ST. JOSEPH'S UNIVERSITY

BENGALURU-27

School of Physical Sciences

DEPARTMENT OF STATISTICS

Syllabus for Bachelor of Science

(With Statistics as one of the three Major subjects with Practical)

Under State Education Policy

For Batch 2024 onwards

Name of the Degree Program: B.Sc.

Discipline Core: Statistics

Total Credits for the Program: 132 (till 6 semesters)

Starting Year of implementation: 2024-25

Pre requisites for the Course:

1. Only those candidates who have passed Pre-University course or an equivalent course with Mathematics/Business Mathematics/Basic Mathematics/Applied Mathematics as one of the optional subjects are eligible to take Statistics as one of the optional subjects in BSc course.
2. Any student taking Statistics as one of the optional subjects in the BSc course shall take Mathematics as another optional subject.
3. The subject of Statistics in this course has to be taught by M.Sc degree holders in Statistics / Applied Statistics.

Program Outcomes

By the end of the program the students will be able to:

PO1	Understand the basic concepts of statistics
PO2	Apply the key concepts of statistical knowledge in real world applications
PO3	Analyze datasets to identify patterns and inconsistencies
PO4	Identify, evaluate and validate statistical models to address research problems
PO5	Apply ethical guidelines in data analysis and reporting and privacy.
PO6	Evaluate implications of statistical misuse in social research
PO7	Communicate statistical findings to technical and non-technical audiences
PO8	Collaborate with peers in cross-functional teams to interpret data
PO9	Create independent research studies using statistical tools
PO10	Design an environment for continuous upskilling in data science tools and techniques

Assessment

Weightage for assessments (in percentage)

Type of Course	Formative Assessment / IA	Summative Assessment
Theory	**40**	**60**
Practical	**15**	**35**
Projects	**30**	**70**
Experiential Learning (Internships, etc.)	**30**	**70**

Summary of Credits of Core Courses			
Semester	**Course Code**	**Title of the Paper**	**Credits**
I	ST 124	Basic Statistics - I	3
	ST 1P24	Practical based on ST 124	2
II	ST 224	Basic Statistics - II	3
	ST 2P24	Practical based on ST 224	2
III	ST 325	Statistical Inference - I	3
	ST 3P25	Practical Based on ST 325	2
IV	ST 425	Statistical Inference- II	3
	ST 4P25	Practical based on ST 425	2
V	ST 5126	Sampling theory & Statistical Quality Control	3
	ST 5P126	Practical based on ST 5126	2
	ST 5226	Quantitative Forecasting Techniques	3
	ST 5P226	Practical based on ST 5226	2
	ST 5326	Applied Statistical Methods in Lifesciences & Economics	3
	ST 5P326	Practical based on ST 5326	2
VI	ST 6126	Analysis of Variance & Design of Experiments	3
	ST 6P126	Practical based on ST 6126	2
	ST 6226	Operations Research	3

	ST 6P226	Practical on ST 6226	2
	ST 6326	Introduction to Data Analytics	3
	ST 6P326	Practical based on ST 6326	2

Practical Knowledge / Skill Enhancement Courses		
Title	**Subject Code**	**No. of Credits**
Data Querying and Visualization using SQL and Power BI	ST5S126	02
Applied Reliability Analysis and Predictive Modelling	ST6S226	02

Multi-Disciplinary Elective Courses		
Title	**Subject Code**	**No. of Credits**
Exploratory Data Analysis using software	STMDE 3125	02
Applied Statistics	STMDE 4125	02

Elementary Research Methodology		
Title	**Subject Code**	**No. of Credits**
Research Methodology	ST5RM26	02

List of Discipline Specific Electives (DSE)	
• Actuarial Statistics • Advanced Statistical Inference • Analysis of Categorical Data • Analysis of Clinical Trials • Artificial Intelligence with R • Bayesian Inference • Bio-Statistics • Computational Statistics • Data Analytics with R/Python • Data Science: Multivariate Techniques with R /Python • Data Science with R/Python • Demography • Extreme value Theory • Econometrics	• Financial Statistics • Multivariate Techniques • Nonparametric and Semiparametric Methods • Operations Research • Project Work • Reliability Analysis • Reliability and Statistical Quality Control • Statistical Learning and Data Mining with R/Python • Statistical Quality Control • Stochastic Models in Finance • Survival Analysis • Time Series Analysis • Sampling Theory and Applications

Curriculum Structure for the Undergraduate Degree

Program: B.Sc., in Statistics

Total Credits for the Program: 132
Starting year of implementation: 2024-25
Name of the Degree Program: B. Sc.
Discipline/Subject: Statistics (Major)

Program Articulation Matrix

This matrix lists only the core courses. Core courses are essential to earn the degree in that discipline/subject. They include courses such as theory, laboratory, project, internships etc.

Elective courses may be listed separately.

Semester	Title / Name of the course	Program outcomes that the course addresses (not more than 3 per course)	Pre-requisite course(s)	Pedagogy##	Assessment$
1	**Basic Statistics - I**	PO1,PO2,PO8	Mathematics of 12th level	1. The course is taught using traditional chalk andtalk method using problemsolving through examples and exercises. 2. Students are encouraged to use resources available on open sources.	The assessment is done using continuous assessment through written test, openbook examination, viva-voce, seminars, and group discussions.
1	**Practical**	PO5, PO6	Mathematics of 12th level	The course is taught usingExcel software and/or manually to carry out descriptive statistical analysis.	Assessment of learning through experiments
2	**Basic Statistics - II**	PO7,PO9,PO 10	Mathematics of 12th level	1. The course is taught using traditional chalk andtalk method using problemsolving through examples and exercises. 2. Students are encouraged to use resources available on open sources	The assessment is done using continuous assessment through written test, open book examination, viva-voce, seminars, and group discussions.
2	**Practical**	PO5, PO6	Mathematics of 12th level	The course is taught using Rprogramming software and/or manually to carry out descriptive statisticalanalysis	Assessment of learning through experiments

Pedagogy for student engagement is predominantly lectures. However, other pedagogies enhancing better student engagement to be recommended for each course. The list includes active learning/course projects/ problem or project-based learning/ case studies/self-study like seminar, term paper or MOOC course.
$ Every course needs to include assessment for higher order thinking skills (Applying/ Analyzing/ Evaluating/ Creating). However, this column may contain alternate assessment methods that help formative assessment (i.e., assessment for learning).

Programme Outcomes:

PO1: Understand the basic concepts of statistics

PO2: Apply the key concepts of statistical knowledge in real world applications

PO3: Analyse datasets to identify patterns and inconsistencies

PO4: Identify, evaluate and validate statistical models to address research problems

PO5: Apply ethical guidelines in data analysis and reporting and privacy.

PO6: Evaluate implications of statistical misuse in social research

PO7: Communicate statistical findings to technical and non-technical audiences

PO8: Collaborate with peers in cross-functional teams to interpret data

PO9: Create independent research studies using statistical tools

PO10: Design an environment for continuous upskilling in data science tools and techniques.

Course Pre-requisite(s): II PUC with Mathematics

Course Outcomes (CO's):

At the end of the course the student should be able to:

1. Acquire knowledge of introductory statistics, its scope and importance in various areas such as Medical, Engineering, Agricultural and Social Sciences etc.
2. Get knowledge of various types of data, their organization and evaluation of summary measures such as measures of central tendency and dispersion etc.
3. Perceive the knowledge of correlation, regression analysis, regression diagnostics, partial and multiple correlations.
4. Learn different of types of data reflecting independence and association between twoor more attributes.
5. Develop ability to critically assess a standard report having graphics, probability statements.
6. Conceptualize the probabilities of events including frequentist and axiomatic approach. Simultaneously, they will learn the notion of conditional probability including the concept of Bayes' Theorem,
7. Get knowledge related to concept of discrete and continuous random variables and their probability distributions including expectation and moments,
8. Learn knowledge of important discrete and continuous distributions such as Binomial, Poisson, Normal distributions.
9. Acquire knowledge on R-programming in the descriptive statistics and probability models.

Course Articulation Matrix: Mapping of Course Outcomes (COs) with Program Outcomes (POs 1-12)

Course Outcomes (COs) / Program Outcomes (POs)	1	2	3	4	5	6	7	8	9	10	11	12
1. Knowledge of introductory statistics, its scope and importance in various areas such as Medical, Engineering, Agricultural and Social Sciences etc.	X	X			X	X						
2.Knowledge of various types of data, their organization and evaluation of summary measures suchas measures of central tendency and dispersion etc.			X	X	X	X				X	X	
3. Knowledge of correlation, regression analysis,regression diagnostics, partial and multiple correlations.				X	X	X		X		X	X	
4.Knowledge of types of data reflecting independence and association between two or more attributes				X	X	X				X		X
5. Develop ability to critically assess a standard reporthaving graphics, probability statements.					X	X	X		X			
6. Knowledge to conceptualize the probabilities of events including frequentist and axiomatic approach. Simultaneously, they will learn the notion of conditional probability including the concept of Bayes' Theorem.					X	X			X	X		
7. Knowledge related to concept of discrete and continuous random variables and their probability distributions including expectation and moments.					X	X			X	X		
8. Knowledge of important discrete and continuous distributions such as Binomial, Poisson, Normal, distributions.					X	X			X	X		
9. Knowledge on R-programming in the descriptive statistics and probability models.					X	X			X	X		

Course Articulation Matrix relates course outcomes of course with the corresponding program outcomes whose attainment is attempted in this course. 'X' in the intersection cell indicates that particular course outcome addressesthat particular program outcome.

Bachelor of Science (Statistics)

First Semester
ST 124: Basic Statistics - I

Semester	**I**
Paper Code	**ST 124**
Paper Title	**Basic Statistics - I**
Number of teaching hours per week	03
Total number of teaching hours per semester	45
Number of credits	03
Formative Assessment Marks	40
Summative Assessment Marks	60

Unit – 1: Introduction to Statistics **11 Hours**

Statistics: Definition and scope. Concepts of statistical population and sample. Census and survey; Types of Sampling (Simple, Stratified and Systematic random sampling- Definitions only). Data: quantitative and qualitative, cross sectional and time-series, discrete and continuous. Scales of measurement: nominal, ordinal, interval and ratio. Classification and Tabulation of data. Frequency distributions: frequency and Cumulative frequency distributions. Diagrammatic and graphical representations of data: Bar diagram, Pie chart, Histogram, frequency Polygon, frequency curve, Ogives and Stem and Leaf.

Unit – 2: Univariate Data Analysis **14 Hours**

Measures of Central Tendency: Arithmetic Mean, Weighted mean, Median, Mode, Geometric and harmonic means, properties, merits and limitations, relation between these measures. Partition values; Quartiles, Deciles, Percentiles. Measures of Dispersion: Range, Quartile deviation, Mean deviation, Standard deviation and their relative measures, properties, merits and limitations. Moments, Skewness and Kurtosis. Box Plot - Outliers.

Unit – 3: Bivariate Data Analysis **10 Hours**

Bivariate Data, Scatter plot, Correlation: Karl Pearson's correlation coefficient and its properties, Spearman's Rank correlation. Curve fitting: Concept of errors, Principle of least squares, polynomial and exponential curves. Simple linear regression: fitting and interpretation of coefficients, Coefficient of determination (R-Square).

Unit –4: Introduction to Probability **10 Hours**

Experiment: Deterministic and Random experiment, trial, sample space and events. Classical, empirical and axiomatic approaches to probability – illustrations and applications. Addition rule, Conditional probability, Independence of events and Multiplication rule. Total probability rule, Bayes theorem - applications.

References

1. Balakrishnan, N., Koutras, M. V., & Politis, K. G. (2019), Introduction to probability: models and applications, John Wiley & Sons.
2. Freedman, D., Pisani. R and Purves. R. (2014), Statistics, 4th Edition, W. W. Norton & Company.
3. Gupta, S.C. (2018), Fundamental of Statistics, Himalaya Publishing House, 7thEdition.
4. Gupta S.C. and V.K. Kapoor (2020), Fundamental of Mathematical Statistics, Sultan Chandand Co. 12th Edition.
5. Hogg, R. V. McKean J. W. and Craig, A. T. (2012), Introduction to Mathematical Statistics, Pearson 7th Edition.
6. Joao Mendes Moreira, Andre C P L F de Carvalho, Tomas Horvath (2018), General Introduction to Data Analytics, Wiley.
7. Johnson, R.A. and Bhattacharyya, G.K. (2006), Statistics: Principles and methods. 5th Edition, John Wiley & Sons, New York.
8. Medhi, J. (2005), Statistical Methods, New Age International.
9. Ross, S.M. (2014), Introduction to Probability and Statistics for Engineers and Scientists, 5th Edition, Academic Press.
10. Rao, B. P. (2009), A first course in probability and statistics, World Scientific.

Pedagogy

1. The course is taught using traditional chalk and talk method using problem solving through examples and exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**ST 124**
Title of the paper:		**Basic Statistics - I**
Unit number	**Number of hrs.**	**Total marks**
1	11	22
2	14	27
3	10	19
4	10	20
Total	45	88
Maximum marks for the paper (Excluding bonus questions)		60

<u>Note:</u> For the current batch, blueprint of having weightage 35% pre Mid-Sem and 65% post Mid-Sem portions will be followed.

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

CO1	Apply knowledge of statistical concepts and data collection methods to effectively to organize and presentdata using appropriate tabular and graphical techniques.
CO2	Demonstrate the ability to analyze and apply various measures of central tendency anddispersion, including mean, median, mode, standard deviation, and quantiles, as well as interpret data distributions using box plots etc.
CO3	Analyze the relationship between two variables using bivariate data analysis techniques likescatter plots, correlation coefficients, and linear regression and draw meaning conclusions.
CO4	Explain the fundamental concepts of probability theory, including random experiments, sample spaces, events, different approaches to defining probability (classical, statistical, subjective, axiomatic), and their applications in various scenarios.

Course Outcomes (COs)	**Program Outcomes (POs)**									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*		*		*			
CO2	*	*	*							
CO3	*	*	*	*			*			
CO4	*	*	*							

Practical I
ST 1P24 – Practical on ST 124

Number of teaching hours per week	03
Total number of teaching hours per semester	45
Number of credits	02
Formative Assessment Marks	15
Summative Assessment Marks	35

Note: Computing all the practical's manually and using MS Excel

1. Presentation of data by frequency tables, diagrams and graphs, stem and leaf, partition values.
2. Measures of Central Tendency: Arithmetic Mean (AM), Geometric mean, Harmonic mean, Weighted AM,Corrected mean.
3. Computation of Median, Mode, Partition values.
4. Problems on Absolute and relative measures of dispersion, Box plots.
5. Problems on moments, skewness and kurtosis.
6. Fitting of curves by least squares method.
7. Computation of Product moment correlation coefficient and rank correlation.
8. Fitting Regression line of two variables.
9. Computation of probability using combinatorial methods.
10. Computation of probability using addition and multiplication theorem. Application of conditional probability, Bayes theorem.

Bachelor of Science (Statistics)

Second Semester
ST 224: Basic Statistics – II

Semester	II
Paper Code	**ST 224**
Paper Title	**Basic Statistics - II**
Number of teaching hours per week	03
Total number of teaching hours per semester	45
Number of credits	03
Formative Assessment Marks:	40
Summative Assessment Marks:	60

Unit –1: Random Variables and Mathematical Expectations **12 Hours**

Discrete and continuous random variables, Distribution function, probability mass function (p.m.f) and probability density functions (p.d.f) – properties and illustrations, Expectation of a random variable and rules of expectation, variance and related results.
Joint, marginal and conditional distributions, Independence of random variables, Moments, covariance and correlation coefficient, Addition and multiplication rules of expectation, Mean and variance of linear combination of random variables, Conditional expectation. Moments and moment generating function (m.g.f) – properties and uses.

Unit –2: Discrete and Continuous Probability Distributions **16 Hours**

Discrete probability distributions:
Discrete Uniform, Bernoulli, Binomial distributions: mean, variance, m.g.f., recurrence relation for probabilities, additive property of Binomial distribution. Poisson, Geometric distributions: mean, variance, and m.g.f. recurrence relations for probabilities, additive property of Poisson and geometric distributions, Lack of memory property of geometric distribution. Negative-Binomial and Hyper geometric distributions: mean, variance.

Continuous probability distributions:
Uniform, exponential distributions – definition through p.d.f.s, mean, variance, m.g.f., lack of memory property and additive property of exponential distribution. Gamma, Beta, distributions: definition through p.d.f.s, Normal distribution: definition through p.d.f., standard normal distribution, median, mode, m.g.f, area property, odd order and even order moments and linear combination of normal variates.

Unit –3: Limit Theorems and Random Number Generation **10 Hours**

Chebyschev's inequality – proof and its use in approximating probabilities for various discrete and continuous distributions. Convergence of Binomial, Poisson, Gamma distributions to Normal distribution. Statement of Central Limit Theorem and its applications.

Introduction to Simulation, Monte Carlo method. Generation of random observations from uniform, exponential, Normal, Binomial, Poisson distributions, simple illustrations.

Unit –4: Introduction to R software **7 Hours**

Introduction to R: Installation, command line environment, overview of capabilities, brief mention of open-source philosophy. Use of parentheses nesting up to arbitrary level. The power operation. Evaluation of simple expressions. Quotient and remainder operations for integers. Standard functions, e.g., exp, log. The different types of numbers in R: Division by zero leading to Inf or -Inf. NaN, NA. No need to go into details. Inserting Variables, Creating a vector using c(), seq() and colon operator. Functions to summarize a vector: sum, mean, sd, median etc. Extracting a subset from the vector (by index, by property). Introduction to plotting. Plot(), lines(), abline(). No details about the graphics parameters except colour and line width. Barplot, Pie chart and Histogram. Box plot, Scatter plot and simple linear regression using lm(y~x).

References

1. Balakrishnan, N., Koutras, M. V., & Politis, K. G. (2019), Introduction to probability: models and applications, John Wiley & Sons.
2. Dudewitz. E.J. and Mishra. S. N. (1998), Modern Mathematical Statistics. John Wiley.
3. Goon A.M., Gupta M.K., Das Gupta. B. (1991), Fundamentals of Statistics, Vol. I, World Press, Calcutta.
4. Gupta. S.C and V.K. Kapoor (2020), Fundamentals of Mathematical Statistics, Sultan Chand and Co, 12th Edition.
5. Hogg, R.V., Tanis, E.A. and Rao J.M. (2009), Probability and Statistical Inference, Seventh Edition, Pearson Education, New Delhi.
6. Mood, A.M., Graybill, F.A. and Boes, D.C. (2007), Introduction to the Theory of Statistics, 3rd Edition. (Reprint), Tata McGraw-Hill Pub. Co. Ltd.
7. Mukhopadhyay, N. (2020). Probability and statistical inference. CRC Press.
8. Ross, S. (2002), A First Course in Probability, Prentice Hall.
9. Ross, S. M. (2014). Introduction to probability models. Academic press.
10. Sudha G. Purohit, Sharad D. Gore, Shailaja R Deshmukh, (2009), Statistics using R, Narosa Publishing House.
11. R for beginners by Emmanuel Paradis (freely available at https://cran.rproject.org/doc/contrib/Paradisrdebuts_en.pdf)
12. Rao, B. P. (2009), A first course in probability and statistics, World Scientific.
13. Dhanavanthan, P and Sakthivel, K.M (2024): Elementary theory of probability distributions, Ane Publishers and distributors LLP, New Delhi.
14. Venables W.N., Smith D.M.,and the R development core team (2008): An Introduction to R. Network publishers.
15. Varzani, J (2014): Using R for Introductory statistics, 2nd edition, CRC Press, Chapman & Hall book.

Pedagogy

1. The course is taught using traditional chalk and talk method using problem solving through examplesand exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**ST 224**
Title of the paper:		**Basic Statistics - II**
Unit number	**Number of hrs.**	**Total marks**
1	12	23
2	16	31
3	10	20
4	07	14
Total	45	88
Maximum marks for the paper (Excluding bonus questions)		60

Note: **For the current batch, blueprint of having weightage 35% pre Mid-Sem and 65% post Mid-Sem portions will be followed.**

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

CO1	Analyze the characteristics of random variables (discrete or continuous) using probability distributions, calculate expected values and interpret moments and moment generating functions to understand the central tendency and spread of data.
CO2	Apply their knowledge of common probability distributions (Bernoulli, Binomial, Poisson, Uniform, Exponential, Normal) to real-world scenarios by identifying the appropriate distribution based on data characteristics and fitting the distribution to the data.
CO3	Demonstrate knowledge of the convergence of Binomial, Poisson, and Gamma distributions to the Normal distribution under specific conditions and understand the basic principles and purpose of simulation in statistical analysis.
CO4	Use R programming to perform basic data analysis tasks, including creating and manipulating data structures (vectors), calculating summary statistics, and generating informative visualizations (plots, histograms, scatter plots) to explore and understand data.

Course Outcomes (COs)	Program Outcomes (POs)									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*							
CO2	*	*	*	*						
CO3	*	*	*	*						
CO4	*	*	*	*			*	*	*	*

Practical II

ST 2P24 – Practical on ST 224

Number of teaching hours per week	03
Total number of teaching hours per semester	45
Number of credits	02
Formative Assessment Marks	15
Summative Assessment Marks	35

Note: Computing all the practical's manually and using R Software

1. Descriptive statistics (Presentations, Summarizations, correlations, regression and Graphs) using R.
2. Univariate probability distribution: Expectation, moments, skewness and kurtosis.
3. Bivariate Probability Distribution: Joint, Marginal, Conditional, and correlation coefficient, Conditional Expectation.
4. Applications of discrete probability distributions: Binomial, Poisson and Geometric
5. Fitting of discrete distributions: Binomial, Poisson, Geometric
6. Applications of Continuous probability distributions: Uniform and Exponential
7. Fitting of continuous probability distributions: Uniform and Exponential
8. Fitting and Applications of Normal distribution
9. Generation of random observations: Binomial, Poisson, Geometric, Uniform, Exponential and Normal using Simulation
10. Applications of Chebyschev's inequality and Central Limit Theorem

Bachelor of Science (Statistics)

Third Semester
ST 325: Statistical Inference–I

Semester	**III**
Paper Code	**ST 325**
Paper Title	**Statistical Inference-I**
Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**03**
Formative Assessment Marks:	**40**
Summative Assessment Marks:	**60**

Unit -1: Family of Distributions and Theory of Point Estimation: 15 Hours

Family of distributions: Single parameter exponential family, K - parameter exponential family, Location - Scale family, Pitman family – Definitions and examples.

Point Estimation - Concepts of estimator and estimate, criteria of a good estimator - Unbiasedness, Consistency, sufficient conditions for consistency, Invariance property of consistent estimator, Sufficient Statistic - Statement of Neyman Factorization theorem, Minimum variance unbiased estimator, Efficiency and Relative efficiency, mean square error. Problems on estimation from some Standard distributions (Bernoulli, Binomial, Poisson, Geometric, Negative Binomial, Uniform, Normal and Exponential distributions only).

Unit -2: Methods of Point Estimation: 8 Hours

Methods of Point Estimation: Maximum likelihood estimator (MLE) and method of moments: Properties, Illustration for non-uniqueness and invariance property of MLE. Estimation of parameters for Bernoulli, Binomial, Poisson, Geometric, Negative Binomial, Normal, Exponential, Continuous Uniform distributions.

Unit - 3: Sampling Distributions and Interval Estimation: 13 Hours

Sampling distributions and Standard error of a statistic and its uses. Sampling distribution of the sample mean, sample variance, and their independence under normality assumption. Chi-square, t, and F distribution and their statistic under normality assumption.

Confidence Interval, Confidence Coefficient, Pivotal quantity method of constructing Confidence Intervals, construction of confidence intervals for: the mean, the difference between two means, variance and ratio of variances, proportion, the difference between proportions, and correlation coefficient (small and large samples) .

Unit - 4: Testing of Hypotheses **6 Hours**

Definition of Statistical hypotheses and its types - Null and Alternative, Simple and Composite hypotheses. Type-I and Type-II Errors, Size of the test, Level of significance, Power of the test and Power function and its computations. Test function - Randomized and non-randomized test (definition only). Critical region, P-value and its interpretation.

References:

1. Freund J.E. (2001): Mathematical Statistics, Prentice Hall of India.
2. Nitis Mukhopadhyay (2000): Probability and Statistical Inference, Marcel Dekker, Inc, New York.
3. Goon A.M., Gupta M.K., Das Gupta.B. (1991): Fundamentals of Statistics, Vol. I, World Press, Calcutta.
4. Hogg R.V. and Tannis E.A. (1988): Probability and Statistical Inference, Collier MacMillan.
5. Hodges J.L and Lehman E.L (1974): Basic Concepts of Probability and Statistics, Holden-Day.
6. Mood A.M, Graybill F.A and Boes D.C. (1974): Introduction to the Theory of Statistics, McGraw Hill.
7. Gupta. S.C. and V.K.Kapoor (2001): Fundamentals of Mathematical Statistics. Sultan Chand & Co.
8. Bhattacharya and Johnson (1986): Statistical Concepts, Wiley Int. Ed.
9. Rohatgi. V.K. and A.K. Md. Ehsanes Saleh (2002): An Introduction to Probability Theory and Mathematical Statistics. John Wiley (WSE).
10. Ross S.M (1987): Introduction to Probability and Statistics for Engineers & Scientists, John Wiley & Sons.
11. Rajagopalan M & Dhanavanthan P (2012): Statistical Inference, Eastern Economy Edition, PHI Learning Private Limited, New Delhi.

Pedagogy

1. The course is taught using traditional chalk and talk method using problem solving through examples and exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**ST 325**
Title of the paper:		**Statistical Inference-I**
Unit number	**Number of Hrs.**	**Total Marks**
1	15	31
2	8	17
3	13	27
4	6	13
Total	42	88
Maximum marks for the paper (Excluding bonus questions)		60

<u>Note</u>: For the current batch, blueprint of having weightage 35% pre Mid-Sem and 65% post Mid-Sem portions will be followed.

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

CO1	Describe the family of distributions, including the single-parameter exponential family, location-scale family and explain their applications with examples.
CO2	Learn the concepts of point estimation to standard distributions, such as estimating parameters from Bernoulli, Poisson, Normal, Exponential distributions, etc.
CO3	Acquire the concept of confidence intervals to estimate population parameters based on sample data.
CO4	Apply the concepts of hypothesis testing to formulate and analyze real life datasets and draw meaningful conclusions.

Course Outcomes (COs)	**Program Outcomes (POs)**									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*							
CO2	*	*	*	*						
CO3	*	*	*	*						
CO4	*	*	*	*	*	*	*		*	

Practical - III

ST 3P25 –Practical on ST 325

Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**02**
Formative Assessment Marks	**15**
Summative Assessment Marks	**35**

Computing all the practical manually and using Excel/R

1. Properties of estimators on unbiasedness, consistency and efficiency.

2. Comparison of estimators by plotting mean square error using MS-Excel and R.

3. Estimation of parameters by Maximum Likelihood method.

4. Estimation of parameters by Method of Moments.

5. Estimation of parameters by Maximum Likelihood method and Method of Moments (Using R).

6. Construction of confidence intervals-I (large samples).

7. Construction of confidence intervals-II (small samples).

8. Construction of confidence intervals using R & MS-Excel.

9. Testing of Hypothesis-I (Formulation of hypotheses, understanding types of hypotheses, understanding types of tests).

10. Testing of Hypothesis-II (Understanding Type I error and Type II Error)

Bachelor of Science (Statistics)

Fourth Semester

ST 425: Statistical Inference-II

Semester	**IV**
Paper Code	**ST 425**
Paper Title	**Statistical Inference–II**
Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**03**
Formative Assessment Marks:	**40**
Summative Assessment Marks:	**60**

Unit – 1: MP and UMP tests **14 Hours**

Most Powerful (MP) test. Statement of Neyman – Pearson Lemma and its applications. MP test for parameters of Bernoulli, Binomial, Poisson and Normal distributions.

Monotone likelihood ratio (MLR) property. Uniformly most powerful (UMP) test. Statement of the theorem on UMP test for testing one-sided hypothesis for distributions with MLR property. (Only parameter of Bernoulli and mean of Normal Distribution)

Unit – 2: Likelihood Ratio tests **6 Hours**

Likelihood ratio tests (LRT) and its properties. LRT's for mean and variance of normal distribution (one-sample problem), LRT for the mean of Exponential distribution.

Unit - 3: Tests of Significance **12 Hours**

Large and small sample tests of significance with corresponding assumptions: Tests for proportions and differences between proportions, Tests for single mean, equality of two means, single variance, and the ratio of two variances. Tests for correlation coefficient and regression coefficients (only slope and intercept (t-test)). Fisher's Z-transformation and its applications (test for non-zero correlation coefficient).
Analysis of categorical data- contingency table, Chi-square test for independence of attributes in a contingency table, Chi-square test for goodness of fit, and equality of several proportions using Chi-square test.

Unit – 4: Non-Parametric tests **10 Hours**

Need for distribution-free tests - non-parametric tests. one sample and two samples sign tests, Wilcoxon signed- rank test, Median test, Wald Wolfowitz run test, Mann Whitney Wilcoxon test, run test for randomness, test for independence based on Spearman's rank correlation coefficient (small and large samples), Kolmogorov-Smirnov one-sample test and two-sample test, test for more than two samples (Kruskal Wallis test and Friedman test), Normal probability plot and QQ Plot.

References:

1. Rohatgi.V.K.and A.K.Md. Ehsanes Saleh (2002): An Introduction to Probability theory and Mathematical Statistics, John Wiley(WSE).
2. Hogg R.V and Tannis, E.A.(1988): Probability and Statistical Inference, Collier MacMillan.
3. Mukhopadhyay. P (1996): Applied Statistics, Calcutta Publishing House.
4. Gupta S. C and V.K.Kapoor (2001): Fundamentals of Mathematical Statistics, Sultan Chand & Co.
5. Gopal K.Kanji (2006): 100 Statistical tests,3rd Edition, Sage Publications
6. Mood A.M, Graybill F.A and Boes D.C. (1974): Introduction to the Theory of Statistics, McGraw Hill.
7. Nitis Mukhopadhyay (2000): Probability and Statistical Inference, Marcel Dekker, Inc, New York.
8. Sheldon M Ross (2017): Introductory Statistics, 4th Edition, Academic Press Publishers.
9. Dudewicz. E.J.and Mishra.S.N.(1980): Modern Mathematical Statistics, John Wiley.
10. Kale B.K (2005): A First Course on Parametric Inference,2nd Edition, Narosa Publishers.
11. Randles R.H and Wolfe DA(1979):Introduction to the Theory of Non-parametric Statistics, John Wiley
12. Rajagopalan M & Dhanavanthan P (2012): Statistical Inference, Eastern Economy Edition, PHI Learning Private Limited, New Delhi.

Pedagogy

1. The course is taught using traditional chalk and talk method using problem solving through examples and exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**ST 425**
Title of the paper:		**Statistical Inference–II**
Unit Number	**Number of Hrs.**	**Total Marks**
1	14	29
2	06	13
3	12	25
4	10	21
Total	42	88
Maximum marks for the paper (Excluding bonus questions)		60

<u>Note</u>: For the current batch, blueprint of having weightage 35% pre Mid-Sem and 65% post Mid-Sem portions will be followed.

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

CO1	Understand and apply the concepts of MP and UMP tests for standard distributions.
CO2	Explain the concept of likelihood ratio test and its importance.
CO3	Apply the statistical inference tools in real data analysis.
CO4	Understand and apply the concept of nonparametric tests to analyze data when the assumptions of normality for parametric tests are not met.

Course Outcomes (COs)	**Program Outcomes (POs)**									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*	*						
CO2	*	*	*	*						
CO3	*	*	*	*	*	*	*		*	
CO4	*	*	*	*		*	*			

Practical - IV
ST 4P25 – Practical on ST 425

Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**02**
Formative Assessment Marks	**15**
Summative Assessment Marks	**35**

Computing all the practical manually and using Excel/R

1. MP test, UMP test and LR test for parameters of Bernoulli distribution & Normal Distributions.

2. Tests concerning means and proportions (small and large sample)

3. Tests concerning variances (small and large sample)

4. Tests for correlation coefficients and Regression coefficient (slope and intercept)

5. Tests for Independence of attributes and test for homogeneity (Chi-square test)

6. Tests for goodness of fit.(Uniform, Binomial, Poisson and Normal)

7. Small and Large sample test using MS Excel and R

8. Nonparametric tests–1 (for single and related samples)
(Sign test and Wilcoxon signed rank test, Kolmogorov Smirnov test and one sample Run test)

9. Nonparametric tests–2 (for two independent samples)
(Median test, Wilcoxon Mann Whitney –U test, Wald-Wolfowitz's Run test and Kruskal Wallis test)

10. Tests for Normality, Histogram, QQ Plot using R.

Bachelor of Science (Statistics)

Fifth Semester

ST 5126: Sampling Theory and Statistical Quality Control

Semester	**V**
Paper Code	**ST 5126**
Paper Title	**Sampling Theory and Statistical Quality Control**
Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**03**
Formative Assessment Marks:	**40**
Summative Assessment Marks:	**60**

UNIT 1: Preliminaries and Data Collection Methods **06 Hrs**

Basic definitions - Population, Sample, Sampling unit, Sampling fraction, Sampling Frame, Population Parameter, Estimator, Standard error, Accuracy, Precision, margin of error. Data Collection methods- Census, Sample surveys, Experiments and Observational studies. Methods of collecting primary data, Drafting of questionnaires, schedules and their pre-test. Requisites of a good questionnaire. Pilot surveys. Sampling and Non-Sampling errors and simple methods of controlling them.

UNIT 2: Basic Sampling Techniques **15 Hrs**

Simple random Sample (SRS) with and without replacement. Unbiased estimators of population mean and total. Derivation of the variances of the estimators of mean and their estimation. Standard errors of the estimators. Construction of confidence intervals for mean(s). Determination of sample size for mean (one sample).

Linear Systematic sampling - Advantages and limitations, estimation of population means.

Stratified random sampling-Need for stratification. Unbiased estimators of population mean and total. Derivation of the variance of the estimators and their estimation. Proportional, Optimum and Neyman allocations (without derivation). Comparison of variances with SRSWOR. Estimation of gain in precision due to stratification.

Other sampling techniques: Cluster Sampling and two stage sampling (Concept only).

UNIT 3: Statistical Process Improvement **13 Hrs**

Meaning and scope of statistical quality control, Causes of quality variation, The seven QC tools, Control charts for variables and attributes - Rational subgroups - Construction and operation of $\bar{x}$, σ, R, np, p, c and u charts – OC and ARL curve of control charts. Specification limits and tolerance limits, Process capability ratios and their interpretations. Design and implementation of SPC and six-sigma programmes, BIS and ISO certification.

UNIT 4: Acceptance Sampling and Quality Assurance **8 Hrs**

Acceptance sampling inspection by attributes – single and double, Measures of performance: OC, ASN, ATI and AOQ functions. Concepts of AQL, LTPD and IQL. Dodge – Romig tables and MIL-STD-105D tables.

References:

1. Cochran W G., (2007): Sampling Techniques, 3rd edition, Wiley Eastern.
2. Grant, E.L., & Leavenworth, R.S. (1988). Statistical Quality Control. McGraw-Hill.
3. Oakland, J., Oakland, R. J. (2018). Statistical Process Control. United Kingdom: Taylor & Francis.
4. Montgomery, D. C. (2020). Introduction to Statistical Quality Control, United Kingdom: Wiley.
5. Murthy M N (1967): Sampling Theory & Methods, 2nd Edition, Statistical Publishing Society, Calcutta.
6. Mukhopadhyay P (2008): Theory and Methods of Survey Sampling, 2nd Edition, Prentice Hall India Learning Private Limited.
7. Sampath, S. (2005). Sampling theory and methods, 2nd Edition, Alpha Science International Ltd, Harrow, U.K.
8. Singh. D and Chaudhary, F. S (1986): Theory and Analysis of sampling survey design. (Wiley Eastern).

Pedagogy:

1. The course is taught using traditional chalk and talk method using problem solving through examples and exercises.

2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**ST 5126**
Title of the paper:		**Sampling Theory & Statistical Quality Control**
Unit number	**Number of Hrs.**	**Total Marks**
1	06	13
2	15	31
3	13	27
4	08	17
Total	42	88
Maximum marks for the paper (Excluding bonus questions)		60

Course Outcomes (COs):

At the end of the Course, the student should be able to

CO1	Define and differentiate key concepts related to sampling and data collection methods, including methods for collecting primary data and drafting questionnaires.
CO2	Apply sampling techniques (SRS, systematic, stratified, cluster, two stage) to estimate population parameters and determine appropriate sample sizes.
CO3	Understand the fundamentals of Statistical Quality Control (SQC), including the meaning and scope of SQC, identifying causes of quality variation, applying the seven QC tools, and constructing and interpreting control charts for variables and attributes.
CO4	Evaluate acceptance sampling plans by attributes (single and double), calculate and interpret performance measures, and apply concepts of AQL, LTPD, and IQL, leveraging industry standards like Dodge-Romig and MIL-STD-105D tables.

Course Outcomes	**Program Outcomes(POs)**									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*								
CO2	*	*	*				*			
CO3	*	*	*	*						
CO4	*	*		*			*			

Practical - V

ST 5P126 –Practical on ST 5126

Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**02**
Formative Assessment Marks	**15**
Summative Assessment Marks	**35**

Computing all the practical's manually and using Excel/R

1. Simple random sampling (SRSWR)
2. Simple random sampling (SRSWOR)
3. Stratified random sampling
4. Systematic random sampling
5. Methods of sampling using R software
6. Variable control charts
7. Attributes Control Charts
8. OC and ARL curves of X bar and R control chart
9. Single sampling plans for attributes
10. Double sampling Plans for attributes

Bachelor of Science (Statistics)

Fifth Semester

ST 5226: Quantitative Forecasting Techniques

Semester	**V**
Paper Code	**ST 5226**
Paper Title	**Quantitative Forecasting Techniques**
Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**03**
Formative Assessment Marks:	**40**
Summative Assessment Marks:	**60**

Unit 1: Fundamentals of Regression Analysis **08 Hrs**

Simple Linear Regression: model and assumptions, Fitting the model using least squares method, interpretations. Coefficient of determination and its interpretation. Hypothesis testing: Overall significance and significance of individual regression coefficients, confidence intervals for parameters, standard errors and prediction intervals.

Unit- 2: Multiple Linear Regression **08 Hrs**

Multiple Linear Regression Models: Assumptions, Least-squares estimates and their properties, Model fitting, residuals, Coefficient of determination. Tests for significance of regression, tests on individual regression coefficients. Confidence intervals in multiple linear regression and prediction of a new observation.

Unit- 3: Model Adequacy and Generalized Linear Model **16 Hrs**

Different methods of variable selections – forward, backward, step-wise (only concept), Model Adequacy – Residual Analysis, tests for normality – QQ Plot. Introduction to Heteroscedasticity, Multicollinearity, and Autocorrelation, measures of identifying above situations, and remedial measures. Introduction to logistic regression.

Unit 4: Exploratory Time Series Analysis **10 Hrs**

Time series analysis: Introduction, components of time series. Additive and multiplicative models.

Test for randomness, Tests for trend and seasonality Measurements of trend by using moving averages and least squares methods. Construction of seasonal indices by simple averages and ratio to moving averages and its applications. Forecasting trend and seasonality. De-seasonalizing and detrending an observed time series.

References:

1. Draper N. R. and Smith H. (2003). Applied Regression Analysis, 3rd Ed., John Wiley.
2. Goon, A.M., Gupta, M.K. and Dasgupta, B., (2005): Fundamentals of Statistics. Vol. II, 8th Edition. World Press, Kolkata.
3. Gupta, S. C., Kapoor, V. K. (2007). Fundamentals of Applied Statistics. India: Sultan Chand & Sons.
4. Hosmer, D. W., Lemeshow, S., Sturdivant, R. X. (2013). Applied Logistic Regression. United Kingdom: Wiley.
5. Montgomery, D. C., Peck, E. A., Vining, G. G. (2021). Introduction to Linear Regression Analysis. United States: Wiley.
6. Parimal Mukhopadhyay (2022). Applied Statistics, 3rd Edition, Books and Allied (P) Ltd, Kolkata.

Pedagogy

1. The course is taught using traditional chalk and talk method using problem solving through examples and exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**ST 5226**
Title of the paper:		**Quantitative Forecasting Techniques**
Unit number	**Number of Hrs.**	**Total Marks**
1	08	17
2	08	17
3	16	33
4	10	21
Total	42	88
Maximum marks for the paper (Excluding bonus questions)		60

Course Outcomes (COs):

At the end of the Course, the student should be able to

CO1	Explain the concepts, assumptions, and interpretations of simple linear regression models, including hypothesis testing and confidence intervals.
CO2	Apply multiple linear regression models, interpret coefficients, check assumptions, and evaluate significance tests for model parameters and predictions.
CO3	Evaluate model adequacy through residual analysis and normality checks, diagnose issues, propose remedies, introduce logistic regression concepts, and demonstrate variable selection methods with interpretation of their relevance in model building.
CO4	Perform exploratory time series analysis, decompose series into components, detect trend and seasonality, and construct forecasts using moving averages and least squares methods.

Course Outcomes (COs)	**Program Outcomes (POs)**									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*		*					
CO2	*	*	*	*	*					
CO3	*	*	*	*						*
CO4	*	*	*	*	*	*				*

Practical - VI

ST 5P226 – Practical on ST 5226

Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**02**
Formative Assessment Marks	**15**
Summative Assessment Marks	**35**

Computing all the practical's manually and using Excel/R

1. Simple linear regression
2. Multiple linear regression
3. Fitting Regression model for Multivariate Data using MS-Excel and R
4. Model Adequacy-Residual Analysis, t-test, and F-test
5. Multicollinearity and variable selection techniques
6. Heteroscedasticity
7. Autocorrelation
8. Tests for trend and seasonality
9. Measurement of trend
10. Measurement of Seasonal variation

Bachelor of Science (Statistics)

Fifth Semester

ST 5326: Applied Statistical Methods in Life Sciences & Economics

Semester	**V**
Paper Code	**ST 5326**
Paper Title	**Applied Statistical Methods in Life Sciences and Economics**
Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**03**
Formative Assessment Marks:	**40**
Summative Assessment Marks:	**60**

Unit 1: Basics of Survival Analysis **12 Hrs**

Introduction to survival analysis, censoring and its types (Type-I and Type-II). Functions of survival time and life time distributions – Exponential and Gamma. Non-parametric estimation of survival functions: Kaplan-Meier and Nelson-Aalen estimator.

Introduction to Proportional Hazards, Cox-Proportional hazards model, Interpretation of hazard ratios.

Unit 2: Index Numbers and Inflation Measurement **10 Hrs**

Index numbers: Introduction. Price and quantity index numbers. Construction of index numbers: Simple and weighted methods. Tests for consistency of index numbers. Consumer price index numbers -Construction of consumer price index numbers and its methods, uses and limitations, Wholesale price Index number, Definition and measurement of Inflation rate - CPI and GNP Deflator.

Unit 3: Demand Analysis **10 Hrs**

Introduction, Demand, Supply, Price elasticity of supply and demand, Partial and Cross elasticities, Methods of determining demand and supply curves; Family Budget data, Engel curves, Pareto Law of income distribution.

Unit 4: Official Statistics and National Income Analysis **10 Hrs**

Official Statistics and national income: History of Indian Statistical System. Pre and post-independence era. CSO NSSO and their activities. National income - Basic concepts of GNP, GDP, NNP. National Income at factor cost – NDP, per capita income. Real national income. Methods of estimating national income. Problems in estimating national income. Uses of national income statistics. National accounts statistics of CSO. Measuring inequality in incomes, Lorenz Curve, Gini coefficient, and Theil's measure.

References

1. Collett D., (2003). Modelling Survival Data in Medical Research, Second Edition Chapman & Hall/CRC Texts in Statistical Science.
2. Goon A.M., Gupta, M.K., Das Gupta, B. (1991). Fundamentals of Statistics, Vol. II, World Press, Calcutta.
3. Lee E. T., Wang J., (2003). Statistical Methods for Survival Data Analysis, 3rd edition, John Wiley & Sons, Inc.
4. Moore D F., (2016): Applied Survival Analysis using R, 1st edition, Springer, Berlin.
5. Saluja, M. R. (1972). Indian Official Statistical Systems, Statistical Publishing Society, Calcutta.
6. CSO (1980). National Accounts Statistics - Sources and Health, Govt. of India, New Delhi.
7. UNESCO: Principles for Vital Statistics Systems. Series M -12.
8. Sen, A. (1997). Poverty and Inequality, Stanford University Press, USA.
9. Mukhopadhyay, P. (2015). Applied Statistics, Books and Allied Pvt Ltd., Kolkata.

MODEL BLUEPRINT		
Code number:		**ST 5326**
Title of the paper:		**Applied Statistical methods in Life Sciences & Economics**
Unit number	**Number of Hrs.**	**Total Marks**
1	12	25
2	10	21
3	10	21
4	10	21
Total	42	88
Maximum marks for the paper (Excluding bonus questions)		60

Course Outcomes:

At the end of the Course, the student should be able to

CO1: Apply survival analysis techniques, including censoring concepts, survival function estimation, and proportional hazards modeling, to analyze and interpret time dependent data.
CO2: Explain the concepts of index numbers and inflation, and describe methods for constructing and interpreting price indices and inflation measures.
CO3: Interpret demand and supply concepts and apply elasticity measures to analyze market behavior.
CO4: Develop awareness about various statistical organizations and estimation methods of national income.

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*	*					*	*
CO2	*	*				*				
CO3	*	*	*							
CO4	*	*			*	*	*	*		

Practical - VII

ST 5P326 –Practical on ST 5326

Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**02**
Formative Assessment Marks	**15**
Summative Assessment Marks	**35**

Computing all the practical's manually and using Excel/R

1. Lifetime distribution
2. Actuarial and Kaplan Meier estimate
3. Hazard curves of lifetime distributions using R
4. Construction of index numbers and consumer price index numbers
5. Tests for consistency of index numbers
6. Price elasticity of demand and supply
7. Fitting Engel Curves from family budget data
8. Pareto Law of income distribution
9. Methods of estimating national income
10. Lorenz Curve, Gini coefficient and Theil's measure

Bachelor of Science (Statistics)

Sixth Semester

ST 6126: Analysis of Variance & Design of Experiments

Semester	**VI**
Paper Code	**ST 6126**
Paper Title	**Analysis of Variance & Design of Experiments**
Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**03**
Formative Assessment Marks:	**40**
Summative Assessment Marks:	**60**

UNIT 1: Analysis of Variance **8 Hrs**

Meaning, assumptions and applications. Fixed, random and mixed effect models. Analysis of One -way and two-way classified data with and without interaction effects. Multiple comparison tests: Tukey's and Critical difference method.

UNIT 2: Basics of Experimental Designs and CRD **10 Hrs**

Basic concepts and principles of design of experiments. Completely randomized design (CRD)- Implementation of Basic principles of design in CRD, layout formation, estimation of parameters, analysis using fixed effect model. Determination of sample size, ANCOVA for CRD.

UNIT 3: Randomized block and Latin square design **14 Hrs**

Randomized block design (RBD) - Implementation of Basic principles of design in RRD, layout formation, estimation of parameters, analysis using fixed effect model.
Latin square design (LSD) - Implementation of Basic principles of design in LSD, layout formation, estimation of parameters, analysis using fixed effect model. Comparison of efficiencies of CRD, RBD and LSD.
Estimation of one missing observation in RBD and LSD and its analysis.

UNIT 4 – Factorial Experiments **10 Hrs**

Basic concepts – main and interaction effects, contrasts and orthogonal contrasts in 2^2 and 2^3 factorial experiments. Yates' method of computing factorial effects total. Analysis of 2^2 and 2^3 factorial experiments in RBD,
Confounding in factorial Experiments: Need for confounding, Types of confounding - Complete and partial, Confounding in a 2^3 - factorial experiment in RBD and its analysis.

References

1. Cochran, W.G. and Cox, G.M. (1959): Experimental Design. Asia Publishing House.
2. Das, M.N. and Giri, N.C. (1986): Design and Analysis of Experiments. Wiley Eastern Ltd.
3. Goon, A.M., Gupta, M.K. and Dasgupta, B. (2005): Fundamentals of Statistics. Vol. II, 8th Edition. World Press, Kolkata.
4. Gupta S C and Kapoor V K (2014): Fundamentals of Applied Statistics, 4th Edition, Sultan Chand & Sons
5. Montgomery, D. C. (2008): Design and Analysis of Experiments, John Wiley.
6. Parimal Mukhopadhyay (2022). Applied Statistics, 3rd Edition, Books and Allied (P) Ltd, Kolkata.
7. Wu, C. F. J. And Hamada, M. (2009). Experiments, Analysis and Parameter Design Optimization, 2nd Edition, John Wiley & Sons.

Pedagogy:

1. The course is taught using traditional chalk and talk method using problem solving through examples and exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**ST 6126**
Title of the paper:		**Analysis of Variance & Design of Experiments**
Unit number	**Number of Hrs.**	**Total Marks**
1	8	17
2	10	21
3	14	29
4	10	21
Total	42	88
Maximum marks for the paper (Excluding bonus questions)		60

Course Outcomes (COs):

At the end of the Course, the student should be able to

CO1	Understand the core concepts, assumptions, and practical applications of Analysis of Variance (ANOVA), differentiate model types, and perform one-way and two-way analysis.
CO2	Comprehend the fundamental principles of experimental design and implement, and analyze experiments using Completely Randomized Design (CRD), determine sample sizes and use ANCOVA for CRD when appropriate.
CO3	Design, implement, and analyze experiments using Randomized Block Design (RBD) and Latin Square Design (LSD), including estimating missing observations and comparing the efficiencies of CRD, RBD, and LSD.
CO4	Understand main and interaction effects in factorial experiments, compute effects using Yates' method, analyze 2^2 and 2^3 factorial experiments in RBD, and explain complete and partial confounding.

Course Outcomes	**Program Outcomes (POs)**									
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*							
CO2	*	*	*				*			
CO3	*	*	*	*			*			*
CO4	*	*	*	*			*			*

Practical - VIII

ST 6P126 –Practical on ST 6126

Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**02**
Formative Assessment Marks	**15**
Summative Assessment Marks	**35**

Computing all the practical's manually and using Excel/R

1. One way ANOVA and CRD
2. Two-way ANOVA with and without interaction effect
3. Multiple comparison test
4. ANCOVA for CRD
5. RBD
6. LSD
7. Missing plot technique in RBD and LSD
8. 2^2 and 2^3 factorial experiments
9. Complete and Partial confounding in 2^3 factorial experiments
10. Design of experiments using R software

Bachelor of Science (Statistics)

Sixth Semester
ST 6226: Operations Research

Semester	**VI**
Paper Code	**ST 6226**
Paper Title	**Operations Research**
Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**03**
Formative Assessment Marks:	**40**
Summative Assessment Marks:	**60**

Unit-1: Introduction to OR and LPP **10 Hrs**

Definition and scope of operations research (OR), Modelling and solution.

Linear programming problem (LPP) – Definition, Standard and canonical forms. Formulation of LPP. Basic solutions, Degenerate and non-degenerate solution. Fundamental theorem of LPP (Statement Only). Graphical solution and Simplex algorithm for solving an LPP, Criteria for unbounded, multiple and infeasible solutions. Introduction of Artificial variable -Charnes' Big- M Method, applications.

Unit-2: Transportation problem, Assignment Problem and Game theory **12 Hrs**

Transportation problem: Mathematical formulation, finding an initial basic feasible solution by North West corner rule, and Vogel's method. Transportation algorithm and test for optimality by U-V method (MODI Method). Problem of degenerate solution and its resolution, applications.

Assignment problem*:* Mathematical formulation and Hungarian algorithm, Job sequencing, optimal sequence of N jobs on two and three machines, applications.

Game theory: Basic concepts. Two-person zero sum game, Pure and mixed strategies. Maximin–Minimax principle, Games with saddle point. Principle of dominance. Games without Saddle point. -mixed strategies, Determination of optimum solution for (2x2) game. Solution by graphical method for (2xn) and (mx2) games, applications.

Unit-3: Network Analysis and Inventory Theory **13 Hrs**

Basic elements of Network, Drawing of project network. Project planning with CPM and PERT. Critical path calculation - Critical path, slack time, floats. PERT – three time estimate approach. Calculation of probabilities of completing a project within a specified period.

Introduction to Inventory theory: Types of Inventory costs. Demand, lead time and reorder level. Inventory models - EOQ model with and without shortages, P – system and Q – system of Inventory.

Unit-4: Queuing theory **07 Hrs**

Introduction to Queuing theory - basic concepts, description of a queuing system and measures of effectiveness. M/M/1 queuing system - Steady state solution, waiting time distributions, Little's formula, derivation of expressions for Queue length, and system size(length), applications. M/M/C and M/M/∞ Queuing models (Concepts only)

References:

1. Churchman, C.W, Ackoff, R.L and Arnoff, E.L. (1957): Introduction to Operations Research. John Wiley.
2. Hillier, F. S., & Liebermann, G. J. (2010). Introduction to Operations Research*, 9th edition,* McGraw-Hill, New York.
3. Kalavathy, S. (2004): Operations research, Vikas Publishing House Pvt Ltd, New Delhi.
4. Kanthiswaroop, Manmohan and P.K. Gupta (2003). Operations Research. Sultan Chand & Co.
5. Mustafi, C.K. (2006). Operations Research: Methods and Practice, 3rd Edition, New Age International, New Delhi.
6. Mital, K.V. and Mohan, C. (2004). Optimization Methods, 3/e, New Age International, New Delhi. Kapoor, V. K. Operations Research. Sultan Chand & Co.

7. N. D. Vohra (2017). Quantitative Techniques in Management, 5th edition, Tata McGraw Hill.
8. S D Sharma (2009), Operations Research: Theory and Applications, Fourth Edition, Macmillan Publishers India Limited.
9. S. D. Sharma, R. K. Malhotra, (1997). Operations Management, 4th edition, Anmol publications Pvt Ltd.
10. Shenoy, G.V., Srivastava, U.K and Sharma, S.C.,(2009) : Operations Research for management. New Age Publications.
11. Taha, H. A. (2019): Operations Research*:* An introduction, 10th Edition, Pearson Education, India.

Pedagogy

1. The course is taught using traditional chalk and talk method using problem solving through examples and exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**ST 6226**
Title of the paper:		**Operations Research**
Unit number	**Number of Hrs.**	**Total Marks**
1	10	21
2	12	25
3	13	27
4	07	15
Total	42	88
Maximum marks for the paper (Excluding bonus questions)		60

Course Outcomes:

By the end of the course, students will be able to:

CO1: Apply graphical and simplex methods to solve Linear Programming Problems, including handling cases like infeasibility, unboundedness, and degeneracy.
CO2: Solve optimization problems using transportation, assignment, and game theory.
CO3: Plan and manage projects using CPM and PERT, and optimize operations using inventory and queuing models.
CO4: Analyze queuing models to evaluate system performance and efficiency.

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*		*						
CO2	*	*		*						
CO3		*	*	*				*	*	*
CO4	*	*	*	*						

Practical - IX

ST 6P226 –Practical on ST 6226

Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**02**
Formative Assessment Marks	**15**
Summative Assessment Marks	**35**

Computing all the practical's manually and using Excel/R

1. Formulation of Linear programming problem (L.P.P) and Graphical solution
2. Solution of L.P.P – Using simplex algorithm.
3. Solution of L.P.P – Using Big M Method
4. Transportation Problem-I
5. Transportation Problem-II
6. Assignment Problem and Job Sequencing problem
7. Game theory problem
8. PERT and CPM
9. Inventory problem
10. Queuing Problem

Bachelor of Science (Statistics)

Sixth Semester

ST 6326: Introduction to Data Analytics

Semester	**VI**
Paper Code	**ST 6326**
Paper Title	**Introduction to Data Analytics**
Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**03**
Formative Assessment Marks:	**40**
Summative Assessment Marks:	**60**

Unit 1: Overview of Data Analytics **05 Hrs**

Introduction and scope of Data Analytics. Classification of analytics: Descriptive, Diagnostic, Predictive, and Prescriptive types, applications in various fields. Ethics in data analytics, Difference between data analytics & data science. Quantity of data, Measurement scales.

Unit 2: Introduction to Python **10 Hrs**

Introduction to Python environments: Introduction to Python programming and the Jupyter Notebook interface. Overview of variables and data types: integers, floats, strings, Booleans, type conversion. Operators: Arithmetic, Relational, Logical, Assignment, etc.

Packages: numpy, pandas, datetime, scipy, S-klearn, seaborn, Matplotlib.

Data structures: lists, tuples, dictionary, Data-frames. Dictionary comprehension, List comprehensions, Lambda functions.

Control Structures: Conditional statements (if, else, elif) and loops (for, while, break, continue). Functions: In-built and user-defined functions.

File handling: reading from and writing to text and csv files.

Unit 3: Fundamentals of Data Management and Exploration **17 Hrs**

Common Data Quality Concepts and Issues, Data Pre-processing: removing duplicates, handling missing values- Deletion and imputation techniques. Data transformation: Standardisation, Dummy encoding, One-hot, Label encoding.

Descriptive statistics: Measures of central tendency, partition values, Measures of dispersion, Outlier detection-boxplots, etc.

Visual representation: Scatter plots, histograms, box plots, strip plots, count plots, and heatmaps, time-series plot, etc.

Correlation and Regression analysis, cross- tabulations (for variables & attributes).

Unit 4: Introduction to Machine Learning **10 Hrs**

Definition and scope of Machine Learning. Overview of learning paradigms: Supervised, Unsupervised learning with illustrative examples. Description of the training-validation-testing pipeline. Core challenges in ML including trade-offs between prediction accuracy and interpretability, the bias–variance trade-off (underfitting vs. overfitting), and impacts of insufficient or non-representative data. Role and relevance of Feature Engineering.

Supervised Classification models: KNN, Naive Bayes, Decision Tree, Logistic Regression with examples. Unsupervised Classification models: K-Means clustering algorithm.

References:

1. Arun K Pujari (2010). Data Mining Techniques, Second Edition, Universities Press India Pvt. Ltd.
2. Davy Cielen, Arno D. B. Meysman, Mohamed Ali (2016). Introducing Data Science, First Edition, Manning Publications Co.
3. D J Patil, Hilary Mason, Mike Loukides (2018). Ethics and Data Science, First Edition, O' Reilly.
4. E. Alpaydin (2014). Introduction to Machine Learning, 2nd Edition, MIT Press.
5. James et al., (2023). An introduction to statistical learning with applications in Python, Springer nature, New York.
6. Jake Vander Plas (2016). Python Data Science Handbook - Essential Tools for Working with Data, O'Reily Media Inc.
7. Jiawei Han, Micheline Kamber, Jian Pie (2012). Data Mining Concept and Techniques, 3rd Edition, Morgan and Kaufmann Publisher.
8. Josh Starmer., (2022). The StatQuest Illustrated guide to machine learning, Qurate Books Pvt. Ltd.
9. McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. O'Reilly Media, Inc.
10. Murphy K P (2012). Machine Learning: A Probabilistic Perspective, MIT Press.
11. OpenStax. (2019). Introductory Statistics (2nd Ed.). OpenStax, Rice University.
12. Tyagi A. Kumar (2021). Data Science and Data Analytics. CRC Press.
13. Zhang.Y (2016). An Introduction to Python and Computer Programming, Springer Publications.

MODEL BLUEPRINT		
Code number:		**ST 6326**
Title of the paper:		**Introduction to Data Analytics**
Unit number	**Number of Hrs.**	**Total Marks**
1	05	10
2	10	21
3	17	36
4	10	21
Total	42	88
Maximum marks for the paper (Excluding bonus questions)		60

Course Outcomes:

At the end of the Course, the student should be able to

CO1: Identify analytics types, data sources, and quality issues.
CO2: Demonstrate basic Python programming using Jupyter Notebook.
CO3: Apply data cleaning and visualization techniques in Python.
CO4: Apply machine learning techniques to regression, classification and clustering problems.

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*			*	*				
CO2		*								*
CO3	*	*	*							*
CO4		*	*	*			*	*		*

Practical - X

ST 6P326 –Practical on ST 6326

Number of teaching hours per week	**03**
Total number of teaching hours per semester	**42**
Number of credits	**02**
Formative Assessment Marks	**15**
Summative Assessment Marks	**35**

Computing all the practical's using R and Python

1. Data Visualization using Python
2. Working With Data Structures using Python
3. Data Cleaning and Handling Missing observations
4. Correlation and Regression Analysis using Python
5. KNN and Naïve Bayes models using R
6. KNN and Naïve Bayes models using Python
7. Logistic regression and Decision Trees using R
8. Logistic regression and Decision Trees using Python
9. K-means Clustering using R
10. K-means Clustering using Python

Bachelor of Science (Statistics)

Fifth Semester

ST 5S126: Data Querying and Visualization using SQL and Power BI

Semester	V
Course Code	ST 5S126
Course Title	Data Querying and Visualization using SQL and Power BI
Number of teaching hours per week	2
Total number of teaching hours per semester	30
Number of credits	2
Formative Assessment Marks:	15
Summative Assessment Marks:	35
Course Type	Skill Enhancement / Practical Knowledge Course

Course Objectives

1. Train students to extract and manipulate data using SQL, as used in real industry databases.
2. Enable students to prepare and structure datasets for analysis and visualization.
3. To introduce data cleaning, modelling, and visualization using Power BI.
4. To develop interactive dashboards for statistical data analysis and decision-making.

Unit – 1: Foundations of SQL and Data Extraction **12 Hours**

Data and databases, limitations of file-based systems, database management system and its advantages, types of databases, relational database concepts including relation, tuple, attribute, schema and instance, tables, records and fields, keys such as primary key, candidate key, composite key and foreign key, relationships between tables, Difference between flat files (Excel/CSV) and relational databases.

Introduction to SQL, syntax rules, different data types- numeric, character, date/time and boolean. Constraints: NOT NULL, UNIQUE, CHECK and DEFAULT, Data Definition Language (DDL): CREATE, ALTER, DROP, TRUNCATE. Data Manipulation Language (DML): INSERT, UPDATE, DELETE, transactions: COMMIT and ROLLBACK and Data Query Language (DQL): SELECT, FROM, WHERE, ORDER BY, LIMIT

Data aggregation: Core Aggregate Functions- COUNT, SUM, AVG, MIN, MAX, VARIANCE, VAR_POP, VAR_SAMP, STDDEV, STDDEV_POP, STDDEV_SAMP. Grouping (Grouped Data)-GROUP BY. Filtering and ordering: HAVING, ORDER BY. Conditional Aggregation-CASE WHEN;

Exporting query results to Excel/CSV for visualizations.

Unit – 2: Data Preparation and Relational Queries **8 Hours**

Joins: SQL Commands- INNER JOIN, LEFT JOIN, RIGHT JOIN, FULL JOIN. Importance of joins in real-world datasets; Handling missing values in SQL; Removing duplicates; Creating derived variables using CASE WHEN for real-world datasets; Saving and exporting cleaned data for Tableau.

Unit - 3: Power BI – Data Visualization and Dashboarding **10 Hours**

Introduction to Business Intelligence and Power BI. Power BI Desktop: interface and data sources. Data cleaning and transformation using Power Query. Data modelling basics. Data visualization tools: Tables, charts, slicers, filters. Introduction to DAX: calculated columns and measures. Designing interactive dashboards. Power BI and SQL integration, Case study: Statistical dashboard using real-world data.

References

SQL:

1. Trueblood, Robert P., and John N. Lovett. *Data Mining and Statistical Analysis Using SQL*. Vol. 1. Berkeley, CA: Apress, 2001.
2. Linoff, Gordon S. *Data analysis using SQL and Excel*. John Wiley & Sons, 2015.
3. Beaulieu, Alan. *Learning SQL: Generate, manipulate, and retrieve data*. O'Reilly Media, 2020.

Tableau:

1. Machiraju, Suren, and Suraj Gaurav. Power Bi data analysis and visualization. DeG Press, 2018.
2. Dunlop, Neil. Beginning Big Data with Power BI and Excel 2013: Big Data Processing and Analysis Using PowerBI in Excel 2013. Apress, 2015.
3. Microsoft. Power BI Guided Learning.
4. Wexler, Steve, Jeffrey Shaffer, and Andy Cotgreave. *The big book of dashboards: visualizing your data using real-world business scenarios*. John Wiley & Sons, 2017.
5. Tableau Software. *Tableau Desktop and Web Authoring Help*. Tableau Official Documentation. https://help.tableau.com/current/pro/desktop/en-us/default.htm

Pedagogy

1. The course is taught using the traditional chalk and talk method, using problem solving through examples, using various statistical software and case-study.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT	
Code number:	**ST 5S26**
Title of the paper:	Data Querying and Visualization using SQL and Power BI

Unit number	Number of hrs.	Total marks
1	12	22
2	08	15
3	10	18
Total	30	55
Maximum marks for the paper (Excluding bonus questions)		**35**

Course Outcomes (COs)

CO1: Write basic and intermediate SQL queries to extract and manipulate data from databases.
CO2: Clean, transform, and structure real datasets using SQL for analysis.
CO3: Create interactive visualizations and dashboards using Power BI.
CO4: Communicate statistical insights effectively through reports and dashboards.

CO-PO Mapping:

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*							*
CO2	*	*	*	*	*					*
CO3		*	*				*	*		*
CO4	*		*		*	*	*	*		*

Bachelor of Science (Statistics)
Sixth Semester
ST 6S226: Applied Reliability Analysis and Predictive Modelling

Semester	**VI**
Paper Code	**ST 6S26**
Paper Title	**Applied Reliability Analysis and Predictive Modelling**
Number of teaching hours per week	02
Total number of teaching hours per semester	30
Number of credits	02
Formative Assessment Marks	15
Summative Assessment Marks	35
Course Type	Skill Enhancement / Practical Knowledge Course
Course Prerequisite(s)	Basic Statistics, Probability Theory, Distribution Theory and Statistical Inference

Course Objectives:

By the end of this course, students will be able to:

1. Understand reliability data and basic reliability measures used in industrial applications.
2. Develop practical skills in modelling failure-time data using standard reliability distributions.
3. Apply simulation techniques for reliability estimation of components and systems.
4. Perform predictive estimation of failures and Remaining Useful Life (RUL).

Unit I: Introduction to Reliability 8 Hours

Reliability concepts and measures: reliability and failure, failure-time data, reliability function, hazard rate, mean time to failure (MTTF). Components and systems. Series and parallel systems. Reliability of systems with independent components. Reliability block diagrams. Notions of ageing. Censored data and its types. Structure and characteristics of industrial reliability datasets. Demonstration of selected reliability case studies.

Unit II: Lifetime Models, Estimation and Simulation 12 Hours

Lifetime Models: Exponential, Gamma, Weibull and Log-normal. Interpretation of model parameters. Reliability measures and their computation. Estimation of parameters using MLE. Estimation of reliability, hazard function and MTTF. Estimation under Censored Data.

Monte Carlo simulation for component and system reliability, Generation and analysis of simulated lifetime data. Comparison of analytical and simulation-based reliability estimates.

Unit III: Predictive Estimation 10 Hours

Predictive estimation: concept, Predictive density, Point prediction of lifetimes; Interval prediction of lifetimes; Remaining Useful Life (RUL), Remaining Quantile Life. Prediction under censored data.

References

1. Rausand, M., & Hoyland, A. (2003). *System reliability theory: models, statistical methods, and applications* (Vol. 396). John Wiley & Sons.
2. Meeker, W. Q., Escobar, L. A., & Pascual, F. G. (2021). *Statistical methods for reliability data*. John Wiley & Sons.
3. Lawless, J. F. (2011). *Statistical models and methods for lifetime data*. John Wiley & Sons.
4. Barlow, R. E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models.* Holt.
5. Sinha, S. K. (1986). *Reliability and Life Testing*. Wiley Eastern Limited.
6. Dersin, P. (2023). *Modeling remaining useful life dynamics in reliability engineering*. CRC Press.
7. Saunders, S. C. (2007). *Reliability, Life Testing and the Prediction of Service Lives,* Springer.
8. Everitt, B. S, and Hothorn, T. (2009). A Handbook of Statistical Analyses Using R. Chapman and Hall/CRC.

Pedagogy

1. The course is taught using the traditional chalk and talk method, using problem solving through examples, using various statistical software and exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**ST 6S26**
Title of the paper:		Applied Reliability Analysis and Predictive Modelling
Unit number	**Number of hrs.**	**Total marks**
1	08	15
2	12	22
3	10	18
Total	30	55
Maximum marks for the paper (Excluding bonus questions)		**35**

Course Outcomes (COs)

After successful completion of the course, the student will be able to:

CO1: Understand reliability concepts and characteristics of industrial failure-time data.
CO2: Model failure-time data using standard reliability distributions and estimate reliability measures.
CO3: Estimate reliability of components and systems using analytical and simulation-based methods.
CO4: Perform predictive estimation and RUL analysis for reliability-based decision making.

CO-PO Mapping:

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*	*			*			*
CO2	*	*	*	*			*		*	*
CO3	*	*	*	*			*	*	*	*
CO4	*	*	*	*	*	*	*	*	*	*

Bachelor of Science (Statistics)

Fifth Semester

ST 5RM26: Research Methodology

Semester	V
Paper Code	**ST5RM26**
Paper Title	**Research Methodology**
Number of teaching hours per week	**02**
Total number of teaching hours per semester	**30**
Number of credits	02
Formative Assessment Marks:	20
Summative Assessment Marks:	30
Course Type	**Elementary research methodology**

Course Objectives:

By the end of this course, students will be able to:

1. Understand research and its significance in academic and applied settings.
2. Identify and define research problems.
3. Conduct a comprehensive literature review.
4. Use statistical tools and software to analyse data.
5. Write well-organised research reports and present findings effectively.

Unit 1: Introduction to Research Methodology 8 Hours

Meaning and Definition: Research methods and Methodology. Types of research: Descriptive Vs Analytical, Fundamental Vs Applied, Quantitative Vs Qualitative, Conceptual Vs Empirical. Research Formulation: Defining and formulating the research problem. Systematic literature review, motivation of the study, identification of the research gap and construction of research objectives. Meaning and importance of research ethics. Development of a working hypothesis. Characteristics of good research problems in Statistics.

Unit 2: Research Design, Data Collection and Sampling Techniques 12 Hours

Research design: Basic Principles, Need and features of good research design. Developing a research plan, Research questions and formulation of objectives. Hypotheses: Framing research questions, Null and alternative hypotheses in a research context. Collection of data: Methods of data collection and validation. Missing data: Types of missing data, consequences of missing observations, Basic handling methods (deletion, imputation – conceptual). Sampling Techniques: Probability and non-probability sampling. Sampling errors and non-sampling errors. Data fabrication, falsification and plagiarism. Data processing and analysis strategies. Introduction to Statistical Software for Research.

Unit 3: **Research Report Preparation and Presentation** **10 Hours**

Report and report writing: Structure and components of scientific reports. Different steps in report preparation. Layout, structure, and Language of typical reports. Illustrations and tables. Bibliography: Citation indexing and impact factor, h-index and i10-index, Databases: Google Scholar, Scopus, Web of Science. Referencing, Plagiarism, and Footnotes.

Meaning and importance of a research proposal. Components: Title, problem statement, objectives, methodology, analysis and conclusion. Introduction to software: MS Word and LaTeX for report writing. Introduction to Research Dissemination: Conference papers and posters (basic idea), Research seminars and presentations, Open-access journals and repositories. Basics of Bibliographic Management Tools: Mendeley / Zotero (overview), Importance of reference management in research.

References:

1. Kothari, C. R., & Garg, G. (2019). *Research methodology: Methods and techniques* (4th ed.). New Age International Publishers.
2. Kumar, R. (2018). Research methodology: A step-by-step guide for beginners.
3. Sekaran, U., & Bougie, R. (2020). *Research methods for business: A skill-building approach* (8th ed.). Wiley.
4. Creswell, J. W., & Creswell, J. (2003). *Research design* (pp. 155-179). Thousand Oaks, CA: Sage publications.
5. Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.
6. Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
7. Zikmund, W. G., Babin, B. J., Carr, J. C., & Griffin, M. (2013). *Business research methods* (9th ed.). Cengage Learning.
8. Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). SAGE Publications.
9. Booth, W. C., Colomb, G. G., & Williams, J. M. (2016). *The craft of research* (4th ed.). University of Chicago Press.
10. Leedy, P. D., & Ormrod, J. E. (2020). *Practical research: Planning and design* (12th ed.). Pearson.

Pedagogy

1. The course is taught using the traditional chalk and talk method, using problem solving through examples, using various statistical software and exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**ST 5RM26**
Title of the paper:		**Research Methodology**
Unit number	**Number of hrs.**	**Total marks**
1	08	15
2	12	22
3	10	18
Total	30	55
Maximum marks for the paper (Excluding bonus questions)		**35**

Course Outcomes: At the end of the Course, the Student should be able to

CO1: Understand the meaning, types, and formulation of research problems.
CO2: Conduct a literature review, identify research gaps, and set objectives
CO3: Apply suitable research designs, data collection, and sampling methods.
CO4: Analyse data using statistical techniques and hypothesis testing.
CO5: Prepare and present a scientific research report using standard tools and formats.

CO- PO mapping table

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	BTL
CO1	*	*					*		*		Understand
CO2			*	*					*		Analyse
CO3		*	*		*				*		Apply
CO4		*	*	*		*			*	*	Analyse
CO5				*	*		*	*	*	*	Create

Bachelor of Science (Statistics)

STMDE 3125: Exploratory Data Analysis using Software

Course Objectives: After completing this course, students will be able to:

1. **Understand** the basic concepts of statistical data, its types, collection, classification, and graphical representation.
2. **Learn and apply** measures of central tendency, dispersion, moments, skewness, and kurtosis to summarize univariate data.
3. **Develop** bivariate analysis techniques including correlation and simple linear regression.
4. **Use** MS Excel and R to organize, visualize, and analyse univariate and bivariate data.
5. **Perform** statistical analysis to draw meaningful conclusions from real datasets.

Course Pre-requisite: II PUC or equivalents with Mathematics

Semester	**IV**
Paper Code	**STMDE 3125**
Paper Title	**Exploratory Data Analysis using Software**
Number of teaching hours per week	**02**
Total number of teaching hours per semester	**28**
Number of credits	**02**
Formative Assessment Marks	**20**
Summative Assessment Marks	**30**

UNIT-I

Introduction to Statistics **18 Hours**

Concept of statistical population and sample from a population. Types of data: Quantitative and Qualitative, Collection of data: Primary and Secondary. Classification and Tabulation, Diagrammatic and Graphical representation of a data.

Univariate Data Analysis: Measures of Central Tendency - Arithmetic Mean, Median and Mode and partition values, characteristics of a good average, merits and demerits. Dispersion: Relative and Absolute measures - Range, Quartile Deviation, Mean Deviation and Standard Deviation. Coefficient of variation and its application. Moments: raw and central moments. Skewness and Kurtosis – definition, examples and applications.

Bivariate Data Analysis: Bivariate data: Concept of Correlation - Types, Measurements and Properties. Concept of Regression – Simple linear regression and its fitting, Regression Coefficients and its properties, Coefficient of determination (R^2).

UNIT-II

Univariate and Bivariate data analysis using MS Excel and R **10 Hours**

Understanding data classification. Constructing frequency tables. Data Visualization tools - Bar charts, Pie Charts, Histogram, Box plots, Stem-and-leaf plot.

Measures of Central Tendency: Arithmetic Mean, Weighted Arithmetic Mean, Median, Mode, Skewness and Kurtosis**.**
Measures of Dispersion: Absolute and Relative Measures of Dispersion**.**
Correlation & Regression Analysis: Simple linear regression - Fitting regression lines and interpreting the results and Coefficient of determination (R^2).

References

1. Balakrishnan, N., Koutras, M. V., &Politis, K. G. (2019): Introduction to probability: Models and Applications, John Wiley & Sons.
2. Sheldon M Ross (2017): Introductory Statistics, 4th Edition, Academic Press Publishers.
3. Neil A Weiss (2012): Introductory statistics, 9th edition, Addison-Wesley, Pearson publishers, New York.
4. Gupta, S.C. (2018), Fundamental of Statistics, Himalaya Publishing House, 7thEdition.
5. Gupta S.C. and V.K. Kapoor (2020), Fundamental of Mathematical Statistics, Sultan Chand and Co. 12th Edition.
6. Hogg, R. V. McKean J. W. and Craig, A. T. (2012), Introduction to Mathematical Statistics, Pearson 7th Edition.
7. Sudha G. Purohit, Sharad D. Gore, Shailaja R Deshmukh, (2009): Statistics using R, Narosa Publishing House.
8. Paul Newbold, William L. Karlin, and Betty Thorne (2012): Statistics for Business and Economics, 8th Edition, Pearson publishers.
9. Hadley Wickham and Garrett Grolemund (2017): R for Data Science, Import, Tidy, Transform, Visualize, and Model Data (Greyscale Indian Edition), O'Reilly Media.
10. R for beginners by Emmanuel Paradis (freely available at https://cran.rproject.org/doc/contrib/Paradisrdebuts_en.pdf)
11. Venables W.N., Smith D.M., and the R-development core team: An Introduction to R.
12. Varzani, J (2014): Using R for Introductory statistics, 2nd edition, CRC Press, Chapman & Hall book.

Note:

1. This is an open elective course for other than statistics students.
2. The students will learn the elements of descriptive statistics, statistical methods such as correlation and regression along with its analysis using software.

Online Resources:

- Microsoft Excel Help Center
- R Documentation and Tutorials (https://cran.r-project.org/manuals.html)

Pedagogy

1. The course is taught using traditional chalk and talk method using problem solving through examples and exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**STMDE 3125**
Title of the paper:		**Exploratory Data Analysis using Software**
Unit number	**Number of hrs.**	**Total marks**
1	18	27
2	10	15
Total	28	42
Maximum marks for the paper (Excluding bonus questions)		30

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

Course Outcome (CO)	Description
CO1	Learn and understand how to describe and summarize the distribution of data using measures such as mean, median, mode and variance.
CO2	Apply skills in visually representing data through histograms, scatter plots and other graphical methods.
CO3	Implement techniques to quantify and interpret the strength and direction of relationships between variables, assess model fit and make predictions based on regression analysis.
CO4	Examine how to perform regression analysis and correlation using Excel and R for real-world datasets

CO-PO Mapping:

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*							
CO2	*	*	*				*			
CO3	*	*	*	*			*		*	
CO4		*	*	*			*	*	*	*

Bachelor of Science (Statistics)

STMDE 4125: Applied Statistics

Course Objectives: After completing this course, students will be able to

1. **Understand** the demographic measures of mortality, fertility, and life tables for population analysis.
2. **Familiarize** with clinical trial designs, measures of risk, and diagnostic test performance using sensitivity, specificity, and ROC curves.
3. **Develop** statistical techniques for scaling psychological and educational test scores, including Z-scores, T-scores, percentile and normalized scores.
4. **Analyse and interpret** health, demographic, and psychological data to support research and decision-making.

Course Pre-requisite: II PUC or any equivalents with Mathematics

Semester	IV
Paper Code	**STMDE 4125**
Paper Title	**Applied Statistics**
Number of teaching hours per week	02
Total number of teaching hours per semester	28
Number of credits	02
Formative Assessment Marks	20
Summative Assessment Marks	30

UNIT 1: Demography and Clinical Trials **18 hours**

Demography: Sources of demographic data. Measurement of Mortality: crude, specific and standardized death rates. Infant mortality rate, maternal mortality rate. Measurement of fertility: crude, age specific, general and total fertility rates. Reproduction rates, Life table: Components of a life table, force of morality and expectation of life. Construction of a life table, uses of a life table.

Clinical Trails: Introduction, therapeutic trials and prophylactic trials. Observational studies – Cross sectional studies, prospective studies, retrospective studies and randomized control studies. Odds ratio and its confidence interval. Relative risk and its confidence interval. Diagnostic efficacy. Sensitivity, specificity, false negative and false positive rates. Receiver operating characteristic (ROC) curve. Body Mass Index.

UNIT 2: Psychological and Educational Statistics **10 hours**

Scaling of mental test and psychological data, Scaling of scores on a Test – Z Score and Scaling , Standardized scores , Normalised scores, Computation of T – Scores for a given frequency distribution , Comparison of T – Scores and Standardised scores , Percentile scores, Scaling of ranking and ratings in terms of Normal curves Intelligent test – Intelligent Quotient and Educational Quotient.

References

1. Goon A.M., Gupta, M.K., Das Gupta, B. (1991). Fundamentals of Statistics, Vol. II, World Press, Calcutta.
2. Gupta S.C. and V.K. Kapoor (2014), Fundamental of Applied Mathematical Statistics, Sultan Chand & Sons, 4th Edition.
3. Henry E Garrett (2005): Statistics in Psychology and Education, Pearson International Publishers, New Delhi.
4. Mukhopadhyay, P. (2015). Applied Statistics, Books and Allied Pvt Ltd., Kolkata.
5. Sundar Rao, P. S. S. and Richard, J. (2012). Introduction to Biostatistics and Research methods, 5/e, Prentice Hall of India, New Delhi.

Pedagogy

1. The course is taught using traditional chalk and talk method using problem solving through examples and exercises.
2. Students are encouraged to use resources available on open sources.

MODEL BLUEPRINT		
Code number:		**STMDE 4125**
Title of the paper:		**Applied Statistics**
Unit number	**Number of hrs.**	**Total marks**
1	18	27
2	10	15
Total	26	42
Maximum marks for the paper (Excluding bonus questions)		30

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

Course Outcome (CO)	Description
CO1	Understand the sources of demographic data and apply them to construct and interpret life tables, including measures such as force of mortality and life expectancy.
CO2	Differentiatebetween various types of clinical trials, including therapeutic, prophylactic, and observational studies.
CO3	Evaluate the accuracy of diagnostic tests using sensitivity, specificity, false positive/negative rates, and ROC curves.
CO4	Analyse and interpret various psychological test scores for research and practical applications.

CO-PO Mapping:

CO \ PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10
CO1	*	*	*	*			*			
CO2	*	*		*	*	*				
CO3		*	*	*	*	*	*		*	
CO4		*	*	*			*	*	*	