## UNIT I: Introduction to Statistics

Statistics: Definition and scope. Concepts of statistical population and sample. Methods of sampling: SRS, Stratified, Systematic and Cluster sampling (Definitions only). Data: quantitative and qualitative, cross sectional and time-series, discrete and continuous. Scales of measurement: nominal, ordinal, interval and ratio. Classification, tabulation of data and diagrammatic representation of data. Frequency distributions, cumulative frequency distributions and their graphical representations. Stem and leaf displays.

## UNIT II: Univariate Data Analysis

Measures of Central Tendency: Mean, weighted mean, trimmed mean, Median, Mode. Geometric and harmonic mean, properties, merits and demerits, relation between these measures. Measures of Dispersion: Range, Quartile deviation, Mean deviation, Standard deviation and their relative measures. Moments, Skewness and Kurtosis. Quintiles' and measures based on them.

## Unit-III: Bivariate and Multivariate Analysis

Bivariate Data, Scatter diagram, Correlation, Karl Pearson's correlation coefficient, Rank correlation – Spearman's and Kendall's measures. Simple linear regression and its properties. Fitting of linear regression line and coefficient of determination.

Analysis of Categorical Data: Contingency table, independence and association of attributes, measures of association - odds ratio, Pearson's and Yule's measure. Multivariate Data Visualization: Mean vector and Dispersion matrix, Multiple linear regression (Three variables only), and Residual variance. Multiple and partial correlation coefficients.

## Unit-IV: Elements of Probability

Random experiment, sample space and events, algebra of events. Definitions of Probability- Classical, statistical, subjective and axiomatic approaches – illustrations and applications, Addition rule, Conditional probability, independence of events and multiplication rule, Total probability rule, Bayes theorem- applications.

| | Teaching Hrs |
|---|---|
| **UNIT I: Random Variables and Mathematical Expectation (One Dimension)** | 12 |
| Definitions of discrete and continuous random variables, distribution function, probability mass and density functions – properties and illustrations. Expectation of random variable and its properties. Probability generating function, Moments and moment generating function and cumulant generating function-properties and uses. | |
| **UNIT II: Standard discrete distributions** | 12 |
| Uniform, Bernoulli, Binomial, Poisson, Geometric, Negative Binomial, Hyper geometric distributions, definition, mean, variance, moments, moment generating functions, recurrence relation for probabilities and moments for binomial, Poisson, and Negative binomial distributions, additive property, Cumulant generating function, theoretical examples. | |
| **Unit-III: Standard Uni-variate continuous distributions** | 12 |
| Rectangular, Beta, Gamma, and Exponential distributions, definitions through p.d.f's, Mean, variance, moments, recurrence relations, Additive property of exponential and gamma variates, Normal distribution and its properties, Cauchy distribution, Uni-variate and Bi-variate transformation of variables of discrete and continuous random variables. | |
| **Unit-IV: Data Analysis Using R** | 12 |
| Introduction to R: Installation, command line environment, overview of capabilities, brief mention of open source philosophy. R as a calculator: The four basic arithmetic operations. Use of parentheses nesting up to arbitrary level. The power operation. Evaluation of simple expressions. Quotient and remainder operations for integers. Standard functions, e.g., sin, cos, exp, log. The different types of numbers in R: Division by zero leading to Inf or -Inf. NaN. NA. No need to go into details. Variables. Creating a vector using c(), seq() and colon operator. How functions map over vectors. Functions to summarize a vector: sum, mean, sd, median etc. Extracting a subset from the vector (by index, by property). R as a graphing calculator: Introduction to plotting. Plot(), lines(), abline(). No details about the graphics parameters except colour and line width. Barplot, Pie chart and Histogram. Box plot. Scatter plot and simple linear regression using lm(y~x). Problems on discrete and continuous probability distributions. | |

## DSCT 3.1 Calculus and Probability Distribution

| Content of Theory Paper 3 |
|---|
| **UNIT 1: Calculus of one and more variables** |
| Review of calculus of one variable: continuity, differentiability, mean value theorem and Taylor series expansion. Functions of several variables: Continuity, directional derivatives, differentials of functions of several variables, the gradient vector. The mean value theorem, a sufficient condition for the existence of the differential, partial derivatives of higher order and Taylor's formula. Applications of partial differentiation, Jacobians, integration by parts, Multiple integrals and evaluation of multiple integrals by repeated integration, Beta and Gama integrals. (Only results and applications) |
| **UNIT 2: Distribution of Random Variables (Two-dimensional)** |
| Two dimensional random variables: Joint distribution, Marginal distribution and Conditional distributions of random variables, conditional expectation, covariance, correlation and moments. Distribution of functions of random variables using m.g.f. and distribution function. Transformation of variable technique (one and two variables).<br>Chebyshev's inequality- proof and its use in approximating probabilities; Statements of Weak Law of Large Numbers; Convergence in law and Central Limit theorems – De-Moivre. ( Some simple examples) |
| **UNIT 3: Probability Distributions-II** |
| Discrete distributions: Rectangular, Geometric, Negative Binomial, Hypergeometric- definition through probability mass function, mean, variance, moments, p.g.f., m.g.f., other properties and applications.<br>Continuous distributions: Uniform, Gamma, Exponential, Beta (type 1 and type 2), Cauchy, – definition through probability density function, mean, variance, moments, m.g.f., other properties and applications. |
| **UNIT 4: Sampling Distributions and Simulation** |
| Definitions of random sample, parameter and statistic, sampling distribution of sample mean, standard error of sample mean.<br>Exact sampling distributions: Chi square distribution- mean, variance, moments, mode, additive property. Student's and Fisher's t-distribution- mean, variance, moments and limiting form of t distribution. Snedecor's F-distribution: mean, variance and mode. Distribution of 1/F. Relationship between t, F and $\chi^2$ distributions.<br>Introduction to simulation. Generation of random observations from Uniform, Exponential, Normal, Binomial, Poisson distributions using R-codes. |

**Contents:**

**Unit 1: Economic Statistics** **12Hours**

**Index numbers**: Definition, Criteria for a good index number, different types of index numbers Construction of index numbers of prices and quantities, consumer price index number Uses and limitations of index numbers Consumer price index number construction of consumer price index numbers. Applications of consumer price index numbers

**Time Series Analysis**: Components of time series, Decomposition of time series- Additive and multiplicative model with their merits and demerits, Illustrations of time series. Measurement of trend by method of free-hand curve, method of semi-averages and method of least squares (linear). Measurement of seasonal variations by method of ratio totrend.

**Unit 2: Vital Statistics** **10Hours**

Sources of demographic data, errors in data.
Measurement of mortality: crude death rate, specific death rates, and standardized death rates, infant mortality rate, maternal mortality rate, neo natal mortality rates, merits and demerits and comparisons of various mortality rates.

Measurement of Fertility and Reproduction: Fecundity, fertility, measurement of fertility, crude birth rate, general fertility rate, age specific fertility rate and total fertility rates, merits and demerits of each measure of fertility, comparative study of these measures of fertility, Growth rates: Gross reproduction rate and Net reproduction rates.

**Unit 3: Sampling Theory** **10Hours**

Population and Sample. Need for sampling, Complete Enumeration versus Sample Surveys, Merits and Demerits, Non – Probability and Probability Sampling, Need and illustrations. Use of random numbers, Principal steps in sample survey.Requisites of a good questionnaire. Pilot surveys, Sampling and non – sampling errors, Description of SRS, simple random sampling with and without replacement procedures, Merits and demerits of Simple random sampling.

Need for stratification, stratifying factors, Merits and demerits of stratified random sampling. Systematic random sampling procedure of obtaining sample, Merits and demerits of systematic random sampling.

# DSCT 4.1:Statistical Inference-I

| Content of Theory Paper 4 | 56 Hrs |
|---|---|
| **UNIT- 1: Point Estimation-I** | **16 Hrs** |
| Families of distributions- location and scale families. Single parameter exponential family. Concept of order statistics, Distribution of maximum and minimum order statistics (with proof ) and $r^{th}$ order statistic (without proof). Concepts of estimator and estimate. Criteria for estimators: Unbiasedness, Consistency. Invariance property of consistent estimators. Efficiency and relative efficiency. Mean squared error as a criterion for comparing estimators. Sufficient statistics. Statement of Neyman-Factorization theorem. | |
| **UNIT-2: Point Estimation-II** | **12 Hrs** |
| Fisher information function. Statement of Cramer–Rao inequality and its applications. Minimum Variance Unbiased Estimator and Minimum Variance Bound Estimator. Maximum likelihood and method of moment estimation; Properties of MLE and moment estimators and examples. Method of Scoring | |
| **UNIT- 3: Testing of Hypotheses** | **18 Hrs** |
| Statistical hypotheses - null and alternative, Simple and composite hypotheses. Type-I and Type-II errors, test functions. Randomized and non-randomized tests. Size, level of significance, Power function, power of tests. Critical region, p- value and its interpretation. Most Powerful (MP) and UMP test. Statement of Neyman-Pearson Lemma and its applications. Likelihood ratio tests. Large and small samples tests of significance. Tests for single mean, equality of two means, single variance and equality of two variances for normal populations. Tests for proportions. | |
| **UNIT- 4: Interval Estimation** | **10 Hrs** |
| Confidence interval, confidence coefficient, shortest confidence interval. Methods of constructing confidence intervals using pivotal quantities. Construction of confidence intervals for mean, difference of two means, variance and ratio of variances, proportions, difference of two proportions and correlation coefficient. | |

## References


1. Chihara, L. and Hesterberg, T. (2011) Mathematical Statistics with Resampling and R

| | | | |
|---|---|---|---|
| Course Title | | | |
| Course Code | DSCT 5.2-1 | No. of Credits | 4 |
| Contact hours | 60 Hours | Duration of SEA Exam | 2 hours |
| Formative Assessment Marks | 40 | Summative Assessment Marks | 60 |

**Course Pre-requisite(s):**

**Course Outcomes (COs):** After the successful completion of the course, the student willbe able to

CO1: Learn fixed and random effect models and one-way and two-way classified data

CO2 Understand different designs (CRD, RBD, LSD) and missing plot techniques

CO3: Understand the different factorial experiments

CO4: Develop complete and partial confounding for factorial experiments

## CONTENTS

**60 Hrs**

### UNIT 1: ANALYSIS OF VARIANCE — 15 Hrs

Meaning and assumptions. Fixed and random effect models. Analysis of One way and two way classified data with and without interaction effects. Multiple comparison tests: Least Significance difference and Duncan's multiple range test.

### UNIT 2: EXPERIMENTAL DESIGNS — 15 Hrs

Principles of design of experiments. Completely randomized, randomized block and Latin square designs (CRD, RBD, LSD) layout formation and the analysis using fixed effect models. Comparison of efficiencies of CRD, RBD and LSD. Estimation of one and two missing observations in RBD and LSD and analysis.

### UNIT 3: Operations Research — 15 Hrs

Introduction to Operations Research, phases of O.R., model building, various types of O.R. problems. Linear Programming Problem, Mathematical formulation of the L.P.P, graphical solutions of a L.P.P. Simplex method for solving L.P.P.

### UNIT 4: Transportation Problem and Assignment problem — 15 Hrs

Transportation Problem. Initial solution by North West corner rule, Least cost method and Vogel's approximation method (VAM), MODI's method to find the optimal solution, special cases of transportation problem. Assignment problem. Hungarian method to find optimal assignment, special cases of assignment Problem.

| | | | |
|---|---|---|---|
| Program Name | **B.Sc in STATISTICS** | Semester | **V** |
| Course Title | **Matrix algebra and regression analysis (Theory)** | | |
| Course Code: | **DSCT 5.1-T** | | |
| Contact hours | **60 Hours** | No. of Credits | **04** |
| Formative Assessment Marks | **40** | Duration of SEA/Exam | **2 hours** |
| **Course Pre-requisite(s):** | | Summative Assessment Marks | **60** |

**Course Outcomes (COs):** After the successful completion of the course, the student will be able to:

CO1. Demonstrate and understanding of basic concepts of matrix algebra, including determinants, inverse and properties of various types of matrices.

CO2. Apply matrix algebra and linear algebra techniques to solve systems of linear equations, determine the rankof matrix, understanding quadratic forms and their applications in statistics, characteristic roots and vectors.

CO3. Develop and understanding of simple and multiple regression models, including the assumptions underlying these models, techniques for inference and hypothesis testing and diagnostics checks and corrections.

CO4. Apply regression analysis techniques to real world data sets.

**Contents**

| **Unit 1: Algebra of matrices and determinants** | **60 Hrs** |
|---|---|
| A review of matrix algebra, theorems related to triangular, symmetric and skew symmetric matrices, idempotent matrices, orthogonal matrices, singular and non-singular matrices and their properties. Trace of a matrix, unitary matrices. Adjoint and inverse of a matrix and related properties. Determinants ants of Matrices: Definition, properties and applications of determinants for 3rd and higher orders, evaluation of determinants of order 3 and more using transformations. Symmetric and Skew symmetric determinants. Jacobi's Theorem, product of determinants. | 15 Hrs |
| **Unit 2: Linear Algebra** | |
| Linear algebra: Use of determinants in solution to the system of linear equations, row reduction and echelon forms, the matrix equations AX=B, solution sets of linear equations, linear independence, Applications of linear equations. Inverse of a matrix. Rank of a matrix, row-rank, column-rank, standard theorems on ranks, rank of the sum and the product of two matrices. Characteristic roots and Characteristic vector. Properties of characteristic roots, Cayley Hamilton theorem, Quadratic forms, nature of quadratic form and properties. Linear orthogonal transformation and their digitalization. | 15 Hrs |
| **Unit 3: Simple linear regression** | |

| | |
|---|---|
| Assumptions, inference related to regression parameters, standard error of prediction, tests on intercepts and slopes, extrapolation, diagnostic checks and correction: graphical techniques, tests for normality, uncorrelatedness, homoscedasticity, lack-of-fit testing, transformations on Y or X (Box- Cox, square root, log etc.), method of weighted least squares, inverse regression. | |
| **Unit 4: Multiple linear regression** | 15 Hrs |
| Standard Gauss Markov setup, Gauss-Markov theorem (without proof), least squares (LS)estimation, variance-covariance of LS estimators, estimation of error variance, LS estimation withrestriction on parameters. Simultaneous estimation of linear parametric functions. Tests of hypotheses for one and more than one linear parametric functions, confidence intervals, Variable selection problems. | |

| | | | |
|---|---|---|---|
| Program Name | **BSc in STATISTICS** | Semester | **VI** |
| Course Title | **Statistical Inference – II (Theory)** | | |
| Course Code | **DSCT 6.1-T** | No. of Credits | **4** |
| Contact hours | **60 Hours** | Duration of SEA/Exam | **2 hours** |
| Formative Assessment Marks | **40** | Summative Assessment Marks | **60** |

**Course Pre-requisite(s):**

**Course Outcomes (COs):** After the successful completion of the course, thestudent will be able to:

CO1. Understand expected loss, decision rules, decision principles and Bayes and minimax decision rule.

CO2. Learn about UMP test, MLR property and Likelihood ratio tests.CO3. Explore about sequential inference.

CO4. Learn about one sample and two sample nonparametric tests.

| **Contents** | **60 Hrs** |
|---|---|
| **Unit-1: Statistical Decision Theory** | 15 Hrs |
| Basic elements of Statistical Decision Problem. Expected loss, decision rules (nonrandomized and randomized), decision principles (conditional Bayes, frequentist), inference as decision problem, Loss function, squared error loss, Bayes and minimax decision rule. | |
| **Unit-2: Testing of Hypothesis-II** | 15 Hrs |
| Definition of UMP test, monotone likelihood ratio (MLR) property, Examples of distributions having MLR property, Construction of UMP test using MLR property. UMP test for single parameter exponential family of distributions. Likelihood ratio (LR) tests, LR test for normal, exponential. | |
| **Unit -3: Sequential Inference** | 15 Hrs |
| Need for sequential analysis, Wald's SPRT, OC and ASN functions, examples based on Bernoulli, Poisson, Normal and exponential distributions. | |
| **Unit-4: Nonparametric tests** | 15 Hrs |
| Nonparametric and distribution-free tests, one sample problems: Sign test, Wilcoxon signed rank test, Kolmogorov-Smirnov test. Test of randomness using run test. General two sample problems: Wolfowitz runs test, Kolmogorov Smirnov two sample test (for sample of equal size), Median test, Wilcoxon-Mann-Whitney U-test. Several sample problems:Friedman's test, Kruskal Wallis test | |

| | |
|---|---|
| Simple random sampling with and without replacement, definition, and procedure of selecting a sample, estimates of population mean, total and proportion, variances and SE of these estimates, estimates of their variances related proofs, sample size determination. | 15 Hrs |
| **Unit 3: Stratified sampling and systematic sampling**<br>Stratification and its benefits, basis of stratification, Technique, estimates of population mean and total, variances of these estimates, proportional, optimum allocations, Neyman's allocation, allocation with cost functions and their comparison with SRS. Practical difficulties in allocation, derivation of the expressions for the standard errors of the above estimators when these allocations are used, estimation of gain in precision, post stratification and its performance.<br>Systematic Sampling: Linear systematic sampling Technique; estimates of population mean and total, variances of these estimates (N=n x k).<br>Comparison of systematic sampling with SRS and stratified sampling in the presence of linear trend and corrections. | |
| **Unit-4: Official Statistics** | 15 Hrs |
| Present official statistical system in India, Methods of collection of official statistics, De-juro and De-facto method and their reliability and limitations. Role of Ministry of Statistics & Program Implementation (MoSPI), concept of Central Statistical Office (CSO), National Sample Survey Office (NSSO), and National Statistical Commission. Government of India's Principal publications containing data on the topics such as population, industry, Health, Economy, finance & Agriculture, RBI.<br>International Data Sources: World Bank, WHO. | |

CO6. Understand the role statistics in national development.

| Contents | 60 Hrs |
|---|---|
| **Unit 1: Introduction to sampling theory** | 15 Hrs |
| Objectives and principles of sampling theory; Concept of population and sample; complete enumeration versus sampling; Planning, execution and analysis of a sample survey; practical problems at each of these stages; basic principle of sample survey; sampling and non-sampling errors; Types of sampling; non-probability and probability sampling, pilot survey. | |
| **Unit 2: Simple random sampling** | 15 Hrs |