Anekant Education Society's
Tuljaram Chaturchand College of Arts, Science and Commerce, Baramati
(Autonomous)

Course Structure for M.Sc. - II Semester- III (STATISTICS)

Semester	Paper Code	Title of the Paper	No. of Credits
III	STAT-5301	Asymptotic Inference	4
	STAT-5302	Design and Analysis of Experiments	4
	STAT-5303	Time Series Analysis	4
	STAT-5304 (A) STAT-5304 (B) STAT-5304 (C)	Data Mining Or Design and Analysis of Clinical Trials Or Optimization Techniques	4
	STAT-5305	Practical-V	4
	STAT-5306	Practical-VI	4

SYLLABUS (CBCS) FOR M.Sc.-II Sem.-III STATISTICS

(With effect from Academic Year 2020-2021)

Paper Code : STAT-5301
Paper : I Title of Paper : Asymptotic Inference
Credit : 4 credits No. of lectures : 60

A) Learning 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) Learning outcomes:

1. Students will be able to understand concept of Consistent estimator, CAN estimator, Likelihood Ratio Test (LRT), Wald's test, Score test, large sample test and asymptotic confidence interval.
2. Students will be able to find asymptotic distributions of moment estimators, percentile estimators, maximum likelihood estimator.
3. Students will be able to 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, Best asymptotically normal (BAN) estimator, 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 and its asymptotic distribution, 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, Barlett's Test for homogeneity of variances, Likelihood Ratio Test for Multinomials, variance stabilizing transformation and large sample tests, comparison of tests: asymptotic relative efficiency of a test, Consistency of Large Sample test.

(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.-II Sem.-III STATISTICS

(With effect from Academic Year 2020-2021)

Paper Code : STAT-5302
Paper : II Title of Paper : Design and Analysis of Experiments
Credit : 4 credits No. of lectures : 60

A) Learning 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) Learning outcomes:

1. Students will be able to understand basic principles and various terms of Design of Experiments.
2. Students will be able to apply Factorial design, fractional factorial design, confounding in real life problems.
3. Students should be able to analyze the data of various experimental design.

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. **(15L)**

Unit-2

Review of 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 2^p 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. **(12L)**

Unit-3

3^k design: contrasts for linear and quadratic effects, statistical analysis of 3^k design, confounding and fractional experiments in 3^k design, Response Surface Methodology (RSM): linear and quadratic model, stationary point, central composite designs (CCD), ridge systems, rotatability, multiple responses, blocking in RSM, Box-Behnken design. **(18L)**

Unit-4

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. **(15L)**

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, Englewood Cliffs, New Jersey.
6) Wu, C.F. Jeff and Hamada M. (2000). Experiments: Planning, Analysis and Parameter Design Optimization, John Wiley and Sons.
7) Bapat, R. B. (2012). Linear algebra and linear models. Springer Science & Business Media.

SYLLABUS (CBCS) FOR M.Sc.-II Sem.-III STATISTICS

(With effect from Academic Year 2020-2021)

Paper Code : STAT-5303
Paper : III Title of Paper : Time Series Analysis
Credit : 4 credits No. of lectures : 60

A) Learning 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) Learning outcomes:

1. Students can model time series data by various time series models like ARMA, ARIMA, SARIMA, ARCH and GARCH properties.
2. Students will be able to use the Box-Jenkins approach to model and forecast time series data empirically.
3. Students will be able to analyses time series data and use multivariate time series models such as vector auto regression (VAR).
4. Students will be able to 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. **(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) auto covariance and auto correlation functions and its properties stationary and invertibility. Introduction to spectral analysis of weakly stationary process. Periodogram and correlogram analysis. **(15L)**

Unit 3

Non-stationary: Unit root, non-stationary unit root test, Integrated ARMA (ARIMA) model, seasonal ARIMA (SARIMA) models, Yule-Walker estimation. Estimation of ARIMA models parameters, Maximum likelihood method for estimation, residual analysis and diagnostic checking. Filter and transfer function. Durbin-Levison algorithm, innovation algorithm, Box-Jenkins model (Without proof). **(15L)**

Unit 4

Graphical method for choosing AR and MA lags, FPE, AIC, BIC, residual analysis, conditional hetroschedastic models, volatility models, ARCH and GARCH properties, examples, estimation and forecasting. Multivariate Time series model, VAR models, vector ARMA models. Conintegration models, use of statistical software for time series analysis. **(15L)**

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.-II Sem.-III STATISTICS

(With effect from Academic Year 2020-2021)

Paper Code : STAT-5304 (A)
Paper : IV Title of Paper : Data Mining
Credit : 4 credits No. of lectures: 60

A) Learning 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) Learning outcomes:

1. Students can propose data-mining solutions for different applications.
2. Students would be able to evaluate different models used for data processing.

TOPICS/CONTENTS:

Unit- 1

Introduction to big data, supervised and unsupervised learning, review of linear discriminant analysis, logistic regression, clustering procedure- k means, hierarchical, principal component analysis. **(10 L)**

Unite -2

Bayes classifier, nearest neighbour classifier, Classification and Regression tree (CART): information gain, gain ratio, 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), 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. **(20 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.-II Sem.-III STATISTICS

(With effect from Academic Year 2020-2021)

Paper Code : STAT-5304(B)

Paper : IV Title of Paper : Design and Analysis of Clinical Trials

Credit : 4 credits No. of lectures : 60

A) Learning 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) Learning outcomes:

1. Understand the basic statistical principles, methods for clinical data analysis and reporting.
2. 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.-II Sem.-III STATISTICS

(With effect from Academic Year 2020-2021)

Paper Code : STAT-5304(C)

Paper : IV Title of Paper : Optimization Techniques

Credit : 4 credits No. of lectures : 60

A) Learning objectives:

1. The main objective is to understand the mathematical tools that are needed to solve optimization problems.
2. Students should be able to understand discrete event simulation and decision analysis with inclusion of modeling based on random events involving uncertainties.

B) Learning outcomes:

1. Students can identify and develop operational research models from the verbal description of the real system.
2. Students will be able to understand the characteristics of different types of decision-making environments and decision-making approaches.
3. Students can apply optimization techniques to take correct decision.

TOPICS/CONTENTS:

Unit-1

Linear Programming Problem: Review of simplex algorithm and simplex method, Existence of unbounded solution, optimality conditions, for other related theorems (statement only), artificial variable technique: Two phase and Big-M method, Revised simplex method.

(12L)

Unit 2

Duality in LPP: Concept of duality, duality theorems, dual simplex method, duality theory of testing optimality of solution in and transshipment problem. Sensitivity analysis: Bellman's optimality principle, changes affecting feasibility and optimality. **(12L)**

Unit 3

Integer linear programming problem (ILPP): The concept of cutting plane, Gomory's method of cutting plane for all ILPP and mixed ILPP, Branch and Bound method, Nonlinear programming: Kuhn-Tucker conditions of optimality, Quadratic programming, Wolfes, Beales methods, applications of ILPP. **(18L)**

Unit 4

Dynamic programming: Nature of dynamic programming, deterministic processes, Non-sequential discrete optimization allocation problems, sequential discrete optimization long-term planning problem, multi-stage production processes. Inventory model: Inventory model building, single items inventory control models without shortages, EOQ model with constant rate of demand, with different rate of demand and economic production quantity model when supply is gradual. **(18L)**

References:

1) Bertsekas, D. (1999). *Nonlinear Programming*, 2nd Edn. Athena Scientific.
2) Chong, E. K. P. and Zak, S. (2004). *An Introduction to Optimization*, Wiley.
3) Fletcher, R. (2000). *Practical Methods of Optimization*, Wiley.
4) Hadley, G. (1987). *Linear Programming*. Addison-Wesley.
5) Kambo, N.S. (1991). *Mathematical Programming Techniques*. Affiliated East-West press.
6) Panneerselvam, R. (2012). *Operations Research*, 2nd Edn. Prentice Hall of India.
7) Taha, H.A. (1992). *Operations Research*, 5th ed. Macmillan.

SYLLABUS (CBCS) FOR M.Sc.-I Sem.-II STATISTICS

(With effect from Academic Year 2020-2021)

Paper Code : STAT-5305

Paper : V Title of Paper : Practical-V

Credit : 4 credits No. of lectures : 60

Sr. No.	Title of Experiments
1.	Analysis of general linear model
2.	Balance Incomplete Block Design (Intra block analysis)
3.	Analysis of Covariance in one way and two-way model
4.	Analysis of 3^k factorial experiments
5.	Total Confounding in 3^k factorial experiment
6.	Partial Confounding in 3^k factorial experiment
7.	Analysis of 3^k fractional factorial experiment
8.	Fitting first and second order response surface model, central composite design contour, surface plots, canonical analysis of stationery points, Blocking in RSM
9.	Random effect model with one factor, estimation of variance
10.	Taguchi methods: S/N ratio, orthogonal arrays, triangular tables, linear graphs, inner and outer arrays
11.	Verification of consistency and asymptotic normality of the estimators
12.	Comparing Consistent estimator, MSE and sample size considerations
13.	Asymptotic Confidence Intervals
14.	Power functions of large sample test (LR, Wald, Rao)
15.	Analysis of three-dimensional contingency tables
16.	Plotting Likelihood function, MLE by methods of scoring

SYLLABUS (CBCS) FOR M.Sc.-I Sem.-II STATISTICS

(With effect from Academic Year 2020-2021)

Paper Code : STAT-5306

Paper : VI Title of Paper : Practical-VI

Credit : 4 credits No. of lectures : 60

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.
3.	Stationarity, Causality and invertibility of ARMA model
4.	Order selection in time series: use of ACF/PACF and AIC, BIC, fitting of AR, MA models (conditional least squares or maximum likelihood).
5.	Fitting of ARMA, ARIMA and SARIMA models (conditional least squares or maximum likelihood).
6.	Forecasting using fitted linear models (recursively), Holt -Winters forecasts construction of forecast intervals.
7.	Fitting heteroscedastic models: checking for heteroscedecity from residuals, ARCH, GARCH modeling.
8.	Mini project based on time series analysis paper (2 Practicals)
9.	Case study based on optional paper (2 Practicals)
10.	Research Paper Review (4 Practicals)