Program: M.Stat./M.Sc.(Statistics)

(Designed under NEP 2020 and based on CBCS)

(Effective from academic year 2022-23)

Duration: 4 Semesters/2 years **Total Credits:** 100

Minimum Credits: I&II Sem.(first year): 52 credits + III&IV Sem.: 48 Credits

Total Credits = 100

First year Exit: Min 52 credits: **B.Sc.(Research) in Statistics**

Second year completion: Min 100 credits: **M.Stat./M.Sc.(Statistics)**

Semester I

Major & Minor Courses

S. No	Course No.	Course Title	Minor for other Faculty	Credits	L:T:P	Internal	External	Teaching Hours
1	STAT-101	Real Analysis & Matrix Algebra		4	3:1:0	25	75	60 hours
2	STAT-102	Probability and Distribution Theory		4	3:1:0	25	75	60 hours
3	STAT-103	Sampling Theory	Minor	4	3:1:0	25	75	60 hours
4	STAT-104	Computer Science	Minor	4	3:1:0	25	75	60 hours
5	STAT-105	Practical		4	0:0:4	---	100	120 hours
		Minor Elective: Other Faculty (to be completed in I or II semester)						
6		The students may choose this course from a list given below or from a list provided by the university*/ College		4/5/6	3:1:0	25	75	60 hours/ 75 hours/ 90 hours
		Major: Industrial Training/Survey/Research Project						
8	STAT-RP01	Research Project Work		4		100	---	120 hours

Semester II

Major & Minor Courses

S. No.	Course No.	Course Title	Minor for other Faculty	Credits	L:T:P	Internal	External	Teaching Hours
1	STAT-201	Statistical Inference-I		4	3:1:0	25	75	60 hours
2	STAT-202	Linear Estimation and Design of Experiments		4	3:1:0	25	75	60 hours
3	STAT-203	Multivariate Analysis		4	3:1:0	25	75	60 hours
4		Elective-I	Minor	4	3:1:0	25	75	60 hours
5	STAT-207	Practical		4	0:0:4	---	100	120 hours

(i)

List of Elective Courses		Major/Minor
STAT-204(E)	Data Analysis using SPSS	Minor Elective for other faculty and value Added Courses
STAT-205(E)	Data Analysis using R	Minor Elective for other faculty and value Added Courses
STAT-206(E)	Data Analysis using Python	Minor Elective for other faculty and value Added Courses

Minor Elective: Other Faculty (to be completed in I or II semester)							
6		The students may choose this course from a list given below or from a list provided by the university*/College	4/5/6	3:1:0	25	75	60 hours/ 75 hours/ 90 hours
Major: Industrial Training/Survey/Research Project							
7	STAT-RP02	Research Project Work	4		---	100	120 hours

***Minor Elective**

1. Social Problems in India: Faculty of Arts
2. Sociology of Population Studies.: Faculty of Arts
3. NGO Management: Faculty of Arts
4. Disaster Management: Faculty of Arts

Note: The minor elective course chosen by the student will be decided by the department to be offered either in I semester or in the II semester.

Semester III

Major & Minor Courses

S. No.	*Course No.*	*Course Title*	*Minor for other Faculty*	*Credits*	*L:T:P*	*Internal*	*External*	*Teaching Hours*
1	STAT -301	Statistical Inference-II		4	3:1:0	25	75	60 hours
2	STAT -302	Operations Research	Minor	4	3:1:0	25	75	60 hours
3		Elective II	Minor	4	3:1:0	25	75	60 hours
4		Elective III	Minor	4	3:1:0	25	75	60 hours
5	STAT -308	Practical		4	0:0:4	---	100	120 hours
Major: Industrial Training/Survey/Research Project								
6	STAT - RP03	Research Project Work		4		100	---	120 hours

List of Elective Courses		Major/Minor
STAT- 303(E)	1. Data Mining	Minor Elective for other faculty and value Added Courses
STAT- 304(E)	2. Population Studies	Minor Elective for other faculty and value Added Courses
STAT- 305(E)	3. Medical Statistics	Minor Elective for other faculty and value Added Courses
STAT- 306(E)	4. Official Statistics	Minor Elective for other faculty and value Added Courses
STAT- 307(E)	5. Econometrics	Minor Elective for other faculty and value Added Courses
STAT- 308(E)	6. Actuarial Statistics	Minor Elective for other faculty and value Added Courses

Manjika

(ii)

Signature Not Verified
RAJEEV KUMAR
Digitally Signed By: Registrar

Semester IV

Major Courses

S. No.	Course No.	Course Title	Minor for other Faculty	Credits	L:T:P	Internal	External	Teaching Hours
1		Elective IV		4	3:1:0	25	75	60 hours
2		Elective V		4	3:1:0	25	75	60 hours
3		Elective VI		4	3:1:0	25	75	60 hours
4		Elective VII		4	2:0:2	25	75	60 hours
5	STAT -410	Practical		4	0:0:4	---	100	120 hours
Major: Industrial Training/Survey/Research Project								
6	STAT - RP04	Research Project Work		4		---	100	120 hours

List of Elective Courses		Major/Minor
STAT-401(E)	Decision Theory and Bayesian Inference	
STAT-402(E)	Advanced Sample Survey	
STAT-403(E)	Sequential Analysis	
STAT-404(E)	Limit Theorems and Stochastic Processes	
STAT-405(E)	Data Science	Minor Elective for other faculty and value Added Courses
STAT-406(E)	Reliability Theory and Survival Analysis	
STAT-407(E)	Linear Models and Regression Analysis	
STAT-408(E)	Research Methodology	Minor Elective for other faculty and value Added Courses
STAT-409(E)	Marketing Research Statistics	Minor Elective for other faculty and value Added Courses

Summary

S. No.	Semester	Major Courses	Minor Elective (other faculty)	Industrial Training/Survey/ Research Project	Total
1	Sem. I	20	4/5/6	4	Min 52
2	Sem. II	20		4	
3	Sem. III	20		4	Min 48
4	Sem. IV	20		4	
	Total				100

Duration: Six Months: 90 Days= 90(days)/6(days)=15 weeks
1 credit theory=15 hours of teaching in a semester
1 credit practical=30 hours of teaching in a semester

(iii)

Real Analysis and Matrix Algebra

Course No. STAT–101

Unit 1

Real Number System, Algebra of Sets, Relations, Functions, Countability, Metric Spaces, Open and Closed Sets, Limit Points, Compact Sets.

Unit 2

Sequence, Cauchy Sequence, Cauchy's Theorem, Monotonic Sequence, Nested Interval Theorem. Infinite Series, Comparison Test, Cauchy's Root Test, D'Alembert's Ratio Test, Raabe's Test, Lograthmic Test, Cauchy's Condensation Test, Integral Test, Gauss Test, Aeterating Series, Absolute Convergence. Limits and Continuity of Functions, Uniform Continuity, Differentiation.

Unit 3

Riemann Integrals, Properties of Darboux Sums, Conditions of Integrability, Classes of Integrable Functions, Algebra of Integrable Functions, Riemann-Stieltges Integral, Existence Theorem, Uniform Convergence Test for Uniform Convergence, Properties of Uniformly Convergent Sequence and Series.

Unit 4

Algebra of Matrices, Trace, Determinants, Inverse, Generalised Inverse, Rank, Linear Equations. Characteristic Roots and Vectors.

Unit 5

Vector Spaces, Subspaces, Linear Independence and Dependence of Vectors, Dimension and Basis of a Vector Space, Gram-Schmidt Orthogonalization.

Books Recommended :

1. Mathematical Analysis—T.M. Apostol
2. Principles of Mathematical Analysis—Walter Rudin
3. Topics in Algebra of Matrices—S. Biswas
4. Linear Algebra—A.R. Rao and P. Bhimasankaram
5. Matrix Algebra Useful for Statistics—S.R. Searle
6. A Course of mathematics Analysis—Shanti Narayan and P.K. Mittal

Probability and Distribution Theory

Course No. STAT–102

Unit 1

Classes of sets, sequences of sets, limit superior and limit inferior of a sequence of sets, fields, sigma field, minimal sigma field, Borel sigma field on real line. Event and event space, sample space, probability measure, properties of measure, independent events, conditional probability and Bayes' theorem.

Unit 2

Measurable functions, random variables, functions of random variables, induced probability measure. Distribution function, joint and marginal and conditional distribution in R^n. Expected values, moments, some related inequalities.

Unit 3

Probability generating function, moment generating function, characteristic function and their properties: uniqueness, continuity and inversion with application.

Unit 4

Bernoulli, Binomial, Multinomial, Hyper geometric, Poisson, Geometric, and negative Binomial distribution. Uniform, Exponential, Cauchy, Beta, Gamma, Normal, Lognormal and Pareto distributions.

Unit 5

Transformation technique. Probability sampling distributions of Chi-square, t and F statistics and their distributions. Order statistics and their distributions.

Books Recommended :

1. Bhat, B.R.(1981).Modern Probability Theory, III Edition, New Age International(P).
2. Das,K.K and Bhattacharyajee,D.(2008). A Treatise On Statistical Inference And Distributions, Asian Books, New Delhi.
3. Feller,W.(1969). Introduction To Probability And Its Applications, Vol.II, Wiley Eastern Ltd.
4. Hogg,R.V.,Craig,A. and Mckean,J.W.(2005).Introduction To Mathematical Statistics, Sixth Edition, Pearson.
5. Johnson,S.and Kotz(1995).Distributions In Statistics, Vol.-I,II And III,Houghton And Mifflin.
6. Loeve,M.(1978). Probability Theory (Springer Verlag), Fourth Edition.
7. Mood,A.M., Graybill,F.A. and Boes,D.C.(1974).Introduction To Theory Of Statistics, Third Edition, Mc Graw Hill.
8. Mukhopadhyaya,P.(1996). Mathematical Statistics, Calcutta publishing house.Rohatgi,V.K.(1984), An introduction to probability theory and mathematical statistics, Wiley Eastern.

Sampling Theory

Course No. STAT-103

Unit 1

Review of simple random sampling with and without replacement. Determination of sample size. Sampling with varying probabilities, ordered and unordered estimates. Des Raj estimator , Horvitz-Thompson estimator.

Unit 2

Stratified random sampling, different types of allocation, allocation problems, problem of allocation with more than one item. Effect of deviation from optimum allocation, construction of strata, number of strata, method of collapsed strata, post stratification.

Unit 3

Ratio method of estimation- concept of ratio estimators, ratio method of estimation in simple random sampling, their bias, variance/MSE, Conditions under which ratio estimators are BLUE, ratio estimators in stratified random sampling. Regression method of estimation-concept of regression estimators, difference estimator, regression estimator in SRS, their bias, variance/MSE, regression estimator in stratified random sampling.

Unit 4

Systematic sampling (circular, population with trend), domain estimation in SRS. Comparison with SRS and stratified random sampling. Cluster sampling with equal and unequal cluster sizes, estimation of mean and variances. Efficiency of cluster sampling in terms of intra- class correlation coefficient.

Unit 5

Two stage sampling with equal first stage units and unequal first stage units- estimator of population mean and variance/MSE. Determination of optimal sample sizes at both the stages. Double sampling and its use in ratio and regression method of estimation. Non-sampling errors, error in surveys, observational error: mathematical model of measurement of observational error. Effect of non-response.

Books Recommended:

1. Cochran,W.G.(1997). Sampling Techniques, Wiley Eastern, New Delhi.
2. Des Raj and Chandok,P. (1998). Sampling Theory, Narosa, New Delhi.
3. Mukhopadhyay, P.(1998). Theory And Methods Of Survey Sampling, Pentice Hall of India, New Delhi.
4. Murthy,M.N.(1977). Sampling Theory And Methods, Statistical Publishing Society, Kolkata.
5. Sukhatme,P.V.,Sukhatme,B.V., Sukhatme, S. and Asok ,C.(1984).Sampling Theory Of Surveys And Applications,Iowa State University press and IARS.

Computer Science

Course No. STAT–104

Unit 1

Introduction to Computer, Evolution of Computer Technology, Types of Computers. Basic elements of Computer, Components of Computer, Generations of Computer and Computer languages.

Unit 2

MS Office, Introduction of MS word, MS Excel, MS PowerPoint, Basic Features and applications.

Unit 3

Knowledge and familiarity with statistical package SPSS, The Fundamental Mechanics of SPSS. Getting Data into and out of SPSS, Graphical representation of data. Tabulation of data. Descriptive Statistics. Summarizing Data, Creating & Editing Charts, Modifying data values. Sorting & Selecting Data Values. Chi- Square and t-test.

Unit 4

Advance features of SPSS, Correlation & Regression, One-way ANOVA, Factorial ANOVA, Nonparametric Tests, Discriminant Analysis, Factor Analysis, Cluster Analysis.

Unit 5

R-Programming, Overview of R, R data Types & Objectives. Reading & Writing Data. Control Structures. Function, Scoping Rules, Loop Functions, Simulation.

Books Recommended:

1. Argyrous, G. (2012). Statistics for Research: With a Guide to SPSS. Sage South Asia: Third Edition.
2. Cox & Lambert (2010). Microsoft Word 2010: Step by Step. Microsoft Press.
3. George Darren: SPSS for Window Step by Step.
4. Griffith, A. (2007). SPSS For Dummies. Published by Wiley Publishing, Inc.
5. Hothorn,T and Everitt, B.S.(2014). A Handbook of Statistical Analyses Using R. Chapman & Hall/CRC Press, Boca Raton, Florida, USA, 3rd edition.
6. Knell, R.J. (2013), Introductory R: A Beginner's Guide to Data Visualisation and Analysis using R.
7. Norton, P. (2010). Introduction to Computers. McGraw Hill Education (India) Private Limited.
8. Patric L. . A. K. and Feeney B. C.: A Simple Guide to SPSS.
9. Sheridon J Coaks:, SPSS.

Statistical Inference-1

Course No. STAT-201

Unit 1

Data Reduction: Data reduction, Sufficiency, Sufficient partition, Completeness, Minimal sufficiency and ancillarity, ancillarity and Basu's theorem, Exponential families and Pitman families. Invariance property of sufficiency under one-one transformations of sample and parameter spaces. Canonical form, convexity property, minimal sufficiency and completeness, Neymann factorization theorem(Proof for discrete case only), examples.

Unit 2

Point Estimation: Estimability of parametric functions, Unbiased Estimator, Rao-Blackwell and Lehmann-Scheffe theorem, Some special class of distributions admitting complete sufficient statistics, extension of results to multi parameter case, Fisher Information for one and several parameters models, Estimation of bias and standard deviation of point estimators by Jackknife, the bootstrap methods with examples, Minimum Variance Unbiased Estimators (UMVUEs)

Unit 3

Variance Lower Bound: Lower bounds for variance of estimators, Frechet Cramer and Rao (FCR), Chapman, Robbins and Kiefer (CRK) and Bhattacharya lower bounds, necessary and sufficient conditions for MVUE

Unit 4

Testing of Hypotheses: Review of notions of randomized and nonrandomized tests, level, size, p-value, power function, Neyman-Pearson fundamental lemma and its applications, UMP tests for simple null hypothesis against one-sided alternatives and for one-sided null against one-sided alternatives in one parameter exponential family. Extension of these results to Pitman family when only upper or lower end depends on the parameters and to distributions with MLR property. Families of distributions with Monotone Likelihood Ratio, UMP tests for one-sided testing problems, UMP tests for one-parameter exponential families, Generalized Neyman-Pearson Lemma, Non-existence of UMP tests for simple null against two-sided alternatives in one parameter exponential family, Unbiased Test, UMPU tests for one-parameter exponential family, one- and two-sided testing problems

Unit 5

Interval Estimation: Interval estimation, confidence sets, relation with hypothesis testing, confidence level, construction of confidence intervals with pivots and shortest expected length, Uniformly most accurate (UMA) one-sided confidence interval and its relation to UMP tests for one-sided null against one-sided alternative hypotheses, Confidence intervals for the parameter for Normal, Exponential, Binomial and Poisson distributions. CI for quintiles, concept of tolerance limits and examples.

Suggested Readings:

1. Bartoszynski, R. and Bugaj, M.N. (2007). *Probability and Statistical Inference*, John Wiley & Sons.
2. BradelyEfronand Robert J Tibshirani: *An Introduction to the Bootstrap*, Chapman and Hall
3. Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series, Second Edition.
4. Cramer, H.(1974). *Mathematical Methods in Statistics*. Princeton Univ. Press.
5. Dudewicz, E. J. and Mishra, S.N.(1988). *Modern Mathematical Statistics*, John Wiley.
6. Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*. Cambridge University Press
7. Ferguson, T.S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press.
8. Gray and Schucany. *Generalized Jakknife*; Marcel Decker,
9. Gray, Schncory and Watkins. *Generalized Jakknife*. Dovenpul
10. Kale, B.K. &Muralidharan, K. (2015) *Parametric Inference: An Introduction*. Alpha Science International Ltd.

11. Kale, B.K. (2005). *A First Course on Parametric Inference*. Second Edition. Narosa.
12. Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*. Springer. New York
13. Lehmann, E. L. and Romano, J. (2005). *Testing Statistical Hypotheses*. Springer
14. Rao, C. R.(1995). *Linear Statistical Inference and its Applications*. Wiley Eastern Ltd..
15. Rohatgi, V.K. and Saleh, A.K. Md. E. (2005). *An Introduction to Probability and Statistics*. Second Edition, John Wiley.
16. Roussas, G. G. (1973). *First Course in Mathematical Statistics*. Addison Wesley. Ltd.. New Delhi.
17. Silvey, S. D. (1975). *Statistical Inference*. Chapman and Hall.
18. Srivastava, M. K. & Srivastava, N. (2014). *Statistical Inference: Testing of Hypotheses*. PHI Learning. Second Edition. New Delhi.
19. Srivastava, M. K. & Srivastava, N. (2018). *Statistical Inference: Theory of Estimation*. PHI Learning. Second Edition, New Delhi.
20. Shao, J. (2003). *Mathematical Statistics*, Springer-Verlag. New. New York.
21. Wilks, S. S. (1962). *Mathematical Statistics*. John Wiley.
22. Zacks, S. (1971). *Theory of Statistical Inference*, John Wiley & Sons

Linear Estimation and Design of Experiment

Course No. STAT–202

Linear Estimation :

Unit 1

Linear Transformation and Projections, The Linear Model, Estimable functions, Estimation and Error space. Best estimates, Normal Equations, Variance and Covariance of Estimates, relation with least squares.

Unit 2

Sums of Squares, Degrees of Freedom, Estimate and Error Sums of Square, Calculation and Distribution of Sums of Squares, Linear Hypothesis, Estimable Linear Hypothesis. The Generalised t-test and Generalised F-test, Calculation of the sums of squares for testing estimable linear hypothesis by Least Square Theory.

Design of Experiment :

Unit 3

Planning of experiment, Completely RandomisedDesign, Randomised Block Design, Latin Square Design.Analysis of Covariance with One Concomitant Variable.

Unit 4

General Incomplete Block Design, Balanced Incomplete Block Design, Partially Balanced incomplete Block Design (with two associate classes).

Unit 5

Symmetric and Asymmetrical Factorial Design. Yates method of analysis for 2^n and 3^n Design. Partial and total confounding in 2^2, 3^2 and 3^3 Design. Fractional Replication in 2^n Design.

Books Recommended :

1. Cochran, W.G. and Cox, G.M. (1959). Exponential Designs. Asia Publishing House, Singapore.
2. Das, M.N. and Giri, N.C. (1986). Design and Analysis of Experiments, Wiley Eastern Limited.
3. Dean, A. and Voss, D. (1999). Design and Analysis of Experiments, Springer. FirstIndian Reprint 2006.
4. Joshi, D.D. (1987). Linear Estimation and Design of Experiments, Wiley Eastern, NewDelhi.
5. Montgomery, D.C. (2005). Design and Analysis of Experiments, Sixth Edition, JohnWiley and Sons.

Multivariate Analysis

Course No. STAT-203

Unit 1

Definite and semi definite quadratic forms, rank, index, and signature of a quadratic form. Equivalence, congruence relation of quadratic forms. Simultaneous reduction of quadratic forms. Differentiation of functions of vectors. Distribution of quadratic forms, characteristic function, independence of quadratic forms.

Unit 2

Motivation to take up multivariate data analysis. Concept of random vectors, its expectation, and variance covariance matrix. Singular and non-singular multivariate distributions. Marginal and conditional distributions. Joint distribution of linear forms.

Unit 3

Regression, Total, partial and multiple correlation coefficients and their distributions. Recursion formulae. Maximum likelyhood estimates of mean vector and dispersion matrix, independence and joint sufficiency of these estimates. Hotelling T^2-statistics as a function of likelyhood ratio criterion, its distribution, optimum properties and applications.

Unit 4

Wishart distribution (without derivation) and its properties. Generalised variance, distribution of sample generalised variance. Classification problems, Fisher's discriminant function, D^2-statistics and its application. Analysis of dispersion, and testing of general linear hypothesis, equality of mean vectors, Wilk's lamda effect, equality of dispersion matrices.

Unit 5

Principal components, maximum liklyhood estimates of principal components and their variances. Canonical variates and correlation- use, estimation and computation. Cluster analysis, factor analysis.

Books Recommended:

1. Anderson,T.W.(1958). An Introduction To Multivariate Statistical Analysis, Second Edition.Wiley.
2. Giri,N.C.(1977). Multivariate Statistical Inference, Academic Press, New York.
3. Johnson,R.A. and Wichern,D.W.(2003). An Introduction To Applied Multivariate Analysis, 5/e.Pearson Education.
4. Johnson,R.A. and Wichern,D.W.(1986). Applied Multivariate Analysis, Wiley.
5. Kshirsagar,A.M.(1972). Multivariate Analysis, Marcel-Dekker.
6. Morrison,D.F.(1976). Multivariate Statistical Methods, Mc Graw-Hill.
7. Singh,B.M.(2002). Multivariate Statistical Analysis, South Asian Publishers, New Delhi.
8. Srivastava,M.S. and Khatri,C.G.(1979). An Introduction To Multivariate Statistics, North Holland.

Data Analysis using SPSS

Course No. STAT-204(E)

Unit 1

Data Management: Types of Data, Metric and Nonmetric Data, Variables, Attributes, Mutually Exclusive Attributes, Independent and Dependent Variable, Extraneous Variable, Sources of Research, Primary and Secondary Data, Data Cleaning, Detection of Errors

Data Preparation in SPSS: Defining Variables and Their Properties Under Different Columns, Defining Variables for a Data File, Entering the Data, Importing Data in SPSS, Importing Data from an ASCII File, Importing Data File from Excel Format

Descriptive Analysis: Measures of Central Tendency, Mean, Median, Mode and their interpretation. Measures of Variability, Range, Interquartile Range, Standard Deviation, Variance, The Index of Qualitative Variation, Standard Error, Coefficient of Variation (CV), Moments, Skewness, Kurtosis, Percentiles, Percentile Rank . Situation for Using Descriptive Study, Computation of Descriptive Statistics using SPSS, Interpretation of the Outputs, Developing Profile Chart

Chi-Square Test and Its Application: Advantages of Using Crosstabs, Statistics Used in Cross Tabulations, Chi-Square Statistic, Chi-Square Test, Application of Chi-Square Test, Contingency Coefficient, Lambda Coefficient, Phi Coefficient, Gamma, Cramer's V, Kendall Tau, Situation for Using Chi-Square, Chi-square for Testing an Equal Occurrence Hypothesis, Computation of Chi-Square Using SPSS, Chi-square for Testing the Significance, Interpretation of the Outputs, Association Between Two Attributes, Computation of Chi-Square for Two Variables Using SPSS, Interpretation of the Outputs

Unit 2

Correlation Matrix and Partial Correlation: Explaining Correlation Matrix and Partial Correlation, Product Moment Correlation Coefficient, Situation for Using Correlation Matrix and Partial Correlation, Construction of Research Hypotheses to Be Tested, Statistical Test, Computation of Correlation Matrix Using SPSS, Interpretation of the Outputs, Computation of Partial Correlations Using SPSS, Interpretation of Partial Correlation

Regression Analysis and Multiple Regression Analysis: Terminologies in used in Regression Analysis, Multiple Correlation, Coefficient of Determination, The Regression Equation, Multiple Regression, Application of Regression, Computation of Regression Coefficients, Multiple Correlation, and Other Related Output in the Multiple Regression Analysis and Interpretation of the Outputs: Multiple Collinearity, Normality of residuals, Variance Distribution of residuals, Outlying observations and other related concepts. Multicollinearity and Varimax test.

Logistic Regression: Developing a Model for Risk Analysis, What Is Logistic Regression?, Important Terminologies in Logistic Regression, Outcome Variable, Natural Logarithms and the Exponent Function, Odds Ratio, Maximum Likelihood, Logit, Logistic Function, Logistic Regression Equation, Judging the Efficiency of the Logistic Model, Understanding Logistic Regression, Graphical Explanation of Logistic Model, Logistic Model with Mathematical Equation, Interpreting the Logistic Function, Assumptions in Logistic Regression, Important Features of Logistic Regression, Research Situations for Logistic Regression, Steps in Logistic Regression, Computation of Logistics Analysis Using SPSS, Interpretation of Various Outputs Generated in Logistic Regression, Explanation of Odds Ratios

Unit 3

Hypothesis Testing for Decision-Making: Hypothesis Construction, Null and Alternative Hypothesis, Test Statistic, Rejection Region, Steps in Hypothesis Testing, Type I and Type II Errors, One-Tailed and Two-Tailed Tests, Criteria for Using One-Tailed and Two-Tailed Tests, Strategy in Testing One-Tailed and Two-Tailed Tests, What Is p Value?, Degrees of Freedom, One-Sample t-Test, Application of One-Sample Test, Two-Sample t-Test for Unrelated Groups, Assumptions in Using Two-Sample t-Test, Application of Two-Sampled t-Test, Assumptions in Using Paired t-Test, Testing Protocol in Using Paired t-Test, Testing for Single Group Mean, Computation of t-Statistic and Related Outputs, Interpretation of the Outputs, Two-Sample t-Test for Unrelated Groups, Computation of Two-Sample t-Test for Unrelated Groups, Interpretation of the Outputs, Paired t-Test with SPSS, Computation of Paired t-Test for Related Groups, Interpretation of the Outputs

Unit 4

One-Way ANOVA: Comparing Means of More than Two Samples. Principles of ANOVA Experiments. One-Way ANOVA. Factorial ANOVA, Repeated Measure ANOVA. Multivariate ANOVA. One-Way ANOVA Model and Hypotheses Testing. Assumptions in Using One-Way ANOVA. Effect of Using Several t-tests Instead of ANOVA, Application of One-Way ANOVA, One-Way ANOVA with Equal Sample Size Using SPSS, Computations in One-Way ANOVA with Equal Sample Size. Interpretations of the Outputs. Example of One-Way ANOVA with Unequal Sample, Computations in One-Way ANOVA with Unequal Sample Size. Interpretation of the Outputs

Two-Way Analysis of Variance: Examining Influence of Two Factors on Criterion Variable.

Principles and Classification of ANOVA Experiment, Factorial Analysis of Variance, Repeated Measure Analysis of Variance, Multivariate Analysis of Variance (MANOVA). Advantages of Two-Way ANOVA over One-Way ANOVA. Important Terminologies Used in Two-Way ANOVA. Factors. Treatment Groups. Main Effect, Interaction Effect, Within-Group Variation, Two-Way ANOVA Model and Hypotheses Testing. Assumptions in Two-Way Analysis of Variance, Situation Where Two-Way ANOVA Can Be Used. Computation of Two-Way ANOVA Using SPSS, Model Way of Writing the Results of Two-Way. ANOVA and Its Interpretations

Analysis of Covariance: Increasing Precision in Comparison by Controlling Covariate

Concepts of ANCOVA, Graphical Explanation of Analysis of Covariance. Analysis of Covariance Model. What We Do in Analysis of Covariance? When to Use ANCOVA. Assumptions in ANCOVA. Efficiency in Using ANCOVA over ANOVA. Computation of ANCOVA Using SPSS. Model Way of Writing the Results of ANCOVA and Their

Interpretations

Unit 5

Cluster Analysis: For Segmenting the Population. What Is Cluster Analysis?. Terminologies Used in Cluster Analysis. Distance Measure. Clustering Procedure, Standardizing the Variables. Icicle Plots, The Dendrogram. The Proximity Matrix. What We Do in Cluster Analysis. Assumptions in Cluster Analysis. Research Situations for Cluster Analysis Application. Steps in Cluster Analysis. Computation of Cluster Analysis Using SPSS. Stage 1. Stage 2. Stage 1: SPSS Commands for Hierarchal Cluster Analysis. Stage 2: SPSS Commands for K-Means Cluster Analysis. Interpretations of Findings

Application of Factor Analysis: To Study the Factor Structure Among Variables. What Is Factor Analysis? Terminologies Used in Factor Analysis. Principal Component Analysis. Factor Loading. Communality. Eigenvalues. Kaiser Criteria. The Scree Plot. Varimax Rotation. What Do We Do in Factor Analysis? Assumptions in Factor Analysis. Characteristics of Factor Analysis. Limitations of Factor Analysis. Research Situations for Factor Analysis Computation of the Factor Analysis using SPSS. Interpretation of Various Outputs Generated in Factor Analysis. CFA and Structural Equation Models using AMOS.

Application of Discriminant Analysis: For Developing a Classification Model

What Is Discriminant Analysis? Terminologies Used in Discriminant Analysis. Variables in the Analysis. Discriminant Function. Classification Matrix. Stepwise Method of Discriminant Analysis. Power of Discriminating Variables. Box's M Test. Eigenvalues. The Canonical Correlation. Wilks' Lambda. What We Do in Discriminant Analysis.. Assumptions in Using Discriminant Analysis. Research Situations for Discriminant Analysis. SPSS Commands for Discriminant Analysis. Interpretation of Various Outputs Generated in Discriminant Analysis

Multidimensional Scaling for Product Positioning: What Is Multidimensional Scaling. Terminologies Used in Multidimensional Scaling. Objects and Subjects. Distances. Similarity vs. Dissimilarity Matrices. Stress. Perceptual Mapping.

Dimensions. What We Do in Multidimensional Scaling? Procedure of Dissimilarity-Based Approach of Multidimensional Scaling. Procedure of Attribute-Based Approach of Multidimensional Scaling. Assumptions in Multidimensional Scaling. Limitations of Multidimensional Scaling. Calculation of Multidimensional Scaling (Dissimilarity-Based Approach of Multidimensional Scaling) Using SPSS. Interpretation of Various Outputs Generated in Multidimensional Scaling

Data Analysis Using SPSS

(Practical)

CourseCode:STAT-204(E)

Lab Work Based on theory of STAT-204(E)

Credit:6(Th:3+Pr:3)
Teaching: 90 hours
Th:45 hours, Pr:45 hours
MaxMarks: 60+40
End Semester: 60=Th:30+Pr:30
Internal: 40=Th:20+Pr:20

Course objectives: To learn real life/Social Sciences/Management/industry based applications of theory

Data Analysis Using R

Course No. STAT-205(E)

Unit 1

Preparatory/Background/Prerequisite:

Probability - The Science of Uncertainty and Data

(i) Introduction to Probability models and axioms, sets, sequences, limits, and series, (un)countable sets, (iii) Conditioning and independence: Conditioning and Bayes' rule, Independence, (iv) Counting: Counting (v) Discrete random variables, Probability mass functions and expectations, Variance, Conditioning on an event, Multiple random variables, Conditioning on a random variable, Independence of random variables, (vi) Continuous random variables: Probability density functions, Conditioning on an event: Multiple random variables, Conditioning on a random variable: Independence, Bayes' rule (vii) Further topics on random variables: Derived distributions, Sums of independent random variables, Covariance and correlation, Conditional expectation and variance revisited, Sum of a random number of independent random variables, (viii) Bayesian inference: Introduction to Bayesian inference, Linear models with normal noise, Least mean squares (LMS) estimation, Linear least mean squares (LLMS) estimation, (ix) Limit theorems and classical statistics: Inequalities, convergence, and the Weak Law of Large Numbers, The Central Limit Theorem (CLT), An introduction to classical statistics (x) Bernoulli and Poisson processes: The Bernoulli process, The Poisson process, More on the Poisson process (xi)Markov chains: Finite-state Markov chains, Steady-state behavior of Markov chains, Absorption probabilities and expected time to absorption.

Presentationofdata:Classification.Tabulation.Diagrammatic&GraphicalRepresentationofGroupeddata.Frequency distributions.Cumulativefrequencydistributionsandtheirgraphicalrepresentations.Histogram.Frequencypolygonand dOgives.StemandLeafplot. BoxPlot.

TheoryofProbability: Randomexperiment,Trial,SamplepointandSamplespace, Events, Operations of events, Concept of equallylikely, Mutuallyexclusiveand Exhaustive events.

DefinitionofProbability:Classical,RelativefrequencyandAxiomaticapproaches.

DiscreteProbabilitySpace,PropertiesofProbabilityunderSetTheoryApproach,Independenceofl:Events.Conditiona l Probability, Total and Compound Probabilitytheorems.Bayestheoremanditsapplications.

RandomVariables–DiscreteandContinuous,Probability Mass Function (pmf) and Probability densityfunction(pdf).Cumulativedistributionfunction(cdf).

Joint distribution of two random variables, Marginal andConditionaldistributions,Independenceofrandomvariables.

Expectationofarandomvariableanditsproperties,Expectation of sum of random variables and product ofindependent random variables, Conditional expectationandrelatedproblems.

Moments,Moment generating function(m.g.f.) &theirproperties. Continuity theorem for m.g.f. (without proof).Chebyshev's inequality. Weak law of large numbers for asequenceofindependentlyandidenticallydistributedrandomvariablesandtheirapplications.(StatementOnly)

Fundamentals of Statistics

(i) Construction of estimators using method of moments and maximum likelihood, and decide how to choose between them. (ii) Quantify uncertainty using confidence intervals and hypothesis testing. (iii) Choose between different models using goodness of fit test; Make prediction using linear, nonlinear and generalized linear models (iv) Perform dimension reduction using principal component analysis (PCA)

Unit 2

Fundamentals of R

(i) Introduction to R, importance of R, features of R, installation of R, starting and ending R session, getting help in R, R commands and case sensitivity, (ii) Data types : Logical, numeric and complex Vectors and vector arithmetic, (iii) Variables/Objects in R: creation of vectors using functions c, assign,seq,rep: Arithmetic operations on vectors using operators +, -, *, /, ^, (iv) Numerical functions: log10, log, sort, max, min, unique, range, length, var, prod, sum, summary, fivenumete, (iv) Conditional statements (v) accessing vectors, alternative ways to create vectors by scan function, (vi) Data frames: creation using data.frame, subset and transform commands, (vii) Resident data sets: Accession and summary (viii) Loops (ix) Functions (x) Graphics

using R: (a) High level plotting functions. (b) Low level plotting functions. (c) Interactive graphic functions. (d) Using R as a calculator

Unit 3

Machine Learning with R: from Linear Models to Deep Learning

(i) Linear classifiers. separability. perceptron algorithm: (ii) Maximum margin hyperplane. loss. regularization: (iii) Stochastic gradient descent. over-fitting, generalization: (iv) Linear regression: (v) Recommender problems. collaborative filtering: (vi) Non-linear classification. kernels: (vii) Learning features. Neural networks: (viii) Deep learning. back propagation; (ix) Recurrent neural networks: (x) Generalization. complexity. VC-dimension: (xi) Unsupervised learning: clustering; (xii) Generative models. mixtures: (xiii) Mixtures and the EM algorithm; (ivx) Learning to control: Reinforcement learning: (xv) Reinforcement learning continued: (xvi) Applications: Natural Language Processing (xvii) R Projects: (Only one of these projects would be offered to the students) Automatic Review Analyzer: Digit Recognition with Neural Networks: Reinforcement Learning

MeasuresofCentraltendencyandDispersionandtheirproperties.MeritsandDemerits oftheseMeasures

Unit 4

Applications of Data Science using R-I

(i) Sampling methods: Drawing a sample from population using SRSWR. SRSWOR. stratified random sampling. systematic sampling. (ii) Diagrams: Simple bar diagram. Subdivided bar diagram. multiple bar diagram, Pie diagram. Stem and leaf diagram, (iii) Graphs: Box plot,rod or spike plot. histogram (both equal and unequal class intervals); frequency polygon. o give curves. empirical distribution function. (iv) Computation of measures of central tendency. dispersion. skewness and kurtosis: (a) central tendency: mean. mode. median. quartiles. deciles. percentiles. g.m. and h.m (b) Dispersion: variance. standard deviation. coefficient of variation. mean deviation (c) Skewness : Bowley's coefficient and Karl Pearson's coefficient of skewness (d) Moments : Computations of raw and central moments. measure of skewness and kurtosis based on it. (v) Probability distributions : Simulation of random experiment. Hypergeometric distribution: computation of probabilities. Binomial distribution: computation of probabilities. model sampling. fitting. Poisson distribution: computation of probabilities. model sampling; fitting Normal distribution: computation of probabilities. model sampling. Fitting. testing normality using Anderson- Darling or Wilks- Shapero test. (vi) Correlation and regression: Fitting of lines of regression. computation of correlation coefficient. Fitting of parabola: Multiple regression : Fitting of regression plane for trivariate data. (vii) Testing of hypotheses: Large sample tests for means and proportions. t.test. var.test. chisq.test (viii) Analysis of variance: one way. two way

Unit 5

Applications of Data Science using R-II

Choose any one of following real life problem:

1. Stock Exchange Data Analysis and Building of Assessment and Forecasting Model
2. Bank Fraud Networks and Network Analysis
3. Market price and sale dynamics of house consumable products and role of online stores such as Amazon.com etc
4. Web data analysis
5. Prices. Economies and Time Series
6. Environmental Data and Spatial Statistics

Discuss one or more of the following procedures to analyse data and prepare report

(i) Model. form hypotheses. perform statistical analysis on real data (ii) Use dimension reduction techniques such as principal component analysis to visualize high-dimensional data and apply this to genomics data (ii) Analyze networks (e.g. social networks) and use centrality measures to describe the importance of nodes. and apply this to criminal networks (iii) Model time series using moving average. autoregressive and other

stationary models for forecasting with financial data (iv) Use Gaussian processes to model environmental data and make predictions (v) Report on analysis results effectively

Boods Recommended:

1- S.G. Purohit, S.D. Gore, S.R. Deshmukh (2008). Statistics using R, Narosa Publishing House(First edition.

2- Gareth James, Daniela Witten et.al.(2013). An Introduction to Statistical Learning with Application in R, Springer Series.

SuggestedOnline Links/Readings: http://heecontent.upsdc.gov.in/SearchContent.aspxhttps://swayam.gov.in/explorer?searchText=statisticshttps://nptel.ac.in/course.htmlhttps://www.edx.org/search?q=statisticshttps://www.coursera.org/search?query=statistics&

Statistical Inference-II

Course No. STAT-301

Unit 1

Asymptotic Inference: Consistency and asymptotic relative efficiency of estimators. Consistent and asymptotically normal (CAN) estimators for real vector valued parameters, BAN estimators and their properties. Invariance of consistency under continuous transformation. Invariance of CAN estimators under differentiable transformations. methods of generation of CAN estimators for real and vector valued parameters using method of moments and method of percentiles and using central limit theorem. Comparison of consistent estimators. minimum sample size required by the estimator to attain certain level of accuracy. examples. Maximum Likelihood Estimation and their large sample properties. restricted parameter space. Inconsistent MLEs. MLEs in irregular cases. Asymptotic distribution of MLE in special class of distributions: Cramer regularity conditions. Special cases such as exponential class of densities and multinomial distribution. CAN estimator for one parameter Cramer family. Cramer-Huzurbazar theorem. Solutions of likelihood equations. method of scoring. Fisher lower bound to asymptotic variance. MLE in Pitman family and double exponential distribution. MLE in censored and truncated distributions. Asymptotic distribution and properties of Likelihood ratio test. Rao's score test and Wald's test in simple hypothesis case.

Tests based on MLEs. Likelihood ratio tests. asymptotic distribution of log likelihood ratio.

Wald Test; Score Test, locally most powerful tests. Pearson's chi-square test and LR test. Consistent Test. Applications to categorical data analysis. three dimensional contingency tables

Unit 2

Testing of Hypothesis (Multi-parameter Case): Similar tests. Neyman Structure. UMPU tests for composite hypotheses. Invariance tests and UMP invariant tests. Likelihood ratio test. Asymptotic distribution of LR I statistic. Consistency of large sample test. Asymptotic power of large sample test

Unit 3

Sequential Analysis: Sequential tests-SPRT and its properties. Wald's fundamental identity. OC and ASN functions. Sequential estimation. Optimality of SPRT(under usual approximation)

Unit 4

Non- parametric Inference: Non-parametric Estimation. U-statistics and their asymptotic properties. UMVU estimator. Rosenblatt's naïve density estimation. its bias and variance. consistency of kernel density estimators and its MSE. nonparametric tests-single sample location. location-cum-symmetry. One-sample problem of location. Sign test. Wilcoxon Sign Rank test. run test. Kolmogorov-Smirnov test. randomness and goodness of fit problems; Rank order statistics. Linear rank statistics. Asymptotic relative efficiency. Two sample problem of location. Wilcoxon signed rank test for pair comparison. Wilcoxon Mann-Whitney test. Kolmogorov-Smirnov test (Expectation and variance of these test statistics, except for Kolmogorov–Smirnov test, statement about their exact and asymptotic distribution). Wald-Wolfowitz Runs test and Normal Sores test. Chi-square test of goodness of fit and independence in contingency tables. Tests for independence based on Spearman's rank correlation and Kendall's tau. Ansari-Bradley test for two-sample dispersions. Kruskal-Wallis test for one-way layout (k-samples). Friedman test for two way layout(randomised block)

ARE and Pitman theorem. ARE of one sample, paired sample and two sample location tests. The concept of Rao's second order efficiency and Hoges-Lehman's deficiency with examples

Unit 5

Bayesian Inference: Elements of decision theory-Preliminary ideas of decision theory and Bayesian overview and comparisons of two-paradigms- Classical statistical analysis and Bayesian analysis. Relative advantages and disadvantages. Motivation for choices of different priors. Family of conjugate prior distributions.

posteriordistributions, loss function, principles of minimum expected posterior loss, quadratic and other loss functions, conjugate prior distribution, asymptotic normality of posterior distributions, common examples. Bayesian HPD confidence Intervals. Bayesian Estimation, Bayes and Minimax decision functions with their application to estimation with quadratic loss functions, testing, interval estimation and prediction for some common models and common priors. Hierarchical Bayes, Brief discussion on Bayesian computational techniques and their properties.

Book Recommended:

1. Bartoszynski, R. and Bugaj, M.N. (2007). *Probability and Statistical Inference*, John Wiley & Sons.
2. Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Second Edition, Springer - Verlag.
3. BradelyEfronand Robert J Tibshirani: *An Introduction to the Bootstrap*. Chapman and Hall
4. Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series, Second Edition.
5. Conover, W. J.: *Practical Non-parametric Statistics for Behavioural Science*. Mc. Graw Hill
6. Cramer, H.(1974). *Mathematical Methods in Statistics*. Princeton Univ. Press.
7. Dudewicz, E. J. and Mishra, S.N.(1988). *Modern Mathematical Statistics*. John Wiley.
8. Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*. Cambridge University Press
9. Ferguson, T.S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press.
10. Ferguson, T.S. (1996). *A Course in Large Sample Theory*. Chapman and Hall.
11. Ghosh, J., Delampady M. and Samanta T.: *Bayesian Inference*
12. Ghosh, B. K.: *Sequential Tests of Statistical Hypotheses*
13. Gibbons (1978). *Non-Parametric Statistical Inference*
14. Gibbons, J.D. and Chakraborti, S. (1992). *Nonparametric Statistical Inference*, Marcel Dekker.
15. Gray, Schucany and Watkins. *Generalized Jakknife*, Dovenpul
16. Gupta Anirban Das (2008), *Asymptotic Theory of Statistics and Probability*, Springer
17. Kale, B.K. &Muralidharan, K. (2015) *Parametric Inference: An Introduction*, Alpha Science International Ltd.
18. Kale, B.K. (2005). A First Course on Parametric Inference. Second Edition. Narosa.
19. Lee, P.: Bayesian Statistics- An Introduction
20. Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*. Springer, New York
21. Lehmann, E. L. and Romano, J. (2005). *Testing Statistical Hypotheses*. Springer
22. Lehmann, E. L. *Large Sample Theory*
23. Mukhopadhyay, N. & de Silva, B. M. *Sequential Methods and their Applications*
24. Myles Hollander and Douglas A Wolfe. *Non-parametric Statistical Methods* (John Wiley and Sons)
25. Randles, R.H. and Wolfe, D.S. (1979). *Introduction to the Theory of Non-parametric Statistics*. John Wiley & Sons.
26. Rao, C. R.(1995). *Linear Statistical Inference and its Applications*. Wiley Eastern Ltd.
27. Rohatgi, V.K. and Saleh, A.K. Md. E. (2005). *An Introduction to Probability and Statistics*. Second Edition, John Wiley.
28. Roussas, G. G. (1973). *First Course in Mathematical Statistics*. Addison Wesley, Ltd., New Delhi.
29. Serfling, R. J.: *Approximation Theorems of Mathematical Statistics*
30. Silvey, S. D. (1975). *Statistical Inference*, Chapman and Hall.
31. Sinha, S. K. (1986). *Probability and Life Testing*, Wiley Eastern Ltd.
32. Srivastava, M. K. & Srivastava, N. (2014). *Statistical Inference: Testing of Hypotheses*. PHI Learning, Second Edition, New Delhi.
33. Srivastava, M. K. & Srivastava, N. (2018). *Statistical Inference: Theory of Estimation*. PHI Learning, Second Edition, New Delhi.
34. Shao, J. (2003). *Mathematical Statistics*, Springer-Verlag, New, New York.
35. Wald, A. *Sequential Analysis*, Dover Publication
36. Wilks, S. S. (1962). *Mathematical Statistics*, John Wiley.
37. Zacks, S. (1971). *Theory of Statistical Inference*, John Wiley & Sons

Operations Research

Course No. STAT–302

Unit 1

Definitions of Operation Research, History of Operations Research, Scope of Operations Research, Models in Operations Research.

Hyperplane, Convex Sets, Convex Functions, Convex Null, Local and Global Extrema.

Unit 2

Introduction to Linear Programming Problem, Mathematical Formulation of L.P.P., Graphical Method, Simplex Method, Big-M Method, Duality in L.P.P, Integer Programming Problem.

Transportation Problem, Unbalanced Transportation Problem, Degeneracy and Its Solution, Assignment Problem, Unbalanced Assignment Problem, Travelling Salesman Problem.

Unit 3

Theory of Games, Two-Person Zero Games, Minimax Rule, Solution of Games Without Saddle Point, Graphical Method, Dominance Method.

Inventory Control—Meaning and Importance, Various Costs Involved in Inventory Control, Deterministic Models, Multi-item Deterministic Models With Restrictions, Probabilistic Inventory Models, Models With Lead time, Inventory Models with Price Breaks.

Unit 4

Essential Features of Queuing System, Steady State, Transient State, Distribution of Arrivals, Inter-Arrivals and Waiting Time, Queuing Models M/M/1: (∞/FIFO), M/M/1: (N/FIFO), M/M/C: (∞/FIFO).

Unit 5

Replacement, Simulation, Sequencing Problem—*n* Jobs on Two Machines, *n* Jobs on Three Machines, *n* Jobs on *m* Machines. Mrkov Chains, Dynamic Progra-mming, PERT/CPM..

Books Recommended :

1. Operations Research—B.S. Goel and S.K. Mittal
2. Operations Research—K. Swarup, P.K. Gupta and M.Mohan
3. Operations Research—S.D. Sharma
4. Operations Research : Theory and application—J.K. Sharma
5. Operations Research : An Introduction—H.A. Taha
6. Linear Programming—G. Hedley

Data Mining

Course No. STAT–303(E)

Unit 1

Introduction, Relational Databases, Data Warehouse, Transactional Databases, Advanced Data and Information Systems and Advanced Applications. Data Mining Functionalities. Concept/Class Description: Characterization and Discrimination. Mining Frequent Patterns, Associations, and Correlations. Classification and Prediction. Cluster Analysis, Outlier Analysis, Evolution Analysis. Classification of Data Mining Systems. Data Mining Task Primitives. Integration of a Data Mining System with a Database or Data Warehouse System. Major Issues in Data Mining.

Unit 2

Data Preprocessing.Descriptive Data Summarization: Measuring the Central Tendency. Measuring the Dispersion of Data, Graphic Displays of Basic Descriptive Data Summaries. Data Cleaning: Missing Values. Noisy Data, Data Cleaning as a Process. Data Integration and Transformation: Data Integration. Data Transformation. Data Reduction: Data Cube Aggregation. Attribute Subset Selection. Dimensionality Reduction. Numerosity Reduction. Data Discretization and Concept Hierarchy Generation: Discretization and Concept Hierarchy Generation for Numerical Data. Concept Hierarchy Generation for Categorical Data.

Unit 3

Mining Frequent Patterns, Associations, and Correlations.Market Basket Analysis: Frequent Itemsets. Closed Itemsets and Association Rules, Frequent Patterns Mining: Efficient and Scalable Frequent Itemset Mining Methods: The Apriori Algorithm: Finding Frequent Itemsets Using Candidate Generation. Generating Association Rules from Frequent Itemsets. Improving the Efficiency of Apriori. Mining Frequent Itemsets without Candidate Generation, Mining Frequent Itemsets Using Vertical Data Format. Mining Closed Frequent Itemsets. Mining Various Kinds of Association Rules: Mining Multilevel Association Rules. Mining Multidimensional Association Rules from Relational Databases and Data Warehouses. From Association Mining to Correlation Analysis: From Association Analysis to Correlation Analysis. Constraint-Based Association Mining: Metarule-Guided Mining of Association Rules. Constraint Pushing: Mining Guided by Rule Constraints.

Unit 4

Issues Regarding Classification and Prediction: Preparing the Data for Classification and Prediction. Comparing Classification and Prediction Methods. Classification by Decision. Tree Induction: Decision Tree Induction. Attribute Selection Measures. Tree Pruning. Scalability and Decision Tree Induction. Bayesian Classification : Bayes' Theorem, Naïve Bayesian

Classification. Bayesian Belief Networks. Training Bayesian Belief Networks. Rule-Based Classification: Using IF-THEN Rules for Classification, Rule Extraction from a Decision Tree. Rule Induction Using a Sequential Covering Algorithm.Associative Classification: Classification by Association Rule Analysis. *k*-Nearest-Neighbor Classifiers. Case-Based Reasoning.

Unit 5

Prediction: Linear Regression, Nonlinear Regression. Other Regression-Based Methods. Accuracy and Error Measures: Classifier Accuracy Measures. Predictor Error Measures. Evaluating The Accuracy of a Classifier or Predictor: Holdout Method and Random Subsampling. Cross-Validation. Bootstrap. Ensemble Methods—Increasing the Accuracy: Bagging, Boosting. Model Selection: Estimating Confidence Intervals. ROC Curves.

Books Recommended:

1. Data Mining: Concepts & Techniques (Second Edition,Jiawei Han &MichelineKamber (Morgan Kaufman Publisher, 2006)
2. Principles of Data Mining, David Hand, HeikkiMannila, Padhraic Smyth (Printice Hall of India Private Limited, 2007)
3. Data Mining, VikramPudi& P. Radha Krishna(Oxford University Press. 2009)

Population Studies

Course No. STAT–304(E)

Unit 1

Definition and scope of demography. Demography, demographic analysis and population studies. Sources of demographic data: census, registration and sample surveys. Errors in census and vital statistics and their adjustments, Whipples and Myer's indices. Chandrasekharan Deming formula to check completeness of registration data.

Unit 2

Fertility, its measures. Measures of reproduction and replacement. Fertility models. Distribution of time of first birth/conception, number of births/ conceptions in a specified time: inter-live birth intervals (for both homogeneous and non-homogeneous groups of women), estimation of parameters.

Unit 3

Mortality, its measures. Standerdised death rates. Life table, its type. Construction of complete and abridged life tables. Makeham's and Gompertz curve.

Unit 4

Migration rates and ratios. Method to estimate intercensal migration using vital statistics, survival ratio and growth rate. Migration models.

Unit 5

Theory of stable population, quasi and stationary population. Stochastic models of population growth. Growth curves and methods of their fitting. Population estimates and projection. Component method of population projection.

Books Recommended :

1. Benjamin, B.(1969). Demographic Analysis. George , Allen And Unwin.
2. Biswas, S. (1988). Stochastic Processes In Demography And Applications. Wiley Eastern, New Delhi.
3. Chiang, C.L.(1968). Introduction To Stochastic Processes In Biostatistics, John Wiley, New York.
4. Cox, P.R. (1970). Demography, Cambridge University Press.
5. Keyfitz, N.(1977). Appplied Mathematical Demography, Springer Verlag.
6. Kumar,R. (1986). Technical Demography, Wiley Eastern Ltd.
7. Pathak, K.B. and Ram, F. (1992). Techniques Of Demographic Analysis, Himalayan Publishing House, Bombay.
8. Shryock,H.S. (1976). The Methods And Materials Of Demography, Academic Press , New York.
9. Spiegelman, M. (1969). Introduction To Demographic Analysis. Harvard University Press.
10. Wolfenden,H.H. (1954). Population Statistics And Their Compilation. American Acturial Society.

Official Statistics

Course No. STAT–306(E)

Unit 1

Introduction to Indian and International Statistical System, Methods of Collection of Official Statistics, Their Reliability and Limitations. Role, Functions and Activities of Central and State Statistical Organisations, Organisation of Large Scale Sample Surveys.

Unit 2

Role of national Sample Survey Organisation, Other Agencies Responsible for Data Collection and Their Main Functions, Principal Publications on various Topics of Data Collection, Scope and Contents of Population Census of India.

Unit 3

Population Statistics, Agricultural Statistics, Medical Statistics, Industrial Statistics, Trade Statistics, Price Statistics.

Unit 4

Statistics of Labour & Employment, Statistics of Transport and Communication, Financial and banking Statistics, Miscellaneous Statistics.

Unit 5

National Income and Its Computation, Utility and Difficulties in Estimation of National Income.

Books Recommended :

1. Basic Statistics Relating to Indian Economy (CSO) 1990
2. Guide to Official Statistics (CSO) 1999
3. Statistical System in India (CSO) 1995
4. Principles and Accommodation of National Population Censuses (UNESCO)
5. National Accounts Statistics—Sources and Health (CSO) 1980

Econometrics

Course No. STAT–307(E)

Unit 1

Definitions of Econometrics, Scope of Econometrics, Division of Econometrics, Models and Methodology of Econometric Research.

Unit 2

Two Variable Linear Model, Least Square and Maximum Likelihood Estimation, Properties of Estimators, Tests of Significance and Confidence Intervals of Parameters, Analysis of Variance in Regression, Extension of Two Variable Model.

Unit 3

General Linear Model, Least Square and M.L.E. Estimation, Properties, Significance Tests and Confidence Intervals, Generalized Least-Squares (AITKEN) Estimator.

Unit 4

Nature of Heteroscedasticity, Consequences, Detection and Remedial Measures of Heteroscedasticity, Introduction of Auto Correlation, OLS Estimation in Presence of Auto Correlation, Detection and Remedial Measures of Auto Correlation.

Unit 5

Nature of Multi Collinearity, Practical Consequences, Detection and Remedial Measures of Multicollinearity, Simultaneous Equalition Systems.

Books Recommended :

1. Basic Econometrics—D.N. Gujarati
2. Econometric Methods—J. Johnston
3. Theory of Econometrics—A. Koutsoyiannis

Decision Theory and Bayesian Inference

Course No. STAT–401(E)

Unit 1

Decision Problem and 2-Person Game, Utility Theory, Loss Functions, Expected Loss, Decision Rules (Non-Randomized and Randomized), Decision Principles (Conditional Bayes, Frequentist), Inference Problems as Decision Problems, Optimal Decision Rules. Concepts of Admissibility and Completeness, Bayes Rules, Admissibility of Bayes Rules.

Unit 2

Supporting and Shyperplane Theorems, Minimax Theorem of for Finite Parameter Space, Minimax Estimators of Normal and Poisson Means, Admissibility of Minimax Rules.

Unit 3

Subjective Interpretation of Probability in Terms of Fair Odds. Evaluation of (i) Subjective Probability of an Event Using a Subjectively Unbaised Coin (ii) Subjective Prior Distribution of a Parameter. Bayes Theorem and Computation of the Posterior Distribution.

Unit 4

Bayesian Point Estimation : As a Prediction Problem from Posterior Distribution. Bayes Estimators for (i) Absolute Error Loss (ii) Squared Error Loss (iii) 0-1 Loss. Generalization to Convex Loss Functions. Evaluation of the Estimate in Terms of the Posterior Risk.

Bayesian Interval Estimation : Credible Intervals. Highest Posterior Density Regions. Interpretation of the Confidence Coefficient of an Interval and Its Comparison with the Interpretation of the Confidence Coefficient for a Classical Confidence Interval.

Unit 5

Bayesian Testing of Hypothesis : Specification of the Appropriate Form of the Prior Distribution for a Bayesian Testing of Hypothesis Problem. Prior Odds, Posterior Odds. Bayes Factor for Various Types of Testing Hypothesis Problems Depending Upon the Null Hypothesis and the Alternative Hypothesis are Simple or Composite.

Books Recommended :

1. Berger, J. O. (1985). Statistical Decision Theory and Bayesian Analysis. 2nd Ed. Springer.
2. Ferguson, T. S. (1967). Mathematical Statistics - A Decision Theoretic Approach, Academic Press.
3. Leonard T. and Hsu, S.J. Bayesian Methods. Cambridge University Press.

Advanced Sample Surveys

Course No. STAT – 402(E)

Unit 1

Fixed Population Approach :

a) Unified Theory of Sampling : Inference Aspect of Sample Survey. Sampling Designs andSampling Schemes. Correspondence; Classes of Estimators. Homogeneous Linear Estimators and Condition of Unbiasedness; Godambe's UMY Non-Existence and Existence Theorems Under Restricted Conditions Relating to Labelled Populations. Basu's Difference Estimator. Sufficiency. Completeness. Aadmissibility of Estimators inSurvey Sampling. Optimal Estimators in Usful Sampling Designs.

b) Unbiased Ratio and Regression Type Estimators. Multivariate Ratio and Regression Methods of Estimation. Product Estimator. Optimum Properties of Ratio and Regression Estimators. Regression Analysis and Categorical Data Analysis With Data From Complex Surveys. Bias Adjustment in Ratio Estimator Due to Murthy. Beale and Tin. JackknifeRatio Estimator. Olkin's Multivariate Ratio Estimator. Self Weighting Designs.

c) Integration of Different Principles and Methods of Sampling in Adopting CompositeSampling Procedures in Actual Practice. Integration of Surveys-Lahiri and Keyfitz's Procedures. Variance Estimation with Complex Designs. Taylor's Series Linearization. Balanced Repeated Replication. Jackknife and Bootstrap Methods.

Unit 2

Model Based Approach and Prediction Approach : Inference Under Superpopulation Model. Concept of Designs and Model Unbiased Estimation. Traditional Model-Based and Optimal Estimators Under Various Useful Sampling Designs. Prediction Approach. Predicting a Super Population Mean.

Unit 3

Bayesian Theories in Finite Population : Non-Informative Bayesian Approach. Extension of Polya Posterior. Empirical Bayes Estimation. Estimation of Stratum Means. Hierarchical Bayes Estimation.

Unit 4

Small Area Estimation : Small Area Estimation-Direct Estimators. Synthetic Estimators. Composite Estimators. Repeated Sampling. Balanced Repeated Replication. Jackknife and Bootstrap Methods. Calibration Approach : Introduction to Calibration Estimators. Calibration Estimators Based on Functional Form. with Restricted Weights. Robustness Aspects. Extended Calibration Estimators. Cosmetic and Calibration Estimators. Model Based Calibration Estimators. Estimation of Distribution Function and Quadratic Finite Population Function.

Unit 5

Large-Scale Surveys : Organizational Aspects of Planning Large-Scale Sample Surveys. NonSampling Errors. Non-Response. Familiarity with NSS Work and Some Specific Large-Scale Surveys. Special Topics Include Wildlife Surveys. Non-Sampling Error Adjustment. Categorical Data Analysis and Practical Survey Examples.

Books Recommended:

1. Cochran. W.G. (1977): Sampling Techniques
2. Des Raj and Chandak (1999): Sampling Theory
3. Mukhopadhyay. P. (1998): theory and Methods of Survey Sampling
4. MukhopadhyayP. (2007): Survey Sampling
5. Sarndal. C.E. and Swensson. B. and Wretman. J.H. (1992) : Model Assisted SurveySampling
6. Sukhatme. P.V. and Sukhatme. B.V. (1992) : Sampling Theory of Surveys With Applications
7. Meeden. G. Ghosh. Malay (1997) : Bayesian Methods in Finite Population Sampling
8. Cassel. Sarndal. Wretman (1977): Foundations of Inference in Survey Sampling

Limit Theorems and Stochastic Processes

Course No. STAT–404(E)

Unit 1

Sequences of Independent Events, Continuity of Probability Functions, First Borel-Contelli Lemma, Second Borel-Contelli Lemma, Zero-One Law.

Unit 2

Probability Inequalities, Chebshev's Law, Sequences of Random Variables, Convergence in Distribution, Convergence in Probability, Mean Convergence, Strong Convergence, Cauche Convergence Criterion.

Unit 3

Weal Law of Large Numbers, Some Variations of Weak Law of Large Numbers, Weak Law of Large Numbers for Bounded Variates, Strong Law of Large Numbers, Central Limit Theorems of Lindberg Levy, De-Moivre, Lindberg-Fellor, Lyapounov, Cramer's Theorem.

Unit 4

Notion Of Stochastic Processes, Specification Of Stochastic Processes, Processes With Independent Increments, Marcov Processes, Stationary Processes, Guassian Processes. Marcov Chain, Transition Matrix, Higher Transition Probabilities, Classification Of States And Chains, Limiting Behaviour, Statistical Inference For Marcov Chains, Marcov Chain With Continuous State Space.

Unit 5

Poisson Process and Its Properties, Pure Birth Process, Yule Process, Death Process, Birth and Death Process, Application to Queues, Random Walk, Gamblers' Ruin Problem.

Books Recommended :

1. The Theory of Probability—B. Gnedenko
2. Probability Theory—M. Loeve
3. Probability Theory and Mathematical Statistics—M. Fisz
4. Modern Probability Theory and Its Applications—E. Parzen
5. An Introduction to Probability Theory—V.K. Rohatgi
6. Stochastic Process—J. Medhi
7. Stochastic Models: Analysis and Applications—B.R. Bhat
8. An Introduction To Finite Marcov Processes—S.R. Adke
9. Stochastic Processes—E. Parzen
10. First Course in Stochastic Processes—S. Karlin
11. Introduction to Stochastic Process—E. Cinlar

Research Methodology

Course No. STAT–408(E)

Unit 1

Introduction to Research Methods : Definition of Research, Role and Objectives of Research Applications and Types of Research, Research Process and Steps in it. Collecting and Reviewing the Literature, Conceptualization and Formulation of a Research Problem, Identifying Variables, Constructing Hypothesis.

Unit 2

Research Design : Selecting and Defining a Research Problem, Need for Research Design, Features of a Good Research Design, Different Research Designs (Exploratory, Descriptive Experimental and Diagnostic Research).

Design of sample Survey : Census v/s Sample Enumerations, Objectives and Principles of Sampling, Types of Sampling, Sampling and Non-Sampling Errors. Designing Questionnaires and Interview. Determination of the Sample Size.

Unit 3

Measurement of Scaling Concepts : Scales of Measurements, Nominal, Ordinal, Internal and Ratio Scales. Errors in Measurements. Validity and reliability in Measurement, Scale Construction Techniques.

Unit 4

Data Collection & Analysis : Primary and Secondary Data, Validity and reliability of Data Collection Procedures, Data Preparation, Exploratory Data Analysis, Parametric and non-parametric Tests, Correlation and Regression Analysis, ANOVA.

Unit 5

Report Writing : Discussions, Conclusion, Referencing and Various Formats for Reference Writing, Bibliography, Formats of Publications in Research Journals including Subject Classification, Impact Factor, Citation Index.

Books Recommended :

1. Management Research Methodology—K.N. Krishna Swamy, A.I. Shiv Kumar
2. Research Methodology : Methods and Techniques—C.R. Kothari
3. Research Methodology : A Step by Step Guide for Beginners—Ranjit Kumar
4. Research Methods for Business Students—Mark Saunders, Philip Lewis
5. Research Methods—Ram Ahuja

Marketing Research Statistics

Course No. STAT–409(E)

Unit 1

Nature, Scope and Importance of Marketing, Approaches to the Study of Marketing, Changing Concepts of Marketing, Recent Innovations in Modern Marketing, Marketing Environment and Marketing System. Consumer Decision Behaviour, Consumer Buying Decision Pocess, Theory of Consumer Behaviour, Marketing Segmentation and Marketing Mix, Physical Distribution of Goods, Meaning and Objectives of Advertisement, Characteristics of an Effective Advertisement, Different Media of Advertisement, Choice of Advertising Media, Advertising Budget, Measuring the Effectiveness of Advertising.

Unit 2

Meaning and Nature of Marketing Research, Objective and Functions of Marketing Research, Elements of Marketing Research, Advantages and Limitations of Marketing Research, Process of Marketing Research, Marketing Research Design, Marketing Information System, Sampling Plan for Marketing Surveys.

Unit 3

Introduction to the Theory and Models in Marketing, Science and Marketing Models, Complexity of Marketing Models, Decision Support Models, Theoretical Modeling in Marketing, Purchase Incidence Models, Stochastic Models of Brand Choice.

Unit 4

Introduction to Advertisement and New Product Planning, The Effects of Advertising, Objective Setting and Budgeting—Practice and Models, Media Selection and Scheduling–Modelling Approaches, New Product Planning—Types of New Product Situations, Adoption Process for New Products, Models of First Purchase, repeat Purchase Models of New Products.

Unit 5

Brand Preference Index, Stability Index Based on Duration of Use, Stability Index Based on Quantity Ratio and Duration of Use, Brand Awareness Index, Consumer Satisfaction Measure, etc.

Books Recommended :

1. Marketing Models—Gary L. Lilien, Philip Kotler and K. Sridhar Moorthy
2. Marketing Management—C.B. Gupta
3. Marketing Analysis and Decision Making : Text and Cases—G.C. Darral
4. Advertising Management—A.A. David and J.G. Myers