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

Paper Code	Title of Paper	No. of Credits
PSST231	Asymptotic Inference	4
PSST232	Design and Analysis of Experiments	4
PSST233	Time Series Analysis	4
PSST234 (A) PSST234 (B)	Data Mining Or Design and Analysis of Clinical Trials	4
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

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

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

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

Course Outcomes:

Students will be able to

1. understand the asymptotic behavior of the estimators, concept of consistent estimator, CAN estimator, Likelihood Ratio Test (LRT), Wald's test, Score test, large sample test and asymptotic confidence interval.
2. find and verify the consistent estimator and consistency and asymptotic normality (CAN) estimator, asymptotic distributions of moment estimators, percentile estimators, maximum likelihood estimator.
3. test the hypothesis.

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, 2nd Ed.).
7. Fergusson T.S. (1996), A course in Large Sample Theory, Chapman and Hall.

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

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

Course Outcomes:

Students will be able to

1. learn and understand various designs of experiments.
2. design and carryout various experiments and analyze the data.
3. apply appropriate design in real life situation.
4. apply Factorial design, fractional factorial design, confounding in real life problems.

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*
3. *Surfaces*, Wiley.
4. Kshirsagar A.M. (1983). *Linear Models*, Marcel Dekker.
5. Montgomery, D.C. (2001). *Design and Analysis of Experiments*, Wiley.
6. Phadke, M.S. (1989). Quality Engineering using Robust Design, Prentice Hall,
7. 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.

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

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

Course Outcomes:

Students will be able to

1. model time series data by various time series models like ARMA, ARIMA, SARIMA, ARCH and GARCH properties.
2. use the Box-Jenkins approach to model and forecast time series data empirically.
3. analyses time series data and use multivariate time series models such as vector auto regression (VAR).
4. use ITSM, R and Python to fit an appropriate time series model and infer the results.

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 π-weights and ψ-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:

1. Brockwell, P.J. and Davis, R. A. Introduction to Time Series Analysis, Springer.
2. Chatfield, C. (2001). Time Series Forecasting, Chapmann & hall, London.
3. Fuller, W. A. (1996). Introduction to Statistical Time Series, 2nd Ed. John Wiley.
4. Hamilton N. Y. (1994). Time Series Analysis. Princeton University press. Princeton.
5. Kendall, Sir Maurice and Ord, J. K. (1990). Time Series (Third Edition), Edward Arnold.

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

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

Course Outcomes:

Students will be able to

1) understand Big Data, Data Warehouse and Data Mining Principles.
2) 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.
4) propose data-mining solutions for different applications.
5) evaluate different models used for data processing.

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, gradiant 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.
3. Galit Shmueli, Nitin Patel, Peter Bruce, (2010). Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner, Wiley
4. Hastie T., Tibshirani R. and Friedman J. H., (2003). The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer.
5. Mitchell Tom, (1997). Machine Learning. McGraw-Hill.
6. Ripley, B.D. (1996). Pattern Recognition and Neural Networks. (Cambridge University Press).
7. Gareth M. James, Trevor Hastie, Daniela Witten, Robert Tibshirani, Introduction to Statistical Learning using R, Springer.
8. Julia Silge and David Robinson, (2017) Text Mining with R, a Tidy Approach, O'Reilly Publication..

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

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

Course Outcomes:

Students will be able to

1. 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.
4. understand the basic statistical principles, methods for clinical data analysis and reporting.
5. demonstrate an understanding of the essential principles of modern bio-statistical methods and statistical software and how to apply them.

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, patent 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, 3rd Edn. CRC Press.
2. Chow S. C. and Liu J.P. (2004) Design and Analysis of Clinical Trials, 2nd Edn. Marcel Dekkar.
3. Fleiss J. L. (1989) The Design and Analysis of Clinical Experiments, Wiley.
4. Friedman L. M., Furburg C., Demets D. L. (1998). Fundamentals of Clinical Trials, Springer.
5. Jennison. C. and Turnbull B. W. (1999) Group Sequential Methods with Applications to Clinical Trials, CRC Press.
6. Marubeni .E. and Valsecchi M. G. (1994) Analyzing Survival Data from Clinical Trials and Observational Studies, Wiley.

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

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

Course Outcomes:

Students will be able to

1. learn and understand various designs of experiments.
2. design and carryout various experiments and analyze the data.
3. find and verify the consistent estimator and consistency and asymptotic normality (CAN) estimator.
4. understand concept of Likelihood Ratio Test (LRT), Wald's test, Score test, large sample test and asymptotic confidence interval.

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

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

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

Course Outcomes:

Students should be able to:

1. Forecast the trend pattern exhibited by the given data by using various methods
2. Run and interpret time series models for time series
3. Use the Box-Jenkins approach to model and forecast time series data empirically.
4. Apply ideas to real time series data and interpret outcomes of analyses.

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.	
Practical Based on Optional Paper		
	PSST235(A): Data Mining	**PSST235(B): Design and Analysis of Clinical Trials**
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	