Department of Statistics, M.D. University, Rohtak
Scheme of Examination for One Year (Two Semesters)
Post Graduate Diploma in Data Analytics
w.e.f. 2023-24
(Under NEP-2020)

Sr. No.	Course Code	Nomenclature of the course	Credits	Hours/Week (L + T + P)	Maximum Marks			Examination Hours
					Internal	End Semester	Total	
Semester 1								
1	Discipline Specific Course-I **(23STAD101DS01)**	Descriptive Statistics and Testing of Hypothesis	4	(4+0+0)	30	70	100	**3 hours**
2	Discipline Specific Course-II **(23STAD101DS02)**	Sampling Techniques	3	(3+0+0)	25	50	75	**3 hours**
3	Discipline Specific Course-II (P) **(23STAD101DS02P)**	Practical Based on (23STAD101DS02)	1	(0+0+1)	5	20	25	**3 hours**
4	Discipline Specific Course-III **(23STAD101DS03)**	Regression Models	4	(4+0+0)	30	70	100	**3 hours**
5	Discipline Specific Course-IV **(23STAD101DS04)**	Data Analysis Using SPSS (Based on 23STAD101DS01& 23STAD101DS03)	4	(0+0+4)	30	70	100	**3 hours**
6	Discipline Specific Course-V **(23STAD101DS05)**	Data Handling & Visualization	4	(0+0+4)	30	70	100	**3 hours**
7	Skill Enhancement Course-I **(23STAD101SE01)**	DBMS & SQL	3	(3+0+0)	25	50	75	**3 hours**
8	Skill Enhancement Course-I (P) **(23STAD101SE01P)**	Practical Based on (23STAD101SE01)	1	(0+0+1)	5	20	25	**3 hours**

Sr. No.	Course Code	Nomenclature of the course	Credits	Hours/Week (L + T + P)	Maximum Marks			Examination Hours
					Internal	End Semester	Total	
Semester 2								
1	Discipline Specific Course-VI **(23STAD102DS01)**	Multivariate Analysis	4	(4+0+0)	30	70	100	**3 hours**
2	Discipline Specific Course-VII **(23STAD102DS02)**	Time Series Analysis & SQC	4	(4+0+0)	30	70	100	**3 hours**
3	Discipline Specific Course-VIII **(23STAD102DS03)**	Machine Learning Using Python	3	(3+0+0)	25	50	75	**3 hours**
4	Discipline Specific Course-VIII (P) **(23STAD102DS03P)**	Practical Based on (23STAD102DS03)	1	(0+0+1)	5	20	25	**3 hours**
5	Discipline Specific Course-IX **(23STAD102DS04)**	Data Analysis Using R (Based on 23STAD102DS01& 23STAD102DS02)	4	(0+0+4)	30	70	100	**3 hours**
6	Skill Enhancement Course-II **(23STAD102SE01)**	Project Work	4				100	**3 hours**

Note:

Project Work: The project work will start in the beginning of 2nd semester under the approved supervisors from among the faculty members of the department. Each student shall submit a project report based on the work done using real life data (Primary or Secondary) related to Industry, Government or Private Organizations, Business Organizations and any other area. The student will be required to submit the project work within one month after the theory papers of 2nd semester. The evaluation will be done jointly by the internal examiner and external examiner on the basis of project work report and viva-voce. In case the supervisor of the student(s) shows his/her inability to act as internal examiner, the Head of the Department will work as internal examiner. The distribution of marks as follows:

1. Project Report Evaluation Marks – 70
2. Viva-Voce–30

Name of the Department: STATISTICS
Name of the Major Course: Descriptive Statistics and Testing of Hypothesis
Semester: I

Course Code	**23STAD101DS01**	**Course Credits**	4(L: 4 T: 0 P: 0)
Max. Marks	100 {External (term-end exam) – 70} (Internal – 30)	**Time of end term examination**	3 Hours

Note: The examiner is to set nine (09) questions in all into five sections A, B, C, D and E of the question paper from all the four units- I, II, III and IV. In section A, there will be a compulsory question of 14 marks consisting of 08 short answer type questions two from each unit (each of marks 02) and out of which the student will be required to attempt any 07 questions. In each section B, C, D and E there will be two questions of 14 marks each from Units I, II, III, & IV respectively. The students will be required to attempt five questions in all selecting at least one question from each section.

Course Objectives:

Students will able to:

- Get Awareness of Different Types of Data Set and Their Presentation Including Statistical Measures.
- Understand the Concept of Random Variables.
- Make the Use of Different Probability Distributions in Real Life Situations.
- Gain the Knowledge about Fundamentals of Hypothesis Testing.
- Identify the Situations where Non-Parametric Tests can be used.

Course Outcomes:

Students acquired the:

- Knowledge to Understand Presentation and Interpretation of the Data.
- Understanding of Measures of Central Tendency, Dispersion and Correlation.
- Skill to Make the Use of Probability Distributions in Real Life Situations.
- Ability to Estimate Unknown Parameters of a Given Probability Distribution.
- Knowledge to Apply the Parametric and Non-Parametric Tests.

Unit - I

Basic Statistics: Presentation of Data, Measures of Central Tendency and Dispersion, Skewness and Kurtosis. Analysis and Consistency of Categorical Data, Correlation: Multiple and Partial Correlation and Scatter Plot.

Unit – II

Random Variables (One and Two Dimensional) and Probability Functions, Mathematical Expectation and its Properties. Discrete Probability Distributions: Binomial, Poisson, Geometric. Continuous Probability Distributions: Normal, Uniform and Exponential.

Unit – III

Statistical Estimation: Standard Error, Point Estimation, Characteristics of a Good Estimator, Methods of Estimation: Moments, Least Square and Maximum Likelihood. Testing of Hypothesis: Null and Alternative Hypothesis. Simple and Composite Hypothesis, Critical Region, Level of Significance, Types of Error, Neyman-Pearson Lemma.

Unit – IV

Parametric Tests: t-test, z-test, Chi-square test, F-test and their Applications with Confidence Intervals, Analysis of Variance for One-Way, Two-Way With One Observations Per Cell. Non-Parametric Test: Sign Test, Mann-Whitney U Test, Wilcoxon Signed Rank Test, Krushal Wallis H Test and Spearman Correlation Tests.

Suggested Readings:

1. Gupta, S.C. and Kapoor, V. K. (2020): Fundamentals of Mathematical Statistics, Sultan Chand & Sons, New Delhi.
2. Mukhopadhyay, P. (2020): Mathematical Statistics, Books and Allied Private Limited, Kolkata.
3. Kapoor, J.N. and Saxena, H.C. (2020): Mathematical Statistics, Sultan Chand & Sons, New Delhi.
4. Ross, S.M. (2017): Introductory Statistics, Academic Press, Elsevier.
5. Gun, A.M., Gupta, M.K. and Dasgupta, B. (2016): Fundamental of Statistics, Vol. I, The World Press Private Limited, Kolkata.
6. Gun, A.M., Gupta, M.K. and Dasgupta, B. (2016): Fundamental of Statistics, Vol. II, The World Press Private Limited, Kolkata.

Name of the Department: STATISTICS
Name of the Major Course: Sampling Techniques
Semester: I

Course Code	**23STAD101DS02**	**Course Credits**	4(L: 3 T: 0 P: 1)
Max. Marks	75 {External (term-end exam) – 50} (Internal – 25)	**Time of end term examination**	3 Hours

Note: The examiner is to set nine (09) questions in all into five sections A, B, C, D and E of the question paper from all the four units- I, II, III and IV. In section A, there will be a compulsory question of 10 marks consisting of 08 short answer type questions two from each unit (each of marks 02) and out of which the students will be required to attempt any 05 questions. In each section B, C, D and E there will be two questions of 10 marks each from Units I, II, III, & IV respectively. The students will be required to attempt five questions in all selecting at least one question from each section.

Course Objectives:

Students will able:

- To Understand the Techniques for Conducting Sample Surveys
- To Identify or Control the Sampling and Non-Sampling Errors
- To Understand Various Methods for Generating of Random Samples
- To Understand the Methods of Sampling
- To Understand the Use of Auxiliary Information at the Estimation Stage

Course Outcomes:

Students will:

- Able to Use the Techniques for Conducting Sample Surveys
- Achieved Knowledge about the Sampling Scheme including Simple Random, Stratified, Systematic, Double and Cluster Samplings
- Able to Identify or Control the Sampling and Non-Sampling Errors
- Gain Knowledge to Use of Auxiliary Information for Estimation of the Parameters
- Ability to Generate the Random Sample using Different Methods

Unit - I

Concepts of Census and Sample Surveys: Basic Concepts in Sampling, Sampling and Non-Sampling Errors, Principal Steps Involved in a Sample Survey, Bias, Precision, Accuracy and Mean Squared Error, Limitation of Sampling, Basic Principle of Sampling Survey, Types of Sampling, Selection of a Simple Random Sample: Lottery and Random Number Methods.

Unit – II

Simple Random Sampling With and Without Replacement, Estimation in Simple Random Sampling, Merits and Demerits of SRS, Estimation of Population Proportion for Attributes, Determination of Sample Size in SRS, Ratio and Regression Estimators: Use of Auxiliary Information, Ratio Estimator, Bias of Ratio Estimator, Unbiased Ratio Type Estimator, Regression Estimator, Bias in the Linear Regression Estimator.

Unit – III

Double Sampling (Two-Phase Sampling), Stratified Random Sampling: Principal Advantages of Stratified Random Sampling, Allocation of Sample Size, Optimum Allocation, Cost Function, Relative Precision Between Stratified Random and Simple Random Sampling..

Unit – IV

Systematic Sampling: Linear and Circular, Advantages of Systematic Sampling over Simple Random Sampling Cluster Sampling: Single Stage, Two Stage, Principal Advantages of Two Stage Sampling, Estimation of Population Mean and Variance, Comparison of Two Stage Sampling with Single Stage Sampling.

Suggested Readings:

1. Goon, A.M., Gupta, M.K., & Gupta, B.D. (2016). Fundamentals of Statistics, Vol.-II, World Press.
2. Singh, D., & Chaudhary, F.S. (2018). Theory & Analysis of Sample Survey Designs. New Age International Private Limited.
3. Gupta, S.C., & Kapoor, V.K. (2014). Fundamentals of Applied Statistics, Sultan Chand & Sons.
4. Raj, D., & Chandhok, P. (2013). Sample Survey Theory. Createspace Independent Publication.
5. Hansen, M.H., Hurwitz, W.N., & Madow, W.G. (1993). Sample Survey Methods and Theory. Wiley.

PRACTICAL PAPER

Course Code	**23STAD101DS02P**	**Course Credits**	1 (L: 0 T: 0 P: 1)
Max. Marks	25 {External (term-end exam) – 20} (Internal – 05)	**Time of end term examination**	3 Hours

Note: The practical question paper will consist of five questions and the students will be required to attempt any three questions. The question paper will be set on the spot jointly by the internal and external examiners.

Distribution of Marks will be as follows:

Marks for Question Paper:	12
Marks for Practical Record Book:	05
Marks for Viva-Voce:	03
Total:	20

List of Practicals:

1. To Select a SRS With and Without Replacement.
2. For a Population of Size 5, Estimate Population Mean, Population Mean Square and Population Variance. Enumerate All Possible Samples of Size 2 by WR & WOR
3. For SRSWOR, Estimate Mean, Standard Error and the Sample Size
4. In SRSWR, Show that the Sample Mean Variance are an Unbiased Estimator of Population Mean and Variance Respectively.
5. Determination of Sample Size in SRS.
6. Stratified Random Sampling with Proportional and Optimum Allocation
7. Systematic Sampling with N= nk. Comparison of Systematic Sampling with Stratified
8. Estimate the Gain in Precision Due to Stratification
9. Estimate the Ratio of Two Population Characteristics
10. Estimation of Population Parameters for the given data using Ratio and Regression Estimators. Compare the Efficiencies of Ratio and Regression Estimators Relative to SRS.
11. Estimation of Mean or Total, Variance of the Estimate, Estimate of Intra-Class Correlation Coefficient for Cluster Sampling.

Name of the Department: STATISTICS
Name of the Major Course: Regression Models
Semester: I

Course Code	**23STAD101DS03**	**Course Credits**	4(L: 4 T: 0 P: 0)
Max. Marks	100 {External (term-end exam) – 75} (Internal – 25)	**Time of end term examination**	3 Hours

Note: The examiner is to set nine (09) questions in all into five sections A, B, C, D and E of the question paper from all the four units- I, II, III and IV. In section A, there will be a compulsory question of 14 marks consisting of 08 short answer type questions two from each unit (each of marks 02) and out of which the students will be required to attempt any 07 questions. In each section B, C, D and E there will be two questions of 14 marks each from Units I, II, III, & IV respectively. The students will be required to attempt five questions in all selecting at least one question from each section.

Course Objectives:

Students will able:

- To Know the Concept of Linear Regression Model
- To Understand OLS Estimators and Their Properties
- To Know the Use of Logistic and Poisson Regression Models
- To Handle the Problems of Autocorrelation, Non-Normality and Heteroscedasticity
- To Understand the Consequences of Multicollinearity and Possible Solutions

Course Outcomes:

Students acquainted with the:

- Theoretical Background for the Standard Methods and Properties of OLS
- Knowledge About Regression Analysis for Analyzing the Data
- Concepts of the Elementary Procedures for Generalized & Weighted Least Square Estimators
- Knowledge of the Concepts of Non-Normality & Heteroscedasticity and Possible Solutions
- Ideas to Deal with Multicollinearity, Autocorrelation and How to Resolve Them

Unit - I

Curve Fitting Using Method of Least Square, Simple Linear Regression Model: Least Square Estimators of Coefficients and Their Properties, Inference in Simple Linear Regression Model, General Linear Regression Model, Ordinary Least Square Estimators and Its Properties, Inference in General Linear Regression Model.

Unit – II

Generalized and Weighted Least Squares, Variable Selection and Model Building, Computational Techniques for Model Selection, Stepwise Regression, Forward Selection, Backward Elimination, Logistic and Poisson Regression, Diagnostics for Leverage and Influence, Mallow's C_p.

Unit – III

Heteroscedasticity, Tests for Heteroscedasticity, Multicollinearity, Types of Multicollinearity, Consequences and Detection of Multicollinearity, Farrar-Glauber Test, Remedies for Multicollinearity, Autocorrelation, Sources and Consequences, AR, MA and ARMA Processes for Autocorrelation, Durbin-Watson Test.

Unit – IV

Model Adequacy Checking, Residual Analysis, Methods of Scaling Residuals: Standardized and Studentized Residuals, Press Residual, Residual Plots, PRESS Statistic, Variance Stabilizing Transformations, Analytical Methods for Selecting a Transformation on Study

Variable.

Suggested Readings:

1. Johnston, J. (1984): Econometric Methods. McGraw-Hill, New York.
2. Gujarati, D. N. (2004): Basic Econometrics. Tata McGraw Hill.
3. Koutsyannis, A. (2004): Theory of Econometrics. Macmillan Publishers Limited
4. Maddala, G.S., & Lahiri, K. (2012): Introduction to Econometrics. Wiley.
5. Madnani, GMK. (2015). Introduction to Econometrics: Principles and Applications. Oxford & IBH Publishing Co Pvt. Ltd.
6. Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., & Lee, T-C. (1985): Introduction to the Theory and Practice of Econometrics. Wiley.

Name of the Department: STATISTICS
Name of the Major Course: Data Analysis Using SPSS
(Based on 23STAD101DS01&23STAD101DS03)
Semester: I

Course Code	**23STAD101DS04**	**Course Credits**	4(L: 0 T: 0 P: 4)
Max. Marks	100 {External (term-end exam)-70} (Internal – 30)	**Time of end term examination**	3 Hours

Note: The practical question paper will consist of seven questions and the students will be required to attempt any five questions. The question paper will be set on the spot jointly by the internal and external examiners.

Distribution of Marks will be as follows:

Marks for Question Paper:	45
Marks for Practical Record Book:	15
Marks for Viva-Voce:	10
Total:	70

Course Objectives:

Students will Able:

- To Plot Different Types of Graphs to Visualize the Data
- To Compute the Measures of Central Tendency & Dispersion of a Dataset
- To Fit the Discrete and Continuous Distributions to the Dataset
- To Test the Hypothesis Based on Parametric and Non-Parametric Tests
- To Perform Different Types of Regression Techniques such as Simple, Multiple, Logistic and Poisson

Course Outcomes:

Students Acquired the:

- Knowledge to Understand Presentation and Interpretation of the Data in SPSS
- Understanding of the Measures of Central Tendency and Dispersion, Correlation and Regression Analysis in SPSS
- Knowledge to Understand Fitting of Probability Distributions in SPSS
- Ability to apply Parametric and Non-Parametric Tests
- Skill to Perform Different Types of Regression Techniques in SPSS

List of Practicals:

1. Presentation of the Data through Different Graphs
2. Compute the Measures of Central Tendency and Dispersion for a Dataset
3. Determine the Correlation Coefficient.
4. Determine the Spearman's Rank Correlation.
5. Check the Independence and Association of Attributes.
6. Check the Consistency of Categorical Data.
7. Perform Small Sample Tests
8. Perform Large Sample Tests
9. Find Confidence Interval to Estimate the Parameters.
10. Perform One-Way and Two-Way ANOVA
11. Find the Best Fitted Line using the Method of Curve Fitting.
12. Fit the Simple Linear Regression and Assess the Significance of Obtained Model.
13. Fit the Multiple Linear Regression and Assess the Significance of Obtained Model.

14. Find the Best Fitted Regression Model using Backward Elimination & Forward Selection Criteria.
15. Fit Logistic Regression Model and Assess the Significance of Obtained Parameters.
16. Fit Poisson Regression Model and Assess the Significance of Obtained Parameters.
17. Test the Heteroscedasticity in the Data.
18. Test the Multicollinearity in the Data.
19. Fit Auto-Regressive and Moving Average Models of Autocorrelation.
20. Test Autocorrelation using Durbin Watson Test.
21. Test Model Adequacy Based on Regression Models.
22. Residual Analysis using QQ Plots and Histograms.
23. Evaluate the PRESS Statistic for Model Validation and Implementing Variance Stabilizing Transformations for Improved Model Performance.
24. Exploring Analytical Methods to Select Appropriate Transformations and Comparing the Effectiveness of Different Transformation Techniques.

Suggested Readings:

1. Mukhopadhyay, P. (2020): Mathematical Statistics, Books and Allied Private Limited, Kolkata.
2. Kapoor, J.N. and Saxena, H.C. (2020): Mathematical Statistics, Sultan Chand & Sons, New Delhi.
3. Cunningham, B.J. (2012): Using SPSS: An Interactive Hands-on approach.
4. Field, A. (2013): Discovering Statistics Using SPSS, Fourth Edition, SAGE.

Name of the Department: STATISTICS
Name of the Value Added Course: Data Handling & Visualization
Semester: I

Course Code	**23STAD101DS05**	**Course Credits**	4(L: 0 T: 0 P: 4)
Max. Marks	100 {External (term-end exam) – 70} (Internal – 30)	**Time of end term examination**	3 Hours

Note: The practical question paper will consist of seven questions and the students will be required to attempt any five questions. The question paper will be set on the spot jointly by the internal and external examiners.

Distribution of Marks will be as follows:

Marks for Question Paper:	45
Marks for Practical Record Book:	15
Marks for Viva-Voice:	10
Total:	70

Course Objectives:

Students will able:

- To Gain the Basic Knowledge of Excel and Tableau
- To Deal with Data in Excel and Tableau
- To Carry out Data Visualization and Explanatory Analysis
- To Use the Data Analysis ToolPak to Perform Statistical Analysis
- To Create Interactive Dashboard using Multiple Visualizations in Tableau

Course Outcomes:

Students:

- Understand Different Types of Data and Scales of Their Measurement
- Able to Perform a Wide Range of Data Management Tasks
- Acquired Knowledge to Handle Data Visualization and Descriptive Statistics
- Get Hands on Data Analysis ToolPak to Perform Statistical Analysis
- Acquainted the Technique to Handle Data Visualization using Tableau
- Able To Perform Interactive Dashboards using Tableau

List of Practicals:

1. A Program on the Import and Export Data from Excel file.
2. Sort the Data in Ascending or Descending Order and Apply Filters to Display Specific Data Subsets. Use Text Functions like CONCATENATE, LEFT, RIGHT, MID, and FIND to Manipulate Text Data in Excel Sheet.
3. Set up Data Validation Rules for a Range of Cells to Restrict Input Based on Criteria like Dates, Numbers, or Specific Values.
4. Utilize Complex Nested Formulas involving Various Functions like INDEX-MATCH, SUMIFS, COUNTIFS, and Nested IF Functions.
5. Determine the Mean, Median, Mode, Range, Variance, and Standard Deviation for a Dataset through Excel Formulae.
6. Use Pivot Tables to Generate insights through Filtering, Sorting, and Grouping Data.
7. To Split the Text into Different Columns with the Convert Text to Columns Wizard and Specify the Delimiters and Adjust Its Column Formatting.
8. Apply Conditional Formatting to Highlight Specific Cells based on Rules or Conditions and

Remove the Duplicate Entries in the Sheet.

9. Determine the Rank and Percentile of Data Points in a Dataset.
10. Create Various Types of Charts (Bar, Line, Pie, etc.) to Represent Data Trends.
11. Use the Random Number Generation Tool to Generate Random Data for Simulations or Testing.
12. Use Data Analysis ToolPak to Generate Random Numbers Following Specified Distributions like Uniform, Normal, Binomial, etc.
13. Perform Linear Regression Analysis using Data Analysis ToolPak to Determine the Relationship between Variables and Generate a Regression Summary Report. Interpret the Regression Output, including Coefficients and Significance Levels.
14. Perform Paired t-test for Comparing Means of Two Samples.
15. Perform t-tests for Comparing Means of Two Samples by Assuming Equal and Unequal Variances.
16. Perform z-test for Comparing Means of Two Samples.
17. Utilize Data Analysis ToolPak for Chi-Square Tests to Analyze Categorical Data for Associations or Independence.
18. Perform F-tests to Compare Variances or Test the Significance of the Relationship between Two Variables.
19. Conduct ANOVA Tests to Compare Means between More Than Two Groups in a Dataset with the Help of Data Analysis ToolPak.
20. Import Various Datasets into Tableau, Clean and Prepare the Data for Visualization.
21. Create Bar Charts, Line Graphs, Scatter Plots, and Pie Charts using Different Data Sets and Customize Colors, Labels, and Formatting.
22. Compare Different Categories within a Dataset using Visualizations like Stacked Bar Charts, Side-by-Side Bar Charts, or Box Plots to Understand Relationships and Variances.
23. Analyse a Dataset to Identify Trends Over Time. Use Line Graphs or Area Charts to Visualize How Specific Metrics Change Over a Period.
24. Create Calculated Fields and Understand Their Applications in Visualizations.
25. Use IF Statements and Mathematical Functions.
26. Use Parameters to Create Dynamic, Interactive Visualizations.
27. Build an Interactive Dashboard from Scratch using Multiple Visualizations. Allow Users to Filter, Highlight, and Interact with the Data to Draw their Insights.
28. Use Geographic Data to Create Maps that Showcase Regional Variations or Patterns. Explore Demographics, Sales Data, or Any Other Geospatial Dataset to Create Meaningful Visualizations.
29. Combine Data from Multiple Sources to Create a Blended Worksheet.
30. Create a Narrative using Data and Tableau Visualizations to Convey a Compelling Story and Use Annotations, Text, and Images to Enhance Storytelling

Suggested Readings:

1. Nigam, M. (2019): Data Analysis with Excel, BPB Publisher, India.
2. Ken, B. (2020): Microsoft Excel Formulas & Functions for Dummies, Wiley.
3. Arora, R. (2023): Mastering Advanced Excel, BPB Publications, India.
4. Guillevin, G. (2019): Getting Started with Tableau, Packt Publishing Limited, India.
5. Milligan, J.N. (2022): Learning Tableau, Ingram Short Title, India.
6. Sleeper, R. (2018): Practical Tableau, Shroff/O'Reilly, India.

Name of the Department: STATISTICS
Name of the Skill Enhancement Course: DBMS & SQL
Semester: I

Course Code	**23STAD101SE01**	**Course Credits**	4(L: 3 T: 0 P: 1)
Max. Marks	75 {External (term-end exam) – 50} (Internal – 25)	**Time of end term examination**	3 Hours

Note: The examiner is to set nine (09) questions in all into five sections A, B, C, D and E of the question paper from all the four units- I, II, III and IV. In section A, there will be a compulsory question of 10 marks consisting of 08 short answer type questions two from each unit (each of marks 02) and out of which the students will be required to attempt any 05 questions. In each section B, C, D and E there will be two questions of 10 marks each from Units I, II, III, & IV respectively. The students will be required to attempt five questions in all selecting at least one question from each section.

Course Objectives:

Students Will Able:

- To Understand the Concepts of Database Management System
- To Organize, Maintain and Retrieve Information from DBMS
- To Study of Information Concepts and the Realization of those Concepts using the Relational Data Model
- To Gain Practical Experience of Designing and Constructing Data Models and using SQL to Interface to User DBMS Packages
- To Enhance the Knowledge of Advanced Concepts such as Logical Data, Parallel & Distributed Database and Warehousing and Data Mining

Course Outcomes:

Students will :

- Able to Differentiate Database Systems from File Systems by Enumerating the Features Provided by Database Systems
- Acquired Skill to Model an Application's Data Requirements using Conceptual Modeling Tools like ER Diagrams and Design Database Schemas Based on the Conceptual Model
- Able to Write SQL Commands to Create Tables and Indexes, Insert/Update/Delete Data, and Query Data in a Relational DBMS
- Able to Understand the Normalization Theory and Apply such Knowledge to the Normalization of a Database
- Use an SQL Interface of a Multi-User Relational DBMS Package to Create, Secure, Populate, Maintain and Query a Database

Unit - I

Basic Concepts of DBMS, Data Abstraction, Data Independence, Data Aggregation, Three Tier Architecture, E-R Diagram (Entity Relationship), Mapping Constraints, Keys, Reduction of E-R Diagram into Tables.

Unit – II

Relational Data Base: Relational Algebra, Tuple and Domain Relational Calculus, Armstrong's Axioms, Normal Forms, Functional Dependency, Lossless Design SQL3, DDL and DML Constructs, Open Source and Commercial DBMS - MYSQL, ORACLE, DB2, SQL Server.

Unit – III

File Organization: Sequential Files, Index Sequential Files, Direct Files, Hashing, B-trees, B+

trees, Index Files.
Query Processing and Optimization: Evaluation of Relational Algebra Expressions, Query Equivalence, Join Strategies, Query Optimization Algorithms.

Unit – IV

Transaction Processing: Concurrency Control, ACID Property, Serializability of Scheduling, Locking and Timestamp-Based Schedulers, Multi-Version and Optimistic Concurrency Control Schemes, Database Recovery
Advanced Topics: Parallel and Distributed Data Base, Object Oriented and Object Relational Databases, Logical Databases, Web Databases, Data Warehousing and Data Mining

Suggested Readings:

1. Silberschatz, A., Korth, H.F., & Sudarshan, S. (2019): Database System Concepts, McGraw-Hill Education, New York.
2. Elmasri, R. & Navathe, S.B. (2017): Fundamentals of Database Systems, Pearson Education, India.
3. Date, C. & Darwen, H. (1996): A Guide to SQL Standard, Addison Wesley, Boston.
4. Majumdar, A.K. & Bhattacharya, P. (2017): Database Management Systems, McGraw-Hill Education, New York.
5. Hansen, G.W. & Hansen, J.V. (1995): Database Management and Design, Pearson Education, India.

PRACTICAL PAPER

Course Code	**23STAD101SE01P**	**Course Credits**	1 (L: 0 T: 0 P: 1)
Max. Marks	25 {External (term-end exam)-20} (Internal – 5)	**Time of end term examination**	3 Hours

Note: The practical question paper will consist of five questions and the students will be required to attempt any three questions. The question paper will be set on the spot jointly by the internal and external examiners.
Distribution of Marks will be as follows:

Marks for Question Paper:	12
Marks for Practical Record Book:	05
Marks for Viva-Voce:	03
Total:	20

List of Practicals:

1. To Create a Database and Writing SQL Queries to Retrieve Information from the Database.
2. To Retrieve all Records and Specific Columns from a Specific Table.
3. To Insert, View, Delete, Alter, Modify and Update Records based on Conditions.
4. To do the INNER JOIN, LEFT JOIN, RIGHT JOIN to Combine Data from Multiple Tables.
5. To Create Views, Synonyms, Sequence, Indexes and Save the file.
6. To Create an Employee Database to Set Various Constraints.
7. To Create the Relationship Between the Database.
8. Use the WHERE Clause to Filter Records Based on Specific Conditions and Sort the Results in Ascending and Descending Order.
9. To Determine the Total, Average, Minimum, and Maximum Values for a Column.
10. To Write a Subquery to Find Information Based on the Result of Another Query.

Name of the Department: STATISTICS
Name of the Major Course: Multivariate Analysis
Semester: II

Course Code	**23STAD102DS01**	**Course Credits**	4 (L: 4, T: 0, P:0)
Max. Marks	100 {External (term-end exam)-70} (Internal – 30)	**Time of end term examination**	3 Hours

Note: The examiner is to set nine (09) questions in all into five sections A, B, C, D and E of the question paper from all the four units- I, II, III and IV. In section A, there will be a compulsory question of 14 marks consisting of 08 short answer type questions two from each unit(each of marks 02) and out of which the students will be required to attempt any 07 questions. In each section B, C, D and E there will be two questions of 14 marks each from Units I, II, III, & IV respectively. The students will be required to attempt five questions in all selecting at least one question from each section.

Course Objectives:

Students will Able to:

- Understand the Basic Concepts Required to Build the Foundations of Multivariate Data
- Understand the Extension of Univariate Techniques to Multivariate Framework
- Perform Exploratory Analysis on Multivariate Dataset
- Hypothesis Testing on Multivariate Data
- Perform the Dimension Reduction Techniques such as PCA, Factor, Cluster and Discriminant Analysis

Course Outcomes:

Students Acquired the:

- Knowledge to Deal with Multivariate Datasets
- Skill to Analyze the Multivariate Data with Mean Vector
- Ability to Test the Hypothesis for Means, Correlation and Regression Coefficients
- Ability to Find Major Factors and the Variability Using Multivariate Techniques including Principal Component Analysis, Factor Analysis, Discriminant and Cluster Analysis
- Knowledge to Judge the Situations Where Multivariate Analysis Techniques are Suitable in Different Environment

Unit - I

Matrix: Definition, Types of Matrices and their Basic Properties, Multivariate Normal Distribution, Distribution of Linear Combinations of Normal Vector, Maximum Likelihood Estimators of Mean Vector and Covariance Matrix, Distribution of Sample Mean Vector, Distribution of Quadratic Form.

Unit – II

Wishart Matrix: Its Distribution and Properties, Distribution of Sample Generalized Variance, Hotelling's T^{2-}Statistic: Definition, Distribution and Uses, Beheran - Fisher's Problem, Wilk's Lambda Criterion, Likelihood Ratio Test Criteria for Testing Independence of Sets of Variables, Likelihood Ratio Criteria for Testing Equality of Covariance Matrices.

Unit – III

Multivariate Linear Regression Models: Estimation of Parameters and their Properties. Distribution of the Matrix of Sample Regression Coefficients and Tests of Linear Hypothesis

about Regression Coefficients, One-way Multivariate Analysis of Variance (MANOVA)

Unit – IV

Fisher's Discriminant Function, Mahalanobis' Distance, Principal Component Analysis (PCA) and Its Uses & Importance, Cluster and Factor Analysis, Canonical Variables and Canonical Correlations.

Suggested Readings:

1. Anderson, T.W. (2009): An Introduction to Multivariate Statistical Analysis, Wiley India Pvt Ltd, India.
2. Johnson, R. A. & Wichern, D. W. (2012): Applied Multivariate Statistical Analysis, Prentice Hall India Learning Private Limited, India.
3. Johnston, J. (1996): Econometric Methods, McGraw-Hill Education.
4. Damodar, N. G. (2017): Basic Econometrics, McGraw-Hill Education.
5. Montgomery, D.C., Peck, E.A., & Vining, G.G. (2006): Introduction to Linear Regression Analysis, Wiley India Pvt. Ltd., India.
6. Draper, N.R. & Smith, H. (1998): Applied Regression Analysis, Wiley India Pvt. Ltd., India.

Name of the Department: STATISTICS
Name of the Major Course: Time Series Analysis & SQC
Semester: II

Course Code	**23STAD102DS02**	**Course Credits**	4(L: 4 T: 0 P: 0)
Max. Marks	100 {External (term-end exam) – 70} (Internal – 30)	**Time of end term examination**	3 Hours

Note: The examiner is to set nine (09) questions in all into five sections A, B, C, D and E of the question paper from all the four units- I, II, III and IV. In section A, there will be a compulsory question of 10 marks consisting of 08 short answer type questions two from each unit(each of marks 02) and out of which the students will be required to attempt any 05 questions. In each section B, C, D and E there will be two questions of 10 marks each from Units I, II, III, & IV respectively. The students will be required to attempt five questions in all selecting at least one question from each section.

Course Objectives:

Students will Able:

- To Know About the Consistency in Time Series Data by Eliminating the Effect of Time Series Component (Trend, Seasonal, Cyclic and Random Variation)
- To Acquire Knowledge About Stationary, Strong and Weak Stationary Time Series and Different Schemes and Models
- To Get Knowledge About Control Charts in Manufacturing Processes
- To Impart Skills for Preparing Control Charts to Control the Quality of Process and Manufactured Products
- To Understand the Significance of Using 3-σ Control Limits in Quality Control to Detect Variations

Course Outcomes:

Students Acquired the:

- Ability to Understand How the Forecasting Can be Used in Economic Analysis
- Knowledge to Drive the Concepts of Auto-Covariance and Autocorrelation for Estimation of Parameters in ARIMA Models
- Ability to Know About Box-Jenkins Models for Estimation of Parameters in Time Series Analysis
- Comprehensive Understanding of the Fundamental Principles and Concepts of Statistical Quality Control
- Skill to Identify the Various Purposes and Applications of SQC in Manufacturing and Service Industries

Unit - I

Introduction to Time Series Data and Its Applications, Components of a Time Series, Decomposition of Time Series. Trend & Its Types of Estimation, Measurement of Seasonal Fluctuations, Measurement of Cyclical Component: Periodogram Analysis, Variate Difference Method and its Uses.

Unit – II

Concept of Stationary Time Series: Strong and Weak Stationary, Auto Covariance & Auto Correlation and their Properties. Correlogram of Auto Regressive Scheme. Box-Jenkins

Models, Estimation of Parameters in ARIMA Models, Estimation of the Parameters of AR(1) and AR(2) – Yule-Walker Equations. Forecasting: Exponential and Adaptive Smoothing Models.

Unit – III

Statistical Quality Control and Its Applications, 3-Sigma Control Limits, Shewhart's Control Chart. Control Charts for Variables and Attributes, Analysis of Patterns on Control Chart, Estimation of Process Capability, Acceptance Sampling Inspection Plans by Variables (Known and Unknown Sigma Case), Natural Tolerance Limits and Specification Limits: Modified Control Limits.

Unit – IV

Sampling Inspection Plan, Producer's and Consumer's Risk OC and ASN Function, AQL. LTPD and ATI. Sampling Inspection Plans for Attributes: Single, Double and Sequential Sampling Plans and Their Curves including AOQ, OC, ASN and ATI Curves. Testing for Sequential Sampling.

Suggested Readings:

1. Montgomery, D. C. (2009): Introduction to Statistical Quality Control, 6th Edition, Wiley India Pvt. Ltd.
2. Goon, A.M., Gupta, M.K., and Dasgupta, B. (2002): Fundamentals of Statistics, Vol. I & II, 8th Edition. The World Press, Kolkata.
3. Mukhopadhyay, P. (2011): Applied Statistics, 2nd Edition Revised Reprint, Books and Allied (P) Ltd.
4. Montgomery, D. C. and Runger, G.C. (2008): Applied Statistics and Probability for Engineers, 3rd Edition Reprint, Wiley India Pvt. Ltd.
5. Gupta, S.C. and Kapoor, V.K. (2007): Fundamentals of Applied Statistics. 4th Edition, Sultan Chand and Sons, New Delhi.

Name of the Department: STATISTICS
Name of the Major Course: Machine Learning Using Python
Semester: II

Course Code	**23STAD102DS03**	**Course Credits**	4(L: 3 T: 0 P: 1)
Max. Marks	75 {External (term-end exam) – 50} (Internal – 25)	**Time of end term examination**	3 Hours

Note: The examiner is to set nine (09) questions in all into five sections A, B, C, D and E of the question paper from all the four units- I, II, III and IV. In section A, there will be a compulsory question of 10 marks consisting of 08 short answer type questions two from each unit(each of marks 02) and out of which the students will be required to attempt any 05 questions. In each section B, C, D and E there will be two questions of 10 marks each from Units I, II, III, & IV respectively. The students will be required to attempt five questions in all selecting at least one question from each section.

Course Objectives:

Students will Able:

- To Understand the Basics of Python Programming Language
- To Explore Real-World Applications of Machine Learning
- To Differentiate Between Supervised, Unsupervised and Semi-Supervised Learning
- To Understand the Benefits of Decision Trees
- To Understand Clustering Techniques: K-means Clustering, Hierarchical Clustering

Course Outcomes:

Students Acquired the:

- Knowledge About Implementing the Conditional Statements to Control the Flow of a Python Program
- Ability to Perform EDA Using Data Frames, Handle Missing Values and Visualize Data
- Skill to Identify, Handle and Troubleshoot Errors in Python Programs
- Knowledge About Binary Logistic Regression, Decision Tree Classifier and KNN Classification.
- Understanding About Support Vector Machine for Classification Tasks

Unit - I

Python Programming: Introduction to Python, Declaring Variables, Conditional Statements, Generating Sequence Numbers, Loops, Functions, List, Tuples, Set, Dictionary, Dealing with Strings, Map, Filter and Reduce, Modules and Packages. Comments in the Program, Errors and Exceptions, Handling Exceptions, Modules

Unit – II

Introduction to Machine Learning, Real-World Applications of Machine Learning, Types of Machine Learning: Supervised, Unsupervised and Semi-Supervised, Python Libraries for Machine Learning: PANDAS, NUMPY, SCIKIT-LEARN, MATPLOTLIB.

Exploratory Data Analysis: Working with Data Frames, Handling Missing Values, Data Exploration Using Visualization.

Hypothesis Testing: z-Test, One Sample t-Test, Two Sample t-Test, Paired Sample t-Test, Chi-Square Goodness of Fit and One-Way Analysis of Variance.

Unit – III

Supervised Learning: Simple and Multiple Linear Regression, Steps in Building a Regression

Model, Model Building, Model Diagnostics, Classification Overviews, Binary Logistic Regression, Model Building: Model Diagnostics, Creating Confusion Matrix, Gain Chart and Lift Chart, Classification Tree, Building Decision Tree Classifier using Gini Criteria. Gini Impurity, Benefits of Decision Tree, KNN Classification, Support Vector Machine.

Unit – IV

Unsupervised Learning: Clustering, Techniques: K-means Clustering and Hierarchical Clustering

Forecasting: Moving Average, Exponential Smoothing, AR Models, Moving Average Processes, ARMA Model, ARIMA Model and Their Diagnostics.

Suggested Readings:

1. Bishop, C.M. (2016): Pattern Recognition and Machine Learning, Springer.
2. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2017): Introduction to Statistical Machine Learning with Applications in R, Springer.
3. Tom, M. (2017): Machine Learning, McGraw Hill Education, New York.
4. Kulkarni, P. (2012): Reinforcement and Systemic Machine learning for Decision Making, Wiley-IEEE Press.
5. Pradhan M. & Kumar, U.D. (2019): Machine Learning using Python, Wiley.

PRACTICAL PAPER

Course Code	**23STAD102DS03P**	**Course Credits**	1(L: 0 T: 0 P: 1)
Max. Marks	25 {External (term-end exam)-20} (Internal – 5)	**Time of end term examination**	3 Hours

Note: The practical question paper will consist of five questions and the students will be required to attempt any three questions. The question paper will be set on the spot jointly by the internal and external examiners.

Distribution of Marks will be as follows:

Marks for Question Paper:	12
Marks for Practical Record Book:	05
Marks for Viva-Voce:	03
Total:	20

List of Practicals:

1. Write a Program to Declaring Variables of Different Types (Int, Float, String).
2. Write a Program Using If, Else and Elif Statements.
3. Create a Program to Generate a Sequence of Random Numbers.
4. Write a Program to Implement For, While and Nested Loops in Python.
5. Write a Program to Define and Call Functions with Different Parameters.
6. Write a Program to Perform Operations on These Data Structures, such as Appending, Slicing and Updating.
7. Write a Program Manipulate Strings, Concatenate them and use String Methods.
8. Write a Program to Apply Map, Filter, and Reduce Functions on Lists.
9. Write a Program to Add Comments for Code, Intentionally Introduce Errors and Handle Exceptions.
10. Explore Python Libraries for Machine Learning.
11. Perform Exploratory Data Analysis (EDA) for a Given Dataset
12. Write a Program to Implement Z-test, One Sample t-test, Two Sample t-test, Paired Sample t-test, Chi-Square Goodness of Fit and One-Way ANOVA using Python.
13. Write a Program for Simple and Multiple Linear Regression Models using a Dataset.
14. Write a Program to Build a Binary Logistic Regression Model using a Dataset.
15. Write a program to Build a Decision Tree Classifier using Gini Criteria on a Dataset.

16. Write a Program to Implement K-Nearest Neighbor and SVM Classifiers on Dataset.
17. Write a Program to Perform K-Means and Hierarchical Clustering on a Dataset.
18. Write a Program to Implement Moving Average, Exponential Smoothing, AR Models and ARIMA Models on a Dataset.

Name of the Department: STATISTICS
Name of the Major Course: Data Analysis Using R
(Based on 23STAD102DS01 & 23STAD102DS02)
Semester: II

Course Code	**23STAD102DS04**	**Course Credits**	4(L: 0 T: 0 P: 4)
Max. Marks	100 {External (term-end exam) – 70} (Internal – 30)	**Time of end term examination**	3 Hours

Note: The practical question paper will consist of seven questions and the students will be required to attempt any five questions. The question paper will be set on the spot jointly by the internal and external examiners.

Distribution of Marks will be as follows:

Marks for Question Paper:	45
Marks for Practical Record Book:	15
Marks for Viva-Voice:	10
Total:	70

Course Objectives:

Students will Able:

- To Understand the Fundamentals of R Programming
- To Detect the Trends in Time Series Data through Graphs
- To Check the Stationarity & Seasonality in the Time Series Data
- To Prepare Different Types of Control Charts
- To Perform Exploratory Analysis and Dimension Reduction Techniques on Multivariate Dataset

Course Outcomes:

Students Acquired the:

- Working knowledge of R Programming
- Ability to Perform Various Forecasting Techniques for the Time Series Data Using R
- Skill to Construct the Different Types of Control Charts Using R
- Knowledge to Deal with Multivariate Datasets in R
- Knowledge to Perform Exploratory Analysis and Dimension Reduction Techniques on Multivariate Dataset in R

List of Practicals:

1. Write a Program to Create Data Structure that can represent Time Series Data and, also Plot One or More Time Series Data.
2. Write a Program to Create a Multivariate Time Series Data.
3. Write a Program to Select One or More Elements from a Time Series Data, Merge Two or More Time Series Data and Fill the Data with the Missing Observations.
4. Write a Program to Shift a Time Series in Time, Either Forward or Backward and, also Compute the Difference between Successive Observations.
5. Write a Program to Compute the Moving Average of a Time Series Data and, also Apply the

Function to Group the Contents by a Calendar Period (like Week, Month or Year).

6. Write a Program to Test the Presence of Autocorrelation and Plot the Autocorrelation Function (ACF) of a Time Series Data.
7. Write a Program to Find a Lagged Correlation Between Two Time Series Data.
8. Write a Program to Remove the Trend from a Time Series Data.
9. Write a Program to Fit an ARIMA Model to the Time Series Data.
10. Write a Program to Forecast the Next Few Observations in the Series from an ARIMA model.
11. Write a Program to Find Mean Vector and Variance Covariance Matrix for a Given Set of Data
12. Write a Program to Find Maximum Likelihood Estimate for Given Mean Vector and Covariance Matrix
13. Write a Program to Perform the Hypothesis Testing for Equality of Mean Vectors.
14. Write a Program to Estimate the Matrix of Regression Coefficients and Variance Covariance Matrix for Given Set of Vectors.
15. Write a Program to Perform the Linear Hypothesis about Regression Coefficients for Given Level of Significance
16. Write a Program to Carry Out Multivariate Analysis of Variance and Construct One Way MANOVA Table.
17. Write a Program to Calculate Fisher's Discriminant Function for a Given Set of Vectors.
18. Write a Program to Find Out Principal Components for Given Variance Covariance Matrix.
19. Write a Program to Calculate Mahalanobis' Distance for Given Set of Data.
20. Write a Program to Extract Factors from a Multivariate Data Set and Their Interpretation.
21. Write a Program to Perform Cluster Analysis to Discover Patterns and Groupings Within a Multivariate Dataset.
22. Write a Program to Find Canonical Variables and Canonical Correlations to Explore the Relationships between Different Sets of Variables in a Multivariate Dataset.

Suggested Readings:

1. Sandip, R. (2018): Statistics with R Programming, McGraw Hill Education, New York.
2. Srinivasa, K.G. & Siddesh, G.M. (2017): Statistical Programming in R, Oxford University Press, India.
3. Hardley, W. (2014): Advanced R, Chapman and Hall/CRC, Florida.
4. Long, J.D. & Paul, T. (2019): R Cookbook, O' Rielly, Columbus.
5. Randall, E.S. (2015): Using R with Multivariate Statistics, SAGE Publications, New York.