# MASTER OF SCIENCE

# STATISTICS

PROGRAMME STRUCTURE AND SYLLABUS

2019-2020 ADMISSIONS ONWARDS

(UNDER MAHATMA GANDHI UNIVERSITY PGCSS REGULATIONS 2019)

BOARD OF STUDIES IN STATISTICS (PG)

MAHATMA GANDHI UNIVERSITY

2019

**SYLLABII RESTRUCTIONG WORKSHOP ORGANIZED BY THE PG BOARD OF STUDIES (STATISTICS)-A REPORT**

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, the discussions 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 finalize 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, | Dr.Priya P. Menon, |
| Chairperson, | Workshop Coordinator, |
| BoS(PG), Statistics. | 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. Special thanks are to Dr P.G.Sankaran, Pro-Vice Chancellor, Cochin University of Science and Technology, for his patience in checking the draft syllabus and valuable guidance. 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 (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, Kozhikode. |
| 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); | Principal(Retd) and HoD, Department of Statistics, St.Thomas College, Pala. |
| 12. | Prof. Rajeevan Pillai.K.A(Subject Expert) : | Associate Professor(Retd),Maharajas College, Ernakulam. |

**TABLE OF CONTENTS**

## M.Sc (STATISTICS) DEGREE PROGRAMME

### (Mahatma Gandhi University Regulations PGCSS2019 From 2019-20 Academic Year)

**1. Aim of the Program** :

Apart from teaching core Statistics subjects, the students are also taught programming languages and also exposed to various statistical soft wares such as R, SPSS, trained to handle real life problems through the practical classes. As a part of the course the students are taught some MATLAB, SAS etc. The course prepares the students for UPSC Examinations like UGC-CSIR-NET, Indian Statistical Services (ISS), Indian Economic Services (IES) as well as Civil Services. The course is so designed that on successful completion, the students would be able to pursue higher studies in the areas of Statistics, Mathematics, Computer Science ,Economics, Management and allied fields. Moreover, emerging areas like Bayesian Inference, Actuarial science and official Statistics are included in the curriculum. There has been much recent interest in Bayesian Inference in Data Analysis. It is a way to get sharper predictions from the data , particularly when there is not much data available and when one want to juice every last bit of information from it. During the last three decades, Actuarial Science has gone through revolutionary changes due to the implementation of high speed computers and modern theory. It applies mathematical and statistical methods to assess risk in insurance, finance and other industries. Official Statistics make information on economic and social development accessible to the public, allowing the impact of government polices to be assessed and thus improving accountability.

**2. Eligibility for admissions:**

B.Sc. Degree in mathematics or statistics main or B.Sc.(triple main) with Mathematics Statistics and Computer science as main subject with at least 50% marks for the optional subjects taken together .**No private /distant students will be admitted for the programme.**

**3. Examination :** Credit and Semester System (CSS)
Direct Grading system with 7 point scale

**Medium of instruction and assessment** : English

**Duration of the Course : 4 Semesters**

This is a regular course in which no private / distance mode will not be conducted. The duration of PG program shall be 4 semesters. The duration of each semester shall be 90 working days. Odd semesters extend from June to October and even semesters from December to April. A student shall be permitted to register for the program at the time of admission. A student who registered for the course shall complete the course within a period of 4 years from the date of commencement of the program.

The program includes two types of courses namely Program Core courses and Program Elective Courses. In the fourth semester each College can select one bunch of electives from the 3 bunches A,B, and C that will suit the needs of students there, from the electives specified in the syllabus (with course codes ST80 04 01-03, ST 81 04 01-03 and ST 82 04 01-03) and the courses offered in that bunch only. The selection of courses from different groups are not possible.

There shall also be a Program Project with dissertation to be undertaken by all students. Every Program conducted under Credit Semester System shall be monitored by the College Council.

**Credits of the programme**

1. Total credit of the programme is 80.
2. Semester wise total credits vary from19-23.
3. Credit for project/dissertation is 3 and that for viva-voce is 4 ( Subject viva-voce – 2, project viva-voce – 2)

**Evaluation and Grading**

Evaluation scheme for each course shall contain 2 parts (a) Internal evaluation (ISA) and (b) external evaluation (ESA. 25% weightage is given to internal evaluation and 75%to external evaluation to maintain the ratio between internal and external evaluation as 1:3. Both internal and external evaluation is carried out using direct grading system.

**Weightage For Internal & External Evaluation**

The weight for internal evaluation of theory/practical /project/comprehensive viva-voce is 5. The maximum weighted grade point (WGP) is 25. The weight for external evaluation is 15. The maximum weighted grade point (WGP) for external is 75.

**Direct grading system**

Direct grading system based on a 7 point scale is used to evaluate the performance (External and Internal examination of the students). Letter grades and **GPA /SGPA/CGPA** are given in the following scale.

| Range | Grade | Indicator |
|---|---|---|
| 4.50 to 5.00 | A+ | Outstanding |
| 4.00 to 4.49 | A | Excellent |
| 3.50 to 3.99 | B+ | Very good |
| 3.00 to 3.49 | B | Good(Average) |
| 2.50 to 2.99 | C+ | Fair |
| 2.00 to 2.49 | C | Marginal |
| Upto 1.99 | D | Deficit (Fail) |

No separate minimum is required for internal evaluation for a pass , but a minimum **C grade** is required for a pass in an external evaluation. However a minimum **C grade** is required for a pass in a course

**Evaluation first stage**

| Grade | Grade points |
|---|---|
| A+ | 5 |
| A | 4 |
| B | 3 |
| C | 2 |
| D | 1 |
| E | 0 |

**Theory- External**

Maximum weight for external evaluation is **30.** Therefore maximum weighted Grade point (WGP) is **150**

**Theory- Internal**

| | Components | Weightage |
|---|---|---|
| 1. | Assignment | 1 |
| 2. | Seminar | 2 |
| 3. | Best two test papers | 1 each(2) |
| | Total | 5 |

For test papers all questions shall be set in such a way that the answers can be awarded grades A+, A, B,C,D,E

**Assignments:**

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

**Seminar Lectures**

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

**Class Tests**

Every student shall undergo at least two class tests as an internal component for every course. The grade points of the best two test papers (weight 1 each)shall be taken for the internal assessment component.

**Project work**

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

**Project Internal evaluation**

| Components | Weightage |
|---|---|
| Relevance of the topic and analysis | 2 |
| Project content and presentation | 2 |
| Project viva | 1 |
| **Total** | **5** |

**Project External evaluation**

| Components | Weightage |
|---|---|
| Relevance of the topic and analysis | 3 |
| Project content and presentation | 7 |
| Project viva | 5 |
| **Total** | **15** |

**Viva Voce**

Comprehensive Viva-voce shall be conducted at the end of the fourth semester of the program and it shall cover questions from all courses in the programme ., presentation of the project work/ dissertation and viva-voce based on project/ dissertation

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

**Comprehensive Viva-voce Internal**

| **Components** | **Weightage** |
|---|---|
| **Course viva ( all courses from first semester to fourth semester)** | **4** |
| **Project presentation** | **1** |
| **Total** | **5** |

4. **Faculty under which the Degree is awarded** : Science

5. **Specializations offered if any** : List enclosed in Page 17

**Note on compliance with UGC Minimum Standards for the conduct and award of Post Graduate Degrees** :

Present syllabus is in compliance with UGC Minimum Standards to award Post Graduate Degree. The present course is intended to provide a platform for talented students to undergo higher studies in the subject as well as to train them to suit for the needs of the society. It is ideal if one enjoys Mathematics and Statistics and would like to use his skills to model future events and risk. It also allows more flexibility to branch out into other areas of Statistics, Mathematics and Computer Science. The curriculum draws together a variety of subject areas to enable you model real-world effects and their financial implications. One will explore a blend of applied Mathematics and Statistics with appropriate computing support. Equal attention is given to areas in describing, exploring and comparing data.

6. **The Program structure** :

| Semester | Course code | Course title | Teaching hours per week | Credit | Total credit |
|---|---|---|---|---|---|
| I | ST 01 01 01 | Probability and Measure Theory | 5 | 4 | 19 |
| | ST 50 01 01 | Distribution Theory | 5 | 4 | |
| | ST 50 01 02 | Analytical Tools for Statistics | 5 | 4 | |
| | ST 50 01 03 | Sampling Theory | 5 | 4 | |
| | ST 01 01 02 | Statistical Computing I using R | 5 | 3 | |
| II | ST 50 02 01 | Estimation Theory | 5 | 4 | 19 |
| | ST 50 02 02 | Stochastic Processes | 5 | 4 | |
| | ST 50 02 03 | Multivariate Distributions | 5 | 4 | |
| | ST 01 02 01 | Advanced Probability Theory | 5 | 4 | |
| | ST 01 02 02 | Statistical Computing II Using R | 5 | 3 | |
| III | ST 50 03 01 | Testing of Hypotheses | 5 | 4 | 19 |
| | ST 50 03 02 | Design and Analysis of Experiments | 5 | 4 | |
| | ST 50 03 03 | Multivariate Analysis | 5 | 4 | |
| | ST 50 03 04 | Time Series | 5 | 4 | |

| | | | | | |
|---|---|---|---|---|---|
| | | Analysis | | | |
| | ST 01 03 01 | Statistical Computing III using R/ SPSS/MATLAB | 5 | 3 | |
| IV | ST 50 04 01 | Econometrics Methods | 5 | 4 | 23 |
| | ST 80/81/82 04 01 | Elective 1 | 5 | 3 | |
| | ST 80/81/82 04 02 | Elective 2 | 5 | 3 | |
| | ST 80/81/82 04 03 | Elective 3 | 5 | 3 | |
| | ST 01 04 01 | Statistical Computing IV using R/ SAS/MATLAB | 5 | 3 | |
| | ST 01 04 02 | Project/Dissertation | | 3 | |
| | ST 01 04 03 | Viva-voce * | | 4 | |
| | Total | | | | 80 |

Courses with code ST 50 -- -- are common to MSc Statistics and MSc Statistics (Applied).
* The Viva Voce examination in ST 01 04 03 is to be conducted externally with at least one external examiner .The project in ST 01 04 02 shall be evaluated internally and the project evaluation based on a dissertation/project Report of ST 01 04 02 shall be done externally in semester IV .

**<u>Table of Elective Courses</u>:**

**Three bunches A,B and C each having 3 programmes are given.A college can select any one bunch. Selection of programmes between bunches is not allowed.**

| BUNCH | Course code | NAME OF THE COURSE | Credit | Teaching Hours |
|---|---|---|---|---|
| A | ST 80 04 01 | Operations Research | 3 | 5 |
| | ST 80 04 02 | Statistical Quality control | 3 | 5 |
| | ST 80 04 03 | Advanced Bayesian Computing with R | 3 | 5 |
| B | ST 81 04 01 | Survival Analysis | 3 | 5 |
| | ST 81 04 02 | Population Studies | 3 | 5 |
| | ST 81 04 03 | Categorical Data Analysis | 3 | 5 |
| C | ST 82 04 01 | Actuarial Statistics | 3 | 5 |
| | ST 82 04 02 | Applied Regression Analysis | 3 | 5 |
| | ST 82 04 03 | Data Mining | 3 | 5 |

**FIRST SEMESTER COURSES-TOTAL CREDITS 19**

| Course code | NAME OF THE COURSE | Credit | Teaching Hours |
|---|---|---|---|
| ST 01 01 01 | PROBABILITY AND MEASURE THEORY | 4 | 5 |
| ST 50 01 01 | DISTRIBUTION THEORY | 4 | 5 |
| ST 50 01 02 | ANALYTICAL TOOLS FOR STATISTICS | 4 | 5 |
| ST 50 01 03 | SAMPLING THEORY | 4 | 5 |
| ST 01 01 02 | STATISTICAL COMPUTING I-USING R | 3 | 5 |
| | TOTAL CREDITS | 19 | |

**ST01 01 01 PROBABILITY AND MEASURE THEORY**
**Total credits-4 Total hours-25 Weightage-30**

**Objectives of the course: To give basic knowledge in measure theory and probability.**
**UNIT 1**

**1.1** Finite and countable operations on sets. **1.2** Sequences of sets, monotone sequence and limit of a sequence of sets, **1.3** field and sigma field, monotone class, generated sigma field, minimal sigma field, Borel field of $R$ and of $R^n$, **1.4** measurable space, measure, measure space, finite and sigma finite measures, monotone and continuity properties of measures, Counting measure, Lebesgue measure and Probability measure. **1.5** CartheodoryExtention theorem (statement only) .LebesgueStieltjes measures and distribution functions.

**UNIT 2**

**2.1** Measurable functions and their properties, indicator functions, simple functions, measurable function as limit of simple functions. **2.2** Integrals of indicator function, simple function and measurable functions,**2.3** basic integration theorems. Monotone convergence theorem, Fatou's Lemma, Bounded convergence theorem and Lebesgue dominated convergence theorem,

**2.4** Lebesgue and Lebesgue-Stieltjes Integral, comparison of Lebesgue and Riemann Integral.

**UNIT 3**

**3.1** Discrete and Continuous probability spaces and their properties.monotone, continuity and other properties. 3.2 Conditional probability, multiplication theorem, total probability and Bayes theorem. Independence of events, **3.3** Borel 0-1 criterion.Random variable, vector and sequence of random variables, properties of random variables and vectors, distribution of random variables.Distribution function and its properties..**3.4** Jordan decomposition theorem, Correspondence theorem (statement only).**3.5** Independence of random variables.Mathematical expectation, moments and its properties.

**UNIT 4**

**4.1** Basic, Chebychev's, Markov's, Liaponov's, Jensen's, Cr, Cauchy-Swartz's, Holder's, Minkowski'sinequalities.**4.2** The four modes of convergence-convergence almost surely, convergence in probability, convergence in distribution and convergence in $r^{th}$ mean of a sequence of random variables, properties, counter examples and their inter-relationships. **4.3** Weak and complete convergence of distribution functions.Helly-Bray Lemma and Helly-Bray Theorem(statements only).

**Text Books**


1) Ash R.B. and Doléans-Dade C.A. (2000) Probability and measure theory, Academic Press(For Units 1 and 2)
2) Bhat B.R. (2014) Modern Probability theory (An introductory text book), Fourth edition, New Age International.


**Reference Books**


1) Basu A.K. (2012). Measure Theory and Probability, Second Edition, PHI Learning Pvt. Ltd, New Delhi.
2 Billingsley P. (2012) Probability and Measure, Anniversary edition, Wiley Eastern ltd.
3 Laha R.G. and Rohatgi V.K. (1979) Probability theory, John Wiley.

4 Loeve M. (1977) Probability Theory, Fourth edition, Springer-Verlag.
5 Rohatgi V.K. and SalehM. (2015) An introduction to probability and statistics, Third edition, Wiley.
6 Robert G. Bartle (2001), A Modern Theory of Integration, American Mathematical Society(RI),ISBN: 978-0-8218-0845-0

## ST 50 01 01-DISTRIBUTION THEORY

**Total credits-4 Total hours-25 Weightage-30**

**Objectives of the course: To acquaint the students familiar with basic probability distributions and their basic properties.**

**UNIT 1**

**1.1**Probability Generating functions, Moment generating functions and their properties, **1.2**Quick review of Discrete Distributions:-(Degenerate, Bernoulli, Binomial, Uniform, Geometric, Poisson) **1.3** Negative binomial and Hyper geometric , Logarithmic series,. **1.4** Modified Power series and Generalized Power series (Binomial, Poisson, Negative binomial, Logarithmic series etc as special cases)

**UNIT 2**

**2.1**Continuous Distributions:-(Quick review of Rectangular,Triangular, Exponential,Normal, Lognormal) **2.2** Weibull, Beta, Gamma, Pareto, Cauchy, Laplace, Logistic, Inverse Gaussian. **2.3** Pearson family and Exponential family of distributions – Definition and Identification of members.

**UNIT 3**

**3.1** Functions of Random variables and their distributions.**3.2** Probability integral transform, Distributions of sums, products and ratios of independent random variables distributions, **3.3** Compound, Truncated and mixture distributions .

**UNIT 4**

**4.1**Sampling distributions:- Chi-square, *t* and *F* distributions (concept of non-central forms$\chi 2$, t, F (definition only), **4.2**Sampling distributions of mean and variance, independence of sample mean and variance, **4.3**Order statistics and their distributions:- joint , marginalsand conditional distributions; **4.4**Distributions of sample median, range and mid–range (Exponential and Uniform).

**Text Books**

1) Hogg R.V and Craig A.T. (2013) Introduction to Mathematical Statistics, Macmillian publishing company.
2) Johnson N.L, Kotz S. and Kemp A.W. (1992) Univariate discrete distributions, Wiley.
3) Johnson N.L, Kotz S. and Balakrishnan N. (1991) Continuous Univariate distributions I & II, Wiley.

**Reference Books**

1) Arnold B.C, Balakrishnan N. and Nagaraja H.N. (1992) A first Course in Order Statistics.
2) Biswas S. and Srivastava G.L (2008)Mathematical Statistics: A text book, Alpha Science
 International Ltd
3) Gupta S.C. and Kapoor V.K. (2000) Fundamentals of Mathematical Statistics, S. Chand & Co, New Delhi.
4) Rohatgi V.K. and Saleh M. (2015) An introduction to probability and statistics, Third edition, Wiley.
5) Robert G. Bartle (2001), A Modern Theory of Integration, American Mathematical Society(RI),ISBN: 978-0-8218-0845-0

**ST 50 01 02-ANALYTICAL TOOLS FOR STATISTICS**

**Total credits-4 Total hours-25 Weightage-30**

**Objectives:** By the end of this course students are expected to well conversant with basics of linear Algebra and Matrix theory.

**UNIT 1**

**1.1**Basics of linear algebra Definition of a vector space, sub spaces,**1.2** linear dependence and independence, basis and dimensions,**1.3** direct sum and compliment of a subspace, caution spaces, inner product and orthogonality.

**UNIT 2**

**2.1** Algebra of Matrices Linear transformations and matrices, **2.2**Matrices with special structures – triangular matrix, idempotent matrix, Nilpotent matrix, symmetric **2.3** Hermitian and skew Hermitian matrices unitary matrix.**2.4** Row and column space of a matrix, inverse of a matrix. Inverse of a partitioned matrix, **2.5**Elementary operations and reduced forms. linear transformations. Change of basis.

**UNIT 3**

**3.1**Eigen values, spectral representation and singular value decomposition **3.2**Characteristic roots, Cayley-Hamilton theorem, **3.3**minimal polynomial, eigen values and eigen spaces,**3.4** spectral representation of a semi simple matrix, algebraic and geometric multiplicities, Diagonal forms, triangular forms ,Jordan canonical form, **3.5**spectral representation of a real symmetric, singular value decomposition.

**UNIT 4**

**4.1** Linear equations generalized inverses and quadratic forms Homogenous system, general system, Rank Nullity Theorem, **4.2**generalized inverses, properties of g-inverse, Moore-Penrose inverse, properties, computation of g-inverse,**4.3** definition of quadratic forms, classification of quadratic forms, **4.4**rank and signature, positive definite and non negative definite matrices, extreme of quadratic forms

**Text Books :**

1. Gilbert Strang (2014) Linear Algebra and its Applications, 15th Re-Printing edition, Cengage Learning.
2. Hoffman K. and Kunze R. (2014) Linear Algebra, Second edition, Phi Learning.

**Reference Books:**

1) Rao A.R. and Bhimasankaram P. (2000) Linear Algebra, Second edition, Hindustan Book Agency.
2) Rao C.R. (2009) Linear Statistical Inference and its Applications, Second edition, Wiley Eastern.

## ST 50 01 03 - SAMPLING THEORY

**Total credits-4  Total hours-25 Weightage-30**

**Objectives** : By the end of this course students are expected to be able to apply and use the basic concepts related to sampling techniques, to determine sample size so as the estimator will have a desired precision and to use appropriate sampling method and determine optimum sample sizes.

**UNIT1**

**1.1** Official Statistical Systems in India – Role of NSSO and CSO and their activities – For general awareness of students ( 1 or 2 hours) **1.2**Census and Sampling methods, Advantages and disadvantages, Principles of sampling theory, Principal steps in a sample survey, **1.3** probability sampling and non probability sampling, sampling and non sampling errors, bias, variance and MSE , **1.4** simple random sampling with and without replacement - estimation of population mean, total and proportions, estimation of sample size -. **1.5** Properties of the estimators, variance and standard error of the estimators, confidence intervals, determination of the sample size.

**UNIT 2**

**2.1** Stratified random sampling, estimation of the population mean, total and proportion, properties of estimators, various methods of allocation of a sample, comparison of the precisions of estimators under proportional allocation, optimum allocation and SRS. **2.2** Systematic sampling – Linear and Circular, estimation of the mean and its variance, intraclass correlation coefficient , **2.3** comparison of systematic sampling, SRS and stratified random sampling for a population with a linear trend.

**UNIT 3**

**3.1** Ratio method of estimation, estimation of population ratio, mean and total,**3.2** Bias and relative bias of ratio estimator, comparison with SRS estimation. Unbiased ratio type estimators- Hartly- Ross estimator, Regression method of estimation. Comparison of ratio and regression estimators with mean per unitmethod,**3.3** Cluster sampling, single stage cluster sampling with equal and unequal cluster sizes, estimation of the population mean and its standard error. **3.4** Two- stage cluster sampling with equal and unequal cluster sizes, **3.5**Multi stage and Multiphase sampling (Basic Concepts), estimation of the population mean and its standard error.

**UNIT 4**

4.1Varying probability sampling, PPS sampling with and without replacement,**4.2** cumulative total method, Lahiris method, Midzuno-Zen method and its inclusion probabilities ., estimation of the population total and its estimated variance under PPS wr sampling, **4.3** ordered and unordered estimators of the population total under PPS wor, Horwitz – Thomson estimator and its estimated S. E,**4.4** Des-Raj's ordered estimator, Murthy's unordered estimator (properties of these estimators for n=2 only). Inclusion probability proportional to size Sampling Procedures.

**Text Books**


1) **Cochran W.G** (1992): Sampling Techniques, Wiley Eastern, New York.

2)**Singh ,DandChowdhary,F.S.** (1999): Theory and Analysis of Sample Survey Designs, Wiley Eastern (New Age International), New Delhi.


**References**


1) **Parimal Mukhopadhyay (**2009) Theory and Methods of Survey Sampling, Second Edition,PHI Learning (P) Ltd

2) **P.V.Sukhatme**et.al. (1984): Sampling Theory of Surveys with Applications. IOWA State University Press, USA.

3) . **M.N. Murthy (**1977) Sampling Theory and Methods, Statistical Publishing Society,

4) . **Sampath S. C**. (2001) Sampling Theory and Methods, Alpha Science International Ltd., India.

5) Thomas Lumley (1969) Complex Surveys . A guide to analysis using R , Wiley eastern Ltd.

6) Desraj (1967) Sampling theory . Tata McGraw Hill ,NewDelhi

7. MOSPI website.


## ST 01 01 02 STATISTICAL COMPUTING I-USING R

**Total credits-3 Total hours-25 Weightage-30**

**Objectives:** Students are expected to learn the basics in R programming and complete the practicals by the R software.

**UNIT 1**

**1.1**Introduction to statistical software R, Data objects in R, Manipulating vectors, matrices, lists, importing of files, data frame, and computations of descriptive statistics measures.**1.2** R-Graphics- Histogram, Box-plot, Stem and leaf plot, Scatter plot, Plot options;**1.3** Multiple plots in a single graphic window, frequency table,**1.4** Controlling Loops- For, repeat, while, if , if else etc, lm and glm functions. Analysis of variance using lm function.

**UNIT 2**

**Distribution Theory using R(essential theory)-Implementation of numerical problems using R:** ,**2.1** Plotting of probability distributions and sampling distributions, P-P plot, Q-Q Plot, **2.2** Simulation of random numbers, **2.3**Fitting of discrete and continuous distributions. Chi Square goodness of fit. **2.4**Test for correlation and regression coefficients.

**UNIT 3**

**Analytical tools for Statistics using R(essential theory)-Implementation of numerical problems using R: 3.1** Writing user defined functions for statistical methods studied. **3.2**Use of the apply group of functions. **3.3**Vectorized functions. Use of matrix methods in statistical analysis-spectral decomposition,**3.4** inverse, g inverse, Moore Penrose g inverse. **3.5**Use of R in linear algebra and numerical analysis.

**UNIT 4**

**Sampling theory using R(essential theory)-Implementation of numerical problems using R**

**4.1**Use of statistical packages in survey sampling. Computations using the survey package. **4.2**Use of other related packages in sampling theory. **4.3**Writing user defined functions for various computations in sampling theory.

**References.**

1. Introductory Statistics with R by Peter Dalgaard, Springer, 2nd edition, 2008.

2. The R Book by Michael J.Crawley, John Wiley and Sons, Ltd., 2007.

3. An Introduction to R by W. N. Venables, D. M. Smith and the R Core Team

4. The Art of R Programming by Norman Matloff, no starch press, San Francisco

5. Complex Surveys, A Guide to Analysis Using R. Thomas Lumley. John Wiley & Sons.

**SECOND SEMESTER COURSES-TOTAL CREDITS 19**

| Course code | NAME OF THE COURSE | Credit | Teaching Hours |
|---|---|---|---|
| ST 50 02 01 | ESTIMATIONTHEORY | 4 | 5 |
| ST 50 02 02 | STOCHASTIC PROCESSES | 4 | 5 |
| ST 50 02 03 | MULTIVARIATE DISTRIBUTIONS | 4 | 5 |
| ST 01 02 01 | ADVANCED PROBABILITY THEORY | 4 | 5 |
| ST 01 02 02 | STATISTICAL COMPUTING II-USING R | 3 | 5 |
| | TOTAL CREDITS | 19 | |

## ST 50 02 01- ESTIMATION THEORY

**Total credits-4 Total hours-25 Weightage-30**

### UNIT 1

**1.1**Point estimation-properties of estimators – unbiasedness - consistency, sufficient condition for consistency – Sufficiency, minimal sufficiency,**1.2** completeness, bounded completeness, Fisher-Neymann factorization theorem, **1.3** exponential families, UMVUE estimators and their characterization, **1.4** Rao- Black well theorem, Lehmann -Scheffe theorem, **1.5** ancillary statistics, Basu's theorem.

### UNIT 2

**2.1** Fisher information measure and its properties, Fisher information matrix , **2.2** Lower bound to the variance of an unbiased estimator, Cramer -Rao inequality, Bhattacharyya's bounds, **2.3** Efficiency , minimum variance

### UNIT 3

**3.1** Methods of estimation: method of moments, method of maximum likelihood & their properties, Cramer-Huzurbazar theorem, Fisher's scoring method, **3.2** method of minimum chi-square and method of modified minimum chi-square- **3.3** Interval estimation – Pivotal method of construction - shortest confidence intervals and their construction (minimum average width) **-3.4** Construction of shortest confidence intervals in large samples..

### UNIT 4

**4.1** Basic elements of Bayesian inference, Loss function and risk functions, Standard forms of loss functions, **4.2** Prior distribution, Bayes Theorem, Posterior distribution, **4.3** Bayes risk, Bayes principle, Bayes estimators, Minimax estimators.

### Text Books

1) Rohatgi V.K. and Saleh A.K. (2015) An Introduction to Probability Theory and Mathematical Statistics, Wiley.
2) Berger J.O. (1993) Statistical Decision Theory and Bayesian Analysis, Third Edition, Springer.
3) Casella, G and Berger, R.L (2007) Statistical Inference, Second Edition, Cengage Learning.

### Reference Books

1) Hogg R. V. and Craig A. T. (2013) Introduction to Mathematical Statistics, Pearson
2) Kale B. K. (2005) A First Course on Parametric Inference, Alpha Science International.

3) Lehmann E.L. (1983) Theory of point estimation – Wiley, New York.
4) Lindgren B.W (1976) Statistical Decision Theory (3rd Edition), Collier Macmillian, New York.
**5)** Rao C.R (2009) Linear Statistical Inference and its Applications, John Wiley, New York.

**ST 50 02 02 - STOCHASTIC PROCESSES**

**Total credits-4 Total hours-25 Weightage-30**

**Objectives:** To impart basic knowledge & skills in Stochastic Models and their applications in Statistics.

**UNIT 1**

**1.1** Introduction to stochastic processes:- classification of stochastic processes according to state space and time space, wide sense and strict sense stationary processes, processes with stationary independent increments, **1.2** Markov process, Markov chains-transition probability matrices, Chapman-Kolmogorov equation, **1.3** first passage probabilities, generating functions, classification of states, criteria for recurrent and transient states, **1.4** mean recurrence time, mean ergodic theorem, the basic limit theorem of Markov chains (statement only),**1.5** reducible and irreducible Markov chains, stationary distributions, limiting probabilities and absorption probabilities.

**UNIT 2**

**2.1** Random walk, gambler's ruin problem; **2.2** Galton-Watson branching process, generating function relations, **2.3** mean and variance functions, extinction probabilities, criteria for extinction.

**UNIT 3**

**3.1** Continuous time Markov chains, Poisson processes,**3.2** pure birth processes and the Yule processes, birth and death processes, **3.3** Kolmogorov forward and backward differential equations, linear growth process with immigration, **3.4** steady-state solutions of Markovian queuing models--M/M/1, M/M/1 with limited waiting space,**3.5** M/M/s, M/M/s with limited waiting space.

**UNIT 4**

**4.1** Renewal processes– concepts, examples, **4.2** Poisson process viewed as a renewal process, renewal equation, elementary renewal theorem,**4.3** asymptotic expansion of renewal function, central limit theorem for renewals,**4.4** key renewal theorem (statement only), delayed renewal processes.

**Text Books**

1. Medhi J. (2017) Stochastic Processes, Second Edition , Wiley Eastern, New Delhi

2. Ross S.M. (2007) Stochastic Processes. Second Edition, Wiley Eastern, New Delhi

**Reference Books**

1. Feller W. (1968) Introduction to Probability Theory and its Applications, Vols. I & II, John Wiley, New York.

2. Karlin S. and Taylor H.M. (1975) A First Course in Stochastic Processes, Second edition, Academic Press, New-York.
3. Cinlar E. (1975) Introduction to Stochastic Processes, Prentice Hall, New Jersey.
4. Basu A.K. (2003) Introduction to Stochastic Processes, Narosa, New-Delhi.
5. Bhat U.N. and Miller G. (2003) Elements of Applied Stochastic Processes. (Third edition), John Wiley, New York.

## ST 50 02 03 -MULTIVARIATE DISTRIBUTIONS

**Total credits-4 Total hours-25 Weightage-30**

**Objectives:** To impart the general knowledge of bivariate and multivariate distributions in Statistics and their applications.

**UNIT 1**

**1.1** Notions of bivariate distributions, Gumbel's bivariate exponentials and basic properties., **1.2** Bivariate normaldistribution- marginals and conditionals, independence of random vectors, **1.3** multinomial distribution and its basic properties.

**UNIT 2**

**2.1** Multivariate normal (singular and non-singular), characteristic function, marginals and conditionals–**2.2** properties, characterizations, **2.3** estimation of mean vector and dispersion matrix, independence of sample mean vector and sample dispersion matrix.

**UNIT 3**

**3.1** Jacobian of matrix transformations of Y= AXB; Y= AXA`; X=TT`,**3.2** matrix variate gamma and beta distributions. **3.3** Wishart distribution and its basic properties, characteristic function,**3.4** generalized variance and its distribution.

**UNIT 4**

**4.1** Quadratic forms and their distributions (both scalar and vector forms), **4.2** Independence of quadratic forms, Cochran's theorem. **4.3** Simple, partial and multiple correlation-distributions ,properties.and their inter-relationships, tests**.4.4** Null and non-null distribution of simple and partial correlations, null distribution of multiple correlation.

**Text Books:**


1) Anderson T.W.(1984): An introduction to multivariate statistical analysis, Second edition, John Wiley.
2) Seber G.A.F. (1983): Multivariate Observations, John Wiley.


**Reference Books :**


3) GiriN.(1984) : Multivariate Statistical Inference, Academic publishers.
4) Kollo T and Rosen D.V. (2005): Advanced Multivariate Statistics with Matrices, Springer.

5) Kotz S, Balakrishnan N , and Johnson N.L.(2000): Continuous Multivariate Distributions, Models and Applications,Volume 1, Second Edition, John Wiley.

6) Mathai A.M. (1996): Jacobins of Matrix Transformations and functions of Matrix Argument, World Scientific Pub CoPvt.Ltd

7) Rao.C.R(2009): Linear statistical inference and its applications, Second Edition, Wiley Eastern.

8) Laha R C and Rohatgi VK (1979): Probability theory, John Wiley.

## ST 01 02 01 - ADVANCED PROBABILITY THEORY

**Total credits-4 Total hours-25 Weightage-30**

**Objectives:** To ensure that the students are familiar with modern probability theory and related applications.

### UNIT 1

**1.1** Characteristic function of a random variable, properties, continuity and inversion theorems of characteristic functions, **1.2** convex combinations of characteristic functions and distribution functions, characteristic function of a vector random variable. **1.3** uniform continuity and non-negative definiteness, statement of Bochner's Theorem.

### UNIT 2

**2.1** Law of large numbers: Weak law of large numbers - Bernoulli, Chebychev, Poisson and Khinchine WLLN, Necessary and sufficient condition for weak law of large numbers. **2.2** Strong law of large numbers, Kolmogrov strong law of large numbers for iidrandomvariables .

### UNIT 3

**3.1** Central limit theorem, Demoivre-Laplace central limit theorem, Lindberg-Levy central limit theorem, Liaponov's central limit theorem, **3.2** Lindberg-Feller central limit theorem (Without proof), Statement of Multivariate central limit theorem, **3.3** Domain of attraction, Infinitely divisible distributions and Stable distributions-definition and elementary properties.

**UNIT 4**

**4.1** Signed measure, Hahn decomposition theorem, Jordan decomposition theorem, **4.2**Statement and applications of Radon-Nikodym Theorem (without proof), Lebesgue decomposition theorem, **4.3** Product space, Fubini's theorem (Statement only), Conditional expectation and its properties, Martingales and its simple properties.

**Text Books**

1) Bhat B.R (1999) Modern Probability theory, Third Edition, Wiley Eastern Ltd, New Delhi.
2) Laha R.G and Rohatgi V.K (1979) Probability theory, Wiley.

**Reference Books**

1) Ash R.B (1972) Real Analysis and Probability, Academic press.
2) Billingsley P (2012) Probability and Measure, Third Edition, Wiley Eastern Ltd.
3) Luckas E (1970) Characteristic functions, Second Edition, Hafner Publishing Company, NewYork.
4) Parthasarathy K.R (2005) Introduction to Probability and Measure, Hindustan Book Agency.

## ST 01 02 02 - STATISTICAL COMPUTING II USING R

**Total credits-3 Total hours-25 Weightage-30**

**Objectives:** To make the student capable to do practical problems in more advanced area of Statistics using R software.

Applications of topics covered in

1. ST 50 02 01: ESTIMATION THEORY
2. ST 50 02 02: STOCHASTIC PROCESSES
3. ST 50 02 03 : MULTIVARIATE DISTRIBUTIONS

**Evaluation**: 6 numerical questions each with weightage 10 ( marks per question is 50) are to be asked. The student is expected to answer 3 full questions. At least one question from each of the section must be answered. Use of the package R is only allowed for answering the questions in this paper. Examination of 3 hour duration must be conducted in the computer lab under the supervision of an external examiner appointed by the Controller of Examinations.

## THIRD SEMESTER COURSES-TOTAL CREDITS 19

| Course code | NAME OF THE COURSE | Credit | Teaching Hours |
|---|---|---|---|
| ST 50 03 01 | TESTING OF HYPOTHESES | 4 | 5 |
| ST 50 03 02 | DESIGN AND ANALYSIS OF EXPERIMENTS | 4 | 5 |
| ST 50 03 03 | MULTIVARIATE ANALYSIS | 4 | 5 |
| ST 50 03 04 | TIME SERIES ANALYSIS | 4 | 5 |
| ST 01 03 01 | STATISTICAL COMPUTING III – USING R/ SPSS/MATLAB | 3 | 5 |
| | TOTAL CREDITS | 19 | |

## ST 50 03 01 - TESTING OF HYPOTHESES

**Total credits-4 Total hours-25 Weightage-30**

**Objectives:** To make the student understand the concepts of testing of hypothesis and to develop appropriate tests for testing certain Statistical hypotheses.

**UNIT 1**

**1.1** Basic concepts in statistical hypotheses testing-simple and composite hypothesis, critical regions, Type-I and Type-II errors, significance level, p-value and power of a test; **1.2** Neyman-Pearson lemma and its applications; **1.3** Construction of tests using NP lemma-Most powerful test, uniformly most powerful test; **1.4** Monotone Likelihood ratio and testing with MLR property; Testing in one-parameter exponential families-one sided hypothesis,**1.5** Unbiased and Uniformly Most Powerful Unbiased tests for different two-sided hypothesis; Extension of these results to Pitman family when only upper or lower end depends on the parameters.

**UNIT 2**

**2.1** Similar regions tests, Neymann structure tests, Likelihood ratio (LR) criterion and its properties, **2.2** LR tests for testing equality of means and variances of several normal populations. Testing in multi-parameter exponential families-tests with Neyman structure,**2.3** UMP and UMPU similar size-tests; **2.4** Confidence sets, UMA and UMAU confidence sets, Construction of UMA and UMAU confidence sets using UMP and UMPU tests respectively.

**UNIT 3**

**3.1** Sequential probability ratio tests (SPRT), Properties of SPRT,Determination of the boundary constants **3.2** Construction of sequential probability ratio tests, Wald's fundamental identity, **3.3** Operating characteristic (OC) function and Average sample number (ASN) functions for Normal Binomial, Bernoulli's, Poisson and exponential distribution.

**UNIT 4**

**4.1** Non-parametric tests-- Sign test, Chi-square tests, Kolmogorov-Smirnov one sample and two samples tests, Median test, WilcoxonSignedRanktest,Mann- Whitney U-test,**4.2** Test for Randomness,Runs up and runs down test, Wald–Wolfowitz run test for equality of distributions, **4.3** Kruskall–Wallis one-way analysis of variance, Friedman's two-way analysis of variance,Power and asymptotic relative efficiency.

**Textbooks**

1) RohatgiV.K.(1976)AnIntroductiontoProbabilityTheoryandMathematicalStatistics,JohnWiley&Sons,NewYork.
2) Gibbons J.K. (1971) Non-Parametric Statistical Inference, McGraw Hill.

**References Books**

1. Casella G. and Berger R.L. (2002) Statistical Inference, Second Edition Duxbury, Australia.
2. Lehman E.L. (1998) Testing of Statistical Hypothesis. John Wiley, New York.
3. Wald A. (1947) Sequential Analysis, Wiley, Doves, New York.
4. ParimalMukhopadhyay(2006):Mathematical Statistics, 3/e, Books and Allied (P) Ltd, Kolkata.
5. Siegel S. and Castellan Jr. N. J. (1988) Non-parametric Statistics for the Behavioral Sciences, McGraw Hill, New York.
6. Rao C.R. (1973) Linear Statistical Inference and its Applications, Wiley.

**ST 50 03 02 -DESIGN AND ANALYSIS OF EXPERIMENTS**

**Total credits-4 Total hours-25 Weightage-30**

**Objectives**: By the end of the course the students will be able to conduct experiment by using appropriate design, to test related hypotheses and estimate the parameters.and to compare different designs and will be capable to use the Analysis Covariance technique for data analysis

**UNIT 1**

**1.1** Linear estimation: Gauss Markovset up, Estimability of parameters, **1.2** Method of least squares, best linear unbiased Estimators, Gauss-MarkovTheorem, Tests of linear hypotheses , **1.3** Analysis of variance- one-way, two-way and three-way classification models.

**UNIT 2**

**2.1** Planning of experiments: Basic principles of experimental design, Uniformity trails, **2.2** Completely randomized design (CRD), Randomized block design (RBD), **2.3** Latin square design (LSD) and Graeco-latin square designs, **2.4** Analysis of covariance (ANACOVA), ANACOVA with one concomitant variable in CRD and RBD

**UNIT 3**

**3.1** Incomplete block design: Balanced incomplete block design (BIBD); Incidence Matrix, C- Matrix, Parametric relations;**3.2** Intra-block analysis of BIBD, Connectedness, Construction of BIBD by developing initial blocks,**3.3** Basic ideas of partially balanced incomplete block design (PBIBD).

**UNIT 4**

**4.1** Factorial experiments, $2^n$ and $3^n$ factorial experiments, Analysis of $2^2$, $2^3$ and $3^2$ factorial experiments,**4.2** Confounding in $2^n$ and $3^n$ factorial experiments, Construction of confounded scheme in $2^n$ factorial experiments,**4.3** Split plot experiments (RBD).

**Text Books**

1) Das M.N. and Giri N.C. (1994) Design and analysis of experiments, Wiley Eastern Ltd
2) Joshi D.D. (1987) Linear estimation and Design of Experiments, Wiley Eastern.

**Reference Books**

1) Agarwal B.L (2010) Theory and Analysis of Experimental Designs, CBS Publishers & Distributers
2) Dean A. and Voss D. (1999) Design and Analysis of Experiments, Springer Texts in Statistics
3) Dey A. (1986) Theory of Block Designs, Wiley Eastern, New Delhi.
4) Gomez K.A. and Gomez A.A. (1984) Statistical Procedures for Agricultural Research,

Wiley Eastern Ltd

5) Kempthrone,O. (1952) Design and Analysis of Experiments,Wiley Eastern, New York
6) Montgomery ,C.D. (2012) Design and Analysis of Experiments, John Wiley, New York.
7) Rangaswamy, R (2010) A textbook on Agricultural Statistics , New Age International publishers

## ST 50 03 03 - MULTIVARIATE ANALYSIS

**Total credits-4 Total hours-25 Weightage-30**

**Objectives:** To impart basic knowledge and skills to the students in applied Multivariate Analysis and their applications in Statistics and also bring the confidence to handle real problems on the spot.

**UNIT** 1

**1.1** Notion of likelihood ratio tests, Hotellings-$T^2$ and Mahalnobis-$D^2$ statistics-Their properties, interrelationships and uses, **1.2** Null distributions (one sample and two sample cases), Testing equality of mean vectors of two independent multivariate normal populations with same dispersion matrix,**1.3** Problem of symmetry ,Multivariate Fisher-Behren problem.

**UNIT 2**

**2.1**Dimension Reduction methods: Profile Analysis and the associated tests,**2.2** Principal component Analysis-Method of extraction-properties, the associated tests, **2.3** Factor Analysis-Orthogonal Model-Estimation of factor loadings, **2.4** Canonical variates and canonical correlation, use, estimation and computation.**2.5** Structural equation models. Hoteling's iterative procedure.

**UNIT 3**

**3.1** Classification problems: Discriminant Analysis-Bayes' procedure, Classification into one of the two populations (Normal distribution only), Classification into several populations (Normal distribution only), **3.2** Fishers linear discriminant function and its associated tests,**3.3** Cluster Analysis: proximity measures, Hierarchical and non-hierarchical methods.

**UNIT 4**

**4.1** Multivariate General linear models-MANOVA (one way and two way), **4.2** Wilk's λ, Rau's U, Pillai's trace,Hotelling-Lawley trace, Roy's Maximum Root Statistics (Concepts only),**4.3** Tests-Independence of sets of variables, Equality of dispersion matrices and Sphericity test.

**Text Books**

1) Anderson T. W. (2010) An Introduction to Multivariate Statistical Analysis (3rd ed.) John Wiley.
2) Seber G. F. (2004) Multivariate Observations, John Wiley.
3) Rencher,A. C. (2012) Methods of Multivariate Analysis.(3rd ed.) John Wiley.

**Reference Books:**

1. Johnson R.A. and Wichern D.W. (2008) Applied Multivariate Statistical Analysis. (6thed.) Pearson education.
2. Rao C. R. (2009) Linear Statistical Inference and Its Applications (2nd Ed.), Wiley
3. Johnson, D. E. (1998) : Applied Multivariate methods for Data Analysts, Duxbury Press,

USA-An International Thomson Publishing Company.

4. Morrison, F (2003): Multivariate Statistical Methods, Brooks/Cole, 4th Revisededn. McGraw Hill Book Company.
5. Kshirsagar A.M. (1972): Multivariate Analysis, M.Dekker.
6. Srivastava M.S.and Khatri C.G.(2002):Methods of Multivariate Statistics,John Wiley & Sons, N.Y.

## ST 50 03 04 TIME SERIES ANALYSIS

**Total credits-4 Total hours-25 Weightage-30**

**Objectives:** By the end of this course the student will be able to analyse time series data and identify and interpret various types of behaviour of the time series .

**UNIT 1**

**1.1** Time series, Components of time series, Additive and multiplicative models, **1.2** Estimation and elimination of trend and seasonality, Moving average, **1.3** Simple Exponential Smoothing, Holt's exponential smoothing, Holt-Winter's exponential smoothing, **1.4** Forecasting based on smoothing.

**UNIT 2**

**2.1** Time series as a discrete parameter stochastic process, Auto-covariance and auto-correlation functions, Partial Auto-correlation function and their properties, **2.2** Stationary processes, Wold representation of linear stationary processes, **2.3** Detailed study of the Box - Jenkins linear time series models: Autoregressive, Moving Average, Autoregressive Moving Average and Autoregressive Integrated Moving Average models.

**UNIT 3**

**3.1** Estimation of ARMA models: Yule-Walker estimation for AR Processes, **3.2**Maximum likelihood and least squares estimation for ARMA Processes. Choice of AR and MA periods, Forecasting using ARIMA models,**3.3** Residual analysis and diagnostic checking.

**UNIT 4**

**4.1** Spectral density of a stationary time series and its elementary properties, Periodogram, **4.2** Spectral density of an ARMA process. Seasonal ARIMA models(Basic concepts only),**4.3** ARCH and GARCH models (Basic concepts only).

**Text Books:**


1. Abraham B. and Ledolter J.C. (2005) Statistical Methods for Forecasting, Second edition Wiley.
2. Box G.E.P, Jenkins G.M. and Reinsel G.C. (2008) Time Series Analysis: Forecasting and Control,Fourth Edition, Wiley.
3. Brockwell P.J and Davis R.A. (2002) Introduction to Time Series and Forecasting Second edition, Springer-Verlag.


**Reference Books**


1) Cryer, J. D. and Chan, K. (2008). Time Series Analysis with Applications in R, Second Edition, Springer-Verlag.
2) Shumway, R. H. and Stoffer, D. S. (2011) Time Series Analysis and Its Applications with R Examples, Third Edition, Springer-Verlag.

**ST 01 03 01- STATISTICAL COMPUTING III –USING R/SPSS/MATLAB**

**Total credits-3 Total hours-25 Weightage-30**

**Objectives:** To make the students able to handle practical problems in testing of hypotheses, design and analysis of experiments and also the multivariate techniques. They can use softwares such as R, SPSS or MATLAB. It will enable them to handle practical situations.

Applications of topics covered in

1. ST 50 03 01: TESTING OF HYPOTHESES
2. ST 50 03 02: DESIGN AND ANALYSIS OF EXPERIMENTS
3. ST 50 03 03 : MULTIVARIATE ANALYSIS

**Evaluation**: 6 numerical questions each with weightage 10 ( marks per question is 50) are to be asked. The student is expected to answer 3 full questions. At least one question from each of the section must be answered. Use of the packages R / SPSS / MATLABare allowed for answering the questions in this paper. Examination of 3 hours duration must be conducted in the computer lab under the supervision of an external examiner appointed by the Controller of Examinations.

**FOURTH SEMESTER COURSES-TOTAL CREDITS 23**

| Course code | NAME OF THE COURSE | Credit | Teaching Hours |
|---|---|---|---|
| ST 50 04 01 | ECONOMETRIC METHODS | 4 | 5 |
| | ELECTIVE I* | 3 | 5 |
| | ELECTIVE II* | 3 | 5 |
| | ELECTIVE III* | 3 | 5 |
| ST 01 04 01 | STATISTICAL COMPUTING IV – USING R/ SAS/MATLAB | 3 | 5 |
| ST 01 04 02 | PROJECT/ DISSERTATION | 3 | |
| ST 01 04 03 | VIVA-VOCE | 4 | |
| | TOTAL CREDITS | 23 | |

* Each College can select one bunch of electives from the bunches A,B, and C(with course codes ST80 04 01-03, ST 81 04 01-03 and ST 82 04 01-03) and the courses offered in that bunch only.The selection of courses from different groups are not possible.

## ST 50 04 01 -ECONOMETRIC METHODS

**Total credits-4 Total hours-25 Weightage-30**

**Objectives:** To enable the students to handle models of econometrics and Mathematical Economics. To apply and use the basic concepts related to the economy of a nation and to interpret various parameters used to measure economic status of a nation.

### UNIT 1

**1.1** Simple linear regression models, Multiple linear regression models, estimation of the model parameters, **1.2** tests concerning the parameters, confidence intervals, prediction, use of Dummy variables in regression, **1.3** polynomial regression models, step-wise regression.

### UNIT 2

**2.1** Multicollinearity- consequences, Detection, Farrar-Glauber test, remedial measures. **2.2** Heteroscedasticity- consequences, Detection, tests, remedial measures Aitken's generalized least square method. **2.3** Auto-correlation-tests for auto correlation, consequences, and estimation procedures, **2.4** Errors in variables-consequences, detection, remedial measures, Stochastic regressors. **2.5** Diagnostics, outlier, Influential observations, Leverage, Non parametric regression basics.

### UNIT 3

**3.1** Demand and supply functions, Cobweb model, elasticity of demand, equilibrium of market, **3.2** indifference curves, Cost Function, Utility, Firms, Marginal analysis of firms,**3.3** production functions- elasticity of production, homogeneous functions, Cobb-Douglas Production function, **3.4** constraint maximization of Profit, Revenue, output, **3.5** input- output analysis-Open and closed system.

### UNIT 4

**4.1** Simultaneous equation models, instrumental variables, recursive models, **4.2** distributed-lag models identification problems, rank and order condition, **4.3** methods of estimation-indirect least squares, least variance ratio and two-stage least squares, FIML- methods.

### Text Books


1) Damodar N Gujrati, Sangeeth (2007) Basic Econometrics $5^{th}$ Ed., McGraw Hill Education Private Ltd.
2) Montgomery D.C., Peck E.A. and Vining G.G. (2007) Introduction to Linear Regression Analysis, John Wiley, India.
3) Johnston J. (1984) Econometric Methods (Third edition), McGraw Hill, New York.

**References Books**


1) AllenR.G.D. ( 2008) Mathematical Analysis For Economists, Aldine Transaction
2) Apte P.G. (1990) Text book of Econometrics, Tata Me Graw Hill.
3) *Jeffrey M. Wooldridge (2012)* Introductory Econometrics: A Modern Approach 5th Edition, South-Western College Pub.
4) Koutsoyiannis A. (2008) Modern Microeconomics, Second Edition, Macmillan Press Ltd
5) Kutner M. H, Nachtsheim C.J, Neter J and Li W. (2005), Applied Linear Statistical Model, Fifth edition. McGraw Hill
6) Theil H. (1982) Introduction to the Theory and Practice of Econometrics, John Wiley.


**ST 01 04 01 STATISTICAL COMPUTING IV- USING R/SAS/MATLAB**

**Total credits-3 Total hours-25 Weightage-30**

**Objectives:** To impart the practical skills in the students in the theories of Econometrics and other elective papers. To make them familiar with the software packages.

Applications of topics covered in

1. ST 01 04 01: ECONOMETRIC METHODS
2. ST 8- 04 01: Elective I
3. ST 8- 04 02 : Elective II

**Evaluation**: 6 numerical questions each with weightage 10 ( marks per question is 50) are to be asked. The student is expected to answer 3 full questions. At least one question from each of the section must be answered. Use of the packages R / SAS / MATLABare allowed for answering the questions in this paper. Examination of 3 hours duration must be conducted in the computer lab under the supervision of an external examiner appointed by the Controller of Examinations.

# ELECTIVES

## Bunch A

### ST 80 04 01 OPERATIONS RESEARCH

**Total credits3 Total hours-25 Weightage-30**

**Objectives:** To make the students able to deal with optimization problems and the mathematical theorey involved in them.

**UNIT 1**

**1.1** Linear programming: convex sets and associated theorems, **1.2** Simplex method, Artificial variables technique-Big M method,**1.3** Two phase method; Dual simplex method. Concept and theorems of duality, **1.4** Transportation problems, Assignment problems, **1.5** Sequencing, Traveling sales man problems.

**UNIT 2**

**2.1** Dynamic and Quadratic programming: Bellman's principle of optimality, single additive constraint- additively separable return,**2.2** single multiplicative constraint- additively separable return, single additive constraint-multiplicatively separable return, **2.3** General non-linear programming problem, Constrained optimization with equality constraints -necessary conditions for a general NLPP, sufficient conditions for a general NLPP with one constraint, sufficient conditions for a general problem with $m(<n)$constraints,**2.4** Constrained optimization with inequality constraints, Kuhn-Tucker conditions for general NLPP with $m(<n)$ constraints, **2.5** Wolfe's modified simplex method and Beale's method.

**UNIT 3**

**3.1** Inventory models:-Deterministic inventory models -general inventory model,**3.2** Economic-order quantity (EOQ) models -classic EOQ model, EOQ with price breaks, multi-item EOQ with storage limitation,**3.3** Probabilistic inventory models:- Single period stochastic models without setup cost, General single period models,

**UNIT 4**

**4.1**Theory of Games, Two person zero sum games, fundamental theorem of matrix games, **4.2** Rectangular games as a Linear programming problem, Dominance property,**4.3** Graphical Method of solution 2 x n and m x2 games.

**Text Books:**


1. KantiSwarup, Gupta, P.K. and Man Mohan (2001) Operations Research, Ninth edition, Sultan Chand & Sons

2. Sharma J.K. (2013) Operations Research: Theory and Applications, Fifth edition,Laxmi Publications-New Delhi.
3. K. V. Mital (2016) Optimization methods in Operations research and systems analysis.

**Reference Books**

1) Taha H.A. (2007) Operations Research -An introduction, Eighth edition, Prentice-Hall of India Ltd.
2) Gass S.I. (1985) Linear Programming -methods and applications, Fifth edition, McGraw Hill, USA,
3) Ravindran A, Philips D.T and Soleberg J.J. (1997) Operation Research-Principles and Practice, John Wiley & Sons.
4) Sinha, S.M. (2006) Mathematical programming theory and methods, Elsevier, a division of Reed Elsevier India Pvt. Ltd., New Delhi.
5) Paneerselvam, R. (2008) Operations Research, Second edition, Prentice Hall of India Pvt. Ltd., New Delhi.

## ST 80 04 02 - STATISTICAL QUALITY CONTROL

**Total credits-3 Total hours-25 Weightage-30**

**Objectives:** To make the students aware of the modern quality assurance techniques and methods.

**UNIT 1**

**1.1** Meaning of quality, and need for quality control. Meaning and scope of statistical process control.**1.2** General theory of control charts, Shewhart control charts for variables- mean charts, R-charts, and S-charts,Moving-average control charts. **1.3** Attribute control charts - p, np, c, u charts.**1.4** OC and ARL curves of control charts.

**UNIT 2**

**2.1** Modified control charts. Control charts with memory - EWMA charts, **2.2** CUSUM charts. Process capability analysis, process capability indices – Cp and $C_{pk}$. **2.3** Economic design of mean charts.

**UNIT 3**

**3.1** Statistical product control- basic ideas. Acceptance sampling for attributes - single sampling, double sampling, multiple sampling and sequential sampling plans.**3.2** ASN curves. Measuring performance of sampling plans through OC curves.**3.3** Rectifying inspection plans.AOQ and ATI curves,

**UNIT 4**

**4.1** Acceptance sampling by variables. Sampling plan for a single specification limit with known and unknown variance. **4.2** Performance evaluation through OC curves. **4.3** Designing a variable sampling plan with a specified OC curve.

**Text Books**

1) Montgomery, D.C. (2012). Introduction to Statistical Quality Control, Seventh edition, Wiley.
2) Duncan, A.J. (1986) Quality control and Industrial Statistics, Irwin, Homewood
3) Grant E.L. and Leaven Worth, R.S. (1980) Statistical Quality Control, McGraw Hill.

**Reference Books**

1) Mittag, H.J. and Rinne, H. (1993) Statistical Methods for Quality Assurance, Chapman & Hall, Chapters 1, 3 and 4.
2) Rabbit, J T and Bergle, P.A. The ISO 9000 book, Second Edition, Quality resources, Chapter-I
3) Schilling, E.G. (1982) Acceptance Sampling in Quality Control, Marcel Dekker.

## ST 80 04 03 -Advanced Bayesian Computing with R

**Total credits-3 Total hours-25 Weightage-30**

**Objectives :** To introduce the basic Bayesian Inference methods and computations using computer packages.

**UNIT 1**

**1.1Bayesian Inference:** Parametric family and likelihood, exponential family, Bayes' theorem for inference, prior and posterior densities, **1.2** conjugate priors, non-informative prior, beta prior for binomial proportion, histogram prior, **1.3** discrete prior, single parameter models, normal distribution with known variance and unknown mean, normal with known mean and unknown variance, **1.4** Poisson model, introduction to LearnBayes package, Examples using LearnBayes package.

**UNIT 2**

**2.1 Multi-parameter Models:** Various methods for prior selection, normal distribution with both parameters unknown, multinomial model, Dirichlet prior, **2.2** Bioassay experiment, comparing two proportions, **2.3** predictive distribution, beta-binomial distribution, **2.4** multivariate normal distribution, examples using LearnBayes package.

**UNIT 3**

**3.1 Bayesian Computation:** Computing integrals using Monte-Carlo simulation, approximation based on posterior mode, importance sampling, multivariate- t distribution, **3.2** Markov Chain Monte Carlo methods, Metropolis-Hastings algorithm, random walk,Gibbs sampling,**3.3** MCMC Output Analysis.

**UNIT 4**

**4.1 Model Comparison and Regression models:** Hierarchical models, shrinkage estimators, posterior predictive model checking, **4.2** comparison of hypotheses, Bayes factor, one sided test for normal mean, two sided test for normal mean, normal linear regression model, **4.3** prediction of future observations, examples and R codes introduction to WinBUGS package.

**Text Books.**


1. **Albert, J. (2007). *Bayesian Computation with R,* New York: Springer-Verlag**
2. Berger, J. (2000).*Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag


**Reference Books.**


1. Bolstad, W. (2004). *Introduction to Bayesian Statistics*, Hoboken, NJ: John Wiley
2. Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2003). *Bayesian Data Analysis*, New York: Chapman and Hall
3. Gilks, W.R., Richardson, S and Spiegelhalter, D.J. (1996). Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC, New York
4. Robert, C. and Casella, G. (2004).*Monte Carlo Statistical Methods*, New York: Springer
5. Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003), WinBUGS 1.4 Manual.

## Bunch B

### ST 81 04 01-SURVIVAL ANALYSIS
### Total credits-3 Total hours-25 Weightage-30

**Objectives:** Survival Analysis is highly applied in clinical data. This course will help them in handling clinical data and related analysis.

**UNIT 1**

**1.1** Basic Quantities and Models - Survival function, Hazard function, Mean residual life function and Median life, **1.2** Common Parametric Models for Survival Data; **1.3** Censoring and Truncation - Right Censoring, Left or Interval Censoring, Truncation,**1.4**Likelihood Construction for Censored and Truncated Data

**UNIT 2**

**2.1** Nonparametric Estimation of a Survivor Function and Quantiles**,** **2.2** The Product-Limit Estimator, Nelson-Aalen Estimator,**2.3** Interval Estimation of Survival Probabilities or Quantiles, **2.4** Asymptotic Properties of Estimators, Descriptive and Diagnostic Plots, Plots Involving Survivor or Cumulative Hazard Functions, Classic Probability Plots,**2.5** Estimation of Hazard or Density Functions**,** **2.6** Methods for Truncated and Interval Censored Data, Left-Truncated Data, Right-Truncated Data,Interval-Censored Data.

**UNIT 3**

**3.1** Semi-parametric Proportional Hazards Regression with Fixed Covariates - Coding Covariates, **3.2** Partial Likelihoods for Distinct-Event Time Data, Partial Likelihoods when Ties are present,**3.3** Local Tests, Discretizing a Continuous Covariate, Model Building using the Proportional Hazards Model, Estimation for the Survival Function; **3.4**Introduction to Time-Dependent Covariates; Regression Diagnostics :- Cox-Snell Residuals for assessing the fit of a Cox Model, **3.5** Graphical Checks of the Proportional Hazards Assumption, Deviance Residuals, Checking the Influence of Individual Observations

**UNIT 4**

**4.1** Inference for Parametric Regression Models - Exponential, Gamma and Weibull Distributions,**4.2** Nonparametric procedure for comparison of survival function, **4.3** Competing risk models – Basic Characteristics and Model Specification

**Text Books:**

1. Klein J.P. and Moeschberger M.L. (2003) Survival Analysis - Techniques for censored and truncated data, Second Edition, Springer-Verlag , New York,

**Reference Books**

1. Lawless J.F (2003) Statistical Models and Methods for Lifetime Data, Second Editon, John Wiley & Sons

2. Kalbfleisch J.D and Prentice, R.L. (2002) The Statistical Analysis of Failure Time Data, Second Edition, John Wiley & Sons Inc.

3. Hosmer Jr. D.W and Lemeshow S (1999) Applied Survival Analysis - Regression Modelling of Time to event Data, John Wiley & Sons. Inc. 3. Nelson. W (1982) Applied Life Data Analysis.
4. Miller, R.G. (1981) Survival Analysis, John Wiley.

## ST 81 04 02 POPULATION DYNAMICS

**Total credits-3Total hours-25 Weightage-30**

**Objectives:** By the end of this course students are expected to be able to understand and use various mortality rates , to construct life tables ,to calculate and use various characteristics of life time models and population growth models.

**UNIT 1**

**1.1** Sources of mortality data-mortality measures-ratios and proportions, crude mortality rates, specific rates- **1.2** standardization of mortality rates, direct and indirect methods,**1.3** gradation of mortality data, fitting Gompertz and Makeham curves.

**UNIT 2**

**2.1** Life tables-complete life table-relation between life table functions,**2.2** abridged life table-relation between abridged life table functions,**2.3** construction of life tables, Greville's formula, Reed and Merrell's formula- sampling distribution of life table functions, **2.4** multivariate pgf –estimation of survival probability by method of MLE.

**UNIT 3**

**3.1** Fertility models, fertility indices **3.2**-relation between CBR,GFR,TFR and NRR **3.3** stochastic models on fertility and human reproductive process, Dandekar's modified binomial and Poisson models, Brass, Singh models **3.4** models for waiting time distributions, Sheps and Perrin model.

**UNIT 4**

**4.1** Population growth indices, logistic model, fitting logistic, other growth models,**4.2** Lotka's stable population, analysis, quasi stable population, **4.3** effect of declining mortality and fertility on age structure, **4.4** population projections, component method-Leslie matrix technique, properties of time independent Leslie matrix-models under random environment.

**Text Books** :

1. Biswas S (2007) Applied Stochastic Processes-A Biostatistical and Population Oriented Approach, Second Edition, New Central Book Agency.

2. Pollard J.H (1975) Mathematical Models for the growth of Human population, Cambridge University Press.

**Reference Books**

1) Biswas S (1988) Stochastics processes in Demography and applications, Wiley Eastern.
2) Keyfitz N (1977) Applied Mathematical Demography A Wiley Interscience publication.
3) Ramkumar R (1986) Technical Demography, Wiley Eastern.
4) Srinivasan K (1970) Basic Demographic Techniques and Applications.

## ST 81 04 03-CATEGORICAL DATA ANALYSIS

**Total credits-3 Total hours-25 Weightage-30**

**Objectives:** To enable the students familiar with categorical data and various probability models associated with it.

**UNIT 1**

**1.1** Categorical variables, Introduction to Binary data, The linear probability models, The logit model, The Probit model, **1.2** the latent variable approach, the odds ratio, Relarive risks, Sensitivity and specificity, MNemar's test,**1.3** Binomial response models, log-log models, Likelihood ration Chi-squared statistic, Log-rate models, Time Hazard models, **1.4** Semi-parametric rate models.

**UNIT 2**

**2.1** Logistic Regression Analysis: Logit Models with Categorical Predictors Logistic Regression models, **2.2** regression diagnostics, Predictions, Interpreting parameters in logistic Regression. Inference for logistic Regression,**2.3** Multiple logistic regression.

**UNIT 3**

**3.1** Poisson regression: interpretations, regression diagnostics, Predictions**,3.2** negative binomial regression**, 3.3**Proportional hazards regression.

**UNIT 4**

**4.1** Principles of Bayesian statistics, Inference using simulations - Standard distributions**, 4.2** Understanding Markov Chain Monte Carlo, **4.3** The Gibbs sampler and the WinBUGS [Necessary topics from Chapter 1-5 of Ioannis Ntzoufras (2009)]

**Reference Books**

1) Agresti, A. (1990) Categorical Data Analysis. New York: John Wiley
2) Carlin, B.P. and Louis, T.A. (2000) Bayes and Emperical Bayes Methods for Data Analysis, Second Edition

3) Congdon P. (2006) Bayesian Statistical Modelling, Second Edition, John Wiley & Sons, Ltd. ISBN: 0-470-01875-5
4) Ntzoufras I. (2009) Bayesian Modeling using WinBUGS John Wiley & Sons Inc.
5) Powers D.A. (1999) Statistical methods for Categorical data analysis. Academic press Inc.
6) Shewhart,W.A. and Wilks, S.S. (2013) Case Studies in Bayesian Statistical Modelling and Analysis. Wiely.

## Bunch C

### ST 82 04 01 ACTUARIAL STATISTICS
**Total credits-3 Total hours-25 Weightage-30**

**Objectives:** To enable the students to get basics in the emerging field of actuaries and insurance and to determine the annuity, and determine the same based of the residual life

**UNIT 1**

**1.1** Insurance Business – Introduction, Insurance Companies as Business Organizations,**1.2** Concept of Risk; Future Lifetime Distribution and Life Tables –**1.3** Future Lifetime Random Variable, Curtate Future Lifetime,**1.4** Life Tables, Assumptions for Fractional Ages, Select and Ultimate Life Tables.

**UNIT 2**

**2.1** Actuarial Present Values or Benefit in Life Insurance Products –**2.2** Compound Interest and Discount Factor,**2.3** Benefit Payable at the Moment of Death, Benefit Payable at the End of Year of Death, , Relation between $A$ and $\bar{A}$ .

**UNIT 3**

**3.1** Annuities – Annuities Certain, Continuous Life Annuities, Discrete Life Annuities, Life Annuities with *m*thly Payments;**3.2** Premiums – Loss at Issue Random Variable, Fully Continuous Premiums, Fully Discrete Premiums,**3.3** True *m*thly Payment Premiums, Gross Premiums.

**UNIT 4**

**4.1** Reserves – Fully Continuous Reserves, Fully Discrete Reserves;**4.2** Multiple Life Contracts – Joint Life Status, **4.3** Last Survivor Status.

**Text Books**

1) Deshmukh, S.R. (2009) Actuarial Statistics – An Introduction using R, University Press (India) Pvt Ltd., Hyderabad, Chapters 1, 4, 5, 6, 7, 8 and 9.

**Reference Books**


1) Daykin, C.D, Pentikainen,T. et al, Practical Risk Theory of Acturies, Chapman and Hill .
2) Promislow, S.D (2006) Fundamentals of Actuarial Mathematics, John Wiley. Chapters 2-11 &14
3) Neill, A (1977) Life Contingencies, Heinemann , London.
4) King,G. Institute of Actuaries Text Book. Part 11, Second Edition, Charles and Edwin Layton, London.
5) Donald D.W.A.(1970) Compound Interest and Annuities, Heinemann, London.
6) Jordan, C.W.Jr.(1967) Life Contigencies, Second Edition, Chicago Society of Actuaries.
7) Spurgeen, E.T. Life Contigencies, 3rd Edition, Cambridge University Press.
8) Benjamin, B. and Pollard, J.H.(1980) Analysis of Mortality and other Actuarial Statistics, Second Edition, Heinemann, London.
9) Freeman,H.(1960) Finite Differences for Actuarial Students, Cambridge University Press.
10) Biandt-Johnson, R.C.andJohnson ,N.L(1980) Survival Models and Data Analysis, John Wiley


## ST 82 04 02 :APPLIED REGRESSION ANALYSIS

**Total credits-3Total hours-25 Weightage-30**

**Objectives:** By the end of this course they will be able to deal with various regression models and their interpretations.

**UNIT 1**

**1.1** Mathematical & Statistical models, Linear Model- estimability of parameters,**1.2** Linear Regression Model, Least squares estimation, Gauss Markov Theorem, BLUE,**1.3** Properties of the estimates, Distribution Theory, Maximum likelihood estimation, Estimation with linear restrictions, **1.4** Generalised least squares; Hypothesis testing - likelihood ratio test, F-test; **1.5** Confidence intervals,Residual analysis, Departures from underlying assumptions.

**UNIT 2**

**2.1** Polynomial regression in one and several variables, Orthogonal polynomials, **2.2** Indicator variables, Subset selection of explanatory variables, **2.3** stepwise regression and Mallows Cp -statistics, **2.4** Introduction to non-parametric regression.

**UNIT 3**

**3.1** Introduction to nonlinear regression, Least squares in the nonlinear case and estimation of parameters, **3.2** Models for binary response variables, estimation and diagnosis methods for logistic and Poisson regressions**.3.3** Prediction and residual analysis, **3.4** Generalized Linear Models – estimation and diagnostics.

**UNIT 4**

**4.1** Transformations and weighting to correct model inadequacies, Analytical methods for selecting a transformation, The Box-Cox method, Transformation on the regress or variables, **4.2** Ridge regression, Basic form of ridge regression , Robust regression Least absolute deviation regression,Least median of squares regression,**4.3** Inverse estimation- The calibration problem, Resampling procedures for regression models(Bootstrapping)

**Text Books:**

1. Seber, A.F. and Lee, A.J. (2003) Linear Regression Analysis, John Wiley, Relevant sections from chapters 3, 4, 5, 6, 7, 9, 10.

2. Montegomery, D.C., Peck, E.A. and Vining, G.G. (2001) Introduction to Regression Analysis, Third edition. Wiley.

3. B.Abraham and Ledotter, J. (1983) Statistical Methods for Forecasting, John Wiley & Sons.

**Reference Books:**

1. Searle, S.R. (1971) Linear models, John Wiley & Sons, Inc.

2. N.Draper and H. Smith (1986) Applied Regression Analysis – John Wiley & Sons.

3. Fox, J. (1984) Linear Statistical Models and Related methods, John Wiley, Chapter 5.

4. Christensen, R. (2001) Advanced Linear Modeling, Chapter 7.

## ST 82 04 03 :DATA MINING

### Total credits-3 Total hours-25 Weightage-30

**Objectives:** To enable the students to handle data mining and the related methodologies and problems.

**Unit 1**
**Introduction to Data Mining**

**1.1** Data Mining for Business Intelligence, Data Mining Goes to Hollywood**!, 1.2** Data Mining Concepts and Definitions,**1.3** Characteristics, and Benefits, How Data Mining Works,**1.4** Data Mining Applications.

**Unit 2**
**Data Mining Process**

**2.1** Data Mining Process, Step 1: Business Understanding, Step 2: Data Understanding, Step 3: Data Preparation, Step 4: Modelling Building, Step 5: Testing and Evaluation, Step 6: Deployment, **2.2** Other Data Mining Standardized Processes and Methodologies.

**Unit 3**
**Data Mining Methods**

**3.1** Data Mining Methods, Classification, **3.2** Estimating the True Accuracy of Classification Models**, 3.3** Cluster Analysis for Data Mining.

**Unit 4**
**Artificial Neural Networks**

**4.1** Association Rule Mining, Artificial Neural Networks for Data Mining, Elements of ANN, Applications of ANN. **4.2** Data Mining Software Tools, Data Mining Myths and Blunders.

References

1. Turban, Sharda Efraim, Ramesh, Dursun Delen and King, David. (2011).Business Intelligence : A Managerial Approach, 2nd Edition. Publisher :Prentice Hall.

2. Han,Jiawei and Kamber, Micheline. (2012). Data Mining: Concepts and Techniques, 3rd edition. Morgan Kaufman Publishers.

3. Tang, P.N., Steinbackm, M. And Kumar, V. (2006). Introduction to Data Mining. Addison Wesley.

4. Myatt, Glenn and Johnson, Wayne. (2009). Making Sense of Data II. John Wiley&Sons.
5. Rajaraman, Anand. (2011). Mining of Massive Datasets. New York: Cambridge University Press.

# MODEL QUESTION PAPERS

# SEMESTER 1

**QP Code**

**Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION , ……**

**First Semester**

**Faculty of Science**

**ST 01 01 01 - PROBABILITY AND MEASURE THEORY**
**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

**Section -A**

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

1. Show by an example that union of two fields need not be a field.
2. Define a measureable function.
3. Show that f(x) = $e^{-x^2}$, -∞< x < ∞ is a Lebesgue measureable function.
4. Show that Borel functions of random variables are random variables. Give an example.
5. Show that all Borel sets will be independent if the class of intervals {(- ∞, x), x ϵ R} is mutually independent. State the result you used.
6. Show that convergence almost surely implies convergence in probability. State the result you used.
7. Define weak and complete convergence of distribution functions.
8. Define limit of a sequence of sets
9. State Helly –Bray theorem
10. State Jensen's inequality (8 x 1 =8)

## Section B

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

11. . Examine whether limit of the following sequence exists, if so find the limit:

$$A_n = \begin{cases} (0,5), & if\ n\ is\ odd \\ (10,15), & if\ n\ is\ even \end{cases}$$

12. If f and g are measureable functions, show that f + g and f.g are measureable functions.
13. State and prove continuity property of measures.
14. Explain how probability measure defined for a class of intervals of R which can generate the Borel field be uniquely extended to all Borel sets.

15. Obtain expectation of the mixture random variable with distribution function

$$F(x) = \begin{cases} 0, & if\ x < 0 \\ x, & if\ 0 \leq x < \frac{1}{2} \\ x^2 + \frac{3}{8}, & if\ \frac{1}{2} \leq x < \sqrt{\frac{5}{8}} \\ 1, & if\ x \geq \sqrt{\frac{5}{8}} \end{cases}$$

16. Show that Student's 't' statistic converges in distribution to normal . State the result you used.
17. Let $X_1, X_{2, \ldots,} X_n$ be a random sample from the uniform distribution in $(0, \theta)$. If $Y_n$ = Max($X_1, X_{2, \ldots,} X_n$), show that $Y_n \rightarrow \theta$ almost surely. State the result you used.

( 4 x 3 = 12 )

18. State and prove Bayes' theorem for a finite number of events. (6 x 2 =12)

## Section C

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

19. (a) Let Ω be a countably infinite set, and let $\mathcal{F}$ be all subsets of Ω. Define μ(A) = 0, if A is finite; μ(A) = ∞, if A is infinite. Show that μ is finitely additive, but not countably.

(b) Let μ be a finite measure on a σ-field and let $\lim A_n$ = A. Show that μ(A) = lim.μ($A_n$).

20. (a) State and prove domonated convergence theorem.

(b) Let F be the distribution function on R given by F(x) = 0, x < -1; F(x) = 1 + x, -1 ≤ x < 0; F(x) = 2 + $x^2$, 0 ≤ x < 2; F(x) = 10, x ≥ 2. If μ is the Lebesgue-Stieltje's measure induced by F, compute the measure of the sets: (i) {2}, (ii) $[\frac{-1}{2}, 3)$, (iii) $\{ x : |x| + 2x^2 > 1 \}$, (iv) $[0, \frac{1}{2}) \cup (1, 2]$

21. (a) Show that the set of discontinuity points of a distribution function is countable. Show that every distribution function can be writtten as a convex combination of a discrete and continuous distribution functions.

(b) What is the chance of observing the sequence of 1,3,5 when an unbiased die is thrown indefinitely? State the result used.

22. (a) Show how Techebycheff's inequality be obtained from Markov's inequality. Also, show that values of Karl Pearson's correlation coefficient lies between -1 and 1.

(b) Show that convergence in probability implies that in distribution under conditions to be stated, and give an example. Establish the converse of this result.

( 5 x 2 = 10 )

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied) DEGREE (C.S.S) EXAMINATION ,……**

**First Semester**

**Faculty of Science**

**ST 50 01 01 DISTRIBUTION THEORY**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

**Section -A**

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

1. Define probability generating function and derive that of negative binomial distribution
2. If X follows B( n, p), then obtain the distribution of Y= n-X
3. Find the first moment of a Poisson distribution which is truncated at zero

4. Let $X_1$ ,$X_2$ be a random sample of size two from a lognormal population with parameters $\mu$and $\sigma^2$ .What is the distribution of $\sum_{i=1}^{2}(\frac{logX_i-\mu}{\sigma})^2$.

5. Show that the two parameter Gamma distribution belongs to the Exponential family of distributions.

6. Give any distribution for which moments exist but MGF does not.

7. If X is a $U(-\frac{\pi}{2},+\frac{\pi}{2})$ write the pdf of $Y = \tan X$.

8. Define a non-central $t$ statistic, mentioning its parameters.

9. If $X_1, X_2, \ldots$, Xn are iid random variables having Logistic distribution, obtain the pdf of

$X_{(1)} = \min(X_1, X_2, \ldots, Xn)$.

10. If X and Y are independent chi-square random variables with m and n degrees of freedom respectively , find the distribution of X-Y if m > n

(8x1= 8)

## SECTION B

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

11. If $P(X = n) = p_{n,}$and $P(X \leq n) = q_{n,}$ so that $q_n = p_0 + p_1 + \cdots + p_n$, then show that $\sum_{n=0}^{\infty} P(X \leq n)s^n = \frac{P_X(s)}{1-s}, |s| \leq 1$ and $P_X(s)$ is the probability generating function of X.

12. Show that X and Y are Geometric random variables, if and only if they are independent and identically distributed with $P(X = s/X + Y = s) = P(X = s - 1/X + Y = s) = \frac{1}{s+1}$ .

13. Explain the Banach's match box problem and hence write the pdf of the Negative Binomial distribution.

14. Derive the moment generating function of a Laplace/ Double exponential distribution.

15. Identify the distribution, when an exponential r.v is compounded with a 2 parameter Ga

16. If $X_{1,}X_{2,}X_3$are iid standard Normal random variables then show that $Y_1 = \left(\frac{X_1-X_2}{\sqrt{2}}\right)$, $Y_2 = (\frac{X_1+X_2-2X_3}{\sqrt{6}})$ and $Y_3 = (\frac{X_1+X_2+X_3}{\sqrt{3}})$ are independent and identify the marginal pdfs.

17. Define an F ($n_1$,$n_2$) statistic. Show that $Y = \frac{1}{1+\frac{n_1 X}{n_2}}$ has a $\beta_1(\frac{n_2}{2},\frac{n_1}{2})$ distribution. Deduce that $P(X \leq x) = 1 - P\left[Y \leq \left(1+\frac{n_1 x}{n_2}\right)^{-1}\right], 0 \leq x < \infty.$

18. Derive the distribution of the Sample Median while a random sample of size n is taken from a continuous U(a,b)population. Find the mean and variance when n is odd.

**(6x2=12)**

## SECTION C

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

19. Define a Modified Power Series distribution?. Establish the recurrence relation for the central moments of it. Also find the expressions of the mean and variance of MPSD. Hence show that it is a special case of the Generalized Power Series Distribution.

20. Define the Pearson family of distributions. Derive the four linear equations by which the Pearson's system of distributions is completely specified.

21. If a random sample of size n is taken from a U(0,1) population, and if $X_{(1)} = Y_1 Y_2 \dots Y_n$, $X_{(2)} = Y_1 Y_2 \dots Y_{n-1}$, ….,$X_{(n)} = Y_1$ , then show that $Y_1, Y_2, \dots, Y_n$ are independent. Also find their marginal pdfs .

22. Derive the pdf of a non-central $\chi^2$ random variable. Write its characteristic function and hence obtain expressions for the mean and variance.

**(2x5=10)**

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied)DEGREE (C.S.S) EXAMINATION , ……**

**First Semester**

**Faculty of Science**

**ST 50 01 02 --ANALYTICAL TOOLS FOR STATISTICS**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

## SECTION A

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

1. Define Idempotent and Hermitian matrices.
2. Do the vectors $a_1=(3,0,2)$, $a_2=(7,0,9)$, $a_3=(4,1,2)$ form a basis for $R_3$?
3. Show that the vectors (1, 2, 3) and (2,-2,0) form a linearly independent set.
4. Define dimension of a vector space.
5. When do you say a quadratic form $X'AX$ to be positive definite and positive semi-definite.
6. Prove that every non- singular matrix is a product of elementary matrices.
7. Explain Moore-Penrose inverse $A^+$ of A.
8. Show that the trace of a matrix is the sum of its eigen values
9. If A is a positive definite matrix, then show that $|A| > 0$
10. Show that a real symmetric matrix has only real characteristic roots .

(8 X 1=8)

## SECTION B

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

11. Let V be a finite dimensional vector space. Show that all bases of V have same number of elements.

12. Let S be a subspace of a finite dimensional vector space. Then prove that every generating set C of S contains a basis of S.
13. If A is an m*m Idempotent matrix, then show that (a) $I_m$- A is also idempotent. (b) Each eigen value of A is 0 or 1.
14. Using Cayley –Hamilton theorem obtain the inverse of the matrix $\begin{matrix} 6 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 2 \end{matrix}$
15. Prove that the geometric multiplicity of a characteristic root cannot exceed algebraic multiplicity of the same.

16. Classify the following quadratic form as positive definite, positive semi-definite and indefinite $2x2+ 2y2 + 3z2$ - 4yz- 4zx+ 2xy.
17. Show that $A$ is the g-inverse of A if and only if $A\bar{A}A = A$
18. If $A^-$ is the g-inverse of A , the show that rank (A) = rank($AA^-$)= trace($AA^-$)

(6 X 2 = 12)

## SECTION C

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

7 (a) Let $X1, X2$ ,…$Xn$be the characteristic vectors corresponding to distinct characteristic roots of a matrix. Prove that $X$i 'sare linearly independent.

b) For a real symmetric matrix show that characteristic vectors corresponding to distinct characteristic roots are orthogonal.

8 State and prove Cayley –Hamilton theorem .

9 What do you mean by matrix mapping. Write a short note on change of basis in matrix mapping.

10 (a) Define the rank of a matrix. Prove that the rank of the product of two matrices cannot exceed the rank of either matrix.

(b) Reduce the following matrix to its normal form and hence find its rank , $\begin{matrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 0 & -1 & -1 \end{matrix}$

(2 X 5 = 10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied)DEGREE (C.S.S) EXAMINATION , ……**

**First Semester**

**Faculty of Science**

**ST 50 01 03 SAMPLING THEORY**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

## SECTION A

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

1. Describe the random number table method of selection of a simple random sample.

2. What is a sampling frame.?

3. Distinguish between stratified and cluster sampling.

4. Explain Census and Sampling. Why sampling is preferred?.

5. Describe the situation in which two stage sampling is better than simple random sampling.

6. Describe the use of auxillary information in sampling.

**7.** Explain Lahiri's method under PPS.

8. What is a difference estimator?

9. Explain linear systematic sampling.

10. What are non- response errors?

( 8 x 1 = 8)

## SECTION B

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

11. Show that in SRSWOR Sample mean $\overline{y}$ is the BLUE of $\bar{Y}$.

12. Explain Principles of Sampling theory.(b) Explain various factors of non sampling errors.

13.(a) Explain Horvitz-Thompson estimator.(b) Explain Hartley –Ross estimator; obtain the corresponding unbiased estimator of the population total.

14. (a) Show that Var ($\overline{y_{sys}}$) = $\frac{N-1}{Nn}$ (1+ (n-1) ρ)$S^2$, where ρ is the interclass correlation between the units of the same systematic sample.(b)Explain Circular systematic sampling with the help of an example.

15. (a) Carry out a comparison between the mean per unit and ratio estimator.(b) Distinguish between stratum and cluster. Also give suitable examples.

16. Differentiate between Cumulative Total Method and Lahiri's method with the help of an example.

17. What is Multi-Phase Sampling? Why it is differ from Multistage Sampling; Explain?

18. Obtain the mean and its variance in equal cluster sampling. Suppose NM units in the population are grouped at random into N clusters of M units each. Show that the sampling of n clusters by srswor should have the same efficiency as sampling of nM units by srswor.

(6 x 2 = 12)

## SECTION C

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

19. (a) Explain the methods of allocation in stratified sampling and find efficiency of variances.(b) If the population consists of liner trend, then prove that

Var $\overline{(Yst)}$ ≤ Var $(\overline{Ysys})$ ≤ Var $(\overline{Yran})$.

20.(a) Explain Principle Steps in a Sample Survey.(b)A shelf in a library contains 48 books, numbered serially. Select (i) a simple random sample of books by 8 draws with replacement, and (ii) a simple random sample of 8 books without replacement.

21. (a)Give any three estimators of population mean in cluster sampling where clusters are of unequal size and discuss their properties.(b) Show that sample proportion, $p$ is an unbiased estimate of population proportion, P. Also obtain the confidence interval for the population proportion.

22 (a) Prove that in PPS sampling without replacement, Desraj ordered estimator is unbiased for population total. Derive its sampling variance.(b) Explain Murthy's unordered estimator.(c) For an SRSWOR with population size *N* and sample size *n*, show that the probability of a specified unit being selected at any given draw is 1/N.

(5 x 2 =10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION**

**First Semester**

**Faculty of Science**

**ST 01 01 02 STATISTICAL COMPUTING I- USING R**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

(Answer any **THREE** questions. Each question carries a weight of **10)**

1**.(A)**Heights(in cm) of father and son are given as follows.Calculate karl pearson's correlation coefficient and Fit a regression line predict the height of son given the height of father and predict the height of son when height of father is 158.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Father(X) | 150 | 152 | 155 | 157 | 160 | 161 | 164 | 165 |
| Son(Y) | 154 | 156 | 158 | 159 | 160 | 162 | 161 | 164 |

**(B)**Explain different control loops and statements in R. Write a program in R to find factorial of a number using while loop.

2. **(A)**Using R functions, write a program to generate random sample of size 100 from a normal population with mean 2 variance 9 and then construct P-P plot of the generated sample.

**(B)** Fit a Negative Binomial distribution to the following data

| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| f | 5 | 7 | 12 | 15 | 11 | 9 | 6 | 5 |

Also test the adequacy of model using goodness of fit in R.

3. **(A)** A sample survey is to be under taken to ascertain the mean annual income of farms in a certain area. The farms where stratified according to their principal products. A Census conducted several years earlier gave the following information

| Type of Farm | Number of Farm | Mean Annual Income | S.D |
|---|---|---|---|
| Sheep | 161 | 10946 | 2236 |
| Wheat | 195 | 6402 | 2614 |
| Dairing | 274 | 2228 | 606 |
| Others | 382 | 1458 | 230 |

For a sample of 12 farms compute the sample sizes in each stratum under Proportion allocation and optimum allocation. Compare the precision of these methods with that of simple random sampling.

**(B)** Write R code to find the optimum sample size and variance under proportion allocation and optimum allocation for the above data.

4. **(A)** For estimating the total number of absentees in 325 factories situated in a state, a sample of 40 factories was drawn with S. R. S without replacement. The data given below shows the no of workers (x) and the no. of absentees (y) for each of the 40 factories. Total no. of workers in all the factories is known to be 27,000.

| Sl. No | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 96 | 80 | 31 | 46 | 29 | 141 | 124 | 80 | 42 | 52 |
| Y | 9 | 7 | 8 | 3 | 2 | 9 | 8 | 11 | 5 | 3 |

| Sl. No | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| X | 141 | 90 | 56 | 130 | 46 | 44 | 116 | 64 | 102 | 51 |
| Y | 14 | 5 | 4 | 12 | 3 | 10 | 11 | 9 | 8 | 8 |

Estimate the total number of absentees in all a 325 factories of the state using regression method of estimation with number of workers as auxiliary variable.

Compute the relative precision of the regression estimate to the ration estimate and comment.

**(B)** Write R code to estimate the population mean using linear regression method for the above data. Also write R code to find the variance of the estimates under regression method and ratio method of estimation.

5 **(A)** Examine the definite nature of the quadratic form

$$24x^2 + 15(y^2 + z^2) + 24w^2 - 4(xw + yz) + 6(y - z)(x + w)$$

**(B)**Obtain Moore-Penrose g-inverse of $\begin{pmatrix} 1 & -1 & 2 \\ 2 & 6 & 3 \\ 3 & 13 & 4 \end{pmatrix}$

6 **(A)** Describe different types of data objects/structures and their operations in R. Explain the difference between data frame and a matrix in R?

**(B)** Write down the spectral decomposition of

A= $\begin{pmatrix} 5 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{pmatrix}$ and obtain $A^{100}$.

## SEMESTER 2

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied)DEGREE (C.S.S) EXAMINATION , ……**

**Second Semester**

**Faculty of Science**

# ST 50 02 01 ESTIMATION THEORY

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

## S ECTION A

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

1. What is the difference between an estimate and an estimator?. Give an example of an estimate which is consistent but biased.
2. Define ancilliary statistic with an example. State and prove Basu's theorem.
3. Define completeness. Give an example of a family of distribution which is not complete.
4. Differentiate between the method of minimum chi-square with modified minimum Chi square method.
5. What is loss function? Explain commonly used loss functions.
6. Explain Fisher's Scoring method.
7. Explain posterior and prior distributions with examples.
8. What do you mean by Fisher information?
9. Define a two parameter exponential family of distributions? Is it complete always?
10. Define risk function associated with a decision rule.

(8 X1=8)

## SECTION B

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

11. Define sufficiency and State and prove Fisher Neyman Factorization theorem. State the condition under which the sample mean is a sufficient estimator of the population mean for a normal population and establish it.

**12.** S.T if $t_n$ is an estimator of $\theta$, then $\frac{\partial \log L}{\partial \theta}$ is a function of $t_n$ and $\theta$ only. Give an example of mle which is not unbiased but consistent.

.13. . Find the CRLB for the estimation of $\theta$ in $f(x,\theta) = \frac{1}{\theta} e^{-x/\theta}; 0 < x < \infty, 0 < \theta < \infty$.

14.(a) State and prove Rao-Blackwell theorem. (b) Define UMVUE and give one example.

15. What do you mean by exponential family .Check whether Cauchy distribution belongs to this family?

16. Explain the shortest confidence interval. Let $X_1, X_2, X_3 ... X_n$ be a sample from $U(0,\theta)$. Find the shortest C.I for $\theta$ .

17. State the properties of mle. Find the mle for $\theta$ based on n observations for the frequency distribution $f(x,\theta)=(1+\theta)x^{\theta};0<x<\theta$ . $f(x,\theta)=(1+\theta)x^{\theta};0<x<\theta$

18. Find Baye's estimator for p in B(n,p) when the loss function is $(p-d(X_1))^2$ and prior is uniform.

(6 X 2= 12)

## SECTION C

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

19. State and prove Fisher- Neyman factorization theorem.

20. State and prove Cramer Rao inequality.

21 (a) S.T minimum chi-square estimate implies mle, when the sample size n is large.

(b) Explain the concept of mle. S.T mle need not be unique.

22. (a) Define loss function and mention commonly used loss functions.

(b) P.T the Baye's risk can be obtained by minimizing posterior risk.

(c) Define randomized and non randomized decision rules.

(5 X2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied)DEGREE (C.S.S) EXAMINATION , ……**

**Second Semester**

**Faculty of Science**

**ST 50 02 02 STOCHASTIC PROCESSES**

**( 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 state space and time space of a stochastic process

2. Define : (i) Markov chain (ii) Transition Probability Matrix .

3.Explain non homogeneous Poisson processes.

4. What is stationary distribution.

5. Describe a discrete time branching process

6. Write down the postulates of a Poisson process.

7. Define renewal process and renewal function

8. What do you mean by a queue? Briefly explain Kendall's notation.

**9.** Define a Brownian motion process.

10. Define weakly stationary and strictly stationary process

(8 X 1=8)

## SECTION B

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

11. (a) Show that recurrence is a class property (b) Check whether a one dimensional random walk is recurrent.

12..(a) Find the stationary distribution for the transition probability matrix $\begin{bmatrix} 0 & \frac{2}{3} & \frac{1}{3} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$

.

.(b) Bring out the relation between Poisson process and Binomial distribution.

13.(a) Compute the density function Tx , the time until Brownian motion hits x.(b) Describe a renewal reward process.

14.(a) Establish Wald's equation.(b) State and prove Chapman-Kolmogorov equations for markov chains.

15. . What do you meant by steady state solution in a queuing process? Explain the role of Poisson process and Exponential distribution in queuing models.

16. .Find the steady-state probability distribution for $M/M/1$ queue

17. Explain Yule-Furry process. Obtain its probability distribution.

18. Explain Galton- Watson branching process. Is it a Markov chain? Establish.

(6 X 2=12)

**SECTION C**

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

19.(a) State and prove elementary renewal theorem. (b) Examine the nature of the Markov chain $\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{3}{8} & \frac{1}{8} & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}$

20.(a) Find the steady-state probability distribution for $M/M/S$ queue.(b) When do you say that a stochastic process has independent increments? Show that a process with independent increments is Markovian.

21. (a) Establish the relation between probability generating functions of off spring random variable and $n^{th}$ generation size in Galton –Watson branching Process. Derive its mean and variance also.

22.(a) Differentiate between Hidden Markov Chains and Semi markov process.(b) Explain conditional mixed Poisson processes.(c) Prove: The stochastic matrix and the initial distribution completely specify a Markov chain.

(5 X2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied)DEGREE (C.S.S) EXAMINATION , ……**

**Second Semester**

**Faculty of Science**

**ST 50 02 03 :MULTIVARIATE DISTRIBUTIONS**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

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

**SECTION A**

1. Show that marginal distribution of multinomial probability distribution is also multinomial.
2. Define a singular multivariate normal distribution

3 . Give . the criterion for testing $H_0$ : μ = $\mu_0$ against μ ≠ $\mu_0$when X ~$N_p$(μ, ∑ ), ∑ known.

4. What is the form of the characteristic function if a three-component vector follows a multivariate normal distribution with mean vector $\mu = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ and dispersion matrix $\sum = \begin{bmatrix} 2 & 5 & 7 \\ 5 & 1 & 2 \\ 7 & 2 & 3 \end{bmatrix}$

5. Give the Jacobian of the transformationY = AXB where X is a square matrix of real random variables.

6. Define matrix variate gamma distributions with clearly stating the arguments.

8. State Cochran's theorem on the distribution of quadratic forms.

9..Obtain the expression of the partial correlation coefficient $\rho_{12.3}$in terms of total correlation coefficients.

10. Define Marshall –Olkin bivariate exponential distribution. Comment on it.

(8 X1=8)

## SECTION B

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

11. .If (X, Y) ~ $N_2$(-3, 10: 25, 9: 0.6), find (a) P[-5 <X< 5] and (b) P[-5 < X < 5 / Y = 13]

12. .If $\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \sim N_3(\mu, \Sigma) with\ \mu = \begin{pmatrix} -2 \\ 5 \\ -1 \end{pmatrix}$ and $\Sigma = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \\ 1 & -1 & 6 \end{bmatrix}$, find the probability distribution of $X_1$-3$X_2$+4$X_3$ and 2$X_1$ + $X_2$ - $X_3$

13. Find the means, variances and correlation coefficient if the exponent of a bivariate normal distribution is $\frac{-2}{3}$ [( x - 2)$^2$ – 3( x - 2)( y + 3 ) + ( y + 3 )$^2$ ]

14..Explain four properties of Wishart distribution.

15.Show that $\bar{X}$ and S are independently distributed when sampling is fom a multivariate normal population.

16. If $X \sim N_p(0, I)$, give a necessary and sufficient condition for the independence of a quadratic form $X'AX$and the linear form $B'X$.

17 .If X~ $N_p$(0,I), obtain a necessary and sufficient condition for the quadratic form X$'$AX to be distributed as $\chi^2$ distribution.

19. .Let X~$N_p(0, \Sigma)$, then state and prove a necessary and sufficient condition for the independence of the quadratic forms $X'AX$ and $X'BX$ where $A$ and $B$are real symmetric matrices .

(6 X 2=12)

## SECTION C

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

19. Obtain the marginal and conditional distributions of Gumbel's bivariate exponential distribution. Also, give two properties of the distribution.

20 (a) Obtain the MLEs of the parameters when sampling is from Multivariate Normal population.

(b) Establish a necessary and sufficient condition for the independence of any two sub-vectors of a multivariate normal random vector.

21(a)Derive the Jacobian of the transformation X = TT′ where T is a lower triangular matrix of

variables.

(b) What is generalized variance? Derive its distribution.

22.(a) If r is the sample correlation coefficient from a bivariate distribution, derive the distribution of $U = \frac{r\sqrt{n-1}}{\sqrt{1-r^2}}$. assuming the population correlation as 0.

(b) Derive the null distribution of the multiple correlation coefficient.

( 5 X 2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION**

**Second Semester**

**ST 01 02 01 ADVANCED PROBABILITY THEORY**

**Faculty of Science**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

**SECTION A**

(Answer any **8** questions .Each question carries a Weight of **1** )

1. Define non negative definiteness. Show that the characteristic function of a r.v is non negative definite.
2. Derive characteristic function of standard Laplace distribution.
3. Explain Domain of attraction and stable distributions.
4. Illustrate a situation where WLLN doesn't hold.
5. Show that every sequence of independent r.vs with uniformly bounded variances obeys SLLN.
6. Examine whether CLT holds for the sequence of independent r.vs with $P[X_n = 0] = 1 - n^{1/2}$ and $P[X_n = n] = P[X_n = -n] = 1/2n^{1/2}$, n= 1,2,…
7. State Radon-Nikodym theorem.
8. State De Moiver- Laplace limit theorem
9. State Bochner's theorem
10. Check whether $\emptyset(t) = cos^2\ t + i sin^2\ t$ is a characteristic function

(8X1=8)

(

**SECTION B**

(Answer any **6** questions . Each question carries a Weight of **2** )

11. State inversion theorem for characteristic functions. If X is an integer valued r.v, show that P(X= j) = $\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{itj}\, \emptyset(t)dt$
12. State and prove uniqueness theorem of characteristic functions.
13. State Khinchines WLLN. Show that Weak Law of Large numbers does not hold for standard Cauchy distribution.
14. Distinguish between WLLN and SLLN. Give an example of a sequence of independent r.vs obeying SLLN.
15. State Lindberg –Levy CLT. Explain how CLT can be viewed as a generalization of law of large numbers.
16. Describe product measure space and state Fubini's theorem.
17. Define martingales and briefly describe its simple properties. Show that a sequence of partial products of independent r.vs with mean one is a martingale.
18. Prove that $|E(X|\mathcal{G})| \leq E(|X||\mathcal{G})$, Where $\mathcal{G}$ is a sigma field.

(6X 2=12)

**SECTION C**

(Answer any **2** questions .Each question carries a Weight of **5** )

19. State and prove Hahn-Jordan decomposition theorem.
20. State and prove continuity theorem on characteristic functions.
21. State and prove Kolmogorov SLLN for independent r.v's.
22. State and prove Liapounov's CLT.

(5X2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION**

**Second Semester**

**Faculty of Science**

**ST 01 02 02 STATISTICAL COMPUTING II–USING R**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

(Answer any **three** questions without omitting any part **.Each question carries a Weight of 10**)

**Section A**

1. The mean vector and dispersion matrix of 4 anthropometric characteristics for a district are given below. Find marginal distribution of (i) $X_3$ ( ii) $(X_1, X_2)$ (iii) $3X_2 + 5X_3 + X_4$ and (iv) the conditional distribution of $X_2$ given $X_1 = 4, X_3 = 5, X_4 = 4$.

Characteristics Mean Vector Dispersion matrix

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} -6 \\ 1 \\ 0 \\ 3 \end{bmatrix} \qquad \begin{bmatrix} 2 & 0 & 3 & 0 \\ 0 & 5 & 0 & 2 \\ 3 & 0 & 5 & 0 \\ 0 & 2 & 0 & 1 \end{bmatrix}$$

2.(a) A random sample of size 72 from a 4 variate Normal distribution gave the sample mean vector (22.3 14.5 7.9 12.8)' and sample dispersion matrix as:

$$\begin{bmatrix} 13.0552 & 4.1740 & 8.9620 & 2.7332 \\ & 4.8250 & 4.0500 & 2.0190 \\ & & 10.8200 & 3.8520 \\ & & & 1.1962 \end{bmatrix}$$

Test whether the population mean vector is ( 20 15 8 14)' at 1% level.

(b) Obtain $R_{1.23}$ and $r_{12.3}$ for the following data.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Profit(in lakhs): | 2 | 6 | 4 | 3 | 7 | 5 | 9 |
| Production(in 1000 units): | 16 | 14 | 18 | 21 | 15 | 14 | 23 |
| Sales(in 1000 units): | 13 | 11 | 14 | 18 | 13 | 10 | 17 |

Also write the regression equation of Profit on production and Sales. Estimate the Profit when production =30 and Sales = 25.

**Section B**

3. Based on the following 20 independent observations from an exponential distribution with pdf $f(x, \theta) = 1/\theta\, e^{-x/\theta}$ , x>0 , θ >0. Obtain the UMVUE of (1) θ (2) 1/ θ (3) P(X>1).
The observations are 2 , 8, 16, 4, 5, 9, 21, 22, 13, 7
3, 19, 17, 1, 6, 10, 20, 23, 14, 11

4. Let X be a random variable with pdf $f(x) = \frac{x}{\theta} e^{-x^2/2\theta}$ , x>0, θ>0. Estimate θ from the following.

(a) n= 10 observations using the method of moments

16.88, 10.23, 4.59, 6.66, 13.69,

14.23, 19.87, 9.4, 6.51, 10.95

(b) Also estimate θ by the method of MLE.

(C) Compare their standard errors

**Section C**

5. (a)Consider two urns with 6 White and 5 Red balls respectively. One ball is taken from each urn and interchanges them. Let $X_n$ denote the number of White balls in Urn II after $n^{th}$ transition and $Y_n$ denote the number of White balls in Urn I after $n^{th}$ transition. Write down the State space, Time space and One step transition probabilities of $\{X_n\}$ and $\{Y_n\}$.

(b)A Markov Chain has states $\{0,1,2,3\}$ and transition probability matrix is given by

$$\begin{bmatrix} \frac{1}{2} & \frac{7}{24} & \frac{1}{6} & \frac{1}{24} \\ \frac{1}{6} & \frac{11}{24} & \frac{1}{6} & \frac{5}{24} \\ \frac{1}{2} & \frac{7}{24} & \frac{1}{6} & \frac{1}{24} \\ \frac{1}{6} & \frac{5}{24} & \frac{1}{6} & \frac{11}{24} \end{bmatrix}$$ . Find the Stationary distribution

6. A Markov chain has states {1, 2, 3, 4, 5} and transition probability matrix

$$P = \begin{bmatrix} .40 & .10 & .20 & .10 & .20 \\ .20 & .30 & .25 & .1 & .15 \\ .55 & .15 & .05 & .1 & .15 \\ .20 & .20 & .30 & .15 & .15 \\ .25 & .15 & .30 & .10 & .20 \end{bmatrix}.$$

(i) Find 3 step transition probability matrix (ii) Find out

$P$ (system reaches state 5 from state 2 at the 3$^{rd}$ step for the 1$^{st}$ time)

# SEMESTER 3

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied)DEGREE (C.S.S) EXAMINATION , ……**

**Third Semester**

**Faculty of Science**

**ST 50 03 01 TESTING OF HYPOTHESES**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

## SECTION A

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

1. Define two types of errors and power of a test.
2. Explain how the best critical region is determined?
3. Obtain the MP test for testing the mean $\mu = \mu_0$ against $= \mu_1$ , $\mu_1 > \mu_0$ when $\sigma^2=1$ in normal population.
4. Give an example of a family of distributions with MLR property and justify it.
5. Explain the role of Neyman Structure in deriving UMPU test.
6. Describe Wald's SPRT. Derive the approximate expression for the O.C. function.
7. Derive the relation connecting the boundary values A, B and strength $(\alpha, \beta)$ of an SPRT.
8. Explain Wilcoxon signed rank test.
9. Describe Kolmogorov – Smirnov two sample test statistic
10. Explain the test procedure for testing the equality of means of two normal populations.

( 8 X 1=8)

## SECTION B

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

11. (a) State and prove Neyman-Pearson lemma.

(b) Show that if a sufficient statistic T exists for a family, then Neyman-Pearson MP test is a function of T.

12. State and prove a set of sufficient conditions for a similar test to have Neyman structure.
13. Define M.P region and U.M.P region. Show that a M.P. region is necessarily unbiased.
14. Show that for a normal population with zero mean and variance $\sigma^2$., the best critical region for $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma = \sigma_1$ is of the form , $\sum_{i=1}^{n} x_i{}^2 \leq a_\alpha$ for $\sigma_0 > \sigma_1$ and $\sum_{i=1}^{n} x_i{}^2 \geq b_\alpha$ for $\sigma_0 < \sigma_1$.
15. Show that the LR test for testing the equality of variances of two normal populations is the usual F-test.
16. Using Walds fundamental identity,derive the ASN function when E(z) = 0.
17. (a)Describe median test.

    (b)What are the advantages and drawbacks of non-parametric methods over parametric methods.

18.Explain Kruskall Walli's test for one way ANOVA.

(6 X 2=12)

## SECTION C

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

18. Given a random sample of size n from the distribution with p.d.f f(x,$\theta$) = $\theta e^{-\theta x}$, x$>$ 0Show that there exists no UMP test for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$
19. (a)Explain how the sequential test procedure differs from the Neyman –Pearson test procedure.

    (b)Develop SPRT for testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda = \lambda_1$ ,($\lambda_1$>$\lambda_0$)when $\lambda$ is the mean of a Poisson distribution. Also obtain its OC and ASN functions.
20. Describe the general methodof construction of Likelihood Ratio test. Discuss the properties of the test.
21. (a) Compare Chi-square test and Kolmogorov-Smirnov Test.

(b) Explain Kruskall–Wallis one-way analysis of variance and Friedman's two-way analysis of variance.

(5 X 2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied)DEGREE (C.S.S) EXAMINATION , ……**

**Third Semester**

**Faculty of Science**

**ST 50 03 02 –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. Explain standard Gauss- Markov model in linear estimation
2. Explain the concept of confounding and its need in factorial experiments?
3. The random variables $Y_2, Y_2, Y_3, Y_4$ are independent with common variance $\sigma^2$ and $E(Y_1)=E(Y_3)=\theta_1+\theta_2$ $E(Y_2)=E(Y_4)=\theta_1+\theta_3$. Verify whether $\theta_1-\theta_2$ and $2\theta_1+\theta_2+\theta_3$ are estimable
4. Analysis of Covariance improves the precision of the experiments- Justify this.
5. Explain Graeco latin square design.
6. Identify the interaction confounded in the following $2^4$ factorial experiment.
   Block1: b ab c ac d ad bcd abcd
   Block2: (1) a bc abc bd abd cd acd
7. Explain fractional factorial experiments
8. Explain the concept of connectedness in incomplete block design.
9. Explain the situations where split plot experiments can be adopted.
10. Under usual notations , show that $\lambda$ (v-1) = r(k-1) for a BIBD

(8 X1=8)

**SECTION B**

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

11.. Develop the procedure to test the general linear hypothesis based on a linear model , stating clearly the assumptions.

12. State and prove Gauss Markov theorem.

13. . Obtain the relative efficiency of RBD in comparison to CRD

14.. Explain the technique of estimation of two missing observations in a LSD. Also explain its analysis.

15. Explain the concept of linear and quadratic effect in $3^2$ factorial experiment. Explain its analysis and write down the ANOVA table of the same with linear, quadratic effects and their interactions.

16. If $e'\beta$ and $m'\beta$ are estimable , find $V(e'\beta)$ and $Cov(e'\hat{\beta} , m'\hat{\beta})$ , where $\hat{\beta}$ is the least square estimate of β. Also find the unbiased estimate for $\sigma^2$ based on Gauss Markov model $(Y, X\beta, \sigma^2 I)$

17. Define BIBD. Prove the Fisher's inequality. Also Construct a BIBD with the following parameters. b=v=4, r=k=3 and λ=2.

18. Explain Yates procedure for obtaining various effect totals in a $2^3$ factorial experiment

(6X2 =12)

**SECTION C**

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

19. Construct a $2^5$ design in blocks of 8 plots by confounding ABC, ADE and BCDE Give the analysis of such a design with r replications.

20. Outline the ANACOVA for RBD with one concomitant variable, stating clearly the assumptions.

21. Develop the intra block analysis of a BIBD assuming a suitable model.

22. Develop the analysis of split plot design with RBD layout for main plot treatments..

(5 X 2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied)DEGREE (C.S.S) EXAMINATION , ……**

**Third Semester**

**Faculty of Science**

## ST 50 03 03 :MULTIVARIATE ANALYSIS

**( 2019 admissions onwards)**

**Time : 3 hours** **Maximum Weightage 30**

**SECTIO N A**

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

1. How $T^2$can be regarded as a Generalization of Student's t Statistic.

2. Establish the invariance property of Hotelling's $T^2$.

3. What do you mean by principal component analysis?

4. Define canonical correlation.

5. How is the Fisher's Linear Discriminant function related to Mahalnobis$D^2$.

6.What is the divisive method of clustering

7. Explain the Wilk's $\lambda$statistic useful in the likelihood ratio tests.

8. Discuss the role of similarity measures in cluster analysis

9. Deiscribe a test for testing (statement only) the hypothesis of equality of two covariance matrices.

11. Explain profile analysis

(8 X1=8)

## SECTION B

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

11. . Explain the Fisher-Behren problem.

12. Obtain a 100(1-∝)% simultaneous confidence interval for all linear combinations of the mean vector μ of a multivariate normal distribution.

13. Describe the orthogonal Factor model in Factor Analysis.

14. Describe the test procedures in Profile analysis for testing: (a) parallelism, (b) level and (c) coincidence.

15. Describe the Baye's classification procedure.

16. Distinguish between Single linkage and complete linkage clustering methods.

17 Explain the sphericity test in multivariate analysis

18. Explain Fisher's linear discriminent function and derive it.

(6X 2=12)

## SECTION C

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

19.. Obtain the null distribution of one sample Hotelling's $T^2$ statistic.

20.. Describe an iterative procedure for obtaining the principal components.

21..Explain the procedure of classification into one of the several multivariate normal population.

22. .Stating the assumptions to be satisfied, explain the Two-way MANOVA.

(5 X2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied)DEGREE (C.S.S) EXAMINATION , ……**

**Third Semester**

**Faculty of Science**

**ST 50 03 04 TIME SERIES ANALYSIS**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

## SECTION A

(Answer any 8 questions. Each question carries a weight of 1.)

1. Explain a time series as a stochastic process.
2. Distinguish between additive and multiplicative models of time series.
3. Define auto-covariance and auto-correlation functions.
4. Distinguish between a strict stationary and weak stationary process.
5. Explain Wold Decomposition of a linear stationary process.
6. Describe (i) AR(p) model and (ii) MA(q) model .
7. What is an ARIMA model?
8. Define spectral density function and state its 3 properties.
9. What is a periodogram? What are its uses?
10. Define an ARCH model stating all assumptions. (8 x 1=8)

## SECTION B

(Answer any 6 questions. Each question carries a weight of 2).

11. Describe the components of a time series. Explain how you will estimate trend.
12. Describe simple exponential smoothing.
13. Define partial autocorrelation function. What are its properties and uses?
14. Show that an AR(1) model can be expressed as an infinite order MA model. Hence obtain its ACF.
15. What are Yule-Walker equations? Explain how they are used for estimating partial autocorrelations of an AR(k) model.
16. Explain the least square method for estimating the parameters of an ARMA(p,q) model. Illustrate it for an ARMA(1,1) model.
17. Derive the spectral density function of an (i) AR(2) process (ii)MA(2) process.
18. Explain a GARCH model and describe its importance in time series modeling.

(6x 2=12)

## SECTION C

( Answer any 2 questions. Each question carries a weight of 5.)

19. Explain Holt-Winter's exponential smoothing and forecasting. What are its advantages and disadvantages?
20. Explain the important steps in Box-Jenkin's approach to time series modeling.
21. Describe residual analysis and diagnostic checking of an ARIMA model. Explain how you will choose the AR and MA periods.
22. Derive the spectral density of an ARMA(p,q) process. Also describe the basics of seasonal ARIMA modeling. (5 X2 =10)

**QP Code** **Reg.No.**

**…….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION**

**Third Semester**

**Faculty of Science**

**ST 01 03 01 STATISTICAL COMPUTING III –USING R/SPSS/MATLAB**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

(Answer any **three** questions without omitting any part **. Each question carries a Weight 10** )

**Section A**

1.(a) A sample of size 38 from a trivariate Normal population gave the following results.

$$\bar{X} = \begin{bmatrix} 31.467 \\ 39.464 \\ 91.467 \end{bmatrix}; \qquad S = \begin{pmatrix} 17.81 & 19.42 & 13.43 \\ & 11.57 & 11.32 \\ & & 9.82 \end{pmatrix}$$

Test the hypothesis that the mean vector is $\mu_0 = [\ 28 \quad 35 \quad 87]'$ at 5% level.

(b)Test whether the following are contents of 2 samples drawn from the same multivariate normal populations assuming that the population dispersion matrices are the same.

$$N_1 = 12 , \bar{X}^{(1)\prime} = [\ 60.9 \ \ 4.35 \ \ 70.39 \ \ 76.42\ ]$$

$$S_1 = \begin{pmatrix} 17.55 & -1.1 & 4.55 & 1.61 \\ & 2.25 & -3.3 & -0.55 \\ & & 12.52 & -0.66 \\ & & & 6 \end{pmatrix}$$

$$N_2 = 15 , \bar{X}^{(2)\prime} = [\ 60.52 \ \ 4.3 \ \ 69.86 \ \ 73.96\ ]$$

$$S_2 = \begin{pmatrix} 8.24 & -1.21 & -1.29 & -1.5 \\ & 1.49 & 0.31 & 0.25 \\ & & 8.56 & 1.6 \\ & & & 8.8 \end{pmatrix}$$

2. The following data describes measures on several variables corresponding to 10 plants.

(i) Construct the Principal Components .

(ii) Use Factor Analysis to identify the important factors .

(iii) Group the plants by cluster analysis.

| Sl.No. | Leaf length | Stem length | Leaf breadth | Stem breadth | Root length |
|---|---|---|---|---|---|
| Plant1 | 12 | 32 | 6 | 11 | 75 |
| Plant2 | 40 | 34 | 7 | 13 | 73 |
| Plant3 | 30 | 35.5 | 5 | 12 | 76 |
| Plant4 | 50 | 42 | 6 | 14.5 | 77 |
| Plant5 | 12.8 | 39 | 8 | 14 | 78 |
| Plant6 | 13.2 | 38.5 | 7 | 16 | 74 |
| Plant7 | 14.5 | 37 | 9 | 16.5 | 73 |
| Plant8 | 16.1 | 31 | 5 | 13.8 | 75 |
| Plant9 | 18.5 | 33 | 8 | 12 | 72 |
| Plant10 | 15.9 | 34 | 7 | 15 | 78 |

## Section B

3. In the table given below are the yields of 6 varieties of 4 replicates of an experiment in which two values are missing. Estimate the missing values and analyse the data.

| Blocks | Treatments | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 18.4 | 15.4 | 9.6 | 13.6 | 17.1 | 14.6 |
| 2 | 11.7 | --- | 12.8 | 15.7 | 16.4 | 11.9 |
| 3 | 13.4 | 16.5 | 17.3 | 18.4 | --- | 22.6 |
| 4 | 16.5 | 13.6 | 12.6 | 15.7 | 16.3 | 17.9 |

3. The following data gives the yields of potatoes in a manurial experiment with 3 manures each at 2 levels along with the layout of the design.

| | Replicate 1 | | | |
|---|---|---|---|---|
| Block 1 | npk<br>33.9 | n<br>27.3 | p<br>33.4 | k<br>30.9 |
| Block2 | np<br>26.5 | pk<br>26.6 | nk<br>33.5 | (1)<br>33.9 |

| | Replicate 2 | | | |
|---|---|---|---|---|
| Block 1 | npk<br>27.9 | np<br>22.4 | k<br>32.5 | ( 1)<br>32.3 |
| Block2 | pk<br>34.8 | nk<br>27.1 | n<br>34.1 | p<br>30.7 |

## Section C

5. a. Let X1, X2,…,Xn be a random sample from N(μ,1). Let the observations be 115,121,136.8,128.7,133.5, 144.6,164.7,171.5. Using the best available test procedure, test H0: μ=142 against H1: μ=165, at 5 % level of significance. Also draw the power curve.

b. Derive SPRT of strength (0.1,0.1) for testing H0: p=0.5 against H1: p=0.7 when n=9 of B(n,p). Draw decision boundaries on a graph paper. Also draw the OC and ASN curves associated with the test.

c. Test the equality of the probability laws that have generated the two samples given below using K S test.

Sample I : 15,13,11,10,8,14,13,8,17,9,16.

Sample II: 13,10,6,13,9,11,6,8.

6. For fitting a production function Y= $AK^{\alpha}L^{\beta}M^{\gamma}$ Y = production, L= labour, K= capital and M=other factors of production, data summary are given below

N=10. If y, l, k are the logarithms of Y, K and M respectively, then $\sum$y=30.27, $\sum L$=31.72, $\sum K$=29.78, $\sum M$=21.66 and the matrix of corrected sum of squares and cross product is:

| y | l | k | m |
|---|---|---|---|
| 92.95 | 95.93 | 90.39 | 65.69 |
| | 120.68 | 100.74 | 67.94 |
| | | 100.96 | 57.1 |
| | | | 57.39 |

(a) Evaluate the elasticities of production

(b) Test whether the inclusion of the variable m is significant

(c) Obtain the value of $R^2$

(d) Test the hypothesis of constant return to scale

Test whether the partial elasticities of production with respect to the three inputs are the same

## SEMESTER 4

**QP Code** **Reg.No. …….**

**Name**

**M.Sc STATISTICS/ STATISTICS (Applied)DEGREE (C.S.S) EXAMINATION , ……**

**Fourth Semester**

**Faculty of Science**

**ST 50 04 01 ECONOMETRIC METHODS**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

**SECTION A**

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

1. Explain equilibrium analysis of market model.
2. Explain constant product curves.
3. Define price elasticity of demand. Calculate price elasticity of demand for the demand law $x = ap^{-\alpha}$
4. Define coefficient of determination and adjusted $R^2$.
5. What is Variance Inflation Factor (VIF)? What is the significance of high value of VIF?
6. Discuss instrumental variable technique in regression analysis.
7. Explain the least variance ratio method.
8. Describe autoregressive and disturbed lag models
9. Discuss problems encountered in estimation of a linear probability model
10. Discuss Von-Neumann ratio test for autocorrelation.

(8X1=8)

**SECTION B**

(Answer **any 6** questions. Each question carries a weight of 2)

11. Discuss on the learners approach to model selection

12. Define elasticity of substitution for the production function. Obtain same for the production function $f(a,b) = (Aa^{-\alpha} + Bb^{-\alpha})^{\frac{-1}{\alpha}}$.

13. For the linear regression model $Y = X\beta + \varepsilon$, where $\beta$ is an p X 1 vector, obtain maximum likelihood estimator of $\beta$ if $\varepsilon$ is normally distributed.

14. Explain Aitken Generalised least square method of estimation.

15. Explain Koyck distributed lag model and estimation procedure.

16. Define auto correlation. Explain the Durbin-Watson Test.

17. Discuss multicollinearity. What are the consequences in OLS estimation?

18. Describe the two stage least squares method of estimation. Obtain the asymptotic properties of the estimates so obtained.

(6 X2=12)

**SECTION C**

(Answer **any 2** questions. Each question carries a weight of 5)

19. Discuss in detail stating all assumptions, Leontief Input – Output model for open system and Walras-Leontief closed system.

20. Obtain a necessary and sufficient condition for identification.

21. Explain heteroscedasticity. What are the consequences? Discuss how to detect heteroscedasticity.

22. For the linear regression model $Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I)$, obtain the distribution of residual sum of squares. Find a $100(1-\alpha)\%$ confidence interval for $\sigma^2$.

(5X2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION**

**Fourth Semester**

**Faculty of Science**

**ST 01 04 01- STATISTICAL COMPUTING IV –USING R/SAS/MATLAB**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

(Answer any **three** questions without omitting any part**.** Each question carries a Weight of 10 )

## SECTION A

1. Consider the seasonal time series data shown in the following table.

| | Q1 | Q2 | Q3 | Q4 |
|---|---|---|---|---|
| 2011 | 318 | 380 | 358 | 423 |
| 2012 | 379 | 394 | 412 | 439 |
| 2013 | 413 | 458 | 492 | 493 |
| 2014 | 461 | 468 | 529 | 575 |
| 2015 | 441 | 548 | 561 | 620 |

a. Calculate the seasonal indices by the "Ratio to Moving Average" method.

b. Use Holt-Winters' method to develop a forecasting method for this data($\alpha = 0.1$, $\beta = 0.2$, $\gamma = 0.1$).

2.Consider the following autocorrelation and partial autocorrelation coefficients using 500 observations for a weakly stationary series

| Lag | ACF | PACF |
|---|---|---|
| 1 | 0.307 | 0.307 |
| 2 | -0.013 | 0.264 |
| 3 | 0.086 | 0.147 |
| 4 | 0.031 | 0.086 |
| 5 | -0.079 | 0.049 |

a. Determine which, if any, of the ACF and PACF coefficients are significant at the 5% level.

b. Use both the Box-Pierce and Ljung-Box statistics to test the joint null hypothesis that the autocorrelation coefficients are jointly zero.

c. What process would you tentatively suggest could represent the most appropriate model for this series?

d. How could you estimate the model you suggest in part (c)?

## SECTION B

3. Control chart for $\bar{X}$ and R are maintained on the tensile strength of a metal fastener . After 30 samples of size n=6 are analysed , we find that

$\sum_1^{30} \overline{Xi}$ = =12870 and $\sum_1^{30} Ri = 1350$

(a)Compute control limits on the R chart

(b) Assuming that the R chart exhibits control , estimate the parameters μ and $\sigma$

(c) If the process output is normally distributed , and if the specifications are 440 ± 40, can the process meet this specifications? Estimate the fraction non – conforming

(d) If the variance remains constant , where should the mean be located to minimize the fraction non- conforming?

4. Suppose that a single sampling plan with n=150 and c=2 is being used for receiving inspection where the vendor ships the products in lots of size N=3000.

(a) Draw the OC curve for this plan.

(b) Draw the AOQ curve and find the AOQ.

(c)Draw the ATI curve for this plan

## SECTION C

5. Use Wolfe's method to solve the following quadratic programming problem:

Maximize $Z = 2x_1 + 3x_2 - 2x_1^2$

subject to $x_1 + 4x_2 \leq 4$

$x_1 + x_2 \leq 2$

$x_1, x_2 \geq 0$

6. A consumable item has a demand of 12,000 units per year. The cost of one procurement is Rs. 200 and the holding cost is Rs. 4 per unit per year. The replacement is immediate on procurement and hence there are no stocks out. Determine the following

(i) The Economics lot size
(ii) The number of orders per year.
(iii) The time between orders.
**(iv)** The total cost per year, including the purchase cost if the cost of 1 unit is Rs. 2.

# ELECTIVE PAPERS

## BUNCH A (ST 80 04 01-03)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION,………**

**Fourth Semester**

**Faculty of Science**

## ST80 04 01 -Operations Research

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

### *SECTION A*

***(Answer any 8 questions. Each question carries a weight of 1)***

1. Define slack, surplus and unrestricted variables
2. Show that the set of feasible solutions to an LPP is a convex set.
3. What is traveling salesman problem? How is it related to assignment problem?
4. Explain the various costs involved in inventory problems.
5. Explain the following terms in inventory(a) Lead time (b) Re - order point (c) stockout cost (d) Buffer stock.
6. State the general non linear programming problem
7. Discuss Bellman's principle of optimality.
8. Explain the characteristics of dynamic programming.
9. Explain the concept of dominance in game problems.
10. Define the term ' Strategy' and 'Optimal Strategy'with reference to game theory

(1 X 8 = 8)

### *SECTION B*

***(Answer any 6 questions. Each question carries weight 2.)***

11. What is sequencing problem. Give an algorithm for processing n jobs through 2 machines.
12. Explain graphical method of solving LPP.

13. Discuss EOQ problems with price breaks. Analyze the inventory model with two price breaks.
14. Analyze a single period probabilistic inventory model with instantaneous and continuous demand and no set - up cost.
15. Use dynamic programming to show that $\sum_{i=1}^{n} p_i \log p_i$ subject to the constraint $\sum_{i=1}^{n} p_i = 1$ and $p_i \geq 0$ is minimum when $p_i = \frac{1}{n}, \; i = 1, 2, ..., n$ .
16. Explain Beal's method for solving a quadratic programming problem.
17. Explain the graphical method of solving m×2 games with the help of an example..
18. Describe the concept of maximin- minimax principle in game theory

(6 X 2 = 12)

***SECTION C***

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

19. Explain transportation problem.Give its mathematical model. Explain MODI method of solving a transportation problem.

20. Derive EOQ formula for an inventory model with a finite rate of production, constant rate of demand incorporating set-up cost, holding cost and shortage cost.
21. Explain quadratic programming problem. Derive Kuhn- Tucker conditions. Describe Wolfe's method for solving the same.

22. Describe the steps involved in transforming an m×n game to an LPP. Hence or otherwise state and prove fundamental theorem of game theory.

(2 X 5 = 10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION,………**

**Fourth Semester**

**Faculty of Science**

**ST 80 04 02 -STATISTICAL QUALITY CONTROL**

**( 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 different causes of variation in the quality of a manufactured product?
2. What are rational sub groups?
3. Explain warning limits. Justify the 3σ limits as control limits in any control chart.
4. Define operating characteristic function of a control chart.
5. What is a *U* – chart. When and where it is used.
6. What is the role of C charts in statistical process control?
7. Explain *CUSUM* charts.
8. Define the terms: *AQL, LTPD*, Producer's risk and Consumer's risk.
9. Distinguish between multiple sampling plans and sequential sampling plans.
10. Write a short note on MIL STD 414 standard in a lot by lot acceptance sampling plan

**(8x1=8)**

**SECTION B**

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

11. (a) Describe the construction of P chart. (b) What is moving average control charts and set up the control limits.

12. (a) What is sampling inspection? Explain the technique of curtailed inspection.

(b) Explain the terms: control limits, tolerance limits and specification limits.

13. Control chart of $\bar{x}$ and $R$ in use with the following measures.

$\bar{x}$ chart : $CL = 420$, $LCL = 390$, $UCL = 450$.

$R$ chart : $CL = 67.05$, $LCL = 54.15$, $UCL = 61.16$, $d_2 = 2.326$

The sample size is 5. Both charts exhibit control. The quality characteristic is normally distributed. (a) What is the α – risk associated with the $\bar{x}$ chart (b) specification of the quality characteristic is 415 ± 20. What are your conclusions regarding the ability of the process to produce within specifications?

14. (a) Distinguish between defects and defectives. Explain the construction and operation of a*p* chart.

15.. (a)What are acceptance sampling plans? Explain the *SSP.*(b) Describe a procedure to derive a SSP using attributes with a specified α and β.

16 (a) What is meant by rectifying inspection? Obtain the *AOQ* functionof a *SSP*.

(b) Explain the method of construction of the *O. C.* curves foran attribute *DSP*.

17. Describe an item by item sequential sampling plan by attributes. Derive the acceptance and rejection lines of such a plan with a given producers risk and consumers risk.

18. . Explain the construction of C chart. Give the situations in which C chart can be used.

(6 x 2=12)

**SECTION C**

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

19. (a) Explain the construction and interpretation of mean chart and range chart.

(b) Describe various ways in which a control chart may be modified t meet special situations.

20. How will you study the process capability of a production process? What are the important indices for measuring the process capability?

21. Derive the *ASN* and *ATI* functions for a *DSP* and draw their general shapes.

22.The measurement $X$ on an item follows a normal distribution with known standard deviation. The item is considered acceptable if $X$is large. Derive a *SSP* for a specified α and β.

(5 x2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION,………**

**Fourth Semester**

**Faculty of Science**

**ST80 04 03 - ADVANCED BAYESIAN COMPUTING WITH R**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

## SECTION A

Answer any **EIGHT** questions. Each question caries a weight of 1)

1. What do you mean by a conjugate prior? Give an example.
2. Derive Jeffrey's prior for Θ in Bernoulli (Θ).
3. What do you mean by improper prior? How can you normalize it?
4. If X has a Poisson distribution with parameter , find a non-informative prior for $\lambda$.
5. Distinguish between prior predictive distribution and posterior predictive distribution.
6. Derive posterior distribution for binomial proportion using conjugate prior.
7. Define predictive distribution.
8. What do you mean by burn-in-period?. How do you fix it?
9. What do you mean by a posterior-predictive model?
10. What do you mean by a shrinkage estimator?.

(8 x1=8)

## SECTION B

(Answer any **SIX** questions. Each question caries a weight of 2.)

11. Explain how do you use LearnBayes package to obtain the posterior summary in the case of binomial proportion using a discrete prior.
12. Derive the Bayes estimate of mean of a normal distribution with known variance assuming (a) a non-informative prior (b) a conjugate prior.
13. Explain how do you construct conjugate prior in the case of exponential family of distributions.
14. Derive Bayes estimate of the Poisson parameter using Gamma ( 10, 15) as the prior distribution.
15. Explain Monte-Carlo method of estimation of the integral of a function, when the integral exists finitely.
16. Explain Gibb's sampling method to simulate a random sample from a bi-variate population with density function $f(x, y) = k\,(x^2+y^3)$, $0 < x < 1$, $0 < y < 2$.
17. What do you mean by Bayes factor?. Explain Bayesian one sided test for normal mean.
18. Explain Normal Linear Regression model and prediction of future observations.

(6x2=12)

## SECTION C

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

19. (a) State and prove Bayes theorem for inference and explain on the basis of a suitable loss function how do you estimate and test a statistical hypothesis using this theorem with a suitable example.

(b) Explainany four important methods for selection of prior distribution.

20. Derive the Bayes estimates of $\mu$ and $\sigma^2$ in $N(\mu, \sigma^2)$ based on sample of size n and assuming a normal prior for $\mu$ and an inverse Gamma prior for $\sigma^2$.
21. (a) Explain Metropolis-Hastings algorithm and show that the algorithm leads to the stationary distribution as the posterior density.
(b) Write R codes to generate a random sample from the standard Cauchy distribution using standard normal distribution as the candidate density in the Metropolis algorithm.
22. (a) Explain important MCMC output analysis.

(b) Explain Bayesian linear regression model and prediction of future observations.

( 5x 2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION,………**

**Fourth Semester**

**Faculty of Science**

**ST 81 04 01 SURVIVAL 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 censoring and truncation.
2. Establish the relationship between failure and mean residual life function.
3. Define discrete time hazard function and give the expression of survival function in terms of hazard function in discrete case.
4. Distinguish between probability plots and hazard plots.
5. What do you mean by deviance residuals.
6. Write a short note on Nelson- Aalen estimate.
7. Explain why Cox regression model is called proportional hazard model.

8. What do you mean by baseline hazard function?
9. Write a short note on Cox-Snell residual method for assessing the fit.
10. Discuss the methods of coding covariates in regression analysis.

(8 x1=8)

## Part B

Answer any **six** questions. **Weight 2** for each question

11. Derive the Greenwoods formula for the variance of the estimates of the survival function.
12. Find out the observed likelihood functions for the type I and Type II censoring mechanism.
13. Explain any one method for estimating confidence intervals for quantiles.
14. What do you mean by double censoring? Explain how to estimate survival function for double censored data.
15. Derive the two sample tests for comparing hazard rates and survival functions.
16. Discuss the estimation of baseline hazard function and base line survivor function.
17. Describe various types of residuals in proportional hazard models.
18. Explain the concept of competing risk models.

(6 x2 =12)

## PART C

Answer any **TWO** questions. **Weight 5** for each question.

19. Show that for the Type I and Type II censoring mechanism ,the observed likelihood function takes the form $\prod_{i=1}^{n} f(t_i)^{\delta_i} \ S(t_i+)^{1-\delta_i}$, where $\delta_i$ is the indicator function.
20. Derive the Kaplan- Meier estimate for the survival function and discuss its properties.
21. Obtain the partial maximum likelihood estimator for the proportional hazards model on distinct event time data.
22. Explain how regression models can be used for comparing or testing the equality of distributions.

(5x2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION,………**

**Fourth Semester**

**Faculty of Science**

**ST 81 04 02 POPULATION DYNAMICS**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

**Time: 3 hours** **Max. Weight 30**

**SECTION A**

(Answer any **8** questions. Each question carries a weight of **1**.)

1. Define (i) infant mortality rate, (ii) neo-natal mortality rate
2. Establish the relationship between the force of mortality and the central mortality rate.
3. What is the need for the gradation of mortality rates?
4. Explain the Makeham's model for mortality gradation.
5. What are life tables? What are its uses?
6. Distinguish between a complete life table and an abridged life table.
7. Establish the Greville's abridged life table formula.
8. Stationary population is a special case of the stable population. Discuss.
9. Distinguish between GRR and NRR. What does it imply, if NRR = 1?
10. What are the important population growth models?

(8x1=8)

**SECTION B**

(Answer any **6** questions. Each question carries a weight of **2.**)

11. What is the purpose of standardization of a mortality data? Explain the direct and indirect methods of standardization.

12. What do you mean by Specific Death Rates?. Mention the important among them and their advantages over the Crude Death Rate.

13. Explain the Gompertz model of mortality.

14. Establish one method of fitting the logistic law for explaining population growth.

15. Describe the William Brass model for human fertility.

16. Discuss the important indices of Fertility measures.

17. What is the effect of mortality and fertility changes on the age distribution of a stable population?

18. Discuss the Leslie Matrix technique for projecting the population briefly.

(6x2=12)

## SECTION C

(Answer any **2**questions. Each question carries a weight of **5**)

19. Establish the Reed and Merrell's formula for the construction of life tables.

20. Derive the sampling distribution of the life table functions.

21. Explain the Shep and Perrin model of human reproductive process. What is the average waiting time between two successive live births?

22. Derive Alfred Lotka's fundamental equation of stable population. Show that age structure and birth rate of stable population are independent of time.

(5 x 2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION,………**

**Fourth Semester**

**Faculty of Science**

**ST 81 04 03 CATEGORICAL DATA ANALYSIS**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

## Section A

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

1. What you mean by a Categorical data analysis?.
2. Define relative risk
3. Define Odds Ratio

4. What you mean by sensitivity and specificity?.
5. What is the relation between Odds and P(Y=1) in a logistic regression model.
6. Give the formula for the computation of the 95% confidence interval for the population odds ratios
7. What are theAdvantages of GLMs over traditional (OLS) regression?
8. What is the problem of over dispersion in Poisson Regression model.
9. Define hazard rate.
10. What you mean by prior and posterior distributions?

**(1 X 8 =8)**

**SECTION B**

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

11. What are the four data types based on the measurement?.

12. Explain the latent variable approach in categorical data analysis.

13. Explain three components of a generalized linear model.

14. Imagine that the incidence of gun violence is compared in two cities,one with relaxed gun laws (A), the other with strict gun laws (B). In the city with relaxed gun laws, there were 50 shootings in a population of 100,000 and in the other city, 10 shootings in a population of 100,000.

    1) What is the relative risk of gun violence in the city with relaxed gun laws (A)? (2)Compute the 90% confidence interval for RR

15. What you mean by deviance measure?

16. Define logodds and derive the standard error of logodds

17. Explain the Hosmer-Lemeshow test statistics.

18. Explain how negative Binomial regression model is useful when over dispersion exists.

(2 X 6 = 12)

**SECTION C**

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

19. Explain logistic regression model with assumptions if any. Derive the likelihood function of the logistic regression model. Explain how you will proceed to estimate the parameters from the likelihood equation.

20. Explain proportional Hazard regression model. Derive the likelihood function

21. Explain different methods of checking model adequacy in generalized linear models ?

22. Explain Gibbs sampler. ( 2 x 5 = 10)

**BUNCH C (ST82 04 01-03)**

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION,………**

**Fourth Semester**

**Faculty of Science**

**ST 82 04 01 ACTUARIAL STATISTICS**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

## SECTION A

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

1. Explain surrender value and paid up policyitcep
2. For a whole life policy with unit sum assured, show that prospective and retrospective reserves are equal.
3. What is a life table?
4. Explain the terms (a) time-until-death, (b) age-at-death
5. Define (i) Survival function and ( ii) curtate-future- lifetime.
6. Develop the models for the life insurances with death benefits payable at the moment of death.
7. Define " force of mortality" . Show that it can be used to specify the distribution of lifetime.
8. Define life annuities.
9. Show that under the assumptions of uniform distribution of death ( in the usual notation)

   $\mathrm{Var}(T) = \mathrm{Var}\,(K) + \frac{1}{12}$

10. Explain the control rates in a multiple decrement contest

(8 x 1 =8)

## SECTION B

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

11. If the random variable T has pdf , $f_t\,(T) = \lambda\,(\exp -\lambda t)$ , $t \geq 0$ , $\lambda > 0$ . Calculate ( a) $e^0_x$ (b) V(T) (c) Median (T) (d) Mode (T)
12. Explain benefit premiums. Derive the expression for level annual benefit premium. Also give a short account of accumulative type benefits.
13. Explain Future Lifetime Random Variables. Comment on the situations when they are used.
14. Describe 'annuities due ' and 'annuities immediate' and obtain the present value random

variables for these two annuities.

15. Derive premium difference formula for n- year term insurance in the continuous case .

16. Derive the values of the following (a) ${}_{n}P_{xy}$ (b) ${}_{n}Q_{xy}$ (c) $\mu_{xy}$

17. What is benefit reserves ? Derive the formulas for fully continuous reserves. Also explain the reserves for general insurance.

18. How do you model a single life, subject to multiple contingencies? Explain the construction of a multiple decrement table .

(6 x2 =12)

**Part C**

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

19. (a) Differentiate between complete and abridged life table

(b) A life subject to a force of mortality of 0.02178. Calculate the probability that

(i) he will live for 10 years

(ii) he will die between 15 and 20 years

20. . Find the annual premium for an assurance granted to a life aged 20 where the sum assured payable at the end of the year of death is Rs.20,000 in the first year increasing by Rs. 2000 each so that the sum assured on death during $20^{th}$ year is Rs. 58,000. The sum assured payable on maturity at the end of 20 years is Rs. 60,000. Basis : A 1967-70 ult . and 4% pa interest

21. Derive the expected present value of ill health retirement pension of 1/k of final salary near retiral for a member of the pension scheme now aged X and put in n years of past service . Define the symbols clearly.

22. (a) Discuss the benefits defined in terms of the time of the last death in the multiple life case.

(b) Determine the survival function and pdf of $T(\overline{xy}) = \max[T(x), T(y)]$ for two lives (x) and (y) with the joint pdf of their future lifetimes ,

$f_{T_{(x)}T_{(y)}}{}^{(s,t)} = 0.0006(t-s)^2$ , $0<s<10$ , $0< t< 10$

$= 0$ elsewhere

(5 x2=10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION,………**

**Fourth Semester**

**Faculty of Science**

**ST82 04 02 - APPLIED 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. Stating the assumptions , describe the multiple linear regression models.

2. What is meant by estimability of a parametric function? Illustrate with an example.

3. Explain how non-constant variance in linear models are dealt with

4. What is " Collinearity" ? Explain its consequences.

5. What is polynomial regression model? Explain a method for its estimation.

6. Explain logit model and explain its advantages.

7. What are kernal smoothers in non parametric regression?

8. What are orthogonal polynomials?

9. State Gauss- Markov theorem.

10. Explain Mallows cp statistic (1 X 8=8)

**SECTION B**

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

11. In the linear model Y= Xβ + e ( in usual notation) , obtain the least square estimate of β and discuss the properties of the estimate. Also obtain unbiased estimate for $\sigma 2$ in the Gauss Markov setup (Y, Xβ , $\sigma 2$ I )

12. Explain how non-normal errors are detected. Discuss the Box-Cox family of transformations on response to deal with non-normality

13. Explain serial correlation. Explain the α test proposed by Durbin and Watson for testing serial correlation

14. How does the assumption of orthogonality simplify the problem of least squares in polynomial regression? Describe the principle involved in choosing the degree of an orthogonal polynomial based on date.

15. Distinguish between linear and non-linear regression models. Explain the assumptions and methodologies involved in them.

16. Explain the ridge estimator. Compare this estimator with the OLS estimator under mean square error criterion.

17. Explain Poisson Regression. Explain a method of estimation in this setup.

18. Explain an analytical method for selecting a transformation to correct model inadequacies. (6 X 2= 12)

**Part C**

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

19. (a) Explain the problem of regression for binary response variable and develop the method of maximum likelihood to estimate the parameters in a logistic regression model.

(b) Explain the residual analysis in the generalized linear model.

20. (a) Let $Y1= \mu_1 + \epsilon_1$ , $Y_2 = 2\mu_1 - \mu_2 + \epsilon_2$, $Y_3 = \mu_1 + 2\mu_2 + \epsilon_3$ , where $\epsilon \sim N(0, \sigma^2 I_3)$.

Derive an F- statistic for testing $H_0 : \mu_1 = \mu_2$

(b) Discuss briefly the state of affairs and consequences on account of possible departures from the underlying assumptions on a linear model.

21. (a) What is the need for piecewise polynomial fitting? Discuss the method of splines in this context.

(b) Distinguish between bias due to under-fitting and bias due to over-fitting in a multiple regression model , giving an illustrative example

22. (a) Bring out the differences between least median of squares regression and least absolute deviation regression

(b) Explain bootstrapping procedure for regression models emphasizing straight line fit (5 X 2= 10)

**QP Code** **Reg.No. …….**

**Name**

**M.Sc (STATISTICS) DEGREE (C.S.S) EXAMINATION,………**

**Fourth Semester**

**Faculty of Science**

**ST 82 04 03 -DATA MINING**

**( 2019 admissions onwards)**

**Time : Three hours** **Max.Weight :30**

**Time : 3 Hours** **Maximum weight : 30**

**SECTION A**

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

1. Explain how data mining can be categorized.
2. Explain the need for data mining.
3. Explain any two tasks carried out using data mining.
4. Explain a decision tree.

5. Give examples for clustering methods

6. Explain the components of data mining

7. Explain different types of association rules

8. Define data set and give example

9. Does a data mining access of data differ from traditional access?

10. Explain about data mining trends (8 x 1=8)

**SECTION B**

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

11. Explain about clusters in detail.

12. Explain the goals of data mining

13. What are the methodology and user interactions in data mining?

14. Explain about the classification process

15. What are the different data mining software tools used?

16. Explain about clusters in detail

17. Explain the approaches of data mining problems

18. Explain the various types of association rules

(6 x 2 = 12)

**SECTION C**

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

19. Explain about the data mining applications

20. Explain the steps involved in data mining process

21. Explain in detail about the artificial neural network

22. Explain about the characteristics and benefits of data mining

(2 X 5 = 10)

*XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX*

**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