File Ref.No.25539/GA - IV - J2/2013/CU

UNIVERSITY OF CALICUT

Abstract

General and Academic IV - Faculty of Science - Integrated M.Sc Statistics Programme under Regulations for Integrated Programmes - Scheme and Syllbaus - With effect from 2020 Admission - Implemented subject to ratification by the Academic Council - Orders Issued.

G & A - IV - J

U.O.No. 15790/2022/Admn Dated, Calicut University.P.O, 19.08.2022

Read:-1. U.O.No. 7220/2021/Admn dated 23.07.2021

2. The scheme and syllabus of Integrated MSc Statistics Programme, forwarded by the Chairperson, Board of Studies in Statistics (PG) dated 01.08.2022

3. Remarks of Dean, Faculty of Science dated 05.08.2022

ORDER

1. The scheme and syllabus of first and second semester for Integrated M.Sc. Statistics Programme, in tune with the Regulations for Integrated Programmes (affiliated colleges), was implemented, with effect from 2020 Admission , vide paper read (1) above.
2. Vide paper read (2) above, the Chairperson, Board of Studies in Statistics (PG), forwarded the scheme and syllabus of whole semesters for Integrated MSc Statistics Programme, in tune with the Regulations for Integrated Programmes (affiliated colleges), with effect from 2020 Admission, after circulating among the members of the Board of Studies in Statistics (UG) and Board of Studies in Statistics (PG), as per Clause (34) of Chapter 3 of Calicut University First Statutes (CUFS) 1976.
3. The scheme and syllabus forwarded by the Chairperson, have approved by the Dean, Faculty of Science, vide paper read (3) above, and then by the Vice Chancellor, on 08.08.2022, subject to ratification by the Academic Council.
4. The scheme and syllabus for Integrated M.Sc Statistics Programme, in tune with the Regulations for Integrated Programmes (affiliated colleges), is therefore implemented, with effect from 2020 Admission, subject to ratification by the Academic Council.
5. Orders are issued accordingly. (syllabus appended) (U.O.No. 7220/2021/Admn Dated, 23.07.2021 stands modified to this extent)

Ajayakumar T.K

Assistant Registrar

To

The Principals of all Affiliated Colleges offering Integrated MSc Statistics Programme

Copy to: PS to VC/PA to PVC/PA to R/PA to CE/JCE I/JCE V/EX and EG Sections/GA I F/CHMK Library/Information Centres/SF/DF/FC

Forwarded / By Order

Section Officer

INTEGRATED M.Sc. STATISTICS

Programme Objectives:

The main objective of the Integrated M.Sc. Statistics programme is to facilitate higher secondary passed students to learn, practice and make career in the art of information analysis for the purpose of decision making on concerned problems. Analysis can be done by using well accepted principle and scientific methods developed in Statistics. As these students have chosen statistics at an early stage of their learning, they have an opportunity of better understanding fundamentals of statistics and equip themselves to work as a professional statistician. Training in statistical computing will enhance their job opportunities and professional skills.

Programme Specific outcome of Integrated M.Sc. Statistics Programme:

After the completion of Integrated M.Sc. Statistics programme, students will:

1. Get a good employment in Government, Public and Private sectors.
2. Create awareness to become an enlightened citizen with commitment to deliver one's responsibilities within the scope of bestowed rights and privileges.
3. Perform data analyst and make interpretations with the skills attained during the course of the study.
4. Learn the art of representing and dealing with random phenomenon.
5. Learn principles and statistical concepts used in decision making.
6. Learn art of gathering information by sampling and designing experiments and analysing it.
7. Be able to assist researchers for drawing inferences using their experimental out comes .
8. Be able to construct statistical models of real world problems and obtain solutions.
9. Familiarize with Statistical software packages and can serve as a data analyst in the public or private sector.
10. Acquire basic theoretical and applied principles of statistics with adequate preparation to pursue Doctoral (Ph. D.) degree or enter job force as an applied statistician.
11. Be able to develop and validate models on the basis of collected data.
12. Continue to acquire relevant knowledge and skills appropriate to professional activities and demonstrate highest standards of ethical issues in Statistical Sciences.

COURSE OUT LINE FOR THE INTEGRATED

M.Sc. STATISTICS PROGRAMME (Core, Elective & Open)

Semester	Course Code	Course Title	Instructional Hours/Week	Credit	Duration of Exam
I	STA1IB01	Statistical Methods	3	2	2 Hours
II	STA2IB02	Basic Probability Theory	3	2	2 Hours
III	STA3IB03	Bivariate Data Analysis	3	3	2 Hours
	STA3IB04	Probability Distributions	4	3	
IV	STA4IB05	Statistical Inference -I	4	3	2 Hours
	STA4IB06	Sampling Techniques	3	3	
V	STA5IB07	Statistical Inference -II	5	4	2.5 Hours
	STA5IB08	Design and Analysis of Experiments	5	4	
	STA5IB09	Data Analytics with R	5	4	
	STA5IB10	Applied Statistics	5	4	
	STA5ID--	Open Course	3	3	2 Hours
		Project (Foundation)	2	-	
VI	STA6IB11	Statistical Quality Control	5	4	2.5 Hours
	STA6IB12	Basic Regression Analysis	5	4	
	STA6IB13	Operations Research	5	4	
	STA6IB14	Analytical Tools for Statistics	4	4	
	STA6IH01	Statistical Computing -I	4	4	
	STA6IF01	Project (Foundation)	2	3	-
VII	STA7IB15	Mathematical Methods for Statistics -I	5	4	3 Hours
	STA7IB16	Mathematical Methods for Statistics - II	5	4	
	STA7IB17	Probability Theory - I	5	4	
	STA7IB18	Distribution Theory	5	4	
	STA7IB19	Advanced Sampling Theory	5	4	
VIII	STA8IB20	Probability Theory - II	5	4	3 Hours
	STA8IB21	Estimation Theory	5	4	
	STA8IB22	Advanced Design and Analysis of Experiments	5	4	
	STA8IB23	Regression Methods	5	4	
	STA8IH02	Statistical Computing - II	5	4	
IX	STA9IB24	Testing of Statistical Hypothesis	5	4	3 Hours
	STA9IB25	Multivariate Analysis	5	4	
	STA9IB26	Stochastic Processes	5	4	
	STA9IE--	Elective - I	5	4	
	STA9IE--	Elective - II	5	4	
X	STA10IE--	Elective - III	5	4	3 Hours
	STA10IE--	Elective - IV	5	4	
	STA10IH03	Statistical Computing -III	5	4	
	STA10IF02	Project (Advanced)	10	4	
	STA10IG01	Comprehensive Viva	---	4	---

COURSES

SEMESTER -1

Course Code	Course Name	Hours/Week	Credit
ENG1IA01	Common English Course I	5	3
ENG1IA02	Common English Course II	4	3
MAL1IA07(3)/ HIN1IA07(3)/ ARB1IA07(3)/ SKT1IA07(3)	Additional Language Course I	5	4
STA1IB01	Statistical Methods	3	2
MAT1IC01	Allied Core -1: Mathematics – 1	4	3
BCS1IC01	Allied Core -2 : Computer Science (Data Science)- Python Programing	2+2	2
	Audit Course – Environment Studies	0	4
	TOTAL	**25**	**17+4**

SEMESTER –II

Course Code	Course Name	Hours/Week	Credit
ENG2IA03	Common English Course III	5	4
ENG2IA04	Common English Course IV	4	4
MAL2IA08(3)/ HIN2IA08(3)/ ARB2IA08(3)/ SKT2IA08(3)	Additional Language Course II	5	4
STA2IB02	Basic Probability Theory	3	2
MAT2IC02	Allied Core -1 : Mathematics -2	4	3
BCS2IC02	Allied Core -2: Computer Science (Data Science- Introduction to Data Science)	2+2	2
	Audit Course – Disaster Management	0	4
	TOTAL	**25**	**19+4**

SEMESTER –III

Course Code	Course Name	Hours/Week	Credit
IA11	General Course - I	4	4
IA12	General Course - II	4	4
STA3IB03	Bivariate Data Analysis	3	3
STA3IB04	Probability Distributions	4	3
MAT3IC03	Allied Core -1 : Mathematics -3	5	3
BCS3IC03	Allied Core -2: Computer Science (Data Science- Database Management System	3+2	2
	Audit Course – Human Rights/Intellectual Property Rights/Consumer Protection	0	4
	TOTAL	**25**	**19+4**

SEMESTER –IV

Course Code	Course Name	Hours/Week	Credit
IA13	General Course - III	4	4
IA14	General Course – IV	4	4
STA4IB05	Statistical Inference – I	4	3
STA4IB06	Sampling Techniques	3	3
MAT4IC04	Allied Core -1 : Mathematics -4	5	3
BCS4IC04	Allied Core -2: Computer Science (Data Science- Data Visualization)	3+2	2
BCS4IC05	Allied Core -2: Computer Science (Data Science- Programming Lab - Python Programmingand Database Management System Lab/ Data Visualization using Tableau)	-	4
	Audit Course – Gender Studies/Gerontology	0	4
	TOTAL	**25**	**23+4**

SEMESTER –V

Course Code	Course Name	Hours/Week	Credit
STA5IB07	Statistical Inference - II	5	4
STA5IB08	Design and Analysis of Experiments	5	4
STA5IB09	Data Analytics with R	5	4
STA5IB10	Applied Statistics	5	4
STA5ID--	Open Course	3	3
	Project Work	2	-
	TOTAL	**25**	**19**

SEMESTER –VI

Course Code	Course Name	Hours/Week	Credit
STA6IB11	Statistical Quality Control	5	4
STA6IB12	Basic Regression Analysis	5	4
STA6IB13	Operations Research	5	4
STA6IB14	Analytical Tools for Statistics	4	4
STA6IH01	Statistical Computing	4	4
STA6IF01	Project Work	2	3
	TOTAL	**25**	**23**

SEMESTER –VII

Course Code	Course Name	Hours/Week	Credit
STA7IB15	Mathematical Methods for Statistics -I	5	4
STA7IB16	Mathematical Methods for Statistics -II	5	4
STA7IB17	Probability Theory -I	5	4
STA7IB18	Distribution Theory	5	4
STA7IB19	Advanced Sampling Theory	5	4
STA7II01	Ability Enhancement Course	--	4
	TOTAL	**25**	**20+4**

SEMESTER –VIII

Course Code	Course Name	Hours/Week	Credit
STA8IB20	Probability Theory -II	5	4
STA8IB21	Estimation Theory	5	4
STA8IB22	Advanced Design and Analysis of Experiments	5	4
STA8IB23	Regression Methods	5	4
STA8IH02	Statistical Computing - II	5	4
STA8II02	Professional Competency Course	---	4
	TOTAL	**25**	**20+4**

SEMESTER –IX

Course Code	Course Name	Hours/Week	Credit
STA9IB24	Testing of Statistical Hypothesis	5	4
STA9IB25	Multivariate Analysis	5	4
STA9IB26	Stochastic Processes	5	4
STA9IE--	Elective - I	5	4
STA9IE--	Elective - II	5	4
	TOTAL	**25**	**20**

SEMESTER –X

Course Code	Course Name	Hours/Week	Credit
STA10IE--	Elective - III	5	4
STA10E--	Elective - IV	5	4
STA10IH03	Statistical Computing - III	5	4
STA10F02	Project(Advanced)	10	4
STA10G01	Comprehensive Viva	---	4
	TOTAL	**25**	**20**

The Open course shall chosen from the following list:

Course Code	Course Title	Hours/ Week	Credits
D01	Introductory Probability	3	3
D02	Basics of Statistical Inference	3	3
D03	Statistical Techniques for Research Methods	3	3

The courses Elective – I, Elective – II, Elective –III and Elective IV shall chosen from the following list:

Course Code	Course Title	Hours/ Week	Credits
E01	Time Series Analysis	5	4
E02	Advanced Operations Research	5	4
E03	Queuing Theory	5	4
E04	Life time Data Analysis	5	4
E05	Advanced Distribution Theory	5	4
E06	Statistical Decision Theory	5	4
E07	Reliability Modeling	5	4
E08	Actuarial Statistics	5	4
E09	Official Statistics	5	4
E10	Biostatistics	5	4
E11	Econometric Models	5	4
E12	Demographic Techniques	5	4
E13	Stochastic Finance	5	4
E14	Longitudinal Data Analysis	5	4
E15	Data Mining Techniques	5	4
E16	Statistical Machine Learning -I	5	4
E17	Statistical Machine Learning - II	5	4
E18	Advanced Statistical Machine Learning Techniques	5	4
E19	Non-Parametric Statistical Methods	5	4
E20	Statistical Modeling and Data Mining Techniques	5	4
E21	Applied Algorithms and Analysis of Multi type and Big Data	5	4

Assessment & Evaluation

Mark system is followed for each question. For each course in the semester letter grade and grade point are introduced in 10-point indirect grading system as per regulation.

Course Evaluation: The evaluation scheme for each course shall contain two parts

1) Internal assessment 2) External Evaluation

20% weight shall be given to the internal assessment. The remaining 80% weight shall be for the external evaluation, except for project. For project evaluation, existing criteria in the UG/PG Regulations (CBCSS UG 2019) will be made applicable to the foundation and advanced programmes respectively, in the Integrated Programmes.

Internal Assessment

20% of the total marks in each course are for internal examinations. The marks secured for internal assessment only need to be sent to University by the colleges concerned. The internal assessment shall be based on a predetermined transparent system involving written tests, Class room participation based on attendance in respect of theory courses and lab involvement/records attendance in respect of Practical Courses. Internal assessment of the project will be based on its content, method of presentation, final conclusion and orientation to research aptitude.

Components with percentage of marks of Internal Evaluation of Theory Courses are-Test paper 40%, Assignment 20%, Seminar 20% and Class room participation based on attendance 20%. For practical courses - Record 60% and lab involvement 40% as far as internal is concerned. (if a fraction appears in internal marks, nearest whole number is to be taken). For the test paper marks, at least one test paper should be conducted. If more test papers are conducted, the mark of the best one should be taken. To ensure transparency of the evaluation process, the internal assessment marks awarded to the students in each course in a semester shall be notified on the notice board at least one week before the commencement of external examination. There shall not be any chance for improvement for internal marks. The course teacher(s) shall maintain the academic record of each student registered for the course, which shall be forwarded to the University by the college Principal after obtaining the signature of both course teacher and Head of the Department.

The Split up of marks for Test paper and Class Room Participation (CRP) for internal evaluation of Integrated Programmes (for both foundation and advanced) are as follows:

Split up of marks for test papers

Range of marks in test papers	Out of 8 (Maximum internal marks is 20)	Out of 6 (Maximum internal mark is 15)
Less than 35%	1	1
35% -- 45%	2	2
45%--55%	3	3
55%--65%	4	4
65% -- 85%	6	5
85% -- 100%	8	6

Split of class room participation:

Range of CRP	Out of 4 (Maximum internal marks is 20)	Out of 3 (Maximum internal mark is 15)
$50\% \leq CRP \leq 75\%$	1	1
$75\% \leq CRP \leq 85\%$	2	2
85% and above	4	3

External Evaluation

For Foundation Programme, External evaluation carries 80% of marks. All question papers shall be set by the University. The external question papers may be of uniform pattern with 80/60 marks. The courses with 2/3 credits will have an external examination of 2 hours duration with 60 marks and courses with 4/5 credits will have an external examination of 2.5 hours duration with 80 marks.

For Advanced Programme, External evaluation carries 80% of marks. All question papers shall be set by the University. The external question papers may be of uniform pattern with 80 marks having duration of 3 hours.

The external examination in theory courses is to be conducted by the University with question papers set by external experts/Question bank prepared by the Board of Studies.

The evaluation of the answer scripts shall be done by examiners based on a well-defined scheme of valuation and answer keys shall be provided by the University. The external examination in practical courses shall be conducted by two examiners – one internal and an external, the latter appointed by the University. The project evaluation with viva can be conducted either internal or external which may be decided by the Board of Studies concerned. After the external evaluation only marks are to be entered in the answer scripts. All other calculations including grading are done by the University. Regarding the internal/external proportion of advanced programme, the existing pattern of respective discipline farmed by the respective Board of Studies is applicable. The different criteria set for the internal evaluation of advanced programme (from 7th semester to 10 semesters) are same as that of a non integrated PG programme.

Question Paper Pattern:

Foundation Programme(I to VII Semester)

Courses having 2/3 Credits- Duration of the examination is two hours

Question number (From….. To …..)	**Type of Questions and Marks**
01 to 12	Short answer type carries 2 marks each - 12 questions **(Maximum Marks 20)**
13 to 19	Paragraph/ Problem type carries 5 marks each – 7 questions **(Maximum Marks 30)**
20 to 21	Essay type carries 10 marks (1 out of 2) **(Maximum Mark 10)**
01 to 21	**Total Marks: 60**

Courses having 4 Credits- Duration of the examination is two and half hours

Question number (From….. To …..)	**Type of Questions and Marks**
01 to 15	Short answer type carries 2 marks each - 15 questions **(Maximum Marks 25)**
16 to 23	Paragraph/ Problem type carries 5 marks each – 8 questions **(Maximum Marks 35)**
24 to 27	Essay type carries 10 marks (2 out of 4) **(Maximum Mark 20)**
01 to 27	**Total Marks: 80**

Advanced Programme(VII to X Semester)

Duration of the examination is three hours.

	Question number (From….. To …..)	**Type of Questions and Marks**
Section A	I (i) to (viii)	Answer any **Four** questions; each question carries **4** marks
Section B	II A a) b) or B a) b) **to** V A a) b) or B a) b)	Answer either **Part A** or **Part B** of all questions; each question carries **16** marks
		Maximum Marks: 80

CORE COURSE – 1

Semester	I
Course Code	STA1IB01
Course Name	Statistical Methods
Hours per Week	03
Credit	02
Examination Hours	Two Hours

Objective of the Course:

To make the students aware of different type of data sets and their graphical representations introducing of descriptive statistical measures.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To understand the concepts of statistical population and sample, variables and attributes.
CO2 : To understand tabular and graphical representation of data based on variables.
CO3 : To compute various measures of central tendency, dispersion, skewness and kurtosis.
CO4 : To understand moments and their use in studying various characteristics of data.

Contents :

Module 1 - (20 Hours)

Meaning and scope of Statistics : General uses, relation with other discipline, Limitation and misuse of Statistics, population, census, parameter, sample, sample survey, statistic; Data types : quantitative, qualitative, geographical and chronological, nominal, ordinal, ratio, interval scale, time series cross sectional and longitudinal data; Methods of collecting data editing of primary data, designing of a questionnaire and a schedule, sources and editing of secondary data ; Classification and tabulation of data : basic principles of tabulation, construction of frequency table; Diagrammatic representation- line diagram, bar diagram and pie diagram, pictogram, cartogram; Graphical representation of frequency distribution - histogram, frequency polygon, frequency curve and ogives.

Module 2 - (10 Hours)

Measures of central tendency – Concept, requirement of a good measure, arithmetic mean, weighted arithmetic mean, median, positional averages – quartiles, deciles and percentiles, mode, geometric mean, harmonic mean, graphical method of determination of median, mode and quantiles, properties and scope of these averages

Module 3 – (10 Hours)

Measures of dispersion- Concept , requirement of a good measures of dispersion, range, quartile deviation, mean deviation, mean square deviation and standard deviation and their relative measure, effect of change of origin and scale, standard deviation of pooled data; coefficient of variation.

Module 4 - (8 Hours)

Moments- Raw moments and central moments, relation between central moment and raw moments, Sheppard correction for moments; Skewness- measures of skewness and types of skewness; Kurtosis – measure of kurtosis and types of kurtosis.

Books for Study

1. Gupta S.C and Kapoor V .K. (2020). *Fundamentals of Mathematical Statistics (12 Edn.)*. Sultan Chand & Co., New Delhi.
2. Saxena, H.C. (1983). *Elementary Statistics*. S. Chand & Co., New Delhi.

Books for References

1. Agarwal B. L. (2013). *Basic Statistics*. New Age International Publishers.
2. Goon A.M., Gupta M.K., Das Gupta B. (1999). *Fundamentals of Statistics*, Vol. I, World Press, Calcutta.
3. Spiegel M. R. (1961). *Theory and problems of statistics*. Schaum's outline series, New York.

CORE COURSE – 2

Semester	II
Course Code	STA2IB02
Course Name	Basic Probability Theory
Hours per Week	03
Credit	02
Examination Hours	Two Hours

Objective of the Course

To introduce the notion of probability, random variable and expectation based on which statistical theory and tools have been developed.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To evaluate the probability of events.
CO2 : To understand the concept of random variable and classification.
CO3 : To understand the probability mass function and probability density function.
CO4 : To know the probability distribution function and its properties.
CO5 : To know the mathematical expectation and generating function and their properties.

Contents :

Module 1 - (16 Hours)

Concept of experiments – Deterministic and probabilistic; Outcome of experiment – Sample space, Discrete (finite and countably finite) and continuous sample space, event, elementary event, compound event, algebra of events, mutually exclusive events, exhaustive events, impossible events, and certain events, Venn diagram.

Classical definition of probability- real life problems, merits and demerits; Probability as an approximation to the relative frequency; Computation of events based on permutations and Combinations, with and without replacement; Axiomatic definition of probability- Addition theorem (two and three events) and properties.

Module 2 – (10 Hours)

Conditional probability- Multiplication theorem for two events, independence of events, pairwise and mutual independence of events; Partition of sample space; Bayes' theorem – Statement, proof and applications.

Module 3 – (12 Hours)

Definition of random variable – Discrete, continuous and mixture random variable; Probability mass function and probability density function – definition and properties; Distribution function- Properties, relation between distribution function and mass/density function; Median and mode of a univariate discrete and continuous random variable; Change of variable (univariate case only).

Module 4 – (10 Hours)

Definition of expectation of a random variable –Properties, expectation of functions of random variables, moments, relation between raw moments and central moments, skewness and kurtosis of random variable; Generating functions – Definition, properties of moments generating function, characteristic function and probability generating function .

Books for Study

1. Gupta S.C and Kapoor V .K. (2020). *Fundamentals of Mathematical Statistics (12 Edn.)*. Sultan Chand & Co., New Delhi.
2. Rohatgi V K and Saleh (2008) . *An Introduction to Probability and Statistics*, John Wiley and Son Inc.

Books for References

1. Mood A.M., Graybill F.A. and Bose D.C. (2007). Introduction to the Theory of Statistics, 3rd Edn. (Reprint), Tata McGraw –Hill Pub. Co. Ltd.
2. Goon A. M., Gupta M.K. and Dasgupta B. (2002). Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkatta.
3. Ross S.M. (2019). Introduction to Probability Models. Academic Press.
4. Spiegel M.R. and Stephens L.J. (2010). Statistics – Schaum's Outline series- Fourth edition. Tata Mcgraw Hill Publishing Company Ltd.

CORE COURSE – 3

Semester	III
Course Code	STA3IB03
Course Name	Bivariate Data Analysis
Hours per Week	03
Credit	03
Examination Hours	Two Hours

Objective of the Course

To understand the relationship between two or more variables.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To analyse the data and predict the future values using curve fitting
CO2 : To measure and interpret the degree of relationship between variables.
CO3 : To estimate the average relationship using regression.
CO4 : To interpret the association of attributes applying different methods.
CO5 : To understand probability distribution for bivariate data.

Contents :

Module 1 - (8 Hours)

Scatter diagram – Curve fitting – Principle of least squares – Fitting of straight line, parabola, exponential and power curves.

Module 2 – (15 Hours)

Bivariate linear correlation – Pearson correlation coefficient – Spearman rank correlation coefficient – Bivariate linear regression – Regression lines – Coefficient of regression-Multiple and partial correlation for three variables (without proof)

Module 3 – (5 Hours)

Association of attributes- Relation between class frequencies- Consistency of data- Independence of attributes –Criterion of Independence –Yule's coefficient of association- Yule's coefficient of colligation.

Module 4 – (20 Hours)

Bivariate random variables: Joint pmf and joint pdf, marginal and conditional probability, independence of random variables, function of random variable. Bivariate Expectations, conditional mean and variance, covariance, independence of random variables based on expectation.

Books for Study

1. Gupta S.C and Kapoor V .K. (2020). *Fundamentals of Mathematical Statistics (12 Edn.)*. Sultan Chand & Co., New Delhi.
2. Gupta S.P. (2011) Statistical Methods, 4th Edition, Sultan Chand & Sons (Publisher), New Delhi, India
3. Rohatgi V K and Saleh (2008) . *An Introduction to Probability and Statistics*, John Wiley and Son Inc.

Books for References

1. Mood A.M., Graybill F.A. and Bose D.C. (2007). Introduction to the Theory of Statistics, 3rd Edn. (Reprint), Tata McGraw –Hill Pub. Co. Ltd.
2. Goon A. M., Gupta M.K. and Dasgupta B. (2002). Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkatta.
3. Ross S.M. (2019). Introduction to Probability Models. Academic Press.
4. Spiegel M.R. and Stephens L.J. (2010). Statistics – Schaum's Outline series- Fourth edition. Tata Mcgraw Hill Publishing Company Ltd.

CORE COURSE – 4

Semester	III
Course Code	STA3IB04
Course Name	Probability Distributions
Hours per Week	04
Credit	03
Examination Hours	Two Hours

Objective of the Course

To understand various types of probability distributions, study their properties and explore the applications of probability distributions.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To equip the students to use discrete standard distributions to explain random phenomena.
CO2 : To equip the students to interpret probabilistic nature of data using continuous probability distributions.
CO3 : To Apply laws of large numbers and central limit theorems in solving problems.
CO4 : To understand functions of normal variable leading to chi square, 't' and 'F' distribution and their interrelation.

Contents :

Module 1 - (14 Hours)

Discrete distributions – Uniform, Bernoulli, Binomial, Geometric, Poisson- mean, variance, m.g.f and their properties- Negative binomial, Hyper geometric distribution (Definition only) .

Module 2 – (20 Hours)

Continuous Distribution – Normal distribution- properties, standard normal table, fitting of normal distribution, - Uniform, exponential, Gamma, Beta (two types) – mean, variance, m.g.f and their properties- Cauchy, Laplace , Weibull, Pareto distribution (Definition only)

Module 3 – (15 Hours)

Limit theorems- Chebyshev's inequality- convergence in probability –Weak law of large numbers (i.i.d. case)-Bernoulli law of large numbers- Central limit theorem (Lindberg-Levy i.i.d. case)- Normal distribution as a limiting case of binomial and Poisson under suitable assumptions.

Module 4 – (15 Hours)

Concept of random sample and statistic- Sampling distribution of a statistic, standard error, sampling distributions of the mean and variance of a random sample taken from normal

population- Chi square, ‘t’ and ‘F’ distributions – Definition, properties, uses and inter relationship

Books for Study

1. Gupta S.C and Kapoor V .K. (2020). *Fundamentals of Mathematical Statistics (12 Edn.)*. Sultan Chand & Co., New Delhi.
2. Rohatgi V K and Saleh (2008) . *An Introduction to Probability and Statistics*, John Wiley and Son Inc.

Books for References

1. Mood A.M., Graybill F.A. and Bose D.C. (2007). Introduction to the Theory of Statistics, 3rd Edn. (Reprint), Tata McGraw –Hill Pub. Co. Ltd.
2. Goon A. M., Gupta M.K. and Dasgupta B. (2002). Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkatta.
3. Ross S.M. (2019). Introduction to Probability Models. Academic Press.
4. Spiegel M.R. and Stephens L.J. (2010). Statistics – Schaum’s Outline series- Fourth edition. Tata Mcgraw Hill Publishing Company Ltd.

CORE COURSE – 5

Semester	IV
Course Code	STA4IB05
Course Name	Statistical Inference -I
Hours per Week	04
Credit	03
Examination Hours	Two Hours

Objective of the Course

To introduce the concept of estimation, learn properties of estimators and to distinguish point estimation and interval estimation and to introduce the notion of order statistics.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To understand the concept of Statistical inference and parametric estimation.
CO2 : To understand the point estimation, estimator and its properties.
CO3 : To know the different method of estimation.
CO4 : To know the interval estimation and solving problems in interval estimation.
CO5: To know the distribution of order statistics.

Contents :

Module 1 - (14 Hours)

Concept of Statistical inference- Parametric estimation – Point estimation – Estimator, properties of estimator- unbiasedness, consistency, efficiency and sufficiency of the estimator- Cramer-Rao inequality- Rao , Blackwell theorem (Statement only).

Module 2 – (20 Hours)

Methods of point estimation-Method of moments, Method of maximum likelihood and method of minimum chi-square-properties of estimators obtained by these methods.

Module 3 – (15 Hours)

Interval estimation – Large sample confidence interval for mean, equality of means, equality of proportions, Derivations of exact confidence intervals for means, variance and ratio of variance based on Normal, "t", chi-square and "F" distribution- Numerical problems in interval estimation.

Module 4 – (15 Hours)

Order statistics- Definition, derivation of p.d.f. of i^{th} order statistics for a random sample of size n from a continuous distribution-Density of smallest and largest observations.

Books for Study

1. Gupta S.C and Kapoor V .K. (2020). *Fundamentals of Mathematical Statistics (12 Edn.)*. Sultan Chand & Co., New Delhi.
2. Rohatgi V K and Saleh (2008) . *An Introduction to Probability and Statistics*, John Wiley and Son Inc.

Books for References

1. Mood A.M., Graybill F.A. and Bose D.C. (2007). Introduction to the Theory of Statistics, 3rd Edn. (Reprint), Tata McGraw –Hill Pub. Co. Ltd.
2. Goon A. M., Gupta M.K. and Dasgupta B. (2002). Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkatta.
3. Ross S.M. (2019). Introduction to Probability Models. Academic Press.
4. Spiegel M.R. and Stephens L.J. (2010). Statistics – Schaum's Outline series- Fourth edition. Tata Mcgraw Hill Publishing Company Ltd.
5. Mukhopadhyay P (1996). Mathematical Statistics, New central Book Agency (P) Ltd. Culcutta.

CORE COURSE – 6

Semester	IV
Course Code	STA4IB06
Course Name	Sampling Techniques
Hours per Week	03
Credit	03
Examination Hours	Two Hours

Objective of the Course

To introduce the concept of census and sample surveys, learn various methods of sampling and study the properties of estimators under different sampling methods.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To understand the basic of sampling and related terminologies.
CO2 : To understand various types of sampling schemes, with their advantages and disadvantages and estimation of population parameters with their standard errors.

Contents :

Module 1 - (10 Hours)

Concepts of population and sample –Sample frame-Sampling design- Need for sampling-Principle steps in sample survey-Advantages of sample survey over census survey- Probability sampling and non probability sampling-Basic concepts in sampling –Organizational aspects of survey sampling – Sampling and non-sampling errors- Sample selection and sample size.

Module 2 – (12 Hours)

Simple random sampling with and without replacement- Estimation of population mean and variance-Expectation and variance of estimators- Confidence interval for population mean - Estimation of sample size based on desired accuracy for variables and attributes.

Module 3 – (14 Hours)

Concepts of stratified population- Stratified sample estimation of population mean and total-Mean and variance of estimator of population mean assuming SRSWOR with in strata-Proportional allocation – Optimum allocation with and without varying costs-Comparison of simple random sampling with proportional and optimum allocation.

Module 4 – (12 Hours)

Concepts of systematic population - systematic sample-estimation of population mean and total- expectation and variance of estimators-circular systematic sampling- comparison with stratified sampling-population with linear trend.

Books for Study

1. Cochran W G (2017). Sampling Techniques, Wiley Eastern Limited, New Delhi.
2. Gupta, S.C. and Kapoor V.K. (2002). Fundamentals of Applied Statistics, Sultan Chand & Co. New Delhi.

Books for References

1. Daroga Singh and F S Chaudhary(2020) . Theory and Analysis of Sample Survey Design, Wiley Eastern Limited.
2. Sukhatme P V, Sukatme B V (2012). Sampling Theory of Surveys with Applications, Indian Society of Agricultural Statistics, New Delhi.

CORE COURSE – 7

Semester	V
Course Code	STA5IB07
Course Name	Statistical Inference -II
Hours per Week	05
Credit	04
Examination Hours	Two and half Hours

Objective of the Course

To gain knowledge on the methods of testing statistical hypothesis and understand the meaning and applications of tests of significance.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To understand testing of statistical hypothesis.
CO2 : To understand test of significance test based on normal, chi-square, 't' and 'F' distribution.
CO3 : To understand Contingency table and goodness of fit.
CO4 : To understand free and non-parametric tests, Mann-Whitney tests

Contents :

Module 1 - (20 Hours)

Testing of Statistical hypothesis-Simple and Composite hypothesis-Null and alternative hypothesis-Sample and parameter space- Two types of errors- Critical region- Power of a test-Neymann-Pearson lemma- Simple applications.

Module 2 – (10 Hours)

Most powerful tests-Uniformly most powerful and unbiased tests based on normal, chi-square, 't' and 'F' distributions – likelihood ratio criterion –definition and simple applications.

Module 3 – (20 Hours)

Test of significance –Asymptotic and exact tests based on Normal, chi-square, 't', and ' F' distributions with regard to mean, proportion, variance, Standard deviation, coefficient of correlation, regression coefficients, partial and multiple correlation coefficients-Concept of observed significance level.

Module 4 – (30 Hours)

Contingency table –Test for independence by contingency tables –goodness of fitness tests – tests of homogeneity of variances, correlation and proportions .Test of Normality (application only).

Elementary ideas on distribution – free and non-parametric tests – Kolmogrov-Smirnov test, Sign test & Wilcoxon signed rank test (One sample test only).

Books for Study

1. Gupta S.C and Kapoor V .K. (2020). *Fundamentals of Mathematical Statistics (12 Edn.).* Sultan Chand & Co., New Delhi.
2. Rohatgi V K and Saleh (2008) . *An Introduction to Probability and Statistics*, John Wiley and Son Inc.

Books for References

1. Mood A.M., Graybill F.A. and Bose D.C. (2007). Introduction to the Theory of Statistics, 3rd Edn. (Reprint), Tata McGraw –Hill Pub. Co. Ltd.
2. Goon A. M., Gupta M.K. and Dasgupta B. (2002). Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkatta.
3. Spiegel M.R. and Stephens L.J. (2010). Statistics – Schaum's Outline series- Fourth edition. Tata Mcgraw Hill Publishing Company Ltd.
4. Mukhopadhyay P (1996). Mathematical Statistics, New central Book Agency (P) Ltd. Culcutta.

CORE COURSE – 8

Semester	V
Course Code	STA5IB08
Course Name	Design and Analysis of Experiments
Hours per Week	05
Credit	04
Examination Hours	Two and half Hours

Objective of the Course

To impart the knowledge of performing analysis of variance, propose methods of performing analysis of variance under different types of experimental design and lean how to estimate missing observations and compare the efficiencies of various design.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To understand the linear design models, analysis of variance.
CO2 : To understand the fundamentals of experimentation.
CO3 : To analyse the different experiments CRD, RBD and LSD.
CO4 : To understand factorial design 2^2 and 2^3 factorial design .

Contents :

Module 1 - (20 Hours)

Linear estimation- Estimability of parametric functions and BLUE- Gauss-Markov theorem- Linear Hypothesis

Module 2 – (10 Hours)

Analysis of one way classified data – Analysis of two way classified data with one observation per cell- Analysis of two way classified data with 'm' observations per cell. Analysis of two way classified data with unequal number of observations in cells under fixed effect model.

Module 3 – (20 Hours)

Basic terms in design of experiments: Experimental unit, treatment, layout of an experiment. Basic principles of design of experiments: Replication, randomization and local control. Choice of size and shape of a plot for uniformity trials, the empirical formula for the variance per unit area of plots.

Module 4 – (15 Hours)

Complete randomized design, randomized block design and Latin square design. Layout, model, assumptions and interpretations: Estimation of parameters, expected values of mean sum of squares, components of variance. Tests and their interpretations, test for equality of two specified treatment effects, comparison of treatment effects using critical difference

Module 5 – (15 Hours)

Basic concepts of factorial experiments-2^2 and 2^3 factorial experiments-Basic concepts of Incomplete block design-Balanced incomplete block design (Concept only).

Books for Study

1. Joshy D D (2020). Linear Estimation and Design of Experiments, New Age International Publications
2. Gupta S.C and Kapoor V .K. (2020). Applied Statistics, Sultan Chand & Co., New Delhi.

Books for References

1. Das M N and Giri N (2017) Design of Experiments, New Age International.
2. Montegomery, D C (2012). Design and Analysis of Experiments, John Wiley & Sons, New York
3. Parimal M (1999). Applied Statistics, 2nd Edition, Books & Applied Ltd. Kolkata, India.

CORE COURSE – 9

Semester	V
Course Code	STA5IB09
Course Name	Data Analytics with R
Hours per Week	05
Credit	04
Examination Hours	Two and half Hours

Objective of the Course

To understand the basics R programming and apply exploratory data analytics using data visualization.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To understand the basics of Fundamentals of R
CO2 : To understand the loading, retrieval techniques of data
CO3 : To understand how data is analysed and visualized using statistic function

Contents :

Module I - (22 Hours)

Introduction to R:What is R? – Why R? – Advantages of R over Other Programming Languages - R Studio: R command Prompt, R script file, comments – Handling Packages in R: Installing a R Package, Few commands to get started: installed.packages(), packageDescription(), help(), find.package(), library() - Input and Output – Entering Data from keyboard – Printing fewer digits or more digits – Special Values functions : NA, Inf and –inf.

R Data Types: Vectors, Lists, Matrices, Arrays, Factors, Data Frame – R - Variables: Variable assignment, Data types of Variable, Finding Variable ls(), Deleting Variables - R Operators: Arithmetic Operators, Relational Operators, Logical Operator, Assignment Operators, Miscellaneous Operators - R Decision Making: if statement, if – else statement, if – else if statement, switch statement – R Loops: repeat loop, while loop, for loop - Loop control statement: break statement, next statement.

Module II – (20 Hours)

R-Function : function definition, Built in functions: mean(), paste(), sum(), min(), max(), seq(), user-defined function, calling a function, calling a function without an argument, calling a function with argument values - R-Strings – Manipulating Text in Data: substr(), strsplit(), paste(), grep(), toupper(), tolower() - R Vectors – Sequence vector, rep function, vector access, vector names, vector math, vector recycling, vector element sorting - R List - Creating a List, List Tags and Values, Add/Delete Element to or from a List, Size of List, Merging Lists, Converting List to Vector - R Matrices – Accessing Elements of a Matrix, Matrix Computations: Addition, subtraction, Multiplication and Division- R Arrays: Naming Columns and Rows, Accessing Array Elements, Manipulating Array Elements, Calculation Across Array Elements - R Factors –creating factors, generating factor levels gl().

Module III (20 Hours)

Data Frames –Create Data Frame, Data Frame Access, Understanding Data in Data Frames: dim(), nrow(), ncol(), str(), Summary(), names(), head(), tail(), edit() functions - Extract Data from Data Frame, Expand Data Frame: Add Column, Add Row - Joining columns and rows in a Data frame rbind() and cbind() – Merging Data frames merge() – Melting and Casting data melt(), cast(). Loading and handling Data in R: Getting and Setting the Working Directory – getwd(), setwd(), dir() - R-CSV Files - Input as a CSV file, Reading a CSV File, Analyzing the CSV File: summary(), min(), max(), range(), mean(), median(), apply() - Writing into a CSV File – R -Excel File – Reading the Excel file.

Module IV (18 Hours)

Descriptive Statistics: Data Range, Frequencies, Mode, Mean and Median: Mean Applying Trim Option, Applying NA Option, Median - Mode - Standard Deviation – Correlation - Spotting Problems in Data with Visualization: visually Checking Distributions for a single Variable - R –Pie Charts: Pie Chart title and Colors – Slice Percentages and Chart Legend, 3D Pie Chart – R Histograms – Density Plot - R – Bar Charts: Bar Chart Labels, Title and Colors.

Books for Study

1. Sandip Rakshit (2017) R Programming for Beginners, McGraw Hill Education (India)
2. Seema Acharya (2018) Data Analytics using R, Mc Graw Hill Education (India).

Books for References

1. Tutorials Point (I) simply easy learning, Online Tutorial Library (2018) R Programming , Retrieved from https:// www.tutorialspoint.com/r/r_tutorial.pdf
2. Anddrie de Vries, Joris Meys (2015). R for Dummies , John Wiley and Sons Inc.

CORE COURSE – 10

Semester	V
Course Code	STA5IB10
Course Name	Applied Statistics
Hours per Week	05
Credit	04
Examination Hours	Two and half Hours

Objective of the Course

To be acquainted with the knowledge of time series analysis, understand the significance of index numbers and its type and describe demographic data and vital statistics measures.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To identify the components of time series and the method of measuring trend.
CO2 : To apply the different measures of variations to forecast the data.
CO3 : To construct index numbers, evaluate the cost of living index and interpret.
CO4 : To understand the vital statistics and its importance in the civic society.

Contents :

Module 1 - (15 Hours)

Concept – components of time series –additive and multiplicative models-Resolving components of a time series-measuring trend: Graphic, semi-averages, moving average and principle of least squares methods

Module 2 – (15 Hours)

Seasonal variation- measuring seasonal variation: method of simple averages, ratio to trend method, ratio to moving average method and link relative method- Cyclical and Random fluctuations- variate difference method.

Module III – (25 Hours)

Index numbers and their definitions – construction of index numbers -simple and weighted index numbers - Laspeyre's, Paache's, Fisher's, Marshall-Edgeworth, Dorbish-Bowley's and Kelly's index numbers – Quantity index numbers.

Test on index numbers-factor reversal test, time reversal test, circular test, chain indexnumbers-base shifting, splicing and deflating of index numbers- Consumer price index number.

Module IV (25 Hours)

Demography – definition-sources of demographic data: vital registration-population census - population register-demographic surveys-population data as aid to social, economic andhealthy planning - process of Indian Civil registration and census.

Fertility measurements: Fertility as a component of population change - Crude Birth Rate (CBR)- General, Specific and Total Fertility Rates(GFR, ASFR,TFR) - Gross and Net Reproduction Rates (GRR & NRR) the relationships and interpretation.

Books for Study

1. Gupta S C, Kapoor V K (2019) Fundamentals of Applied Statistics, Fourth Edition, Sultan Chand & Sons, NewDelhi, India
2. Goon A M , Guptha M K , Dasgupta B (2016). Fundamentals of Statistics, Vol. II World Press Kolkatta, India.

Books for References

1. Agarwal B L (2006). Basic Statistics, New Age International Private Limited, New Delhi.
2. Parimal M (1999). Applied Statistics 2nd Edition, Books & Applied Ltd. Kolkatta, India.

CORE COURSE – 11

Semester	VI
Course Code	STA6IB11
Course Name	Statistical Quality Control
Hours per Week	05
Credit	04
Examination Hours	Two and half Hours

Objective of the Course

To understand the concept of statistical quality control

Course Outcome :

After completing the Course, the Student should be able

CO1 : To understand need for SQC
CO2 : To understand control chart for attributes np, p, c chart.
CO3 : To know the acceptance sampling for attributes –single, double and sequential sampling plans
CO4 : To make aware of the use of statistical techniques in decision making.

Contents :

Module 1 - (20 Hours)

Concept of Quality – Quality movement in India – Standardization for Quality – Quality movement – Quality management – Quality circles – Total Quality Management – ISO 9001.

Module 2 – (20 Hours)

Need for SQC in industries; Process control: Chance and assignable causes of variation - specification and tolerance limits; process capability- Statistical basis for control charts: X , R and standard deviation charts - their construction and analysis

Module III – (20 Hours)

Control charts for attributes – p, np, c and u charts – their construction and analysis

Module IV (20 Hours)

Product control: Acceptance sampling by attributes; Producer's and Consumer's risk-Notions of AQL, LTPD and AOQL – Concepts of Single and Double sampling plans

OC, AOQ, ASN, ATI curves for Single and Double sampling plans – Concept of Sequential sampling plan for attributes.

Books for Study

1. Montogomery, D. C. (2013): Introduction to Statistical Quality Control, 7th Edition, Wiley India Pvt. Ltd.
2. Gupta and Kapoor V K (2000). Fundamentals of Applied Statistics, Sultan Chand and Sons.
3. Goon A.M., Gupta M.K. and Dasgupta B. (2002): Fundamentals of Statistics, Vol. I & II, 8th Edn. The World Press, Kolkata

Books for References

1. Mukhopadhyay, P (2011):Applied Statistics, 2nd edition revised reprint, Books and Allied(P) Ltd.
2. Montogomery, D. C. and Runger, G.C. (2008): Applied Statistics and Probability for Engineers, 3rd Edition reprint, Wiley India Pvt. Ltd.
3. Ehrlich, B. Harris (2002): Transactional Six Sigma and Lean Servicing, 2nd Edition, St. Lucie Press.
4. Hoyle, David (1995): ISO Quality Systems Handbook, 2nd Edition, Butterworth Heinemann Publication. India.

CORE COURSE – 12

Semester	VI
Course Code	STA6IB12
Course Name	Basic Regression Analysis
Hours per Week	05
Credit	04
Examination Hours	Two and half Hours

Objective of the Course

To identify an appropriate relationship between two variables using scatter plot and fitting the same by the method of least squares and perform regression analysis

Course Outcome :

After completing the Course, the Student should be able

CO1 : To understand the regression model
CO2 : To know residual plot.
CO3 : To understand multiple regression and correlation coefficient.
CO4 : To identify multicollinearity.

Contents :

Module 1 - (25 Hours)

Simple Regression model: Description of data model – Assumption about the explanatory variable - Estimation and test of hypotheses - Confidence Intervals – Predicted values and standard errors – Evaluation of fit.

Module 2 – (20 Hours)

Analysis of residual - Residual plots – Transformation of variables – transformation to stabilize variance – Removal of hetroscedasticity

Module III – (25 Hours)

Multiple regression model: Description of data model – Properties of least square estimators – Predicted values and standard errors – Multiple correlation coefficient - Selection of variables – Forward selection procedure – Backward elimination procedure – Stepwise method (algorithms only).

Module IV (10 Hours)

Test of hypothesis on the linear model — Testing a subset of regression coefficients equal to zero – Testing of equality of regression coefficients.

Multicollinearity and its effects on inference and forecasting – Detection of multicollinearity – Remedial measures.

Books for Study

1. Montegomerry D C, Peak E A and Vining G G (2007). Introduction to Linear regression analysis, John Wiley & Sons (Asia) Pvt. Ltd. Singapore.
2. Chatterjee S and Price B (2013). Regression Analysis by Example, John Wiley & SOns, New York.

Books for References

1. Johnston J (1997). Econometric Methods, McGraw-Hill Education - Europe
2. Thomson P Ryan (2006). Modern Regression Methods, John Wiley and Sons, Inc.

CORE COURSE – 13

Semester	VI
Course Code	STA6IB13
Course Name	Operations Research
Hours per Week	05
Credit	04
Examination Hours	Two and half Hours

Objective of the Course

To know the methods of solution of linear programming, understand the concept and solution of Transportation problems, assignment problem game theory and sequencing problem.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To understand methods of solving linear programming problem
CO2 : To know the concept and solution of Transportation problems.
CO3 : To understand the concept and solving of assignment problem
CO4 : To understand the concept of game theory.
CO4 : To understand the concept and solving sequencing problem

Contents :

Module 1 - (25 Hours)

Introduction to Operations Research, phases of O.R., model building, various types of O.R. problems. Linear Programming Problem, Mathematical formulation of the L.P.P, graphical solutions of a L.P.P. Simplex method for solving L.P.P. Charne's M-technique for solving L.P.P. involving artificial variables. Special cases of L.P.P. Concept of Duality in L.P.P: Dual simplex method

Module 2 – (20 Hours)

Transportation Problem: Initial solution by North West corner rule, Least cost method and Vogel's approximation method (VAM), MODI's method to find the optimal solution, special cases of transportation problem.

Assignment problem: Hungarian method to find optimal assignment, special cases of assignment problem.

Module III – (25 Hours)

Game theory: Rectangular game, minimax-maximin principle, solution to rectangular game using graphical method, dominance and modified dominance property to reduce the game matrix and solution to rectangular game with mixed strategy.

Module IV (10 Hours)

Sequencing: Introduction – n jobs and two machines, n jobs and three machines, n jobs and m machines – idle times and total elapsed time calculations.

Books for Study

1. Taha, H. A. (2007): Operations Research: An Introduction, 8th Edition, Prentice Hall of India.
2. Kanti Swarup, Gupta, P.K. and Manmohan (2007): Operations Research, 13th Edition, Sultan Chand and Sons

Books for References

1. Hadley, G: (2002) : Linear Programming, Narosa Publications
2. Sharma, S. D. (2017). Operations Research: Theory, Methods and Applications, Kedar Nath, Ram Nath and Co, Meerut

CORE COURSE -14

Semester	VI
Course Code	STA6IB14
Course Name	Analytical Tools for Statistics
Hours per Week	04
Credit	04
Examination Hours	Two and Half Hours

Objective of the Course

To understand the mathematical concept required to learn theoretical Statistics.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To understand the limits of sequence and series for real numbers.
CO2 : To understand the concept of complex numbers.
CO3 : To understand the concept of analytic function.
CO4 : To know about pole and singularity.

Contents :

Module 1 - (18 Hours)

Limits of Sequences of numbers- Theorems for calculating limits of sequences(Excluding Picard's Method), Infinite series, Comparison test for series of non negative terms, Ratio and root test for series of non negative terms. Continuous function-Boundedness theorem-Maximum –Minimum theorem-Intermediate value theorem.

Module 2 – (14 Hours)

Limit and Continuity- Limit of a complex function, condition for non existence of limit, real and imaginary parts of limit, properties of complex limits, continuity, discontinuity of principal square root function, properties of continuous functions, continuity of polynomial and rational functions- Differentiability and Analyticity – Derivative of a complex Function, rules of differentiation, function that is nowhere differentiable, Analytic functions, entire functions, singular points.

Module 3 – (16 Hours)

Cauchy Riemann Equations- Necessary condition for analyticity, Criterion for non analyticity, sufficient condition for analyticity, sufficient condition for differentiability, Cauchy Riemann equations in polar coordinates. Harmonic Functions- definition, analyticity and harmonic nature, harmonic conjugate functions, finding harmonic conjugate. Zeros and Poles- classification of isolated singular points, removable singularity, pole, essential singularity, order of zeros and poles.

Books for Study

1. George B T Jr, Ross L F (1998) Calculus and Analytic Geometry , LPE, 9 Edition, Pearson Education.
2. Dennis Z L S (2015). Complex Analysis A First Course with Applications, 3rd Edition, Jones and Bartlett Learning.

Books for References

1. Malik S.C and Savitha Arora (2005). Mathematical , New Age International.
2. Shanti Narayanan (2003). Elements of Real Analysis, Sultan Chand Publications, New Delhi.
3. James Ward Brown, Ruel Vance Churchill (2009) : Complex variables and applications, McGrawHill Higher Education and Applications (2/e),
4. Alan Jeffrey (2006) : Complex Analysis and Applications, Chapman and Hall/CRC Taylor Francis Group.
5. Saminathan Ponnusamy, Herb Silverman (2006) : Complex Variables with Applications, Birkhauser Boston(2006) .

CORE COURSE – PRACTICAL -I

Semester	VI
Course Code	STA6IH01
Course Name	Statistical Computing -I
Hours per Week	04
Credit	04
Examination Hours	Two and half Hours

Objective of the Course

Statistical computing is a practical course. Its objectives are to develop scientific and experimental skills of the students and to correlate the theoretical principles with application based studies.

Course Outcome :

After completing the Course, the Student should be able

CO1 : To develop skills to statistical problems using R.
CO2 : To undertake statistical analysis works of real life problems.
CO3 : To develop ability to correlate theoretical knowledge with applied problems.
CO4 : To develop capacity for working as a statistical data analyst in industries and organizations.

Practical is to be done using R package. At least five statistical data oriented/supported problems should be done from each course. Practical record shall be maintained by each student and the same shall be submitted for verification at the time of external examination.

The practical is based on the following courses they studied from 3rd to 6 semesters.

STA3IB03: Bivariate Data Analysis
STA3IB04: Probability Distributions
STA4IB05: Statistical Inference-I
STA4IB06: Sampling Techniques
STA5IB07 : Statistical Inference – II
STA5IB08 : Design and Analysis of Experiments
STA5IB10: Applied Statistics
STA6IB11 : Statistical Quality Control
STA6IB12: Basic Regression Analysis

The Board of Examiners (BoE) shall decide the pattern of question paper and the duration of the external examination. The external examination shall be conducted and evaluated on the same day jointly by two examiners – one external and one internal, appointed at the centre of the examination by the University on the recommendation of the Chairman, BoE. The question paper for the external examination at the centre will be set by the external examiner in consultation with the Chairman, BoE and the HoDs of the centre. The questions are to be evenly distributed over the entire syllabus. Evaluation shall be done by assessing each candidate on the scientific and experimental skills, the efficiency of the algorithm/program implemented, the presentation and interpretation of the results.

CORE COURSE – PROJECT(FOUNDATION)

Semester	VI
Course Code	STA6IF01
Course Name	Project Work
Hours per Week	02 Hours in V^{th} and VI^{th} Semester
Credit	03

Objectives of the Course:

The project work will helps the students to enhance their Research attitude and also helps in applying the theory of research in real life situations. Students get an exposure to study the working atmosphere of an enterprise or they can undertake research on any socially relevant area based on their various courses.

The following guidelines may be followed for project work.

1. The project is offered in the fifth and sixth semester of the degree course and the duration of the project may spread over the complete year.
2. A project may be undertaken by a group of students, the maximum number in a group shall not exceed 5. However the project report shall be submitted by each student.
3. There shall be a teacher from the department to supervise the project and the synopsis of the project should be approved by that teacher. The head of the department shall arrange teachers for supervision of the project work.
4. As far as possible, topics for the project may be selected from the applied branches of statistics, so that there is enough scope for applying and demonstrating statistical skills learnt in the degree course.

OPEN COURSE

Semester	V
Course Code	STA5ID01
Course Name	Introductory Probability
Hours per Week	03
Credit	03
Examination Hours	Two Hours

Objective of the Course

To know the basic concept of probability and standard probability distributions.

Course Outcome :

After completing the Course, the student should be able

CO1 : To understand different definitions of probability and their simple applications.
CO2 : To understand Baye's theorem and its applications in real life.
CO3 : To understand random variable and its associating probability distributions.
CO4 : To understand various standard probability distributions.

Contents :

Module 1 - (15 Hours)

Probability: Introduction, random experiments, sample space, events and algebra of events. Definitions of Probability – classical, statistical, and axiomatic. Conditional Probability, laws of addition and multiplication, independent events, theorem of total probability, Bayes'theorem and its applications.

Module 2 - (15 Hours)

Random Variables: Discrete and continuous random variables, p.m.f., p.d.f. ,c.d.f. Illustrations of random variables and its properties. Expectation, variance, moments and moment generating function.

Module 3 – (18 Hours)

Standard probability distributions: Binomial, Poisson, geometric, uniform, normal and exponential.

Books for Study

1. Gupta S.C and Kapoor V .K. (2020). Fundamentals of Mathematical Statistics (12 Edn.). Sultan Chand & Co., New Delhi.
2. Saxena, H.C. (1983). Elementary Statistics. S. Chand & Co., New Delhi.

Books for References

1. Hogg, R.V., Tanis, E.A. and Rao J.M. (2009): Probability and Statistical Inference, Seventh Ed, Pearson Education, New Delhi. 2
2. Miller, Irwin and Miller, Marylees (2006): John E. Freund's Mathematical Statistics with Applications, (7th Edn.), Pearson Education, Asia.
3. Gupta S.C (2021). Statistical Methods, Sultan Chand & Co., New Delhi.

OPEN COURSE – 02

Semester	V
Course Code	STA5ID02
Course Name	Basics of Statistical Inference
Hours per Week	03
Credit	03
Examination Hours	Two Hours

Objective of the Course

To know the basic concept of probability and standard probability distributions.

Course Outcome :

After completing the Course, the student should be able

CO1 : To understand different definitions of probability and their simple applications.
CO2 : To understand Baye's theorem and its applications in real life.
CO3 : To understand random variable and its associating probability distributions.
CO4 : To understand various standard probability distributions.

Contents :

Module 1 - (15 Hours)

Estimation of population mean, confidence intervals for the parameters of a normal distribution (one sample and two sample problems). The basic idea of significance test. Null and alternative hypothesis. Type I & Type II errors, level of significance, concept of p-value. Tests of hypotheses for the parameters of a normal distribution (one sample and two sample problems).

Module 2 – (18 Hours)

Categorical data: Tests of proportions, tests of association and goodness-of-fit using Chi-square test, Yates' correction. Tests for the significance of correlation coefficient. Sign test for median, Sign test for symmetry, Wilcoxon two-sample test.

Module 3 – (15 Hours)

Analysis of variance, one-way and two-way classification. Brief exposure of three basic principles of design of experiments, treatment, plot and block. Analysis of completely randomized design, randomized complete block design.

Books for Study

1. Gupta S.C and Kapoor V .K. (2020). Fundamentals of Mathematical Statistics (12 Edn.). Sultan Chand & Co., New Delhi.
2. Saxena, H.C. (1983). Elementary Statistics. S. Chand & Co., New Delhi.
3. Gupta and Kapoor V K (2002). Fundamentals of Applied Statistics, Sultan Chand and Sons.

Books for References

1. Goon, A.M., Gupta M.K. & Das Gupta (2005). Fundamentals of statistics, Vol.-I &). II The World Press Kolkotta.
2. Dass, M. N. & Giri, N. C.(2019) : Design and analysis of experiments. New Age publications, New Delhi.

OPEN COURSE – 03

Semester	V
Course Code	STA5ID03
Course Name	Statistical Techniques for Research Methods
Hours per Week	03
Credit	03
Examination Hours	Two Hours

Objective of the Course

To understand the concept of Research methodology using Statistics.

Course Outcome :

After completing the Course, the student should be able

CO1 : To understand the basis of research.
CO2 : To understand the method of collection, tabulation and analysis of the data
CO3 : To estimate the descriptive measures of Statistics.
CO4 : To understand various types of presentation of data and their inferences.

Contents :

Module 1 - (15 Hours)

Introduction: Meaning, objection and motivation in research, types of research, research approach, significance of research. Research problems: definition, selection and necessity of research problems.

Module 2 – (20 Hours)

Survey Methodology and Data Collection, inference and error in surveys, the target populations, sampling frames and coverage error, methods of data collection, non-response, questions and answers in surveys.
Develop a questionnaire, collect survey data pertaining to a research problem (such as gender discriminations in private v/s government sector, unemployment rates, removal of subsidy, impact on service class v/s unorganized sectors), interpret the results and draw inferences.

Module 3 – (13 Hours)

Processing, Data Analysis and Interpretation: Review of various techniques for data analysis covered in core statistics papers, techniques of interpretation, precaution in interpretation.

Books for Study

1. Kothari, C. R. (2001). Research Methodology-Methods and Techniques, 2nd Ed. Viswa Prakashan, New Delhi.
2. Kumar, R (2011): Research Methodology: A Step - by - Step Guide for Beginners, SAGE publications.

Books for References

1. Gopal, M. H. (2019). An Introduction to Research Procedure in Social Sciences. Asia Publishing House, Mumbai.

SYLLABI OF ADVANCED CORE COURSES

STA7IB15: Mathematical Methods for Statistics – I (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Explain the concept of Reimann – Stieltjes Integral and evaluate the same under different conditions.

CO2: Discuss the concepts of sequence and series of functions, and determine limits of sequences and test the convergence and series of functions.

CO3: Evaluate the Limit and check the continuity of multivariable functions.

CO4: Compute the derivatives, partial and total derivatives and maxima and minima of multivariable functions.

CO5: Solve systems of linear equations, diagonalize matrices and Classify quadratic forms.

CO6: Compute g inverse of matrices.

CO7: Compute algebraic and geometric multiplicity of characteristic roots.

Unit-I. Reimann – Stieltjes Integral- Definition, existence and properties. Integration by parts. Change of variable - Step functions as integrators. Reduction to finite sum. Monotone increasing integrators. Riemann's condition. Integrators of bounded variations. Mean value theorems. Improper integrals.

Unit-II. Sequences and Series of Functions – Point wise convergence and uniform convergence. Tests for uniform convergence. Properties of uniform convergence. Weirstrass theorem.

Unit-III. Matrices- Diagonal reduction. Transformations. Idempotent matrices. Generalized inverse. Solution of linear equations. Special product of matrices. Characteristic roots and vectors. Definition and properties. Algebraic and geometric multiplicity of characteristic roots. Spectral decomposition. Quadratic forms. Classification and reduction of quadratic forms.

Unit-IV. Vector space with real and complex scalars. Subspaces, liner dependence and independence, basis, dimension. Linear transformations and matrices. Jacobean of matrix transformations, functions of matrix argument.

Text Books

1. **Apostol, T.M.** (1974). Mathematical Analysis -Second Edition. Narosa Publishing House, New-Delhi. Chapters 7 & 9.
2. **Khuri, A. T.** (2003). Advanced Calculus with Applications in Statistics. John Wiley & Sons, New York. Chapter 7.
3. **Rao, C.R.** (2002). Linear Statistical Inference & Its Applications- Second Edition. John Wiley & Sons, New York.
4. **Graybill, F. A.** (2001). Matrices with Applications in Statistics. John Wiley & Sons, New York.

References

1. **Malik, S.C. & Arora, S.** (2006). Mathematical Analysis- Second Edition. NewAge International, New Delhi.
2. **Lewis, D. W.** (1995). Matrix Theory. Allied Publishers, Bangalore.

STA7IB16: Mathematical Methods for Statistics –II (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Recollect basic concepts of set theory and explain the concepts of class of sets& set function.
CO2: Discuss measure and different types of measures such as Outer measure, Lebesgue measure, Lebesgue-Steiltjes measure together with their properties.
CO3: Describe Measurable function and related results.
CO4: Discuss the basic concepts of Product space and product measure
CO5: Explain Multiple integral, Absolute continuity and singularity of measures
CO6: Discuss vector space and linear transformation.

Unit-I. Classes of Sets – Field of sets, sigma field, monotone class and minimal sigma field. Borel sigma field and Borel sets in R and R^p. Set functions. Additivity and sigma additivity – Measures - examplesand properties. Outer measure. Lebesgue measure in R and R^p . Lebesgue-Steiltjes measure

Unit-II. Measurable function. Properties. Sequence of measurable functions, convergence, Egoroff's theorem.Integrals of simple, non-negative and arbitrary measurable functions. Convergence of integrals. Monotone convergence theorem, dominated convergence theorem and Fatou's lemma.

Unit III. Product space and product measure .Multiple integral. Fubini's theorem (without proof).Absolute continuity and singularity of measures. Radon-Nikodym theorem (without proof) and its applications.

Unit-IV. Multivariable functions. Limit and continuity of multivariable functions. Derivatives, directional derivatives and continuity. Total derivative in terms of partial derivatives, Taylor's theorem. Inverse and implicit functions. Optima of multivariable functions. Determinants.

Text Books

1. **Royden, H. L. (2010).** Real Analysis- Fourth Edition. Prentice Hall of India, New Delhi.
2. **Bartle, R.G. (2016).** The Elements of Integration. John Wiley & Sons, New York.
3. **Lewis, D.W. (1996).** Matrix Theory. Allied Publishers, Bangalore.
4. **Rao,A.R. and Bhimsankaram, P.** (2000). Linear Algebra.Tata McGraw Hill, NewDelhi.
5. **Rao, C.R. (2002).** Linear Statistical Inference and Its Applications-Second Edition. John Wiley & Sons, NewYork.
6. **Mathai, A. M.** (1999). Linear Algebra Part-IIII : Application of Matrices and Determinants, Lecture Notes -Module 3, Centre for Mathematical Sciences, Trivandrum.

References

1. **Kingman, J.F.C. and Taylor, S.J. (2008).** Introduction to Measure and Probability. Cambridge University Press, UK.
2. **Bapat, R.B. (2012).** Linear Algebra and Liner Models. Hindustan Book Agency, New Delhi.

STA7IB17: Probability Theory – I (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Explain the concepts of probability measure, random variables and decomposition of distribution functions.

CO2: Describe induced probability space and notion of vector valued random variables

CO3: Explain Expectation of simple, non-negative and arbitrary random variables

CO4: Distinguish between different types of inequalities such as Cr-inequality, Basic inequality, etc.

CO5: Illustrate convergence of sequence of random variables with examples

CO6: Discuss independence of events, Borel Cantelli Lemma, Borel and Kolmogrov zero-one criteria.

CO7: Describe properties of Characteristic Function, inversion theorem and its applications.

Unit-I. Probability measure, measure, probability space, random variable. Inverse function and properties. Sequence of random variables and limit. Extension of probability measure - Caratheodory extension theorem (without proof). Distribution function. decomposition of distribution function. Vector valued random variables and its distribution function. Induced probability space of a random variable.

Unit-II. Mathematical expectation of simple, non-negative and arbitrary random variables - properties of expectation. Moment generating functions-moments. Inequalities. Cr- inequality, Jenson's inequality, Basic inequality, Markov inequality.

Unit-III. Different modes of convergence. Convergence in probability, convergence in distribution, r^{th} mean convergence, almost sure convergence and their mutual implications.

Unit-IV. Independence of events, classes of events. Independence of random variables. Kolmogorow's 0-1 law, Borel's 0-1 criteria. Borel-Cantelli Lemma. Characteristic Functions- definition, properties, inversion theorem, inversion formula for lattice distributions, Characteristic functions and moments, Taylor's series for characteristic functions, Bochner's theorem (without proof).

Text Book

1. **Bhat, B.R. (2014).** Modern Probability Theory- Third Edition. New-Age International, New Delhi.

References

1. **Resnick, S.I. (1999).** Probability Paths. Birkhauser, Boston.
2. **Laha, R.G. and Rohatgi, V.K**. (2020). Probability Theory. John Wiley& Sons, New York.
3. **Billingsly, P. (1995).** Probability and Measure- Third Edition. John Wiley & Sons, New York.
4. **Basu, A.K. (1999).** Measure Theory and Probability. Printice Hall of India, New-Delhi.
5. **Rohatgi, V.K**. (2015). An Introduction to Probability Theory and Mathematical Statistics. .John Wiley & Sons , New York.

STA7IB18: Distribution Theory (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Distinguish different discrete distributions and illustrate their role in modeling count data.

CO2: Explain Pearson system and its different members and how they arise from the defining differential equation.

CO3: Describe different positive valued and real valued random variables along with their properties and role in modeling real life data.

CO4: Find marginal and conditional distributions, explain distribution of functions of random vectors , order statistics and their distributions.

CO5: Explain sampling distributions- t, Chi-square and F and their applications.

Unit-I. Review of Discrete and Continuous distributions. Weibull, Pareto, lognormal, Laplace, Cauchy, logistic, Rayleigh distribution their properties and applications.

Unit II. Discrete and continuous bivariate random variables: Definitions, Computation of probabilities of various events, marginal, conditional, product moments and correlations. Conditional expectation and conditional variance. The p. d. f. of a bivariate normal distribution, Marginal and conditional distributions, conditional expectation and conditional variance, regression lines of Y on X and X on Y., independence and uncorrelated-ness imply each other, m. g. f and moments.

Unit-III. Functions of random variables and their distributions using Jacobian of transformation and other tools. Distribution of distribution function. Bivariate exponential distributions. Concept of a sampling distribution. Sampling distributions of t, $\chi 2$ and F (central and non central), their properties and applications

Unit-IV. Compound, truncated and mixture distributions. Convolutions of two distributions. Order statistics: their distributions and properties. Joint, marginal and conditional distribution of order statistics. The distribution of sample range and sample median. Text Books

1. **Rohatgi, V.K. (2015).** An Introduction to Probability Theory and Mathematical Statistics. JohnWiley & Sons, New York. Chapter 4- Sections 2, 4 and 5; Chapter 5- Sections 2, 3 and 4; Chapter 7- Sections 3, 4 and 5.
2. **Johnson, N.L., Kotz. S. and Kemp, A.W. (1992).** Univariate Discrete Distributions. John Wiley & Sons, New York
3. **Johnson, N.L., Kotz. S. and Balakrishnan, N.(2004).** Continuous Univariate Distributions-Vol. I- Second Edition. John Wiley &Sons, New York.
4. **Johnson, N.L., Kotz. S. and Balakrishnan,N**. (1995). Continuous Univariate Distributions-Vol. II- Second Edition. John Wiley & Sons, New York.
5. **Mukhopadhyay P** . (1996). Mathematical Statistics, New Central Book Agency (P) Ltd. Calcutta.

References

1. **Goon, A.M. ,Gupta, M.K. and Das Gupta, B. (2019).** Fundamentals of Statistics- Vol. I and Vol. II (2001). World Press, Calcutta.
2. **David, H. A., &Nagaraja, H. N. (2003).** Order statistics. John Wiley & Sons, Inc.

STA7IB19: Advanced Sampling Theory (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Distinguish between the concepts of probability and non-probability sampling;

CO2: Learn the principles underlying sampling as a means of making inferences about a population and the estimation methods for population mean, total and proportion under various sampling schemes;

CO3: Apply various sampling procedures like SRS, Stratified, systematic, Cluster etc., and estimate the population parameters for attributes and variables;

CO4: Describe the use of auxiliary information for the estimation of population characteristics using ratio, product and regression estimators;

CO5: Employ sampling strategies under varying probability sampling;

CO6: Explain various types of errors in surveys, and procedures to rectify them;

CO7: Analyse data from multi-stage and multiphase surveys;

CO8: Have an appreciation of the practical issues arising in sampling studies.

Unit-I. Review of Probability Sampling – SRSWR & SRSWOR- Stratified sampling – Systematic sampling (linear and circular).

Unit-II. PPS sampling with and without replacement. Estimation of population mean, total and variance in PPS sampling with replacement. Desraj's ordered estimator. Murthy's unordered estimator. Horvitz – Thomson estimator. Their variances and standard error. Yates – Grundy estimator. Sen – Midzuno scheme of sampling. ΠPS sampling.

Unit-III. Ratio estimators and Regression estimators. Comparison with simple arithmetic mean estimator. Optimality properties of ratio and regression estimators. Hartly – Ross unbiased ratio type estimator.

Unit-IV. Cluster sampling with equal and unequal clusters. Multi stage and multiphase sampling . Comparison with simple random sampling and Stratified random sampling. Relative efficiency of cluster sampling. Two-stage sampling. Non-sampling errors.

Text Books

1. **Cochran, W.G. (2017).** Sampling Techniques. Wiley Eastern, New Delhi.
2. **Singh, D. and Chaudhury, F.S. (2020).** Theory and Analysis of Sample Survey Designs.Wiley Eastern, New Delhi.

References

1. **Des Raj (2020).** Sampling Theory. McGraw Hill, New York.
2. **Murthy, M. N. (1967).** Sampling Theory and Methods. Statistical Publishing Society, Calcutta.
3. **Mukhopadhyay, P. (1999).** Theory and Methods of Survey Sampling. Printice Hall India, New Delhi.

STA8IB20: Probability Theory – II (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Describe the role of characteristic functions in the study of weak convergence;

CO2: Explain Helly's Convergence theorem, Helly-Bray lemma, Scheffe's theorem and convergence of moments;

CO3: Distinguish Weak Law of Large Numbers in iid and non-iid set up;

CO4: Illustrate the applications of Weak Laws of large numbers;

CO5: Check whether different types of Central Limit Theorems hold ;

CO6: Define Conditional expectation, martingales and describe the smoothing properties of martingales;

CO7: Define the concept of infinite divisibility, Explain its properties and check whether a given distribution is id or not.

Unit-I. Weak Convergence and Characteristic Functions – Helly's convergence theorem, Helly- Bray lemma, Scheffe's theorem, convergence of distribution functions and characteristic functions, Convergence of moments.

Unit-II. Laws of Large Numbers –Convergence in probability of sequence of partial sums, Kolmogorov inequality and almost sure convergence, almost sure convergence of a series, criterion for almost sure convergence, stability of independent random variables, WLLN(iid and non-iid cases), strong law of large numbers.

Unit-III. Central Limit Thseorem (CLT) – CLT as a generalization of laws of large numbers, Lindeberge-Levy form, Liapounov's form, Lindeberg-Feller form (without proof). Examples and relation between Liapounov's condition.

Unit-IV. Conditioning and Infinite Divisibility: Conditional expectation, properties, Martingales, smoothing properties, Infinite divisibility: Definition, Elementary properties and examples.

Text Books

1. **Bhat. B. R. (2014).** Modern Probability Theory- Third Edition. New Age International (P) Limited, Bangalore.
2. **Laha, R.G. and Rohatgi, V.K. (2020).** Probability Theory. John Wiley & Sons, New York.(Chapter-4, Section-1)

References

1. **Rohatgi, V. K. (2015).** An Introduction to Probability Theory and MathematicalStatistics. John Wiley & Sons, New York.
2. **Feller. W. (1993).** An Introduction to Probability Theory and Its Applications. Wiley Eastern, New Delhi.
3. **Rao, C.R. (2002).** Linear Statistical Inference and Its Applications- Second Edition . John Wiley & Sons, New York.
4. **Basu, A.K. (1999).** Measure Theory and Probability. Prentice Hall of India, New Delhi.

STA8IB21: Estimation Theory (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Explain Sufficiency, Minimal Sufficiency, Unbiasedness & BLUE.
CO2: Describe the ways of obtaining MUVE.
CO3: Check consistency of estimators and how to choose among consistent estimators.
CO4: Apply different methods of estimation such as method of percentiles, method of moments and method of maximum likelihood.
CO5: Describe various loss functions, Risk function and Bayesian estimation under squared error, absolute error and zero-one loss functions
CO6: Derive SELCI. Explain Bayesian and Fudicial intervals

Unit-I. Fisher Information- Sufficient statistic-Minimal sufficient statistic-Exponential family and minimal sufficient statistic. Unbiasedness – best Linear Unbiased estimator – MVUE – Cramer-Rao inequality and its application – Rao-Blackwell theorem-Completeness-Lehman-Scheffe theorem and its application.

Unit-II. Consistent estimator-examples and properties-CAN estimator-invariance property-asymptotic variance- Multiparameter case- choosing between Consistent estimators.

Unit-III. Method of moments-method of percentiles-method of maximum likelihood-MLE in exponential family-Solution of likelihood equations-Bayesian method of estimation-Prior information-Loss functions (squared error absolute error and zero-one loss functions) – Posterior distribution-estimators under the above loss functions.

Unit-IV. Shortest expected length confidence interval-large sample confidence intervals-unbiased confidence intervals-examples-Bayesian and Fiducial intervals.

Text Books

1. **Kale, B.K. (2019).** A First Course on Parametric Inference- Second Edition, Narosa Publishing, New-Delhi.
2. **Casella, G. and Berger, R.L. (2007).** Statistical Inference- Second Edition. Duxbury, Australia.

References

1. **Rohatgi,V. K. (2015).** An Introduction to Probability Theory and Mathematical Statistics. John Wiley & Sons, New York.
2. **Rohatgi,V.K. (2010)**. Statistical Inference. John Wiley & Sons, New York.
3. **Lehman, E.L. (1998).** Theory of Point Estimation. John Wiley & Sons, New York
4. **Rao, C.R. (2002).** Linear Statistical Inference and Its Applications- Second Edition. John Wiley & Sons, New York.

STA8IB22: Advanced Design and Analysis of Experiments (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Explain ANOVA, It's Regression approach, fixed and random effect models and their analysis.
CO2: Check model adequacy.
CO3: Explore designs like RBD, CRD, LSD, Greaco- LSD, BIBD, Youden square, Lattice design and Factorial designs.
CO4: Understanding of ANOCOVA

CO5: Understanding the concept, use and analysis of factorial experiments.

Unit-I. Basic principle of experimental design, overview of RBD, CRD and LSD, Missing plot techniques in RBD with one and two missing observations, Analysis of LSD with one missing observation

Unit-II. General theory of intra block analysis of block design, connectedness and balancing block design, incomplete block design, intra block analysis of BIBD and its properties.

Unit-III. Purpose of analysis of covariance. Practical situations where analysis of covariance is applicable. Model for analysis of covariance in CRD and RBD. Estimation of parameters (derivations are not expected).Preparation of analysis of covariance (ANOCOVA) table, test for $\beta = 0$, test for equality of treatment effects (computational technique only).

Unit-IV. General description of factorial experiments, factorial effects, analysis of factorial experiment (2^n , 3^n), main and interaction effects, advantages and disadvantages, total and partial confounding, split plot experiment.

Text Books

1. **Montgomery, D.C. (2001).** Design and Analysis of Experiments- Fifth Edition. John Wiley & Sons, New York.

References

1. **Das, M. N. and Giri, N. C. (2002).** Design and Analysis of Experiments- Second Edition. New AgeInternational (P) Ltd., New Delhi.
2. **Cochran, W.G. and Cox, G.M.(1992):** Experimental Design, John Wiley and Sons, Inc., New York
3. **Dey Alok (1986):** Theory of Block Design; Wiley Eastern.

STA8IB23: Regression Methods (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Illustrate the concept of linear regression model.
CO2: Estimate regression parameters and explain the properties of estimators.
CO3: Describe the procedure for testing the significance of regression parameters and construct confidence intervals.
CO4: Build various Non parametric Regression models.
CO5: Examine model diagnostics techniques and remedies to overcome violating assumptions.

Unit-I. Least square estimation-properties of least square estimates-unbiased estimation of σ^2 – distribution theory – maximum likelihood estimation – estimation with linear restrictions- design matrix of less that full rank-generalized least squares.

Unit-II. Hypothesis testing; Likelihood ratio test—F-test – multiple correlation coefficient-Confidence intervals and regions. Simultaneous interval estimation- confidence bands for the regression surface – prediction intervals and band for the response.

Unit-III. The straight line – weighted least squares for the straight line- Polynomials in one variable – piecewise polynomial fitting – Polynomial regression in several variables.

Unit-IV. Bias-incorrect variance matrix-effect of outliers-Diagnosis and remedies: residuals and hat matrix diagonals – nonconstant variance and serial Correlations-departures from normality – detecting and dealing with outliers- diagnosing collinearity, Ridge regression and principal component regression.

Text Books

1. **Seber, G. A. F. and Lee, A.J. (2012).** Linear Regression Analysis- Second Edition. John Wiley & Sons, New York.
2. **Draper, N.R. and Smith, H. (1988).** Applied Regression Analysis- Third Edition. John Wiley &Sons, New York.

References

1. **Searle, S.R.** (2014). Linear Models. Wiley Paperback Edition. Wiley Inter Science, New Jersey.
2. **Rao, C.R.**(2002). Linear Statistical Inference and Its Applications. Wiley Eastern, New Delhi.
3. **Abraham, B. and Ledolter, J.** (2005). Introduction to Regression Modeling. Duxbury Press, New York.
4. **Sengupta, D. and Jammalamadaka, S.R.** (2003). Linear Models: An Integrated Approach. World Scientific Press, New Jersey.
5. **Montgomery, D.C., Peck, F.A. and Vining, G**. (2001). Introduction to Linear Regression Analysis- Third Edition. John Wiley & Sons, New York.

STA8IH02: Statistical Computing – II (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Apply the principles of Distribution Theory, Sampling Theory, Estimation, Design & Analysis of Experiments and Regression Methods using real data sets.

CO2: Know the formulas to be applied for the analysis.

CO3: Write the R codes for the analysis of the given data.

CO4: To install and load the packages required to run the R codes.

CO5: Enter the data given for analysis

CO6: Explain how to make conclusions and write the inference for the data analysis based on the output obtained.

The practical is based on the following core papers in the first and the second semesters:

1. STA7IB18 : Distribution Theory
2. STA7IB19 : Advanced Sampling Theory
3. STA8IB21 : Estimation Theory
4. STA8IB22 : Advanced Design and Analysis of Experiments
5. STA8IB23 : Regression Methods

Practical are to be done using scientific programmable calculators or personal computers. The question paper for the external examination will be set by the external examiner in consultation with the chairman. The practical will be valued on the same day the examination is held out and the marks will be finalized on the same day.

STA9IB24 : Testing of Statistical Hypothesis (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Compute MP and UMP tests corresponding to any given testing problem

CO2: Formulate LR test, unbiased tests and similar tests corresponding to any given testing problem.

CO3: Apply different non parametric testing methods.

CO4: Construct SPRT corresponding to any given testing problem.

Unit-I. Tests of hypotheses – error probabilities – Most powerful tests – Neyman. Pearson Lemma – Generalized Neymann – Pearson Lemma.

Unit-II. Method of Finding Tests – Likelihood ratio tests – Bayesian tests – Union – intersection and intersection-union tests. Unbiased and invariant tests – Similar tests and locality most powerful tests.

Unit-III. Non-parametric Tests – Review of Single sample tests. Two sample tests – the chi-square test for homogeneity – Kolmogorov – Smirnov test the median test – Mann-Whitney-Wilcoxon test-Test for independence – Kendall's tau – Spearman's rank correlation coefficient – robustness. Kruskall-Wallis Test – Test for randomness.

Unit-IV. Sequential Inference – Some fundamental ideas of sequential sampling – sequential unbiased estimation – sequential estimation of mean of a normal population – the sequential probability tests (SPRT) – important properties – the fundamental identity of SPRT.

Text Books

1. **Casella, G. and Berger, R.L, (2002).** Statistical Inference -Second Edition. Duxbury, Australia..
2. **Rohatgi, V.K. (2015).** An Introduction to Probability Theory and Mathematical Statistics,. John Wiley & Sons, New York.

References

1. **Fraser, D.A. (1957).** Non – parametric Methods in Statistics. John Wiley & Sons, New York.
2. **Lehman, E.L. (2005).** Testing of Statistical Hypotheses. John Wiley & Sons, New York.
3. **Forguson, T.S**. (2014). Mathematical Statistics: A Decision – Theoretic Approach. Academic Press, New York.

STA9IB25 : Multivariate Analysis (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Describe the properties and applications of multivariate normal distribution.
CO2: Explain partial and multiple correlation coefficients.
CO3: Derive the ML Estimates of the mean vector and the dispersion matrix of multivariate normal.
CO4: Describe the genesis of Wishart distribution with its properties.
CO5: Define Hotelling T^2 and Mahalanobis D^2 statistics and able to apply them in testing problems.
CO6: Classify multivariate normal populations.
CO7: Be familiar with principal components and their analysis.

Unit-I. Multivariate Normal Distribution – Definition properties, conditional distribution, marginal distribution. Independence of a linear form and quadratic form, independence of two quadratic forms, distribution of quadratic form of a multivariate vector. Partial and multiple correlation coefficients, partial regression coefficients, Partial regression coefficient.

Unit-II. Estimation of mean vector and covariance vector – Maximum likelihood estimation of the mean vector and dispersion matrix. The distribution of sample mean vector inference concerning the man vector when the dispersion matrix is known for single and two populations. Wishart distribution – properties – generalized variance..

Unit-III. Testing Problems – Mahalnobis D^2 and Hotelling's T^2 Statistics Likelihood ratio tests – Testing the equality of mean vector, equality of dispersion matrices, testing the independence of sub vectors, sphericity test.

Unit-IV. The problem of classification – classification of one of two multivariate normal population when the parameters are known and unknown. Extension of this to several multivariate normal populations. Population principal components – Summarizing sample variation by principal components – Iterative procedure to calculate sample principal components.

Text Books

1. **Anderson, T.W. (2003).** Multivariate Analysis. John Wiley & Sons, New York.
2. **Rao, C.R.(2002).** Linear Statistical Inference and Its Applications- Second Edition. John Wiley & Sons, New York.

References

1. **Giri, N.C. (2003).** Multivariate Statistical Analysis. Marcel Dekker, New York.
2. **Kshirasagar, A.M. (2006).** Multivariate Analysis. Marcel Dekker**,** New York
3. **Rencher, A.C. (2002).** Multivariate Statistical Analysis. John Wiley & Sons, New York**.**

STA9IB26 : Stochastic Processes (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Explain the concepts of stochastic processes;
CO2: Classify stochastic processes
CO3: Distinguish between strict stationarity and wide-sense stationary .
CO4: Describe the concepts of Discrete time Markov chains, classification of its states and limiting probabilities;
CO5: Describe and use the recurrence relation for generation sizes in a Branching Process and determine the probability of ultimate extinction;
CO6: Explain continuous time Markov chains, Poisson processes and its generalizations;
CO7: Employ birth-death methodology for solving queueing problems;
CO8: Explain the renewal processes and Brownian motion processes.

Unit-I. Concept of Stochastic processes, examples. Specifications. Markov chains- ChapmanKolmogorov equations – classification of states – limiting probabilities Gamblers ruin problem – mean time spent in transient states – branching processes Hidden Markov chains.

Unit-II. Exponential distribution – counting process – inter arrival time and waiting time distributions. Properties of Poisson processes – Conditional distribution of arrival times. Generalization of Poisson processes – non –homogenous Poisson process, compound Poisson process, conditional mixed Poisson process. Continuous time Markov Chains – Birth and death processes – transition probability function-limiting probabilities.

Unit-III. Renewal processes-limit theorems and their applications. Renewal reward process. Regenerative processes, semi-Markov process. The inspection paradox Insurers ruin problem.

Unit-IV. Basic characteristics of queues – Markovian models – network of queues. The M/G/I system. The G/M/I model, Multi server queues. Brownian motion Process – hitting time – Maximum variable – variations on Brownian motion – Pricing stock options – Gaussian processes – stationary and weakly stationary processes.

Text Books

1. **Ross, S.M.(2014).** Introduction to Probability Models- Eleventh Edition. Academic Press, New York.

References

1. **Medhi,J. (2019).** Stochastic Processes- Second Edition. Wiley Eastern, New Delhi**.**
2. **Karlin, S. and Taylor, H.M.** (1975). A First Course in Stochastic Processes- Second Edition. AcademicPress, New York.
3. **Cinlar, E.** (2013). Introduction to Stochastic Processes. Prentice Hall, New Jersey.
4. **Basu, A.K**.(2003). Introduction to Stochastic Processes. Narosa Publishing House, New Delhi.

STA10IH03 : Statistical Computing - III (Four Credits)

Course Outcomes:

On successful completion of the course, students should be able to:

CO1: Apply the principles of Testing of Hypotheses, Multivariate Analysis and the two electives offered in Semester X using real data sets.

CO2: Know the formulas to be applied for the analysis.

CO3: Write the R codes for the analysis of the given data.

CO4: To install and load the packages required to run the R codes.

CO5: Enter the data given for analysis

CO6: Explain how to make conclusions and write the inference for the data analysis based on the output obtained.

The practical is based on the following courses in the third and fourth semesters**.**

1. STA9IB24 : Testing of Statistical Hypothesis
2. STA9IB25 : Multivariate Analysis
3. Elective – III
4. Elective – IV

Practical is to be done using scientific programmable calculators or personal computer. The question paper for the external examination will be set by the external examiner in consultation with the chairman.. The practical will be valued on the same day the examination is carried outand the mark sheet will be given to the chairman on the same day.

STA10IF02 : Project (Advanced) (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Discuss the applications of various statistical techniques learned in the entire course in the form of project work.

CO2: Manage a real practical situation where a statistical analysis is sought.

CO3: Develop professional approach towards writing and presenting an academic report.

CO4: Get more insight about the opportunities in research/career.

CO5: Know the works presented in various journals and current trends in their project/dissertation area.

CO6: Get an idea of how new developments in the topic have arose and why new computational techniques are needed.

As a part of the course work, during the fourth semester each student has to undertake a project work in a selected area of interest under a supervisor in the department. The topic could be a theoretical work or data analysis type. At the end of the tenth semester the student shall prepare a report/dissertation which summarizes the project work and submit to the H/D of the parent department positively before the deadline suggested in the Academic calendar. The project/ dissertation is of 4 credits for which the following evaluation will be followed:

The valuation shall be jointly done by the supervisor of the project in the department and an External Expert appointed by the University, based on a well defined scheme of valuation framed by them. The following break up of weightage is suggested for its valuation.

2 Review of literature, formulation of the problem and defining clearly the objective: 20%
3 Methodology and description of the techniques used: 20%
4 Analysis, programming/simulation and discussion of results: 20%
5 Presentation of the report, organization, linguistic style, reference etc.: 20%
6 Viva-voce examination based on project/dissertation: 20%.

STA10IG01 : Comprehensive Viva (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Communicate the concepts of each course precisely

CO2: Communicate the importance and applications of the subject Statistics in a broad sense

CO3: Get more insights into the subject areas.

CO4: Face interviews without fear and communicate their ideas effectively.

There shall be a comprehensive Viva Voce examination based on all courses of the programme with 4 credits, internal and external being in the ratio 1:4. The Viva-Voce shall be conducted by a board of examiners consisting of at least on external expert and internal examiners.

ELECTIVE COURSES

E01: Time Series Analysis (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe the basics of time series data, its auto-covariance, auto-correlation and stationarity.

CO2: Illustrate test for trend and seasonality.

CO3: Explain the smoothing methods for determining trend of the data.

CO4: Describe the properties of linear time series models.

CO5: Fit linear models for time series data sets.

CO6: Describe the maximum likelihood, Yule-Walker and least square estimation methods.

C07: Learn to validate a model using residual analysis.

CO8: Define ARCH and GARCH models and derive their properties.

CO9: Analyse spectral density and periodogram.

Unit-I- Motivation, Time series as a discrete parameter stochastic process, Auto – Covariance, Auto-Correlation and spectral density and their properties. Exploratory time series analysis, Test for trend and seasonality, Exponential and moving average smoothing, Holt – Winter smoothing, forecasting based on smoothing, Adaptive smoothing.

Unit-II- Detailed study of the stationary process: Autoregressive, Moving Average, Autoregressive Moving Average and Autoregressive Integrated Moving Average Models. Choice of AR / MA periods.

Unit-III- Estimation of ARMA models: Yule – Walker estimation for AR Processes, Maximum likelihood and least squares estimation for ARMA Processes, Discussion (without proof) of estimation of mean, Auto-covariance and auto-correlation function under large samples theory, Residual analysis and diagnostic checking. Forecasting using ARIMA models, Use of computer packages like SPSS.

Unit-IV- Spectral analysis of weakly stationary process. Herglotzic Theorem. Periodogram and correlogram analysis. Introduction to non-linear time Series: ARCH and GARCH models.

Text Books

1. **Box, G.E.P and Jenkins, G.M. (1970).** Time Series Analysis, Forecasting and Control. Holden Day, San fransisco.
2. **Brockwell, P.J. and Davis, R.A. (1987).** Time Series: Theory and Methods. Springer Verlag. New York.
3. **Abraham, B and Ledolter, J.C. (1983).** Statistical Methods for Forecasting. John Wiley & Sons, New York.

References

1. **Anderson ,T.W. (1971).** Statistical Analysis of Time Series. John Wiley & Sons, New York.
2. **Fuller, W.A. (1978).** Introduction to Statistical Time Series. John Wiley & Sons, New York..
3. **Kendall ,M.G. (1978).** Time Series. Charles Griffin, London.
4. **Tanaka, K. (1996).** Time Series Analysis. Wiley Series, New York.

E02: Advanced Operations Research (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Discuss the Non linear programming problems and methods to solve the problems.

CO2: Understand and solve quadratic programming problem.

CO3: Explain Dynamic and Geometric programming.

CO4: Discuss inventory management, deterministic and probability models.

CO5: Understand Replacement models.

CO6: Understand simulation modeling and random number generation

Unit-I-Non-linear programming, Lagrangian function, saddle point, Kuhn-Tucker Theorem, Kuhn- Tucker conditions, Quadratic programming, Wolfe's algorithm for solving quadratic programming problem

Unit-II- Dynamic and Geometric programming: A minimum path problem, single additive constraint, additively separable return; single multiplicative constraint, additively separable return; single additive constraint, multiplicatively separable return, computational economy in DP. Concept and examples of Geometric programming.

Unit-III-Inventory management; Deterministic models, the classical economic order quantity, nonzero lead time, the EOQ with shortages allowed, the production lot-size model. Probabilistic models. the newsboy problem, a lot size. reorder point model.

Unit-IV-Replacement models; capital equipment that deteriorates with time, Items that fail completely, mortality theorem, staffing problems, block and age replacement policies. Simulation modeling: Monte Carlo simulation, sampling from probability distributions. Inverse method, convolution method, acceptance-rejection methods, generation of random numbers, Mechanics of discrete simulation.

Text Books

1. **Mital, K.V. and Mohan, C. (1996).** Optimization Methods in Operations Research and Systems Analysis-Third Edition. New Age International (Pvt.) Ltd., New Delhi.

2. **Sasieni, M., Yaspan, A. and Friendman, L.(1959).** Operations Research- Methods and Problems. John Wiley & Sons, New York.

3. **Taha. H.A. (2007).** Operations Research -An Introduction-Eighth Edn. Pearson Printice Hall, New Jersey.

4. **Ravindran, A., Philips, D.T. and Solberg, J.J. (2006).** Operations Research- Principles and Practice-Second Edition. John Wiley & Sons, New York.

References

1. **Sharma, J.K. (2003).** Operations Research-Theory & Applications, Macmillan India Ltd., New Delhi.

2. **Man Mohan, Kanti swarup and Gupta(1999).** Operation Research, Sultan Chand & Sons, New Delhi.

E03: Queueing Theory (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1:Discuss the basic characteristic features of a queuing system and acquire skills in analyzing queuing models.

CO2:Explain the theory of Markovian queueing models, M/M/I, M/M/∞ and M/M/c queueing systems.

CO3: Define and explain basic concepts in the theory ofAdvanced Markovian models, $M^{[X]}$ /M/1, M/$M^{[Y]}$ /1, M/E_k/1 , E_k/M/1 and Erlangian models.

CO4: Analyze a network of queues with Jackson network.

CO5:Apply models with general arrival pattern and M/G/1 queueing model.

CO6: Discuss and apply queuing models in real life problem.

Unit-I- Introduction to queueing theory, Characteristics of queueing processes, Measures of effectiveness, Markovian queueing models, steady state solutions of the M/M/I model, waiting time distributions, Little's formula, queues with unlimited service, finite source queues.

Unit-II- Transient behavior of M/M/1 queues, transient behavior of M/M/∞. Busy period analysis for M/M/1and M/M/c models. Advanced Markovian models. Bulk input $M^{[X]}$ /M/1 model, Bulk service M/$M^{[Y]}$ /1 model, Erlangian models, M/E_k/1andE_k/M/1. A brief discussion of priority queues.

Unit-III- Queueing networks-series queues, open Jackson networks, closed Jackson network, Cyclic queues, Extension of Jackson networks. Non Jackson networks.

Unit-IV- Models with general arrival pattern, The M/G/1 queueing model, The Pollaczek-Khintchine formula, Departure point steady state systems size probabilities, ergodic theory, Special cases M/E_k/1 and M/D/1, waiting times, busy period analysis, general input and exponential service models, arrival point steady state system size probabilities.

Text Books/References

1. **Gross, D. and Harris, C.M.(1985).** Fundamentals of Queueing Theory- Second Edition. John Wiley and Sons, New York.
2. **Kleinrock, L. (1975).** Queueing Systems, Vol. I & Vol 2. John Wiley and Sons, New York.
3. **Ross, S.M. (2007).** Introduction to Probability Models- Nineth Edition. Academic Press, New York.
4. **Bose, S.K. (2002).** An Introduction to Queueing System., Kluwer Academic/Plenum Publishers, New York.

E04: Lifetime Data Analysis (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Discuss life time distributions and important parametric models.

CO2: Explain censoring and estimation of parameters using censored data.

CO3: Estimate the survival probabilities and cumulative hazard function using product limit and Nelson-Aalen estimate respectively.

CO4: Describe inference under exponential model and discuss the comparison of distributions.

CO5: Explain important hazard models and apply Rank test, Log-rank test and Generalized Wilcox on test

CO6: Describe the concept of regression models and Cox PH model & AFT model.

CO7: Discuss multivariate lifetime models.

Unit-I- Lifetime distributions-continuous and discrete models-important parametric models: Exponential Weibull, Log-normal, Log-logistic, Gamma, Inverse Gaussian distributions, Log location scale models and mixture models. Censoring and statistical methods

Unit-II- The product-limit estimator and its properties. The Nelson-Aalen estimator, interval estimation of survival probabilities, asymptotic properties of estimators, descriptive and diagnostic plots, estimation of hazard function, methods for truncated and interval censored data, Life tables.

Unit-III- Inference under exponential model – large sample theory, type-2 censored test plans, comparison of two distributions; inference procedures for Gamma distribution; models with threshold parameters, inference for log-location scale distribution: likelihood based methods: Exact methods under type-2 censoring; application to Weibull and extreme value distributions, comparison of distributions.

Unit-IV- Log-location scale (Accelerated Failure time) model, Proportional hazard models, Methods for continuous multiplicative hazard models, Semi-parametric maximum likelihood-estimation of continuous observations, Incomplete data; Rank test for comparing Distributions, Log-rank test, Generalized Wilcox on test. A brief discussion on multivariate lifetime models and data.

Text Books

1. **Lawless, J.F.(2003).** Statistical Methods for Lifetime-Second Edition. John Wiley & Sons, New York.

2. **Kalbfiesche, J.D. and Prentice, R.L. (1980).** The Statistical Analysis of Failure Time Data. John Wiley & Sons, New York.

References

1. **Miller, R.G.(1981).** Survival Analysis, John Wiley & Sons, New York.
2. **Bain, L.G.(1978).** Statistical Analysis of Reliability and Life testing Models. Marcel Decker, New York.
3. **Nelson, W.(1982).** Applied Life Data Analysis. Wiley Series in Probability and Statistics, New York.
4. **Cox, D.R. and Oakes, D.(1984).** Analysis of Survival Data. Chapman and Hall, New York.
5. **Lee, E. T. (1992).** Statistical Methods for Survival Data Analysis. John Wiley & Sons, New York.

E05: Advanced Distribution Theory (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Generate generalized distribution model's by combining any two discrete distributions by the wayof stopped sum.

CO2: Familiarize bivariate discrete distributions / family of distributions, their statistical properties and generation procedure,

CO3: Apply bivariate discrete distributions for modeling bivariate discrete data sets.

CO4: Describe bivariate and multivariate continuous distributions, Derive its marginal's, conditionals and sub cases and thereby able to apply these distributions for modeling bivariate/ multivariate continuousdata sets.

CO5: Develop various record value distributions and analyze their statistical properties and through which, realize the importance of sole existence of extreme observations

Unit-I- Stopped sum distributions: Poisson stopped sum, Neyman type A, Poission-binomial, Poisson-negative binomial, Legrangian Poisson distributions, Distributions of order Poisson, negative binomial, Logarithmic series, Binomial.

Unit-II-Bivariate discrete distributions: bivariate power series distributions, bivariate Poisson, negative binomial and logarithmic series distributions, properties of these distributions, bivariate hypergeometric distribution and its properties.

Unit-III- Bivariate continuous models, bivariate Pearson system, Farlie Morgenstern distribution; distributions with specified conditionals, bivariate Pareto of I, II, III and IV kind, multivariate Liouville distributions.

Unit-IV- Record values - definition, properties, distribution of nth record, record values from exponential, Weibull and logistic; Moments relationships, characterizations.

References

1. **Johnson, N.L., Kotz, S. and Kemp, A.W. (1992).** Univariate Discrete Distributions-Second Edition. John Wiley & Sons, New York.
2. **Kocherlakota, S. and Kocharlakota, K. (1992).** Bivariate Discrete Distributions. Marcel-Dekker, New York.
3. **Johnson, N.L., Kotz, S. and Balakrishnan, N. (1997).** Discrete Multivariate Distributions- Second Edition. John Wiley & Sons, New York.
4. **Kotz, S. ,Balakrishnan, N. and Johnson, N.L. (2000).** Continuous Multivariate Distributions- Vol. I. John Wiley & Sons, New York.
5. **Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1998).** Records. John Wiley & Sons, New York.

E06: Statistical Decision Theory (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Discuss the need and role of statistical decision theory to solve real decision problems

CO2: Explain various decision principles and associated loss functions.

CO3: Differentiate between frequentist and Bayesian decision theory.

CO4: Describe the prior and posterior distributions.

CO5: Handle prior selection problem

CO6: Describe the development of optimal strategies for actions in competitive situations involving two or more intelligent antagonists.

CO7: Explain the general techniques for solving games using game theory.

Unit-I-Statistical decision Problem – Decision rule and loss-randomized decision rule. Decision Principle – sufficient statistic and convexity. Utility and loss-loss functions- standard loss functions, vector valued loss functions.

Unit-II- Prior information-subjective determination of prior density-Non-informative priors- maximum entropy priors, the marginal distribution to determine the prior-the ML-II approach to prior selection. Conjugate priors.

Unit-III-The posterior distribution-Bayesian inference-Bayesian decision theory-empirical Bayes analysis – Hierarchical Bayes analysis-Bayesian robustness Admissibility of Bayes rules.

Unit-IV-Game theory – basic concepts – general techniques for solving games Games with finite state of nature-the supporting and separating hyper plane theorems. The minimax theorem. Statistical games.

Text Book

1. **Berger, O,J.(1985).** Statistical Decision Theory and Bayesian Analysis- Second Edition. Springer, New York.

References

1. **Ferguson, T.S. (1967).** Mathematical Statistics- A Decision Theoretic Approach. Academic Press, New York.
2. **Lehman, E.L.(1983).** Theory of Point Estimation. John Wiley & Sons, New York.
3. **Parmigiani,G., Inoue, L.Y.T. and Lopes, H.F. (2009).** Decision Theory- Principles and Approaches, John Wiley & Sons, New York.

E07: Reliability Modeling (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe the basic concepts of system reliability

CO2: Discuss the lifetime of a system based on ageing properties

CO4: Explain shock models and stress-strength models using reliability theory.

CO5: Describe Maintenance and replacement policies

CO6: Explain reliability growth models.

Unit-I-Reliability concepts and measures; components and systems; coherent systems; reliability of coherent systems; cuts and paths; modular decomposition; bounds on system reliability; structural and reliability importance of components.

Unit-II- Life distributions; reliability function; hazard rate; common life distributions- exponential, Weibull, Gamma etc. Estimation of parameters and tests in these models.Notions of ageing; IFR, IFRA, NBU, DMRL, and NBUE Classes and their duals; closures or these classes under formation of coherent systems, convolutions and mixtures.

Unit-III-Univariate shock models and life distributions arising out of them; bivariate shock models; common bivariate exponential distributions and their properties. Reliability estimation based on failure times in variously censored life tests and in tests with replacement of failed items; stress-strength reliability and its estimation.

Unit-IV-Maintenance and replacement policies; availability of repairable systems; modelling of a repairable system by a non-homogeneous Poisson process. Reliability growth models; probability plotting techniques; Hollander- Proschan and Deshpande tests for exponentiality; tests for HPP vs. NHPP with repairable systems. Basic ideas of accelerated life testing.

Text Books / References

1. **Barlow R.E. and Proschan F.(1985).** Statistical Theory of Reliability and Life Testing. Holt, Rinehart and Winston
2. **Bain L.J. and Engelhardt (1991).** Statistical Analysis of Reliability and Life Testing Models. Marcel Dekker, New York.
3. **Aven, T. and Jensen,U. (1999).** Stochastic Models in Reliability, Springer, New York.
4. **Lawless, J.F. (2003).** Statistical Models and Methods for Lifetime -Second Edition. John Wiley & Sons., New York.
5. **Nelson, W (1982).** Applied Life Data analysis. John Wiley & Sons, New York.
6. **Zacks, S. (1992).** Introduction to Reliability Analysis: Probability Models and Statistics Methods. Springer, New York.

E08: Actuarial Statistics (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Apply the elements of interest.

CO2: Discuss regular pattern of cash flows and related topics.

CO3: Illustrate and apply individual and collective risk models for a short period.

CO4: Discuss survival distributions and derive survival functions.

CO5: Explain and apply life insurance models.

CO6: Discuss and apply annuity models.

Unit I- Elements of the Theory of Interest -Compound interest - Nominal rate - Discount andannuities - Accumulated value - Effective and nominal discount rates. Cash flows - An analogy with currencies - Discount functions - Calculating the discount function - Interest and discount rates - Constant interest - Values and actuarial equivalence – Regular pattern cash flows -Balances and reserves -Time shifting and the splitting identity - Change of discount function - Internal rates of return - Forward prices and term structure – Economics of Insurance – Utility – Insurance and Utility.

Unit II-An Individual Risk Model for a Short Period: The distribution of individual payment – The aggregate payment (convolutions) – Premiums and solvency – Some general premium principles. A Collective Risk Model for a Short Period: The distribution of aggregate claim (Single homogeneous and several homogeneous groups) – Premiums and solvency.

Unit III-Survival Distributions - Survival functions and force of mortality - The time-until-death for a person of a given age - Curate-future-lifetime- Survivorship groups- Life tables and interpolation- Analytical laws of mortality - A Multiple Decrement Model - Multiple Life Models.

Unit IV- Life Insurance Models: The present value of a future payment- The present value of payments for a portfolio of many policies – Whole life insurance - Deferred whole life insurance - Term insurance – Endowments - Varying Benefits - Multiple Decrement and Multiple Life Models. Annuity Models: Continuous and discrete annuities - Level Annuities (certain and random annuities)- whole life annuities – Temporary annuities - Deferred annuities - Certain and life annuities - Varying Payments – annuities with monthly payments - Multiple Decrement and Multiple Life Models – Premiums and reserves.

Text books:

1. **Rotar, V.I. (2015).** Actuarial Models – The mathematics of Insurance – Second Edition. CRC Press, New York.
2. **Promislow, S.D. (2015).** Fundamentals of Actuarial Mathematics- Third Edition. John Wiley & Sons, New York.
3. **Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A.& Nesbitt**, **C.J. (1997).** Actuarial Mathematics, Society of Actuaries.

E09: Official Statistics (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe Indian and International Statistical systems, their role, functions and activities;

CO2: Discuss the scope and contents of population census of India;

CO3: Explain the population growth in developed and developing countries and evaluate the performance of family welfare programs;

CO4: Identify Statistics related to industries, foreign trade, balance of payment, cost of living inflation, educational and social statistics;

CO5: Illustrate economic development and national income estimation using product approach, income approach and expenditure approach;

CO6: Discuss the measures of inequality in income and measures of incidence and intensity.

Unit I: Introduction to Indian and International Statistical systems. Role, function and activities of Central and State Statistical organizations. Organization of large-scale sample surveys. Role of National Sample Survey Organization. General and special data dissemination systems. Scope and Contents of population census of India.

Unit II: Population growth in developed and developing countries, Evaluation of performance of family welfare programmes, projections of labour force and man power. Statistics related to Industries, foreign trade, balance of payment, cost of living, inflation, educational and other social statistics.

Unit III: Economic development: Growth in per capita income and distributive justice indices of development, human development index. National income estimation- Product approach, income approach and expenditure approach.

Unit IV: Measuring inequality in incomes: Gini Coefficient, Theil's measure; Poverty measurements: Different issues, measures of incidence and intensity; Combined Measures: Indices due to Kakwani, Senetc.

Suggested Readings:

1. Basic Statistics Relating to Indian Economy (CSO) 1990
2. Guide to Official Statistics (CSO) 1999
3. Statistical System in India (CSO) 1995
4. Principles and Accommodation of National Population Census, UNEDCO.
5. **Panse, V.G:** Estimation of Crop Yields (FAO)
6. Family Welfare Year Book. Annual Publication of D/O Family Welfare.
7. Monthly Statistics of Foreign Trade in India, DGCIS, Calcutta and other Govt. Publications.
8. **CSO (1989)a:** National Accounts Statistics- Sources and Methods.
9. **Keyfitz, N (1977):** Applied Mathematical Demography- Springer Verlag.
10. **Sen, A(1977):** Poverty and Inequality.

11. **UNESCO:** Principles for Vital Statistics Systems, Series M-12.
12. **CSO (1989)b:** Statistical System in India
13. **Chubey, P.K (1995):** Poverty Measurement, New Age International.

E10: Biostatistics (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Discuss types of Biological data and Principles of Bio Statistical design of medical studies.

CO2: Explain the concepts of survival time functions of important parametric models and compare two survival distributions using LR test and Cox's F-test.

CO3: Explain censoring and estimation of parameters using censored data.

CO4: Describe competing risk theory and estimate the probabilities of death by ML method.

CO6: Discuss the Basic biological concepts in genetics and clinical trials.

Unit-I: Biostatistics-Example on statistical problems in Biomedical Research-Types of Biological data-Principles of Biostatistical design of medical studies- Functions of survival time, survival distributions and their applications viz. exponential, gamma, Weibull, Rayleigh, lognormal, distribution having bath-tub shape hazard function. Parametric methods for comparing two survival distributions (L.R test and Cox's F- test).

Unit-II: Type I, Type II and progressive or random censoring with biological examples, Estimation of mean survival time and variance of the estimator for type I and type II censored data with numerical examples. Non-parametric methods for estimating survival function and variance of the estimator viz. Acturial and Kaplan – Meier methods.

Unit-III: Categorical data analysis (logistic regression) - Competing risk theory, Indices for measurement of probability of death under competing risks and their inter-relations. Estimation of probabilities of death under competing risks by ML method. Stochastic epidemic models: Simple and general epidemic models.

Unit-IV: Basic biological concepts in genetics, Mendel's law, Hardy- Weinberg equilibrium, random mating, natural selection, mutation, genetic drift, detection and estimation of linkage in heredity. Planning and design of clinical trials, Phase I, II, and III trials. Sample size determination in fixed sample designs. Planning of sequential, randomized clinical trials, designs for comparative trials; randomization techniques and associated distribution theory and permutation tests (basic ideas only); ethics behind randomized studies involving human subjects; randomized dose-response studies(concept only).

Text Books / References

1. **Biswas, S. (1995).** Applied Stochastic Processes. A Biostatistical and Population Oriented Approach. Wiley Eastern Ltd., New Delhi.
2. **Cox, D.R.** and **Oakes, D. (1984).** Analysis of Survival Data. Chapman & Hall, New York.
3. **Elandt, R.C.** and **Johnson (1975).** Probability Models and Statistical Methods in Genetics. John Wiley & Sons, New York.

4. **Ewens, W. J.** and **Grant, G.R. (2001).** Statistical methods in Bioinformatics: An Introduction. Springer, New York.
5. **Friedman, L.M., Furburg, C.** and **DeMets, D.L. (1998).** Fundamentals of Clinical Trials. Springer, New York.
6. **Gross, A. J.** and **Clark V.A. (1975).** Survival Distribution; Reliability Applications in Biomedical Sciences. John Wiley & Sons, New York.
7. **Lee, E. T. (1992).** Statistical Methods for Survival Data Analysis. John Wiley & Sons, New York.
8. **Li, C.C. (1976).** First Course of Population Genetics. Boxwood Press, California.
9. **Daniel, W.W.(2006).** Biostatistics: A Foundation for Analysis in the Health Sciences. John Wiley & Sons, New York.
10. **Fisher, L.D.** and **Belle, G.V. (1993).** Biostatistics: A Methodology for the Health Science. John Wiley & Sons, New York.
11. **Lawless, J.F.(2003).** Statistical Methods for Lifetime - Second Edition. John Wiley & Sons, New York.
12. **Chow, S.C. and Chang, M. (2006).** Adaptive Design Methods in Clinical Trials. Chapman & Hall/CRC Biostatistics Series, New York.
13. **Chang, M. (2007).** Adaptive Design Theory and Implementation Using SAS and R. Chapman & Hall/CRC Biostatistics Series, New York.
14. **Cox, D.R. and Snell, E.J. (1989).** Analysis of Binary Data- Second Edition. Chapman &Hall / CRC Press, New York.
15. **Hu, F. and Rosenberger, W.F. (2006).** The Theory of Response-Adaptive Randomization in Clinical Trials. John Wiley & Sons, New York.
16. **Rosenberger, W.F. and Lachin, J. (2002).** Randomization in ClinicalTrials: Theory and Practice. John Wiley & Sons, New York.

E11: Econometric Models (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Explain the meaning and methodology of econometrics.

CO2: Discuss the Leontief input output models and explain the optimization problems in Economics.

CO3: Explain the optimization problems with equality constraints and discuss various production functions like Cobb-Douglas production function and CES production function.

CO4: Discuss the Domar growth model, Solow growth model and Cobweb model.

CO5: Explain the meaning of Multi collinearity, Heteroscedasticity, Autocorrelation and discuss various dynamic econometric models.

CO6: Describe the Simultaneous equation models and Approaches to econometric forecasting..

Unit-I: Basic economic concepts: Demand, revenue, average revenue, marginal revenue, elasticity of demand, cost function, average cost, marginal cost. Equilibrium analysis: Partial market equilibrium- linear and nonlinear model, general market equilibrium, equilibrium in national income analysis. Leontief input output models. Optimization problems in economics, Optimization problems with more than one choice variable:multi product firm, price discrimination.

Unit-II: Optimization problems with equality constraints: utility maximization and consumer demand, homogeneous functions, Cobb-Douglas production function, least cost combination of inputs, elasticity of substitution, CES production function. Dynamic analysis: Domar growth model, Solow growth model, Cobweb model.

Unit-III: Meaning and methodology of econometrics, regression function, multiple regression model, assumptions, OLS and ML estimation, hypothesis testing, confidence interval and prediction. Multicollinearity, Heteroscedasticity, Autocorrelation: their nature, consequences, detection, remedial measures and estimation in the presence of them. Dynamic econometric models: Auto regressive and distributed lag- models, estimation of distributed lag- models, Koyck approach to distributed lag- models, adaptive expectation model, stock adjustment or partial adjustment model, estimation of auto regressive models, method of instrumental variables,detecting autocorrelation in auto regressive models: Durbin- h test, polynomial distributed lag model.**Unit-IV:** Simultaneous equation models: examples, inconsistency of OLS estimators, identification problem, rules for identification, method of indirect least squares, method of two stage least squares. Time series econometrics: Some basic concepts, stochastic processes, unit root stochastic processes, trend stationary and difference stationary stochastic processes, integrated stochastic processes, tests of stationarity, unit root test, transforming non-stationary time series, cointegration. Approaches to economic forecasting, AR, MA, ARMA and ARIMA modeling of time series data, the Box- Jenkins methodology.

Text Books

1) Chiang, A.C. (1984). Fundamental Methods of Mathematical Economics –Third Edition. McGraw Hill, New York.

2). Gujarati, D.N. (2007). Basic Econometrics -Fourth Edition. McGraw-Hill, New York.

References

1. **Johnston, J. (1984).** Econometric Methods -Third Edition. McGraw–Hill, New York.

2.**Koutsoyiannis, A (1973).** Theory of Econometrics, Harper & Row, New York.

3.Maddala ,G.S. (2001). Introduction to Econometrics - Third Edition. John Wiley & Sons, New York.

4.Yamane, T. (1968). Mathematics for Economists- An Elementary Survey- Second Edition. Prentice-Hall,India.

E12: Demographic Techniques (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe the sources of demographic Statistics and explain basic demographic measures.

CO2: Construct a life table.

CO3: Explain the measures of fertility.

CO4: Find the point estimates and population projections based on mortality, fertility and migration basis.

CO5: Discuss the ageing of the population

CO6: Estimate the demographic measures from incomplete data

Unit-I- Sources of demographic Statistics, Basic demographic measures: Ratios, Proportions and percentages, Population Pyramids, Sex ratio Crude rates, Specific rates, Labour force participation rates, Density of population, Probability of dying.

Unit-II- Life tables: Construction of a life table, Graphs of lx, qx, dx, Functions Lx, Tx and Ex. Abridged life tables Mortality: Rates and Ratios, Infant mortality, Maternal mortality, Expected number of deaths, Direct and Indirect Standardization, Compound analysis, Morbidity.

Unit-III- Fertility: Measures of Fertility, Reproductivity formulae, Rates of natural increase, Fertility Schedules, Differential fertility, Stable populations, Calculation of the age distribution of a stable population, Model Stable Populations.

Unit-IV- Population estimates, Population Projections: Component method, Mortality basis for projections, Fertility basis for projections, Migration basis for projections, Ageing of the population, Estimation of demographic measures from incomplete data.

Text book

1. **Pollard, A.H., Yusuf, F. and Pollard, G.N (1990).** Demographic Techniques. Pergamon Press, New York. Chapters 1-8, 12.

References

1. **Keyfitz, N. (1977)**. Applied Mathematical Demography. Wiley-Inter Science Publication.
2. **Keyfilz, N. (1968).** Introduction to the Mathematic of Population. Addition-Wesley, London.
3. **Keyfilz, N. and Caswell, H. (2005).** Applied Mathematical Demography- Third Edition. Springer, New York.

E13: STOCHASTIC FINANCE (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1.Expain basic concepts of financial markets and market lines.

CO2. Learn the usage of Statistical models in modeling Financial data.

CO3. Interpret and apply the black Scholes theorem and its properties.

CO4. Describe the pricing of European and American options by Monte-Carlo and finite difference methods.

CO5. Discuss on the modeling security market and price process models

CO6. Describe the special features of the financial time series, their models and its estimation.

Unit-I: Basic concepts of financial markets. Forward contracts, futures contracts, options-call and put options, European option and American options. Hedgers, speculators, arbitrageurs. Interest rates, compounding, present value analysis, risk free interest rates. Returns, gross returns and log returns. Portfolio theory – trading off expected return and risk, one risky asset and one risk free asset. Two risky assets, estimated expected return Optimal mix of portfolio CAPM, capital market line, betas and security market line.

Unit-II: Options, pricing via arbitrage, law of one price. Risk neutral valuation. Binomial model- single and multi period binomial model, martingale measure. Modeling returns: lognormal model, random walk model, geometric Brownian motion process. Ito lemma (without proof). Arbitrage theorem. The Black-Scholes formula. Properties of the Black-Scholes option cost, the delta hedging arbitrage strategy. Some derivatives, their interpretations and applications.

Unit-III: Volatility and estimating the volatility parameter. Implied volatility. Pricing American options. Pricing of a European option using Monte-Carlo and pricing an American option using finite difference methods. Call options on dividend-paying securities. Pricing American put options, Modeling the prices by adding jumps to geometric Brownian motion. Valuing investments by expected utility. Modeling security market: Self-financing portfolio and no arbitrage, price process models, division rule, product rule.

Unit-IV: Financial Time Series – Special features of financial series, Linear time series models: AR(1),AR(p), ARMA(p,q) processes, the first and second order moments, estimation and forecasting methods. Models for Conditional heteroscedasticity: ARCH(1), ARCH(p), GARCH(p,q) models and their estimation. Comparison of ARMA and GARCH processes.

References

1. **Ross, S.M. (2003).** An Elementary Introduction to Mathematical Finance- Second Edition. Cambridge University Press, UK.
2. **Ruppert, D. (2004).** Statistics and Finance- An Introduction. Springer, New York.
3. **Kijima, M. (2003).** Stochastic Process with Applications to Finance. Chapman Hall/CRC Press, New York.

4. **Tsay, R.S. (2005).** Analysis of Time Series – Third Edition. John Wiley & Sons, New York.
5. **Hull, J.C. (2008).** Options, Futures and other derivatives. Pearson Education India, New Delhi.
6. **Gourieroux, C. and Jasiak, J. (2005).** Financial Econometrics. New Age International (P) Ltd., New Delhi.
7. **Cuthbertson, K. and Nitzsche, D. (2001).** Financial Engineering - Derivatives and Risk Management. John Wiley & Sons, New York.

E14: Longitudinal Data Analysis (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe the basic concepts of Linear Model in longitudinal data analysis

CO2: Analyze numerical methods to solve the problems in Linear Model

CO3: Explain basic concepts of Generalized Linear Model

CO4: Illustrate and study on missing data mechanism in longitudinal data analysis

CO5: Describe Multivariate and Time-dependent covariates in longitudinal data analysis

Unit-I- General Linear Model for Longitudinal Data.ML and REML estimation, EM algorithm: General linear mixed-effects model, Inference for ; the random effects, BLUPs, Empirical Bayes , Bayes, Shrinkage Model building and diagnostic, Relaxing parametric assumptions: generalized additive mixed model.

Unit-II-Generalized Linear Model for Longitudinal Data: Marginal models, for binary, ordinal, and count data: Random effects models for binary ordinal and count data: Transition models: Likelihood-based models for categorical data; GEE; Models for mixed discrete and continuous responses.

Unit-III-Dropouts and missing data: Classification missing data mechanism; Intermittent missing values and dropouts; Weighted estimating equations; Modelling the dropout process (Selection and patternmixture models).

Unit-IV- Time-dependent covariates and special topics: Dangers of time-dependent covariates: Lagged covariates; Marginal Structural models; Joint models for longitudinal and survival data; Multivariate longitudinal data; Design of randomized and observational longitudinal studies.

Text book

1. **Diggle, P.J., Heagerty, P., Liang, K.Y and Zeger. S.L (2003).** Analysis of Longitudinal Data-Second Edition. Oxford University Press, London.

References

2. **Crowder, M.J. and Hand, D.J. (1990).** Analysis of Repeated Measures. Chapman and Hall/ CRC Press, London .
3. **Davidian,M. and Giltinan, D.M. (1995**). Nonlinear Models for Repeated Measurement Data. Chapman and Hall/CRC Press, London.
4. **Hand,D and Crowder, M. (1996).** Practical Longitudinal Data Analysis. Chapman and Hall/CRC Press, London..
5. 4.**Lindsey, J.K. (1993)** Models for Repeated Measurements. Oxford University Press, London.
6. **Little, R.J.A, and Rubin, O.B. (2002).** Statistical Analysis with Missing Data- Second Edition. John Wiley & Sons, New York.
7. **McCullagh, P. and Nelder, J.A (1989).** Generalized Linear Models- Second Edition. Chapman and Hall/CRC Press, London.
8. **Weiss, R.E. (2005).** Modeling Longitudinal Data. Springer, New York.

E15 : Data Mining Techniques (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Apply classification techniques and concept of decision trees.

CO2: Discuss clustering techniques in statistical and data mining viewpoints.

CO3: Explain and apply unsupervised and unsupervised learning and data reduction techniques.

CO4: Explain and apply artificial neural networks and extensions of regression models.

CO5: Discuss data warehousing and online analytical data processing.

CO6: Explain and apply the techniques of association rules and prediction.

Unit-I- Review of classification methods from multivariate analysis; classification and decision trees. Clustering methods from both statistical and data mining viewpoints; vector quantization.

Unit-II- Unsupervised learning from univariate and multivariate data; Dimension reduction and featureselection. Supervised learning from moderate to high dimensional input spaces;

Unit-III- Artificial neural networks and extensions of regression models, regression trees. Introduction todata bases, including simple relational databases.

Unit-IV- Data warehouses and introduction to online analytical data processing. Association rules and prediction; data attributes, applications to electronic commerce.

Text books / References

1. **Berson, A. and Smith, S.J. (1997).** Data Warehousing, Data Mining and OLAP. McGraw-Hill, New York.
2. **Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984).** Classification and Regression Trees. Wadsworth and Brooks/Cole, New York.
3. **Han, J. and Kamber,M. (2000).** Data Mining-Concepts and Techniques. Morgan Kaufmann Publishers, USA.
4. **Mitchell, T.M. (1997).** Machine Learning. Mc Graw-Hill, New York.
5. **Ripley, B.D. (1996).** Pattern Recognition and Neural Networks. Cambridge University Press, UK.

E16 : Statistical Machine Learning – I (Four Credits)

Course Outcomes:

On completion of the course, students should be able :

CO1: To explain the concept of supervised learning and its connections to Statistics

CO2: To use the linear methods of classification, linear regression, logistic regression, piecewise regression and lasso regression in machine learning.

CO3: To understand kernel smothers and related concepts for machine learning.

CO4: To learn various model assessment techniques and inferential procedures.

Unit I: Introduction and overview of supervised learning Variable Types and Terminology, Least Squares and Nearest Neighbours, Local Methods in High Dimension, Supervised Learning and Function Approximation, A Statistical Model for the Joint Distribution of input and output vectors, Function Approximation, Structured Regression Models, Classes of Restricted Estimators: Roughness Penalty and Bayesian Methods, Kernel Methods and Local Regression, Basis Functions and Dictionary Methods. Model Selection and the Bias–Variance Tradeoff. Linear Methods for Regression: Least squares, Subset selection, Shrinkage Methods, Methods using derived input directions, Multiple outcome shrinkage and selection, Lasso and related path algorithms

Unit II: Linear methods for classification Linear methods for classification using linear regression of an indicator matrix, linear discriminant analysis, logistic regression and separating hyperplanes. Basis expansions and regularizations: Piecewise polynomials and splines, Filtering and feature extraction, smoothing splines, Automatic Selection of the Smoothing Parameters, Nonparametric Logistic Regression, Multidimensional Splines, Regularization and Reproducing Kernel Hilbert Spaces, Wavelet smoothing.

Unit III: Kernal smoothing One-Dimensional Kernel Smoothers, Selecting the Width of the Kernel, Structured Local Regression Models in $\mathbb{R}^p$, Local Likelihood and Other Models, Kernel Density Estimation and Classification: Kernel Density Estimation, Kernel Density classification and the Naïve Bayes classifier. Radial Basis Functions and Kernels, Mixture Models for Density Estimation and Classification.

Unit IV: Model assessment, inference and averaging Bias, Variance and Model Complexity, The Bias–Variance Decomposition, Optimism of the Training Error Rate, Estimates of In-Sample Prediction Error, The Effective Number of Parameters, The Bayesian Approach and BIC, Minimum Description Length, Vapnik–Chervonenkis Dimension, Cross-Validation, Bootstrap Methods, Conditional or Expected Test Error, introducing Model Inference and averaging: Local regression in $\mathbb{R}$, The Bootstrap and Maximum Likelihood Methods, Maximum Likelihood Inference, Bootstrap versus Maximum Likelihood, Bayesian Methods, Relationship Between the Bootstrap

and Bayesian Inference, The EM Algorithm, MCMC for Sampling from the Posterior, Bagging, Model Averaging and Stacking, Stochastic Search: Bumping.

Text Books/ References:

1. **Hastie, T., Tshibirai, R. and Friedman, J. (2017)** The Elements of Statistical Learning : Data Mining, Inference and Prediction, 2nd edition. Springer, New York.
2. **James, G., Witten, D., Hastie, T. and Tibshirani, R.(2013)** An Introduction to Statistical Learning with Applications in R. Springer, New York.
3. **James, G., Witten,D., Tibshirani, R. and Hastie, T.** Neural Networks and Deep Learning: A Textbook.
4. Introduction to Machine Learning The Wikipedia Guide.
5. **Vinoth, B, Rajarathian, A. and Manju Bargavi, S.K. (2016)** Nonlinear Regression and Artificial Neural Network Based Model for Forecasting Paddy (Oryza Sativa) Production in Tamil Nadu. *IOSR Journal of Mobile Computing & Application (IOSR-JMCA)e-ISSN: 2394-0050, P-ISSN: 2394-*0042.*Volume 3, Issue 3. (May. - Jun. 2016), pp. 01-06.*

E 17: Statistical Machine Learning - II (Four Credits)

Course Outcomes:

On completion of the course, students should be able :

CO1: To explain the concept of unsupervised learning and its connections to Statistics

CO2: To be familarised with Random Forest techniques in unsupervised learning

CO3: To be acquainted with the advanced unsupervised learning techniques like ensemble learning,undirected graphical method and to deal with high dimensional problems.

Unit I: Unsupervised Learning

Introduction to unsupervised learning, Association Rules like the Market Basket Analysis, The Apriori Algorithm, Unsupervised as Supervised Learning, Generalized Association Rules, Choice of Supervised Learning Method; Cluster Analysis: Proximity Matrices, Clustering algorithms, Combinatorial algorithms, K-means, Gaussian mixtures as Soft K-means clustering, Vector quantization, K-medoids, Hierarchical clustering. Self-Organizing Maps; Principal Components, Curves and Surfaces, Non- negative Matrix Factorization: Archetypal Analysis. Independent Component Analysis(ICA) and Exploratory Projection Pursuit, Exploratory Projection Pursuit, Multidimensional Scaling, Nonlinear Dimension Reduction and Local Multidimensional Scaling, The Google Page Rank Algorithm.

Unit II: Random forests

Definition of Random Forests, Details of Random Forests: Out of bag samples, variable importance, proximity plots, Random Forests and Over fitting. Analysis of Random Forests: Variance and the De- Correlation Effect, Bias, Adaptive Nearest Neighbours.

Unit III: Ensemble Learning and undirected graph models

Ensemble Learning, Boosting and Regularization Paths: Penalized Regression, The “Bet on Sparsity” Principle, Regularization Paths, Over-fitting and Margins. Learning Ensembles: Learning a Good Ensemble, Rule Ensembles. Undirected Graphical Models: Markov Graphs and Their Properties. Undirected Graphical Models for Continuous Variables: Estimation of the Parameters when the Graph Structure is Known, Estimation of the Graph Structure. Undirected Graphical Models for Discrete Variables: Estimation of the Parameters when the Graph Structure is Known, Hidden Nodes, Estimation of the Graph Structure, Restricted Boltzmann Machines.

Unit IV: High dimensional problem $p \gg N$ (p is Much Bigger than N)

Diagonal Linear Discriminant Analysis and Nearest Shrunken Centroids, Linear Classifiers withQuadratic Regularization: Regularized Discriminant Analysis, Logistic Regression with Quadratic Regularization, The Support Vector Classifier, Feature Selection, Computational Shortcuts When $p \gg N$. Linear Classifiers with $L1$ Regularization,The Fused Lasso for Functional Data. Classification When Features are Unavailable: Classification and Other Models Using Inner-Product Kernels and Pairwise Distances. High-Dimensional Regression: Supervised Principal Components, Connection to Latent- Variable Modelling, Relationship with Partial Least Squares, Pre-Conditioning for Feature Selection:

Feature Assessment and the Multiple-Testing Problem: The False Discovery Rate(FDR), Asymmetric Cutpoints and the (significance analysis of microarrays) SAM Procedure, A Bayesian Interpretation of the FDR.

Text Books/ References:

1. **Hastie, T., Tshibirai, R. and Friedman, J. (2017)** The Elements of Statistical Learning : Data Mining, Inference and Prediction, 2nd edition. Springer, New York.
2. **James, G., Witten, D., Hastie, T. and Tibshirani, R.(2013)** An Introduction to Statistical Learning with Applications in R. Springer, New York.
3. **James, G., Witten,D., Tibshirani, R. and Hastie, T.** Neural Networks and Deep Learning: A Textbook.
4. Introduction to Machine Learning The Wikipedia Guide.
5. **Vinoth, B, Rajarathian, A. and Manju Bargavi, S.K. (2016)** Nonlinear Regression and Artificial Neural Network Based Model for Forecasting Paddy (Oryza Sativa) Production in Tamil Nadu. *IOSR Journal of Mobile Computing & Application (IOSR-JMCA)e-ISSN: 2394- 0050, P-ISSN: 2394-*0042.*Volume 3, Issue 3. (May. - Jun. 2016), pp. 01-06.*

E 18: Advanced Statistical Machine Learning Techniques
(Four Credits)

Course Outcomes:

On completion of the course, students should be able :

CO1: To familiarize with the use of generalized additive model, trees and related methods.

CO2: To be well equipped with the Neural Networks and deep learning techniques

CO3: To be acquainted with the prototype methods like K-means clustering and nearest neighbours

Unit I: Additive models, Trees and related methods

Generalized Additive Models: Fitting Additive Models. Tree-Based Methods: Regression Trees, Classification Trees. The Patient Rule Induction Method(PRIM):Bump Hunting. Multivariate Adaptive Regression Splines (MARS), Hierarchical Mixtures of Experts, Missing Data, Boosting and Additive Trees: Boosting Methods, Boosting Fits an Additive Model, Forward Stagewise Additive Modelling, Exponential Loss and Ada Boost, Loss Functions and Robustness, "Off-the-Shelf" Procedures for Data Mining, Boosting Trees. Numerical Optimization via Gradient Boosting: Steepest Descent, Gradient Boosting, Implementations of Gradient Boosting. Right-Sized Trees for Boosting, Regularization: Shrinkage, Sub sampling. Interpretation: Relative Importance of Predictor Variables, Partial Dependence plots.

Unit II: Neural networks and deep learning

Projection Pursuit Regression, Neural Networks, Fitting Neural Networks. Some Issues in Training Neural Networks: Starting Values, Over fitting, Scaling of the Inputs, Number of HiddenUnits and Layers, Multiple Minima. Bayesian Neural Nets and the NIPS 2003 Challenge: Bayes, Boosting and Bagging, Performance Comparisons. Artificial Neural networks, Non-linear regression and deep learning

Unit III: Support vector machines & flexible discriminants

The Support Vector Classifier: Computing the Support Vector Classifier, Support Vector Machines and Kernels: Computing the SVM for Classification , The SVM as a Penalization Method, Function Estimation and Reproducing Kernels, SVMs and the Curse of Dimensionality. A Path Algorithm for the SVM Classifier, Support Vector Machines for Regression, Regression

and Kernels, Discussion; Generalizing Linear Discriminant Analysis, Flexible Discriminant Analysis, Computing the FDA Estimates, Penalized Discriminant Analysis , Mixture Discriminant Analysis

Unit IV: Prototype methods and nearest neighbours Prototype Methods:- K-means Clustering, Learning Vector Quantization, Gaussian Mixtures; k-Nearest- Neighbour Classifiers , Examples, Invariant Metrics and Tangent Distance; Adaptive Nearest-Neighbour Methods , Global Dimension Reduction for Nearest- Neighbours, Computational Considerations

Text Books/ References:

1. **Hastie, T., Tshibirai, R. and Friedman, J. (2017)** The Elements of Statistical Learning : Data Mining, Inference and Prediction, 2nd edition. Springer, New York.
2. **James, G., Witten, D., Hastie, T. and Tibshirani, R.(2013)** An Introduction to Statistical Learning with Applications in R. Springer, New York.
3. **James, G., Witten,D., Tibshirani, R. and Hastie, T.** Neural Networks and Deep Learning: A Textbook.
4. Introduction to Machine Learning The Wikipedia Guide.
5. **Vinoth, B, Rajarathian, A. and Manju Bargavi, S.K. (2016)** Nonlinear Regression and Artificial Neural Network Based Model for Forecasting Paddy (Oryza Sativa) Production in Tamil Nadu. *IOSR Journal of Mobile Computing & Application (IOSR-JMCA)e-ISSN: 2394- 0050, P-ISSN: 2394-*0042.*Volume 3, Issue 3. (May. - Jun. 2016), pp. 01-06.*

Elective 19: Non-Parametric Statistical Methods (Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1 Understand the concept of random and non parmetric tests for random-ness.

CO2 Apply various nonparametric statistical methods for testing normality .

CO3 Use the nonparametric alternative of one way and two way analysis ofvariance.

CO4 Get basic knowledge of nonparametric density estimation.

Unit-I Measurement scales: Nominal Scale, Ordinal Scale, Interval Scale, Ra- tio Scale; Order Statistics, Quantiles, QQplot, Empirical distribution, Kaplan-Meier estimator; Measures of dependence: Cramer's contingency coefficient, Pearson's contingency coefficient, Pearson's mean-square con-tingency coefficient, Phi coefficient.

Unit-II Lilliefors's test for normal and exponential distributions, Cramer-Von Mises test for identical populations, Anderson Darling test, Shapiro-Wilk test, Jarque-Bera test, Test for randomness: Wald Wolfowitz run test.

Unit-III McNemar test, Cochran's test, K samples procedures; Kruskal-Wallis test, Friedman's test, Resampling techniques; Jackknife method for bias and variance of an estimator, Nonparametric Bootstrap method for estimating variance and distribution of an estimator.

Unit-IV Smoothing, Bias variance, trade off, Nonparametric density estimation; Cross validation, Histogram and Kernel density estimation, Kernel density rug plot in R.

Practical problems Using R

Text Books:

1. **Gibbons, J. D. and Chakraborti, S. (2003)**.Nonparametric Statis- tical Inference, 4th edition(Sections 4.5,4.6,6.2 and 14.5). Marcel Dekker, New York.
2. **Wasserman, L. A.(2006)**. All of Nonparametric Statistics. Springer, New York(Sections 3.1, 3.2, 4.1,6.1,6.2 and 6.3).
3. **Hollander , M. , Wolfe, D. A. and Chicken, E. (2013)**. Non para- metric Statistical Methods. Wiley, New York(Section 11.6)

4. **Sprent, P. (1989)**. Applied Nonparametric Statistical Methods . Springer,Netherlands.(Section 5.5.2)
5. **Conover, W. J. (1999)**. Practical Nonparametric Statistics, 3rd Edi- tion. (Sections 2.1, 4.4, 4.6 and 6.2) Wiley.
6. **D'Agostino, R. B. and Stephans, M. A.(1986)**. Goodness-of-fit Techniques. Marcel Dekker. (Section 4.2.2)

References:

1. **Kloke, J. and McKean, J. W.(2015)**. Nonparametric Statistical Methods Using R . CRC Press, London.
2. **Siverman, B. W.(1986)**. Density Estimation for Statistics and DataAnalysis . Springer, New York.
3. **Jarque, C.M and Bera, A. K(1987)**. A test for normality of ob- servations and regression residuals, *International Statistical Review* **55**, 163-172.
4. **Anderson, T. W., and Darling, D. A. (1954)**. A test of goodness of fit. *Journal of the American Statistical Association*, **49**, 765–769.

E20: Statistical Modeling and Data Mining Techniques
(Four Credits)

Course Outcomes:

On completion of the course, students should be able to :

CO1 Acquire skills in advanced statistical modeling.

CO2 Get acquainted with association analysis in data mining techniques.

CO3 Learn the data pre-processing steps in data mining.

CO4 Understand various measures of patterns, data warehousing and the concepts of online transaction and analytical processes in data mining.

Unit I Statistical Modeling, Steps in Statistical Modeling, Regression Analysis: Transformations and weighting to correct model inadequacies, Analytical methods for selecting a transformation, The Box-Cox method, Transformation on the regressor variables, Ridge regression, Basic form of ridge regression, Robust regression; Least absolute deviation regression, Least median of squares regression.

Unit II Data mining; Introduction, Data types for Data mining, Database and Data warehouse, Data mining functionalities - Concept/class description: characterization and discrimination, Association analysis, Classification and Prediction, Clustering analysis, Evolution and Deviation analysis, Data Pre-processing, Data cleaning, Data Integration and transformation, Data reduction, Discretization and concept hierarchy generation.

Unit III Measures of pattern interestingness; Objective measures of pattern- sup- port and confidence, Classification of data mining systems; Classifications according to databases, Knowledge and techniques, Major issues in data mining .

Unit IV Data warehouse, On-line transaction process(OLTP) and On-line analytical processing(OLAP), Distinguishing features between OLTP and OLAP.

Text Books:

1. **Montgomery, D.C., Peck, E.A. and Vining, G.G.(2007)**. Introduction to Linear Regression Analysis. Wiley, New York.
2. **Han, J., Kamber, M. and Pei, J.(2000).** Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers (Relevant sections of Chapters 1, 2 and 3).

References:

1. **Draper, N. R. and Smith, H.(1998)** Applied Regression Analysis, 3*rd* Edition, Wiley.
2. **Berson, A. and Smith, S.J. (1997)**. Data Warehousing, Data Mining, and OLAP. McGraw-Hill, New York.

E21: Applied Algorithms and Analysis of Multi type and Big Data
(Four Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1 Explain the concept of EM clustering algorithms.

CO2 Understand the classification techniques and the concept of support vector machines.

CO3 Get acquainted with basic concepts related to big data.

CO4 Learn the multidimensional scaling techniques in unsupervised learning.

Unit I EM Algorithm: Two-Component Mixture Model, Gaussian Models, The EM Algorithm in General, EM as a Maximization–Maximization Procedure.

Unit II Support Vector Machines: Maximal Margin Classifier, Support Vector Classifiers, Support Vector Machines, SVMs with More than Two Class- One- Versus-One Classification and One-Versus-All Classification.

Unit III Big Data: Definition, Characteristics, Data Analytics, General Categories of Data Analytics, Structured, Unstructured and Semi Structured Data, Met data, Big Data Analytics Life Cycle.

Unit IV Multi-Dimensional Scaling; Definition, Perceptual Map, Decision Frame- work for Perceptual Mapping, Non-metric versus Metric methods, Similarities Data, Preferences Data, Aggregate and Disaggregate Analysis, Decompositional and Compositional approaches, Interpreting the MDS results.
Practical problems using R.

Text Books:

1. **Hastie, T., Tibshirani, R. and Friedman, J. (2017)**. The Elements of Statistical Learning : Data Mining, Inference and Prediction, 2nd edition. Springer, New York(Section 8.5).
2. **James, G., Witten, D., Hastie, T. and Tibshirani, R.(2013)**. An Introduction to Statistical Learning with Applications in R. Springer, New York (Sections 9.1-9.4).
3. **Erl, T. and Khattak, W. (2016)**. Big Data Fundamentals Concepts, Drivers & Techniques. Prentice Hall. (Chapters 1 and 3)

4. **Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E.(2009)**. Multivariate Data Analysis, 7th edition. Prentice Hall, New York (Chapter 10).

Audit courses:

In addition to the core and elective courses of the programme there will be two Audit Courses (Ability Enhancement Course & Professional Competency Course) with 4 credits each. These have to be done one each in the seventh & eigth semesters. These courses are mandatory for all programmes but their credits will not be counted for evaluating the overall SGPA & CGPA. The Department/College shall conduct examination for these courses and have to intimate /upload the results of the same to the University on the stipulated date during the Third Semester. Students have to obtain only minimum pass requirements in the Audit Courses. The details of Audit courses are given below.

Ability Enhancement Course (AEC) 4 Credits

The objective of this course is to enhance the ability and skill of students in the core and elective areas of statistics, through hands on experience, internship, industrial visits, case study, community linkage, book/research paper review, scientific word processing etc.

The faculty members in the department collectively or a particular faculty member shall be in charge of this course for students of the semester, which shall be decided by the Department council. The following are the requirements in this course:-

1. Short term internships at research institutions/R&D centre/Industry.
2. Seminar presentation on a topic in statistics or related fields that is not normally covered in the in the syllabi of the programme.
3. Case study and analysis on any relevant issues in the nearby society
4. Publication of articles in statistical magazines/journals
5. Interaction with Statistical Organizations/ Industries/ Research Institutions.
6. Any community linking programme relevant to the area of study
7. Book/paper review and summary.
8. English communication skills and technical writing with LATEX.
9. Survey methodology and Data collection- sampling frames and coverage error, non-response.
10. Developing a questionnaire, collect survey data pertaining to a research problem (such as gender discrimination in private vs government sector, unemployment rates, removal of subsidy, impact on service class). Formats and presentation of reports.

After conducting the AEC, the evaluation/examination should be done either common for all students of the semester or individually depending upon the AEC conducted. Evaluation/examination on AEC must contain the following components: MCQ type written examination, Report on study/investigation, Presentation, Viva voce etc. as decided by the Department council. Evaluation/examination must be conducted by 30 weightage pattern, as in the theory courses and the GPA and overall grade of the AEC should be determined.

Professional Competency Course (PCC) 4 Credits

The objective of this course is to get professional competency and exposure in the core areas of statistics. It particularly aims to improve the skill level of students, especially for using software useful in their respective field of study, both related to the core and elective subject area. Also it is a platform for the student community to undertake socially committed statistical investigations and thereby developing a method of learning process by doing through the involvement with society.

The faculty members in the department collectively or a particular faculty member shall be in charge of this course for students of the semester, which shall be decided by the Department council. The following are the requirements in this course:-

1. Working knowledge on different statistical software/utilities like SPSS (or GNU PSPP), R, Python. (Introduction of the software- Use of the software as a calculator, as a graphing (plotting) utility, for matrix operations and for problems on probability distributions)
2. Use of Internet and other technologies - Internet and www, applications, internet protocols.
3. E-commerce and financial statistics- Electronic fund transfer, payment portal, e-commerce security.
4. Mobile commerce, Bluetooth and Wi-Fi
5. Introduction to Data Science and Big-data issues.
6. Trend Analysis (elementary time series analysis) and Index numbers
7. Official Statistics: An outline of present official statistical systems in India, Methods of collection of official statistics, their reliability and limitations, Role of MoSPI, CSO, NSSO and NSC.

8. Monte Carlo methods: Brief look at some popular approaches- simulating a coin toss, a die roll and a card shuffle.
9. CDF inversion method- simulation of standard distributions
10. Monte Carlo Integration- Basic ideas of importance sampling.

After conducting the PCC, the evaluation/examination should be done either common for all students of the semester or individually depending upon the PCC conducted. Evaluation/examination on PCC must contain the following components: MCQ type written examination, Report on study/investigation, Presentation, Viva voce etc. as decided by the Department council. Evaluation/examination must be conducted by 30 weightage pattern, as in the theory courses and the GPA and overall grade of the PCC should be determined.

MODEL QUESTION PAPER

FIRST SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA1IB01 - STATISTICAL METHODS

Time : Two Hours Maximum : 60 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 20.

1. Define qualitative and quantitative data.
2. Explain ordinal and nominal data.
3. What are the limitations of Statistics.
4. What are the requisites for an ideal measures of central tendency.
5. Show that the sum of squares of deviations of the observations taken from arithmetic mean is zero.
6. A driver has to drive 90 miles at an average 30mph. He could drive only at the rate of 20mph for the first half of the journey. What must be his average speed for the second half of the journey so as to make the average speed of his entire journey as 30mph.
7. Define mean deviation about mean.
8. What is the sum of squares of 10 observation with mean 4 and variance 36.
9. Two samples of sizes 40 and 50 respectively have the same mean 53 but different standard deviation 19 and 8 respectively. Find the standard deviation of the combined sample of size 90.
10. Find the mean and variance of first n natural numbers.
11. What is meant by raw moments and central moments.
12. Distinguish between lepto kurtic and platy kurtic curves.

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 30.

13. Define primary data. What are the various methods of collecting primary data.
14. Differentiate questionnaires and schedules.

15. Explain the advantages of sampling over census. Explain simple random sampling and stratified sampling.

16. For two positive numbers, show that $A.M. \geq G.M \geq H.M.$. When does the equality sign hold.

17. The expenditure of 1,000 families is given below:

Expenditure	:	40-59	60-79	80-99	100-119	120-139
Number of families	:	50	–	500	–	50

The median and mean of the distribution are both Rs. 87.50. Find the missing figures.

18. Prove that for any discrete distribution standard deviation is not less than mean deviation from mean.

19. Define moments. Derive an expression for r^{th} central moment in terms of raw moments.

Section C(Essay Type Questions)

Each questions carries 10 marks (1 out of 2)

Maximum marks that can be scored from this section is 10.

20. Find the mean deviation about median and standard deviation for the following data.

Class	:	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	:	2	12	18	25	22	15	6

Also find the percentage of the item in the rage (i) $\bar{x} \pm \sigma$ (ii) $\bar{x} \pm 2\sigma$.

21. Define Skewness. Explain various measures of skewness. Calculate β_1 and comment on skewness of a set of data:

Class	:	1-5	6-10	11-15	16-20	21-25	26-30	31-35
Frequency	:	6	8	136	60	20	12	4

* * * * *

MODEL QUESTION PAPER

SECOND SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA2IB02 - BASIC PROBABILITY THEORY

Time : Two Hours Maximum : 60 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 20.

1. Define random experiment.
2. What are equally likely events.
3. What is the probability of drawing a black queen from a well shuffled deck of cards.
4. Distinguish between classical and empirical definitions of probability.
5. What are the axioms of probability. Using these axioms establish $P(A^c) = 1 - P(A)$.
6. If $P(A) = P(B) = P(B|A) = 0.5$, Examine whether A and B are independent.
7. If $P(.)$ is a probability measure, show that $Q(B) = P(B|A)$ for any event B is also a probability measure.
8. Define distribution function of a random variable X and state any two properties.
9. $f(x) = kx, x = 1, 2, 3$ and zero elsewhere is a p.m.f. Find $P(X \geq 1.8)$
10. A random variable has density $f(x) = kx^2e^{-x^3},\ x > 0$. Determine k
11. Find $E(X)$, X denotes the square of the number shown by a fair coin when it is tossed.
12. Define the moment generating function of a random variable.

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 30.

13. State and prove addition theorem for the probabilities of two events. Using this prove the addition theorem for three events.
14. If $P(A) = 0.3, P(B) = 0.2$ and $P(A \cap B) = 0.1$, find the probability that: (i) Exactly one of the events will happen (ii) At least one of the events will happen (iii) None of the events happen.
15. Three person A, B and C shoot at a target. The chances of their hitting the target are $\frac{1}{3},\ \frac{1}{2}$ and $\frac{4}{5}$ respectively. Find the chances that the target will be hit if three attempt to shoot exactly one each.

16. Define pairwise independence and mutual independence of events. Discuss the implication between them.

17. Given the probability density function $f(x) = \frac{1}{x^2},\ 1 < x < \infty$. Find $P(A \cup B)$ where $A = \{x|1 < x < 2\}$ and $B = \{x|4 < x < 5\}$.

18. If X is a rectangular random variable over the interval [0, 1], find the probability density function of $Y = e^x$.

19. Find the moment generating function a random variable with probability density function $f(x) = 2x, 0 < x < 1$, and hence find its mean and variance.

Section C(Essay Type Questions)

Each questions carries 10 marks (1 out of 2)

Maximum marks that can be scored from this section is 10.

20. State and prove Bayes theorem. Suppose that there is a chance for a newly constructed house to collapse whether the design is faulty or not. The chance that the design is faulty is 10%. The chance that the house collapse if the design is faulty is 95% and otherwise it is 45%. It is seen that the house collapsed. What is the probability that it is due to faulty design?

21. A continuous random variable X has the probability density function,

$$f(x) = \begin{cases} ke^{-\frac{x}{2}}, & \text{if } 0 < x < \infty \\ 0, & \text{otherwise.} \end{cases}$$

Find the value of the constant k. Find the variance of X. Show that for any two positive numbers s and t,

$$P(X > s + t|X > t) = P(X > t).$$

* * * * *

MODEL QUESTION PAPER

THIRD SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA3IB03 - BIVARIATE DATA ANALYSIS

Time : Two Hours Maximum : 60 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 20.

1. Define scatter diagram.
2. Fit a straight line of the form $y = ax + b$ to the following data on x and y:

x	:	4	6	8	10	12
y	:	12	15	22	34	40

3. Explain the fitting of a curve of the form $y = ae^{bx}$ to the data $(x_1, y_1),\ (x_2, y_2), \ldots,\ (x_n, y_n)$.
4. Describe negative and positive correlation. Give example.
5. Define multiple and partial correlation coefficient.
6. Two regression coefficients cannot be grater than 1 simultaneously. Justify
7. What is dichotomy? Show that the continued dichotomy according to n attributes gives rise to 3^n classes.
8. If $f(x, y) = cxy,\ 0 < x < 4;\ 1 < y < 5$ is the joint p.d.f. Determine c.
9. Define Joint probability distribution function and state its properties.
10. If the joint p.m.f. of X and Y is $p(x_1, x_2) = \frac{x_1+x_2}{24},\ x_1 = 0, 1, 2, 3;\ x_2 = 1, 2$. Find the marginal p.m.f's
11. Explain how you can get the joint joint p.d.f. from the marginal and conditional p.d.f.'s.
12. State and prove the addition theorem on expectation for two random variables X and Y.

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 30.

13. Fit a curve of the form $y = ab^x$ to the following data:

x	:	2	3	4	5	6	7	8
y	:	10	14	21	45	72	94	120

14. Show that correlation coefficient is independent of change of origin and scale.

15. Compute Spearman's rank correlation coefficient for the following data:

x	:	92	89	87	86	84	77	71	63	53	50
y	:	86	83	91	77	68	85	52	82	37	57

16. An investigation of 23,713 households was made in an urban and rural mixed locality. Of these 1,618 were farmers 2,015 well-to-do and 770 families were having at least one graduate. Of these graduate families 335 were those of farmers and 428 were well-to-do, also 587 well-to-do families were those of farmers and out of them only 156 were having at least one of their family member as graduate. Obtain all the ultimate class frequencies.

17. Can vaccination be regarded as a preventive measure for small pox from the data given below:
'Of 1482 persons in a locality exposed to small-pox, 368 in all were attacked'
'Of 1482 persons, 343 had been vaccinated and of these only 35 were attacked.'

18. For two random variables X and Y, the joint p.m.f $f(x,y) = \frac{x^2+2y}{22},\ x = 1,2;\ y = 1,2.$ Whether X and Y are independent?

19. Two random variables X and Y have the following joint probability density function:

$$f(x,y) = \begin{cases} 2-x-y, & \text{if} 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find conditional density functions.

Section C(Essay Type Questions)

Each questions carries 10 marks (1 out of 2)

Maximum marks that can be scored from this section is 10.

20. The heights in inches (x) and weights (y) in lbs of 10 college students are given below:

x	:	70	64	72	67	65	69	70	62	72	66
y	:	181	125	178	160	139	145	165	126	180	132

Calculate (i) The correlation coefficient (ii) The regression of y on x.

21. Let X and Y are two random variables with joint p.d.f. $f(x,y) = 8xy,\ 0 < x < y < 1$ and $f(x,y) = 0$ elsewhere. Find correlation between X and Y.

* * * * *

MODEL QUESTION PAPER

THIRD SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA3IB04 - PROBABILITY DISTRIBUTIONS

Time : Two Hours Maximum : 60 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 20.

1. Find the mean of a Uniform random variable with possible values 1,2,3,4 and 5.
2. Find the mean and variance of Bernoulli distribution.
3. Define negative binomial distribution
4. Obtain the m.g.f. of gamma distribution.
5. If X follows normal distribution with mean 10 and variance 9, find $P(7 < X < 13)$.
6. Define Laplace and Weibull distribution.
7. Define Beta I distribution. How it relates to Beta II
8. Define parameter and statistic.
9. Define convergence in distribution.
10. State Central Limit Theorem.
11. Find the probability that the sample mean of a random sample of size 16 taken from a normal population with mean 2 and variance 4 is less than 1.
12. If X and Y are independent standard normal random variables, identify the probability distribution of $\left[\frac{X-Y}{X+Y}\right]^2$.

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 30.

13. If the $(r-1)^{\text{th}}$, r^{th} and $(r+1)^{\text{th}}$ central moments of X following binomial distribution with parameters n and p are μ_{r-1}, μ_r and μ_{r+1}, show that $\mu_{r+1} = pq\left[nr\mu_{r-1} + \frac{d}{dp}\mu_r\right]$.
14. Show that for a Poisson distribution, mean and variance are equal.
15. State and prove lack of memory property of exponential distribution.

16. Examine whether Weak Law of Large Numbers hold for the sequence of random variable $\{X_i\}$, where $P(X_i = \pm\sqrt{2i-1}) = \frac{1}{2}$.

17. State and prove Bernoulli's law of large numbers.

18. Explain two uses of t distribution.

19. If t follows student's t-distribution with n degrees of freedom, show that t^2 follows F distribution with $(1, n)$ degrees of freedom.

Section C(Essay Type Questions)

Each questions carries 10 marks (1 out of 2)

Maximum marks that can be scored from this section is 10.

20. (a) Show that odd cental moments of X following $N(\mu, \sigma^2)$ are zeroes.

    (b) Prove that the mean deviation about the mean of X following $N(\mu, \sigma^2)$ is $\sqrt{\frac{2}{\pi}}\sigma$.

21. (a) State and prove the additive property of Chi-square distribution by deriving m.g.f.

    (b) State the relationship between Normal, χ^2, t and F statistic.

* * * * *

Fourth Semester Integrated M. Sc. Examination

STA4IB05 **- Statistical Inference- 1**

Time: 2 Hours **Maximum Marks: 60**

Part A

Each question carries 2 Marks.

Maximum Marks that can be scored in this Part is 20

1. Define the terms (i) point estimation and (ii) interval estimation.
2. Define an unbiased estimator. Give an example.
3. State the factorization theorem for sufficient statistic?
4. Define an efficient estimator.
5. Find the moment estimator of the parameter p of a Bernoulli distribution.
6. State Rao-Blackwell theorem.
7. What do you mean by Fisher Information?
8. State the Cramer-Rao inequality.
9. If $X_1, X_2, \ldots, X_n$ is a random sample (r.s) from $N(\mu, \sigma^2)$, what is the distribution of $\frac{nS^2}{\sigma^2}$ under the cases μ known and unknown?
10. Write down the confidence interval for the difference of proportions, under large sample setup.
11. Define first-order statistic.
12. Find the cumulative distribution function of the largest order statistic when a r.s of size n is taken from a distribution with cumulative distribution function F(x).

Part B

Each question carries 5 Marks.

Maximum Marks that can be scored in this Part is 30

13. Differentiate between an estimator and estimate with suitable examples.
14. State and prove a sufficient condition for the consistency of an estimator.
15. Estimate the two parameters of a Gamma distribution, using the method of moments.
16. Find the maximum likelihood estimator(m.l.e) of the parameter θ when we have a random sample of size 'n' from

$$f(x) = \begin{cases} 1 & if \theta - \frac{1}{2} < x < \theta + \frac{1}{2} \\ 0 & otherwise \end{cases}$$

17. Random samples of sizes 9 and 7 taken from two normal populations, present the following data:

Sample means:	56	50
Sample variances:	6	12

Obtain 95% confidence interval for the difference of population means.

18. Obtain the confidence interval for the ratio of variances of two normal populations.
19. Derive the density function of the r^{th} order statistic when a r.s of size n is taken from a distribution with cumulative distribution function F(x). Hence derive the distribution of first order statistic.

Part C

Answer any one question and carries 10 Marks.

20. (i) Show that the sample variance $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is a biased estimator of the population variance. Suggest an unbiased estimator for the population variance.

(ii) In a random sample of 8 envelopes is taken from letter box of a post office and their weights in grams are found to be 12·1,11·9, 12·4, 12·3, 11·9, 12·1, 12·4, 12·1. (a) Find 99% confidence limits for the mean weight of the envelop received at that post office.

21. (i) Obtain the sufficient estimator for the parameter θ , when a r.s of size n is taken from the p.d.f

$$f(x) = \frac{1}{\theta}\ , 0 \leq x \leq \theta$$

(ii) If $X_1, X_{2,} \dots, X_n$ is a random sample from exponential with mean λ , find the distribution of $Y = \min(X_1, X_{2,} \dots, X_n)$.

****

MODEL QUESTION PAPER
SIXTH SEMESTER INTEGRATED M.Sc. S EXAMINATION
(CBCSS-UG)
Statistics
STA4IB06 : SAMPLING TECHNIQUES

Time: Two Hours Maximum: 60 Marks.

Section A (Short Answer Type Questions)
Each question carries 2 marks
Maximum marks that can be scored from this section is 20.

1. Define simple random sample.
2. What is meant by sampling design?
3. What is meant by sampling frame?
4. What are the principles of stratification?
5. What is allocation in stratified random sampling?
6. Define linear systematic sampling.
7. Define non response error.
8. How will you draw a circular systematic sample?
9. What is finite population correction?
10. What is a random number table?
11. Mention a situation where sampling method alone can be used.
12. What is meant by probability sampling?

Section B (Paragraph /Problem Type Questions)
Each question carries 5 marks
Maximum marks that can be scored from this section is 30.

13. What are the advantages of sampling over census?
14. What is meant by stratification? What are the basic principles of stratification?
15. Derive an expression for $V(\bar{y}_{sys})$ where a systematic sample of size n drawn from a population of size N=nk units.
16. In SRSWOR, show that the samples mean is an unbiased estimate of population mean. Compare its efficiency with SRSWR.
17. Give the unbiased estimate of population mean in the case of stratified random sampling and obtain an expression for it's variance.
18. Derive the confidence interval for population mean in SRSWOR.

19. Define sampling and non sampling errors.

Section C (Essay Type Questions)

Each question carries 10 marks(1 out of 2)

Maximum marks that can be scored from this section is 10.

20. Describe various steps in planning and execution of a large scale sample survey.
21. Show that $V_{opt} \leq V_{prop} \leq V_{ran}$, where V_{opt} , V_{prop} and V_{ran} are the variances of the sample mean under optimum allocation, proportional allocation and SRSWOR respectively.

****

Fifth Semester Integrated M. Sc. Examination

STA5IB07 -Statistical Inference- 11

Time: Two and a half hours **Maximum Marks: 80**

Part A

Each question carries 2 Marks.

Maximum Marks that can be scored in this Part is 25

1. Distinguish between simple and composite hypotheses.
2. Define (i) Type I error and (ii) Type II error.
3. Define power of a test.
4. State Neyman-Pearson Lemma.
5. Define uniformly most powerful test.
6. Define an unbiased test .
7. What is meant by sampling distribution?
8. List the assumptions of one sample t test.
9. If $X_1, X_2, \ldots, X_n$ is a random sample (r.s) from $N(\mu, \sigma^2)$, what is the test statistic used for testing the population mean when the sample size is large and σ^2 unknown?
10. What do you mean by a contingency table?
11. List the steps involved in testing of single population proportion.
12. Write a suitable situation where Kruskall Wallis test is applicable.
13. Define runs with a suitable example.
14. What do you mean by empirical distribution?
15. Briefly explain one sample sign test.

Each question carries 5 Marks.

Maximum Marks that can be scored in this Part is 35

16. It is desired to test the hypothesis H_0 :θ=1 *against* H_1 : θ=2 using a single observation X. X $\geq$ 0.95 is used as the critical region. Evaluate Type I error and Type II error.
17. If X follows $N(0,\sigma^2)$, show that there does not exist UMP test of $H_0 : \sigma^2 = \sigma_0^{\,2}$ against $H_1 : \sigma^2 \neq \sigma_0^{\,2}$.
18. Explain the likelihood ratio test.
19. A sample of 25 boys who passed SSLC examination are found to have mean marks 50 with standard deviation 5 for English. The mean marks of 18 girls are found to be 48 with standard deviation 4 for the same subject. Does this indicate any significance difference between the marks of boys and girls assuming the population standard deviation are equal?
20. Tests were carried out to assess the strength of single fibre yarn spun on two different machines A and B and the results are given below:

Machine A	4	4.4	3.9	3	4.2	4.4	5
Machine B	5.3	4.3	4.1	4.4	5.3	4.2	3.8

Assuming the samples have been taken from normal population, test the hypothesis that variability is same for both the machines.

21. Explain Chi square test for independence of attributes.
22. Explain the test for Kolmogorov Smirnov two sample test.
23. A sample of 10 men was used in a study to test the effects of a relaxant on the time required to fall asleep for male adults. Data for 10 subjects showing the number of minutes required to fall asleep with and without the relaxant follow. Use a 0.05 level of significance to determine whether the relaxant reduces the time required to fall asleep. Perform sign test and draw your conclusion.

Subject	1	2	3	4	5	6	7	8	9	10
Without Relaxant	15	12	22	8	10	7	8	10	14	9
With Relaxant	10	10	12	10	8	5	9	7	11	6

Maximum Mark = 35

PART C

Each question carries 10 marks (Answer any **TWO** Questions)

24. Use the Neyman Pearson Lemma to obtain the best critical region for testing $H_0 : \mu=\mu_0$ against $H_1 : \mu=\mu_1$, in the case of a normal population with mean μ and variance σ^2. Find the power of the test.
25. In a survey of 70 business firms, it was found that 45 are planning to expand their capacities next year. Does the sample information contradict the hypothesis that 70% the firms are planning to expand next year.

26. Fit a Poisson distribution for the following data and test the goodness of fit.

X	0	1	2	3	4	5	6
frequency	275	72	30	7	5	2	1

27. Explain (i) Wilcoxon signed rank test and (ii) Mann Whitney test

(2 x 10=20 marks)

****

MODEL QUESTION PAPER
SIXTH SEMESTER INTEGRATED M.Sc. S EXAMINATION
(CBCSS-UG)
Statistics
STA5IB08 : DESIGN AND ANALYSIS OF EXPERIMENTS

Time: Two and half Hours Maximum: 80 Marks.

Section A (Short Answer Type Questions)
Each question carries 2 marks
Maximum marks that can be scored from this section is 25.

1. What is meant by linear estimation?
2. What do you mean by fixed effects model?
3. How do you define a linear hypothesis?
4. What do you mean by the statement "treatment effects are significant" in ANOVA?
5. Define Gauss-Markov theorem.
6. Explain "Uniformity trial".
7. Write down the ANOVA table of a CRD.
8. What is the degree of freedom associated error of a RBD with 4 treatments in 'r' replicates?
9. What is the efficiency of RBD over CRD?
10. What are the null and alternative hypotheses stated in LSD?
11. What do you mean by row efficiency of LSD over RBD?
12. What are the demerits of a CRD?
13. What do you mean by partially balanced incomplete block design?
14. Which are the two types of effects measured in a factorial experiment?
15. How many factors are there in a 2^3 factorial experiment? Explain.

Section B (Paragraph /Problem Type Questions)
Each question carries 5 marks
Maximum marks that can be scored from this section is 35.

16. State and prove the necessary and sufficient condition for the estimability of a linear parametric function.
17. What is meant by BLUE? Explain the procedure to obtain it?
18. What are the different types of statistical models for experimental design?

19. Describe CRD. Write the model for the data from such an experiment and state the advantages of CRD.
20. How do you estimate two missing observations in RBD? Explain.
21. Define Latin square design. How do you split up the total sum of squares into components for this design? Give the analysis of variance table?
22. Give any four factors responsible for determining the number of replications?
23. Define BIBD. State the important relations among the parameters of a BIBD.

Section C (Essay Type Questions)

Each question carries 10 marks(2 out of 4)
Maximum marks that can be scored from this section is 20.

24. Explain the basic principle of experimental design.
25. Describe a method of analyzing RBD with 2 factors A and B at p and q levels respectively.
26. Explain 2^2 factorial table and obtain its ANOVA table..
27. Explain the analysis of two way classified data with unequal number of observations in cells under fixed effect model.

****

MODEL QUESTION PAPER
FIFTH SEMESTER INTEGRATED M.Sc. S EXAMINATION
(CBCSS-UG)
Statistics
STA5IB09 : DATA ANALYTICS WITH R

Time: Two and half Hours Maximum: 80 Marks.

Section A (Short Answer Type Questions)
Each question carries 2 marks
Maximum marks that can be scored from this section is 25.

1. Write down advantages of R.
2. Create a vector of 15 elements and access 10^{th} and 12^{th} element.
3. Write down the R command to draw a pie chart.
4. Explain read.table() in R.
5. Explain saving, storing and retrieving work in R.
6. Describe help() in R.
7. Explain while loop in R.
8. What is out put of R command seq(10,25,by=3)?
9. Write down the R command to draw a boxplot.
10. What do you mean by packages in R?
11. Construct a matrix of order 4×3 with first 12 natural numbers as elements.
12. What is the out put of R command rep(1,4)?
13. Explain data.frame() in R.
14. Write down the R command to find mean, median, variance and range of a vector X.
15. How do you install R in your computer?

Section B (Paragraph /Problem Type Questions)
Each question carries 5 marks
Maximum marks that can be scored from this section is 35.

16. Write about some useful built-in functions in R
17. Write down the R command to access the data set "treering" containing tree-ring widths in dimensionless units, from the base package of R .Then write the R command to find the following
    a) How many observations are in the dataset?
    b) What is the minimum observation and maximum observation?

c) List the observation longer than1.8.

18. For the data given below

Age (mid pont):	25	35	45	55	65	75	85	95
Number(frequency):	55	93	113	90	85	73	29	5

write down R command for

a) Enter the data set using data.frame command

b) Add a column of cumulative frequency

c) Plot cumulative frequency v/s mid points.

19. Describe the procedure of import a data file from Excel to R.

20. Create two matrices of order 3×3 and write down R command to find

a) Matrix multiplication

b) Addition and subtraction of matrices

c) Inverse of A

d) Transpose of A

21. Explain the use of Histogram. Give the R command to draw Histogram.

22. Explain the utility of scan() function.

23. Briefly explain if statement, if-else statement and if-elseif statement in R

Section C (Essay Type Questions)

Each question carries 10 marks(2 out of 4)

Maximum marks that can be scored from this section is 20.

24. Write about data input methods in R

25. Explain the following

a) Lists in R programing

b) Data frames in R programing

c) Workspace in R programing

d) Scripts in R programing

26. Write down the R command to determine correlation coefficient, regression line X on Y and Y on X for the following data

X:	34	37	36	32	32	36	35	34	29	35
Y:	37	37	34	34	33	40	39	37	36	35

27. Explain the following,

a) repeat loop

b) for loop

c) Accessing array elements

d) Creating a list, list tags and values

e) Identify NA values

****

MODEL QUESTION PAPER
FIFTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics
STA5IB10 - APPLIED STATISTICS

Time: Two and half Hours Maximum: 80 Marks

Section A (Short Answer Type Questions)
Each question carries 2 marks
Maximum marks that can be scored from this section is 25.

1. What is meant by seasonal variation?
2. Distinguish between assignable and chance causes.
3. Define an index number.
4. Give any four limitations of index numbers.
5. Distinguish between population and sample.
6. Define age specific fertility rate.
7. Explain consumer price index number.
8. Give the formula for Fisher's ideal index. Why it is said to be 'ideal'?
9. What are the uses of index numbers?
10. Explain briefly the concept of whole sale price index number.
11. Explain principle of least squares method.
12. Write short note on vital registration.
13. What is meant by splicing of index number?
14. What is time reversal test?
15. Give the formula for Marshall-Edgeworth index number.

Section B (Paragraph/Problem Type Questions)
Each question carries 5 marks
Maximum marks that can be scored from this section is 35.

16. Briefly explain the sources of demographic data.

17. What are the processes of Indian civil registration?

18. Explain the additive and multiplicative model for a time series.

19. Why are index numbers called economic barometers?

20. Explain the four methods to estimate the secular trend.

21. Expalin the components of a times series with examples.

22. Explain the method of ratio-to-moving average. What are the merits and demerits of this method?

23. Calculate seasonal variation indices from the data given below using simple average method.

Year	Quarter I	Quarter II	Quarter III	Quarter IV
1995	65	58	56	61
1996	68	63	63	67
1997	70	59	56	52
1998	60	55	51	58

Section C (Essay Type Questions)

Each question carries 10 marks (2 out of 4)

Maximum marks that can be scored from this section is 20.

24. Write an essay on "Population data as aid to social,economic and healthy planning".

25. Compute Fisher's ideal index from the following data and show that it satisfies the factor reversal test.

Commodity	Price		Quantity	
	Base year	Current year	Base year	Current year
A	5	8	10	1
B	6	24	18	3
C	8	11	8	1
D	3	12	6	4

26. Write an essay on "Problems in the Construction of Index Numbers".

27. Explain the following:

    (a) Crude Birth Rate.
    (b) General Fertility Rate.
    (c) Total Fertility Rate.
    (d) Net Reproduction Rate.
    (e) Gross reproduction Rate.

MODEL QUESTION PAPER

THIRD SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA6IB11 - STATISTICAL QUALITY CONTROL

Time : Two Hours and a Half Maximum : 80 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 25.

1. What are the causes of variation in quality control?
2. Explain the need for quality control techniques in production.
3. Describe C chart
4. How to read control chart?
5. Explain the usefulness of R-chart. When s-chart used in place of R-chart.
6. What are the difference between defects and defectives?
7. A p-chart indicates that the average is 0.2. If 50 items are inspected each day, what is the probability of detecting a shift of 0.04 on the first day after the shift.
8. What is meant by specification limits and control limits.
9. What is the significance of OC curve?
10. Distinguish between consumer's risk and producer's risk.
11. Define LTPD.
12. What are ASN and ATI for the single sampling plan?
13. Explain how you will draw the O.C. curve for a single sampling plan with sample size 5 if the acceptance number $c = 2$ assuming the lot size to be large.
14. An inspector with 80 percent efficiency uses the sampling plan $n = 100$, $c = 1$. Plot the effective O.C. curve.
15. What you mean by sequential sampling plan for attributes.

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 35.

16. What is meant by a control charts? Explain the applications of these charts.

17. Describe procedure or drawing X bar and R charts.

18. A machine is set to deliver the packets of a given weight. Ten samples of size five each were examined and the following results were obtained:

Sample No.	:	1	2	3	4	5	6	7	8	9	10
Mean	:	43	49	37	44	45	37	51	46	43	47
Range	:	5	6	5	7	7	4	8	6	4	6

Calculate the values for the central line and the control limits for the mean chart and range chart. Comment on the state of control.

19. Explain AQL and ASN.

20. Explain the main control charts for attributes and obtain their control limits.

21. What do you understand by acceptance sampling procedure? State its uses giving illustrations.

22. What are single sampling plan and double sampling plan? Discuss the relative merits and demerits of single and double sampling plans.

23. What is sampling inspection? Distinguish between the rectifying and the non-rectifying types.

Section C(Essay Type Questions)

Each questions carries 10 marks (2 out of 4)

Maximum marks that can be scored from this section is 20.

24. From a lot consisting of 2,200 items, a sample of size 225 is taken. If it contains 14 or less defectives, the lot is accepted otherwise rejected. Plot O.C., ATI, and AOQL curves.

25. Draw the OC curve of the single sampling plan showing the consumers and producers risk.

26. Plot the O.C. curves for single sampling plan where $N = 5000;\ n = 100;\ c = 1, 2, 3.$ Assuming $P_c = 0.10$, determine the lot tolerance fraction defective. Also obtain AOQL.

27. Describe double sampling in detail.

* * * * *

MODEL QUESTION PAPER
SIXTH SEMESTER INTEGRATED M.Sc. S EXAMINATION
(CBCSS-UG)
Statistics
STA6IB12 : Basic Regression Analysis

Time: Two and half Hours Maximum: 80 Marks.

Section A (Short Answer Type Questions)
Each question carries 2 marks
Maximum marks that can be scored from this section is 25.

1. Define response variable and explanatory variable.
2. What are the assumptions about the explanatory variable?
3. What is meant by multicolleniarity?
4. Explain the terms predicted values and standard errors in multiple regression model shortly.
5. Explain the role of residual plots in regression analysis..
6. Find the least square estimates of regression coefficient in simple linear regression model.
7. Explain Gauss markov theorem shortly.
8. Write about multiple correlation coefficient briefly.
9. What is the importance of scatter diagram in regression analysis?
10. Explain the effects of multicolleniarity.
11. Write a note on hetroscedasticity.
12. Write about confidence intervals on slope, intercept and variance in simple linear regression model.
13. What are the uses of regression analysis?
14. Explain backward elimination procedure briefly.
15. Explain the outliers and explain the effect on regression analysis.

Section B (Paragraph /Problem Type Questions)
Each question carries 5 marks
Maximum marks that can be scored from this section is 35.

16. Explain hypothesis testing of slope and intercept in simple linear regression model.
17. Distinguish between simple linear regression model and multiple linear regression model.
18. Describe multiple linear regression models. Estimate the parameters in multiple linear regression models.

19. Explain the concept of residual analysis.
20. What are the properties of least square estimates in multiple regression model.
21. Explain forward selection procedure and stepwise method.
22. Estimate the variance of parameters in simple linear regression model.
23. Fit a simple regression model for the following data and interpret the result.

X	1	2	3	4	5	6	7	8
Y	2	3.2	3.8	4.5	5.6	5.7	7.6	10.1

Section C (Essay Type Questions)

Each question carries 10 marks(2 out of 4)
Maximum marks that can be scored from this section is 20.

24. Explain test procedure for significance of equality of regression coefficients.
25. Explain detection of multicolleniarity and it's remedial measures.
26. Derive the least squares estimate of simple linear regression model and show that they are unbiased.
27. State and prove the properties of least square estimates in a multiple linear regression model.

****

MODEL QUESTION PAPER

THIRD SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA6IB13 - OPERATIONS RESEARCH

Time : Two Hours and a Half Maximum : 80 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 25.

1. Define (i) Basic variable (ii) Degenerate basic solution.
2. Write down the standard form a L.P.P.
3. Define slack and surplus variables.
4. Distinguish between feasible solution and optimal feasible solution.
5. Formulate dual of the L.P.P:
   Maximize $z = 4x_1 + 2x_2$ subject to the constraints: $x_1 + x_2 \geq 3$, $x_1 - x_2 \geq 2$; $x_1 \geq 0$ and $x_2 \geq 0$.
6. Define artificial variable? When do we use it?
7. Explain transportation problem and show that it can be considered as an LPP.
8. Define degeneracy in transportation problem.
9. What is meant by North-West corner rule in LPP
10. Give the mathematical formulation of an assignment problem.
11. Give an example of a sequencing problem.
12. What are the characteristics of a game.
13. Explain the concept of value of a game.
14. Define saddle point. Is it necessary that a game should always possess a saddle point.
15. How will you solve the sequencing of n jobs on three machines

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 35.

16. Describe LPP. Write some of its applications.
17. Explain the graphical method for solving LPP.

18. What is balanced transportation problem? What are applications?

19. Describe Vogel's approximation method for solving transportation problem.

20. Explain assignment problem. Describe any method to solve it.

21. Explain Hungarian algorithm

22. Explain the method of solving a zero-sum two-person game as a linear programming problem.

23. A book binder has one printing press, one binding machine, and the manuscripts of a number of different books. The time required to perform the printing and binding operations for each book is shown below. Determine the order in which books should be processed, in order to minimize the total time required to turn out all the books:

Book	:	1	2	3	4	5	6
Printing time (hrs)	:	30	120	50	20	90	100
Binding time (hrs)	:	80	100	90	60	30	10

Section C(Essay Type Questions)

Each questions carries 10 marks (2 out of 4)

Maximum marks that can be scored from this section is 20.

24. What are the computational procedures of dual simplex method, explain with an example?

25. Maximize $z = 22x_1 + 30x_2 + 25x_3$ subject to constraints :

$$\begin{aligned} 2x_1 + 2x_2 &\leq 100, \\ 2x_1 + x_2 + x_3 &\leq 100 \\ x_1 + 2x_2 + 2x_3 &\leq 100; \quad x_1,\ x_2,\ x_3 \geq 0. \end{aligned}$$

26. Find the initial basic feasible solution of the following transportation problem. There are four origins three destinations. The availabilities are 9,10,8,7 and the requirements are 17,10,7 respectively.

	A	B	C
D	2	3	2
E	1	3	4
F	2	3	1
G	2	4	3

27. Solve the following 2 x 2 game graphically:

	B_1	B_2	B_3	B_4
A_1	2	1	0	-2
A_2	1	0	3	2

* * * * *

MODEL QUESTION PAPER

THIRD SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA6IB14 - ANALYTICAL TOOLS FOR STATISTICS

Time : Two Hours and a Half Maximum : 80 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 25.

1. Show that every convergent sequence is bounded
2. Write any two properties of real numbers.
3. State Taylor's theorem.
4. What is nested intervals?
5. write triangular inequality for two variables.
6. Identify the sixth term in the sequence {7,26,63,124,215,. . . }.
7. Discuss the relationship between continuity and differentiability of a function.
8. State sufficient condition for uniform continuity.
9. Show that $f(z) = Re(z)$ is nowhere differentiable.
10. Define harmonic function and give an example.
11. Obtain Taylor series representation for $\frac{1}{z}$ about $z = 1$.
12. State Laurent's theorem.
13. Define residue of f at the isolated singular point z_0.
14. Identify poles of $\frac{2z}{z^2+1}$.
15. Define essential singularity and give an example.

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 35.

16. Show every bounded sequence of real numbers has a convergent subsequence.
17. Establish Rolle's theorem.
18. State and prove intermediate value theorem

19. State and prove necessary condition for differentiability.

20. Evaluate $\int_C f(z)dz$ where $f(z) = x^2 + i3xy$ and C is the line segment joining $1+i$ to the point $2-i$.

21. State and prove Cauchy's inequality.

22. State and prove Cauchy's residue theorem.

23. Evaluate $\int_{|z|=\frac{3}{2}} \frac{z^4-3z^2+6}{(z+1)^3}dz$.

Section C(Essay Type Questions)

Each questions carries 10 marks (2 out of 4)

Maximum marks that can be scored from this section is 20.

24. State and prove Cauchy's root test.

25. State and prove principle of Mathematical induction.

26. Let $u(x,y) = 4xy - x^3 + 3xy^2$. Show that u is harmonic function. Find its harmonic conjugate and hence find analytic function $f(z)$ with u real part.

27. State and prove Cauchy's integral formula.

* * * * *

MODEL QUESTION PAPER

THIRD SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA3IA11 - BASIC NUMERICAL METHODS

Time : Two Hours and a Half Maximum : 80 Marks

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 25.

1. What is a matrix?
2. What is compound interest?
3. What is a linear equation?
4. What do you mean by a sequence?
5. What is EMI?
6. What do you mean by deferred perpetuity?
7. What do you mean by dispersion?
8. What do you mean by singular and nonsingular matrix?
9. Define Arithmetic Mean?
10. What do you mean by continuous series?
11. What is negative skewness?
12. Find the next number in the sequences 1,4,9,16,25, x.
13. What is range?
14. What do you mean by standard deviation?
15. What is Geometric Mean?

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 35.

16. Solve $4(x-1)+1=5(2x+1)-6$.
17. What is $\begin{bmatrix} 2 & -3 \\ -4 & 2 \end{bmatrix}$- $\begin{bmatrix} -1 & -5 \\ 3 & -2 \end{bmatrix}$ =?
18. At what rate percent per annum will a sum of money double in 8 years?
19. Find out the median from the following data:

Age	10	5	7	12	8
No. of students	15	20	15	28	12

20. Find two natural numbers whose sum is 27 and product is 182.

21. The first term of an Arithmetic Progression is 15 and the last term is 85. If the sum of all terms is 750, what is the 6th term ?

22. What is mean deviation? What are its merits and limitations?

23. A bank offers 5% compound interest calculated on half-yearly basis. A customer deposits Rs. 1,600 each on 1st January and 1st July of a year. Calculate the amount he would have gained by way of interest at the end of the year.

Section C(Essay Type Questions)

Each questions carries 10 marks (2 out of 4)

Maximum marks that can be scored from this section is 20.

24. What are the major measures of central tendency? List out the merits and limitations of each measures.

25. A man constructed his house by taking a home loan of Rs. 15,00,000. He is asked to repay the loan in 5 years and rate of interest is 13% p.a. Calculate EMI.

26. Find a solution to the following system by using Cramer's rule:

$$x - 2y + 3z = 9; \quad -x + 3y - z = -6; \quad 2x - 5y + 5z = 17.$$

27. Find mean, median and mode of the following data:

Wages	0-50	50-100	100-150	150-200	200-250	250-300	300-350
No. of Employees	2	3	5	6	5	3	1

* * * * *

MODEL QUESTION PAPER

THIRD SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA3IA12 - PROFESSIONAL BUSINESS SKILLS

Time : Two Hours and a Half Maximum : 80 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 25.

1. What is Green Computing?
2. What do you mean by transactional written communication?
3. What is virtual classroom?
4. What do you mean by Hybrid education?
5. What is a Webinar?
6. Briefly describe any four popular e-commerce websites.
7. What is White Hat hacking?
8. What is Google Scholar?
9. What are the disadvantages of new generation computers?
10. What you mean by cyber-vandalism?
11. Write a short note on IT ACT 2000.
12. What do you mean by soft skills?
13. What is a virtual assistant?
14. What is video marketing?
15. What is Microlearning?

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 35.

16. What B2C business model? What are the advantages and challenges of B2C model?
17. What are the various social and ethical issues concerned with IT in society?
18. List out the similarities and differences between formal online courses and MOOCs.

19. What are the dimensions of data quality standards? What are the benefits of quality data?

20. Explain the types of PPC marketing ads. What are its advantages and limitations?

21. What are various benefits of using videos in e-Learning?

22. What is written communication? List out the advantages and limitations of written communication.

23. What are advantages and limitations of e-Books?

Section C(Essay Type Questions)

Each questions carries 10 marks (2 out of 4)

Maximum marks that can be scored from this section is 20.

24. What is digital marketing? What are its advantages and disadvantages? Also explain the various types of digital marketing?

25. What are cyber crimes? Explain the major categories of cyber crimes. How such crimes can be prevented or avoided?

26. What do you mean by business data analytics? What are its components? Explain the importance of business data analytics.

27. How online learning is different from classroom learning? What are the advantages and disadvantages of online learning?

* * * * *

MODEL QUESTION PAPER

FOURTH SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA4IA13 - ENTREPRENEURSHIP DEVELOPMENT

Time : Two Hours and a Half Maximum : 80 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 25.

1. Define Entrepreneur
2. Who is a Pure Entrepreneur?
3. Mention four social factors affecting entrepreneurial growth.
4. What do you mean by Entrepreneurship?
5. Mechanism of Medium-term finance to entrepreneur.
6. What is EDP?
7. What is MUN?
8. What is KVIC?
9. What is NIESBUD?
10. What is BOUNTY?
11. What is Functional Industrial Estates?
12. What is Seed Capital?
13. What is Economic feasibility?
14. What is social cost benefit analysis?
15. Four factors affecting location decision?

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 35.

16. Distinguish between Entrepreneur vs Intrapreneur.
17. Write down the importance of Entrepreneurship Development.
18. Mention the objectives of district industries center.

19. Briefly explain NSIC.

20. Write down the need for Incentives and Subsidies.

21. List out the objectives of industrial estates.

22. What is Single Window System?

23. Explain the importance of MSMEs.

Section C(Essay Type Questions)

Each questions carries 10 marks (2 out of 4)

Maximum marks that can be scored from this section is 20.

24. Explain the different classification of Entrepreneur.

25. Narrate the problems of women entrepreneurship and its remedies.

26. Illustrate the Phases or Process of EDP.

27. Narrate the format for Project report for a New Business.

* * * * *

MODEL QUESTION PAPER

FOURTH SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics

STA4IA14 - BANKING AND INSURANCE

Time : Two Hours and a Half Maximum : 80 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 25.

1. What do you mean by a foreign bank?
2. Expand NEFT and RTGS.
3. What is noting in negotiable instruments?
4. What is meant by financial inclusion?
5. What do you mean by repo rate?
6. What is meant no-claim bonus?
7. What do you mean by traveller's cheque?
8. What you mean by a debit card?
9. List out any four insurance companies in India.
10. What is meant by dishonour of cheque?
11. Who is an Insurer?
12. What is Cash Reserve Ratio?
13. What do you mean by bank overdraft?
14. What is tele-banking?
15. What do you mean by social insurance?

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 35.

16. List out the characteristics or features of a bank.
17. What are the various advantages or benefits of having an insurance?
18. What are the various classes or types of endorsements?

19. Write a short note on SWIFT.

20. List out the differences between a cheque and a bank demand draft.

21. List out any five advantages and limitations of credit cards?

22. Distinguish between traditional banking and online or internet banking.

23. What are the various types of life insurance policies?

Section C(Essay Type Questions)

Each questions carries 10 marks (2 out of 4)

Maximum marks that can be scored from this section is 20.

24. What are various types of bank accounts? Also explain the steps or procedures to be followed for opening a bank account.

25. What do you mean by CORE banking? Explain its advantages and limitations.

26. What is an insurance? What are the characteristics of insurance?

27. What do you mean by a bank? Explain in detail the various functions of commercial banks.

* * * * *

MODEL QUESTION PAPER

FIFTH SEMESTER INTEGRATED M.Sc. EXAMINATION

Statistics -Open Course

STA5ID02 - BASICS OF STATISTICAL INFERENCES

Time : Two Hours Maximum : 60 Marks

(2020 Admission)

Section A (Short Answer Type Questions)

Each questions carries 2 marks

Maximum marks that can be scored from this section is 20.

1. What are the desirable properties of an estimator.
2. Define unbiasedness of an estimator
3. Distinguish between simple and composite hypothesis with example.
4. Discuss the two types of errors associated with hypothesis testing.
5. Distinguish between point estimate and interval estimate
6. Discuss the major applications of t-test.
7. Write down the test statistic used in testing of the proportion of success of a population.
8. It is believed that life time of an electronic device is 500 hours. A sample of 100 items yielded mean 450 and variance 16. Based on this data test whether the belief is true.
9. The standard deviation of two samples of sizes 10 and 14 from two normal populations are 3.5 and 3.0 respectively. Examine whether the standard deviation of the first population is more than that of the second population.
10. Explain paired t-test
11. What are the merits and demerits of non-parametric tests?
12. State the basic assumptions of analysis of variance.

Section B(Paragraph/Problem Type Questions)

Each questions carries 5 marks

Maximum marks that can be scored from this section is 30.

13. What do you mean by significance level, power and critical region of a test?
14. A random sample of 50 Mathematics grades showed a mean of 75 and a standard deviation of 10. What are the 95% confidence limits for the population mean?
15. What are the properties of Chi-square distribution?

16. The following data was obtained in an investigation about an effect of vaccination for small pox:

	Vaccinated	Not Vaccinated	Total
Attacked by small pox	3	12	15
Not attacked	8	5	13
Total	11	17	28

Examine whether vaccination is effective in preventing small pox.

17. Explain the procedure for testing equality of two population means.

18. Explain the procedure of Wilcoxon Signed Rank test.

19. write a short note on ANOVA

Section C(Essay Type Questions)

Each questions carries 10 marks (1 out of 2)

Maximum marks that can be scored from this section is 10.

20. A systematic sample of 100 pages was taken from the Oxford Dictionary and the observed frequency distribution of foreign words per pages was found to be as follows:

No. of foreign words per page (X)	:	0	1	2	3	4	5
Frequency	:	42	34	12	7	4	1

Calculate the expected frequencies and test for goodness of fit using Poisson distribution.

21. Three varieties of wheat were sown in 4 plots each and the following yields in quintals per acre were obtained.

	Type I	Type II	Type III
Plot I	10	9	4
Plot II	6	7	7
Plot III	7	7	7
Plot IV	9	5	6

Find out whether there is significant difference between the mean yield and plots of the three varieties.

* * * * *

MODEL QUESTION PAPER
SEVENTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics
STA7IB15-MATHEMATICAL METHODS FOR STATISTICS-I

Time:3 hours Maximum : 80 Marks

Section A

Answer any **four** questions.
Each question carries 4 marks

1. (a) If $f(.)$ is bounded function and $\alpha(.)$ be a non-decreasing function on$[a, b]$, show that the lower Reimann-Stieltjes integral does not exceed the upper Reimann-Stieltjes integral.

   (b) Define beta and gamma functions. Show that beta function is symmetrical in its constants.

   (c) Show that a monotonically increasing sequence which is not bounded above diverges.

   (d) What is an alternating series? State the conditions for testing the convergence of an alternating series.

   (e) State the conditions to be satisfied by the generalized inverse of a matrix.

   (f) Explain the standard form of a system of linear equations.

   (g) Define complex n -space, and vector multiplication on complex n -space.

   (h) State the conditions for linear dependence of vectors in a vector space. Give an example.

(4x4=16 marks)

Section B

Answer **either** part-a **or** part-b.
Each question carries 16 marks.

2. (a) i. State and prove second mean value theorem.

   ii. if $f(.)$ is Reimann-Stieltjes integrable on $[a, b]$ with respect to a monotonically non-decreasing function $\alpha(.)$ on $[a, b]$ and if $|f(x)| \leq K$, find the upper bound for $|\int_a^b f(x)d\alpha(x)|$.

   OR

   (b) i. A. Let $S(P, f, \alpha)$ be the Reimann-Stieltjes sum, where P is the partition of $[a, b]$, f is the bounded function on $[a, b]$ and α is the monotonic non-decreasing function on $[a, b]$. Prove that, $\lim_{||P||\to 0} S(P, f, \alpha)$ exists and is equal to $\int_a^b f(x)d\alpha(x)$.

   B. Examine the convergence of $\int_0^2 \frac{1}{2x-x^2}dx$.

3. (a) i. Define sequence. Show that a sequence can not converge to more than one limit.

   ii. Prove that a series of positive terms either converges or diverges but never oscillates.

   OR

   (b) i. State and prove Cauchy's general principle of convergence for series.

ii. Justify the following statement citing two illustartions: A bounded sequence need not be convergent.

4. (a) i. Define characteristic polynomial and minimal polynomial. State any two properties connecting both the polynomials.

ii. If $A = \begin{pmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{pmatrix}$, verify whether it is diagonalizable. If yes, find the matrix P such that $P^{-1}AP$ is diagonal.

OR

(b) Show that rank of the generalized inverse of a mathrix A equals the rank of A.

(c) Find the characteristic polynomial of $A = \begin{pmatrix} 3 & -1 & 1 \\ 7 & -5 & 1 \\ 6 & -6 & 2 \end{pmatrix}$. Find the algebraic and geometric multiplicities of one of the eigen values of the matrix.

5. (a) Define matrix representation of a liner operator $T : V \to V$ relative to a basis.

(b) Define subspace of a vector space. Show that the intersection of any number of subspaces of a vector space V is a subspace of V.

(6+10=16 marks)

Or

(a) Define linear span. Show that the vectors $(1, 1, 1), (1, 2, 3)$ and $(1, 5, 8)$ span R^3.

(b) Let $F : R^2 \to R^2$ be the linear operator defined by $F(x, y) = (2x + 3y, 4x - 5y)$. Find the matrix representation of F relative to the basis $S = \{u_1, u_2\} = \{(1, -2), (2, -5)\}$.

(8+8=16 marks)

* * * * *

MODEL QUESTION PAPER

SEVENTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics

STA7IB16 - MATHEMATICAL METHODS FOR STATISTICS -II

Time: 3 hrs Max Marks:80.

Section A

(Answer any **four** questions. Each carries **4** marks)

I. (i) Distinguish between Sigma field and Borel sigma field.

(ii) What is a measure? Define finite measure and sigma-finite measure.

(iii) What is Borel measurable function? How does it differ from Lebesgue measurable function?

(iv) Define measurable space. State the conditions for a function defined on a measurable space to be simple.

(v) State Randon-Nikodym theorem. What is its significance?

(vi) Define signed measures and absolute continuity of one measure with respect to another.

(vii) Examine the equality of the partial derivatives f_{xy} and f_{yx} of $f(x,y) = x^3 + e^{xy^2}$

(viii) State Taylor's theorem on partial derivatives.

(4x4=16 marks)

Section B

Answer either part-A or part-B of all questions.
Each question carries **16** marks.

II.A) (a) Define monotone class of sets. Show that sigma field is a monotone class and a monotone field is a sigma field.

(b) Let $\{A_n, n = 1, 2, ..\}$ be a finite, disjoint class of sets in P, each contained in a given set A_o, such that $A_o \subset \bigcup_{i=1}^{\infty} A_i$, where P is the class of all bounded, left closed and right opened intervals, prove that $\mu(A_o) \leq \sum_{i=1}^{\infty} \mu(A_i)$.

(8+8=16 marks)

Or

B) (a) Let μ be a finite, non-negative and additive set function on a sigma field. If μ is either continuous from below at every set in the sigma field or continuous from above at 0, show that μ is a measure.

(b) Define outer measure. State any two properties on outer measures.

(8+8=16 marks)

III.A) (a) Define fundamental in measure and convergence in measure of a sequence of measurable functions. State any two properties relating to convergence of a sequence of measurable functions.

(b) State and prove Fatou's lemma.

(8+8=16 marks)

Or

B) (a) Let $\{f_n\}$ be a sequence of measurable functions which converges in measure to f and to g. Show that $\{f_n\}$ is fundamental in measure and $f = g$ almost everywhere.

(b) Let f be measurable, g be integrable and $|f| \leq |g|$ a.e, Prove that f is integrable.

(8+8=16 marks)

IV.A) (a) Let μ be a signed measure and ν be a finite signed measure, such that $\nu << \mu$. Prove that for every $\epsilon > 0$, there is a $\delta > 0$, such that $|\nu|(A) < \epsilon$ for every measurable set A, for which $|\mu|(A) < \delta$.

(b) Define product space and state Fubini's theorem.

(10+6=16 marks)

Or

B) (a) What are double and iterated integrals?

(b) Show that every section of a measurable set is a measurable set.

(6+10=16 marks)

V. A (a) What are partial derivatives of a function of two variables? State partial derivatives of higher order.

(b) Show that

$$f(x,y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

is continuous at the origin.

OR

(a) Define limit, continuity and differentiability of a function $f(x,y)$ at (x_0, y_0).

(b) Define extreme values of a function. Show that $f(x,y) = x^4 + x^2y + y^2$ has a minimum at (0,0).

(8+8=16 marks)

* * * * *

MODEL QUESTION PAPER

SEVENTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics

STA7IB17 - PROBABILITY THEORY-I

Time: 3 hrs Max Marks:80.

Section A

(Answer any **four** questions. Each carries **4** marks)

I. (i) Define indicator random variable and mention any two properties.

(ii) What is induced probability space? Explain with an illustration.

(iii) Show that the expected value of a bounded variate X always exists.

(iv) State Holder's inequality and mention its relation with Schwarz inequality.

(v) State Cauchy's criterion of convergence.

(vi) Define convergence almost surely and mutual convergence.

(vii) Define tail events. State Kolmogorov's 0-1 law.

(viii) Define independence of events. State Borel-Cantelli lemma.

(4x4=16 marks)

Section B

Answer either part-A or part-B of all questions.
Each question carries **16** marks.

II.A) (a) Show that the probability function defined on all intervals of the form $(a, b] \subseteq \mathbf{R}$ defines an extension uniquely to the minimal field containing all the intervals.

(b) A function $F(x, y)$ of two variate X, Y is defined by $F(x, y) = 1$, if $x + y \geq 0$; $F(x, y) = 1$, if $x + y < 0$. Examine whether $F(x, y)$ can be a distribution function of some two-dimensional random variable.

(8+8=16 marks)

Or

B) (a) Distinguish probability space and induced probability space.

(b) State and prove Jordan decomposition theorem.

(8+8=16 marks)

III.A) (a) Define Gamma distribution. Obtain its moment generating function and hence find $E(X)$.

(b) Show that $E\left(\frac{|X|^r}{1+|X|^r}\right) - \frac{a^r}{1+a^r} \leq P\{|X| \geq a\} \leq \frac{1+a^r}{a^r} E\left(\frac{|X|^r}{1+|X|^r}\right)$

(8+8=16 marks)

Or

B) (a) If $E(X^r)$ exists, then show that $E(X^t)$ need not exist if $t > r$.

(b) State and prove C_r-inequality.

(8+8=16 marks)

IV.A) (a) Define convergence in probability. Show that

$$X_n \xrightarrow{P} 0 \text{ iff } E\left(\frac{|X|^r}{1+|X|^r}\right) \to 0, \text{ as } n \to \infty.$$

(b) Let $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Show that $X_nY_n \xrightarrow{P} XY$.

(10+6=16 marks)

Or

B) (a) Define convergence in distribution. Show that $X_n \xrightarrow{c}$ implies that $F_n(x) \to 0$ for $x < c$, $F_n(x) \to 1$ for $x \geq c$, and conversely.

(b) Define convergence in mean square. Verify a sequence of variates $\{X_n\}$ defined with $P(X_n = 0) = 1 - (1/n^2)$, $P(X_n = n) = 1/n^2, n = 1, 2, 3, ...$ is convergent in mean square.

(10+6=16 marks)

V.A) (a) What is lattice distribution? State inversion formula for lattice distribution.

(b) If $X \sim N(\mu, \sigma^2), Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$, show that $X = \mu + \sigma z$ has its characteristic function given by $\phi_X(t) = exp(it\mu - (t^2\sigma^2)/2)$.

(8+8=16 marks)

Or

B) (a) State and prove Borel 0-1 criterion.

(b) Find the distribution for which characteristic function is $\sigma(t) = e^{-|t|}, -\infty < i < \infty$.

(10+6=16 marks)

* * * * *

MODEL QUESTION PAPER
SEVENTH SEMESTER INTEGRATED M.Sc EXAMINATION
Statistics
STA7IB18- DISTRIBUTION THEORY

Time: Three Hours **Maximum: 80 Marks**

Section A

Answer any four questions.
Each question carries 4 marks.

I (i) Define discrete uniform distribution and derive its mean and variance.

(ii) Write a short note on hypergeometric distribution.

(iii) Derive the M.G.F of gamma distribution.

(iv) Define beta distribution of first kind and second kind. Also state their properties.

(v) Write a short note on bivariate normal distribution.

(vi) Let (X, Y) be a random vector continuous type with p.d.f f. Obtain the p.d.f of:
(a) $U = X - Y$ and (b) $V = \frac{X}{Y}$.

(vii) Define F distribution and state its important properties.

(viii) Define Chi-square distribution and list its applications.

$(4 \times 4) = 16$ *marks*

Section B

Answer either part -A or part -B of all questions.
Each question carries 16 marks.

II A) (a) Define logarithmic distribution and derive its M.G.F.

(b) Write a short note on hypergeometric and its association with other distributions.

$(8 + 8) = 16$ *marks*

Or

B) (a) Derive the r^{th} order moment of discrete uniform distribution.

(b) Explain the multinomial distribution and its marginal distributions.

$(8 + 8) = 16$ *marks*

III A) (a) Define the r^{th} order moment of Pareto distribution.

(b) Explain the role and significance of transformed distribution with illustration.

$(8+8) = 16\ marks$

Or

B) (a) Write the statistical properties of bivariate normal distribution.

(b) Derive the characteristic function of generalized Laplace distribution.

$(8+8) = 16\ marks$

IV A) (a) Let (X, Y) be a bivariate normal random vector with parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$ and ρ. Let $U_1 = \sqrt{X^2 + Y^2}$ and $U_2 = \frac{X}{Y}$. Find the joint density of (U_1, U_2) and find the marginal density of U_1 and U_2.

(b) Derive the p.d.f of the median and mid-range of order statistics from a random sample of size n.

$(8+8) = 16\ marks$

Or

B) (a) Let X_1, X_2, X_3 be i.i.d. random variables with common density function

$$f(x) = \left\{ \begin{array}{ll} 0, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{array} \right\}$$

Find the p.d.f. of $Y = X_1 + X_2 + X_3$.

(b) Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statistics of a random sample from exponential distribution with parameter $\lambda = 1$. Compute the probability of an event $Y_4 \geq 3$.

$(8+8) = 16\ marks$

V A) (a) Obtain the mean and variance of t distribution.

(b) Show that non-sentral Chi-square satisfies additive property.

$(8+8) = 16\ marks$

Or

B) (a) Derive the pdf of non-central F distribution.

(b) Write in detail about the interrelation between F and χ^2 distribution.

$(8+8) = 16\ marks$

* * * * *

MODEL QUESTION PAPER
SEVENTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics
STA7IB19- ADVANCED SAMPLING THEORY

Time: Three Hours **Maximum: 80 Marks**

Section A

Answer any four questions.
Each question carries 4 marks.

I (i) State the principle of statistical regularity and indicate its importance.

(ii) What are the advantages of sampling methods.

(iii) What are the advantages of PPS sampling.

(iv) Explain Lahiris method.

(v) What are regression estimators.

(vi) Obtain the expression for the bias of ratio estimators in the case of simple random sampling.

(vii) What is systematic random sampling.

(viii) What are the advantages of cluster sampling.

$(4 \times 4) = 16$ *marks*

Section B

Answer either part -A or part -B of all questions.
Each question carries 16 marks.

II A) (a) Show that in case of simple random sampling the probability that a specified unit of the population being selected in any given draw is equal to the probability of its being selected at the first draw.

(b) Show that in simple random sampling s^2 is an unbiased estimator of S^2. Also obtain an unbiased estimator of the variance of the $\overline{y}$ in simple random sampling without replacement.

$(8 + 8) = 16$ *marks*

Or

B) (a) Discuss the methods for the estimation of sample size in simple random sampling.

(b) Explain optimum allocation.

$(8+8) = 16\ marks$

III A) (a) Distinguish between simple random sampling and PPS sampling.

(b) Explain Des Raj's ordered estimator and show that it is unbiased. Obtain its sampling variance.

$(8+8) = 16\ marks$

Or

B) (a) Distinguish between ordered and unordered estimators in PPS sampling.

(b) Define the Horwitz-Thompson estimator for population total in case of PPS sampling. Obtain the expression for its variance.

$(8+8) = 16\ marks$

IV A) (a) Compare the ratio estimator with mean per unit.

(b) What are unbiased ratio type estimators. Define Hartley and Ross estimator and show that it is unbiased. Obtain its sampling variance under simple random sampling without replacement.

$(8+8) = 16\ marks$

Or

B) (a) Obtain an expression for the approximate bias of the regression estimator and also obtain the large sample variance of the regression estimator.

(b) When the bias of the ratio estimator will be small.

$(8+8) = 16\ marks$

V A) (a) Explain the multi-stage sampling. What re the situations in which we can use it.

(b) Explain unequal cluster sampling and obtain the unbiased estimator of the mean. Also derive its variance.

$(8+8) = 16\ marks$

Or

B) (a) Explain the multi-phase sampling and mention its advantages. How it differs from multistage sampling.

(b) Explain various sources of non-sampling errors.

$(8+8) = 16\ marks$

* * * * *

MODEL QUESTION PAPER
EIGHTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics
STA8IB20-PROBABILITY THEORY-II

Time:3 hours Maximum : 80 Marks

Section A

Answer any **four** questions.
Each question carries 4 marks

1. (a) State and prove Hellys convergence theorem.

   (b) Show that characteristic function is real if and only if the distribution function is symmetric about zero.

   (c) For the sequence $\{X_n\}$ of independent random variables with $P\{X_k = \pm\sqrt{k}\} = \frac{1}{2}$, does the WLLN hold?

   (d) State Borel Cantelli lemma. Is the converse true? Justify your answer.

   (e) State and prove Kolmogorov three series theorem.

   (f) Show that Liapunov condition implies Lindeberg-Feller condition.

   (g) Define sub-Martingale and super-Martingale. Give examples.

   (h) For the sequence $\{X_n\}$ of independent random variables with $P\{X_k = \pm 2^k\} = \frac{1}{2}$, does the SLLN hold?

(4x4=16 marks)

Section B

Answer **either** part-a **or** part-b.
Each question carries 16 marks.

2. (a) i. State continuity theorem on characteristic function.

   ii. Show that characteristic function is uniformly continuous over R.

OR

   (b) i. State and prove Inversion theorem on characteristic function.

   ii. Find the probability density function corresponding to the characteristic function

$$\phi(t) = \begin{cases} 1 - |t| & \text{if } |t| \leq 1 \\ 0 & \text{other wise,} \end{cases}$$

3. (a) i. State and prove Helly-Bray lemma.

   ii. State and prove Scheffe's theorem.

OR

(b) i. Prove that probability of the tail events of a sequence of independent random variables is either 0 or 1

ii. State and prove Kolmogorov three series theorem.

4. (a) i. Establish Kolmogorov strong law of large numbers for a sequence of independent random variables.

ii. For the sequence $\{X_n\}$ of independent random variables with $P\{X_k = \pm 2^k\} = \frac{1}{2^{2k+1}}$ and $P\{X_k = 0\} = 1 - \frac{1}{2^{2k}}$, does SLLN hold?

OR

(b) i. State and prove Liapunov CLT.

ii. State Lindeberg-Feller form of CLT

5. (a) i. State and prove Kolmogorov inequality.

ii. Define Martingales. Show that if $\{X_n\}$ is a sequence of independent random variables with $E(X_n) = 0$ then $Y_n = \sum_{k=1}^{n} X_k$ is a Martingale sequence.

OR

(b) i. Define conditional expectation. State and prove its important properties.

ii. Define infinite divisibility. Examine the infinite divisibiity of Poisson distribution and Normal distribution.

(4x16=64 marks)

* * * * *

MODEL QUESTION PAPER
EIGTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics
STA8IB21- ESTIMATION THEORY

Time: Three Hours **Maximum: 80 Marks**

Section A

Answer any four questions.
Each question carries 4 marks.

I (i) Let X_1, X_2 be i.i.d. $P(\lambda)$ r.v.s. Show that $T = (X_1 + X_2)/3$ is sufficient but $T_1 = X_1 + 3X_2$ is not sufficient for λ.

(ii) What do you mean by a complete statistic. Give an example of a statistic which is sufficient but not complete.

(iii) Define exponential family of density. Show that $\{N(\mu, \sigma^2), \mu \epsilon R, \sigma > 0\}$ is an exponential family. Obtain a minimal sufficient statistic for (μ, σ^2).

(iv) Define CAN estimator. Examine whether sample median is a CAN estimator for μ in the Cauchy population $C(\mu, \sigma), \mu \epsilon R, \sigma > 0$.

(v) Obtain the moment estimators of α and β in the Gamma distribution $G(\alpha, \beta)$ based on a sample of size n from this population.

(vi) Find out the MLE of θ based on a sample of size n from $U(\theta, \theta + 2), \theta \epsilon R$.

(vii) Distinguish between Bayesian and fiducial intervals.

(viii) Let $X_1, X_2, ..., X_n$ be a random sample from $U(0, \theta)$. Obtain $(1 - \alpha)$ level UMA confidence interval for θ.

$(4 \times 4) = 16$ *marks*

Section B

Answer either part -A or part -B of all questions.
Each question carries 16 marks.

II A) (a) State and prove Cramer-Rao inequality.

(b) Examine whether MVB estimator for θ exists for the exponential population with mean θ. If so obtain the MVB estimator.

$(8 + 8) = 16$ *marks*

Or

B) (a) State and prove Rao-Blackwell theorem.

(b) Obtain UMVUE of $2\mu^2 + 3$ in $N(\mu, 1), \mu \epsilon R$ based on a random sample of size n.

$(8 + 8) = 16$ *marks*

III A) (a) State and prove invariance property of CAN estimators.

(b) Obtain the MLE and moment estimators of θ in $U(\theta, 2\theta), \theta > 0$ and examine whether these estimators are CAN.

$(8 + 8) = 16$ *marks*

Or

B) (a) If T is a consistent estimator of θ show that $1+2e^T$ is a consistent estimator of $1 + 2e^{\theta}$.

(b) Let $X_1, X_2, ..., X_n$ be a random sample from $N(\mu, \sigma^2)$. Show that $T = \overline{X} + S^2$ is a CAN estimator of $\mu + \sigma^2$.

$(8 + 8) = 16$ *marks*

IV A) (a) Explain maximum likelihood method of estimation. Show that under some regularity conditions to be stated MLE is a CAN estimator.

(b) Explain Bayes method of estimation. Obtain Bayes estimate of λ under squared error loss, in $P(\lambda), \lambda > 0$, when the prior distribution of λ is an exponential distribution with mean 5.

$(8 + 8) = 16$ *marks*

Or

B) (a) Let $X_1, X_2, ..., X_n$ be a random sample from $N(\mu, \sigma^2)$. Obtain Bayes estimator under squared error loss for the parameter μ when σ^2 is known and the prior distribution of μ is $N(\mu_0, \sigma_0^2)$.

(b) Let $X_1, X_2, ..., X_n$ be a random sample from $b(1, \theta), \theta \epsilon [1/3, 3/4]$. Find an MLE of θ if it exists.

$(8 + 8) = 16$ *marks*

V A) (a) What do you mean by large sample confidence interval. Obtain confidence interval for the $p_1 - p_2$ based on random samples from two independent binomial populations $b(n, p_1)$ and $b(m, p_2)$.

(b) Let X_1 and X_2 be two independent observations from the exponential population with pdf $f(x; \theta) = e^{-(x-\theta)},\ x > \theta$.
Let $Y = min(X_1, X_2)$. Find the confidence co-efficient of the interval $[Y - 1/2, Y + 1/2]$. $(8 + 8) = 16$ *marks*

Or

B) (a) Let $X_1, X_2, ..., X_n$ be a random sample from
$f_\theta(x) = \frac{1}{2\theta} exp(\frac{-|x|}{\theta}),\ x \epsilon R,\ \theta > 0$. Find the shortest length $(1 - \alpha)$ level confidence interval for θ, based on the sufficient statistics $\sum_{i=1}^{n} |X_i|$.

(b) Let $\overline{X}$ be the mean of a random sample from $N(\mu, 25)$. Find the sample size n such that $(\overline{X} - 2, \overline{X} + 2)$ is a .90 level confidence interval for μ.

$(8 + 8) = 16$ *marks*

* * * * *

MODEL QUESTION PAPER
EIGTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics
STA8IB22- ADVANCED DESIGN AND ANALYSIS OF EXPERIMENTS

Time: Three Hours **Maximum: 80 Marks**

Section A

Answer any four questions.
Each question carries 4 marks.

I (i) What are the Guidelines of Experimental Design. Also note two Applications of the Experimental Design.

(ii) Give a brief note on the Regression approach in ANOVA.

(iii) Explain the analysis of Completely Randomized Design.

(iv) What is Youden square Lattice Design?

(v) Explain the salient features of Fractional Factorial Design.

(vi) Construct a 2^6 design in blocks of 8 plots confounding ABC, ADE and BCDE.

(vii) Describe the effect components of a 2^3 design into seven mutually orthogonal contrast.

(viii) What are Multiple Responses?

$(4 \times 4) = 16$ *marks*

Section B

Answer either part -A or part -B of all questions.
Each question carries 16 marks.

II A) (a) What are the Principles of Design? Explain why RBD is much better than CRD in light of the principles of design

(b) Explain the process of choice of sample size and its importance in the experimental design.

$(8 + 8) = 16$ *marks*

Or

B) (a) Distinguish between Fixed effect and Random effect Model.

(b) Describe the concept of Two-way ANOVA with Interaction and its analysis

$(8 + 8) = 16$ *marks*

III A) (a) Describe the concept and analysis of Latin Square Design.

(b) Explain Graeco-Latin Square Design.

$(8+8) = 16\ marks$

Or

B) (a) Distinguish between BIBD and PBIBD.

(b) Note down the model for BIBD and give its analysis.

$(8+8) = 16\ marks$

IV A) (a) Describe a two-factor factorial design and give its analysis..

(b) Explain Yate's procedure of obtaining various effects and the sum of squares of a 2^k factorial experiments.

$(8+8) = 16\ marks$

Or

B) (a) Describe confounding with suitable example.

(b) Write down the confounded arrangement of a 3^3 design by confounding the Interactions ABC and BC into blocks. Identify other interactions of any which get confounded in our arrangement. Hence analyse the design if there are two replications of the same type of arrangement of the treatment.

$(8+8) = 16\ marks$

V A) (a) What are Response Surface Designs?

(b) Give some Experimental Designs for Fitting and Analysing Response Surface Designs.

$(8+8) = 16\ marks$

Or

B) (a) Explain ANCOVA.

(b) Develop the analysis of RBD with one concomitant variable.

$(8+8) = 16\ marks$

* * * * *

MODEL QUESTION PAPER
EIGTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics
STA8IB23- REGRESSION METHODS

Time: Three Hours **Maximum: 80 Marks**

Section A

Answer any four questions.
Each question carries 4 marks.

I (i) List out the properties of Least Square Estimates.

(ii) Define R^2 and adjusted R^2.

(iii) Describe Likelihood Ratio Test.

(iv) Define Prediction Intervals.

(v) Describe the concepts of lack of fit and pure error.

(vi) Write a short note on spline smoothing.

(vii) What is the need of normal probability plots.

(viii) What is Random Explanatory Variable. Discuss its effect in a Regression Model.

$(4 \times 4) = 16$ *marks*

Section B

Answer either part -A or part -B of all questions.
Each question carries 16 marks.

II A) (a) Describe the assumptions underlying the classical linear regression model.

(b) Prove that least square estimate and MLE of the parameters of normal linear regression model coincides.

$(8 + 8) = 16$ *marks*

Or

B) (a) Explain the methods of estimation in regression models with linear restrictions.

(b) Describe generalized least square method.

$(8 + 8) = 16$ *marks*

III A) (a) Explain the procedure for testing the significance of regression parameters in simple linear regression model.

(b) Find out the Confidence Interval for the slope and intercept in Simple Linear Regression Model.

$(8 + 8) = 16\ marks$

Or

B) (a) Define the Coefficient of Determination. How do you interpret it. Is it related in any way with Multiple Correlation Coefficient.

(b) What is the Confidence bands for the Regression Surface. How they used in linear regression.

$(8 + 8) = 16\ marks$

IV A) (a) Define polynomial regression and discuss its estimation methods.

(b) Describe the method of weighted least square for straight line regression.

$(8 + 8) = 16\ marks$

Or

B) (a) Describe the method of fitting orthogonal polynomials.

(b) Explain logistic regression model. What are its uses.

$(8 + 8) = 16\ marks$

V A) (a) Define collinearity. Discuss the method of Diagnosing it.

(b) Discuss the effect of outliers in a Regression Model.

$(8 + 8) = 16\ marks$

Or

B) (a) Describe how you use residuals for checking the adequacy of the model fitted.

(b) Distinguish between Ridge Regression and Principle Component Regression.

$(8 + 8) = 16\ marks$

* * * * *

MODEL QUESTION PAPER
NINTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics
STA9IB24- TESTING OF STATISTICAL HYPOTHESIS

Time:3 hours Maximum : 80 Marks

Section A

Answer any **four** questions.
Each question carries 4 marks

1. (a) Define most powerful test.

   (b) Does the Laplace family of pdf's $f_\theta(x) = \frac{1}{2}e^{-|x-\theta|}, -\infty < x < \infty, \theta \in R$ posses MLR property.

   (c) Define UMP unbiased test.

   (d) Define locality most powerful tesst.

   (e) Define ASN function. Explain its uses.

   (f) State fundamental identityof SPRT.

   (g) Distinguish between parametric and non parametric tests.

   (h) Define Kendall's tau.

(4x4=16 marks)

Section B

Answer **either** part-a **or** part-b.
Each question carries 16 marks.

2. (a) i. A sample of size one is taken from population with distribution P(λ). To test $H_0 : \lambda = 1$ against $H_1 : \lambda = 2$ consider the non-randomized test:

$$\phi(x) = \begin{cases} 1 & \text{if } x > 3 \\ 0 & \text{other wise.} \end{cases}$$

   Find the probabilities of type-I and type-II errors. If it is required to achieve a size equal to 0.05, how will you modify the test.

   ii. State and prove generalized Neyman-Pearson lemma.

OR

   (b) i. Define MLR property. Prove that U(0,θ) has MLR in $X_{(n)}$.

   ii. Define UMP test. Obtain the UMP test for testing $H_0 : M \leq M_0$ against $H_1 : M > M_0$ based on a single observation from hypergeometric distribution with p.m.f

$$f(x, M) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{x}} & \text{if } x = 0, 1, ..., M. \\ 0 & \text{other wise.} \end{cases}$$

3. (a) i. Define likelihood ratio test. Obtain the asymptotic distribution of the likelihood ratio test statistic.

ii. Obtain the likelihood ratio test for testing the equality of means of two normal populations with equal variances.

OR

(b) i. Define UMP test. Find a UMP test size α test for testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$ based on a sample of n observations from the population with p.m.f

$$f(x,\theta) = \begin{cases} \theta^x(1-\theta)^{1-x} & \text{if } x = 0,1; 0 < \theta < 1 \\ 0 & \text{other wise.} \end{cases}$$

ii. Briefly explain Union-Intersection and Intersection-Union tests.

4. (a) i. Explain Wilcoxon signed rank test.

ii. Explain Mann-Whitney-Wilcoxon test.

OR

(b) i. Explain the Chi-square test for homogeneity.

ii. Define median test. Derive its null distribution.

5. (a) i. Define Sequential Probability Ratio Test and derive the boundary values of it.

ii. Obtain the SPRT for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1, \theta_0 < \theta_1$ based on observations from $B(n,\theta)$ at strength (α,β).

OR

(b) i. Prove that the Sequential Probability Ratio Test terminates with probability one.

ii. Obtain the OC function corresponding to the Sequential Probability Ratio Test for testing $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1, \mu_0 < \mu_1$ based on observations from $N(\mu, \sigma^2)$ at strength (α,β) where σ^2 is known.

(4x16=64 marks)

* * * * *

MODEL QUESTION PAPER
NINTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics
STA9IB25- MULTIVARIATE ANALYSIS

Time: Three Hours **Maximum: 80 Marks**

Section A

Answer any four questions.
Each question carries 4 marks.

I (i) Differentiate between simple and partial correlations.

(ii) If $X \sim Np(\mu, \Sigma)$, derive the characteristic function of X.

(iii) State and prove the additive property of multivariate normal distribution.

(iv) State multivariate central limit theorem.

(v) Is Hotelling's T^2 is a generalization of the square of students-t statistic. Justify the answer.

(vi) Explain the concept of principal components.

(vii) What are canonical variates and canonical correlations.

(viii) Explain spericity test.

$(4 \times 4) = 16$ *marks*

Section B

Answer either part -A or part -B of all questions.
Each question carries 16 marks.

II A) (a) Define multivariate normal distribution. If $X \sim Np(\mu, \Sigma)$, find the distribution of $Y = CX$, where C is a non-singular matrix of order p.

(b) If $X \sim Np(\mu, \Sigma)$, then show that $Q = (X - \mu)\sum^{-1}(X - \mu) \sim \chi_{(p)}$. Explain the hypothesis test concerning the mean vector μ of $Np(\mu, \Sigma)$ where Σ is known.

$(8 + 8) = 16$ *marks*

Or

B) (a) Let $X \sim Np(\mu, \Sigma)$, obtain the MLE's of μ and Σ.

(b) Define one sample Hotelling's T^2 statistics. Show that Hotelling's T^2 statistic is invariant under non-singular transformation.

$(8+8) = 16$ *marks*

III A) (a) Derive the characteristic function of Wishart distribution.

(b) Prove that Wishart distribution is a generalization of $\sigma^2\chi^2$ distribution.

$(8+8) = 16$ *marks*

Or

B) (a) Define Wishart distribution. State and prove any three properties of Wishart distribution.

(b) Describe multivariate Fisher-Behren problem.

$(8+8) = 16$ *marks*

IV A) (a) Establish the relation between principal components and the eigen vectors of the variance covariance matrix.

(b) Derive the distribution of the sample correlation coefficient.

$(8+8) = 16$ *marks*

Or

B) (a) Let $X_1, X_2, ..., X_n$ i.i.d. random variables such that $X_i \sim Np(\mu, \Sigma)$, $i = 1, 2, ..., n$. Prove that the mean vector X and SP matrix A are independent.

(b) Derive the distribution of Hotelling T^2 statistic.

$(8+8) = 16$ *marks*

V A) (a) Let $X_1, X_2, ..., X_n$ be a random sample from $Np(\mu, \Sigma)$. Obtain the distribution of the distribution of the sample generalized variance.

(b) Explain Bayes classification rule.

$(8+8) = 16$ *marks*

Or

B) (a) Derive the linear discriminant function for classifying an observation between two multivariate normal distributions.

(b) Derive the test criterion to test the hypothesis that mean vectors of two multivariate normal populations are equal when they have same unknown covariance matrix.

$(8+8) = 16$ *marks*

* * * * *

MODEL QUESTION PAPER
NINTH SEMESTER INTEGRATED M.Sc. EXAMINATION
Statistics
STA9IB26- STOCHASTIC PROCESSES

Time: Three Hours **Maximum: 80 Marks**

Section A

Answer any four questions.
Each question carries 4 marks.

I (i) Explain different types stochastic processes based on nature of state space and index set. Give one example for each.

(ii) State and prove the Chapman- Kolmogrov equations.

(iii) Define Poisson process. Obtain its relation with uniform distribution.

(iv) Define Renewal reward process and Regenerative process.

(v) Obtain the integral equations satisfied by the renewal function of a renewal process.

(vi) Explain gamblers ruin problem.

(vii) Explain the characteristics of a Queueing system.

(viii) Define weak stationary process. Give an example.

$(4 \times 4) = 16$ *marks*

Section B

Answer either part -A or part -B of all questions.
Each question carries 10 marks.

II A) (a) Explain the concept of communication between the states of a Markov chain. Prove that communication is a class property.

(b) Define transient state and recurrent state. Show that all the states of an irreducible finite chain are positive recurrent.

$(8 + 8) = 16$ *marks*

Or

B) (a) Explain the concept of steady state distribution. If P is the TPM of an ergodic chain and π is its stationary distribution show that $\pi = \pi P$.

(b) Define stationary distribution of a Markov chain. Classify the states of the Markov chain whose TPM is given below.

$$\begin{bmatrix} 1/3 & 0 & 2/3 & 0 \\ 0 & 1/4 & 3/4 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \end{bmatrix}$$

$(8+8) = 16\ marks$

III A) (a) Show that a stochastic process $\{N(t)\}$ is a Poisson process iff its inter arrival time distribution is Poisson.

(b) Let $\{N(t)\}$ is a Poisson process with parameter λ. Suppose that each occurrence of the events has a constant probability p of being recorded independently. If $\{M(t)\}$ is the number of events being recorded in an interval of length t. Then show that $\{M(t)\}$ is a Poisson process with parameters λp.

$(8+8) = 16\ marks$

Or

B) (a) Derive the steady state distribution of the Poisson process.

(b) Describe a one dimensional and two dimensional random walk. Show that states of one dimensional and two dimensional symmetric random walk are recurrent.

$(8+8) = 16\ marks$

IV A) (a) Derive Kolmogrov Equations for continuous time Markov Process.

(b) Define a birth and death process. Derive the forward Kolmogrov differential equation satisfied by the process.

$(8+8) = 16\ marks$

Or

B) (a) Define Galton Watson branching processes. Obtain the recurrence relation between the PGFs of a branching process.

(b) Show that probability of extinction of Galton Watson branching process is the smallest positive root of the equation $s = P(s)$, where $P(s)$ is the PGF of the offspring distribution.

$(8+8) = 16\ marks$

V A) (a) Define weak stationary process. Consider the stochastic process $\{X(t),\ t > 0\}$, where $X(t) = A\cos(\omega t) + B\sin(\omega t)$, where A and B are uncorrelated random variables such that $E(A) = E(B) = 0$ and $V(A) = V(B) = 1$. Show that $\{X(t),\ t > 0\}$ is a weak stationary process.

(b) Define renewal processes. State and prove renewal theorem.

$(8 + 8) = 16\ marks$

Or

B) (a) Describe the characteristics of a Queueing system. Obtain the steady system size distribution of an $M/M/1$ Queueing system.

(b) Define Brownian motion process. Explain how Brownian motion process can be obtained from a random walk through a limiting process.

$(8 + 8) = 16\ marks$

* * * * * *