GANDHI INSTITUTE OF TECHNOLOGY AND MANAGEMENT (GITAM)
(Deemed to be University)
VISAKHAPATNAM * HYDERABAD * BENGALURU

Accredited by NAAC with A^+ Grade

REGULATIONS AND SYLLABUS

OF

M.Sc. Statistics

(for 2021-22 admitted batch)

M.Sc. Statistics
REGULATIONS*
(2020-21 admitted batch)

1. ADMISSION

1.1 Admission into M.Sc. Statistics program of GITAM is governed by GITAM admission regulations.
1.2 Admission into M.Sc. (Statistics) will be based on an All India GITAM Science Admission Test (GSAT) conducted by GITAM and the rule of reservation, wherever applicable.

2. ELIGIBILITY CRITERIA

A pass in B.Sc. with Statistics as one of the Subject(s) and with a minimum aggregate of 50% marks in degree or any other equivalent examination approved by GITAM.

3. CHOICE BASED CREDIT SYSTEM

Choice Based Credit System (CBCS) is introduced with effect from the admitted Batch of 2015-16 based on UGC guidelines in order to promote:

- Student Centered Learning
- Cafeteria approach
- Inter-disciplinary learning

Learning goals/ objectives and outcomes are specified leading to what a student should be able to do at the end of the program.

4. STRUCTURE OF THE PROGRAM

The Program Consists of

i) Foundation Courses (compulsory) which give general exposure to a Student in communication and subject related area.
ii) Core Courses (compulsory).
iii) Discipline centric electives which
    a) are supportive to the discipline
    b) give expanded scope of the subject
    c) give inter disciplinary exposure
    d) Nurture the student skills
iv) Open electives are of general nature either related or unrelated to the discipline.

Each course is assigned a certain number of credits depending upon the number of contact hours (lectures/tutorials) per week.

In general, One credit for each Lecture / Tutorial hour per week are assigned to the courses based on the contact hours per week per semester.

The curriculum of the four semesters M.Sc. program is designed to have a total of **83** credits for the award of M.Sc degree

5. MEDIUM OF INSTRUCTION

The medium of instruction (including examinations and project reports) shall be English.

6. REGISTRATION

Every student has to register himself/herself for each semester individually at the time specified by the Institute / University.

7. ATTENDANCE REQUIREMENTS

A student whose attendance is less than 75% in all the courses put together in any semester will not be permitted to attend the end - semester examination and he/she will not be allowed to register for subsequent semester of study. He/she has to repeat the semester along with his / her juniors.

However, the Vice Chancellor on the recommendation of the Principal / Director of the Institute/School may condone the shortage of attendance to the students whose attendance is between 66% and 74% on genuine grounds and on payment of prescribed fee.

8. EVALUATION

The assessment of the student's performance in a Theory course shall be based on two components: Continuous Evaluation (40 marks) and Semester-end examination (60 marks).

A student has to secure an aggregate of 40% in the course in the two components Continuous Evaluation and Semester-end examination put together to be declared to have passed the course, subject to the condition that the candidate must have secured a minimum of 24 marks (i.e. 40%) in the theory component at the semester-end examination.

Details of Assessment procedure are furnished below in Table 1.

Table 1: Assessment Procedure

S. No.	Component of assessment	Marks allotted	Type of Assessment	Scheme of Examination
1	Theory	40	Continuous Evaluation	(i) Three mid semester examinations shall be conducted for 15 marks each. (ii) 5 marks are allocated for quiz. (iii) 5 marks are allocated for assignments.
		60	Semester-end examination	The semester-end examination shall be for a maximum of 60 marks.
2	Practicals	100	Continuous Evaluation	(i) Forty(40) marks for continuous evaluation is distributed among the components : regularity, preparation for the practical, performance,submission of records and oral presentation (ii) Sixty(60) marks for two tests of 30 marks each conducted by the concerned lab teacher and another faculty member of the department who is not connected to the lab, as appointed by the HoD.
3	Project work	200	Continuous Evaluation	(i) 150 marks for evaluation of the project work dissertation submitted by the candidate. (ii) 50 marks are allocated for the project Viva-Voce. (iii)The project work evaluation and the Viva-Voce shall be conducted by one external examiner outside the University and The internal project work supervisor.

9. SUPPLEMENTARY and SPECIAL EXAMINATIONS

(i) The odd (I and III) semester supplementary examinations will be conducted after conducting regular even semester examinations during April/May.

(ii) The even (II) semester supplementary examinations will be conducted after conducting regular odd semester examinations during November.

(iii) A student who has completed period of study and has 'F' grade in final semester Project is eligible to appear for special examination.

10. GRADING SYSTEM

Based on the student performance during a given semester/trimester, a final letter grade will be awarded at the end of the semester in each course. The letter grades and the corresponding grade points are as given in Table 2.

Table 2: Grades & Grade Points

SI.No.	Grade	Grade Points	Absolute Marks
1	O (outstanding)	10	90 and above
2	A+ (Excellent)	9	80 to 89
3	A (Very Good)	8	70 to 79
4	B+ (Good)	7	60 to 69
5	B (Above Average)	6	50 to 59
6	C (Average)	5	45 to 49
7	P (Pass)	4	40 to 44
8	F (Fail)	0	Less than 40
9	Ab. (Absent)	0	-

A student who earns a minimum of four grade points (P grade) in a course is declared to have successfully completed the course, subject to securing an average GPA (average of all GPAs in all the semesters/Trimesters) of 5 at the end of the Program to declare pass in the program.

.

Candidates who could not secure an average GPA of 5 at the end of the program shall be permitted to reappear for a course(s) of their choice to secure the same.

11. GRADE POINT AVERAGE

A Grade Point Average (GPA) for the semester/trimester will be calculated according to the formula:

$$GPA = \Sigma C \times G / \Sigma C$$

where

C = number of credits for the course,
G = grade points obtained by the student in the course.

To arrive at Cumulative Grade Point Average (CGPA), a similar formula is used considering the student's performance in all the courses taken, in all the semesters up to the particular point of time.

CGPA required for classification of class after the successful completion of the program is shown in Table 3.

Table 3: CGPA required for award of Class

Class	CGPA Required
First Class with Distinction	≥ 8.0*
First Class	≥ 6.5
Second Class	≥ 5.5
Pass Class	≥ 5.0

* In addition to the required CGPA of 8.0 or more the student must have necessarily passed all the courses of every semester in first attempt.

12. ELIGIBILITY FOR AWARD OF THE M.Sc. DEGREE

Duration of the program: A student is ordinarily expected to complete M.Sc. program in four semesters of two years. However a student may complete the program in not more than four years including study period.

However the above regulation may be relaxed by the Vice Chancellor in individual cases for cogent and sufficient reasons.
A student shall be eligible for award of the M.Sc Degree if he / she fulfills all the following conditions.

a.Registered and successfully completed all the courses and projects.
b. Successfully acquired the minimum required credits as specified in the curriculum corresponding to the branch of his/her study within the stipulated time.
c. Has no dues to the Institute, hostels, Libraries, NCC / NSS etc, and
d. No disciplinary action is pending against him / her.

The degree shall be awarded after approval by the Academic Council

13. DISCRETIONARY POWER

Not with standing anything contained in the above sections, the Vice Chancellor may review all exceptional cases, and give his decision, which will be final and binding.

Program Educational Objectives

PEO1	To apply basic knowledge of statistics to understand the data interpretation problems.
PEO2	To establish the methodologies for core statistical problems.
PEO3	To implement computer solution methods for large systems.
PEO4	To perform inter-disciplinary research objectives
PEO5	To imbibe professional and ethical responsibility towards the society.

Program Outcomes (POs) & Program Specific Outcomes (PSOs) for M.Sc. Statistics

PO 1	Apply basic knowledge of statistics to understand the data interpretation problems.
PO 2	Develop complexity data problem solving techniques using statistical tools.
PO 3	Establish the methodologies for core statistical problems.
PO 4	Implement computer solution methods for large systems.
PO 5	Assess the influence of global changes on organization for effective decision making business problems.
PO 6	Acquire knowledge of fast changing methodologies for solving big data problems.
PO 7	Exhibit leadership capabilities
PO 8	Perform inter-disciplinary research objectives
PO 9	Communicate effectively in peer and research related conferences
PO 10	Acquire skills to become a good researcher
PO 11	Engage in life-long learning environment.
PO 12	Imbibe professional and ethical responsibility towards the society.
PSO1	Create Statistical and Mathematical Models (along with solution) for various physical needs.
PSO2	Use Mathematics and Statistics, not only in the discipline of Statistics, but also in other disciplines and in their future endeavours
PSO3	Develop the statistical and computer programming skill for solving various physical problems.

M.Sc. (Statistics) – Curriculum Structure

(2020-21 Admitted Batch)

FIRST SEMESTER

Sl. No.	Course Code	Name of the Course	Category	Credits	Scheme of Instruction			Scheme of Examination		
					Hours per Week		Total	Duration in Hrs.	Maximum Marks	
					L/T	P			Sem. End Exam	Con. Eval
1	20SST701	Probability and Distribution Theory	PC	4	4	0	4	3	60	40
2	20SST703	Estimation Theory	PC	4	4	0	4	3	60	40
3	20SST705	Programming with Python	PC	4	4	0	4	3	60	40
4	20SST707	Sampling Methods	PC	4	4	0	4	3	60	40
5	SSE 701/ SSE 703	Skill Enhancement Course*	SEC	2	0	3	3	3	--	100
6	20SST 721	Statistical Lab - I using Python/SPSS	PP	2	0	3	3	3	--	100
7	20SST 723	Python Programming Lab	PP	2	0	3	3	3	--	100
8	VDC 111	Venture Discovery	VDC	2	2	0	2	---	---	100
Total			---	24	18	9	27	--	240	560

- **Skill enhancement courses (Choose one of the following)**
  SSE701: Basic Computer Concepts
  SSE703: Information Technology Tools

SECOND SEMESTER

Sl. No.	Course Code	Name of the Course	Category	Credits	Scheme of Instruction			Scheme of Examination		
					Hours per Week		Total	Duration in Hrs.	Maximum Marks	
					L/T	P			Sem. End Exam	Con. Eval
1	20SST702	Multivariate Analysis	PC	4	4	0	4	3	60	40
2	20SST704	Stochastic Process	PC	4	4	0	4	3	60	40
3	20SST706	Design of Experiments	PC	4	4	0	4	3	60	40
4	20SST708	Testing of Hypothesis	PC	4	4	0	4	3	60	40
5	SAE 702	Professional Communication Skills	AEC	2	0	3	3	--	--	100
6	20SST 722	Statistical Lab - II using SPSS/Python	PP	2	0	3	3	3	--	100
Total			---	20	16	6	22	--	240	360

THIRD SEMESTER

Sl. No.	Course Code	Name of the Course	Category	Credits	Scheme of Instruction			Scheme of Examination		
					Hours per Week		Total	Duration in Hrs.	Maximum Marks	
					L/T	P			Sem. End Exam	Con. Eval
1	20SST801	Intensive Computational Methodologies using R	PC	4	4	0	4	3	60	40
	Select any **TWO** Courses (each one course from Sl.No. 2 and 3)									
2	**Generic Elective – I**		GE	4	4	0	4	3	60	40
	20SST841	Data Mining Methods								
	20SST 843	Big Data Analytics								
	20SST 845	Machine Learning								
3	**Generic Elective – II**		GE	4	4	0	4	3	60	40
	20SST 851	Statistical Demography								
	20SST 853	Actuarial Statistics								
	20SST 855	Econometrics								
4	XXX	OPEN ELECTIVE	OE	3	3	0	3	3	60	40
5	20SST 821	Programming and Computing Lab with R	PP	2	0	3	3	3	--	100
6	20SST823	Statistical Lab - III using SPSS/Python/ R	PP	2	0	3	3	3	--	100
7	20SST891	Minor Project	PP	2	0	2	2	--	--	100
Total			---	21	15	8	23	--	240	460

FOURTH SEMESTER

Sl. No.	Course Code	Name of the Course	Category	Credits	Scheme of Instruction			Scheme of Examination		
					Hours per Week		Total	Duration in Hrs.	Maximum Marks	
					L/T	P			Sem. End Exam	Con. Eval
1	20SST802	Time Series Analysis and Forecasting Methods	PC	4	4	0	4	3	60	40
2	**Generic Elective**		GE	4	4	0	4	3	60	40
	20SST852	Advanced Operations Research								
	20SST854	Industrial Statistics and Quality Control								
	20SST856	Reliability								
3	20SST822	Statistical Lab – IV using SPSS	PP	2	0	3	3	--	--	100
4	20SST892	Project	PP	8	--	3	3	--	--	200*
Total			--	18	8	6	14	6	120	380

- 150 marks for Evaluation and 50 marks for presentation.

M.Sc (Statistics) – I Semester
20SST701: PROBABILITY AND DISTRIBUTION THEORY

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble: This course emphasizes the mathematical theory of probability. Random events are represented by sets and probability is just a normal measure defined on these sets. The topics covered are Measure theory, concept of random variable, Central limit theorem, Mathematical expectation and Convergence theorems. Distributions like Weibull, Laplace, Compound distributions and sampling distributions are also discussed. The concept of Multivariate normal, bivariate as a special case, partial and multiple correlations and their interrelationships are also discussed.

Course Objectives:

- To learn the concepts of sets, field, sigma field, measure on field, Borel function and induced sigma field.
- To learn the concepts of random variable, mathematical expectation,
- To learn the concepts of Central limit theorem and convergence theorems
- To familiarize with distributions like Weibull, Laplace distributions
- To familiarize with compound distributions, Multivariate normal and distribution of order Statistics

UNIT I
Classes of sets, field, sigma field, minimal sigma field, Borel field. Limit of asequence of sets. Measure on field, extension of measure to sigma field, Lebesquemesure, Lebesque-Stieltjes measures, Measurable functions, Borel function, induced sigma field.
UNIT II
Random variable, convergence of random variables- convergence in probability,almost surely, in the rth mean and in distribution their relationships. Characteristic function, properties, inversion theorem, continuity theorem, Central Limit Theorem, Lindberg-Levy, Liaunoff forms.
UNIT III
Mathematical expectation, Moments of random variable, conditional expectation, problem of moments. Basic Markov's, Chebycheff's, Holder's, Minkovski's and Jensen's inequalities. Law of large numbers. Chebycheff's and Kinchin's forms of WLLN, Kolmogrov's SLLN.

Convergence theorems relating to $X_n + Y_n$, X_nY_n and X_n / Yn Where $X_n \rightarrow X$ and $Y_n \rightarrow C$

UNIT IV

Weibull and Laplace distributions, their m.g.f., c.f. and other properties.Compound distributions – Poisson-Binomial. Sampling distributions . Non-central chi-square, Non-central t and Non-central F distributions and their properties. Distribution of quadratic forms under normality and related distributions.

UNIT V

Multivariate normal, bi-variate as a particular case, moments, c.f., conditional and marginal distributions. Distributions of order statistics from rectangular,exponential and normal distributions. Empirical distribution function, distribution of correlation coefficient, partial and multiple correlations, derivation formulae and inter relationships.

Course Outcomes:

CO1: To understand the concepts of set theory and measure theory.

CO2: To learn the concepts of random variable, Characteristic function and Central limit theorem.

CO3: Implement the concepts of mathematical expectation, Law of large numbers and convergence theorems.

CO4: To understand distributions like Weibull, Laplace, compound distributions and sampling distributions

CO5: To understand concepts of Multivariate normal, order statistics, partial and multiple correlations.

Text Books.

1. Bhat.B.R. . Modern probability theory. Wiley Eastern Ltd.
2. Rohatgi, V.K.(1984) . An introduction to Probability and Mathematical Statistics . Wiley Eastern
3. Goon, A.M., Gupta, M.K., Das Gupta, B. An outline of statistical theory, Vol.I.The World Press Pvt. Ltd., Kolkata.

References:

1. Billingsley, P.(1986) . Probability and Measure. Wiley
2. Kingman, J.F.C., and Taylor, S.J. (1966).Introduction to Measure and Probability. Cambridge University Press.
3. David, H.A. . Order Statistics Feller, W.Introduction to Probability Theory And Its Applications, Vol. II
4. Cramer, H. (1946). Mathematical Methods of Statistics.
5. PrincetonMorrison, D. F.(1976). Multivariate Statistical Methods. 2nd ed., McGraw Hill Mardia

M.Sc. (Statistics) – I SEMESTER
20SST 703: ESTIMATION THEORY

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble: This course is designed to expose students about Point and Interval Estimation and the characteristics of Estimation theory.

Course Objectives:

To enable students to

- Acquire the knowledge of point estimation
- Acquire characteristics of estimation theory.
- Ability to learn sufficiency and related theorems
- Learn about maximum likelihood Estimation and related theorems
- Acquire the knowledge about censored and truncated distributions

UNIT I

Point estimation. Concepts of unbiasedness, consistency, minimum varianceunbiased estimation. Information in a sample, Cramer-Rao inequality, efficiency of an estimator, Chapman-Robin's inequality and Bhattacharya bounds, definition of CAN estimator.

Learning Outcomes:

At the end of the unit, the student will be able to:

- Learn about the concept of unbiasedness, consistency and Minimum Unbiased Variance estimation
- Learn about Cramer-Rao Inequality theorem and its applications
- Acquire the applications under Chapman-Robin's inequality
- Learn about CAN estimator

UNIT II

Concept of sufficiency, single parameter and several parameter cases. Fisher-Neyman Factorization theorem, minimal-sufficient statistic, exponential families and Pitman families. Invariance property of sufficiency under 1-1 transformation of sample space and parameter space.

Learning outcomes:

At the end of the unit, the student will be able to:

- Gain the knowledge about the concept of sufficiency in single and several parameter cases
- Learn about Fisher-Neyman Factorization theorem and its applications
- Learn about the concept of minimal sufficient statistic
- Acquire the knowledge of transformation of sample space and parameter space

UNIT III

Distributions admitting sufficient statistics, Rao-Blackwell theorem, completeness , Lehman-Scheffe theorem, joint sufficiency (regular case).

Learning outcomes:

At the end of the unit, the student will be able to:

- Learn the concept of sufficient statistics
- Acquire the knowledge about Rao-Blackwell theorem and its applications
- Learn the concept of completeness
- Acquire the knowledge about Lehman-Scheffe theorem and its applications

UNIT IV

Method of maximum likelihood, CAN estimators for one-parameter Cramerfamily. Cramer-Huzurbazar theorem, solution of likelihood equations, method of scoring. Connection between MLEs and efficient estimators, MLEs and sufficient estimators.

Learning outcomes:

At the end of the unit, the student will be able to:

- Learn about the method of maximum likelihood and its applications
- Acquire the knowledge about CAN estimators for one parameter
- Learn about Cramer-Huzurbazar theorem and its applications
- Learn about the concept of method scoring

UNIT V

Censored and truncated distributions. Type I and Type II censoring for normal and exponential distributions and their MLEs. Interval estimation. confidence intervals using pivots, shortest expected length confidence intervals.

Learning outcomes:

At the end of the unit, the student will be able to:

- Acquire the knowledge of Censored and truncated distributions
- Learn about censoring some probability distributions and their maximum likelihood estimators
- Learn about the concept of Interval estimation
- Acquire the knowledge of shortest expected length confidence intervals

Course Outcomes:

On successful completion of this course, students will be able to:

- Learn about Cramer-Rao Inequality theorem and its applications
- Acquire the applications under Chapman-Robin's inequality
- Learn about Fisher-Neyman Factorization theorem and its applications
- Learn about the concept of minimal sufficient statistic
- Acquire the knowledge of transformation of sample space and parameter space

Text Books .

1. Goon, A.M., Gupta, M.K., Das Gupta, B. An Outline of Statistical Theory. Vol. II, The World Press PVT. Ltd., Kolkata.
2. Rohatgi, V. (1998). An Introduction to Probability and Mathematical Statistics.Wiley Eastern Ltd., New Delhi.

3. Kale, B.K. (1999). A First Course on Parametric Inference. Narosa PublishingHouse.

References.

1. Lehmann, E.L.(1986). Theory of Point Estimation.
2. Rao, C.R. (1973). Linear Statistical Inference.
3. Dudewicz, E.J. and Misra, S.N(1988) . Modern Mathematical Statistics. Student'sEdition, Wiley.
4. Lawless, J.F., Statistical Models and Methods for Lifetime Data. John Wiley &Sons.

M.Sc. (Statistics) – I SEMESTER
20SST 705: PROGRAMMING WITH PYTHON

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble: *Python is an interpreter oriented, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991. Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales.*

Course Objectives:

- To learn the basic concepts and usage of variables, expressions and practice the use of functions in Python programming language.
- To identify and practice different conditionals and implement recursive functions.
- To understand the concepts of strings, lists and dictionaries, practice the use of classes methods, overloading and polymorphism.
- To learn the basic concepts of raw data and use different statistical methods on the data.
- To implement line properties, use different setter methods and practice different kinds of plots.

UNIT- I

The way of the program: Running Python, Arithmetic Operators, Values and types, Formal and natural Languages, debugging.

Variables, expressions and statements: Assignment statements, variable names, expressions and statements, script mode, order of operations, string operations.

Functions: Function calls, math functions, composition, adding new functions, definitions and uses, flow of execution, parameters and arguments.(8 hours) **Learning Outcomes:**

By the end of the unit the student will be able to

- Explain different types of operators.(L2)
- Develop and run simple Python program.(L3)
- Describe the concepts of variables, expressions and statements.(L2)
- Use functions and develop programs using functions.(L3)
- Extend the concept of functions using parameters.(L2)

UNIT - II

Conditionals and Recursion: Floor division and modulus, Boolean expressions, logical operators, conditional execution, alternative execution, chained conditionals, nested conditionals, recursion, stack diagrams for recursive functions, infinite recursion.

Fruitful Functions: Return values, incremental development, composition, Boolean functions. **Iteration:** Reassignment, updating variables, while statement, break, square roots. (9 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Use the logical operators in programming.(L3)
- Identify the need of recursive functions.(L1)
- Construct programs using while statements.(L3)

- Explain the use of break statement.(L2)

UNIT – III

Strings: string length, traversal with for loop, string slices, searching, looping and counting, string methods, in operator, string comparison.

Lists: traversing a list, list operations, list slices, list methods, map, filter and reduce, deleting elements, lists and strings, objects and values, aliasing, list arguments.

Dictionaries: looping and dictionaries, reverse lookup, dictionaries and lists, memos, global variables.

Classes and objects: programmer defined types, attributes, rectangles, instances as return values.

Classes and methods: object oriented features, init method, str method, operator overloading, polymorphism. 10 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Construct programs to perform operations on strings.(L3)
- Explain basic concepts related to lists.(L2)
- Outline the concepts in dictionaries.(L2)
- Develop simple programs using class.(L3)
- Apply operator overloading and polymorphism.(L3)

UNIT - IV

Getting Started with Raw Data: The world of arrays with NumPy, Empowering data analysis with pandas, Data cleansing, Data operations.

Inferential Statistics: Various forms of distribution, A z-score, A p-value, One-tailed and two- tailed tests, Type 1 and Type 2 errors, A confidence interval, Correlation, Z-test vs T-test, The F distribution, The chi-square distribution, Chi-square for the goodness of fit, The chi-square test of independence, ANOVA. (8 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Explain the basic analytics that can be applied on data.(L2)
- Explain various forms of distribution.(L2)
- Develop programs using different statistical functions.(L3)

UNIT - V

Making Sense of Data through Advanced Visualization : Controlling the line properties of a chart, Using keyword arguments, Using the setter methods, Using the setp() command, Creating multiple plots, Playing with text, Styling plots, Box plots, Heatmaps, Scatter plots with histograms, A scatter plot matrix, Area plots, Bubble charts, Hexagon bin plots, Trellis plots, 3D plot of a surface. **Plotting and Visualization:** A brief matplotlib API primer, Plotting functions in pandas, Plotting Maps, Python Visualization Tool Ecosystem. (10 hours)

Learning Outcomes:

By the end of the unit the student will be able to

- Use keyword arguments and setter methods.(L3)
- Identify the need of plots.(L1)
- Explain basic concepts of charts.(L2)
- Illustrate the usage of visualization Tool Ecosystem.(L3)

Course Outcomes:

Upon completion of this course, student will be able to:

- List the different types of operators.
- Understand the concept of variables, expressions and statements.
- Understand the concept of functions and recursive functions.

- Identify the use of iteration.
- List the operations that can be performed on strings.

Text Books:

1. Mastering Python for Data Science by Samir Madhavan, PACKT Publishing, 2015.
2. Think Python by Allen Downey O'Reilly Publications, 2nd Edition, 2016.

Reference Books:

1. Programming Python by Mark Lutz, O'Reilly Publications, 4th Edition,2011.
2. Python in a nutshell by Alex Martelli, Anna Ravenscroft, Steve Holden, O'Reilly Publications, 3rd Edition, 2017.

M.Sc (Statistics) - I Semester (2020-2021 AB)
20SST707: Sampling Methods

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble : This course is concerned with the design of sample surveys and the statistical analysis of data collected from such surveys. Topics covered are: Simple random sampling with associated estimation and confidence interval methods, Selecting sample sizes, Estimating proportions, Unequal probability sampling, Ratio and regression estimation, Stratified sampling, Cluster and systematic sampling, Multistage designs and Double or Two-stage sampling.The aim of this course is to cover sampling design and analysis methods that would be useful for research and management in many fields.

Course Objectives :

- To learn scientific view to conduct the survey in proper way to collect the data about specific perspective.
- To Learn variety of probability and non-probability sampling methods for selecting a sample from a population.
- To amalgamate the intellectual facts of the sampling techniques to implement in projects
- To motivate the students in carrying out the field projects in scientific manner and statistical skills
- To convey some extended concepts in sampling to encourage the students in industrial and research aspects

UNIT I
Selection with varying probabilities, PPS sampling, Horvitz and Thomsonestimator, Yates' and Grundy's estimator, Midzuno-Sen sampling scheme.

UNIT II
Systematic sampling. Estimation of population mean and its variance, methods for populations with linear trend. Yates correction, modified systematic sampling, balanced systematic sampling, centrally located sampling, circular systematic sampling.

UNIT III
Cluster sampling. Estimation of population mean and its variance, efficiency ofcluster sampling, determination of optimal cluster size, estimation of proportion,cluster sampling with varying sizes. Two-stage sampling . Two-stage samplingwith equal first stage units. Estimation of mean and its variance. Optimumallocation. Three –stage sampling with equal probabilities. Two-stage ppssampling.

UNIT IV

Ratio estimation: Introduction. Bias and mean square error, estimation of variance, confidence interval, comparisons with mean per unit estimator, ratio estimator in stratified random sampling. Difference estimator and regression estimator: Introduction. Difference estimator, difference estimator in stratified sampling. Regression estimator, comparison of regression estimator with mean per unit estimator and ratio estimator. Regression estimator in stratified sampling.

UNIT V

Multi-phase sampling: Introduction. Double sampling for difference estimation,double sampling for ratio estimation, double sampling for regression estimator,optimum allocation varying probability sampling. Non-sampling errors. Sourcesand types of non-sampling errors, non-response errors, techniques for djustmentof non-response, Hansen and Harvitz technique, Deming's model.

Course Outcomes :

CO1: Understand the basic principles underlying survey design and estimation.

CO2: Apply the concept of systematic sampling along with centrally located sampling and circular systematic sampling.

CO3: Implement Cluster sampling, Two Stage Sampling and Opitmum allocation in real life problems.

CO4: Apply Ratio and Regression estimation in real life problems.comparison of regression estimator with mean per unit estimator and ratio estimator. Regression estimator in stratified sampling.

CO5: Apply Multi-phase sampling techniques, double sampling for regression estimator, Non-sampling errors, Hansen and Harvitz technique, Deming's model.

Text Books:

1. F.S.Chaudhary . Theory and Analysis of Sample Survey Designs, New AgeInternational Publishers, Delhi.
2. Des Raj . Sampling Theory.
3. Cochran, W.G. . Sampling Techniques.
4. Murthy, M.N. . Sampling Theory Techniques.
5. Parimal Mukhopadhyay. Theory and Methods of Survey Sampling. Prentice-Hall of India Pvt. Ltd., New Delhi.
6. Sukhatme, P.V. and Sukhatme, B.V. . Sampling Theory of Survey withApplications.

M.Sc. (Statistics)- I SEMESTER

(2020-2021 AB)

SSE 701: BASIC COMPUTER CONCEPTS

Hours per week: 3
Credits: 2 **Continuous Evaluation: 100 Marks**

Preamble: The course gives an understanding about the characteristics and classification of computers, various components of computer along with different operating systems that are available. It gives a hands on training on the packages MS-Word, MS-Power Point and MS-Excel. The course also comprehends AI tools.

Course Objectives:

- To introduce components of digital computer
- To work along with the outline of Operating Systems.
- To give hands on training on MS-Word,
- To give hands on training on Power Point
- To give hands on training on Excel features.

Basics of Computers: Definition of a Computer - Characteristics and Applications of Computers – Block Diagram of a Digital Computer – Classification of Computers based on size and working – Central Processing Unit – I/O Devices, Primary, Auxiliary and Cache Memory – Memory Devices. Software, Hardware, Firmware and People ware – Definition and Types of Operating System – Functions of an Operating System – MS-DOS –MS Windows, UNIX.

MS-Word: Features of MS-Word – MS-Word Window Components – Creating, Editing, ormatting and Printing of Documents – Headers and Footers – Insert/Draw Tables, Table Auto format – Page Borders and Shading – Inserting Symbols, Shapes, Word Art, Page Numbers, Equations – Spelling and Grammar – Thesaurus – Mail Merge.

MS-PowerPoint : Features of PowerPoint – Creating a Blank Presentation - Creating a Presentation using a Template - Inserting and Deleting Slides in a Presentation – Adding Clip Art/Pictures -Inserting Other Objects, Audio, Video- Resizing and Scaling of an Object –Slide Transition – Custom Animation.

MS-Excel : Overview of Excel features – Creating a new worksheet, Selecting cells, Entering and editing Text, Numbers, Formulae, Referencing cells – Inserting Rows/Columns – Changing column widths and row heights, auto format, changing font sizes, colors, shading.

Course Outcomes:

- Able to understand fundamental hardware components that make up a computer's hardware and the role of each of these components
- Understand the difference between an operating system and an application program, and what each is used for in a computer.
- Acquire knowledge about AI tools.
- Create a document in Microsoft Word with formatting that complies with the APA guidelines.
- Write functions in Microsoft Excel to perform basic calculations and to convert number to text and text to number.

Reference Books:

1. Fundamentals of Computers by V.RajaRaman, PHI Learning Pvt. Ltd, 2010.
2. Microsoft Office 2010 Bible by John Walkenbach, Herb Tyson, Michael R. Groh andFaithe Wempen, Wiley Publications, 2010.

M.Sc. (Statistics)-I SEMESTER

(2020-2021 AB)

SSE 703 : INFORMATION TECHNOLOGY TOOLS

Hours per week: 3
Credits: 2 **Continuous Evaluation: 100 Marks**

Preamble: The course enables the student to understand networking concepts related to Internet and introduce the social Networking sites and working of Email. It gives orientation of Block Chain technology. It give hands on training in SPSS, R Programming and creation of simple HTML document**s.**

Course Objectives:

- To enable the student to understand networking concepts related to Internet
- To introduce the social Networking sites and working of Email.
- To give orientation of Block Chain technology.
- To give hands on training in SPSS
- To give hands on training in R Programming and creation of simple HTML documents

Introduction to Internet: Networking Concepts, Data Communication –Types of Networking, Internet and its Services, Internet Addressing –Internet Applications–Computer Viruses and its types –Browser –Types of Browsers.

Internet applications: Using Internet Explorer, Standard Internet Explorer Buttons, Entering a Web Site Address, Searching the Internet– Introduction to Social Networking: twitter, tumblr, Linkedin, facebook, flickr, skype, yahoo!, google+, youtube, WhatsApp, etc.

E-mail : Definition of E-mail, Advantages and Disadvantages,User Ids, Passwords, Email Addresses, Domain Names, Mailers, Message Components, Message Composition, Mail Management, Email Inner Workings.

WWW-Web Applications, Web Terminologies, Web Browsers ,URL–Components of URL, Searching WWW –Search Engines and Examples.

Block Chain technology: What is Block Chain, Blockchain Architecture, How Block chain Transaction Works? Why do we need Blockchain? Block chain versions, Block chain Variants, Block chain Use Cases, Important Real-Life Use Cases of Block chain Bitcoin cryptocurrency: Most Popular Application of Block chain, Block chain vs. Shared Database, Myths about Block chain, Limitations of Block chain technology.

SPSS : SPSS Commands, Descriptive Statistics, Hypothesis Testing, Test of Difference, Analysis of Variance- One Way ANOVA, Non Parametric Tests, Correlation Analysis, Regression Analysis.

R Programming: Becoming familiar with R, Working with Objects, Introduction to Graphical Analysis.

HTML: WEB Terminology, Structure of HTML Document, HTML – Head and Body tags, Semantic tags- HR- Heading, Font, Image & Anchor tags, Different Types of Lists using Tags, Table Tags, Image Formats – Creation of Simple HTML Documents.

Course Outcomes:

- Enable to understand the basic networking concepts, types of networks, Internet Explorer and www.
- Outline the Block chain architecture, Bitcoin Crypto currency and Limitations of Block Chain.
- Choose different statistical tests to be performed on the data sets.
- Demonstrate the R programming with simple graphs.
- To make use of commands to structure HTML document

Reference Books:

- In-line/On-line : Fundamentals of the Internet and the World Wide Web by Raymond Greenlaw and Ellen Hepp, 2nd Edition, TMH.
- Microsoft Office 2010 Bible by John Walkenbach, Herb Tyson, Michael R. Groh and Faithe Wempen, WileyPublications.

M.Sc. (Statistics)-I SEMESTER

(2020-2021 AB)

20SST 721 Statistical Lab – I Using SPSS/Python

Hours per week: 4 **Continuous Evaluation: 100 Marks**
Credits: 2

Any 10 From the Following :

1. Determination of sample size.
2. PPSWR sample selection by (i) Cumulative Total method and (ii) Lahiri's method.
3. PPSWR sampling: Hansen-Hurwitz estimator and its sampling variance and Comparison of PPSWR sampling with SRSWR sampling based on PPS sample.
4. PPSWOR sampling: Horvitz-Thompson sampling strategy.
5. PPSWOR sampling: Sen- Midzuno sampling strategy.
6. PPSWOR sampling: Desraj's ordered estimator and Murthy's unordered estimator and their sampling variance and PPSWOR sampling : Rao-Hartley-Cochran strategy.
7. Cluster sampling with clusters of equal size.
8. Cluster sampling with clusters of unequal size.
9. Two stage sampling with SRSWOR at both the stages.
10. Two stage sampling with SRSWR at the first stage and SRSWOR at the second stage.
11. Ratio Method of estimation.
12. Regression method of estimation.
13. Two Phase sampling.
14. Two stage sampling with PPSWR at the first stage and SRSWOR at the second stage

M.Sc. (Statistics)-I SEMESTER

(2020-2021 AB)

20SST723 PYTHON PROGRAMMING LAB

Hours per week: 4 **Continuous Evaluation: 100 Marks Credits: 2**

1. Find all numbers which are multiples of 17, but not the multiples of 5, between 2000 and 2500?
2. Swap two integer numbers using a temporary variable.
   Repeat the exercise using the code format: a, b = b, a.
   Verify your results in both the cases.
3. Given two pairs of Cartesian points such as (x1, y1) and (x2, y2). Find the Euclidian distance between them.
   Hint: Use math module to find the square root.
4. Print the first 2 and last 3 characters in a given string. Use the string slicing concept. Do not use loops. If the length of the string is less than 5, print a suitable message.
5. Implement bubble sort. Do not use the default sort() method.
   Hint: So as to familiarize with the concept of sorting, and nested looping structures.
6. Implement shallow copy and deep copy of a list. You may use the copy module.
   Hint: While we copy a list, just a reference is copied. Hence if we make any changes to one of the lists, the same will reflect in the other as well. This is called shallow copying. Hence, in some cases we might need to deep copy, where a completely independent copy is created. This can be achieved through the deepcopy() method of copy module.
7. Write a temperature converter program, which is menu driven. Each such conversion logic should be defined in separate functions. The program should call the respective function based on the user's requirement. The program should run as long as the user wishes so.
8. Find the largest of n numbers, using a user defined function largest().
9. Write a function that capitalizes all vowels in a string.
   Hint: Do not use the ASCII concept. Use the upper() method.
10. Write a function leapYear() which receives a four digit year and returns a Boolean value: True if the year is leap, False if the year is not leap.
11. Read a line containing digits and letters. Write a program to give the count of digits and letters. Hint: Instead of checking ASCII, use the in-built methods like isdigit(), isalpha() etc.
12. Write a function myReverse() which receives a string as an input and returns the reverse of the string.
13. Use the list comprehension methodology in python, to generate the squares of all odd numbers in a given list.
    Hint: List comprehension is one of the powerful techniques in python;
14. Check if a given string is palindrome or not.
    Hint: do not use the C philosophy where we compare indices. Instead, copy the string as a new list, reverse the list using reverse(), join the list so that the reversed string is formed, using join(). Compare the new string and the old one.
15. Write a function to see if a given number is prime or not. Do not use any flag variables. Use math

module to find the square root, and its roof which will be fed in to range().
Hint: Just the return statements are enough. No need of flag variables. The loop has to run up to the roof of the square root of the number.

16.Write a function to find the factorial of a number using recursion.

17.Extend the above problem to find the nCr of given values of n and r. Verify the result with the help of itertools module, which helps to find the combinations.

18.Write a program that eliminates duplicates in a list. Do not use the concept of sets. Now, convert the original list into a set. Verify the result in both cases.

19.The user will enter five integers separated by commas. Write a program to read these values, and make a list. Print the list.
Hint: They will need to read the input using raw_input(), and then split the one and only line of input using split(). Then each of the values will need to be appended to a list, which will be empty at first.

20.Generate a dictionary and print the same. The keys of the dictionary should be integers between 1 and 10 (both inclusive). The values should be the cubes of the corresponding keys.

21.Create a nested dictionary. The roll number of a student maps to a dictionary. This inner dictionary will have name, age, and place as keys. Read details of at least three students.
Hint: A sample output should look like the one given below:
{11: {'name': 'Sachin', 'age': 18, 'place': 'Kochi'}, 12:
{'name':'Ammu', 'age': 19, 'place': 'Kannur'}, 13:
{'name':'jishad', 'age':20, 'place':'Calicut'}}

22.Enter a word. Create a dictionary with the letters of this word as keys, and the corresponding ASCII values as values.
Hint: Students may use the ord() function. Further, this is a simple problem, if list comprehension is used.

23.Implement Multiple plots.

24.Implement Scatter plots with histogram.

25.Implement Bubble charts.

Course Outcomes:

Upon completion of this course, student will be able to:

- Able to develop programs in Python.(L4)
- Able to implement functions using parameters.(L)
- Understand the concept of expressions.(L2)
- Construct programs using while statement.(L6)
- List the operations that can be performed on strings.(L4)

Text Books:

1.Mastering Python for Data Science by Samir Madhavan, PACKT Publishing,2015.

2.Think Python by Allen Downey O'Reilly Publications, 2nd Edition,2016.

M.Sc. (Statistics)-I SEMESTER

(2020-2021 AB)

Course Code: VDC111	**Course Title**: Venture Discovery	
Semester: I	**Course Type:** Continuous Evaluation	**Credits**: 2
Program: M.Sc. Statistics		
Course Leader: Venture Discovery Centre		

Course description and learning outcomes

India as part of its Make in India initiative has been focusing on creating incubation centers within educational institutions, with an aim to generate successful start-ups. These start-ups will become employment creators than employment seekers, which is the need of the hour for our country.

This common course for all the disciplines is a foundation on venture development. It is an experiential course that lets students venture and find out what is a business, financial and operating models of a business are. How to design and prototype a solutions that meets their customers' needs and generate revenue for the business.

LEARNING OBJECTIVES

- Discover who you are – Values, Skills, and Contribution to Society.
- Gain experience in actually going through the innovation process.
- Conduct field research to test or validate innovation concepts with target customers.
- Understand innovation outcomes: issues around business models, financing for start-ups, intellectual property, technology licensing, corporate ventures, and product line or service extensions.

On successful completion of this course, students will be able to:

	Learning Outcome	Assessment
1	Understand conceptual framework of the foundation of a venture	A1, A2
2	Understand the concept of purpose, mission and value-add service offered by a venture	A3
3	Analyze design and positioning of the product	A3
4	Demonstrate prototyping	A3
5	Analyze business, revenue and operating models	A3

Course outline and indicative content

Unit I (6 sessions)

Personal Values: Defining your personal values, Excite & Excel, Build a Team, Define purpose for a venture. Four stages: Personal Discovery, Solution Discovery, Business Model

Discovery, Discovery Integration.

Unit II (6 sessions)

Solution Discovery: Craft and mission statement, Experience design, Gaining user insight, Concept design and positioning, Product line strategy, Ideation & Impact.

Unit III (6 sessions)

Business Model Discovery: Prototyping solutions, Reality Checks, Understand your industry, Types of business models, Define Revenue Models, Define Operating Models

Unit IV (6 sessions)

Discovery Integration: Illustrate business models, Validate business models, Define company impact

Unit V (6 sessions)

Tell a Story: Can you make money, Tell your venture story.

Assessment methods

Task	Task type	Task mode	Weightage (%)
A1. Assignments	Individual	Report/Presentation	20
A2. Case / Project/Assignment	Groups* or Individual	Presentations/Report/Assignment	40
A3. Project	Individual/Group	Report/Pitch	40

Transferrable and Employability Skills

	Outcomes	Assessment
	Know how to use online learning resources: G-Learn, online journals, etc.	A1 & A2
2	Communicate effectively using a range of media	A1& A2
3	Apply teamwork and leadership skills	A2
4	Find, evaluate, synthesize & use information	A1 & A2
5	Analyze real world situation critically	A3
6	Reflect on their own professional development	A3
7	Demonstrate professionalism & ethical awareness	A2
8	Apply multidisciplinary approach to the context	A2

Learning and teaching activities

Mixed pedagogy approach is adopted throughout the course. Classroom based face to face teaching, directed study, independent study via G-Learn, case studies, projects and practical activities (individual & group)

Teaching and learning resources

Soft copies of teaching notes/cases etc. will be uploaded onto the G-learn. Wherever necessary, printouts, handouts etc. will be distributed in the class. Prescribed text book will be provided to all. However you should not limit yourself to this book and should explore other sources on your own. You need to read different books and journal papers to master certain relevant concepts to analyze cases and evaluate projects. Some of these reference books given below will be available in our library.

***Prescribed Modules*:**
Access to NU-IDEA online modules will be provided.

***Referential text books and journal papers*:**

Personal Discovery Through Entrereneurship, Marc H. Meyer and Chaewon Lee, The Institute of Enterprise Growth, LLC Boston, MA.

Suggested journals:

Vikalpa, Indian Institute of Management, Ahmedabad
Journal of General Management, Mercury House Business Publications, Limited
Harvard Business Review, Harvard Business School Publishing Co. USA

----------0----------

M.Sc. (Statistics) – II SEMESTER (2020-2021 AB)
20SST 702: MULTIVARIATE ANALYSIS

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble: The course is concerned with statistical methods designed to elicit information from different kinds of data sets. Since the data include simultaneous measurement on many variables, the methodology is called Multivariate analysis.

Course Objectives:

- To learn the techniques of data reduction, sorting and grouping.
- To investigate the dependence among variables.
- To predict values of one or more variables on the basis of observations.
- To understand Principal components
- To understand dimension reduction

Unit I

Hotelling's T^2 and its sampling distribution, application in test on mean vector for one and More multivariate normal population and also on equality of components of a mean vector in multivariate normal population.

Unit II

Definition of Wishart matrix and its properties, Mahalanobis distance, null distribution of Hotelling's T^2 statistic. Its application, tests on mean vector for one and more multivariatenormal populations , equality of the components of a mean vector in a multivariate population.

Unit III

Classification and discrimination procedures: Procedures for discriminating between twomultivariate normal populations, sample discriminant function, tests associated with discriminant functions, probability of mis-classification and their estimation.

Unit IV

Principle components, dimension reduction. Canonical variables and canonical correlation-

definition, use, estimation and computation.

Unit V

Cluster analysis : Hierarchical clustering - single, complete and average linkage methods, centroid and Ward's methods. Non-hierarchical methods – K-means algorithm.

Course Outcomes:

CO1: To understand the concept of Hotelling's T^2 and its sampling distribution.

CO2: To learn the concept of Wishart matrix and its properties.

CO3: To apply classification and discrimination procedures between two multivariate normal distributions.

CO4: To understand Principal components and dimension reduction.

CO5: To sensitize the basic ideas and concepts in Cluster analysis.

Text books:

1. T.W. Anderson, An Introduction to Multivariate Statistical Analysis, 3nd Ed., Willey, 2003.
2. Johnson, R. and Wichern (1992) .Applied Multivariate Statistical Analysis. Prentice- Hall, 3rd edition.
3. Naresh K. Malhotra and Satyabhushan Das, " Marketing Research an applied orientation", 7th edition revised, Pearson education.

Reference books

1. N.C. Giri, Multivariate Statistical Inference, Academic Press, 2014.
2. D.F. Morrison, Multivariate Statistical Methods, 4nd Ed. McGraw Hill, 2014.

M.Sc (Statistics) – II Semester (2020-2021 AB)

20SST704: STOCHASTIC PROCESS

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble : Stochastic models are among the most widely used tools in operations research and management science. Stochastic processes and its applications can be used to analyse and solve a diverse range of problems arising in production and inventory control, resource planning, service systems, computer networks and many others. This course, with an emphasis on model building, covers inventory models, Markov chains, Poisson processes, queuing theory, branching process and renewal process.

Course Objectives :

- To learn and to understand stochastic processes predictive approach.
- To develop an ability to analyze and apply some basic stochastic processes for solving real life situations.
- To inculcate the concepts of Stochastic modeling in-compassing the predictive analytics required for job roles.
- Formulate problems which involve setting up stochastic models
- Solve problems which involve setting up stochastic models

Unit I

Introduction to stochastic processes(sps), classification of sps' according to state space and time domain. Countable state Markov Chains(MCs), Chapman-Kolmogorov equations, calculation of n-step transition probability and its limit. Classification of states, period of state, stationary distribution of MC.

UNIT II

Random walk and gambler's ruin problem. Random walk in one and twodimensions. Gambler's ruin problem, probability of ultimate ruin, expectedduration of the game.

UNIT III

Discrete state space continuous time MC. Poisson process and its properties, birth process, death process, birth and death process.

UNIT IV

Weiner process as a limit of random walk, elementary properties of Weinerprocess. Branching

process. G-W branching process, probability of ultimateextinction, distribution of population size.

UNIT V

Renewal Theory. Elementary renewal theorem and applications. Study of residual and excess life times and their distributions. Stationary process. Weakly stationary process and strongly stationary process.

Course Outcomes :

CO1: Understand the stochastic processes, Markov chains, Transition probability matrix and various types of states.

CO2: Explain Random walk, Gambler ruins problem and apply Poisson process in real life situations.

CO3: Formulate and solve problems which involve setting up stochastic models.

CO4: Understand renewal theory and branching processes with applications.

CO5: Apply the elementary renewal theorem in real life problems and derive the stationery random process.

Text Books.

1. Medhi, J. (1982). Stochastic Processes. Wiley Eastern.
2. Bhat, B.R.(2002). Stochastic Models- Analysis and Applications. New AgeInternational, India.
3. Basu, A.K. . Introduction to Stochastic Process.
4. Srinivasan and Mehta. Stochastic Processes.

References.

1. Adke, S.R. and Manjunath, S.M. (1984). An introduction to Finite MarkovProcesses, Wiley Eastern.
2. Cinlar, E. (1975) . Introduction to Stochastic Processes. Prentice-Hall.
3. Feller, W. (1968). Introduction to Probability and Applications. Vol. I, WileyEastern.
4. Hoel, P.G., Port, S.G. and Stone, C.J. (1972). Introduction to Stochastic Processes. Houghton Miffin and Co.
5. Karlin, S. and Taylor, H.M. (1975) . A First Course in Stochastic Processes. Vol. I
6. Parzen, E. (1962). Stochastic Processes. Holden-Day.

M.Sc (Statistics) - II Semester

20SST706: Designs of Experiments

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble: This course is designed to expose students to the basics of principles of designs in order to develop professionals in factorial design and BIBD and PBIBD.

Course Objectives:
To enable students to

- Brush up principles of design, ANOVA and ANCOVA
- Acquire the knowledge of CRD,
- Acquire the knowledge of RBD and LSD
- Acquire the knowledge of Factorial design
- Acquire the knowledge on BIBD and PBIBD

UNIT I
Principles of designs, analysis of variance and analysis of co-variance, fixed andrandom effect models. Contrasts. Model adequacy checking. Test for normality, test for equality of variances (Bartlett test, Modified Levene method)

Learning Outcomes:
At the end of the unit, the student will be able to:

- learn about test for normality and equality of variance
- the importance of ANOVA and ANCOVA
- learn about the normality and test for equality of variances

UNIT II
C.R.D., R.B.D., estimation of parametric functions and tests of hypothesis,comparison of their efficiencies. Missing plot techniques, testing the equality ofsubsets of block effects or treatment effects. Multiple comparisons tests . Tukey's , Fisher's Least Significant Difference (LSD) method, Duncan's multiple range test.

Learning Outcomes:
At the end of the unit, the student will be able to:

- learn the statistical analysis of CRD and RBD
- comparison of their efficiency
- learn about the missing plot techniques of CRD and RBD

UNIT III
L.S.D., orthogonality in L.S.D. Missing plot technique, Analysis of split plot design.

Learning Outcomes:
At the end of the unit, the student will be able to:

- know the Statistical analysis of LSD
- learn about Orthogonality of LSD
- learn about the missing plot techniques
- learn about the analysis of spilt plot designs

UNIT IV

Factorial designs. Analysis of 2^n and 3^2 designs. Estimation of factorial effects, testing their significance. Total and partial confounding.

Learning Outcomes:

At the end of the unit, the student will be able to:

- acquire the knowledge about Factorial designs
- learn about statistical analysis of 2^n and 3^2 designs
- acquire the knowledge about total and partial confounding

UNIT V

Youdin design, intra block analysis . B.I.B.D., P.B.I.B.D., their analysis, estimation of parameters, testing of hypothesis.

Learning Outcomes:

At the end of the unit, the student will be able to:

- acquire the knowledge about Youdin design
- learn about intra block analysis
- acquire about the knowledge of BIBD and PBID and their statistical analysis

Course Outcomes:

On successful completion of this course, students will be able to:

- learn about the normality and test for equality of variances
- learn the statistical analysis of CRD and RBD
- know the Statistical analysis of LSD
- acquire the knowledge about Factorial designs
- learn about statistical analysis of 2^n and 3^2 designs

Text Books.

1. Das, M.N. and Giri, N.C.. Design and Analysis of Experiments. New AgeInternational Pvt. Ltd.
2. Montgomery, D.C. . Design and Analysis of Experiments. John Wiley and Sons, New York.

References.

1. Cochran and Cox . Experimental Designs. Asia Publishing House, Bombay.
2. Kemp Thorne. Design and Analysis of Experiments, Wiley Eastern Pvt. Ltd., New Delhi.

M.Sc. (Statistics) – II SEMESTER
20SST 708: TESTING OF HYPOTHESIS

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble: This course is concerned with statistical inference which is the mainstream of present-day statistical thinking. One of the main objectives of Statistics is to draw inferences about a population from the analysis of a sample drawn from a population. Statistical inference is classified into 1). Estimation 2). Testing of hypothesis. The theory of testing of hypothesis initiated by J. Neyman and E. S. Pearson is discussed. In N-P theory, we use statistical methods to arrive at decisions in certain situations where there is lack of certainty on the basis of a sample whose size is fixed in advance while in Wald's sequential theory the sample size is not fixed but is regarded as a random variable.

Course Objectives:

- To learn Neyman Pearson theory.
- To perform uniformly most powerful tests (UMP) for one sided and composite hypothesis.
- To study Wald's Sequential probability ratio test.
- To apply nonparametric tests.
- To learn Chi square test for goodness of fit, Kendall's and Spearman's test involving rank correlation.

Unit I
Neyman-Pearson theory. Lemma using critical functions. Uniformly mostpowerful tests, their relation with sufficient statistics,

UNIT II
Monotone likelihood ratio and UMP tests for one-sided hypothesis, compositehypothesis. Unbiased tests, uniformly most powerful unbiased tests. Type-A and Type-A regions.

UNIT III
Likelihood ratio critrion, its asymptotic distribution, one sample, two sample and k-sample problems. Linear hypothesis. Wald's SPRT. Proof that it terminates in a finite number of steps with probability 1. O.C ad A.S.N. functions. Examples of binomial and normal cases for

testing hypothesis on μ and σ^2

UNIT IV
Notion of non-parametric test, different NP tests. Run test, sign test, Wilcoxon and Mann-Whitney test, Median test, derivations of the mean and variance of the above test statistics when null hypothesis is true.

UNIT V
Chi-square test for goodness of fit, its asymptotic distribution, description ofKolmogorov-Smirnov test, tests involving rank correlation (Kendall's and
Spearman's).

Course Outcomes:

CO1: To learn Neyman Pearson lemma and UMP tests.

CO2: To understand unbiased tests and uniformly most powerful unbiased tests.

CO3: To study OC and ASN functions in case of Binomial and Normal distributions.

CO4: To apply Run test, Sign test, Wilcoxon and Mann Whitney test and Median test.

CO5: To understand Chi square test for goodness of fit, Kolmogorov and Smirnov test and Kendall's and Spearman's test

Text Books .

1. Rohatgi, V.K. . Statistical Inference, John Wiley and Sons.
2. Gibbons, J.D. . Non-parametric Inference, McGraw Hill
3. Wald. Sequential Analysis, John Wiley and Sons.
4. Goon, Gupta and Das Gupta . An Outline of Statistical Theory. Vol. 2, The World Press Pvt. Ltd., Kolkata.

References.

1. Lehmann, E.L. . Testing of Statistical Hypothesis. John Wiley and Sons.
2. Rao, C.R.. Linear Statistical Inference and its Applications. John Wiley and Sons.
3. Sidney Siegel . Non-parametric Statistics for the Behavioural Sciences.

M.Sc. Statistics –II semester

SAE 702: PROFESSIONAL COMMUNICATION SKILLS

Hours per week: 3
Credits: 2 **Continuous Evaluation: 100 Marks**

Preamble
This course is designed to expose students to the basics of academic and professional communication in order to develop professionals who can effectively apply communication skills, theories and best practices to meet their academic, professional and career communication needs.

Course Objectives :
To enable students to

- acquaint themselves with basic English grammar
- acquire presentation skills
- develop formal writing skills
- develop creative writing skills
- keep themselves abreast with employment-readiness skills

UNIT I
BACK TO BASICS
Parts of Speech, Tenses, Concord – Subject Verb Agreement, Correction of Sentences-Error Analysis, Vocabulary building
Learning Outcomes:
At the end of the unit, the student will be able to:

- use structures and tenses accurately
- apply the right verb to the right subject in a sentence
- detect incorrect sentences in English and write their correct form
- acquire new vocabulary and use in speaking and writing

UNIT II
ORAL PRESENTATION
What is a Presentation? Types of Presentations, Technical Presentation – Paper Presentation Effective Public Speaking, Video Conferencing,
Learning Outcomes:
At the end of the unit, the student will be able to:

- overcome speaking anxiety prior to presentation
- plan and structure effective presentations that deliver persuasive messages
- prepare slides that can catch the attention of the audience
- engage the audience
- skills in organizing, phrasing, and expressing the ideas, opinions and knowledge.
- facilitate and participate in a video conference effectively

UNIT III

DOCUMENTATION

Letter –Writing, E-mail Writing & Business Correspondence, Project Proposals, Report Writing, Memos, Agenda, Minutes, Circulars, Notices, Note Making

Learning Outcomes:

At the end of the unit, the student will be able to:

- write a business letter, which includes appropriate greetings, heading, closing and body and use of professional tone.
- draft crisp and compelling emails
- draft project proposals, reports and memos
- prepare agenda and draft minutes
- prepare circulars, notices and make notes.

UNIT IV

CREATIVE WRITING

Paragraph Writing, Essay writing, Dialogue Writing, Précis Writing, Expansion of Hints, Story Writing

Learning Outcomes:

At the end of the unit, the student will be able to:

- write paragraphs on familiar and academic topics using a topic sentence, supporting detail sentences and a conclusion sentence.
- learn the structure of a five-paragraph essay and write essays that demonstrate unity, coherence and completeness
- structure natural, lucid and spontaneous dialogues
- draft clear, compact logical summary of a passage
- recognize the elements of a short story and develop their functional writing skills.

UNIT V

PLACEMENT ORIENTATION

Resume preparation, group discussion – leadership skills, analytical skills, interviews –Types of Interviews, Preparation for the Interview, Interview Process.

Learning Outcomes:

At the end of the unit, the student will be able to:

- write a professional resume that highlights skills, specific to the student's career field
- acquire the personality traits and skills required to effectively participate in a G.D
- understand the purpose of interviews
- be aware of the processes involved in different types of interviews
- know how to prepare for an interview
- learn how to answer common interview questions

Course Outcomes:

To enable students to

- acquaint themselves with basic English grammar
- acquire presentation skills
- develop formal writing skills
- develop creative writing skills
- keep themselves abreast with employment-readiness skills

Text Books :

1.Essentials of Business Communication by Rajendra Pal and J S KorlahaHi, Sultan Chand & Sons, New Delhi.
2. Advanced Communication Skills by V. Prasad, Atma Ram Publications, New Delhi.
3. Effective Communication by Ashraf Rizvi, McGraw Hill Education; 1 edition (27 June 2005)
4. Interviews and Group Discussions How to face them, T.S.Jain, Gupta, First Edition, New Delhi.
5. High School English Grammar and Composition, P.C.Wren & Martin, N.D.V.Prasada Rao (Editor), S.Chand, 1995.

M.Sc. Statistics -II semester (2020-21AB)
20SST 722 Statistical Lab - II Using SPSS/Python

- **Practical Problems based on 20SST702, 20SST706 and 20SST708**

I Multivariate analysis

1. Maximum likelihood estimators Mean vector and dispersion matrix
2. Test for Mean Vectors (is known)
3. Hotelling s T2 statistic
4. Discriminant analysis
5. Principal component analysis
6. Canonical correlation and canonical variables

II. Design of Experiments

1. 2^4 , 3^2, 3^3 factorial experiment
2. Complete and partial confounding in 2^4, 3^2, 3^3 factorial experiments
3. Fractional factorial
4. Split plot design
5. Strip plot design
6. BIBD
7. PBIBD with two associate classes
8. Youden Square Design

III Testing of hypothesis

1. Construction of randomized and nonrandomized MP, UMP and UMPU tests of hypotheses and drawing the power curves.
2. Construction of SPRT and its OC and ASN curves.
3. Non parametric tests: Run test, Sign test, Wilcoxon test, Kolmogorov Smirnov test, Median test for k-sample problem, Kruskal Wallis test

M.Sc. (Statistics) – III SEMESTER (2020-2021 AB)

20SST 801 INTENSIVE COMPUTATIONAL METHODOLOGIES USING R

Hours per week: 4 **End Examination: 60 Marks**

Credits : 4 **Continuous Evaluation : 40 Marks**

Preamble: This course is designed to expose to the basic idea about computational methodologies using R like simulation, Markov chain, Bootstrap methods etc.

Course Objectives:

To enable students:

- Acquire the knowledge about simulation
- Importance sampling for integration
- Learn Monte Carlo methods
- Learn Bootstrap methods
- Get knowledge about simulating multivariate distributions

UNIT I

Stochastic simulation. generating random variables, simulating multivariate distributions, simulating stochastic processes such as simple queues.

Learning Outcomes:

At the end of the unit, the student will be able to:

- Acquire the knowledge about the simulation
- Learn how to generate random variables
- Get knowledge about simulating multivariate distributions

UNIT II

Variance reduction: importance sampling for integration, control variates and antithetic variables.

Learning Outcomes:

At the end of the unit, the student will be able to:

- Acquire the knowledge about variance reduction
- Learn the importance sampling for integration
- Learn how to control variates and antithetic variables

UNIT III

Markov Chain Monte Carlo methods. Gibbs sampling for multivariate simulation, simulated annealing for optimization.

Learning Outcomes:

At the end of the unit, the student will be able to:

- Acquire the knowledge about the Markov chain Monto Carlo methods
- Learn Gibbs sampling for multivariate simulation
- Learn simulated annealing for optimization

UNIT IV
Bootstrap methods: re-sampling paradigms, bias and standard errors, confidence intervals, bootstrapping in regression.

Learning Outcomes:
At the end of the unit, the student will be able to:

- Acquire the knowledge about Bootstrap methods
- Learn bias and standard errors
- Learn about confidence intervals

UNIT V
Jackknife and cross-validation: jackknife in sample surveys, cross validation fortuning parameters.

Learning Outcomes:
At the end of the unit, the student will be able to:

- Acquire the knowledge about Jackknife and cross validation
- Learn about Jackknife in sample surveys
- Learn about the cross validation of the parameters

Course Outcomes:
On successful completion of this course, students will be able to:

- Learn the importance sampling for integration
- Learn how to control variates and antithetic variables
- Acquire the knowledge about Bootstrap methods
- Learn bias and standard errors
- Learn about Jackknife in sample surveys

Text Books:
1. Rubinstein (1981). Simulation and the Monte Carlo Method. Wiley.
2. Tanner, M.A. (1996). Tools for Statistical Inference. 3rd ed., Springer.
3. Efron, B. and Tibshirani, R.J. (1993). An introduction to Bootstrap. Chapmann & Hall.
4. Shao, J. and Tu, D. (1995). The Jackknife and the Bootstrap. Springer Verlag.
5. Gnanadesikan, R. (1997). Methods for Statistical data Analysis of Multivariate Observations. 2nd ed., Wiley.

References:
1. Fishman, G.S. (1996). Monte Carlo . Concepts, Algorithms and Applications.Springer.
2. Belsley, D.A., Kuh, E., and Welsch, R.E.(1980). Regression Diagnostics. Wiley
3. McCullagh, P. and Nelder, J.A. (1999) . Generalized Linear Models. 3rd ed.,Chapman and Hall.
4. Seber, G.A.F., and Wild, C.J.(1989).Non-liner Regression. Wiley.
5. McLachlan, G.J. and Krishnan, T.(1997). The EM algorithms and extensions.Wiley.
6. Simon off, J.S. (1996). Smoothing Methods.

M.Sc Statistics- SEMESTER- III

Generic Elective – I

20SST 841 DATA MINING METHODS

Hours per week: 4 **End Examination: 60 Marks**

Credits: 4 **Sessionals: 40 Marks**

Preamble:

Due to advent of technology, internet, and advanced applications like social media, huge amount of digital data has been accumulated in data centers/cloud databases, which has led to a situation "we are drowning in data but starving from knowledge". To find golden nuggets which are useful for decision making process, various data mining functionalities like association analysis, classification, clustering, outlier analysis and web mining used. Data warehousing (DW) is an integral part of knowledge discovery process, where DW is an integration of multiple heterogeneous data repositories under a unified schema at a single site. The students will acquire knowledge in data modeling, design, architecture, data warehouse implementation and further development of data cube technology.

Course Objectives:

- Understand the importance of Data Mining and its applications.
- Introduce various types of data and pre-processing techniques.
- Learn various multi-dimensional data models and OLAP Processing.
- Study concepts of Association Analysis.
- Learn basics of cluster analysis.

UNIT – I

Introduction: Need for Data Mining, Definition of Data Mining, Kinds of data, Kinds of patterns to be mined, Technologies used, applications, Major issues in Data Mining.

Data Preprocessing: Need for Preprocessing the Data, Data Cleaning, Data Integration, Data Reduction, Data Transformation and Data Discretization.(10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Understand basic concepts of data mining. (L2)
- Learn the KDD process. (L2)
- Learn different data mining tasks. (L2)
- Learn major challenges in the field of data mining. (L2)
- Understand various types of data sets and attributes. (L2)
- Apply different statistical techniques on different types of attributes to measure the similarities and dissimilarities. (L3)
- Learn different data preprocessing techniques and apply them on data sets. (L2)

UNIT – II

Data Warehouse and OLAP Technology: Data Warehouse – basic concepts, Data Cube and OLAP Technology, Design and Usage, implementation, Data Generalization by Attribute- Oriented Induction.(8 hours)

Learning outcomes

- After completion of this unit, student will be able to
- Learn the basics of data warehousing and different OLAP operations. (L2)
- Understand the relationship between data warehousing and other data
- generalization methods.(L2)
- Study the methods of data cube computation.(L2)
- Explorations of data cube and OLAP technologies.(L4)

UNIT – III

Mining Frequent Patterns, Associations and Correlations: Basic Concepts and Methods- Basic Concepts, Frequent itemset Mining methods, Pattern Evaluation methods.

Advance Pattern Mining: Pattern mining in multilevel, multidimensional space, Constraint based Frequent Pattern Mining.(8 hours)

Learning outcomes:

After completion of this unit, student will be able to

- Understand the use of frequent patterns in business analysis. (L2)
- Implement A priori algorithm and FP-growth algorithm. (L3)
- Learn different types of association rules.(L2)
- Identify the importance of each pattern evaluation method. (L3)
- Understand measures for mining correlated patterns. (L2)
- Learn Advanced Pattern Mining Methods. (L2)

UNIT –IV

Classification: Basic Concepts, Decision Tree induction, Bayes Classification methods, Rule based Classification, Model Evaluation and selection, Techniques to improve classification accuracy, Support Vector Machines, Classification using Frequent patterns, Lazy Learners. (10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Understand basic concepts of classification. (L2)
- Implement the classification algorithms. (L3)
- Compare the performance of various classification algorithms. (L2)
- Understand model evaluation and selection methods. (L2)
- Identify the method that improves classification accuracy. (L3)

UNIT –V

Cluster Analysis: Definition, Requirements, Basic Clustering methods, Partitioning methods, Hierarchical methods, Density based methods, grid based methods, Evaluation of Clustering.

Outlier Detection: Outliers and Outlier Analysis, Detection methods, Statistical approaches, Proximity Based Approaches(8 hours)

Learning outcomes

After completion of this unit, student will be able to

- Understand the basic concepts of clustering. (L2)
- Implement the clustering algorithms. (L3)
- Compare the performance of various clustering algorithms. (L2)

- Learn various outlier detection methods. (L2)

Course Outcomes:

Upon completion of this course, student will be able to:

- Understand the functionality of various data warehousing and data mining component
- Understand the strengths and limitations of various data mining models.
- Implement data mining algorithms with different datasets.
- Compare various approaches of data mining implementations.
- Identify and apply appropriate data mining technique to solve a problem.

Text Book:

1. Data Mining Concepts and Techniques by Jiawei Han, Michel Kamber, Elsevier,3rd Edition, 2012.

Reference Books

1.Introduction to Data Mining by Pang-Ning Tan & Michael Steinbach, Vipin Kumar, Pearson Publications, 1st edition, 2016 .

2.Data Mining Techniques and Applications: An Introduction by Hongbo Du, Cengage Learning EMEA, 1st edition, 2010.

3.Data Mining : Introductory and Advanced topic by Dunham, Pearson Publications, 1st edition, 2006.

M.Sc Statistics

SEMESTER- III

Generic Elective-I

20SST 843 BIG DATA ANALYTICS

Hours per week: 4 **End Examination: 60 Marks**
Credits: 4 **Sessionals: 40 Marks**

Preamble:

The internet, Big Data, vastly improved computational power, and a wide variety of variables are involved in complex, real-world problems led to a new set of analytic techniques and technologies. The concept of Big Data includes massive volumes of data and huge benefits that can accrue from the analysis of it.

Course Objectives:

- To introduce an in depth understanding of all the concepts related to Big Data.
- To provide a sight into the real life implementations of Big Data solutions
- To evaluate problem solving in data analytics.
- To provide learners with a deep and systematic knowledge of business
- To provide technical strategies for data analytics and the subsequent skills to implement solutions in these areas.

UNIT – I

Getting an overview of Big Data: Big Data definition, History of Data Management, Structuring Big Data, Elements of Big-data, Big Data Analytics.

Exploring use of Big Data in Business Context: Use of Big Data in Social Networking, Use of Big Data in preventing Fraudulent Activities in Insurance Sector & in Retail Industry. **(**8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn various sources of data and forms of data generation. (L2)
- Understand the evolution and elements of Big Data. (L2)
- Explore different opportunities available in the career path. (L3)
- Understand the role and importance of Big Data in various domains. (L2)

UNIT – II

Introducing Technologies for Handling Big Data: Distributed and parallel computing for Big Data, Introducing Hadoop, Cloud computing and Big Data, In-memory Computing Technology for Big Data.

Understanding Hadoop Ecosystem: Hadoop Ecosystem**,** Hadoop Distributed File System, MapReduce, Hadoop YARN, Introducing HBase, Combing HBase and HDFS, Hive, Pig and Pig Latin, Sqoop, ZooKeeper, Flume, Oozie.

Understanding MapReduce Fundamentals and HBase: The MapReduce Framework, Techniques to Optimize Map Reduce Jobs, Uses of Map Reduce, Role of HBase in Big Data Processing.(10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Identify the difference between distributed and parallel computing. (L3)
- Learn the importance of Virtualization in Big Data. (L2)
- Learn the details of Hadoop and Cloud Computing. (L2)

- Learn the architecture and features of HDFS. (L2)
- Understand Hadoop Ecosystem, MapReduce and HBase. (L2)
- Apply the technique in optimizing MapReduce job. (L3)

UNIT- III

Understanding Big Data Technology Foundations: Exploring the Big Data Stack, Virtualization and Big Data, Virtualization approaches.

Processing Data with MapReduce: Recollecting the Concept of MapReduce Framework, Developing Simple MapReduce Applications.

Customizing MapReduce Execution and Implementing MapReduce Program: Controlling MapReduce Execution with Input Format, Reading Data with Custom Record Reader, Organizing Output data with Output Formats, Customizing Data with Record Writer, Optimizing MapReduce Execution with Combiner, Controlling Reducer Execution with Partitioners, Customizing the MapReduce Program for Sorting Text Data, implementing a Map Reduce Program for Sorting Text Data.(10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Explore the layers of Big Data Stack. (L2)
- Learn virtualization approaches in handling Big Data operations. (L2)
- Able to develop simple applications using map and reduce function. (L5)
- Learn the classes available in MapReduce framework. (L2)
- Understand the role of Combiner and Partitioners in a MapReduce applications. (L3)

UNIT – IV

Understanding Hadoop YARN Architecture: Background of YARN, Advantages of YARN, YARN Architecture, Working of YARN, YARN Schedulers, Backward Compatibility with YARN, YARN Configurations, YARN commands, YARN containers, Registry.(8 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the importance of YARN.(L2)
- Understand the use and importance of schedulers and backward compatibility in YARN.(L3)
- Learn the commands, log management and configuration for handling Big Data. (L3)

UNIT – V

Exploring Hive: Introducing Hive, Getting Started with Hive, Hive Services, Data Types , Built- in Functions, Hive-DDL, Data Manipulation, Data Retrieval Queries, Using Joins.

Analyzing Data with Pig: Introducing Pig, Running Pig, Getting started with Pig Latin, working with operators in Pig, Debugging Pig, Working with Functions in pig, Error Handling in Pig. **Understanding Analytics and Big Data**: Comparing Reporting and analysis, Types of Analytics, Developing an Analytic Team, Understanding Text Analytics.(10 hours)

Learning Outcomes:

After completion of this unit, student will be able to

- Learn the working of Hive and query execution. (L2)
- Learn the importance of Pig. (L2)
- Choose the operators in Pig. (L2)

- Understand various types of analytical approaches. (L3)

Course Outcomes:

Upon completion of this course student will be able to

- Able to handle a real time big data application. (L4)
- Able to develop Map Reduce Applications. (L4)
- Perform YARN Configuration for handling Big Data. (L3)
- Learn how Pig Latin is used for programming in Hadoop. (L3)
- Learn to design new analytical approaches for Big Data. (L3)

Text book:

1.Big Data Black Book by Dt Editorial Services, Dreamtech Publications, 2016.

Reference Book :

1.Hadoop The Definitive Guide by Tom White, O'reilly ,4th Edition,2016.

M.Sc Statistics -III semester

Generic Elective-I

20SST 845: MACHINE LEARNING

Hours per week: 4 **End Examination: 60 Marks**
Credits: 4 **Sessionals: 40 Marks**

Preamble:

Machine Learning addresses the question how to enable computers to learn from past experiences. It introduces the field of machine learning describing a variety of learning paradigms, algorithms, theoretical results and applications. Upon successful completion of the course, students will have an understanding the working of various machine learning algorithms which can be implemented through.

Course Objectives:

- To design a learning system and know about the learning tasks.
- To apply decision tree learning in classification tasks.
- To develop neural networks algorithms in machine learning.
- To illustrate Bayesian learning and instance based learning.
- To examine the concepts of genetic algorithms and reinforcement learning.

UNIT – I

Introduction: Well-Posed Learning Problems, Designing a Learning System, Perspectives and Issues in Machine Learning.

Concept Learning and the General-to-Specific Ordering: Introduction, A Concept Learning Task, Concept Learning as Search, FIND-S: Finding a Maximally Specific Hypothesis, Version Spaces and the Candidate-Elimination Algorithm, Remarks on Version Spaces and Candidate- Elimination, Inductive Bias.(10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Define a well-posed learning problem. (L1)
- Illustrate the designing of a learning system. (L1)
- Understand a concept learning task. (L2)
- Name what are version spaces. (L1)

UNIT – II

Decision Tree Learning: Introduction, Decision Tree Representation, Appropriate Problems for Decision Tree Learning, The Basic Decision Tree Learning Algorithm, Hypothesis Space Search in Decision Tree Learning, Inductive Bias in Decision Tree Learning, Issues in Decision Tree Learning.(10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Define a decision tree. (L1)
- Illustrate the decision tree learning algorithm and hypothesis space search. (L2)
- List various issues in decision tree learning. (L1)

UNIT - III

Artificial Neural Networks: Introduction, Neural Network Representations, Appropriate Problems for Neural Network Learning, Perceptrons, Multilayer Networks and the BACKPROPAGATION algorithm, Remarks on the Backpropagation algorithm, Illustrative Example, Advanced Topics in Artificial Neural Network.

Evaluating Hypothesis: Estimating Hypothesis Accuracy, Basics of Sampling Theory, A General

Approach for deriving Confidence intervals, Difference in Error of two Hypothesis, Comparing Learning Algorithms.(10 hours)

Learning outcomes

After completion of this unit, student will be able to

- Define what is a neural network and associated fundamentals. (L1)
- Demonstrate the working of multilayer neural networks. (L2)
- Learn the methods of evaluating the accuracy of hypotheses. (L2)

UNIT - IV

Bayesian Learning: Introduction, Bayes Theorem, Bayes Theorem and Concept Learning, Maximum Likelihood and Least Squared Error Hypothesis, Maximum Likelihood Hypothesis for predicting probabilities, Naive Bayes Classifier, Bayesian Belief Networks, EM Algorithm. **(**8 hours)

Learning outcomes

After completion of this unit, student will be able to

- Define Bayes theorem. (L1)
- Summarize the importance of Bayesian methods in machine learning. (L2)
- Show how Bayes theorem and concept learning are related. (L2)
- Learn how EM algorithm provides a quite general approach in the presence of unobservable variables. (L1)

UNIT-V

Instance-Based Learning: Introduction, k-Nearest Neighbour Learning, Locally Weighted Regression. **Genetic Algorithms:** Motivation, Genetic Algorithms, An Illustrative Example, Hypothesis Space Search, Genetic Programming.

Reinforcement Learning: Introduction, The Learning Task, Q Learning.(8 hours)

Learning outcomes

After completion of this unit, student will be able to

- Contrast instance-based learning with other methods of learning. (L4)
- Model genetic learning method by an analogy to biological evolution. (L3)
- Experiment with hypothesis space search in genetic learning. (L3)

Course Outcomes:

Upon completion of this course student will be able to

- Define a well-posed learning problem. (L3)
- Illustrate the decision tree learning algorithm and hypothesis space search. (L4)
- Use the Bayes theorem and EM Algorithm in machine learning . (L3)
- Contrast instance-based learning with other methods of learning. (L4)
- Model genetic learning method by an analogy to biological evolution. (L3)

Text Book:

1. Machine Learning by Tom M. Mitchell, McGraw Hill Education Private Limited, 2013.

Reference Books:

1.Pattern Recognition and Machine Learning by Christopher Bishop, Springer series,1st edition, 2006.

2.Machine Learning a Probabilistic Perspective by Kevin P Murphy & Francis Bach, MIT Press, 1st Edition, 2012.

M.Sc. (Statistics) – III SEMESTER

Generic Elective-II

20SST 851: STATISTICAL DEMOGRAPHY

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble: This course is designed to explore to the basic idea about demography like population of India and fertility and mortality rates.

Course objectives:

- Acquire the knowledge of the population, composition and ratio in India
- Know the different types of fertility measurements
- Know the different types of mortality measurements
- Know the growth of population
- Learn about the live birth intervals and number of births

UNIT-I

Scope and content of population census of India. Population, Composition, Dependency ratio. Brief Coverage and content errors in demographic data. Adjustment of age data – use of Whipple, Myer and UN indices. Chandrasekhar – Deming formula to check completeness of registration data.

Learning Outcomes:

At the end of the unit, student will be able to:

- Scope of the population
- Knowledge about the census of India
- Different types of errors in demographic data

UNIT-II

Measures of fertility: Stochastic models for reproduction, (Dandekar`s Modified Binomial and Poisson distributions, William Brass Model), distributions of time to first birth, inter-live birth intervals and number of births.

Learning Outcomes:

At the end of the unit, student will be able to:

- Acquire the knowledge about the measures of fertility
- Learn about the distributions of time to first birth
- Learn about the live birth intervals and number of births

UNIT-III

Measures of Mortality: Construction of abridged life tables (l_x–linear, exponential, Reed and Merrell`s, Grevill′s) Relations between functions of Life Tables. Distributions of life table functions.

Learning Outcomes:
At the end of the unit, student will be able to:

- Acquire the knowledge about the measures of mortality
- Learn about the relations between functions of life tables
- Learn about the distributions of life table functions

UNIT-IV
Stable and quasi-stable populations, intrinsic growth rate. Methods for population projection. Use of Leslie matrix.

Learning Outcomes:
At the end of the unit, student will be able to:

- Acquire the knowledge about stable and quasi-stable populations
- Learn about intrinsic growth rate
- Different methods for population projection
- Use of Leslie matrix

UNIT-V
Models for population growth and their fitting to population data. Linear, Exponential, logarithmic, modified logarithmic, Gompertz and Logistic Curves. Stochastic models for population growth (Pure Birth Model, Simple Birth & Death Model, Birth, death and migration model).

Learning Outcomes:
At the end of the unit, student will be able to:

- Learn the models for population growth
- Different types of curves for population growth
- Acquire the knowledge of stochastic models

Course Outcomes:
At the end of the course, student will be able to:

- Scope of the population
- Knowledge about the census of India
- Acquire the knowledge about the measures of fertility
- Acquire the knowledge about the measures of mortality
- Acquire the knowledge of stochastic models

Text Books:

1. Sudhendra Biswas (1995): Applied Stochastic Processes, New Age International Publishers Ltd.
2. Pathak, K.B. & Ram, F. (1998): Techniques of Demographic Analysis, Himalays Publishers
3. K.Srinivisan (1998): Basic Demographic Techniques and Tpplications: Sage publications.

4. Asha A bhande, Tara Kanitkar (2004): Principales of Population Studies; Himalayas publishing House.

References Books:

1. Saxena H.C and Surrendran P.U: Statistical Inference.
2. Bartholomew, D.J.(1982): Stochastic Modals for Social Processes, John Wiley.
3. Benjamin, B. (1969): Demographic Analysis, Geprge. Allen and Unwin.
4. Chain, C.L (1968): Introduction to Stochastic Processes in Biostatisties; John Wiley.
5. Cox, P.R. (1970): Applied Mathematical Demography, Sprinmger Verlag.
6. Spiegelman, M. (1969): Introduction to Demographic Analysis; Harvard University Press.

M.Sc. (Statistics)
III SEMESTER

Generic Elective-II

20SST 853: ACTURIAL STATISTICS

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation : 40 Marks**

Preamble : The aim of Actuarial Statistics is to provide a grounding in mathematical and statistical methods that are of relevance for actuarial work. It equips the student with knowledge of statistical distributions, methods to summarize data, the principles of statistical inference, regression models (including generalized linear models) and the fundamental concepts of Bayesian statistics.

Course Objectives :

- To learn the life tables used in insurance products.
- To learn the concept of interest
- To learn different life insurance products
- To learn life annuities and net premiums.
- To motivate students to prepare for exams required for employment in the actuarial science profession.

UNIT I

Theory of interest rates: ate of interest, nominal rate of interest. Accumulationfactors. Force of interest, present values, Stoodley formula for the force of interest, present value of cash flows, valuing cash flows. Basic compound interest function, equations of values and yield on transaction-annuities certain, present values and accumulation, concepts of different annuities, continuously payable annuities, varying annuities.

UNIT II

Utility theory, insurance and utility theory, models for individual claims an theirsums, approximations for the distribution of the sum. Application to insurance.Survival function, time until death for a person age x, curate future life time ,force of mortality.

UNIT III

Life table and its relation with survival function, examples, the deterministicsurvivorship group, recursion formulas, assumptions for fractional ages, someanalytical laws of mortality, select and ultimate tables.

UNIT IV

Life insurance: insurance payable at the moment of death and at the end of the year of death-

level benefit insurance, endowment insurance, deferred insurance an varying benefit insurance. Life annuities. single payment, continuous life annuities, discrete life annuities, life annuities with monthly payments, recursions, complete annuities-immediate and apportion able annuities-due.

UNIT V

Multiple life functions, joint life and last survivor status, insurance and annuitybenefits through multiple life functions, evaluation for special mortality laws.Multiple decrement models, deterministic and random survivorship groups,associated single decrement tables, central rates of multiple decrements, centralforce assumptions for multiple decrements. Uniform distribution assumption formultiple decrements.

Course Outcomes :

CO1: Understand the utility theory, insurance products and life tables.

CO2: Understand the concept of utility theory and interest.

CO3: Understand the concept of life life tables and to know some analytical laws of mortality.

CO4: Know insurance and the existing insurance products of different insurance company, life annuities, net premium and net premium reserves.

CO5: Know the concept of multiple life functions and decrement models.

Text Books:

1. Bowers, N.L., Gerber, H.U., Hickman, J.C, Jones, D.A., and Nesbitt, C.J. (1986) .Actuarial Mathematics. Society of Actuaries, Ithaca, Illinois, U.S.A. 2nd ed. (1997) CH. 1,2,3,4,5,9 & 10.
2. McCutcheon, J.J. and Scott, W.F. . An Introduction to Mathematics of Finance. Butter Worth & Heinemann.

References:

1. Spurgeon, E.T. (1972). Life Contingencies. Cambridge University press.
2. Nall, A. (1977). Life Contingencies. Heinemann.

M.Sc. (Statistics) – III SEMESTER
(2020-2021 AB)

20SST 823- Statistical Lab III Using SPSS/ Python/R

- Practical Problems based on Generic Elective II

I Actuarial Statistics:

1. Stoodley formula for the force of interest
2. Basic compound interest function, equations of values and yield on transaction-annuities
3. Models for individual claims and their sums
4. The deterministic survivorship group
5. Life annuities. single payment, continuous life annuities, discrete life annuities, life annuities with monthly payments
6. Multiple decrement models, deterministic and random survivorship groups

II Econometrics:

1. Equality of two regression equations, specification errors. Estimation methods
2. Goldfeld-Quandt test, Park test, weighted least square method of estimation
3. Auto-correlation. detection by Durbin-Watson statistic
4. Koyck approach, adaptive expectations model, stock adjustments models, Almon's approach
5. Methods of estimation. 1LS, 2SLS,IV, LIML and 3SLS.

M.Sc (Statistics)
III Semester

20SST855: Econometrics

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble: The focus of this course is on economic models. The models that have been developed in econometrics play an important role in social sciences where there is a concern with building and estimating models and the interconnection between various sets of variables in predominantly nonexperimental situation.

Course Objectives:

- To learn the nature and scope of econometrics and to formulate General linear model.
- To study equality of two regression equations.
- To learn about heteroscedasticity and autocorrelation.
- To understand distributed lag models
- To understand simultaneous equation models.

UNIT I
Nature and scope of Econometrics. General Linear Model. assumptions, OLSmethod of estimation, tests of hypothesis, confidence intervals, prediction,estimation subject to linear restrictions, maximum likelihood estimation.

UNIT II
Tests of structural change: dummy variables and seasonal adjustments, equality of two regression equations, specification errors. Estimation methods.

UNIT III

Generalized least squares: Aitken estimators. Heteroscedasticity. Goldfeld-Quandt test, Park test, weighted least square method of estimation. Auto-correlation. detection by Durbin-Watson statistic, estimation methods. Cochran, Orcutt and Durbin's. SUR system of equations.

UNIT IV

Lagged variables: distributed lag models- Koyck approach, adaptive expectations model, stock adjustments models, Almon's approach. Errors in variables.

UNIT V

Simultaneous equation models: structural form, reduced form and recursive form. Identification problem, order and rank conditions. Methods of estimation. 1LS, 2SLS,IV, LIML and 3SLS.

Course Outcomes:

CO1: To learn General linear model, OLS method of estimation and tests of hypothesis.

CO2: To understand tests of structural change in two variable and K variable linear model.

CO3: To learn about Generalized least squares estimators and Durbin Watson statistic.

CO4: To understand the sources of lagged variables and its estimation methods.

CO5: To learn simultaneous equation models and its estimation.

Text Books:

1. Gujarathi, D.(1979). Basic Econometrics, McGraw Hill.
2. Johnston, J.(1984). Econometric Methods. 3rd ed., McGraw Hill.
3. Koutsoyiannis, A.(1979). Theory of Econometrics. Macmillan Press.
4. Theil, H.(1982). Introduction to the Theory and Practice of Econometrics. JohnWiley.

References:

1. Apte, P.G. (1990). Text Book of Econometrics. Tata McGraw Hill.
2. Cramer, J.S. (1971). Empirical Econometrics. North Holland.
3. Intrilligator, M.D.(1980). Econometric Models-Techniques and Applications, Prentice Hall of India.
4. Klien, L.R. (1962). An Introduction to Econometrics. Prentice Hall of India.
5. Mallnvand, E (1966). Statistical Methods of Econometrics. North Holland.
6. Srivastava, V.K. and Gile, D.A.E. (1987). Seemingly Unrelated RegressionEquation Models. Marcel and Dekker.
7. Walters, A. (1970). An Introduction to Econometrics. McMillan & co

20SST 821 PROGRAMMING AND COMPUTING WITH R

Hours per week: 4 **Continuous Evaluation: 100 Marks**
Credits: 2

1. R Introduction:Reserved Words,Variables & Constants, Operators, Operator Precedence.
2. Decision and Loop Structure : if...else, for loop, while Loop, break & next, repeat Loop.
3. R Functions:Functions, Function Return Value, Environment & Scope, Switch Function.
4. R Data Structures : Vectors, Matrix, List in R programming, Data Frame, Factor.
5. R Objects and Class : Object and Class ,S3 Class,S4 Class, R Reference Class, Inheritance.
6. R Graphs and Charts: Bar Plot, Histogram, Pie Chart, Box Plot, Strip Chart.
7. More on Plotting In R: Plot Function: Subplot, Saving Plot, 3D Plot.
8. R Program using a csv file to plot using the distributions like Normal Distribution, Binomial Distribution, Poisson Distribution, Bernoulli Distribution and inferencing the behavior of data.
9. Program to normalize the data, replacing null columns either with mean or variance.
10. Program to import the data. And calculate covariance, correlation, mean, standard deviation and generate plots.
11. Program to implement linear regression for predictive modeling check how well the model fit the data.
12. Program to implement Multiple Regression for a given data.
13. Program to implement Cross Validation Methods

- Validation set approach (or data split)
- Leave One Out Cross Validation
- k-fold Cross Validation
- Repeated k-fold Cross Validation

Course Outcomes:

Upon completion of this course, student will be able to:

- Understand basic concepts such as data type and index and use them in their work. (L3)
- Demonstrate use of basic functions. (L2)
- Conceptualize and create loops to solve different types of problems. (l3)
- Create their own customized functions. (L4)
- Construct tables and figures for descriptive statistics. (L3)

Text Books:

1.Beginning R–The Statistical Programming Language by Mark Gardener, Wrox publications, 2012.
2.Statistics Using R by Sudha G. Purohit, Sharad D. Gore, Shailaja R. Deshmukh, Narosa publishers, 2nd Edition, 2015.

M.Sc. (Statistics) – IV SEMESTER
(2020-21 AB)
20SST 802: TIME SERIES ANALYSIS & FORECASTING METHODS

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble: This course is concerned with Time series analysis and Forecasting methods which is an important aid in effective and efficient planning.The five basic steps in any forecasting task like problem definition, gathering information, preliminary (exploratory) analysis, choosing and fitting models, using and evaluating a forecasting model are discussed.

Course Objectives:

- To learn about Time series analysis and growth models.
- To understand exponential smoothing methods.
- To study forecasting with ARIMA models.
- To estimate ARIMA model parameters.
- To study spectral analysis of weakly stationary process.

UNIT I:
Review of Time Series Analysis. Growth models: Modified Exponential Curve, Gompertz curve, Logistic curve and their Fitting; Measurement of cyclical component: Harmonic analysis, auto regression series: Markoff and Yule's series, Periodogram and correlogram analysis, measurement of irregular component: variate difference method.

UNIT II:
Exponential smoothing methods: trend adjusted exponential smoothing, double and triple exponential smoothing, Holt and Winters smoothing, chow's adaptive control methods, brown's one parameter adaptive method:

UNIT III:
Detailed study of the stationary processes: Moving Average (MA), Auto Regressive (AR), ARMA and AR Integrated MA(ARIMA)models - forecasting with ARIMA models.

UNIT IV:
Box-Jenkins models: Discussion (without proof) of estimation of mean, auto covariance and auto-correlation functions under large sample theory. Choice of AR and MA periods. Estimation of ARIMA model parameters.

UNIT V:

Spectral analysis of weakly stationary process. Periodogram and correlogram analyses. Computations based on Fourier transform.

Course Outcomes:

CO1: To learn about Exponential, Gompertz and Logistic curves and their fitting.

CO2: To understand trend adjusted exponential smoothing methods.

CO3: To study MA, AR, ARMA and ARIMA models.

CO4: To understand Box-Jenkins models and estimate auto covariance and auto correlation functions.

CO5: To study Periodogram and Correlogram analyses.

Text Books:

1. Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis – Forecasting and Control. Holden Day, San Francisco.
2. Anderson, T.W. (1971). The Statistical Analysis of Time Series.Wiley, N.Y. Makridakis, Wheelwright and McGee. Forecasting. Methods and Applications. John Wiley & Sons.
3. Montgomery, D.C. and Johnson, L.A. (1977). Forecasting an Time Series Analysis. McGraw Hill.

References:

1. Fuller, W.A. (1976). Introduction to Statistical Time Series. John Wiley,
2. N.Y. Granger, C.W.J. and Newbold (1984). Forecasting Econometric Time Series. 3rd ed., Academic Press.
3. Priestley, M.B. (1981). Spectral Analysis and Time Series. Griffin, London.
4. Kendall, S.M. and Ord, J.K. (1990). Time Series Analysis. 3rd ed., Edward.
5. Kendall, M.G. and Stuart, A. (1966). The Advanced Theory of Statistics.Vol.3, Charles Griffin, London.
6. Koopmans, L.H.(1974). The Spectral Analysis of Time Series. Academic Press.

M.Sc (Statistics) - IV Semester

20SST852 Advanced Operations Research

Hours per week: 4 **End Examination: 60 Marks**

Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble : Operations research (OR) have many applications in science, engineering, economics, and industry and thus the ability to solve OR problems are crucial for both researchers and practitioners. Being able to solve the real life problems and obtaining the right solution requires understanding and modelling the problem correctly and applying appropriate optimization tools and skills to solve the mathematical model. In particular, we will cover linear programming, network flow problems, nonlinear programs, dynamic programming, solve specialized linear programming problems like the transportation and assignment problems, solve network models like the shortest path, minimum spanning tree, and maximum flow problems, understand how to model and solve problems using dynamic programming, learn optimality conditions for single- and multiplevariable unconstrained and constrained non-linear optimization problems and corresponding solution methodologies.

Course Objectives :

- To develop the optimization techniques that will be useful in the personal and professional life.
- To learn the mathematical formulation of complex decision-making problems and arrives at optimal or near-optimal solutions using different techniques of operations research.
- To learn advanced methods in operations research course that are used in the systems approach to decision-making problems, in particular emphasizing the roles of uncertainty and risk.
- To grow expertise in optimization techniques, mathematical modeling which are essential in business intelligence.
- To build and simulate the advanced models and skills to implement in real time scenarios and research works

UNIT I
Decision Theory: Decision theory approach, decision theory under uncertainty,under risk, posterior probabilities and Bayesian analysis, decision tree analysis,decision making with utilities.

UNIT II
Game theory: two-person games, pure and mixed strategies, existence of solution and uniqueness of value in zero-sum games, finding solutions in 2x2,2xm and mxn games. Dynamic programming.

UNIT III
Integer programming: Branch and Bound algorithm and cutting plane algorithm.Multi-criterion and goal programming. Replacement problems: block and agereplacement policies, replacement of items with long life.

UNIT IV
Project management: CPM, PERT, probability of project completion, crashing.

UNIT V
Information theory: Communication process, entropy, channel capacity, efficiency, redundancy. Shannon-Fano encoding procedures. Non-linear programming: Kuhn-Tucker conditions, Wolfe and Beale's algorithms for solving quadratic programming problems.
Course Outcomes :

CO1: To the know the concept of decision theory along with practical problems.

CO2: Describe the basic concepts of game theory and demonstrate the formulations of realworld problems as a game theory model

CO3: Describe the basic concepts of integer programming problem and demonstrate the formulations of real-world problems as a integer linear programming model

CO4: Understand and deal with Networking models and know the concept of project completion and crashing.

CO5 : To know the concept of Information theory and apply khun-tucker conditions in real life problems.

Text Books:

1. Kanti Swarup, Gupta, P.K. and Man Mohan (1985) . Operations Research, Sultan Chand and Sons.
2. Sharma, J.K. (2003). Operations Research Theory and Applications. Macmillan, India.

3. Sharma, S.D. . Operations Research. Kedarnath Ramnath Publishers, Meerut

References:

1. Taha, H.A.(1982) .Operations Research-An Introduction. Macmillan.
2. Hillier, F.S. and Leiberman, G.J. . Introduction to Operations Research. HoldenDev.
3. Churchman, C.W., Ackoff, R.L., and Arnoff, E.L. (1957) . Introduction toOperations Research. John Wiley.
4. Gross, D. and Harris, C.M. (1974) . Fundamentals of Queuing Theory. John Wiley.

M.Sc (Statistics), IV Semester

20SST 854: INDUSTRIAL STATISTICS & QUALITY CONTROL

Hours per week: 4 **End Examination: 60 Marks**

Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble: This course is designed to expose students to the basic idea about Industrial Statistics and Quality Control in order to develop professionals in quality and quality improvement in industries.

Course Objectives:

To enable students to

- How to improve quality of a products in Industries
- Acquire the knowledge about the charts to improve to quality of product
- Develop process capability
- Acquire the knowledge about accept sampling plans
- Acquaint themselves with basic reliability concepts

Unit I

Quality improvement: Meaning of quality and quality improvement Different types of Quality costs and their management.

Learning Outcomes:

At the end of the unit, student will be able to:

- Learn the meaning of quality and quality improvement
- Acquire different types of quality costs and their management

Unit II

Control charts: Review of $\bar{X}$, R, p, c, d charts - Modified control charts for mean CUSUM chart technique of V-mask Weighted Moving average charts Slopping control charts and group control charts.

Learning Outcomes:

At the end of the unit, student will be able to:

- Review of basic charts
- Learn about the different types of control charts

Unit III

ProcessCapabilityanalysis:Meaning,Estimationtechniqueforcapabilityofaprocess Capability Indices: C_p, capability ratio and C_{pk} index. Estimation of natural tolerance, limit of a process.

Learning Outcomes:
At the end of the unit, student will be able to:

- Learn estimation technique for capability of a process
- Learn about capability of ratio and index
- Estimation of natural tolerance

Unit IV

AcceptanceSamplingplansforattributes:Single,double,multipleandcontinuoussamplingplansfor attributes (Dodge type).
Acceptance Sampling plans for variables: one sided and two sided specification Standardized plans (ANSI/ANSQ Z1.9) andMIL-STD-414.

Learning Outcomes:
At the end of the unit, student will be able to:

- Acquire the knowledge about different types of sampling plans
- Learn about the accept sampling plans for variables

Unit V

Reliability: Concept, Definition and need - Concepts of Hazard rate, IFR and DFR - Relevance of exponential distribution in Reliability Failure models Taguchi's approach in Quality and reliability- Six sigma approach.

Learning Outcomes:
At the end of the unit, student will be able to:

- Acquire the knowledge about the concept of reliability, Hazard rate, IPR and DFR
- Learn about the reliability failure models
- Learn about the Quality and reliability using Taguchi's approach

Course Outcomes :

- Learn the meaning of quality and quality improvement
- Acquire different types of quality costs and their management

- Acquire the knowledge about different types of sampling plans

- Acquire the knowledge about the concept of reliability, Hazard rate, IPR and DFR
- Learn about the reliability failure models

Text books:

1. DouglasC.Montgomer:IntroductiontoStatisticalQualityControl, 6th edition John WileyandSons, NewYork, 2008.
2. Balagurusamy,B:ReliabilityEngineering,TataMcGrawHillBookCompany,New

Delhi, 2017.

Reference books:

1. Mahajan,M(1998):StatisticalQualityControl,DhanpatRai&CoPrivateLtd.,New Delhi.
2. Smith,G.M(1991):StatisticalProcessControlandQualityImprovement,5th edition, PrinticeHall,NY, 2003

M.Sc. (Statistics) – IV SEMESTER
20SST 856: RELIABILITY

Hours per week: 4 **End Examination: 60 Marks**
Credits : 4 **Continuous Evaluation: 40 Marks**

Preamble : The course will end with a brief introduction on system reliability and its various configurations. Finally, to have hands-on experience of the reliability models, the course will also include a practical session using softwares.

Course Objectives :

- To learn the reliability theory and analysis of survival data.
- To distinguish censored and uncensored data.
- To visualize and communicate time-to-event data,
- To communicate time-to-event data
- To fit and interpret failure time model

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

Unit II
Life distributions; reliability function ; hazard rate; common life distributionsexponential , Weibull, Gamma, etc. Estimation of parameters and tests in these models.

Unit III
Notions of ageing: IFR, IFRA, NBU, DMRL, and NBUE Classes and their duals andImplications

Unit IV
Common bivariate exponential distributions and their properties.

Unit V
Reliability growth models; probability plotting techniques; Hollander- Proschan andDeshpande tests for exponentiality - Basic ideas of accelerated life testing.

Course Outcomes :

CO1: Understand the elements of reliability, hazard function and its applications.

CO2: Understand the concept of censoring, life distributions and ageing classes.

CO3: To know the concept of Notions of ageing: in reliability theory.

CO4: Explain common bivariate exponential distributions and their properties.

CO5: Explain the reliability growth models and basic ideas of accelerated life testing.

Text Books :

1. Barlow R.E. and Proschan F. (1985) Statistical Theory of Reliability and Life Testing; Rinehart and Winston.
2. Lawless J.F. (1982) Statistical Models and Methods of Life Time Data; John Wiley.

References :

1. Bain L.J. and Engelhardt (1991) Statistical Analysis of Reliability and Life Testing Models; Marcel Dekker.
2. Nelson, W (1982) Applied Life Data analysis; John Wiley.
3. Zacks S , Reliability Analysis, Springer Verlag.

M.Sc. (Statistics) – IV SEMESTER (2020-2021 AB)

20SST 822 Statistical Lab IV Using SPSS/ Python/R

- Practical Problems based on 20SST802 and Generic Elective

I TIME SERIES ANALYSIS & FORECASTING METHODS:

1. Modified Exponential Curve, Gompertz curve, Logistic curve and their Fitting
2. Double and triple exponential smoothing, Holt and Winters smoothing
3. Moving Average (MA), Auto Regressive (AR), ARMA and AR Integrated MA(ARIMA)models
4. forecasting with ARIMA models

II Advanced Operations Research:

1. 2x2,2xm and mxn games
2. Branch and Bound algorithm and cutting plane algorithm
3. CPM, PERT, probability of project completion, crashing
4. Wolfe and Beale's algorithms for solving quadratic programming problems

III INDUSTRIAL STATISTICS & QUALITY CONTROL:

Control charts:

1. CUSUM chart
2. Modified Control chart
3. Moving Average Control chart
4. Exponentially Weighted Moving Average chart
5. Sloping Control Chart

Acceptance sampling:

1. Single sampling plans and double sampling plans (for attributes)
2. Variable Sampling plans (Single and double specifications)
3. Standard plans Sloping Control Chart

M.Sc (Statistics) – VI Semester
20SST 892 PROJECT

1. A project work is mandatory and shall be offered in IV semester.

2. A project work may be taken individually or by a group of two students.

3. Project work shall be supervised by a faculty member assigned by the HoD.

4. The project work should be selected in such a way that there is enough scope to apply and demonstrate the statistical techniques learnt in the course.

5. At the end of the semester, before the last working day, a report on the work done should be submitted (two copies). If a team of two students jointly do a project work then they must submit individual report separately (not copy of the same report).

6. The project report shall clearly state the selected problem, the statistical methodologies employed for data collection and analysis and the conclusions arrived. Details of previous studies in the area and related references should also be given.

7. The project work will be assessed for a maximum of 200 marks. 150 marks for Evaluation and 50 marks for presentation.