DEPARTMENT OF STATISTICS
SCHOOL OF MATHEMATICAL SCIENCES
KAVAYITRI BAHINABAI CHAUDHARI
NORTH MAHARASHTRA UNIVERSITY
JALGAON - 425001, INDIA

SYLLABUS STRUCTURE AND SYLLABUS AS PER NEP-2020

FOR

M.Sc. STATISTICS
(Second Year)
with specialization in Industrial Statistics

(Medium of Instructions: English)

WITH EFFECT FROM ACADEMIC YEAR
2024-2025

Syllabus Structure for M.Sc. (Statistics) Second Year
(With effect from Academic Year 2024-25)

Semester-III

Course Type	Course Code	Title of the Course	Credits	Contact hours/week			Distribution of Marks for Examination					
							Internal		External		Total	
				Th(L)	Pr	Total	Th	Pr	Th	Pr	Th	Pr
DSC-33 (Theory)	ST-511	Design, Planning and Analysis of Experiments	04	04	--	04	40	--	60	--	100	--
DSC-34 (Theory)	ST-512	Multivariate Analysis	04	04	--	04	40	--	60	--	100	--
DSC-35 (Theory)	ST-513	Analysis of Clinical Trials using SAS	02	02	--	04	20	--	30	--	50	--
DSC-36 (Practical)	ST-515	Practicals III	04	--	08	04	--	40	--	60	--	100
DSE-7 (Theory)	ST-514 (A)	Statistical Process Control and Reliability	04	04	--	04	40	--	60	--	100	--
	ST-514 (B)	Data Mining										
	ST-514 (C)	Asymptotic and Nonparametric Inference										
RP	ST-516	Research Project I	04	--	08	08	--	40	--	60	--	100
Total			**22**	**14**	**16**	**30**	**140**	**80**	**210**	**120**	**350**	**200**

Semester-IV

Course Type	Course Code	Title of the Course	Credits	Contact hours/week			Distribution of Marks for Examination					
							Internal		External		Total	
				Th(L)	Pr	Total	Th	Pr	Th	Pr	Th	Pr
DSC-37 (Theory)	ST-521	Stochastic Processes	04	04	--	04	40	--	60	--	100	--
DSC-38 (Theory)	ST-522	Time Series Analysis	04	04	--	04	40	--	60	--	100	--
DSC-39 (Practical)	ST-524	Technical Communications and Practicals-IV	04	--	08	08	--	40	--	60	--	100
DSE-8 (Theory)	ST-523 (A)	Actuarial Statistics	04	04	--	04	40	--	60	--	100	--
	ST-523 (B)	Investment Under Uncertainty										
	ST-523 (C)	Survival Analysis										
RP	ST-525	Research Project II	06	--	12	12	--	60	--	90	--	150
Total			**22**	**12**	**20**	**32**	**120**	**100**	**180**	**150**	**300**	**250**

Syllabus for M.Sc. (Statistics) Second Year Semester III

ST-511: DESIGN, PLANNING AND ANALYSIS OF EXPERIMENTS

Course Objectives (CObs):

- How to plan, design and conduct experiments efficiently and effectively by applying the basic principles of designs of experiments
- To study various characterizations, properties and analysis of designs such as CRD, RBD, LSD, BIBD, YSD PBIBD, factorial and fractional factorial designs etc.
- To study Response Surface models, Taguchi designs, orthogonal arrays and matrix experiments, use of signal-to-noise ratio in analysis of Taguchi designs, ANOMs etc.

- **Review of basic principles and simple designs (10L, 10M)**
  - Review of basic principles of design of experiments: Randomization, replication and local control with examples.
  - Review of one-way and two-way classification models
  - Two-way classification model with single observations per cell with interactions.
  - Two-way classification model with r-observations per cell ($r > 1$) with and without interactions.
  - Concept of Fixed effect models, Random effect models and Mixed effect models, random effect model for one way classification, variance component estimation.

- **Theory of general two-way block design (GTWBD) (12L, 12M)**
  - General two-way block designs, normal and reduced normal equations, estimability condition for the linear parametric function, properties of C and D-matrices, adjusted and unadjusted treatment and block sum of squares, ANOVA
  - Various characteristics of GTWBD: connectedness, balancedness and orthogonality, necessary and sufficient conditions, examples.
  - Concept of incomplete block design, Balanced Incomplete Block Design (BIBD), properties, PBIBD, association scheme, PBIBD with two associate classes, examples
  - LSD and Youden Square design.

- **Analysis of covariance (ANCOVA) (8L, 8M)**
  - The concept of concomitant variable/covariate and examples, analysis of covariance (ANCOVA) in a general Gauss-Markov model, applications and need of ANCOVA technique, estimability conditions, BLUE, variances of BLUE, adjusted sum of square for regression, testing the hypothesis of no covariate, analysis of covariance table.
  - Analysis of covariance for one-way and two-way classification model and their ANCOVA tables.

- **Analysis of factorial and fractional factorial designs (12L, 12M)**
  - 2^k-full factorial designs: diagrammatic representation of main effects and first order interactions in the model, estimation of main effects and interactions, their variance and testing significance, analysis of single as well as more than one replicates using ANOVA.
  - Confounding: total and partial confounding technique with examples, confounding of treatments in 2^k design in more than two blocks (2^p blocks with $k > p > 2$), analysis of designs with treatments confounded.
  - Two-level-fractional factorial experiments, half; quarter and general fractions of the design, aliases, generators of the design, complete defining relation, Resolution of a design (III, IV & V), minimum abbreviation designs, etc.

- Three-level factorial designs, models when the factors are physical/categorical/mixed type, diagrammatic representation of runs of the experiment, analysis of 3^2 and 3^3 design, contrasts for linear and quadratic effects, and ANOVA.

- **Response surface methodology (RSM) (10L, 10M)**
  - The concept and need of response surface models, first and second order response surface models, determination of stationary point, contour plots, canonical analysis, ridge systems, multiple responses, blocking in RSM.
  - Test of curvature for the two-level factorial/fractional factorial designs
  - The concept and definition of Central composite designs (CCD), runs, layout and geometric representation of CCD with two, three and k-factors, linear model for CCD and its analysis, Concept of rotatable designs, rotatability condition for CCD.
  - Plackett-Burman designs

- **Taguchi methods (8L, 8M)**
  - Quality, Taguchi's philosophy about quality, traditional quality loss function and a quadratic loss function, loss functions for different types of quality characteristics, S/N ratio as a functions of average quality loss (AQL)
  - The concept of orthogonal arrays and their role in Taguhi designs, standard orthogonal arrays and interactions tables, linear graphs to represent designs, ANOMs, inner and outer arrays, analysis of Taguchi designs.

REFERENCES

1. Kshirsagar A.M. (1983) Linear Models (Marcel Dekkar).
2. John P.W.M. (1971) Linear Models (John Wiley Ltd.)
3. Montogomery D.C. (2008) Design and Analysis of Experiments (John Wiley).
4. Ogawa J. (1974) Statistical Theory of the Analysis of Experimental Design (Marcel Dekker).
5. Phadke, M.S. (1989) Quality Engineering through Robust Design (Prentice Hall).
6. Kuehl R.O. (1994). Statistical Principles of Research Design and Analysis (Duxbury Press).
7. Nigam, A.K.., Puri, P.D.., Gupta, V.K. (1988). Characterizations and Analysis of BlockDesigns, (Wiley).
8. Morris M. (2010). Design of Experiments: An Introduction Based on Linear Models (Chapman & Hall/CRC Texts in Statistical Science).
9. Bauro of Indian Standards IS 10427-1 (1982). Designs for industrial experimentation, Part 1: Standard designs.
10. Bauro of Indian Standards IS 10427-2 (2006). Designs for industrial experimentation, Part 2: Orthogonal Arrays.

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST511.1	Understand how to use designed experiments to achieve breakthrough improvements in process efficiency and quality	2
ST511.2	Critically review basic concepts and models of experimental design.	4
ST511.3	Analyse the results of a designed experiment in order to conduct the appropriate statistical analysis of the data, interpret statistical results from an experiment and report them in non-technical language.	4

ST-512: MULTIVARIATE ANALYSIS

Course Objectives (CObs):

- To understand Multivariate Normal distribution (MVND), its properties, Sampling, Estimation and Hypothesis testing for MVND.
- To learn the sampling distributions, distributions of quadratic forms in MVND and other distributions emerging from MVND.
- To Develop a thorough understanding of multivariate data analysis tools such as, Principal component Analysis, Canonical correlations, Discriminant analysis, Cluster and Factor analysis, correspondence analysis and MANOVA.

- **Multivariate Normal Distribution (MVND): (14 Marks)**
  - Singular and nonsingular MVND, Mean vector and variance covariance matrix. (3L)
  - Characteristic function and mgf of MVND, Additivity property of MVND (1L)
  - Distribution of linear forms of a vector having MVND, Marginal distributions, Conditional distributions (3L)
  - Necessary and sufficient condition for independence of $\underline{X}^{(1)}$ and $\underline{X}^{(2)}$ (two components of $\underline{X}$). (1L)
  - Central and noncentral χ^2distribution, their characteristic function,χ-distribution. (2L)
  - Random sampling from MVND, Unbiased and maximum likelihood estimators of parameters of MVND, their sampling distributions, independence. (3L)
  - Sample correlation coefficients, their maximum likelihood estimators (mle), Correlation matrix and its mle. (1L)

- **Wishart distribution: (10 Marks)**
  - Wishart matrix, Derivation of Wishart distribution in canonical case and in general case, Bartelett Decomposition theorem. (3L)
  - Characteristic function of Wishart distribution, Additive property of Wishart distribution, Moments of Wishart distribution. (3L)
  - Properties of Wishart distribution. (4L)
  - Necessary and sufficient condition for XAX' to have Wishart distribution and its application. (2L)

- **Hotelling's T^2 and its applications: (8 Marks)**
  - Hotelling's T^2statistic as a generalization of square of Student's statistic. (1L)
  - Derivation of Hotelling's T^2statistic from Likelihood Ratio Test, Application of union-intersection principle to obtain Hotelling's T^2statistic, Invariance of T^2statistic under scale transformation. (2L)
  - Distance between two populations, Mahalnobis D^2 statistic and its relation with Hotelling's T^2statistic. (1L)
  - Application of Hotelling's T^2 :
    Test of equality of mean vector for one or more multivariate normal population, Test of equality of components of a mean vector of MVND, Two sample problem. (2L)

- **Discriminant Analysis: (5 Marks)**
  - Classification problem, Misclassification cost and probabilities, Classification Rule by minimizing the expected cost of misclassification. (2L)
  - Classification and discrimination procedure for discrimination between two multivariate normal populations, sample discriminant function, Probabilities of misclassification and their estimation, Optimum error rate, Test associated with discriminant function. (3L)

- **Cluster Analysis: (4 Marks)**
  - Cluster Analysis, Intra and inter cluster distances, similarity and association measures, concept of partitional clustering and Hierarchical Clustering,
  - Hierarchical Clustering methods: singe, complete and average linkage methods, nonhierarchical Clustering method: K-means algorithm, dendrogram for cluster representation. (4L)
- **Principal Component Analysis: (4 Marks)**
  - Introduction and need, population principal components, Finding i^{th} principal component, correlation of i^{th} principal component with k^{th}element of vector $\underline{X}$, principal component when Σ has special structure, Sample principal components. (4L)
- **Canonical Correlation: (4 Marks)**
  - Concept of Canonical correlation as generalization of multiple correlation, Geometrical interpretation and its use, Definition of canonical correlation and canonical variables, Existence of canonical variables, Canonical correlation as a maximum root of characteristic equation of a matrix, Sample canonical correlation and canonical variable. (4L)
- **Factor Analysis: (4 Marks)**
  - Introduction, concept of factor analysis with example, the orthogonal factor model methods of estimation, Factor rotation, factor scores. (4L)
- **Correspondence Analysis (CA): (4 Marks)**
  - Contingency tables, algebraic development of CA for two-way contingency tables, the concept of inertia and biplots, Construction of biplots and interpretation from it. (4L)
- **Multivariate Analysis of Variance (MANOVA): (3 Marks)**
  - MANOVA for one way and two ways classified data, Wilk's Λ criteria. (3L)

REFERENCES

1. Anderson, T.W. (1983). An Introduction to Multivariate Statistical Analysis (2nd Ed. Wiley).
2. Johnson, R.A. and Wichern D.W. (2002). Applied Multivariate Statistical Analysis, (Prentice-Hall)
3. Giri, N.C. (1977). Multivariate Statistical Inference, (Academic Press).
4. Kshirsagar, A.M. (1972). Multivariate Analysis, (Marcel Dekker).
5. Morrison, D.F. (1976). Multivariate Statistical Methods, (2ndEd.McGraw Hill).
6. Muirhead, R.J. (1982). Aspects of Multivariate Statistical Theory, (J.Wiley).
7. Rao,C.R. (2002). Linear Statistical Inference and its Applications, (2ndEd.Wiley).
8. Seber, G.A.F. (1984). Multivariate Observations, (Wiley).
9. Sharma, S. (1996). Applied Multivariate Techniques, (Wiley).
10. Srivastava, M.S.and Khatri, C.G. (1979). An Introduction to Multivariate Statistics. (North Holland).

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST512.1	Understand the link between multivariate techniques and corresponding univariate techniques	2
ST512.2	Apply multivariate techniques appropriately, undertake multivariate hypothesis tests and draw appropriate conclusions	3
ST512.3	Summarize and interpret multivariate data using appropriate multivariate methods to analyse data with statistical software.	5

ST-513: ANALYSIS OF CLINICAL TRIALS USING SAS

Course Objectives (CObs):

- The basic principles for design of randomized clinical trials and how they should be reported.
- Introduction to terminology used in clinical trials and the several common designs used for clinical trials, such as parallel and cross-over designs.
- Techniques of clinical trials, like randomization and blinding of treatment, Analysis and interpretation of the experiments.

- **Unit I: Introduction to SAS** **(3L, 3M)**
  Introduction to SAS: Data Structures, Statistical Testing, Statistical Analysis and Modelling

- **Unit II: Essentials of Clinical Trial Initiation:** **(7L, 7M)**
  Need and ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I-IV trials, multi-center trials. Data management: data definitions, case report forms, database design, data collection systems for good clinical practice. Bioavailability, pharmacokinetics and pharmacodynamics, two-compartment model, Formulation and Routes of Administration, Clinically Importance Differences.

- **Unit III: Strategic Design and Monitoring of Clinical Trials:** **(7L, 7M)**
  Parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, objectives and endpoints of clinical trials, design of Phase I trials, design of single-stage and multi-stage Phase II trials. Design and monitoring of Phase III trials with sequential stopping, design of bio-equivalence trials, Assessment of Bioequivalence, Decision Rules and Regulatory Aspect, Effect estimations and other Inferences for 2x2 crossover design: Classical methods of interval hypothesis testing for bioequivalence, Bayesian methods, nonparametric methods.

- **Unit IV: Estimands in clinical trials:** **(7L, 7M)**
  Power and sample size determination, multiplicative (or log-transformed) model, ML method of estimation, assessment of inter and intra subject variabilities, detection of outlying subjects. Optimal crossover designs: Balaam's design, two-sequence dual design. Optimal four period designs. Assessment of bioequivalence for more than two drugs, Williams design.

- **Unit V: Designs based on clinical endpoints:** **(6L, 6M)**
  Weighted least squares method, log-linear models, generalized estimating equations, drug interaction study, dose proportionality study, steady state analysis. Meta analysis, analysis of categorical data.

REFERENCES

1. Cody, Ron. An introduction to SAS university edition. SAS Institute, 2018.
2. Cody, Ron. Learning SAS by example: a programmer's guide. SAS Institute, 2018.
3. J.L.Fleiss (1989) The Design and Analysis of Clinical Experiments. Wiley & Sons.
4. E.Marubeni and M.G.Valsecchi (1994). Analyzing Survival Data from Clinical Trials and Observational studies, Wiley and Sons.
5. Shein-Chung Chow and Jen-pei Liu, Design and Analysis of Clinical trial, Wiley & Sons.
6. Shein-Chung Chow and Jen-pei Liu, Design And Analysis of Bioavailability and Bioequvalence Studies, Marcel Dekker,Inc

7. S.Paintadosi.(1997) Clinical Trials: A Methodologic Perspective. Wiley & Sons.
8. L.M.Friedman, C.Furburg, D.L.Demets (1998). Fundamentals of Clinical Trials, Springer Verlag.
9. E9-R1_Step4_Guideline_2019_1203.pdf (ich.org)

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST513.1	Reduce the bias and variability involved during conduction of Clinical trials.	3
ST513.2	Estimate the true therapeutic effect of the drug	4
ST513.3	Analyse the outcomes of clinical experiments through various statistical techniques	4

ST-514 (A): STATISTICAL PROCESS CONTROL AND RELIABILITY

Course Objectives (CObs):

- To describe the concept of variation and role of statistics in Quality Improvement through SPC tools.
- To make aware the students about the Industrial production processes and applications of SPC tools for monitoring those processes.
- To prepare students to carry out process capability analysis based on real life process data.
- To describe some mathematical and applied terminologies of reliability theory.

- Total Quality Management. (6L,6M)
  - Concept of Quality, Quality improvement, Quality philosophy.
  - Introduction of TQM, evaluation of Total Quality.
  - Some important TQM concepts.
  - TQM Gurus' Ideas.
  - Six Sigma Process
  - Quality systems, ISO 9000 and other Quality systems.

- Review of some Statistical methods useful in Quality Improvement. (2L,2M)
  - Concept of variation, systematic variation, random variation, stable industrial processes.
  - Describing variation through graphical and numerical methods.
  - Some important Discrete and continuous probability distributions useful in quality control and improvement.
  - Some useful approximations of Distributions.

- Statistical Process Control (SPC).
  - Introduction of SPC. (3L,3M)
    - Basic concept of process monitoring and control.
    - Seven tools of SPC.
    - General theory of Control charts.
    - Different types of limits, Specification limits, Natural tolerance limits, Control limits, Warning limits.
    - OC Curve and ARL of control charts.

- Control Charts for Attributes. (4L,4M)
  - Control chart for fraction nonconforming.
  - Control chart for fraction nonconformities (defects)
  - OC Curves for Attributes control charts.
- Control Charts for Variables. (7L,7M)
  - Statistical basis of the charts for variables.
  - $\overline{X}$, R, S, $\overline{X}$ *and* R, $\overline{X}$ *and* S, $\overline{X}$ *and* S^2 Control Charts.
  - Median chart and Midrange chart.
  - Control charts for Individual Measurements.
  - Special control charts: CUSUM, EWMA control charts.
- Process Capability Analysis. (6L,6M)
  - Capable process and Process capability.
  - Process Capability Analysis using Histogram or Probability plot.
  - Process Capability indices under normal distribution of quality characteristic.
    - Capability indices C_p , C_{pk} , C_{pm} .
    - Potential and Actual proportion of nonconforming.
    - Estimation, C.I. and tests of hypotheses relating to C_p .
  - Process Capability Analysis for non-normal data.
  - Gauge and Measurement system capability studies.
  - Setting specification limits on discrete components, linear and nonlinear combinations.
  - Estimating the Natural tolerance limits of a process.
- SPC for short production. (1L,1M)
- Modified, Acceptance control charts and Group control chart. (2L,2M)
- SPC with autocorrelated process data (2L,2M)
- Multivariate Quality control. (2L,2M)
- Engineering process control(EPC) and SPC (2L,2M)
- Acceptance Sampling. (8L,8M)
  - Single, double and sequential sampling plans for attributes and their properties.
  - Curtailed double sampling plans, operating characteristics functions and other properties of the sampling plans.
  - Sampling plans with rectification. OC, ASN, ATI, AOQ curves, AOQL, Designing of sampling plan. Dodge-Romig acceptance sampling plans.
  - Plan for inspection by variables for one-sided and two-sided specifications; AQL based sampling plans.
- Elements of Reliability: (15L,15M)

  Components and systems, binary coherent structure k-out-of–n: G structure, bridge structure. Cuts and Paths, minimal path sets and minimal cut sets. Reliability of coherent system, bonds on system reliability. Structural and reliability importance of components, Hazard function, distribution with DFR and IFR.

REFERENCES

1. Besterfield, D.H., Besterfield-Michana, c., Besterfield, G.H., Besterfield-Sacre, M. Total Quality Management; Pearson Education (Singapore) Pte. Ltd. India. 2nd Edition 2001.
2. Montgomery, D.C. (2009) Introduction to Statistical Quality Control; Wiley, 6th Edition.
3. Wadsworth H.M.; Stephens K.S. and Godfrey A.B. Modern Methods for Quality Control and Improvement ,2nd Ed. Wiley.

4. Wetherill, G.B. and Brown, D.W. Statistical Process Control, Theory and Practice; Chapman and Hall.
5. Mittag H.J. and Rinne H.(1993). Statistical Methods of Quality Assurance.
6. Barlow R.E. and Proschan F. (1985) Statistical Theory of Reliability and Life Testing; Holt, Rinehart and Winston.
7. Lawless J.F. (1982) Statistical Models and Methods of Life Time Data; John Wiley.

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST514A.1	Describe the concept of variation, Quality, TQM tools and terminologies useful as per standards of ISO.	2
ST514A.2	Describe and apply SPC tools useful for Quality control and Quality improvement.	2
ST514A.3	Apply different graphical tools useful in SPC and interpret their uses.	3
ST514A.4	Describe the role of Process Capability Analysis and its measures useful in SPC.	2
ST514A.5	Explain and Apply different terminologies useful in mathematical reliability based importance of components under different coherent structures.	3

ST-514 (B): DATA MINING

Course Objectives (CObs):

- To apply acquired knowledge for understanding data, data pre-processing and data quality.
- Understand various supervised and unsupervised learning methods and apply for data different sets.
- Apply suitable methods for data analysis and algorithms for data mining.

Unit I: **(12L, 12M)**

Concepts of Supervised and unsupervised learning, inference vs prediction, parametric methods and non-parametric methods, reducible and irreducible error, trade-off between prediction accuracy and model interpretability, train error vs test error, regression vs classification problem, assessing model accuracy, bias-variance trade-off, K-nearest neighbourhood algorithm for classification and regression.

Unit II: **(12L, 12M)**

Supervised learning: overview of classification problems, review on logistic regression, Bayes classifier, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), K nearest neighbors, Naïve Bayes, comparison of classification methods, classifier performance via confusion matrix and related measures, ROC curve.

Unit III: **(12L, 12M)**

Resampling methods, cross-validation approaches, bootstrap, tree-based methods, classification and regression trees, tree-pruning, Gini-index, Cross-entropy, Ensemble methods, bagging, Random forests, boosting, algorithms like Adaboost, XGBoost etc.

Unit IV: **(12L, 12M)**

Hyperplane, separating hyperplane, maximal margin classifier, support vector classifiers, support vector machines (SVM), SVMs with more than two classes, Neural network (NN).

Unit V: **(12L, 12M)**

Unsupervised learning: principal component analysis, Clustering procedures: k-means, hierarchical, k- medoids, self-organizing maps, association rules, feature selection and feature extraction, missing data imputation techniques.

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. James G., Witten, D., Hastie, T. Tibshirani, R. (2013). An Introduction to Statistical Learning: With Applications in R. Springer.
6. James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). An Introduction to statistical learning: With applications in Python. Springer.
7. Mitchell Tom, (1997). Machine Learning. McGraw-Hill.
8. Ripley, B.D. (1996). Pattern Recognition and Neural Networks. (Cambridg University Press).

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST514B.1	Analyze data mining problems and reason about the most appropriate methods to apply to a given dataset and knowledge extraction need.	4
ST514B.2	Implement basic pre-processing, association mining, classification and clustering algorithms.	3
ST514B.3	Apply and reflect on advanced pre-processing, association mining, classification and clustering algorithm.	3

ST-514 (C): ASYMPTOTIC AND NONPARAMETRIC INFERENCE

Course Objectives (CObs):

- To study important features of large sample theory.
- To construct likelihood ratio test (LRT) for testing of hypothesis, obtaining asymptotic confidence interval (ACI) of parameter.
- To study different nonparametric tests.

- Review of convergence in probability and convergence in distribution, Cramer and Slutsky's Theorems. (2L,2M)
- Consistent Estimation of real and vector parameter. Invariance of Consistent estimator under continuous transformation. (3L,3M)
- Consistency of estimators by method of moments, and method of percentiles, Mean squared error criterion. (4L,4M)
- Asymptotic relative efficiency, Error probabilities and their rates of convergence, Minimum sample sizes required to attain given level of accuracy. (4L,4M)
- Consistent Asymptotic Normal (CAN) estimator, Invariance of CAN estimator under differentiable transformation. (4L,4M)
- CAN property of estimators obtained by moments and percentiles. (3L,3M)
- CAN estimators obtained by moment and MLE method in one parameter exponential family, Extension of multi-parameter exponential family. (3L,3M)
- Examples of consistent but not asymptotically normal estimators. (2L,2M)
- Method of maximum likelihood, CAN estimators for one-parameter Cramer family, Cramer-Huzurbazar theorem, Solution of likelihood equations, Method of scoring, Newton-Raphson and other iterative procedures. (6L,6M)
- Fisher Lower Bound to asymptotic variance, extension to multi-parameter cases (without proof.) Multinomial distribution with cell probabilities depending on a parameter. (3L,3M)
- MLE in Pitman Family and Double Exponential distribution, MLE in censored and truncated distribution. (3L,3M)
- Likelihood Ratio Test (LRT), asymptotic distribution of LRT statistic, Wald Test, Rao's score test, Pearson χ^2 test for Goodness of fit, Barlett's Test for homogeneity of variances. Large Sample Tests and confidence intervals based on CAN estimators, Variance stabilizing transformation and large sample tests. Consistency of Large Sample Tests. Asymptotic power of large sample tests. (8L,8M)
- Nonparametric Statistical Inference. (15L,15M)
  - Introduction to Nonparametric Inference.
  - U-Statistics.
  - Some Single-Sample problems.
  - Some Two-Sample problems.
  - Test of Independence.
  - Some Applications of Order Statistics.

REFERENCES

1. Das Gupta, A. (2008). Asymptotic Theory of Statistics & Probability, Springer, New York
2. Deshmukh S. R. and Kulkarni M. G. (2021). Asymptotic Statistical Inference - A Basic Course Using R, Springer
3. Dudewicz, E.J. and Mishra, S.N.(1988). Modern Mathematical Statistics. Wiley
4. Ferguson, T.S. (1996). A course on Large Sample Theory. Chapman and Hall, London.
5. Gibbons, J.D. (1985). Nonparametric Statistical Inference, 2nd ed., Marcel Dekkar, Inc.

6. Kale, B.K. and K. Muraridharan (2015). Parametric Inference: An Introduction, Alpha Sci. Intl Ltd.
7. Lehmann, E.L. (1986). Testing Statistical Hypotheses (Student Edition).
8. Lehmann, E. L. (1999). Elements of Large Sample Theory, Springer.
9. Rao, C.R. (1973). Linear Statistical Inference, Wiley.
10. Rohatgi V.K. and Ehsanes Saleh A. K. MD. (2003). An Introduction to Probability and Statistics (Wiley Eastern, 2nd Ed.).
11. S. Biswas and G. L. Sriwastav (2011) Mathematical Statistics: A Text Book, Narosa Pub.

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST514C.1	Apply large sample properties of an estimator such as consistency, CAN estimator and different methods to construct such estimators.	3
ST514C.2	Understand large sample properties of MLEs.	2
ST514C.3	Construct likelihood ratio test (LRT) for testing of hypothesis with different examples, obtain asymptotic confidence interval (ACI) of a parameter.	3
ST514C.4	Understand knowledge about non-parametric method and some important non-parametric tests.	2

ST-515: PRACTICALS III

(By using statistical software and/or Computer programming language)

Course Objectives (CObs):

- Analyse different designed experiments' data, interpret the results through AMONs, interaction plots, contours and response surface plots, test specific hypothesis of interest etc.
- to deal and process the multivariate data, use R and Matlab to perform exploratory analysis of such a data, apply the multivariate statistical techniques to solve the objective specific problems.
- to apply various data mining techniques to data and draw conclusions.
- to demonstrate CANness of estimators, obtaining MLE by numerical methods, construct ACI, analysing data by appropriate non-parametric tests.
- to teach SPC tools through statistical software.

A. Practicals based on Design, Planning and Analysis of Experiments (30 Hrs, 15M)

1. Estimation of parameters, testing various hypothesis and analysis of variance for the following linear models:
    - Two-way classification model with r observations per cell with/without interaction.
    - Two-way classification model with unequal observations per cell
    - Random effect model for one-way classification model
2. Estimation of parameters, testing various hypothesis and analysis of covariance for the following linear models:
    - One-way classification model with one or more than one concomitant variable.
    - Two-way classification model with one or more than one concomitant variable.
3. Analysis of BIBD
4. Identification/verification of various properties (balancedness/connectedness, orthogonality) of the given design.

5. Generation and analysis of two-level factorial designs, main effect and interaction plots.
6. Analyzing two-level factorial designs with
   (i) Total confounding (ii) Partial confounding (iii) single replicate
7. Generation and analysis of two-level fractional factorial designs.
8. Analysis of 3^2 factorial design using response surface model (RSM), main effect and interaction plots.
9. Generating CCD and analysis of CCD with RSM, contour plots, response surface plots, calculation of stationary point and optimum response.
10. Generation and analysis of Taguchi orthogonal array designs.

B. Practicals based on Multivariate Analysis (30 Hrs, 15M)

1. Model sampling from multivariate normal distribution (including conditional distribution)
2. Estimation of $\underline{\mu}$, Σ-matrix, correlation coefficient
3. Applications of Hotelling's T^2.
4. Discriminant Analysis and Classification problem.
5. Principal components.
6. Canonical Correlation Analysis
7. Cluster Analysis
8. Correspondence Analysis
9. MANOVA

C. Practicals based on Analysis of Clinical Trials using SAS (30 Hrs, 15M)

1. Creating Data Structures in SAS
2. Managing and formatting Data in SAS
3. Exporting Data and Generating Reports from SAS
4. Relation between sample size and power of the test.
5. Randomization Methods in CTs.
6. Statistical Analysis for Parallel Designs in SAS.
7. Statistical Analysis for Standard 2x2 Crossover Designs.
8. Analysis of continuous data based on repeated measures under CTs.
9. Analysis of Categorical Data.
10. Outlier Detection in CTs.
11. Estimation of Pharmacokinetic parameters using SAS
12. Statistical Analysis and Modelling

D. Practicals based on SPC and Reliability (30 Hrs, 15M)

1. Graphical tools used in SPC with their interpretations.
2. Identification of probability distribution of quality characteristics.
3. Plotting and interpretation of Control chart for attribute.
4. Plotting and interpretation of Control chart for variable.
5. Plotting Multivariate Control chart.
6. Process capability analysis for normal data.
7. Process capability analysis for non-normal data.
8. Gauge capability analysis.
9. Control charts for Short Production Runs
10. Single and double sampling plans for attributes: plotting OC, ASN, ATI, AOQ curves, finding AOQL.
11. Single sampling plan for variables.
12. Calculation and/or estimation of system reliability in parallel, series and k-out-of-n structures.
13. Estimation of warranty period.

OR

D. Practicals based on Data Mining (30 Hrs, 15M)

1. Introduction to key libraries, packages, data sets from R/Python related to data mining.
2. Exploratory data analysis
3. Classification techniques: Logistic Regression, LDA, QDA, Naïve bayes, KNN
4. Resampling Methods: Cross-validation, bootstrap
5. Decision tree and Ensemble methods: Classification trees, regression trees, bagging, boosting, random forest
6. Support vector machines (SVM): Support vector classifier, SVM for binary classification, SVM with multiple classes
7. Unsupervised learning methods: Principal component analysis, clustering: K means clustering, Hierarchical clustering.
8. Association rule mining.

OR

D. Practicals based on Asymptotic and Nonparametric Inference (30 Hrs, 15M)

1. Demonstrating consistency and CANness of consistent estimators.
2. Demonstration of consistency and asymptotic non-normality of estimator that is consistent but not CAN.
3. Computation of moment estimators and demonstration of their asymptotic distributions.
4. Verification of invariance property of consistent and CAN estimators under continuous transformation.
5. Generating consistent estimators by method of percentile.
6. MLE by methods of scoring.
7. Comparison of consistent estimator
   On the basis of their MSE's of different estimators.
   On the basis of requirement of minimum sample size.
8. ACI, Testing of hypothesis by likelihood ratio tests, computation and plotting of power function of test.
9. Fitting of distributions to sample data using following tests:
   Chi-square goodness of fit test
   Kolmogorov Smirnov goodness of fit test
   Lilliefor's goodness of fit test
10. Practical on one sample location problem (i) Sign test (ii) Wilcoxon Signed rank test
11. Practical on two sample location problem
    (i) Wilcoxon Rank sum test (ii) Mann Whitney U test
12. Practical on k sample Kruskal Wallis test, Friedman test.

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST515.1	Process and analyse the multivariate data, interpret the results etc.	4
ST515.2	Solve objective specific problems through analysis of multivariate data with continuous response and qualitative and/or quantitative predictors.	3
ST515.3	Analyse the outcomes of clinical experiments through various statistical software.	4

ST515.4	Apply different SPC tools to analyse Industrial production data. OR Fit a suitable data mining techniques to the data and draw conclusions. OR Demonstrate consistency and asymptotic normality of estimators, for testing of hypothesis use of appropriate non-parametric tests using R software	4 3 2

ST-516: RESEARCH PROJECT I

Course Objectives (CObs):

- The intention of the project work is to make students develop deeper knowledge and understanding in the context of a specific topic.
- The overall objective is to enable students think independently on a specific topic, design and perform a set of experiments to display the knowledge and capability required for independent work.
- It also aims at managing time effectively while working independently, appropriate referencing, analyzing the results and develop skills in report writing.
- Able to design research problem and solve it using a range of different data sources, design an intervention and evaluation.

- Project duration (Tentative Dates): 15th July to 15th November. Students are supposed to submit final project report in typed form with binding by November 20.
- General Guidelines:
  - Project work is considered as a special course involving application of knowledge in solving/applying/exploring a real-life situation/difficult problem. The student undertake research in specific area of his Major/Core subject with an advisory support by a teacher/faculty member.
  - Full flexibility is given to the student in identifying the project depending on the resources and infrastructure available in the Department/host organization. Students are encouraged to work on multidisciplinary projects but it is not mandatory.
  - Students should work in a group of 2-3 together, however, in any case, not more than four students should be involved in the same project work.
- Project Guide: Teachers from the Department of Statistics and/or personnel from organization where student is going to visit for field work or training related to project. Each project group will be guided by concerned teacher (guide) for one hour per week throughout the semester.
- Project Topic: Students in consultation with the guide will decide Project Topic/Area. Each research group should have different research topics with some kind of possible level of novelty.
- Project evaluation: Project evaluation will be based on
  - Internal Evaluation by Guide (40 Marks)
  - External Evaluation by two examiners (60 Marks)
    - Project report (20 Marks)
    - Presentation by student or group of students (20 Marks)
    - Viva voce (20 marks)

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST516.1	Understand the real-world problems through statistical angle.	2
ST516.2	Design and execute experiments/surveys independently.	6
ST516.3	Analyze, co-relate and interpret the data and build up skills and temperament of scientific writing.	4

Syllabus for M.Sc. (Statistics) Second Year Semester IV

ST-521: STOCHASTIC PROCESSES

Course Objectives (CObs):

- To study of the basic concepts of the theory of stochastic processes
- Study of various properties and characteristics of stochastic processes
- Study of the methods for describing and analysing complex stochastic models.

- **Introduction: Stochastic Processes, Markov chains (14 Marks)**
  - Introduction to Stochastic Processes, Classification of Stochastic Processes according to state space and time domain. (1L)
  - Finite and countable state space Markov chains (Definitions and examples). (2L)
  - Chapman-Kolmogorov equations, Calculation of n-step transition probability and its limit. (2L)
  - Stationary distribution of Markov chains. (2L)
  - Classification of states, Period of the state, Transient and recurrent Markov chain and related results. (4L)
  - Random walk and gambler's ruin problem. (2L)
  - First passage time and other problems with applications. (2L)
  - Applications of Markov Chains in Social, Biological and Physical Sciences. (1L)
- **Branching Process: (6 Marks)**
  - Golton-Watson branching process. (2L)
  - Probability of ultimate extinction, Distribution of population size. (2L)
  - Applications. (1L)
- **Discrete state space continuous time Markov Chain. (12 Marks)**
  - Definition and examples. (1L)
  - Markov Pure jumps processes. (2L)
  - Kolmogorov's differential equations. (2L)
  - Poisson process (Definitions, properties and applications). (3L)
  - Birth and death processes, Machine repairmen problem. (2L)
  - Wiener process as a limit of random walk, Brownian motion, Gaussian Processes, Ornstein-Uhlenbeck process, geometric Brownian motion (4L)
- **Simple Queuing Systems: (6 Marks)**
  - $M/M/1, M/M/\infty, M/M/S$ queuing systems and their applications. (2L)
  - Stationary solution for these systems using birth & death process approach. (2L)
- **Renewal Theory: (10 Marks)**
  - Renewal process (Definition and examples) (2L)
  - Elementary renewal theorem and its applications (2L)
  - Statement and uses of key renewal theorem (1L)
  - Renewal reward process, Regenerative Process, Semi-Markov process. (2L)
  - Age of renewal process and residual life in renewal processes and their distributions. (2L)
- **MCMC Algorithm:(3 Marks)** (3L)
- **Inference in Markov chains: (9 Marks)**
  - Estimation of transition probabilities, estimation of functions of transition probabilities in Markov chains, Testing of order of a Markov chains, Parametric models and their goodness of fit. (9L)

REFERENCES

1. Adke, S.R. and Manjunath, S.M. (1984). An Introduction to Finite Markov Processes, (Wiley Eastern).
2. Bhat, B.R. (2000). Stochastic Models: Analysis and Applications, (2nd New Age Int., India).
3. Cinlar, E. (1975). Introduction to Stochastic Processes, (Prentice Hall).
4. Feller, W.(1968). Introduction to Probability and its Applications, (Vol.1, Wiley Eastern).
5. Hoel, P.G., Port, S.C.and Stone, C.J.(1972). Introduction to Stochastic Processes, (Houghton Miffin& Co).
6. Jagers, P. (1974). Branching Processes with Biological Applications, (Wiley).
7. Karlin, S.and Taylor, H.M. (1975). A First Course in Stochastic Processes, (Vol.1, Academic Press).
8. Medhi, J. (1994). Stochastic Processes, (2nd Ed. New Age Publisher) .
9. Parzen, E. (1962). Stochastic Processes, (Holden-Day).
10. Ross, S. (2005). Introduction to Probability Models, (6th Ed. Academic Press).
11. Taylor and Karlin (1984). An Introduction to Stochastic Modeling, (Academic Press).

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST521.1	Clarify the power of stochastic processes and their range of applications;	2
ST521.2	Demonstrate essential stochastic modelling tools including Markov chains and queuing theory;	2
ST521.3	Formulate and solve problems which involve setting up stochastic models	6

ST-522: TIME SERIES ANALYSIS

Course Objectives (CObs):

- To Understand the fundamental advantage and necessity of forecasting in various situations, summarize and carry out exploratory and descriptive analysis of time series data, classical decomposition model, different smoothing methods and their advantages and limitations over each other, accuracy measures etc.
- Explain the concepts and basic properties of AR, MA, ARMA and ARIMA models, choosing an appropriate forecasting method in a particular environment of stationarity/non-stationarity.
- Analyzing the frequency domain -Sampling Periodograms, Spectral Density, Identifying the important periodic components of a series, ARCH Models for changing variation and periods of volatility in a series.

- **Introduction: (12L, 12M)**
  - Time series as a discrete parameter stochastic process, objectives of time series analysis (TSA), auto covariance and autocorrelation functions (ACF) and their properties, sample autocovariance function.
  - Exploratory time series analysis, classical decomposition model, tests for trend and seasonally. exponential and moving Average Smoothing. The concept of linear filter,

Holt and Winters smoothing, forecasting based on smoothing, adaptive smoothing, trend/seasonal component elimination using difference operators.
- Testing estimated noise sequence: Portmanteau tests, turning point test, the difference sign test, the rank test, transformation to obtain Gaussian series.

- **Stationary processes: (12L, 12M)**
  - Stationary, strict stationary processes, examples, autocovariance functions and its properties, autocorrelation matrix, linear process and its properties
  - Estimation of mean, auto covariance and autocorrelation functions under large sample theory.
  - $MA(q)$, $AR(p)$ and $ARMA(p,q)$ processes, stationarity condition, calculation of ACVF and ACF; causal and invertible processes; causality and invertibility conditions for $MA(q)$, $AR(p)$ and $ARMA(p,q)$ processes.
  - Estimation of AR and MA parameters, Yule-Walker estimation of AR-parameters

- **Forecasting stationary processes: (12L, 12M)**
  - Best linear prediction operator and its properties, estimation of missing values using prediction operator, one-step and h-step prediction for AR and MA processes, examples.
  - Partial autocorrelation function (PACF), PACF for AR and MA processes, use of ACF and PACF for the choice of AR and MA orders
  - Discussion of Durbin-Levinson algorithm and the innovations algorithms, differences between the two algorithms,
  - Order selection problem for ARMA process, FPE; AIC; AICC and BIC criteria for order selection, Forecasting ARMA processes, Residual analysis and diagnostic checking.

- **Frequency Domain Analysis: (10L, 10M)**
  - The idea of spectral analysis, spectral density function (SDF) and its characterizations, one-one correspondence between ACVF and SDF, SDF of filtered stationary processes, SDF for WN, AR, MA and ARMA processes,
  - Spectral estimation: Sinusoidal model for spectral analysis, time domain verses frequency domain analysis, estimation of parameters, periodogram and correlogram analyses, intensities, spectrum, properties of periodogram.

- **Nonstationary and seasonal time series models: (8L, 8M)**
  - The concept of non-stationarity, role of difference operator to eliminate non-stationarity, Integrated ARMA (ARIMA) models, their ACFs, examples, Unit-root in non-stationarity, Unit-root tests, Augmented Dickey-Fuller Test for AR(1) and AR(p) models, Forecasting in ARIMA models, Estimation of ARIMA model parameters.
  - Seasonal non-stationary models: Seasonal behavior of a time series, assumptions about seasonal component, Seasonal ARIMA (SARIMA) models, seasonal and non-seasonal differencing, examples of SARIMA models and their ACFs, demonstrations through real-life examples.

- **Volatility models (6L, 6M)**
  - The concept of volatility and need to model it, introduction to Conditional Heteroscedastic Models.
  - Conditional and unconditional mean; variances; stationarity condition and ACF for ARCH(1) model, ARCH(p) model, GARCH(p, q) model, stationarity condition for CH(1,1) model, properties, examples

REFERENCES

1. Box, G.E.P and Jenkins G.M. (1994). Time Series Analysis-Forecasting and Control, Holden-day San Francisco.
2. Anderson. T.W. (1971). The Statistical Analysis of Time Series Wiley, N.Y.
3. Montgemory, D.C. Johnson L.A (1990) Forecasting and Time Series Analysis, McGraw Hill.
4. Kendall, Sir Maurice and Ord, J.K. (1990). Time Series (Third Edition), Edward Arnold.
5. Brockwell, P.J. and Davis R.A. (2006) Time Series: Theory and Methods (Second Edition) Springer-Verlag.
6. Fuller, W.A (1996). Introduction to Statistical Time series, John Wiley N.Y.
7. Granger, C.W.J. and Newbold (1984) Forecasting Economic Time Series, Third Edition, Academic Press.
8. Kendall, M.G. and Stuart A. (1966). The Advanced Theory of Statistics, Volume 3, Charles Griffing, London.
9. Koopmans, L.H. (1974), The Spectral Analysis of Time series, Academic Press.

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST522.1	Outline the processes of identification, estimation and diagnosis of a time series, the criteria for choosing between models and the diagnostic tests that might be applied to the residuals of a time series after estimation	2
ST522.2	Choose appropriate statistical model for modelling the time-series data, assessment of the suitability of the model, use a model for forecasting and determining prediction intervals for forecasts, be aware of limitations and possible sources of errors in the analysis	3
ST522.3	Identify the important periodic components of a series through frequency domain analysis	2

ST-523(A): ACTUARIAL STATISTICS

Course Objectives (CObs):

- To make student aware about Indian Insurance business.
- To describe the concept of risk and measurement risk models.
- To describe some basic concepts of Actuarial Mathematics.
- To explain the role of statistical theory while addressing the problem of finding premium under different life insurance contracts.

- Introduction to Insurance Business. (3L,3M)
- Insurance and utility theory. (4L,4M)
- Risk models for Insurance. (5L,5M)
  - Individual and aggregate Risk models for short term.
  - Distribution of aggregate claims, compound Poisson distribution and its applications. Distribution of aggregate claims, compound Poisson distribution and its applications.
- Survival function and Life tables. (10L,10M)
  - Survival function, Distribution function, Density functions and Force of mortality. Time-until-death random variable and Curtate-future lifetime random variable.
  - Life tables, Select and ultimate life tables.
  - Assumptions for fractional ages and some analytical laws of mortality.

- Life Insurance. (10L,10M)
  - Principles of compound interest: Nominal and effective rates of interest and force of interest and discount, compound interest, accumulation factor, continuous compounding.
  - Insurance payable at the moment of death and at the end of the year of death, level benefit insurance, Whole life insurance, endowment insurance, deferred insurance and varying benefit insurance.
  - Recursion equations and commutation functions.
- Annuities. (10L,10M)
  Annuities certain, Continuous and Discrete life annuities. Life annuities with m-thly payments and approtionable annuities. Recursion equations.
- Net premium. (8L,8M)
  - Fully continuous and discrete premiums.
  - True m-thly payment premiums, apportionable premiums and accumulation type benefits. Insurance model including expenses.
- Reserve. (6L,6M)
  Prospective and retrospective reserve. Fully continuous and discrete net premiums reserves. Reserves on a semicontinuous basis and true m-thly premiums. Reserves on an apportionable or discounted continuous basis. Recursive formulates and differential equations for reserves commutation functions.
- Multiple life functions. (4L,4M)
  Joint life and last survivor status, insurance and annuity benefits through multiple life functions evaluation for special mortality laws.

REFERENCE

1. Robin Cunningham, Thomas N. Herzog, Richard L. Models for Quantifying Risk, 4th Edition, ACTEX Publications, 2011.
2. Browers, Newton L et al., Actuarial Mathematics 2nd . Society of Actuaries, 1997.
3. Dickson, David C. M., Hardy, Mary R. and Waters, Howard R., Actuarial Mathematics for life contingent risks, international series on actuarial science, Cambridge 2009.
4. Deshmukh S. R., An Introduction to Actuarial Statistics, University Press, 2009
5. Narang, Uma, Insurance Industry in India: Features, Reforms and Outlook, New Century Publications

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST523A.1	Identify basic risk available in the problem and formulate suitable loss random variable.	2
ST523A.2	Summarize different terms of life tables and their applications in Life Insurance.	2
ST523A.3	Apply tools of Financial Mathematics to solve the real-life problems related to investment, loan repayment and life annuities.	3
ST523A.4	Simulate data from present value random variables from different life insurances or life annuities to estimate required premium.	3
ST523A.5	Formulate new need-based annuity and life insurance contract.	6

ST-523(B): INVESTMENTS UNDER UNCERTAINTY

Course Objectives (CObs):

- To make student aware about investment and investment market.
- To describe the concept of risk in investments and investment market.
- To describe concepts of portfolio management.
- To explain the role of statistical theory in investments under uncertainty.

- Main theme: Risk - Return Trade off. (4L, 4M)
- Money market, Fixed income, equity, stocks and bonds, Treasury notes, Market indexes, Rates of interest, compound interest, inflation, Risk in a portfolio context, law of one price and arbitrage. (12L, 12M)
- Risk and risk aversion, mean variance analysis, allocation between risky and risk-free portfolios. (8L, 8M)
- Diversification and portfolio risk, Markovitz portfolio selection, optimal portfolios. (8L, 6M)
- Capital assets, pricing model, passive strategy, risk premium, Index models and diversification, CAPM and index model. (8L, 8M)
- Options markets, American and European options, call and put options, open strategies, option like instruments, option valuation. Binomial option pricing, Black-Scholes option valuation, uses of Black-Scholes formula. Futures markets, Mechanics and strategies, Futures prices, expected spot prices. (20L, 20M)

REFERENCES

1. Bodie, Z., Kane, A., and Marcus, A.J. (1996), Investments 4th Edition, Irwin. Chapters: 1, 2, 4, 5, 6, 7, 8, 9, 10, 20, 21, 22.
2. Arrow, K. J. (1971), Essays in the Theory of Risk Bearing, North Holland.
3. Hull John C. (1993), Options, Futures and Other Derivative Securities. 2nd Ed. Prentice Hall

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST523B.1	Describe basic concepts of investment and investment market	2
ST523B.2	Identify risk factors in investment	2
ST523B.3	Apply Actuarial and Statistical theory to solve the real-life problems related to investment.	3
ST523B.4	Explain Options markets and its applications	3
ST523B.5	Formulate need-based investment portfolios/plans	6

ST-523 (C): SURVIVAL ANALYSIS

Course Objectives (CObs):

- To learn the different statistical distributions of survival models
- To understand censored data and non-parametric techniques for survival estimation
- To understand semi-parametric regression set up (Cox PH models)

Unit I: **(15L, 15M)**

Survival data, meaning of censoring, concepts of time, order and random censoring (left and right), survival function, density function, hazard function (rate), cumulative hazard rate, mean residual life function, percentile residual life function, Equilibrium distribution function. Exponential distribution and it's no ageing properties: Lack of memory property, constant failure rate, Cauchy-function equation, constant mean residual life function, TTT transform, identity function as a TTT transform.

Ageing classes - IFR, IFRA, NBU, NBUE, DMRL, HNBUE and their duals, and inter relationship among these classes. Bathtub Failure rate, IFRA closure property, bound on reliability function of an IFRA distribution.

Unit II: **(15L, 15M)**

Life-time distributions – Exponential, Gamma, Weibull, Lognormal, Pareto, linear Failure rate, Makeham family, Lehman families (proportional hazard rate family), spacing, normalized spacing and results of an exponential distribution based on normalized spacing.

Parametric inference for complete data:

a) Exponential distribution: Point estimation of parameter of exponential distribution and Fisher information, exact and asymptotic Confidence Intervals for λ, obtaining minimal sufficient and consistent estimator of λ, Graphical method for checking exponentiality of data.
b) Weibull: Obtaining MLE of scale and shape parameter of Weibull distribution and sample information matrix.
c) Gamma: Obtaining MLE of scale and shape parameter of Gamma distribution and sample information matrix.
d) Lognormal: Obtaining MLE of parameter μ and σ, Confidence Interval for μ and σ

Unit III: **(15L, 15M)**

Parametric inference for censored data:

a) Type I censoring: Exponential distribution
b) Type II censoring: Exponential, gamma, Lognormal
c) Random censoring: Exponential, Lehman family, Weibull distribution,

Non-Parametric estimation of survival Function

a) For complete data: Non-parametric estimator of distribution function and survival function, distribution of empirical survival function, confidence band for survival function (by Using Kolmogorov - Smirnov statistics)
b) For censored data: Kaplan-Meier estimator, Actuarial estimator of survival Function, Estimator of variance of actuarial estimator (Greenwoods formula), product limit estimator and its variance, redistribution to right algorithm.

Unit IV: **(15L, 15M)**

Test for Exponentiality: Estimable function of degree r, Kernel, symmetric Kernel, U-statistic, variance of U- Statistic, one sample U-Statistic theorem, Hollander and Proschan Test, Test for exponentiality against positive ageing based n sample spacing, Analytical test for exponentiality against NBUE, Deshpande's Test, Two sample U statistic theorem, Wilcoxon and Mann –Whitney test, Gehan's test, Mantel- Haenzel test, Log rank test

Semi-parametric regression for failure rate, Cox's proportional hazards models, estimation and inference methods for the Cox models.

REFERENCES

1. Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data, Chapman and Hall, New York.
2. Deshpande, J. V., & Purohit, S. G. (2015). Lifetime Data: Statistical Models and Methods (Vol. 16). World Scientific Publishing Company.
3. Elandt - Johnson, R.E., Johnson N.L. (1980). Survival models and Data Analysis, John Wiley and Sons
4. Gross A.J. and Clark, V. A. (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences, John Wiley and Sons.
5. Miller, R.G. (1981) Survival Analysis (Wiley)

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST523C.1	Understand various statistical methods and Data Analysis techniques for Survival models.	2
ST523C.2	Apply parametric and non-parametric methods for estimation of survival Function.	3
ST523C.3	Estimate parameters when lifetime data available for censored observations.	5

ST-524: TECHNICAL COMMUNICATIONS AND PRACTICALS-IV

Course Objectives (CObs):

- To make students aware of some soft skills for technical communication and preparation for interviews.
- To handle data sets on life tables, actuarial mathematics, life insurance contracts.
- To estimate parameters and function of parameters of different lifetime distributions.
- To fit different models to time series data using R, to study the role of diagnostic tests that might be applied to the residuals of the fitted time series model and study the criteria for choosing between models.

Technical Communications (30 Hrs, 40 marks Internal Evaluation)

Each student will have to prepare his/her presentation/lecturer based on any topic from Statistics and deliver/present it before all students and teachers of the department.

Practicals-IV (90 Hrs, 60 marks External Practical Examination)

(Based on software and computer programming)

A. Practicals based on Stochastic Processes (30 Hrs, 20M)

1. Calculation of n-step transition probabilities and limiting distribution in Markov chain.
2. Realization of Markov chain.
3. Realization of Branching process.
4. Simulation of Poisson process.
5. Simulation of Random Walk.
6. Simulation of Renewal process.
7. Simulation of $M/M/1$ queuing system.
8. Simulation of Brownian Motion
9. Estimation of transition probability of Markov chain using realization.
10. MCMC Techniques

B. Practicals based on Time Series Analysis (30 Hrs, 20M)

1. Exploratory Time Series Analysis, multiple time series plots. Time series plots with discontinuities,
2. Simulation of various noise processes, stationary and non-stationary processes
3. Plotting of sample ACF and sample PACF of the time series data, Correlogram analysis and interpretation.
4. Applying different smoothing techniques to the time series data
5. Analyzing classical decomposition model.
6. Simulation of AR, MA and ARMA processes
7. Numerical exercises on AR, MA, ARMA and ARIMA models Forecasting.
8. Numerical exercises on Box-Jenkins models.
9. Residual analysis and diagnostic checking.
10. Periodogram analysis and interpretation.
11. Numerical exercises on non-Stationary time series models
12. Study of volatility models.

C. Practicals based on Actuarial Statistics (30 Hrs, 20M)

2. Calculation of simple interest and compound interest.
3. Relation between nominal, effective and force of interest.
4. Plotting of utility functions.
5. Distribution of total claim in short term risk models.
6. Construction of life tables and Problems based on life tables
7. Life table using analytical laws of mortality.
8. Annuity immediate and due.
9. Calculation of premiums.
10. Calculation of reserves.
11. Multiple life functions.

OR

C. Practicals based on Investment Under Uncertainty (30 Hrs, 20M)

Concerned teacher is expected to design the practicals according to the syllabus.

OR

C. Practicals based on Survival Analysis (30 Hrs, 20M)

1. Plotting of Survival and Hazard function for various families of distributions.
2. Parametric analysis for Survival data (both complete data and censored data)
3. Assess goodness-of-fit for parametric models
4. Comparison of parametric survival models
5. Nonparametric estimation of the Survival function:
   Estimate and plot survival curves using Kaplan-Meier method, actuarial estimator of survival function, log-rank test for comparing survival curves.
6. Fit Cox Proportional-Hazards model to survival data and interpret the results.
7. Machine learning approaches in Survival analysis.

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST524.1	Simulate stochastic processes and reliability systems and estimate different parameters of interest.	3
ST524.2	Fit suitable model to time series data and justify its appropriateness through accuracy measures or different criteria such as AIC, AICC or BIC.	3
ST524.3	Apply knowledge of life tables, Actuarial Mathematics, human mortality laws to calculate premium of life insurance contracts. OR Solve different problems related investment using statistical software. OR Apply appropriate survival models to analyze survival data	3 6 3

ST-525: RESEARCH PROJECT II

Course Objectives (CObs):

- The intention of the project work is to make students develop deeper knowledge and understanding in the context of a specific topic.
- The overall objective is to enable students think independently on a specific topic, design and perform a set of experiments to display the knowledge and capability required for independent work.
- It also aims at managing time effectively while working independently, appropriate referencing, analyzing the results and develop skills in report writing.
- Able to design research problem and solve it using a range of different data sources, design an intervention and evaluation.

- Project duration (Tentative Dates): 15th December to 15th April. Students are supposed to submit final project report in typed form with binding by April 20.
- General Guidelines:
  - Project work is considered as a special course involving application of knowledge in solving/applying/exploring a real-life situation/difficult problem. The student undertake research in specific area of his Major/Core subject with an advisory support by a teacher/faculty member.
  - Full flexibility is given to the student in identifying the project depending on the resources and infrastructure available in the Department/host organization. Students are encouraged to work on multidisciplinary projects but it is not mandatory.
  - Students should work in a group of 2-3 together, however, in any case, not more than four students should be involved in the same project work.
- Project Guide: Teachers from the Department of Statistics and/or personnel from organization where student is going to visit for field work or training related to project. Each project group will be guided by concerned teacher (guide) for one hour per week throughout the semester.
- Project Topic: Students in consultation with the guide will decide Project Topic/Area. Each research group should have different research topics with some kind of possible level of novelty.
- Project evaluation: Project evaluation will be based on
  - Internal Evaluation by Guide (60 Marks)
  - External Evaluation by two examiners (90 Marks)

- Project report (30 Marks)
- Presentation by student or group of students (30 Marks)
- Viva voce (30 marks)

Course Outcomes (COts):

After completing this course, the student will be able to:

Course Outcome	Course Outcome	Cognitive Level
ST525.1	Understand the real-world problems through statistical angle.	2
ST525.2	Design and execute experiments/surveys independently.	6
ST525.3	Analyze, co-relate and interpret the data and build up skills and temperament of scientific writing.	4

List of Equivalent Courses

Course Code	Title of Equivalent Course under Old Syllabus (w.e.f. 2019-20)	Course Code	Title of Equivalent Course under New Syllabus (w.e.f. 2023-24)
ST-101	Real Analysis	-	#
ST-102	Linear Algebra	-	#
ST-103	Sampling Theory and Statistics for National Development	ST-412	Sampling Theory and Statistics for National Development
ST-104	Distribution Theory	ST-411	Distribution Theory
ST-105	R Programming and Numerical Methods	ST-414 (A)	Advanced R Programming and Numerical Methods
ST-106	Practicals- I	ST-415	Practicals- I
ST-201	Probability Theory	ST-421	Probability Theory
ST-202	Linear Models and Regression Analysis	ST-424(A)	Linear Models and Regression Analysis
ST-203	Multivariate Analysis	ST-512	Multivariate Analysis
ST-204	Parametric Inference	ST-422	Parametric Inference
ST-205	Python Programming	-	#
ST-206	Practicals-II	-	#
ST-301	Asymptotic and Nonparametric Inference	ST-514 (C)	Asymptotic and Nonparametric Inference
ST-302	Design, Planning and Analysis of Experiments	ST-511	Design, Planning and Analysis of Experiments
ST-303	Total Quality Management (TQM) and Statistical Process Control (SPC) and Reliability	ST-514 (A)	Statistical Process Control and Reliability
ST-304	Stochastic Processes	ST-521	Stochastic Processes
ST-305(A)	Design and Analysis of Clinical Trials	-	#
ST-305(B)	Econometrics	ST-424 (B)	Econometrics
ST-306	Practicals-III	-	#
ST-401	Optimization Techniques	-	#
ST-402	Actuarial Statistics	ST-523 (A)	Actuarial Statistics
ST-403	Time Series Analysis	ST-522	Time Series Analysis
ST-404(A)	Data Mining	ST-514 (B)	Data Mining
ST-404(B)	Survival Analysis	ST-523 (C)	Survival Analysis
ST-405	Technical Communications and Practicals-IV	-	#
ST-406	Project	-	#
AC-101	Practicing Cleanliness	-	#
AC-201(A)	Soft Skills	-	#
AC-201(B)	Practicing Sports Activities	-	#
AC-201(C)	Practicing Yoga	-	#
AC-201(D)	Introduction to Indian Music	-	#
AC-301(A)	Computer Skills	-	#
AC-301(B)	Cyber Security	-	#
AC-301(C)	Base Statistical Analysis Software (SAS)	-	#
AC-301(D)	Review of Research Paper	-	#
AC-401(A)	Human Rights	-	#
AC-401(B)	Current Affairs	-	#
AC-401(C)	Introduction to LaTeX	-	#
AC-401(D)	History of Statistics	-	#

No equivalent courses available for these papers, so student may be allowed to appear by old courses.