D.C ; Peck, E.A, Vining, G.G (2003). Introduction to Linear Regression Analysis,John Wiley & Sons

Draper, N.R. and Smith, H (2003). Applied Regression Analysis, John Wiley & Sons.

4. **References**

Gupta S.C. and Kapoor V.K.(2012) Fundamentals of Mathematical Statistics, S. Chand and Company, New Dehi

Shanti Narayan (2000) A Text Book of Matrices, S. Chand and Company, New Dehi

Rao, C.R.(2012) Linear Statistical Inference and its Applications; Wiley Asia

Rossi R.J.(2010) Applied Biostatistics for Health Sciences, Wiley.

ST02 0202SAMPLING DISTRIBUTIONS & STATISTICAL ESTIMATION METHODS

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Sampling Distributions & Statistical Estimation& their applications Biostatistics.

2. **UNIT 1. Sampling Distributions :1.1** Concepts of random sample **1.2** Parameter, Statistic **1.3** Standard error **1.4** Distribution of sample mean and sample variance from a normal population**1.5** Student's t, Chi Square and F distributions and Statistics**1.6** Applications with simple illustrations.

3. **UNIT 2. Basic concepts and properties of estimators**: **2.1** Statistical Inference **2.2** Parametric models **2.3** Parameter space **2.4** Sample space **2.5** Parameters **2.6** Random sample and its likelihood **2.7** Statistic and its sampling distribution **2.8** Estimator and estimate **2.9** Mean square error (MSE) **2.10** Properties of estimators- unbiasedness, consistency, efficiency, relative efficiency of an estimator (statement and its applications)

4. **UNIT 3 Sufficiency 3.1** Cramer-Rao lower bound **3.2** Minimum variance unbiased estimator **3.3** Fisher information **3.4** complete and sufficient statistic **3.5** One-parameter exponential family and multi-parameter exponential family **3.6** Rao-Blackwell theorem and UMVUE.(*Statement and applications only)*

5. **UNIT 4 Methods of Estimation &Interval estimation**: **4.1** Method of moments **4.2** method of MLE **4.3** Properties of MLE (statements only) **4.4** Method of minimum chi-square and modified minimum chi-square ***4.5*** linkage estimation ***(****Examples from Genetics).***4.6* **Concepts of confidence interval**: confidence coefficient **4.7** Confidence interval for the parameters of univariate normal, proportion, mean, difference of means **4.8** Small sample and large sample confidence intervals **4.9** Large sample confidence intervals for binomial and Poisson parameters **4.10** Bootstrap methods.

1. **Recommended Text Books :**

Hogg R.V. and Tanis E.A.(2001). Probability and Statistical Inference, Prentice Hall International Inc.

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

2. **Recommended Reference Books:**

Manly, B. F. (2007). Randomization, Bootstrap and Monte Carlo methods in Biology, Chapman & Hall / CRC.

Lehmann E.L and George Casella .(1998). Theory of Point Estimation, Springer.

S.C Gupta & V.K. Kapoor (2002). Fundamentals of Mathematical Statistics, Sultan Chand & Sons, New Delhi

ST02 0203 PARAMETRIC ANDNON-PARAMETRIC TESTS

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills Parametric &Non Parametric Methods& their applications in Biostatistics

2. **UNIT 1. Testing of hypothesis:1.1** Basic concepts **1.2** Simple and composite hypotheses **1.3** Two types of errors **1.4** Critical region **1.5** Significance level **1.6** Size and power of the test **1.7** p-value and its interpretation **1.8**Neymann-Pearson Lemma **1.9** UMP & LRT (Statement only and its application with respect to Binomial,Poisson, Exponential & Normal distributions).

3. **UNIT 2.Parametric&Non-parametric One-Sample Tests 2.1** Test for mean when SD is known & unknown **2.2** Test for Proportion **2.3** Test for variance **2.4** Binomial test **2.5** Ordinary Sign test **2.6** Sign Test for Quintiles**2.7** One Sample Signed Rank Test **2.8**Kolmogrov-Smirnov Goodness of Fit **2.9** Run Test for Randomness

4. **UNIT 3. Parametric & Non-parametric Two-Sample Tests:3.1** Test for equality of variances **3.2** Test for equality of means(independent samples & dependent samples) **3.3** Test for proportions **3.4** Sign Test(2 sample) **3.5** Wilcoxon Signed Rank Test(2-sample) **3.6** Wilcoxon –Mann-Whitney (Rank Sum) **3.7** Wald-Wolfowitz Run Test for identical populations **3.8** Two sample Kolmogrov-Smirnov Test **3.9** Freidman's test **3.10** Kruskal Wallis test (large sample and small sample cases).

5. **UNIT 4.Hypothesis Testing-Categorical Data &Basics of sequential testing:**
   4.1 Chi-Square Tests & its applications- Goodness of Fit **4.2** Test for Independence **4.3** Homogeneity **4.4** Median Test **4.5** Fisher Exact test **4.6**Wald's SPRT with

illustrations **4.7**OC and ASN functions for tests regarding binomial and normal populations **4.8** Linkage estimation **4.9** Partitioning of Chi-square and detection of linkage.

1. **Recommended Text Books:**

Hogg R.V. and Tanis E.A.(2001). Probability and Statistical Inference, Prentice Hall International Inc.

Lehmann,E.L and Romano Joseph P. (2009).Testing Statistical Hypothesis, Springer

Gibbons Jean Dickinson(1976).NonParametric Methods for Quantitative Analysis, Mc Gill Publishers

2. **Recommended ReferenceBooks:**

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

ST 02 02 04 EPIDEMIOLOGY AND STUDY DESIGNS

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Epidemiology & Study Designs & their applications in Biostatistics

2. **UNIT 1. Basic concepts & Measures of exposure and outcome**: **1.1** What is epidemiology **1.2** History of Epidemiology **1.3** Emergence of modern epidemiology **1.4** Measures of Exposures **1.5** Types of exposures **1.6** Sources of exposures **1.7** Measures of outcome **1.8** Disease registries **1.9** Classification of diseases **1.10**

Measures of disease frequency **1.11** Prevalence **1.12** Incidence **1.13** Risk **1.14** Odds of disease **1.15** Incidence time **1.16** Incidence rate **1.17** Relationship between prevalence, rate and risk **1.18** Routine data to measure disease occurrence **1.19** Age standardization **1.20** Direct method of Standardization **1.21** Indirect method of standardization **1.22** Cumulative rate **1.23** Cumulative risk **1.24** Proportional incidence.

3. **UNIT 2**. **Overview of study designs**: **2.1** Type of study design **2.2** Intervention studies **2.3** Cohort studies **2.4** Case-control studies **2.5**Cross-sectional studies **2.6** Ecological studies **2.7** Measures of exposure effect **2.8** Relative and absolute measures of effect **2.9** Confidence intervals and significance tests for measures of occurrence and effect.

4. **UNIT 3.Case-control studies:3.1** Definition of cases and controls **3.2** Methods of selecting cases and controls **3.3** matching **3.4** Sample size **3.5** Power calculations **3.6** Basic methods of analysis of grouped data **3.7** Basic methods of analysis of matched data. **3.8** Logistic regression for case-control studies **3.9** Estimation and interpretation of logistic parameters **3.10** Matched analysis- estimation of logistic parameters **3.11** unmatched analysis of matched data **3.12** confounder score **3.13**Categorical data analysis.

5. **UNIT 4. Cohort studies: 4.1** Prospective cohort studies: planning and execution **4.2** retrospective cohort **4.3**Nested case-control 4.4 Case-cohort studies: planning and execution **4.5** Matching and efficiency in cohort studies **4.6** Cohort studies –statistical analysis **4.7Longitudinal studies:** Design, execution and analysis of longitudinal studies **4.8** Repeated measurement analysis**.**

1. **Recommended Text Books :**

Isabel dos Santos Silva,(1999) Cancer Epidemiology: Principles and Methods, International Agency for Research on Cancer.

2. **Reference Books:**

Ahrens W. and PigcotI.(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

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

ST 02 02 05 STATISTICAL DATA ANALYSIS USING SPSS, R& PYTHON

Total Credits: 3

Total Hours: 25

Weightage: 30

1. **Objective of the course:**

This course is intended to impart the skills in data analysis and biostatistical computation. It also enables to draw valid conclusions and interpretation of the findings.

2. **Syllabus:**

This paper includes practical problems using data from Biostatistical contexts based on all courses during Semester II namely, ST02 0201-0204. There will be 4 questions (1 from each course) of which 3 are to be answered. Questions shall be based on

moderate data sets to illustrate the application of theory. Data Analysis using SPSS, R and Python is expected. The time duration for examination is 3 hours.

Practical training should be imparted to the students in the computer lab for analyzing data sets using SPSS, R and Python. Practical Records are to be submitted by each student.

THIRD SEMESTER COURSES

ST 02 03 01 DESIGN AND ANALYSIS OF EXPERIMENTS

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Design and Analysis in Experiments & their applications in Biostatistics and related areas.

2. **UNIT 1**. **1.1** Introduction to design of experiments **1.2** Gauss-Markov Theorem (meaning and statement only) **1.3** Analysis of variance(ANOVA) – one-way and two-way ANOVA with examples **1.4** Complete and incomplete Block designs **1.5** Basic principles of experimental design **1.6** CRD-design, model, analysis, advantages and disadvantages **1.7** Missing plot techniques, one-missing observation in CRD, simple problems

3. **UNIT 2**. **2.1** RBD- design, model, analysis, advantages and disadvantages **2.2** One-missing and two-missing observations in RBD, simple problems **2.3** LSD- design, model, analysis, advantages and disadvantages **2.4** One-missing and two-missing observations in LSD, simple problems

4. **UNIT 3.3.1 Incomplete block designs**: **3.2** Balanced incomplete block designs (BIBD) **3.3** Partially Balanced incomplete block designs (PBIBD)**3.4** ANOVA Table and simple illustrations**3.5**Hierarchical and nested designs **3.6**Split plot experiments **3.7**Analysis of Covariance (ANACOVA) and examples

5. **UNIT 4**: **4.1** General factorial experiments **4.2** Factorial effects **4.3** 2^n and 3^n factorial experiments in randomized block with simple problems **4.4** Yate's method **4.5** complete and partial confounding.**4.6** Simple examples and ANOVA.

1. **Recommended Text Books :**

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

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

2. **Recommended Reference Books:**

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

   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.

ST 02 03 02 STOCHASTIC MODELS AND TIME SERIES ANALYSIS

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Stochastic Models & Time Series & their applications in Biostatistics

2. **UNIT 1.1.1** Introduction to stochastic processes (sp's) **1.2** Classification of sp's according to state space and time domain **1.3** Countable state Markov chains (MC's) **1.4** Chapman-Kolmogrov equations; calculation of n-step transition probability and its limit **1.5** Stationary distribution **1.6** Classification of states; first –passage time problems.

3. **UNIT 2.2.1** Stationary processes; weakly stationary and strongly stationary processes **2.2** Discrete state space continuous time MC **2.3** Kolmogrov differential equations **2.4** Poisson processes and properties **2.5** Pure birth process, Pure death process,Yule process**2.6** Birth and Death processes and applications.

4. **UNIT 3**. **3.1** Renewal theory: Basic concepts, current life, excess life, total life **3.2** Poisson process as a renewal process **3.3**Elementary renewal theorem and applications **3.4**Statement and uses of key renewal theorem **3.5**Branching process: Galton –Watson branching process, pgf relations, **3.6**Probability of ultimate extinction and conditions **3.7**Distribution of population size, mean and variance

5. **UNIT 4.4.1** Box Jenkins Models **4.2** Moving average processes **4.3** Auto regressive processes **4.4** ARIMA models **4.5** Auto correlation function and correlogram, diagnostic checks, modeling and prediction **4.6** Non-Gaussian time series models **4.7** Applications in biostatistical contexts.

1. **Recommended Text Books:**

Karlin and Taylor (1985) A First Course In Stochastic Processes, Academic Press

Medhi, J, (1982): Stochastic Processes, New AgeInternational Publishers, New Delhi

Brockwell and Davis. (2010) Introduction to Time Series and Forecasting, Springer

2. **Recommended Reference Books:**

Basu A.K. (2003).Introduction to Stochastic Processes, Narosa Publishing House.

Feller, W. (1968): Introduction to Probability and its Applications, Vol.1, Wiley Eastern..

Suddhendu Biswas (1995). Applied Stochastic Processes: A Biostatistical and Population Oriented Approach, Wiley Eastern.

ST 02 03 03 APPLIED MULTIVARIATE ANALYSIS

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Applied Multivariate Analysis & their applications in Biostatistics

2. **UNIT 1. 1.1** Multivariate data **1.2** Multivariate normal distribution **1.3** Random sampling from a multivariate normal distribution **1.4** Maximum likelihood estimators of parameters **1.5** Distribution of sample mean vector.

3. **UNIT 2.2.1**Hotelling's T^2 Statistics (one sample & two sample) **2.2**Mahalanobis D^2 statistics **2.3** Applications in tests on mean vector for one and more multivariate normal populations **2.4** Test the equality of the components of a mean vector in a multivariate normal population **2.5** Wishart distribution and applications.

4. **UNIT3. 3.1** Likelihood ratio test criterion **3.2** Multivariate Analysis of Variance [MANOVA] of one-and two-way classified data **3.3**Dimension reduction **3.4** Principal components–estimation and computation **3.5** canonical correlations and applications.

5. **UNIT 4.4.1** Classification and discrimination procedures for discrimination between two multivariate normal populations **4.2** Sample discriminant function **4.3** Tests associated with discriminant functions **4.4** Classification into more than two multivariate normal populations **4.5** Cluster analysis **4.6** Hierarchical and agglomerative methods

1. **Recommended Text Books:**

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

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

2. **Recommended Reference Books:**

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

   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,

   Morrison, D.F. (1976): Multivariate Statistical Methods. 2nd.Ed. McGraw Hill.

ST 02 03 04 ADVANCED EPIDEMIOLOGY & BIOASSAYS

Total Credits: 4

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills inAdvancedEpidemiology& Bioassays& their applications in Biostatistics

2. **UNIT 1. Confounding1.1** Confounding **1.2** Assessment of confounding **1.3** Mantel-Haenszel summary measures of effect **1.4** Interaction**1.5** Mantel-Haenszel method to adjust for several confounders **1.6** Confidence intervals and statistical tests for adjusted relative measures of effect.

3. **UNIT 2. Bias, Attributable risk and Causation in Epidemiology: 2.1** Sources of bias **2.2** Selection bias **2.3** Measurement bias **2.4**Misclassification of exposure and

outcome **2.5** Differential and non-differential exposure and outcome classification **2.6** Excess risk and Attributable risk **2.7** Causation and Hill's criteria.

4. **UNIT 3. Validity and reliability of measures of exposure and outcome**: **3.1** Sensitivity **3.2** Specificity **3.3** Predictive value method for selecting a positivity criterion **3.4** Receiver operator characteristic (ROC) curve **3.5** Intra and Inter-observer reliability **3.6**Kappa measure of agreement.

5. **UNIT 4. Bioassays:4.1** Types of biological assays **4.2** Direct assays **4.3** Ratio estimators **4.4** Asymptotic distributions **4.5** Regression approaches for estimating dose response relationships **4.6** Quantal responses **4.7** Methods of estimation of parameters **4.8** Dose allocation schemes **4.9** Median dose **4.10** Estimation of points on the Quantal response function **4.11** Estimation of safe doses.

1. **Recommended Text Books:**

Epidemiology (2008)LeonGordis, Elsevier Publishers

Z. Govindarajulu (2001) Statistical Techniques in Bioassays. Karger publication

2. **Recommended Reference Books:**

Ahrens W. and PigcotI.(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

M.Iqbal Choudhary and William J.Thomson (2001) Bioassay techniques for drug development. Harvard Academic Publishers

ST 02 03 05 STATISTICAL DATAMANAGEMENT USING PYTHON, ADVANCED R AND SPSS

Total Credits: 3

Total Hours: 25

Weightage: 30

1. **Objective of the course:**

   This course is intended to impart the skills in data analysis and biostatistical computation. It also enables to draw valid conclusions and interpretation of the findings.

2. **Syllabus:**

   This paper includes practical problems using data from Biostatistical contexts based on all courses during Semester III namely, ST02 0301-0304. There will be 4questions (1 from each course) of which 3 are to be answered. Questions shall be based on moderate data sets to illustrate the application of theory. Data Analysis using Advanced R, SPSS & Python is expected. The time duration of examination is 3 hours.

   Practical training should be imparted to the students in the computer lab for analyzing data sets using Advanced R, SPSS & Python. Practical Record should be submitted by each student.

ST02 0306:MINI PROJECT

Total Credits: 2

In order to give exposure in planning and execution of a biostatistical study, a mini project work will be done during the second semester vacation under the guidance of a medical doctor or scientist in a hospital or a research institute. The data will be collected and analyzed by each student during the third semester and a project report of 20-30 pages will be submitted to the course coordinator before the commencement of Semester III examination and it will be valued by the external examiner for ST02

0305. It is meant for understanding the essential ingredients for the Project Work during the final semester.

FOURTH SEMESTER COURSES

ST 02 04 01 SAS PROGRAMMING, BAYESIAN INFERENCE AND MCMC METHODS

Total Credits: 3

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in R Programming, Bayesian Inference and MCMC methods& their applications in Biostatistics

2. **UNIT 1. Elements of SAS Programming : 1.1** Introduction to SAS: SAS variables, **1.2**Libraries ,Windows, **1.3**Parts of a SAS program, **1.4**Data sets-Creation, **1.5**Reading data from an external file, **1.6**Data step statements like CARDS, DATA, Assignment, Do-loops, **1.7**DROP, KEEP, INPUT, OUTPUT, SET, STOP, IF-THEN-ELSE, **1.1** SAS Operators **1.8**Functions **1.9**Arrays. **1.10**Procedures in SAS, Proc statements, **1.11**CLASS, BY, DROP, FREQ, KEEP, OUTPUT, LABEL etc. **1.12**Procedures-PRINT, FREQ, MEANS, UNIVARIATE, SORT, CONTENTS, CORR, PLOT, REG, ANOVA, LOGISTIC, IMPORT, TABULATE.

3. **UNIT 2. Elements of Bayesian approach : 2.1** Subjective interpretation of probability **2.2** Evaluation of subjective probability of an event **2.3** Prior distribution of a parameter **2.4** Bayes' theorem and computation of the posterior distribution **2.5** Natural Conjugate family of priors for a model **2.6** Conjugate families for (i) exponential family models, (ii) models admitting sufficient statistics of fixed dimension.

4. **UNIT 3. Bayesian Inference: 3.1** Bayes estimators for (i) absolute error loss (ii) squared error loss (iii) 0-1 loss **3.2** Generalization to convex loss functions **3.3**Evaluation of the estimate in terms of the posterior risk **3.4** Bayesian interval estimation **3.5** Credible intervals **3.6** Highest posterior density regions **3.7** Bayesian testing of hypothesis problem **3.8**Prior odds **3.9**Posterior odds **3.10** Bayes factor for various types of testing of hypothesis problems.

5. **UNIT 4. Simulation Techniques : 4.1** Random number generation-Inverse Method: Acceptance-rejection method, Metropolis-Hastings algorithm **4.2**Gibbs sampling **4.3** Monte-Carlo methods **4.4** Markov Chain Monte Carlo (MCMC) methods **4.5** Bootstrap methods, Jackknife method**4.6** EM algorithm.

1. **Recommended Text Books:**

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

Bansal A.K. (2007) Bayesian Parametric Inference, Narosa

BolstadW.M.(2007) Introduction to Bayesian Statistics, Wiley

2. **Recommended Reference Books:**

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

Jim Albert (2009) Bayesian Computation with R, Springer

Lee P.M. (2004). Bayesian Statistics, Arnold Publication.

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

ST 02 04 02 SURVIVAL ANALYSIS AND LIFETIME MODELING

Total Credits: 3

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Survival Analysis& their applications in Biostatistics

2. **UNIT 1. 1.1** Concepts of time **1.2** Order **1.3** Random censoring- right and left **1.4** Likelihood in these cases **1.5** Survival function-Actuarial estimator **1.6** Kaplan-Meier (K-M) estimator **1.7** Graphical Display for Survival **1.8** Median survival time and confidence interval for median survival **1.9** Hazard Ratio **1.10** Relation between Hazard Ratio **1.11** Relative Risk **1.12** Odds ratio **1.13** Relative survival estimation **1.14** Nonparametric Methods for Comparing Survival Distributions - log rank test **1.15** Confidence Interval for Hazard Ratio **1.16** Stratified Log-rank test **1.17**Peto's test **1.18**Gehan test **1.19** Mantel-Haenzel test.

3. **UNIT 2. 2.1** Life Time distributions - Parametric (Exponential, Gamma, Weibull, Log-logistic) **2.2** Linear failure rate **2.3** Parametric inference (point estimation, confidence intervals, scores) **2.4** Likelihood ratio test **2.5** Accelerated failure time model **2.6** Cox-Snell residuals.

4. **UNIT 3**. **3.1** Identification of Prognostic Factors Related to Survival Time **3.2** Cox's proportional hazards regression model with one and several covariates **3.3** Rank test for the regression coefficients **3.4** Adequacy Assessment of the Proportional Hazards Model **3.5** Time dependent Extension of the Cox model **3.6** Tests with non-proportional hazards **3.7** Parametric and nonparametric inference for this model.

5. **UNIT 4. 4.1** Parametric Methods for Comparing Two Survival Distributions **4.2** Likelihood Ratio Test for Comparing Two SurvivalDistributions **4.3** Comparison of

Two Exponential Distributions **4.4** Comparison of Two Weibull Distributions **4.5** Comparison of Two Gamma Distributions**4.6** Preliminary Examination of Data **4.7** General Structure of Parametric Regression Modelsand Their Asymptotic Likelihood Inference **4.8** Partial Likelihood Function for Survival Times **4.9** Identification of Significant Covariates,

1. **Recommended Text Books :**

Elisa T. Leeand John Wenyuwang (2003)Statistical Methods for Survival Data Analysis. John Wiley & Sons

2. **Recommended ReferenceBooks :**

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 MKB.(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

David G. Kleinbaum, Mitchel Klein,(2005) Survival Analysis- A Self-Learning Text; Springer

ELECTIVES

BUNCH A

ST 86 04 01 CLINICAL TRIALS ANDBIOINFORMATICS

Total Credits: 3

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Controlled Clinical Trials & Bioinformatics & their applications in Biostatistics

2. UNIT 1. **1.1** Introduction to clinical trials **1.2** Participants and Sponsors of a trial **1.3** Informed consent **1.4** Benefits and risk of participating in a trial **1.5** Blinding **1.6** Placebo **1.7** Controlled and Uncontrolled trials **1.8** Need and Ethics of clinical trials **1.9** Objectives and End-points of a clinical trial **1.10** Single center and Multi-center trials **1.11** ICH and GCP **1.12** FDA and EMEA guidelines **1.13** Drug Development Process **1.14** Overview of phase I-IV trials(Design and analysis)

3. **UNIT 2. 2.1** Clinical trial study designs **2.2** Bioequivalence trials **2.3** Adaptive trials **2.4** Sample size determination **2.5** Randomization methods **2.6** Handling of missing data **2.7** Handling multiplicity **2.8** Clinical data Management (CDM) **2.9** Understanding protocol **2.10** Clinical study report **2.11** Statistical analysis plan(SAP) **2.12** Data visualization methods **2.13** Data Comprehension **2.14** Data Interpretation **2.15Clinical Data Analysis:** Analysis methods/models for Continuous data **2.16** Categorical data **2.17** Binary data **2.18** Survival data **2.19** Parametric and Non-parametric methods **2.20** Sub-group Analysis **2.21** Sensitivity analysis **2.22** Interim analysis **2.23** Quality of life data analysis

4. **Unit 3:3.1**Basics of Bioinformatics **3.2** Definition, importance and role of bioinformatics in life sciences**3.3** Computational Biology **3.4.** Biological databases: Nucleotide sequence databases (NCBI- GENBANK, DDBJ and EMBL) **3.5** Protein databases - structure and sequence databases (PDB, SWISSPROT and UNIPROT).

5. **Unit 4:4.1**Introduction to Sequences alignments: Local alignment and Global alignment, Pair wise alignment (BLAST and FASTA] and multiple sequence alignment. **4.2**Phylogenetic Tree construction and Analysis Molecular visualization software - RASMOL.**4.3** Basic concepts of Drug discovery pipe line, computer aided drug discovery and its applications.**4.4** Human Genome Project.

1. **Recommended Text Books :**

Friedman L.M.,Furberg C.D. &Demets D.L.(1998). Fundamentals of clinical trials, Springer

Scott Evans , Naitee Ting, Fundamental Concepts for New Clinical Triallists, Chapman & Hall Book

Rajan S S and Balaji R, (2002), Introduction to Bioinformatics, Himalaya Publishing House

2. **Recommended ReferenceBooks :**

Shein-Chung Chow and Jen-Pei Liu(2004). Design and Analysis of Clinical Trials: Concepts and Methodologies (2nd edition) Wiley-Interscience

Stuart J. Pocock (2010)Clinical Trials – A practical approach (Reprint), John Wiley

Stephen Senn (2009) Statistical Issues in Drug Development (2nd edition), John Wiley

David Collett (2003) Modeling Binary Data (2nd edition), Chapman & Hall/CRC

Alan Agresti (2002) Categorical Data Analysis (2nd edition), Wiley-Interscience

Waterman, M.S. (2000). Introduction to Computational Biology, CRC Press

ST 86 04 02 DEMOGRAPHY AND VITAL STATISTICS

Total Credits: 3

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills inDemographyAnd Vital Statistics& their applications in Biostatistics

2. **UNIT 1. Methods of Demographic / Population Analysis: 1.1** Source of population data **1.2** India- Census **1.3** National Sample Survey(NSS)**1.4** Demographic Surveys and other sources **1.5** Data appraisal **1.6** Rates of population growth: arithmetic,geometric and exponential rates of growth **1.7** Lexis diagram **1.8** Cohort and cross sectional indicators **1.9** Crude rates and standardized methods **1.10** Methods of population projections **1.11** Inter–Censual/Post–censual estimates of populations ,ratios, proportions, percentages ,persons, months/years, incidence **1.12** Prevalence.

3. **UNIT 2. Population Composition and Change, Nuptiality and Fertility: 2.1** Spatial and temporal changes in size **2.2** Composition and distribution of population-global perspective with special focus on India.Composition of India's population (Demographic, Social , Economic, Cultural) **2.3** Concept of ageing **2.4** Concept and measures of Nuptiality and Fertility **2.5** Levels **2.6** trends and differentials of fertility in India **2.7** Sources of data on fertility **2.8** Determinants of fertility-Framework of fertility analysis- Davis' &Blake's intermediate variable framework of fertility **2.9**Bongaart's proximate determinants of fertility.

4. **UNIT 3. Mortality , Morbidity and Health: 3.1** Mortality – concept and measures **3.2** Mortality trends **3.3** Levels and determinants in India with special reference to infant mortality and maternal mortality **3.4** Cause of death statistics **3.5** Life table(Basic concepts of life table, types and forms of life table and Model life table) **3.6** Concepts and definitions of health and morbidity **3.7** Sources of data on mortality and morbidity

5. **UNIT 4: Migration, Urbanization, Population Theories and Policies: 4.1** Basic concept and definitions **4.2** Types of migrations (internal and international) **4.3** Trends and differentials of migration **4.4** Determinants and consequences of migration **4.5** Concept and definitions of urban **4.6** Trends and patterns of urbanization in India **4.7** Issues in urbanization and urban problems in developing countries with focus on India **4.8** Theories of population growth(Malthus to modern; limits to population growth) **4.9**Theory of demographic transition **4.10** Theories related to fertility, migration and urbanization **4.11** Population policies in the con text of growth, structure, distribution and quality of life**4.12** National and State population policies in India **4.13** Evolution of family welfare programme in India **4.14** Methods of programme Impact Assessment

1. **Recommended Text Books :**

Shrivastava OS, 1998, Demography and Population Studies, Vikas Publishing House Pvt Ltd(Edition 2)

Drogin Richard, 1975, Vital Statistics, Tata Mc Graw-Hill Publishing

Sharma Rajendra K, 2007, Demography and Population Problems, Atlantic publishers and distributions (Edition 1)

Heuveline Patrick, 2001, Demography, Blackwell Publishers Ltd (First Edition)

2. **Recommended Reference Books :**

Pandey Arvind, 2010, Population GenderandHealth in India, Academic Founder

Wolfender H Hugh, 1925, Population Studies and their Compilation, Routledge

CaoleAnsely .J, 1972, Population Growth and Economic Development, Prentice-hall of India Pvt Ltd

Rajhans, 1998, Population Studies, Surjeet Publications

ST 86 04 03 STATISTICAL COMPUTING AND DATA ANALYSIS USING R AND SAS

Total Credits: 3

Total Hours: 25

Weightage: 30

1. **Objective of the course:**

   This course is intended to impart the skills in data analysis and biostatistical computation. It also enables to draw valid conclusions and interpretation of the findings.

2. **Syllabus:**

   This paper includes practical problems using data from Biostatistical contexts based on all courses during Semester IV namely, ST02 401, ST02 402 and Elective Bunch I (ST02 E4.1.1 and ST02 E4.1.2). There will be 4 questions (1from each course) of which 3 are to be answered. Questions shall be based on moderate data sets to illustrate the application of theory. Data Analysis using SAS& Ris expected. The duration of the examination is 3 hours.

   Practical training should be imparted to the students in the computer lab for analyzing data sets using SAS & R. Practical Records are to be submitted by each student.

BUNCH B

ST 87 04 01 ADVANCED CLINICAL TRIALS AND OPERATIONS RESEARCH

Total Credits: 3

Total Hours: 25

Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Advanced Clinical Trials and Operations Research & their applications in Biostatistics

2. **UNIT 1. 1.1 Recruitment of Study Participants of Clinical Trials** 1.2 Data Collection and Quality Control 1.3 Assessing and Reporting Adverse Effects 1.4 Assessment of Health related quality of Life:, uses of HRQL, Methodological issues 1.5 Participant Adherence: Considerations before participant enrollment, Maintaining good participant Adherence, Adherence monitoring.

3. **UNIT 2**. 2.1 Monitoring Response Variables: repeated testing for significance, decision for early termination, decision to extend a trial 2.2 Statistical Methods Used in Monitoring 2.3 Issues in data analysis 2.4 Closeout 2.5 Termination procedures, 2.6 Post-study Follow –up 2.7 data clean-up and Verification 2.8 Storage of Study material

4. **UNIT 3**. 3.1 Reporting and Interpretation of Results 3.2 Guidelines for reporting 3.3 Multicenter trials 3.4 Reasons for multi-center trials 3.5 Conduct of multi-center trials 3.6 Baseline Assessment 3.7 Blindness 3.8 Publishing trial results

5. **UNIT 4**. **Introduction to Operations Research**, 4.1 Linear programming problems (LPP) 4.2 framing an LPP problem 4.3 graphical solution 4.4 feasible 4.5 basic feasible and optimal basic feasible solutions to an LPP 4.6 simplex method 4.7 dual of linear programming 4.8 transportation problems 4.9 assignment problems 4.10 simple numerical problems as illustration.

1. **Recommended Text Books:**

Friedman L.M., Furberg C.D. & Demets D.L.(1998). Fundamentals of clinical trials, Springer

Kanti Swarup, P.K Gupta, Man Mohan, (2010), Operations Research, Sulthan Chand and Sons

2. Recommended Reference Books:

Scott Evans , Naitee Ting, Fundamental Concepts for New Clinical Triallists, Chapman & Hall Book

Shein-Chung Chow and Jen-Pei Liu(2004). Design and Analysis of Clinical Trials: Concepts and Methodologies (2nd edition) Wiley-Interscience
P.K Gupta and Dr.D.S Hira, Operations Research, S.Chand

ST 87 04 02 QUALITY CONTROL AND RESEARCH METHODOLOGY

Total Credits: 3
Total Hours: 25
Weightage: 30

1. **Objective of the Course:** To impart basic knowledge & skills in Quality Control and Research Methodology& their applications in Biostatistics

2. **UNIT 1**. **1.1Introduction to Quality Control and related concepts** 1.2 ISO Certification 1.3 Six-sigma 1.4 Statistical process control 1.5 Theory of control charts 1.6 Shewhart control charts for variables - $\bar{x}$, R, sigma charts 1.7 Attribute control charts - p, np, c charts 1.8 Modified control charts 1.9 Statistical process control.

3. **UNIT 2**. 2.1 O.C. and ARL curves of control charts 2.2 Moving average control charts 2.3 EWMA charts 2.4 CUSUM charts 2.5 Process capability analysis 2.6 Process capability indices

4. **UNIT 3**. **3.1 Introduction to Research Methodology** 3.2 Meaning and importance of research 3.3 Types of research 3.4 Selection and formulation of research 3.5 Research design 3.6 Development of a research plan 3.7 Analysis of literature review 3.8 Research methods.

5. **UNIT 4.** 4.1 Structure and components of scientific reports 4.2 Types of report 4.3 Different steps in preparation 4.4 Layout, structure and languages of typical reports, bibliography, referencing and foot notes 4.5 Oral presentation 4.6 Preparing research papers for journals, seminars and conferences 4.7 Intellectual rights and patent laws 4.8 Reproduction of published material, Plagiarism, Citation and Acknowledgement 4.9 Calculations of impact factor of a journal, citation index, ISBN and ISSN 4.10 Preparation of Research/Project Proposal.

1. Recommended Text Books:

Douglas C and Mongomery, (2005) Introduction to Statistical Quality Control, Wiley

Sinha , S.C. and Dhtman, A.K., (2002). Research Methodology, Ess Ess Publications.

Day RA (1992) How to write and publish a Scientific paper. Cambridge University press, London

2. Recommended Reference Books:

Grant E.L and Leavenworth R.S. (1980) Statistical Quality Control, McGraw Hill

Ott E.R. (1975) Process Quality control, McGraw Hill

Kothari, C.R. (2008). Research Methodology: Methods and Techniques, New Age International Publishers, New Delhi

ST 87 04 03 STATISTICAL COMPUTING AND DATA MANAGEMENT USING R AND SAS

Total Credits: 3

Total Hours: 25

Weightage: 30

1. **Objective of the course:**

   This course is intended to impart the skills in data analysis and biostatistical computation. It also enables to draw valid conclusions and interpretation of the findings.

2. **Syllabus:**

   This paper includes practical problems using data from Biostatistical contexts based on all courses during Semester IV namely, ST02 0401, ST02 0402 and Elective

Bunch II (ST02 E4.2.1 and ST02 E4.2.2). There will be 4 questions (1from each course) of which 3 are to be answered. Questions shall be based on moderate data sets to illustrate the application of theory. Data Analysis using SAS & R is expected. The duration of the examination is 3 hours.

Practical training should be imparted to the students in the computer lab for analyzing data sets using SAS & R. Practical Records are to be submitted by each student.

ST02 04 03 PROJECT WORK AND ON THE JOB TRAINING (IN A REPUTED INDUSTRY/RESEARCH INSTITUTE/MEDICAL COLLEGE)

Total Credits: 3

Project work shall be executed by working outside the regular teaching hours under the supervision of a reputed researcher/ scientist/ an expert faculty in a reputed research institute/ medical college/industry. At the end of the Project work, the candidate has to submit 3 copies of the Project Report consisting of the Title of the Study, Objectives,Review of Literature, Materials and Methods, Analysis of Data, Presentation of Results, Applications/ Conclusions, Referencesetc of about 40-50 pages. The guidelines for evaluating the Project Report will be issued by the University. There will be an internal assessment and external assessment for the project work. The external evaluation of the Project work is followed by a presentation based on the work and comprehensive Viva-Voce.

ST02 04 04 COMPREHENSIVE VIVA-VOCE

Total Credits: 3

In order to assess the overall knowledge in theory and applications as well as general understanding of the different courses studied as part of the program, a Comprehensive Viva-voce shall be conducted at the end of the fourth semester of the program and it shall cover oral questions from all courses in the program. The Viva-Voce board will consist of an outside expert, the Chairman of the Board of Examiners and the Head of the concerned Department.

MODEL QUESTION PAPERS

Model Question Paper

QP Code (To be assigned by Exam Section) **Reg. No. …….**

Name ……

M Sc (Biostatistics) Degree (C.S.S) Examination, …….

First Semester

Faculty of Science

ST 02 0101 Statistical Methods and Probability Distributions

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. What is a secondary data? What are the sources of it?
2. What is median of sample data? What is its advantage over mean and in what situations?
3. Define skewness of a distribution. What are the various measures of skewness?
4. Give axiomatic definition of probability.
5. Distinguish between Discrete and Continuous random variables with suitable examples
6. Define a Poisson distribution. Find its mean and variance and discuss its skewness.
7. What is Weibull distribution? Mention its application in biostatistics.
8. Define Logistic distribution and mention its applications.
9. Explain any two uses of '*t*' distribution.
10. State the central limit theorem and illustrate it with an example.

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. (a) What is a primary data? What are the methods of collecting primary data?
    (b) Design a questionnaire to assess the incidence of cancer in a locality.

12. (a) Discuss the need for the diagrammatic representation of data in biostatistical studies. (b) What are histograms and Ogives? Discuss the construction and uses of them in biostatistics.
13. (a) Describe the important measures of dispersion.

    (b) Explain a Lorenz curve and point out its applications.
14. There are 9 boys and 9 girls in a class. Their heights are given below :

    Boys (height in cm.) : 170,175,165,162,168,170,172,163,161

    Girls (height in cm.) : 152,165,155,161,168,160,155,154,152

    Calculate variance of height for boys and girls separately. Which is more consistent?
15. State addition and multiplication theorems of probability. Suppose two doctors D1 & D2 test all cases coming into a pulmonary medicine OPD for Tuberculosis (TB).Let us define two events A&B as: A= {doctor D1 makes a positive diagnosis for TB} and B={doctor D2 makes a positive diagnosis for TB}.If doctor D1 diagnoses 12 % of all patients as positive, doctor D2 diagnoses 15 % of all patients as positive, and both doctors diagnoses 9% of all patients as positive for TB, what is the probability that either doctor makes a positive diagnosis for TB?
16. Show that binomial distribution and negative binomial distribution tends to Poisson distribution under certain conditions to be stated.
17. State Weak law of large numbers and its applications.
18. Define chi-square, *t,* F tests and bring out the relations connecting them.

(**6 x 2 = 12**)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. (a) State the Bayes' theorem and mention its applications.

    (b) Only 1 in 1000 adults is afflicted with a rare disease for which a diagnostic test has been developed. The test is such that when an individual actually has the disease, a positive result will occur 99% of the time, whereas an individual without disease will show a positive test result only 2 % of the time. If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease?
20. (a) Define binomial distribution and discuss about the symmetry and kurtosis of the distribution.

    (b) Five individuals from an animal population thought to be near extinction in a certain region have been caught, tagged and released to mix into the population. A

random sample of 10 of these animals is then selected. Suppose X denotes the number of tagged animals in the second sample. If there are actually 25 animals of this type in the region, what is the probability that (i) $X = 2$ (ii) $X \leq 2$?

21. (a) State the central Limit Theorem(CLT).Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution where X_i 's are positive. Using CLT prove that, for large n, the product $Y= X_1X_2 \ldots X_n$ has approximately a log normal distribution.

    (b) It is known that a seed of certain plant germinate with probability 2/3. If 100 seeds are planted, obtain an upper bound of the probability that the number of plants germinating will differ from the expected number by more than 10, using Chebychev's inequality.

22. (a) What are the different measures of central tendency? Explain their merits and demerits.

    (b) Describe the important steps in a biostatistical study on cancer.

(2 x 5 = 10)

Model Question Paper

QP Code (To be assigned by Exam Section) **Reg. No. …….**

Name ……..

M Sc (Biostatistics) Degree (C.S.S) Examination, …….

First Semester

Faculty of Science

ST 02 01 02 Theory and Methods of Sample Surveys

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. What are the advantages of sample survey in comparison with a census survey?
2. Under SRSWR from a finite population what are the basic properties of the estimator of the population mean for a variate X based on distinct elements in the sample.
3. Explain the situations where systematic sampling can be used with advantage.

4. Give a brief account of the problem of formation of strata.
5. Define a ratio estimator and give an example.
6. Explain the randomized response technique.
7. Define cluster sampling and present an example based on biostatistical experiments.
8. What is double sampling? Under what conditions it is used?
9. Obtain an unbiased estimator of the population mean in the case of PPS sampling.
10. What is non probability sampling?

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Explain the principal steps involved in the organization and conduct of a sample survey.
12. (a) Stating the assumptions clearly describe how you would determine the sample size in simple random sampling.

(b) Show that in SRSWR the sample mean square is an unbiased estimate of the

Population variance.

13. Define allocation. Describe proportional allocation in stratified sampling and obtain the expression for sample size and also derive its variance.
14. With the help of an example, explain PPSWR.
15. Briefly explain multiphase sampling with the help of an example from health science.
16. Write a short note on Des Raj ordered estimator in PPSWOR.
17. Distinguish between probability and non-probability sampling.
18. Compare the efficiency of regression estimate with that of ratio estimate.

(6 x 2 = 12)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. (a) Describe the use of auxiliary information in sample surveys with an example

(b) Explain Warner's model related and unrelated questionnaire methods of randomized Responses.

20. Obtain the optimum total size of the sample and its allocation in various strata when:

(a) Total cost is fixed.

(b) The variance is fixed and the cost of surveying a unit differs from stratum to stratum.

21. Explain: (i) Convenient sampling (ii) Non- random sampling and present examples. Bring out their importance in Biostatistical surveys
22. What is meant by Cluster Sampling? Derive the expression for the variance of the Estimator of the population mean in terms of intra-cluster correlation. State the situation under which you recommend it over simple random sampling.

(2 x 5 = 10)

Model Question Paper

QP Code (To be assigned by Exam Section) **Reg. No. ……**

Name

…….

M Sc (Biostatistics) Degree (C.S.S) Examination, …….

First Semester

Faculty of Science

ST 02 01 03 STATISTICAL PROGRAMMING IN R & PYTHON

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. Explain the data objects list and vectors in R with an example
2. What do you mean by built-in functions in R?
3. Write the syntax for calculating mean, illustrate it with an example
4. What are the advantages of using R program in Biostatistics?
5. How will you declare a variable in Python? What are the rules for naming a variable?
6. Explain the data types in Python.
7. Write the syntax for drawing a histogram in Python, illustrate it with an example

8. How will you draw a box plot in python?

9. Explain for loop in R

10. What are the important features of Python?

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Briefly explain the data types in R

12. Explain if and if-else statement in R

13. Explain the computation of Descriptive Statistics using an example in R

14. Write a R program for simulating 50 observations with Normal distribution N(10,1)

15. Explain functions in Python

16. Explain Python Loops

17. Explain the syntax for drawing Line diagram, bar diagram and Scatter plot

18. Write a python program for simulating 50 observations from Normal N(15, 1)

(6 x 2 = 12)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. Explain R Loops with illustrative examples.

20. Write the syntaxes for generating samples from different distributions with examples.

21. What are the decision making statements in python?

22. Write the Python Syntaxes for descriptive statistics with examples.

(2 x 5 = 10)

Model Question Paper - Format

QP Code (To be assigned by Exam Section) **Reg. No. ……**

Name ……..

M Sc. Degree (C.S.S) Examination, …….

First Semester

Faculty of Science

ST 02 01 04 STATISTICAL GENETICS AND ECOLOGY

(2019 admissions onwards)

Time: Threehours Max. Weight: 30

Section- A

(Answer any **eight**questions. Each question carries a weight of 1)

1.Define mutation with example

2.What are the indirect methods of abundance estimation.

3.What are X-linked genes? Explain with example.

4. Write a note on assortative mating

5. How does Simpson's index help in abundance estimation.

6. Define dominant/co-dominant cases in genetics

7. Explain Mendel's laws.

8. What are the properties of evolutionarily stable strategy?

9. How does Shannon-Wiener index help in abundance estimation?

10. Explain re-combination fraction.

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11.Discuss on Hardy-Weinberg equilibrium.

12.What are the approaches to equilibrium for X-linked gene.

13.Explain Hawk-Dove game and Prisoner's dilemma and its relevance in ecology.

14. Explain the concept of linkage and recombination in genetics and its detection.

15. What is species abundance curve? How is it constructed?

16. Explain the concept of Non-random mating and inbreeding

17. Explain monotone and bath tub shaped hazard rates.

18. What are the methods in genetic mapping?

(**6 x 2 = 12**)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19.Write a note on pedigree data using Elston-Stewart algorithm for calculation of likelihood.

20. Explain evolutionarily stable strategy and its properties.

21. Define Leslie matrix model for age and stage structured population.

22. Explain two species interaction with Lotka-Volterra equations.

(**2 x 5 = 10)**

Model Question Paper - Format

QP Code (To be assigned by Exam Section) **Reg. No. …….**

Name

…….

M Sc Degree (C.S.S) Examination, ……

First Semester

Faculty of Biostatistics

ST 02 01 05 STATISTICAL DATA ANALYSIS USING MICROSOFT EXCEL, R AND PYTHON

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

(Answer any **three** questions. Each question carries a weight of **ten**)

1. (a) Find the first three central moments for the following data and find its skewness coefficient.

x_i	0	1	2	3	4	5	6	7	8	9	10
Frequency	3	16	26	17	18	9	4	4	0	1	4

(b) Also fit a Binomial distribution to the above data and compute the expected frequencies. Is it a good fit?

(c) 5 individuals from an animal population to be near extinction in a certain region have been caught, tagged and released to mix into the population. After they have had an opportunity to mix, a random sample of size 10 of these animals is selected. There are actually 25 animals of these types in the region. Find the probability that

(i) There are exactly 2 tagged animals in the second sample

(ii) The number of tagged animals in the second sample is less than or equal to 2.

2. (a) Families in a certain city are divided in six income groups. Basic data for these groups are given below :

Income group	No of families (N_h)	Average Income (T_h)	Standard Deviation (σ_h)	Cost per unit in the strata (C_h)
1	534	486	250	8
2	465	236	260	3
3	365	330	380	4
4	768	440	320	2
5	544	606	374	1
6	297	860	520	4

For a sample of 300 families compute the sample size in each stratum under :

(i) Proportional Allocation

(ii) Optimum allocation- assuming equal costs for all stratum and

(iii) Optimum allocation with given costs C_h.

(b) The following data gives the number of inhabitants for the year 1991(x-value) and 2001 (y-values) for each of the 10 villages included in SRSWOR sample from group of 198 village of a district. The total number of inhabitants in all 198 villages in 1991 was 1, 04,000. Obtain an estimate for the total number of inhabitants in 2001 of all villages using:

(i) Ratio estimator

(ii) Regression estimator. Also compute the standard error of your estimate.

Sl No	1	2	3	4	5	6	7	8	9	10
X	78	48	138	50	29	39	381	45	23	74
Y	80	75	143	38	50	54	464	53	48	93

3. (a) Write a R program to simulate 50 observations from normal (5,2) and find the mean and variance for the following data
11,15,55,24,43,32,31,43,10,40
(b) Write a python program to find the arithmetic mean and variance for the following data and also draw a histogram for the given data.
44,55,35,25,83,29,92,4,5,23,23,56,6,43,25,67,34,345,67,34,45,67,74,43,35,67,34,34,45,46,99.

4. a) Five of the self-fertilized plants that the Mendel observed for segregation of yellow and green seed color plants, showed the following result among the seeds.

Plant	Yellow seed	Green seed
1	45	12
2	28	8
3	25	8
4	20	10
5	33	12
Total	151	30

Test the homogeneity of the five plants for 3:1 ratio

b) Plant species studied by BSGN in a pilot study gave the following data.

Habitat	Species in the sample and its distribution	Sample size
Trimbak	10 [55,6,4,4,2,6,8,2,2,3]	90
Nandur	12 [100,12,8,5,5,5,5,4,2,2,1,1]	150
Torangan	6 [30,10,5,3,1,1]	50

Obtain diversity indices of the above three habitats and give your comments.

Model Question Paper

QP Code (To be assigned by Exam Section) **Reg. No. …….**

Name

……..

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Second Semester

Faculty of Science

ST 02 02 01 Matrix Algebra and Regression Analysis

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. Define linear independence of a set of vectors and give an example.
2. Explain the eigen values and eigen vectors of a square matrix.
3. What is the principle of least squares?
4. Explain the role of residual analysis in regression theory.
5. Explain "multi-collinearity" in linear regression.
6. State the general linear model and the assumptions underlying it.
7. Explain the Poisson regression.
8. How do you detect the outliers in statistical analysis?
9. Describe a non-parametric regression model.
10. Explain correlation ratio and its uses.

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. (a) Define linear dependence of a set of vectors. Check the linear dependence or independence of the following set of vectors.

$V_1=(1,2,1)$; $V_2=(1,-1,1)$; $V_3=(3,3,3)$

(b) Distinguish between the regular inverse and generalized inverse of a matrix. Explain the need and applications of a generalized inverse

12. (a) Show that the characteristic vectors associated with distinct characteristic values of a matrix are linearly independent.

(b) Examine the definiteness of the quadratic form:

$$6x^2 + 3y^2 + 14z^2 + 4yz + 18xz = 4xy$$

13. (a) Distinguish between multiple and partial correlation coefficient. What are their practical applications?

(b) Explain the role of residual analysis in regression problems

14. Write short notes on:

(a) Polynomial regression; and

(b) Power transformations for dependent variable in linear regression

15. With the help of suitable examples discuss random and mixed effects models.

16. Explain the normal probability plot.

17. Give an account on non-parametric regression and generalized linear models.

18. Describe homogeneity of variances.

(**6 x 2 = 12**)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. (a) Solve the following system of equations using the theory of matrices :

$$3x + 2y + 7z = 4$$
$$2x + 3y + z = 5$$
$$3x + 4y + z = 7$$

(b) Explain the concept of Moore Penrose g-inverse. Show that it is unique. When is it used?

20. Explain the analysis of binary and grouped data by using logistic models.

21. (a) Explain the concept of rank correlation. When is it used?

(b) Derive the least square estimators of simple linear regression and prove that they are unbiased.

22. What is a non-linear regression model? Discuss any two methods of estimation in this model. Comment on the properties of the estimators in each case.

(2 x 5 = 10)

Model Question Paper

QP Code (To be assigned by Exam Section) **Reg. No. ……**

Name

……

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Second Semester

Faculty of Science

ST02 02 02 SAMPLING DISTRIBUTIONS & STATISTICAL ESTIMATION METHODS

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. What is the concept of random sample?

2. Define statistic and parameter

3. What is Standard Error?

4. What do you mean by Statistical Inference?

5. Define parametric space and sample space

6. Differentiate between estimate and estimator

7. What do you mean by Fisher information?

8. What is complete sufficient statistic?

9. Write a short note on Bootstrap method

10. What are the Properties of MLE?

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Define Chi-square Distribution and explain it's applications

12. Briefly explain any two uses of F distribution

13. What are the desirable properties of a good estimate? Give an illustrative example ?

14. What do you mean by an unbiased estimator? A random samples X_1 , X_2,…,X_5 is drawn from a normal population with unknown mean μ and σ^2 . Check whether sample mean is a unbiased estimator

15 Distinguish between one parameter and multi-parameter exponential family.

16. Explain UMVUE with example

17. Explain the method of minimum chi-square

18. Write a note on Small sample and large sample confidence interval

(**6 x 2 = 12**)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. Define Student's t, F and Chi-square distributions and write down their sampling distribution. State the important assumptions in respect of them

20. i) Obtain a unbiased, consistent and sufficient estimate for μ in a normal distribution with variance 1

    ii) Obtain the parameter p in Binomial distribution by the method of moments

21. What are the methods of estimation?

22. Derive (100-α)% confidence limits for the parameter λ of the Poisson distribution. Write the 95% and 99% confidence intervals f or the parameter λ

(2 x 5 = 10)

Model Question Paper

QP Code (To be assigned by Exam Section) **Reg. No. ……**

Name

……

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Second Semester

Faculty of Science

ST 02 02 03 PARAMETRIC AND NONPARAMETRIC TESTS

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. Distinguish between Type I and Type II errors.
2. Distinguish between simple and composite hypothesis with suitable examples.
3. Briefly explain uniformly most powerful test.
4. Explain the Kolmogrov Smirnov goodness of fit test.
5. Describe Wald's SPRT.
6. Write two examples where Chi-square is applied in health science research.

7. Explain OC function and ASN function and their applications.
8. Give a situation where you applied paired t test.
9. What is a P-value?
10. What is likelihood ratio test?

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. What do you mean by non-parametric tests? Enumerate the advantages and disadvantages of nonparametric test over parametric tests
12. Briefly explain Friedman's test with an example.

.

13. State Neymann – Pearson lemma and explain its application in testing of hypothesis.
14. Obtain the uniformly most powerful test for testing H_0: $\mu=\mu_0$ Vs H_1 : $\mu=\mu_1$ where $\mu_1 > \mu_0$ when the variance $\sigma^2 = 1$ in a normal population.
15. Comment on test for equality of variances of two normal populations.
16. Explain the test for equality of means of two populations stating the assumptions.
17. Explain Wald's SPRT with illustrations.
18. Discuss OC and ASN functions for tests regarding binomial population.

(**6 x 2 = 12**)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. Describe Kruskal Wallis test with suitable examples.
20. With the help of examples explain how do you perform independent sample t test and Mann Whitney U test.
21. Let $X \sim N(\mu, \sigma^2)$ where σ^2 is known. For testing H_0: $\mu=\mu_0$ Vs H_1 : $\mu=\mu_1$, construct the SPRT.
22. Obtain LRT for testing $H_0 : \Theta=\Theta_0$ Vs $H_1 : \Theta > \Theta_0$ based on a random sample of size $N(\Theta,\sigma^2)$ where σ^2 is unknown.

(2 x 5 = 10)

Model Question Paper

QP Code (To be assigned by Exam Section) **Reg. No. …….**

Name

……

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Second Semester

Faculty of Science

ST 02 02 04 EPIDEMIOLOGY AND STUDY DESIGNS

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. What is epidemiology?

2. What is cumulative risk?

3. What are the measures of disease frequency?

4. Who you mean by exposure?

5. What are the absolute measures of effect?

6. Define the terms (a) Incidence time (b) Incidence rate.

7. What is ecological fallacy?

8. What are different types of exposures?

9. What are disease registries?

10. Explain 'measures of outcome'

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Give a brief note on basic method and analysis of grouped data.

12. What are the indirect methods of standardization?

13. Note on repeated measurement analysis.

14. What are the relative absolute measures of effect?

15. What re age standardization? Explain with example.

16. What are the different classifications of diseases?

17. Explain cumulative risk and proportional incidence.

18. Give the confidence intervals and significance test for measures of occurrence and effect.

(6 x 2 = 12)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. Write a note on different types of cohort study and discuss on planning and execution.
20. What are longitudinal studies? What is the analysis used in longitudinal studies?
21. Explain sample size calculation in the case of different study designs.
22. Describedirect and indirect methods of standardization.

(2 x 5 = 10)

Model Question Paper - Format

QP Code (To be assigned by Exam Section) **Reg. No. …….**

Name

…….

M Sc Degree (C.S.S) Examination, ……

Second Semester

Faculty of Biostatistics

ST 02 02 05 Statistical Data Analysis Using SPSS, R and Python

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

(Answer any **three** questions. Each question carries a weight of **ten**)

1. (i) The following table shows the weight(KG), total Cholesterol (Tot_chol mg/100ml) and triglyceride(Trigly mg/100ml) in 10 patients with primary type II hyperlipoproteinemia just before treatment.

Sl. No	1	2	3	4	5	6	7	8	9	10
Weight	76	97	83	52	70	67	75	78	70	99
Tot_cho	302	336	220	300	382	379	331	337	426	399
Trigly	139	101	57	56	113	42	84	186	164	205

Compute the partial and multiple correlation coefficients and test for significance at 5%.

(ii) Fit a regression model to predict birth weight (KG) based on mothers weight (KG) at first trimester of pregnancy and mothers age (years) for the data given below

Sl. No	1	2	3	4	5	6	7	8	9	10
Birth weight	2.45	2.23	2.65	2.39	2.45	2.48	2.6	2.43	2.55	2.75
Mothers weight	44	45	51	42	52	50	60	51	41	65
Mothers age	23	27	29	25	24	25	25	21	25	20
Sl. No	11	12	13	14	15	16	17	18	19	20
Birth weight	2.4	2.54	2.56	2.64	2.73	3.4	2.6	2.73	2.75	2.93
Mothers weight	46	46	49	55	62	75	69	60	62	70
Mothers age	19	24	23	25	26	20	22	19	22	22

2. (i) Determination of saliva pH levels were made in random sample of 15 seventh standard school children. The results were as follows

   7.14 7.11 7,61 7,98 7.21 7.16 7.89 7.24 7.86 7.47 7.82 7.37 7.66 7.62 7.65

   Assuming that the saliva pH levels follows normal distribution, estimate the mean and variance by method of moments

   (ii) The following data gives the number of mistakes per page observed in a book which follows Poisson distribution.

No. of mistakes	0	1	2	3	4
No. of pages	211	90	19	5	0

   Find the estimate of the parameter by maximum likelihood method

   (iii) Check unbiasedness and consistency of the parameters for both cases ((i) and (ii))

3. (a) In a study pf complaints of fatigue among men with brain injury (BI) , the researchers obtain depression scores from three different samples of subjects. The result of score for three samples were as follows:

   BI & fatigue: 46, 55, 51, 36, 51, 45, 54, 51, 78, 54, 51, 38, 56

   BI & no fatigue: 39, 44, 58, 29, 40, 48, 65, 41, 46

   Normal controls: 36, 34, 41, 29, 31, 26, 33

   Perform an appropriate non parametric test to find whether the depression score differ significantly between the three groups. Report appropriate statistics for this analysis.

   (b) The following sequence of observations are taken from a normal distribution with mean μ and variance 20, -32, 9, -49, 23, -26, -23, -26, -6, 15, -15, 34, -34. Test the null hypothesis $H_{0:}\mu=\mu_0$ Vs $H_{1:}\mu=\mu_{1,}$ constructing an SPRT of strength (0.05, 0.10) and plotting the observations on a graph. Obtain the ASN under H_0 and H_1, also obtain OC curve for this test.

4. Consider the matched case control study of oral contraceptives and congenital heart disease. Assuming the proportion of the exposed controls is p0=0.3 and α=0.05 and β=0.1. In order to detect an increased two folded risk (R=2) with these specifications. Calculate the sample size and power of the study.

Model Question Paper

QP Code (To be assigned by Exam Section) **Reg. No. ……..**

Name ……..

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Third Semester

Faculty of Science

ST 02 03 01 Design and Analysis of Experiments

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. Define Design of Experiments with an example.

2. Explain Gauss-Markov theorem.

3. Give an example to Latin Square design in Biostatistics.

4. What are the advantages and disadvantages of Latin Square design?

5. What is the difference between Complete and Incomplete Block Design?

6. Write a short note on Hierarchical Design.

7. Briefly explain the concept of confounding.

8. Write a note on Factorial Experiments

9. Define Nested Designs

10. Explain Split-plot Designs

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Explain one-way and two-way ANOVA and its applications

12. Explain the basic principles of experimentation

13. With the help of an example explain RBD

14. Three drugs A, B and C are tested in a Randomized block designs with four replications

Drugs	**Blocks**			
	I	**II**	**III**	**IV**
A	6	4	8	6
B	7	6	6	9
C	8	5	10	9

State all Assumptions, Analyze the data and give your conclusions

15. Discuss Analysis of Covariance and its applications in experimental designs

16. Discuss the design and model ANOVA for BIBD

17. Explain Yate's Method with an example.

18. Distinguish between complete and partial confounding. Give an example.

(6 x 2 = 12)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. Explain CRD under the headings design, model, assumptions, analysis, advantages and disadvantages.

20. Explain missing data technique in LSD with an example

21. Explain in detail 2^n factorial design and its analysis.

22. Write a note on Incomplete block Design and its ANOVA.

(2 x 5 = 10)

Model Question Paper

QP Code (to be assigned by Exam Section) **Reg. No. …….**

Name

……..

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Third Semester

Faculty of Science

ST 02 03 02 Stochastic Models and Time Series Analysis

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. Distinguish between recurrent and transient states.
2. How do you obtain the *n*–step transition probabilities of a Markov chain.
3. Explain the stationary distribution of a Markov chain.
4. Show that the inter arrival times in a Poisson process are i.i.d random variables following exponential distribution.
5. Define a weakly stationary process and give an example.
6. If the renewal function is M(t)=2t, then obtain the corresponding renewal process .
7. Show that an AR(1) model can be represented as a MA model.
8. Let $\{X_n\}_{n\geq 0}$ be a discrete time branching process with the offspring p.g.f. as^2+bs+c. Find the probability of ultimate extinction.
9. Define time series, auto correlation and correlogram.
10. Discuss on Exponential AR(1) model.

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Define a Markov Chain and give an example. Derive the Chapman-Kolmogrov equationfor a Markov chain.

12. Explain the postulates of a pure birth process. Derive the differential equations for a Yule-Furry process.

13. (a) State the postulates of Poisson process and obtain its probabilistic structure.

(b) Show that for a Poisson process

$$\{N(t), t \geq 0\}, P\{N(s) = k / N(t) = n\} = \binom{n}{k}\left(\frac{s}{t}\right)^k \left(1-\frac{s}{t}\right)^{n-k}, K = 0,1,2,\ldots\ n\ (s < t).$$

14. Obtain the probability of ultimate extinction in the Galton-Watson Branching process with geometric offspring distribution $P_k=pq^k$, k=0,1,2….($0<p,q<1, p<q=1$)

15.(a) Discuss the various classification of states of a Markov chain (MC).

(b) Classify the states of a MC with the following one–step transition probability matrix:

$$\begin{bmatrix} 0.8 & 0 & 0.2 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0.3 & 0.4 & 0 & 0.3 \end{bmatrix}$$

16. Define MA(q) process. Obtain the ACF of an MA(3) process.

17. Show that in a branching process $\varphi_{n+1}(s) = \varphi_n(\varphi(s))$ under usual notations.

18. Obtain autocorrelation function of AR(1) model.

(6 x 2 = 12)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. Define (i) an ergodic chain and (ii) stationary distribution of a MC, and show that the stationary distribution exists uniquely for an ergodic chain. What happens when the chain is doubly stochastic?

20. Define Birth and death process. Obtain the Kolmogrov forwarddifferential equation for this process. Explain how this is applied to obtain the stationary solution of a M/M/1 queue.

21. (a) State and prove elementary renewal theorem

(b) Prove that probability of extinction is the smallest positive root of $\varphi(s) = s$. Derive the condition that ultimate extinction is sure.

22. Explain in detail ARIMA models and Box-Jenkins approach to time series analysis and modeling.

(2 x 5 = 10)

Model Question Paper

QP Code (To be assigned by Exam Section) **Reg. No. ……..**

Name

……..

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Third Semester

Faculty of Science

ST 02 03 03 APPLIED MULTIVARIATE ANALYSIS

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. Briefly discuss the importance of multivariate statistical analysis
2. If every component of the vector $X= (X_1, X_2, X_3\ldots)'$ has a normal distribution, does it follow that the distribution of X is multivariate normal? Justify your answer.
3. State and establish the invariance property of Hotelling's T^2 statistic.
4. Define a Mahalanobies D^2- Statistic and mention its applications.

5. Show that Wishart distribution is the multivariate analogue of χ^2 distribution.
6. Discuss MANOVA with the help of biostatistical examples.
7. Write a short note on Fisher's discriminant function.
8. Explain single linkage method in cluster analysis
9. Describe the use of dimension reduction techniques in biostatistics
10. Define canonical variables and canonical correlation.

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Derive the characteristic function of a p-component normal variate. Show that the random vector X has a multivariate normal distribution if and only if every linear combination of the components of X is univariate normal
12. Obtain the MLE's of μ and Σ in a p-variate normal distribution, $N_P(\mu, \Sigma)$. Show that they are independent. Are they unbiased? Verify.
13. Obtain the characteristic function of a Wishart distribution.
14. Find the distribution of SP matrix.
15. Explain Hotelling's method of extraction of principal components.
16. Explain LRT. Obtain LRT for testing $H_O : \mu=\mu_0$ Vs $H_1 : \mu = \mu_0$ based on a random
    sample of size n from N (μ, σ^2) where μ & σ are unknown.
17. Derive a discriminant function for classification into one of two known multivariatenormal populations
18. Explain Nearest neighbor methods of cluster analysis.

(**6 x 2 = 12**)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. With the help of an example explain two-way MANOVA.

20. What are the goals of Principal Component Analysis (PCA)? Derive the components in PCA.

21. (a) Explain test of symmetry.

(b) Discuss in detail hierarchical clustering methods.

22. What are Canonical correlations? Explain how Canonical correlations can be obtained as roots of certain determinantal equations.

Model Question Paper - Format

QP Code (To be assigned by Exam Section) **Reg. No. ……**

Name

……

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Third Semester

Faculty of Biostatistics

ST 02 03 04 ADVANCED EPIDEMIOLOGY AND BIOASSAYS

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. Define Confounding?

2. What are measurement biases?

3. Define safe doses.

4. What is attributable risk and Excess risk?

5. Write Note on adjust for several confounders.

6. Define median dose.

7. Explain Interaction in epidemiology.

8. What is predictive value?

9. Briefly explain Specificity and Sensitivity.

10. What are the methods in Assessment of confounding?

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11.Explain MH test.

12.Which are the ratio estimators in bioassay?

13.Write a note on quantal responses.

14.What are the different types of biological assays?

15.Explain kappa measure of agreement

16.Explain dose allocation schemes in bioassay.

17. What is attributable risk in Epidemiology?

18.Confidence intervals and statistical tests for adjusted relative measures of effect.

(6 x 2 = 12)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19.Write a note on causation and Hill's criteria.

20.Describe predictive value method for selecting a positivity criterion and receiver operator characteristic (ROC).

21.What are the different regression approaches for estimating dose response relationships and dose allocation schemes?

22.Explain Mantel-Haenszel method to adjust for several confounders.

(2 x 5 = 10)

Model Question Paper - Format

QP Code (To be assigned by Exam Section) **Reg. No. ……**

Name

…….

M Sc Degree (C.S.S) Examination, ……

Third Semester

Faculty of Science

ST 02 03 05 Statistical Data Management using Python, Advanced R and SPSS

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **three** questions. Each question carries a weight of **ten**)

1. (i) 4 drugs were tested in randomized block design (4 replications). Check whether the 4 drug effects are same or not.

   (ii) Assume that one observation is missing (block 2 and treatment 4). Find the missing value and analyze the data

   (iii) Carry out the analysis when 2 observations are missing (block 2, treatment 4 and block 3, treatment 2)302

		B 1		B 2		B 3		B 4	
T	1	6	6	4	6	6	0	5	0
T	2	5	1	5	9	4	4	4	0
T	3	3	3	4	6	5	9	4	0
T	4	3	4	5	1	2	5	5	2

2. (i) Given under the transition matrix of Markov Chain $\{X_2\}$ describing the generaral transition with states 0,1,2:

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix}$$

   Then obtain

   (a) $P(X_2=1|X_1=0)$

   (b) $P(X_4=0|X_2=2)$

Also obtain the stationary distribution and interpret it.

(ii) Suppose that customers arrived at a facility according to Poisson process having rate $\lambda=2$. Let X(t) be the number of customers that have arrived up to time t. determine the following probabilities and conditional probabilities.

(a) P{X(1)=2}

(b) P{X(1)=2 and X(3)=6}

(c) P{X(1)=2 / X(3)=6}

(d) P{X(3)=6 /X(1)=2}

3. (a) A study was conducted to measure serum alkaline phosphate activity levels in children with seizure disorders who were receiving and convulsant therapy under the care of a physician. 45 subjects were found for the study and categorized into two drug groups

G_1 – Central G_2 –Phenobarbital G_3-corbamazepine G_4- other anticonvulsant

From blood samples collected on each subject the serum alkaline phosphate activity level was determined and recorded in the following table. Test the hypothesis at 0.05 level of significance, that the average serum alkaline phosphate activity level is the same for the four drugs.

G_1– 49, 20, 44, 54, 45, 80, 95, 84, 30, 10, 36, 50,82, 30, 87, 89, 105, 00, 95,22, 97, 50, 105, 00, 58, 09, 86, 60, 58, 60, 58, 39, 72, 80, 116, 70, 45, 15, 70, 35, 77,40

G_2– 97.07, 73, 40, 68, 50, 91, 85, 106, 50, 57, 79, 77, 0.81

G_3 -62.10, 94.95, 142, 50, 53.00, 175.00, 79.50, 29.50, 78.40,127.50

G_4 – 110.6, 57.10, 117.60, 77.71, 150.00, 82.90,111, 20

(b) Perspiration from 20 healthy females were analyzed. Three components X_1: sweat rate, X_2: sodium content and X_3- Potassium content were measured and the results are presented in the following table

Sl No	1	2	3	4	5	6	7	8	9	10
X_1	3.7	5.7	3.8	3.2	3.1	4.6	2.4	7.2	6.7	5.4
X_2	48.5	65.1	47.2	53.2	55.5	36.1	24.8	33.1	47.4	54.1
X_3	9.3	8.0	10.9	12.0	9.7	7.9	14.0	7.6	8.5	11.3
Sl No	11	12	13	14	15	16	17	18	19	20
X_1	3.9	4.5	3.5	4.5	1.5	8.5	4.5	6.5	4.1	5.5
X_2	3.9	58.8	27.8	40.2	13.5	56.4	71.6	42.8	44.1	40.9
X_3	12.7	12.3	9.8	8.4	10.1	7.1	8.2	10.9	11.2	9.4

Assuming that $X^1 = (X_1, X_2, X_3)$ follows N (μ, $\sum$), test the hypothesis

$H_0 : \mu^1 = (\ 4,50,10\ )$ Vs $H_1 : \mu^1 \neq (\ 4,50,10\ )$.

4. In case control study 4 controls are matched to each cases. Calculate OR_{MH} test statistics and 95% CI.

NO	Case	Control
1	YES	1 / 4
2	YES	1 / 4
3	NO	¼
4	YES	0 / 4
5	NO	¼
6	YES	0 / 4
7	YES	1 / 4
8	YES	0 / 4

Model Question Paper

QP Code (to be assigned by Exam Section) **Reg. No. ……**

Name

……

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Fourth Semester

Faculty of Science

ST 02 04 01 SAS PROGRAMMING, BAYESIAN INFERENCE AND MCMC METHODS (2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. What is SAS variable; illustrate with example

2. How will you initialize a Matrix in SAS?

3. Write the syntax to find mean of a set of observations, giving an example

4. How does subjective probability differ from the frequency distribution?

5. Define conjugate prior with example.

6. Define posterior distribution with example

7. Define an exponential family.

8. What is meant by conjugate family ? Give an example.

9. Define Bayes estimate using squared error loss.

10. Explain acceptance-rejection method for generating random numbers.

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. What do you mean by (a) Jeffrey's prior (b) Non-informative prior. Give examples.

12. Suppose a diagnostic test is 95% accurate. If 0.5 % of the population actually has a disease, find the probability that a particular individual has it, given that the test says he has.

13. Explain credibility intervals. How do they differ from confidence intervals?

14. A random variable X follows a binomial distribution with parameters n = 8 and p. Furthermore, we know that p follows a uniform distribution on the interval (0,1). Find the Bayes estimate of p under the squared- error loss function

15 Explain the prior and posterior analysis of Bernoulli process. Find the Baye's estimate of parameter p in a Bernoulli process.

16. Discuss on Bootstrapping and Gibbs sampling.

17. Write a SAS program to simulate 50 observations from N(10,2)

18. Write a short note on control statements in SAS.

(**6 x 2 = 12**)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. In testing a simple H_0 against simple H_1, show that the Bayes factor is in facet a likelihood ratio. Explain Lindley method of testing a composite hypothesis.

20. Explain (i) Monte Carlo and (ii) Makov Chain Monte Carlo simulation techniques with applications

21. (a) Consider a normal population with mean θ and unit variance.Derive posterior distribution of θ assuming a normal priorand obtain its estimate under the 0-1 loss function.

(b) Explain LINEX loss function. Discuss how you will obtain the Bayes estimator ?.

22. Write an essay on various procedures in SAS program. (**2 x 5 = 10)**

Model Question Paper - Format

QP Code (To be assigned by Exam Section) **Reg. No. …….**

Name

……..

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Fourth Semester

Faculty of Biostatistics

ST 02 04 02 SURVIVAL ANALYSIS AND LIFETIME MODELING

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. Explain the concept of censoring and their types.

2. Define hazard rate and hazard function.

3. What are the uses of Leslie matrix?

4. Distinguish between cohort life table and current life table.

5. What is crude birth rate?

6. Write a note on confidence interval for median survival time.

7. What is PPR?

8. Briefly describe the test with non- proportional hazards

9. Define abridged life table.

10. Explain log rank test.

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Derive Kaplan- Meier estimator using suitable example.

12. Obtain the MLE of one parameter (λ) exponential distribution for data with censored observation.

13. Explain a) Mantel- Haenzel test.

    b) Taron- ware test.

14. What is parity- progression ratio?

15. Write a note on Cox- Snell residuals.

16. Discuss on linear failure rate and bath-tub failure rate.

17. What are the methods of identification of prognostic factors related to survival time?

18. How will you compare two survival functions? Explain it with the help of any two tests.

(**6 x 2 = 12**)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. What is accelerated failure time model? Compare accelerated failure time model with proportional hazard model.
20. Describe the different stochastic models for reproduction?
21. Explain the likelihood ratio test for comparing two exponential survival distributions and describe the test procedure
22. Write a note on
    a) Adequacy assessment of the proportional hazard model.
    b) Semi-parametric models of survival distribution.

(**2 x 5 = 10**)

Model Question Paper (Elective BunchA)

QP Code (To be assigned by Exam Section) **Reg. No. ……..**

Name ……..

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Fourth Semester

Faculty of Science

ST 86 04 01 CLINICAL TRIALS AND BIOINFORMATICS

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. What is Clinical Trial?

2. What do you mean by a Protocol?

3. Write a short note on Statistical Analysis Plan (SAP)

4. Briefly explain Bioequivalence trial

5. Define Bioinformatics

6. What is the Role of Bioinformatics in Health Sciences?

7. Write a note on Human Genome Project

8. Explain briefly sequence alignment

9. What is meant by quality of a data?

10. Explain the concept of Blinding

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Explain Drug Development Process

12. What are the benefits and risk of participating in trials?

13. Explain the Randomization Methods

14. Discuss on Binary Data Analysis

15. Explain PDB, SWISSPROT and UNIPROT

16. Write a short note on Computational Biology

17. Distinguish between Local and Global Alignment

18. What is Multiple Sequence Alignment?

(6 x 2 = 12)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. Explain in detail Phase I to IV in clinical trials.

20. What are the techniques for handling of missing data and multiplicity?

21. Explain Nucleotide sequence databases (NCBI- GENBANK, DDBJ and EMBL)

22. Describe Pair wise alignment (BLAST and FASTA]

(2 x 5 = 10)

Model Question Paper

QP Code (To be assigned by Exam Section) **Reg. No. …….**

Name ……..

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Fourth Semester

Faculty of Science

ST 86 04 02 Demography and Vital Statistics

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

11. Explain National Sample Surveys (NSS).
12. What do you mean by data appraisal?
13. Define Lexis diagram with the help of an example.
14. What is fertility? State the factors affecting it.
15. Explain the concept of ageing.
16. What is mortality? Explain mortality trends.
17. Define death statistics. Explain the cause of death statistics.
18. Explain the evolution of family welfare programme in India.
19. Briefly explain the population policies in the context of growth and structure.
20. Distinguish between migration and emigration.

(8 x 1 = 8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Write short notes on:

(a) Importance of population projection.

(b) Method of projection

(c) Limitation of population projection

12. Explain the concept of rates of population growth.

13. Explain the composition of India's population.

14. Write a short note on Davi's and Blake's intermediate variable framework of fertility.

15. Distinguish between mortality and morbidity. Explain the sources of data on mortality and morbidity.

16. Explain the concept of measures of mortality.

17. Explain the types of migrations.

18. Write a short note on Malthus theory on population growth

(**6 x 2 = 12**)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. Write a short note on;

(a) Cohort and cross sectional indicators

(b) Crude rate(c) Prevalence

20 Explain:

(a) The concept and measures of Nuptiality and Fertility.

(b) Bongaart's proximate determinants of fertility.

21. Explain:

(a) The concept of life tables and their uses.

(b) Types and forms of life tables. How are they prepared?

22.Explain National and State population policies in India.

(**2 x 5 = 10)**

Model Question Paper - Format

QP Code (to be assigned by Exam Section) **Reg. No. ……**

Name …….

M Sc Degree (C.S.S) Examination

Fourth Semester

Faculty of Biostatistics

ST 86 04 03 Statistical Computing and Data Analysis using R and SAS

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

(Answer any **three** questions. Each question carries a weight of **ten**)

1. (a) Write a SAS program to simulate 50 observations from a log normal distribution with parameters μ= 30 and σ = 2 and for calculating the mean and standard deviation for the stimulated sample.

   (b) Fifteen items were subject to a life test experiment and following failure time(in hours) were recorded: 12.8, 15.4, 25.8, 30.2, 31.5, 34, 38.1, 40.6, 42.7, 50.8, 55.2, 59.6, 68.3, 72.2, 94.6. Suppose the failure time X follows N(μ, $\sigma = 20$) and the prior knowledge about the product under test suggests that the mean failure time,μ follows N(m = 40, δ=10). Compute Bayes estimate of μ under SELF..
2. (a) Consider the following survival time of 16 patients in weeks 4,20,22,35,38,38,40,44,56,83,89,98,110,138,145 and 27. Assuming that they are following the one parameter exponential distribution with parameter λ obtain

   (i)The MLE of λ (ii) The MLE of mean survival time μ

   (iii) The 95% of confidence intervals for λ and μ

   (b)The following data gives the remission times in week of a group of 21 leukemia patients given placebo. Obtain the Kaplan Meier estimate of survival function. The observations are 1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23.
3. (a) From the following table giving the age specific mortality rates (ASMR), construct a life table

Age	0	1-4	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84
ASMR per thousand	34	1.4	0.7	1.5	1.9	3.8	9.8	23	48.6	100.3

(b) From the above life table find the probability that a couple married at the ages 25 and 20 have a married life of 25 years.

(c) What is the expected life of a person aged 45?

4. (a) With the help of Rasmol generate molecular information from a PDB file. Apply any 10 commands including 4 display commands.

(b) Find the best local alignment between the M. leprae and M. ulcerans chorismate lyase proteins.

Model Question Paper (Elective BunchB)

QP Code (to be assigned by Exam Section) **Reg. No. ……**

Name …….

M Sc (Biostatistics) Degree (C.S.S) Examination, ……

Fourth Semester

Faculty of Science

ST 87 04 01 ADVANCED CLINICAL TRIALS AND OPERATIONS RESEARCH

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

Section- A

(Answer any **eight** questions. Each question carries a weight of 1)

1. Explain the methods used in data collections.
2. How do we recruit study participants in clinical trials?
3. Explain the methods to assess and report adverse effects.
4. What do you mean by monitoring of response variables?
5. Briefly explain the statistical methods used in monitoring.
6. How do you report and interpret the results in clinical trials?
7. What are the reasons for multi-center trials?

8. Explain the term Blindness.
9. What do you mean by Basic solution?
10. Explain the terms slack variables, surplus variables, and artificial variables?

(8 x 1 =8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. How to assess health related quality of life?
12. Explain the uses of HRQL.
13. Write a short note on issues in data analysis of clinical trials
14. Explain the term close out.
15. What do you mean by Multicenter trials?.
16. Briefly explain Baseline Assessment.
17. Explain the graphical method used in;

    (a) LPP with suitable examples.

    (b) Degeneracy and Cycling.
18. Explain the Dual Simplex method.

(6 x 2 = 12)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. Explain the methodological issues facing in clinical trials.
20. (a) Explain Post-study Follow-up.

    (b) How do we store the study materials used in clinical trials?
21. How do we conduct multi-center trials?
22. Explain;

    (a) The Hungarian Method used in Assignment problem.

    (b) MODI method of Transportation problem.

(2 x 5 = 10)

Model Question Paper - Format

QP Code (to be assigned by Exam Section) **Reg. No. ……**

Name …….

M Sc Degree (C.S.S) Examination, ……

Fourth Semester

Faculty of Biostatistics

ST 87 04 02 QUALITY CONTROL AND RESEARCH METHODOLOGY

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

(Answer any **three** questions. Each question carries a weight of **ten**)

1. Define Six-sigma.
2. What is Statistical Process Control?
3. What do you mean by ISO certification?
4. Explain ARL curves of control charts.
5. Explain CUSUM charts and EWMA charts.
6. Briefly explain moving average control charts.
7. Explain the meaning and importance of research.
8. Explain the different types of research.
9. What do you mean by ISBN and ISSN?
10. Briefly explain the structure and components of scientific research.

(8 x 1 =8)

Section B

(Answer any **six** questions. Each question carries a weight of 2)

11. Explain the theory of control charts.
12. Write a short note on sigma chart and R chart
13. What do you mean by process capability analysis?
14. Explain the term process capability.
15. Explain the procedure for selection and formulation of research.
16. Briefly explain the research design.

17. Explain the different steps in the preparation of scientific reports.
18. How do you prepare research papers for journals?

(**6 x 2 = 12**)

Section C

(Answer any **two** questions. Each question carries a weight of 5.)

19. Write a short note on attributable control charts.
20. Explain OC charts and moving control charts.
21. Explain different research methods, and also explain how will you prepare literature review.
22. What are the preparations needed for a project proposal?

(**2 x 5 = 10)**

Model Question Paper - Format

QP Code (to be assigned by Exam Section) **Reg. No. ……**

Name …….

M Sc Degree (C.S.S) Examination, ……

Fourth Semester

Faculty of Biostatistics

ST 87 04 03 Statistical Computing and Data Management using R and SAS

(2019 admissions onwards)

Time: Three hours Max. Weight: 30

(Answer any **three** questions. Each question carries a weight of **ten**)

1. (a) Write a SAS program to simulate 50 observations from a log normal distribution with parameters μ= 30 and σ = 2 and for calculating the mean and standard deviation for the stimulated sample.

   (b) Fifteen items were subject to a life test experiment and following failure time(in hours) were recorded: 12.8, 15.4, 25.8, 30.2, 31.5, 34, 38.1, 40.6, 42.7, 50.8, 55.2, 59.6, 68.3, 72.2, 94.6. Suppose the failure time X follows N($\mu,\ \sigma = 20$) and the prior

knowledge about the product under test suggests that the mean failure time,μ follows N(m = 40, δ=10). Compute Bayes estimate of μ under SELF..

2. (a) Consider the following survival time of 16 patients in weeks 4,20,22,35,38,38,40,44,56,83,89,98,110,138,145 and 27. Assuming that they are following the one parameter exponential distribution with parameter λ obtain:

   (i)The MLE of λ (ii) The MLE of mean survival time μ

   (iii) The 95% of confidence intervals for λ and μ

   (b)The following data gives the remission times in week of a group of 21 leukemia patients given placebo. Obtain the Kaplan Meier estimate of survival function. The observations are 1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23.

3. (a) Use Simplex method to solve the following problem

   Maximize $Z= 3x_1+4x_2+2x_3$

   Subject to $x_1-2x_2+4x_3 \leq 36$

   $2x_1+3x_2-5x_3 \leq 40$

   $3x_1+2x_2+x_3 \leq 28$

   $x_1,x_2,x_3 \geq 0$

   (b) A department head has four subordinates, and four tasks have to be performed. Subordinates differ in efficiency and tasks differ in their intrinsic difficulty. Time each man would take to perform each task is given in the effectiveness matrices.How the tasks should be allocated to each person so as to minimize the total man – hours?

	Subordinates			
	I	II	III	IV
A	8	26	17	11
B	13	28	4	26
C	38	19	18	15
D	19	26	24	10

4. 12 samples of 4 bulbs were selected at regular intervals from a light bulbs manufacturing company. If bulbs have the mean life equal to 1000 hours, it is

considered satisfactory. The standard deviation of life of bulbs is expected to be 522 hours. On testing the samples , the failure times(in hours) were recorded as given below:

Sample number	*Life of bulbs (hours)*			
1	*1081*	*363*	*1092*	*1385*
2	*528*	*330*	*1053*	*945*
3	*984*	*1384*	*1194*	*456*
4	*728*	*972*	*647*	*792*
5	*804*	*845*	*1132*	*1024*
6	*1002*	*804*	*760*	*1035*
7	*994*	*1023*	*1136*	*842*
8	*616*	*832*	*497*	*692*
9	*982*	*1342*	*1132*	*945*
10	*1132*	*998*	*554*	*777*
11	*1134*	*1140*	*756*	*994*
12	*749*	*948*	*1050*	*857*

FORMAT OF AWARDS TO BE ISSUED TO STUDENTS

10.1 GRADE CARDS/ MARK CUM GRADE CARDS FOR EACH SEMESTER

10.2 CONSOLIDATED GRADE CARD

10.3 PROVISIONAL CERTIFICATE

10.4 DEGREE CERTIFICATE