

**Anekant Education Society's  
Tuljaram Chaturchand College of Arts, Science and Commerce,  
Baramati**

**Autonomous**

**Course Structure for M.Sc. Statistics (2022 Pattern)  
(With effect from Academic Year 2023-2024)**

**Name of the Programme : M.Sc. Statistics**

**Program Code : PSST**

**Class : M.Sc. Part – II**

**Semester : III**

<b>Paper Code</b>	<b>Title of Paper</b>	<b>No. of Credits</b>
PSST231	Asymptotic Inference	4
PSST232	Design and Analysis of Experiments	4
PSST233	Time Series Analysis	4
PSST234 (A)	Data Mining	4
	Or	
PSST234 (B)	Design and Analysis of Clinical Trials	
PSST235	Practical Paper -V	4
PSST-236	Practical Paper -VI	4

**Name of the Programme : M.Sc. Statistics**

**Program Code : PSST**

**Class : M.Sc. Part – II**

**Semester : IV**

<b>Paper Code</b>	<b>Title of Paper</b>	<b>No. of Credits</b>
PSST241	Stochastic Processes	4
PSST242	Statistical Process Control	4
PSST243	Survival Analysis	4
PSST244 (A)	Actuarial Statistics	4
	Or	
PSST244 (B)	Optimization Techniques	
PSST245	Practical Paper-VII	4
PSST246	Project	4

### Program Outcomes (POs) for M.Sc. Programme

<b>PO1</b>	<b>Disciplinary Knowledge:</b> Demonstrate comprehensive knowledge of the discipline that forms a part of a postgraduate programme. Execute strong theoretical and practical understanding generated from the specific programme in the area of work.
<b>PO2</b>	<b>Critical Thinking and Problem solving:</b> Exhibit the skill of critical thinking and understand scientific texts and place scientific statements and themes in contexts and also evaluate them in terms of generic conventions. Identify the problem by observing the situation closely, take actions and apply lateral thinking and analytical skills to design the solutions.
<b>PO3</b>	<b>Social competence:</b> Exhibit thoughts and ideas effectively in writing and orally; communicate with others using appropriate media, build effective interactive and presenting skills to meet global competencies. Elicit views of others, present complex information in a clear and concise way and help reach conclusions in group settings.
<b>PO4</b>	<b>Research-related skills and Scientific temper :</b> Infer scientific literature, build a sense of enquiry and able to formulate, test, analyse, interpret and establish hypothesis and research questions; and to identify and consult relevant sources to find answers. Plan and write a research paper/project while emphasizing on academics and research ethics, scientific conduct and creating awareness about intellectual property rights and issues of plagiarism.
<b>PO5</b>	<b>Trans-disciplinary knowledge:</b> Create new conceptual, theoretical and methodological understanding that integrates and transcends beyond discipline-specific approaches to address a common problem.
<b>PO6</b>	<b>Personal and professional competence:</b> Perform independently and also collaboratively as a part of a team to meet defined objectives and carry out work across interdisciplinary fields. Execute interpersonal relationships, self-motivation and adaptability skills and commit to professional ethics.
<b>PO7</b>	<b>Effective Citizenship and Ethics:</b> Demonstrate empathetic social concern and equity centred national development, and ability to act with an informed awareness of moral and ethical issues and commit to professional ethics and responsibility.
<b>PO8</b>	<b>Environment and Sustainability:</b> Understand the impact of the scientific solutions in societal and environmental contexts and demonstrate the knowledge of and need for sustainable development.
<b>PO9</b>	<b>Self-directed and Life-long learning:</b> Acquire the ability to engage in independent and life-long learning in the broadest context of socio-technological changes.

**SYLLABUS (CBCS) FOR M.Sc. Statistics  
(2022 Pattern)  
(With effect from Academic Year 2023-2024)**

<b>Name of the Programme</b>	: M.Sc. Statistics
<b>Program Code</b>	: PSST
<b>Class</b>	: M.Sc. Part – II
<b>Semester</b>	: III
<b>Course Name</b>	: Asymptotic Inference
<b>Course Code</b>	: PSST231
<b>No. of Lectures</b>	: 60
<b>No. of Credits</b>	: 4

**A) Course Objectives:**

1. The main objective of this course is to learn and understand asymptotic behavior of the estimators.
2. Students should be able to find and verify the consistent estimator and consistency and asymptotic normality (CAN) estimator.
3. To learn various methods of finding consistent estimator, CAN estimator, large sample test and asymptotic confidence interval.

**B) Course Outcomes:**

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

- CO1.** understand concept of Consistent estimator, CAN estimator.
- CO2.** obtain consistent estimator and their asymptotic distributions.
- CO3.** choose the ARE estimator among given various consistent estimators.
- CO4.** obtain asymptotic distributions of moment estimators, percentile estimators.
- CO5.** determine maximum likelihood estimator and its asymptotic distributions
- CO6.** derive Likelihood Ratio Test (LRT), large sample test Wald's test, and Score test,
- CO7.** compute asymptotic confidence interval.

**TOPICS/CONTENTS:**

**Unit – 1**

Consistency: real and vector parameters, invariance of consistency under continuous transformation, Consistent estimators by method of moments and method of percentiles,

choosing between consistent estimators, Minimum sample sizes required to attain given level of accuracy, consistency and asymptotic normality (CAN): real and vector parameters, invariance of CAN under differentiable transformations (delta method), generation of CAN estimators using central limit theorem, CAN property of estimators obtained by moments and percentiles, examples of consistent but not asymptotically normal estimators, asymptotic relative efficiency (ARE) of consistent estimator. **(18 L)**

## **Unit – 2**

Maximum likelihood estimation, MLE in exponential family, Cramèr family, Cramèr-Huzurbazar theorem, asymptotic properties of maximum likelihood estimators, Solution of likelihood equations, Method of scoring, Newton-Raphson and other iterative procedures, MLE in case of restricted parameter space, super-efficient estimators, extension to vector-valued parameters, inconsistent MLE, special cases such as exponential class of densities and multinomial distribution, Multinomial with cell probabilities depending on a parameter. **(18 L)**

## **Unit – 3**

The Likelihood Ratio Test (LRT), asymptotic distribution of log likelihood ratio, Bartlett Correction, Wald Test, Rao's score test, Likelihood Ratio Test for Multinomials, variance stabilizing transformation and large sample tests **(15 L)**

## **Unit – 4**

Asymptotic confidence intervals: construction and examples, applications to categorical data analysis. **(9 L)**

## **References:**

1. Kale B. K. and Muralidharan K. (2015) Parametric Inference: An Introduction, Alpha Science International Ltd.
2. Gupta Anirban Das (2008), Asymptotic Theory of Statistics and Probability, Springer, New York.
3. Dudewicz E. J. and Mishra S. N. (1988) Modern Mathematical Statistics, John Wiley and Sons.
4. Casella G. and Berger R. L. (2001). Statistical Inference, 2nd edition, Duxbury press.
5. Lehmann, E.L. (1986). Testing Statistical Hypotheses (Student Edition).

6. Rohatgi V.K. and Ehsanes Saleh A. K. MD. (2003). An Introduction to Probability and Statistics, (Wiley Eastern, 2<sup>nd</sup> Ed.).
7. Fergusson T.S. (1996), A course in Large Sample Theory, Chapman and Hall.

**Programme Outcomes and Course Outcomes Mapping:**

Course Outcomes	Programme Outcomes (POs)								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3								3
CO2	3			3					3
CO3	2				2				3
CO4		3							3
CO5		3							3
CO6		3					1		3
CO7			3			1		1	3

Weight:      1 - Partially related      2 - Moderately Related      3 - Strongly related

**PO1. Disciplinary Knowledge**

**CO1. Understand the concept of Consistent estimator, CAN estimator.** (Weightage: 3 - Strongly Related)

**Justification:** This directly aligns with building a foundational understanding of statistical estimation, specifically focusing on consistency and the concept of CAN estimator.

**CO2. Obtain consistent estimator and their asymptotic distributions.** (Weightage: 3 - Strongly Related)

**Justification:** This builds on the foundational knowledge from PO1, emphasizing the application of consistent estimators and understanding their asymptotic behavior.

**CO3. Choose the ARE estimator among given various consistent estimators.** (Weightage: 2 - Moderately Related)

**Justification:** While related, the emphasis is more on understanding and applying consistent estimators rather than comparing their efficiency (ARE).

## **PO2. Critical Thinking and Problem Solving**

**CO4. Obtain asymptotic distributions of moment estimators, percentile estimators.**

(Weightage: 3 - Strongly Related)

**Justification:** This involves critical thinking in determining the asymptotic distributions for different types of estimators, requiring a deeper understanding of statistical concepts.

**CO5. Determine maximum likelihood estimator and its asymptotic distributions.**

(Weightage: 3 - Strongly Related)

**Justification:** The process of determining maximum likelihood estimators involves critical thinking, and understanding their asymptotic distributions further enhances problem-solving skills.

**CO6. Derive Likelihood Ratio Test (LRT), large sample test Wald's test, and Score test.**

(Weightage: 3 - Strongly Related)

**Justification:** These statistical tests involve critical thinking and problem-solving skills in choosing and applying the appropriate test for a given scenario.

## **PO3. Social Competence**

**CO7. Compute asymptotic confidence interval.** (Weightage: 1 - Partially Related)

**Justification:** While this involves statistical inference, it has less direct relevance to social competence. The focus is more on the technical aspects of estimation and confidence intervals.

## **PO4. Research-related Skills and Scientific Temper**

**CO2. Obtain consistent estimator and their asymptotic distributions.** (Weightage: 3 - Strongly Related)

**Justification:** This aligns with research-related skills, as obtaining and understanding consistent estimators and their asymptotic distributions is fundamental to statistical research.

## **PO5. Trans-disciplinary Knowledge**

**CO3. Choose the ARE estimator among given various consistent estimators.** (Weightage: 2 - Moderately Related)

**Justification:** While not directly trans-disciplinary, the ability to choose the most efficient estimator is a skill that can be applied across various disciplines.

#### **PO6. Personal and Professional Competence**

**CO7. Compute asymptotic confidence interval.** (Weightage: 1 - Partially Related)

**Justification:** The focus here is more on technical competence rather than personal and professional aspects.

#### **PO7. Effective Citizenship and Ethics**

**CO6. Derive Likelihood Ratio Test (LRT), large sample test Wald's test, and Score test.**  
(Weightage: 1 - Partially Related)

**Justification:** The derivation of statistical tests is not directly related to citizenship and ethics, but an understanding of these tests is crucial for ethical and responsible statistical practice.

#### **PO8. Environment and Sustainability**

**CO7. Compute asymptotic confidence interval.** (Weightage: 1 - Partially Related)

**Justification:** The content of this outcome is more focused on statistical methods rather than environmental or sustainability issues.

#### **PO9. Self-directed and Life-long Learning**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** The entire set of outcomes is closely tied to the ability to engage in self-directed learning. Mastering statistical estimation and related concepts requires ongoing learning and adaptation.

**SYLLABUS (CBCS) FOR M.Sc. Statistics  
(2022 Pattern)  
(With effect from Academic Year 2023-2024)**

<b>Name of the Programme</b>	: M.Sc. Statistics
<b>Program Code</b>	: PSST
<b>Class</b>	: M.Sc. Part – II
<b>Semester</b>	: III
<b>Course Name</b>	: Design and Analysis of Experiments
<b>Course Code</b>	: PSST232
<b>No. of Lectures</b>	: 60
<b>No. of Credits</b>	: 4

**A) Course objectives:**

1. The main objective of this course is to learn and understand various designs of experiments.
2. Students should be able to design and carryout various experiments and analyze the data.
3. Students should be able to apply appropriate design in real life situation.

**B) Course outcomes:**

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

- CO1. understand basic principles and various terms of Design of Experiments.
- CO2. apply factorial design in real life problems.
- CO3. apply fractional factorial design in real life problems.
- CO4. implicating total confounding and partial confounding in real life problems.
- CO5. apply appropriate design in real life situation
- CO6. analyze the data of various experimental design.
- CO7. understand the concept of Taguchi methods.

**TOPICS/CONTENTS:**

**Unit – 1**

Estimability of linear parametric function, necessary and sufficient condition for estimability, Best Linear Unbiased Estimator (BLUE), Gauss-Markov set up, Least square estimation, Normal equations, Consistency of system of normal equations and their solution,



Gauss-Markov theorem, Variances and covariances of BLUE's, Estimation space, Error space, their ranks, Orthogonality of estimation space and error space, Simultaneous estimates of linear parametric function, Estimation of error variance, Estimation with correlated observations, Least square estimates with restriction on parameters, Method of generalized least squares. (15 L)

### **Unit – 2**

Incomplete Block Design, Balanced Incomplete Block Design (BIBD),  $2^k$  full factorial experiments, concepts of main effects, interaction effect, their graphical representation, analysis of single replicate and more than one replicates of  $2^k$  design using ANOVA total and partial confounding of  $2^k$  design in  $2p$  blocks  $p = 2, 3$ . Two level fractional factorial experiments, resolution of a design (III, IV and V), aberration of a design, aliases, generators of the design, complete defining relation.

(12 L)

### **Unit – 3**

$3^k$  factorial design: contrasts for linear and quadratic effects, statistical analysis, confounding and fractional confounding experiments in  $3^k$  factorial design, Response Surface Methodology (RSM): linear and quadratic model and stationary point (Practical only), Central Composite Designs (CCD) (Practical only), Box-Behnken design.

(18 L)

### **Unit – 4**

Split plot design and Plackett – Burman designs (Practical only), Taguchi methods: Concept of noise and control factors, inner and outer arrays, concept of loss function, S/N ratio, orthogonal arrays, linear graphs, interaction tables, ANOVA, random effect models and mixed models, Nested design.

(15 L)

### **References:**

1. Dean, A. and Voss, D. (1999). *Design and Analysis of Experiments*, Springer.
2. George E. P. Box, Draper N.R. (1987). *Empirical Model-Building and Response Surfaces*, Wiley.
3. Kshirsagar A.M. (1983). *Linear Models*, Marcel Dekker.
4. Montgomery, D.C. (2001). *Design and Analysis of Experiments*, Wiley.
5. Phadke, M.S. (1989). *Quality Engineering using Robust Design*, Prentice Hall,
6. Englewood Cliffs, New Jersey.

8. Wu, C.F. Jeff and Hamada M. (2000). Experiments: Planning, Analysis and Parameter
9. Design Optimization, John Wiley and Sons.
10. Bapat, R. B. (2012). Linear algebra and linear models. Springer Science & Business Media.

**Programme Outcomes and Course Outcomes Mapping:**

Course Outcomes	Programme Outcomes (POs)								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	3	1	3		1	1	1	3
CO2	3	3	1	3		1	1	1	3
CO3	3	3	1	3		1	1	1	3
CO4	3	3	1	3		1	1	1	3
CO5	3	3	1	3		1	1	1	3
CO6	3	3	1	3		1	1	1	3
CO7	2	3	1	3	2	1	1	1	3

Weight:      1 - Partially related      2 - Moderately Related      3 - Strongly related

**PO1. Disciplinary Knowledge**

**CO1. Understand basic principles and various terms of Design of Experiments.** (Weightage: 3 - Strongly Related)

**Justification:** This directly aligns with building foundational knowledge in the field of experimental design, emphasizing the understanding of principles and terms.

**CO2. Apply factorial design in real-life problems.** (Weightage: 3 - Strongly Related)

**Justification:** This extends disciplinary knowledge into practical application, requiring the application of factorial design principles in solving real-life problems.

**CO3. Apply fractional factorial design in real-life problems.** (Weightage: 3 - Strongly Related)

**Justification:** Similar to CO2, this involves the application of fractional factorial design, deepening the understanding and application of experimental design concepts.

**CO4. Implicate total confounding and partial confounding in real-life problems.**  
(Weightage: 3 - Strongly Related)

**Justification:** Understanding and implicating confounding factors are integral to disciplinary knowledge in experimental design.

**CO5. Apply appropriate design in real-life situations.** (Weightage: 3 - Strongly Related)

**Justification:** This is a direct application of disciplinary knowledge, requiring the selection and application of the appropriate experimental design in various situations.

**CO6. Analyze the data of various experimental designs.** (Weightage: 3 - Strongly Related)

**Justification:** Analyzing data from different experimental designs is a critical aspect of disciplinary knowledge, enhancing the ability to draw meaningful conclusions.

**CO7. Understand the concept of Taguchi methods.** (Weightage: 2 - Moderately Related)

**Justification:** While Taguchi methods are related, they represent a specific aspect of experimental design that may not be as fundamental to basic principles. Hence, the moderately related weightage.

## **PO2. Critical Thinking and Problem Solving**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** The entire set of outcomes involves critical thinking and problem-solving skills, from understanding principles to applying designs and analyzing data.

## **PO3. Social Competence**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The content is more technical and focused on statistical methodologies, with less direct relevance to social competence.

## **PO4. Research-related Skills and Scientific Temper**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Experimental design is a fundamental aspect of research, and the outcomes align closely with research-related skills and the development of a scientific temper.

### **PO5. Trans-disciplinary Knowledge**

**CO7. Understand the concept of Taguchi methods.** (Weightage: 2 - Moderately Related)

**Justification:** Taguchi methods have applications beyond traditional experimental design, making this outcome moderately related to trans-disciplinary knowledge.

### **PO6. Personal and Professional Competence**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** While the outcomes contribute to technical competence, the direct connection to personal and professional aspects is limited.

### **PO7. Effective Citizenship and Ethics**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The outcomes are more technical and less directly related to citizenship and ethics.

### **PO8. Environment and Sustainability**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The focus here is more on statistical methodologies and less on environmental or sustainability aspects.

### **PO9. Self-directed and Life-long Learning**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Experimental design is a dynamic field, and mastering its principles requires ongoing self-directed learning and adaptation.

**SYLLABUS (CBCS) FOR M.Sc. Statistics  
(2022 Pattern)  
(With effect from Academic Year 2023-2024)**

<b>Name of the Programme</b>	: M.Sc. Statistics
<b>Program Code</b>	: PSST
<b>Class</b>	: M.Sc. Part – II
<b>Semester</b>	: II
<b>Course Name</b>	: Time Series Analysis
<b>Course Code</b>	: PSST233
<b>No. of Lectures</b>	: 60
<b>No. of Credits</b>	: 4

**A) Course Objectives:**

1. The main objective of this course is that students should understand various time series models, estimation of its parameters and be able to make predictions.
2. To learn the concepts like Auto-covariance, auto-correlation function and vector auto regression.

**B) Course Outcomes:**

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

- CO1.** fit the model on time series data like ARMA, ARIMA, SARIMA, ARCH and GARCH properties.
- CO2.** apply and understand the techniques for estimating parameters of time series models also the role of maximum likelihood estimation in time series modeling.
- CO3.** perform diagnostic checks on time series models to assess model adequacy.
- CO4.** identify and address issues such as autocorrelation and heteroscedasticity.
- CO5.** analyses time series data and use multivariate time series models such as vector auto regression (VAR).
- CO6.** Gain proficiency in using ITSM, R and Python to fit an appropriate time series model and infer the results.

**CO7.** effectively interpret the results of time series analyses, both in written reports and oral presentations.

**TOPICS/CONTENTS:**

**Unit-1**

Exploratory time Series analysis, Time Series as a discrete parameter stochastic process, tests for trend and seasonality, moving average smoothing, exponential smoothing, double (Holt exponential smoothing), Triple (Holt -Winters exponential smoothing), adaptive smoothing definition and its application. Auto covariance and autocorrelation functions and their properties, Portmanteau tests for noise sequences. (15 L)

**Unit 2**

Stationary process: General linear process, stationary process and strict stationary process, moving average (MA), Auto Regressive (AR) and autoregressive moving average (ARMA). Concept of causality, invertibility, Computation of  $\pi$ -weights and  $\psi$ -weights. computation of ACVF and ACF. Partial auto covariance function. Periodogram and correlogram analysis. (15 L)

**Unit 3**

Non-stationary: Unit root, non-stationary unit root test, Integrated ARMA (ARIMA) model, Analysis of seasonal models: Parsimonious models for seasonal time series, SARIMA models, forecasting, identification, estimation and diagnosis methods for seasonal time series. Yule-Walker estimation. Estimation of ARIMA models parameters, Maximum likelihood method for estimation. Durbin-Levison algorithm, innovation algorithm, Box- Jenkins model (Without proof). (15 L)

**Unit 4**

Graphical method for choosing AR and MA lags, FPE, AIC, BIC, residual analysis, conditional heteroscedastic models, volatility models, ARCH and GARCH properties, examples, estimation and forecasting. Introduction to Multivariate Time series model, VAR models, vector ARMA models. (15 L)

**References:**



**CO4. Identify and address issues such as autocorrelation and heteroscedasticity.**  
(Weightage: 3 - Strongly Related)

**Justification:** Dealing with issues like autocorrelation and heteroscedasticity is crucial in time series modeling, contributing directly to disciplinary knowledge.

**CO5. Analyze time series data and use multivariate time series models such as vector auto-regression (VAR).** (Weightage: 3 - Strongly Related)

**Justification:** Analyzing time series data and using multivariate models extends disciplinary knowledge to a broader and more advanced level.

**CO6. Gain proficiency in using ITSM, R, and Python to fit an appropriate time series model and infer the results.** (Weightage: 3 - Strongly Related)

**Justification:** Proficiency in using specific tools and programming languages is integral to disciplinary knowledge in the modern context of time series analysis.

**CO7. Effectively interpret the results of time series analyses, both in written reports and oral presentations.** (Weightage: 3 - Strongly Related)

**Justification:** The ability to interpret and communicate results is a critical aspect of disciplinary knowledge, especially in the context of time series analysis.

### **PO2. Critical Thinking and Problem Solving**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Time series analysis requires critical thinking and problem-solving skills at every stage, from model fitting to interpretation of results.

### **PO3. Social Competence**

**CO7. Effectively interpret the results of time series analyses, both in written reports and oral presentations.** (Weightage: 1 - Partially Related)

**Justification:** While interpretation skills are essential, the direct social relevance is limited in this technical context.

### **PO4. Research-related Skills and Scientific Temper**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Time series analysis is a research-oriented field, and the outcomes align closely with research-related skills and scientific temper.

### **PO5. Trans-disciplinary Knowledge**

**CO5. Analyze time series data and use multivariate time series models such as vector**



**auto-regression (VAR).** (Weightage: 2 - Moderately Related)

**Justification:** The use of multivariate models extends the knowledge to a more trans-disciplinary context, although not fully.

#### **PO6. Personal and Professional Competence**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The outcomes contribute more to technical competence, with limited direct connection to personal and professional aspects.

#### **PO7. Effective Citizenship and Ethics**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The content is more technical and less directly related to citizenship and ethics.

#### **PO8. Environment and Sustainability**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The focus here is more on statistical methodologies and less on environmental or sustainability aspects.

#### **PO9. Self-directed and Life-long Learning**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Time series analysis is a dynamic field, and mastering its principles requires ongoing self-directed learning and adaptation.

**SYLLABUS (CBCS) FOR M.Sc. Statistics  
(2022 Pattern)  
(With effect from Academic Year 2023-2024)**

<b>Name of the Programme</b>	: M.Sc. Statistics
<b>Program Code</b>	: PSST
<b>Class</b>	: M.Sc. Part – II
<b>Semester</b>	: III
<b>Course Name</b>	: Data Mining
<b>Course Code</b>	: PSST234 (A)
<b>No. of Lectures</b>	: 60
<b>No. of Credits</b>	: 4

**A) Course Objectives:**

1. Students should understand Big Data, Data Warehouse, Data Mining Principles.
2. Students should be able to identify appropriate data mining techniques to analyze big data.
3. Evaluating efficiency of different data mining techniques like classification prediction, clustering and association rule mining

**B) Course Outcomes:**

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

- CO1.** understand the basic concepts, goals, and challenges of data mining.
- CO2.** explore the role of data mining in extracting meaningful patterns and knowledge from large datasets.
- CO3.** Study and apply a variety of data mining techniques, like CART, SVM, PCA, KNN, etc.
- CO4.** understand and implement supervised learning algorithms for classification and regression tasks, and explore unsupervised learning techniques, including clustering algorithms.
- CO5.** Explore ethical issues related to data mining, including privacy concerns and bias in algorithms.
- CO6.** apply data mining techniques to real-world datasets, and interpret the results and draw actionable insights from the analysis.

**CO7.** gain practical experience by working with data mining tools and software like, R, Python.

**TOPICS/CONTENTS:**

**Unit- 1**

Introduction to big data, Data preparation for knowledge discovery: Data understanding and data cleaning tools, Data transformation, Data Discretization, Data Visualization, Imbalanced data, Data Mining Process: CRISP and SEEMA; Concept of training data, testing data and validation of model. supervised and unsupervised learning, review of linear discriminant analysis, logistic regression, clustering procedure - k means, hierarchical, principal component analysis. (15 L)

**Unite -2**

Bayes classifier, nearest neighbor classifier, Classification and Regression tree (CART): information gain, gain ratio, entropy – Gini index, artificial neural network, convoluted neural network (CNN), Support Vector Machine (SVM) for linearly separable data and linearly in separable data. (15 L)

**Unit- 3**

**Model evaluation and selection methods:** Metrics for evaluating classifier performance (confusion matrix), precision, recall F1 score, holdout method and random sampling, cross validation, bootstrap, ROC curves, AIC, BIC, CIC, DIC (information criterion), bias variance tradeoff.

**Techniques to improve classification accuracy:** Bagging, boosting, Ada boosting, Random forest, gradient boosting. (15 L)

**Unit -4**

Self-Organizing Map (SOM), EM algorithm, market basket analysis, text mining: sentiment analysis, word frequency analysis, N-grams and correlation, topic modeling. (15 L)

**References:**

1. Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). Classification and Regression Trees. (Wadsworth and Brooks/Cole).
2. Daniel T.Larose, (2006). Data Mining Methods and Models, Wile-Interscience.



**CO3. Study and apply a variety of data mining techniques, like CART, SVM, PCA, KNN, etc.** (Weightage: 3 - Strongly Related)

**Justification:** Studying and applying various data mining techniques directly enhances disciplinary knowledge by covering a range of methods used in the field.

**CO4. Understand and implement supervised learning algorithms for classification and regression tasks, and explore unsupervised learning techniques, including clustering algorithms.** (Weightage: 3 - Strongly Related)

**Justification:** Understanding and implementing both supervised and unsupervised learning algorithms contributes significantly to disciplinary knowledge in data mining.

**CO5. Explore ethical issues related to data mining, including privacy concerns and bias in algorithms.** (Weightage: 2 - Moderately Related)

**Justification:** While not the primary focus, ethical considerations are moderately related to disciplinary knowledge, as they influence the responsible application of data mining techniques.

**CO6. Apply data mining techniques to real-world datasets, and interpret the results and draw actionable insights from the analysis.** (Weightage: 3 - Strongly Related)

**Justification:** Applying data mining techniques to real-world datasets and drawing actionable insights directly contributes to disciplinary knowledge in a practical context.

**CO7. Gain practical experience by working with data mining tools and software like R, Python.** (Weightage: 3 - Strongly Related)

**Justification:** Gaining practical experience with tools like R and Python is essential for disciplinary knowledge in data mining, emphasizing hands-on skills.

### **PO2. Critical Thinking and Problem Solving**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Data mining involves critical thinking at every stage, from understanding concepts to applying techniques and interpreting results.

### **PO3. Social Competence**

**CO5. Explore ethical issues related to data mining, including privacy concerns and bias in algorithms.** (Weightage: 2 - Moderately Related)

**Justification:** Ethical considerations in data mining have a social impact, making this outcome moderately related to social competence.

### **PO4. Research-related Skills and Scientific Temper**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** The outcomes align closely with research-related skills and the development of a scientific temper in the context of data mining.

#### **PO5. Trans-disciplinary Knowledge**

**CO5. Explore ethical issues related to data mining, including privacy concerns and bias in algorithms.** (Weightage: 2 - Moderately Related)

**Justification:** Ethical considerations in data mining can have trans-disciplinary implications beyond the core technical aspects.

#### **PO6. Personal and Professional Competence**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** While technical skills are emphasized, the outcomes have limited direct connections to personal and professional aspects.

#### **PO7. Effective Citizenship and Ethics**

**CO5. Explore ethical issues related to data mining, including privacy concerns and bias in algorithms.** (Weightage: 2 - Moderately Related)

**Justification:** Ethical issues related to data mining have direct implications for effective citizenship and ethical considerations.

#### **PO8. Environment and Sustainability**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The outcomes are more focused on data mining methodologies than on environmental or sustainability aspects.

#### **PO9. Self-directed and Life-long Learning**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Data mining is a dynamic field, and mastering its principles requires ongoing self-directed learning and adaptation.

**SYLLABUS (CBCS) FOR M.Sc. Statistics**  
**(2022 Pattern)**  
**(With effect from Academic Year 2023-2024)**

<b>Name of the Programme</b>	: M.Sc. Statistics
<b>Program Code</b>	: PSST
<b>Class</b>	: M.Sc. Part – II
<b>Semester</b>	: III
<b>Course Name</b>	: Design and Analysis of Clinical Trials
<b>Course Code</b>	: PSST234 (B)
<b>No. of Lectures</b>	: 60
<b>No. of Credits</b>	: 4

**A) Course Objectives:**

1. Students will be able to understand the basic principles of probability and how they are related to biostatistics.
2. Establishing an objective framework for conducting an investigation for clinical trials.
3. To provide an unbiased evaluation of the merits of using one or more treatment options for a given disease or condition of interest.

**B) Course Outcomes:**

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

- CO1.** understand the principles of Good Clinical Practice.
- CO2.** demonstrate an understanding of the essential principles of modern bio-statistical methods and statistical software and how to apply them.
- CO3.** learn methods for determining the appropriate sample size for a clinical trial.
- CO4.** understand the purpose and importance of clinical trials in medical research.
- CO5.** explore the phases of clinical trials and their objectives.
- CO6.** learn about ethical guidelines and regulatory requirements governing clinical trials.
- CO7.** develop and implement statistical analysis plans for clinical trials, and understand the principles of intention-to-treat analysis.

## **TOPICS/CONTENTS:**

### **Unit-1**

Introduction to Clinical Trials (CTs): epidemiology, need and ethics of CTs, History of clinical trials, New Drug Application, overview of phase I-IV trials, clinical trial protocol, Bias and Random error, Objective and points of CTs. (15L)

### **Unit 2**

Design of clinical trials: Basic design consideration, introduction, patient selection, selection control parallel and cross-over designs, cross-sectional and longitudinal designs, balanced incomplete block and designs, Titration designs, Enrichment Designs. Randomization models, Randomization methods, Implementation of Randomization, Generalization of controlled Randomized trials blinding. (15L)

### **Unit 3**

Bio availability and Bio equivalence studies: History Bioavailability studies, Formulation and Routes of administration, Pharmacokinetic parameter, Clinically importance differences, Assessment of Bioequivalence Statistical inference for standard 2 x 2 crossover designs: The carry-over effect, The direct drug effect, The period effect Analysis of Variance (ANOVA), Assessment of inter and intra subject variability. (15L)

### **Unit 4**

Multicenter trials, nonparametric test, outlier detection in clinical trials, power and sample size determination, drug interaction study, dose proportionality study, steady state analysis, Meta-analysis. (15L)

## **References:**

1. Chow S. C. and Liu J. P. (2009) Design and Analysis of Bioavailability and bioequivalence, 3<sup>rd</sup> Edn. CRC Press.
2. Chow S. C. and Liu J.P. (2004) Design and Analysis of Clinical Trials, 2<sup>nd</sup> Edn. Marcel Dekkar.





**CO4. Understand the purpose and importance of clinical trials in medical research.**  
(Weightage: 3 - Strongly Related)

**Justification:** Understanding the purpose and importance of clinical trials is fundamental to disciplinary knowledge in medical research.

**CO5. Explore the phases of clinical trials and their objectives.** (Weightage: 3 - Strongly Related)

**Justification:** Exploring the phases of clinical trials directly contributes to disciplinary knowledge, providing insights into the systematic process of medical research.

**CO6. Learn about ethical guidelines and regulatory requirements governing clinical trials.**  
(Weightage: 3 - Strongly Related)

**Justification:** Understanding ethical guidelines and regulatory requirements is crucial for maintaining disciplinary knowledge and ensuring the integrity of clinical research.

**CO7. Develop and implement statistical analysis plans for clinical trials, and understand the principles of intention-to-treat analysis.** (Weightage: 3 - Strongly Related)

**Justification:** Developing and implementing statistical analysis plans for clinical trials, including the intention-to-treat analysis, is a key component of disciplinary knowledge in this field.

### **PO2. Critical Thinking and Problem Solving**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Clinical trials involve critical thinking and problem-solving skills at every stage, from study design to statistical analysis.

### **PO3. Social Competence**

**CO6. Learn about ethical guidelines and regulatory requirements governing clinical trials.**  
(Weightage: 2 - Moderately Related)

**Justification:** The understanding of ethical guidelines and regulatory requirements contributes to social competence, as it involves ethical considerations in conducting research involving human subjects.

### **PO4. Research-related Skills and Scientific Temper**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** The outcomes align closely with research-related skills and the development of a scientific temper in the context of clinical research.

**PO5. Trans-disciplinary Knowledge**

**CO6. Learn about ethical guidelines and regulatory requirements governing clinical trials.**

(Weightage: 2 - Moderately Related)

**Justification:** Ethical guidelines and regulatory requirements can have trans-disciplinary implications, as they may extend beyond the strictly clinical or medical domain.

**PO6. Personal and Professional Competence**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The outcomes contribute more to technical competence in the context of clinical research, with limited direct connections to personal and professional aspects.

**PO7. Effective Citizenship and Ethics**

**CO6. Learn about ethical guidelines and regulatory requirements governing clinical trials.**

(Weightage: 2 - Moderately Related)

**Justification:** The understanding of ethical guidelines and regulatory requirements contributes to effective citizenship and ethical considerations in the context of clinical research.

**PO8. Environment and Sustainability**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The focus here is more on clinical and research methodologies than on environmental or sustainability aspects.

**PO9. Self-directed and Life-long Learning**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Clinical research is a dynamic field, and mastering its principles requires ongoing self-directed learning and adaptation.

**SYLLABUS (CBCS) FOR M.Sc. Statistics  
(2022 Patern)  
(With effect from Academic Year 2023-2024)**

<b>Name of the Programme</b>	: M.Sc. Statistics
<b>Program Code</b>	: PSST
<b>Class</b>	: M.Sc. Part – II
<b>Semester</b>	: III
<b>Course Name</b>	: Practical Paper -V
<b>Course Code</b>	: PSST235
<b>No. of Lectures</b>	: 60
<b>No. of Credits</b>	: 4

**A) Course Objectives:**

1. The main objective of this course is to learn and understand various designs of experiments.
2. Students should be able to design and carryout various experiments and analyze the data.
3. Students should be able to apply appropriate design in real life situation.

**B) Course Outcomes:**

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

- CO1.** apply factorial design in real life problems.
- CO2.** apply fractional factorial design in real life problems.
- CO3.** implicating total confounding and partial confounding in real life problems.
- CO4.** understand the concept of Taguchi methods.
- CO5.** obtain and verify consistent estimator and their asymptotic distributions.
- CO6.** derive Likelihood Ratio Test (LRT), large sample test Wald's test, and Score test.
- CO7.** plotting likelihood function and obtain the MLE by scoring method.

Sr. No.	Title of Experiments
1.	Balance Incomplete Block Design (Intra block analysis)
2.	Analysis of Covariance in one way and two-way model
3.	Analysis of $3^k$ factorial experiments, $3^k$ fractional factorial experiment
4.	Total Confounding and Partial Confounding in $3^k$ factorial experiment
5.	Fitting first and second order response surface model, central composite design contour, surface plots, canonical analysis of stationery points, Blocking in RSM

6.	Random effect model with one factor, estimation of variance
7.	Taguchi methods: S/N ratio, orthogonal arrays, triangular tables, linear graphs, inner and outer arrays
8.	Verification of consistency and asymptotic normality of the estimators
9.	Comparing Consistent estimator, MSE and sample size considerations
10.	Asymptotic Confidence Intervals and analysis of three-dimensional contingency tables
11.	Power functions of large sample test (LR, Wald, Rao)
12.	Plotting Likelihood function, MLE by methods of scoring

### Programme Outcomes and Course Outcomes Mapping:

Course Outcomes	Programme Outcomes (POs)								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	3	1	3		1	1	1	3
CO2	3	3	1	3		1	1	1	3
CO3	3	3	1	3		1	1	1	3
CO4	3	3	1	3	2	1	1	1	3
CO5	2	3	1	3		1	1	1	3
CO6	2	3	1	3		1	1	1	3
CO7	2	3	1	3		1	1	1	3

Weight:      1 - Partially related    2 - Moderately Related                      3 - Strongly related

#### PO1. Disciplinary Knowledge

**CO1. Apply factorial design in real-life problems.** (Weightage: 3 - Strongly Related)

**Justification:** Applying factorial design directly contributes to building disciplinary knowledge in experimental design and statistical methodologies.

**CO2. Apply fractional factorial design in real-life problems.** (Weightage: 3 - Strongly Related)

**Justification:** Similar to CO1, applying fractional factorial design enhances disciplinary knowledge in experimental design and statistical techniques.

**CO3. Implicate total confounding and partial confounding in real-life problems.** (Weightage: 3 - Strongly Related)

**Justification:** Understanding and implicating confounding factors in real-life problems directly aligns with building disciplinary knowledge in statistical design.

**CO4. Understand the concept of Taguchi methods.** (Weightage: 3 - Strongly Related)

**Justification:** Understanding Taguchi methods contributes to disciplinary knowledge in the context of experimental design and optimization.

**CO5. Obtain and verify consistent estimator and their asymptotic distributions.**  
(Weightage: 2 - Moderately Related)

**Justification:** While related, the emphasis here is more on statistical estimation, making it moderately related to the broader discipline.

**CO6. Derive Likelihood Ratio Test (LRT), large sample test Wald's test, and Score test.**  
(Weightage: 2 - Moderately Related)

**Justification:** Deriving statistical tests is moderately related to disciplinary knowledge, focusing on the application of statistical methods.

**CO7. Plotting likelihood function and obtain the MLE by scoring method.** (Weightage: 2 - Moderately Related)

**Justification:** Similarly, plotting likelihood functions and obtaining Maximum Likelihood Estimators (MLE) is moderately related to disciplinary knowledge, emphasizing statistical inference.

### **PO2. Critical Thinking and Problem Solving**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** All outcomes involve critical thinking and problem-solving skills, from experimental design to statistical estimation and hypothesis testing.

### **PO3. Social Competence**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The outcomes are more technical and focused on statistical methodologies, with less direct relevance to social competence.

### **PO4. Research-related Skills and Scientific Temper**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** The outcomes align closely with research-related skills and the development of a scientific temper in the context of statistical research and experimentation.

### **PO5. Trans-disciplinary Knowledge**

**CO4. Understand the concept of Taguchi methods.** (Weightage: 2 - Moderately Related)

**Justification:** Taguchi methods, while specific to certain fields, have some trans-disciplinary applications, making this outcome moderately related.

#### **PO6. Personal and Professional Competence**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The outcomes contribute more to technical competence, with limited direct connections to personal and professional aspects.

#### **PO7. Effective Citizenship and Ethics**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The content is more technical and less directly related to citizenship and ethics.

#### **PO8. Environment and Sustainability**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The outcomes are more focused on statistical methodologies than on environmental or sustainability aspects.

#### **PO9. Self-directed and Life-long Learning**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** The outcomes require ongoing self-directed learning and adaptation, aligning with the goal of lifelong learning in statistical research and experimentation.

Top of Form

**SYLLABUS (CBCS) FOR M.Sc. Statistics  
(2022 Pattern)  
(With effect from Academic Year 2023-2024)**

<b>Name of the Programme</b>	: M.Sc. Statistics
<b>Program Code</b>	: PSST
<b>Class</b>	: M.Sc. Part – II
<b>Semester</b>	: III
<b>Course Name</b>	: Statistics Practical – VI
<b>Course Code</b>	: PSST236
<b>No. of Lectures</b>	: 60
<b>No. of Credits</b>	: 4

**A) Course Objectives:**

1. The main objective of this course is that students should understand various time series models, estimation of its parameters and be able to make predictions.
2. To learn the concepts like Auto-covariance, auto-correlation function and vector auto regression.
3. Students should understand Big Data, Data Warehouse, Data Mining Principles.

**A) Course Outcomes:**

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

- CO1.** fit the model on time series data like ARMA, ARIMA, SARIMA, ARCH and GARCH properties.
- CO2.** apply and understand the techniques for estimating parameters of time series models also the role of maximum likelihood estimation in time series modeling.
- CO3.** perform diagnostic checks on time series models to assess model adequacy.
- CO4.** identify and address issues such as autocorrelation and heteroscedasticity.
- CO5.** develop skills in time series forecasting using appropriate models.
- CO6.** gain proficiency in using ITSM, R and Python to fit an appropriate time series model and infer the results.
- CO7.** effectively communicate the results of time series analyses, both in written reports and oral presentations.



Sr. No.	Title of Experiments	
1.	Smoothing time series using various filters (exponential, MA), Box-Cox transformation, differencing, checking stationarity and normality after transformation.	
2.	ACF/PACF of series and residual analysis, stationarity, causality and invertibility	
3.	Order selection in time series: use of ACF/PACF and AIC, BIC, fitting of AR, MA models (conditional least squares or maximum likelihood).	
4.	Fitting of ARMA, ARIMA and SARIMA models (conditional least squares or maximum likelihood).	
5.	Forecasting using fitted linear models (recursively), Holt -Winters forecasts construction of forecast intervals.	
6.	Fitting heteroscedastic models: checking for heteroscedasticity from residuals, ARCH, GARCH modeling.	
<b>Practical Based on Optional Paper</b>		
	<b>PSST235(A): Data Mining</b>	<b>PSST235(B): Design and Analysis of Clinical Trials</b>
7.	Supervised Learning	Research Paper Review – I
8.	Artificial Neural Network	Research Paper Review – II
9.	Support Vector Machine	Case Study – I
10.	Unsupervised Learning	Case Study – II
11.	Case Study – I	Mini Project
12.	Case Study – II	

**Programme Outcomes and Course Outcomes Mapping:**

Course Outcomes	Programme Outcomes (POs)								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
<b>CO1</b>	3	3	1	3		1	1	1	3
<b>CO2</b>	3	3	1	3		1	1	1	3
<b>CO3</b>	3	3	1	3		1	1	1	3
<b>CO4</b>	3	3	1	3	2	1	1	1	3
<b>CO5</b>	2	3	1	3		1	1	1	3
<b>CO6</b>	2	3	1	3		1	1	1	3
<b>CO7</b>	2	3	1	3		1	1	1	3

Weight:      1 - Partially related      2 - Moderately Related      3 - Strongly related

**PO1. Disciplinary Knowledge**

**CO1. Fit the model on time series data like ARMA, ARIMA, SARIMA, ARCH, and GARCH properties.** (Weightage: 3 - Strongly Related)

**Justification:** Fitting various time series models aligns directly with building disciplinary knowledge in time series analysis and forecasting.

**CO2. Apply and understand the techniques for estimating parameters of time series models also the role of maximum likelihood estimation in time series modeling.** (Weightage: 3 - Strongly Related)

**Justification:** Understanding and applying estimation techniques in time series models contributes significantly to disciplinary knowledge in time series analysis.

**CO3. Perform diagnostic checks on time series models to assess model adequacy.** (Weightage: 3 - Strongly Related)

**Justification:** Diagnostic checks are essential for assessing the adequacy of time series models, enhancing disciplinary knowledge in model evaluation.

**CO4. Identify and address issues such as autocorrelation and heteroscedasticity.** (Weightage: 3 - Strongly Related)

**Justification:** Addressing issues like autocorrelation and heteroscedasticity directly contributes to disciplinary knowledge in time series analysis.

**CO5. Develop skills in time series forecasting using appropriate models.** (Weightage: 3 - Strongly Related)

**Justification:** Developing forecasting skills with appropriate time series models is fundamental to disciplinary knowledge in this field.

**CO6. Gain proficiency in using ITSM, R, and Python to fit an appropriate time series model and infer the results.** (Weightage: 3 - Strongly Related)

**Justification:** Proficiency in using software tools to fit time series models and infer results is a practical aspect of disciplinary knowledge in time series analysis.

**CO7. Effectively communicate the results of time series analyses, both in written reports and oral presentations.** (Weightage: 3 - Strongly Related)

**Justification:** Communicating results effectively is an integral part of disciplinary knowledge, ensuring that insights are conveyed to stakeholders.

## **PO2. Critical Thinking and Problem Solving**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Time series analysis requires critical thinking and problem-solving skills at every stage, from model fitting to communication of results.

### **PO3. Social Competence**

**CO7. Effectively communicate the results of time series analyses, both in written reports and oral presentations.** (Weightage: 2 - Moderately Related)

**Justification:** Effective communication of results has a social dimension, making this outcome moderately related to social competence.

### **PO4. Research-related Skills and Scientific Temper**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** The outcomes align closely with research-related skills and the development of a scientific temper in the context of time series analysis.

### **PO5. Trans-disciplinary Knowledge**

**CO6. Gain proficiency in using ITSM, R, and Python to fit an appropriate time series model and infer the results.** (Weightage: 2 - Moderately Related)

**Justification:** Proficiency in using various software tools has trans-disciplinary applications, making this outcome moderately related.

### **PO6. Personal and Professional Competence**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The outcomes contribute more to technical competence, with limited direct connections to personal and professional aspects.

### **PO7. Effective Citizenship and Ethics**

**All COs** (Weightage: 2 - Moderately Related)

**Justification:** Ethical considerations in communicating results and using software tools make this outcome moderately related to effective citizenship and ethics.

### **PO8. Environment and Sustainability**

**All COs** (Weightage: 1 - Partially Related)

**Justification:** The outcomes are more focused on time series analysis methodologies than on environmental or sustainability aspects.

### **PO9. Self-directed and Life-long Learning**

**All COs** (Weightage: 3 - Strongly Related)

**Justification:** Time series analysis is a dynamic field, and mastering its principles requires ongoing self-directed learning and adaptation.