GOVERNMENT DEGREE COLLEGE FOR WOMEN (A)

BEGUMPET

CHOICE BASED CREDIT SYSTEM

(CBCS)

SYLLABUS

Under Graduate Programme

DEPARTMENT OF STATISTICS

Course Outcome:

1. **Descriptive Statistics:**
   - The learning objectives include summering the data and to obtain its salient features from the vast mass of original data. After completing this course, the students should have developed a clear understanding of Concepts of statistical population and sample, variables and attributes.
   - Tabular and graphical representation of data based on variables.
   - 'Conditions for the consistency' and criteria for the independence of data based on attributes. Measures of central tendency, Dispersion, Skewness and Kurtosis.
   - Moments and their use in studying various characteristics of data.
   - Different approaches to the theory of probability.
   - Important theorems on probability and their use in solving problem.
   - Concept of correlation, various correlation coefficients- Pearson's correlation coefficient, Spearman's rank correlation coefficient, partial correlation coefficient

and Multiple correlation coefficient. Concept of Principle of least squares for curve fitting and regression lines.

Paper- I: Descriptive Statistics and Probability
(4 HPW with 4 credits and 100 marks)
Unit –I
Descriptive Statistics: Concept of primary and secondary data, Methods of collection and editing of primary data. Designing a questionnaire and a schedule. Sources and editing of secondary data. Classification and tabulation of data. Measures of central tendency and measures of dispersion with simple applications. Moments:- Importance, central and non-central moments, and their interrelationships, Sheppard's corrections. Skewness and Kurtosis and their measures with real life examples.
Unit –II
Probability: Basic concepts in probability—deterministic and random experiments, trail, outcome, sample space, event, and operations of events, mutually exclusive and exhaustive events, and equally likely and favorable outcomes with examples. Mathematical, statistical and axiomatic definitions of probability with limitations. Properties of probability based on axiomatic definition. Conditional probability and independence of events. Addition and multiplication theorems for n events. Boole's inequality and Bayes' theorem. Problems on probability using counting methods and theorems**.**
Unit-III
Random Variables: Definition of random variable, discrete and continuous random variables, functions of random variables, probability mass function and probability density function with illustrations. Distribution function and its properties. Transformation of one-dimensional random variable (simple 1-1 functions only). Notion of bivariate random variable, bivariate distribution and statement of its properties. Joint, marginal and conditional distributions. Independence of random variables**.**

Unit –IV
Mathematical Expectation: Mathematical expectation of a function of a random variable. Raw and central moments and covariance using mathematical expectation with examples. Addition and multiplication theorems of expectation. Definition of moment generating function (m.g.f), cumalant generating function (c.g.f), probability generating function (p.g.f) and characteristic function (c.f) and statements of their properties with applications. Chebyshev's , and Cauchy-Schwartz's inequalities and their applications.

2.**Probability Distributions**:

- A probability distribution is a statistical model that shows the possible outcomes of a particular event or course of action as well as the statistical likelihood of each event. Probability distribution functions are quite important and widely used in actuarial science (insurance), engineering, physics, evolutionary biology, computer science and even social sciences such as psychiatry, economics and even medical trials.

Paper- II: PROBABILITY DISTRIBUTIONS

(4 HPW with 4 Credits and 100 Marks)

Unit –I

Discrete distributions – I **:** Uniform and Bernoulli distributions : definitions, mean, variance and simple examples. Definition and derivation of probability function of Binomial distribution, Poisson distribution definition, properties of these distributions such as median, mode, m.g.f, c.g.f., p.g.f., c.f., and moments up to fourth order, reproductive property, wherever exists, and their real life applications. Poisson approximation to Binomial distribution.

Unit –II

Discrete distributions – II**:** Negative binomial, Geometric distributions: Definitions and physical condition, properties of these distributions such as m.g.f, c.g.f., p.g.f., c.f. and moments upto fourth order, reproductive property, wherever exists, lack of memory property for Geometric distribution and their real life applications. Poisson approximation to Negative binomial distribution. Hyper-geometric distribution – definition, physical conditions, derivation of probability function, mean, variance and real life applications. Binomial approximation to Hyper-geometric.

Unit-III

Continuous distributions – I: Rectangular and Normal distributions – definition, properties such as m.g.f., c.g.f., c.f. and moments up to fourth order, reproductive property, wherever exists and their real life applications. Normal distribution as a limiting case of Binomial and Poisson distributions.

Unit –IV

Continuous distributions – II : Exponential, Gamma : definition, properties such as m.g.f., c.g.f., c.f. and moments up to fourth order, reproductive property wherever exists and their real life applications. Beta distribution of two kinds : Definitions, mean and variance. Cauchy distribution - Definition and c.f.
Definition of convergence in Law, in probability and with probability one or almost sure convergence. Definition of Weak Law of Large Numbers (WLLN) and Strong Law of Large numbers (SLLN). Definition of Central Limit Theorem (CLT) for identically and independently distributed (i.i.d) random variables with finite variance

3. Sampling Distributions:

- To understand the concept of sampling distributions and their applications in statistical inference.
- To understand the process of hypothesis testing and its significance 3. Importance of Standard Error and to draw conclusions using p-value

Paper- III: Statistical Methods and Theory of Estimation
(4 HPW with 4 credits and 100 marks)

Unit –I

Bivariate data, scattered diagram, Principle of least squares, Fitting of straight line, second degree parabola, quadratic and power curves. Concepts of correlation, computation of Karl Pearson correlation coefficient for grouped and ungrouped data and properties.

Correlation ratio, Spearman's rank correlation coefficient and its properties. Simple linear regression, correlation verses regression, properties of regression coefficients.

Unit –II

Concepts of partial and multiple correlation coefficients (only for three variables). Analysis of categorical data, independence and association and partial association of attributes, various measures of association (Yule's) for two way data and coefficient of contingency (Pearson and Tcherprow), coefficient of colligation**.**

Unit – III

Concepts of population, parameter, random sample, statistic, sampling distribution and standard error.Standard error of sample mean(s) and sample proportion(s). Exact sampling distributions-Statement and properties of 2, t and F distributions and their interrelationships. Independence of sample mean and variance in random sampling from normal distributions.

Point estimation of a parameter, concept of bias and meansquare error of an estimate. Criteria of good estimator- consistency, unbiasedness, efficiency and sufficiency with examples.

Unit – IV

Statement of Neyman's Factorization theorem, derivations of sufficient statistics in case of Binomial, Poisson, Normal and Exponential (one parameter only) distributions.Estimation by method of moments, Maximum likelihood (ML), statements of asymptotic properties of MLE.Concept of interval estimation.Confidence intervals of the parameters of normal population by Pivot method.

4. Statistical Inference:

Drawing conclusions about the whole population on the basis of a sample. Statistical inference is the process of deducing properties of an underlying probability distribution by analysis of data. Inferential statistical analysis infers properties about a population; this includes testing hypotheses and deriving estimates.

Nonparametric Methods.

- Testing of hypothesis using Non-Parametric tests like Median test, Runs test, U test, etc. and ability to use them judiciously for the testing of given data.

Paper- IV: Statistical Inference
(4 HPW with 4 credits and 100 marks)

Unit –I

Concepts of statistical hypotheses, null and alternative hypothesis, critical region, two types of errors, level of significance and power of a test. One and two tailed tests, test function (non-randomized and randomized).Neyman-Pearson's fundamental lemma for Randomized tests. Examples in case of Binomial, Poisson, Exponential and Normal distributions and their powers.

Unit II

Large sample tests for single sample mean , difference of means, single sample proportion, difference of proportion and difference of standard deviations. Fisher's Z- transformation for population correlation coefficient(s) and testing the same in case of one sample and two samples. Definition of order statistics and statement of their distributions.

Unit – III

Tests of significance based on 2- 2 test for specified variance, goodness of fit and test for independence of attributes(r x s, 2x k and 2x2 contingency tables). Tests of significance based on Student's t- t test for single mean , difference of means for independent and paired samples, sample correlation coefficient. F test for equality of population variances.

Unit – IV

Non-parametric tests- their advantages and disadvantages, comparison with parametric tests.Measurement scale- nominal, ordinal, interval and ratio. One sample runs test, sign test and Wilcoxon-signed rank tests (single and paired samples). Two independent sample tests: Median test, Wilcoxon –Mann-Whitney U test, Wald Wolfowitz's runs test**.** Use of central limit theorem in testing.

5. Sampling Techniques and Forecasting Methods

- Survey Sampling provides the tolls/ techniques for selecting a sample of elements from a target population keeping in mind the objectives and nature of population. Most of the research work is done through Sample Survey.. After completing the course, students should have developed clear understanding of : Basic concepts of survey sampling.
- Principles of survey sampling and main steps involved in selecting a sample.
- Simple random sampling.
- Stratified random sampling
- Systematic sampling

Time Series

- Advanced understanding of the concepts of time series and their application to health, climate, finance and other areas.
- Familiarity with a range of examples for the different topics covered in the course.
- An advanced understanding of the underlying concepts in the time series and frequency domains.
- Apply ideas to real time series data and interpret outcomes of analyses.

Demand analysis

- To make student understand the demand and supply analysis in business applications.

- Understand how supply and demand interact to determine equilibrium price and quantity

Paper V- Sampling Theory, Time series and Demand Analysis
(3 Hours Per Week with 3 Credits)

UNIT-I

Sample Surveys: Concepts of population, sample, sampling unit, parameter, statistic, sample frame and standard error. Principal steps in sample surveys - need for sampling, census versus sample surveys, sampling and non- sampling errors, sources and treatment of non-sampling errors, advantages and limitations of sampling.

Sampling Methods: Types of sampling: Subjective, probability and mixed sampling methods. Methods of drawing random samples with and without replacement. Estimates of population mean, total, and proportion, their variances and the estimates of variances in Simple Random Sampling With and Without Replacement

UNIT-II

- Estimates of population mean, total, and proportion, their variances and the estimates of variances in the following methods.
- (i) Stratified Random Sampling with Proportional and Neyman allocation, and
- (ii) Systematic Sampling when N= nk.
- Comparison of relative efficiencies. Advantages and disadvantages of SRS, Stratified and Systematic sampling methods.

UNIT-III

- **Time series:** Time series and its components with illustrations, additive, multiplicative and mixed models. Determination of trend by least squares and moving average methods. Growth curves and their fitting with reference to Modified exponential, Gompertz and Logistic curves. Determination of seasonal indices by Ratio to moving average, ratio to trend and link relative methods.

UNIT-IV

- **Demand Analysis**: Introduction. Demand and supply, price elasticity of supply and demand. Methods of determining demand and supply curves, Leontief's ,Pigous's methods of determining demand curve from time series data, limitations of these methods Pigou's method from time series data. Pareto law of income distribution curves of concentration.

6. Statistical Quality Control :

- Understand the role of statistical tools in quality improvement
- Understand the different types f variability, rational subgroups, and how a control chart is used to detect assignable causes.
- Construct and interpret control charts for variables such as $\bar{x}$, R, S and Individual charts.
- Construct and interpret control charts for attributes such as p, np, U charts.

Index Numbers

- Interpret and use a range of index numbers commonly used in the business sector
- Define an index number and explain its use

- Perform calculations involving simple, composite and weighted index numbers
- Understand the basic structure of the consumer price index (CPI) and perform calculations involving its use

Paper- VI: Statistical Quality Control , Reliability &Index numbers
(3 Hours Per Week with 3 Credits)

UNIT –I
Statistical Quality Control: Importance of SQC in industry. Dimensions of quality, Statistical basis of Shewart control charts. Construction of control charts for variables (mean, range and standard deviation) and attributes (p , np with fixed and varying sample sizes) and their Interpretation.

UNIT –II
Control charts for attributes (c and u charts with fixed and varying sample sizes) and their Interpretation. Construction of control charts for Natural tolerance limits and specification limits, process capability index and modified control charts.

UNIT –III

Acceptance sampling plans:. Concept of AQL and LTPD. Producers risk and consumer's risk Single and Double sampling plans for attributes and their OC and ASN functions. Design of single and double sampling plans for attributes using Binomial and Poisson distributions. Construction of OC and ASN functions.
Reliability: Introduction. Hazard function, Exponential distribution as life model, its memory-less property. Reliability function and its estimation. System reliability - series, parallel and k out of N systems and their reliabilities with simple examples.

UNIT –IV

Index Numbers: Concept, construction, uses and limitations of simple and weighted index numbers. Laspeyer's, Paasche's and Fisher's index numbers, criterion of a good index numbers, problems involved in the construction of index numbers. Fisher's index as an ideal index number. Fixed and chain base index numbers. Cost of living index numbers and wholesale price index numbers. Base shifting, splicing and deflation of index numbers

7. Design of Experiments:

- DOE is a tool to develop an experimentation strategy that maximizes learning using a minimum of resources. Extensively used by engineers and scientists involved in the improvement of manufacturing processes to maximize yield and decrease variability. It is widely used in many fields with broad application across all the natural and social sciences, to name a few: Biostatistics, Agriculture, Marketing, Software engineering. Industry etc. After completing Course in DOE students should have developed a clear understanding of: The fundamental concepts of design ofexperiments.

- Introduction to planning valid and economical experiments within givenresources.
- Completely randomized design.
- Randomized block design.
- Latin square design

Indian Official Statistics: The students are able to know about Indian Official Statistical System.

Paper- VII: Design of Experiments, Vital statistics, Official Statistics and Business Forecasting
(3 HPW with 3 Credits)

UNIT –I

Analysis of Variance and Design of Experiments : Concept of Gauss-Markoff linear model with examples, statement of Cochran's theorem, ANOVA – one-way, two-way classifications with one observation per cell Expectation of various sums of squares, Statistical l analysis, Importance and applications of design of experiments.

UNIT –II

Principles of experimentation: Analysis of Completely randomized Design (C.R.D), Randomized Block Design (R.B.D),Latin Square Design (L.S.D) including one missing observation, expectation of various sum of squares. Comparison of the efficiencies of above designs.

UNIT – III

Vital statistics: Introduction, definition and uses of vital statistics. Sources of vital statistics, registration method and census method. Rates and ratios, Crude death rates, age specific death rate, standardized death rates, crude birth rate, age specific fertility rate, general fertility rate, total fertility rate. Measurement of population growth, crude rate of natural increase- Pearl's vital index. Gross reproductive rate sand Net reproductive rate, Life tables, construction and uses of life tables and Abridged life tables.
UNIT – IV

Official Statistics: Functions and organization of CSO and NSSO. Agricultural Statistics, area and yield statistics. National Income and its computation, utility and difficulties in the estimation of national income.
Business Forecasting: Role of forecasting in Business, Steps in Forecasting, Methods of Forecasting, Choice of a method of Forecasting, Theories of Business Forecasting, Cautions while using Forecasting Techniques.

8. Operations Research:

The 'Operations Research' is not only confined to any specific agency like defence services but today it is widely used in all industrial organizations. It can be used to find the best solution to any problem be it simple or complex. It is useful in every field of human activities. Thus, it attempts to resolve the conflicts of interest among the components of organization in a way that is best for the organization as a whole. Main fields where OR is extensively used are:

1. National Planning and Budgeting
2. Defense Services
3. Industrial Establishment and Private Sector Units
4. Research & Development and Engineering

Paper- VIII: Operations Research
(3 HPW with 3 Credits)

UNIT –I

Operations Research: Meaning and scope of OR. Convex sets and their properties. Definition of general LPP. Formulation of LPP. Solution of LPP by graphical method. Statements of Fundamental theorem of LPP and other related theorems. Simplex algorithm.

UNIT –II

Concept of artificial variables. Big –M /Penalty method and two-phase simplex methods. Concept of degeneracy and resolving it. Concept of duality of LPP. Dual Primal relationship, Statement of Fundamental Theorem of Duality.

UNIT –III

Definition of transportation problem, TPP as a special case of LPP, Initial basic feasible solutions by North-West Corner Rule, Matrix minimum method and VAM. Optimal solution through MODI tableau and stepping stone method for balanced and unbalanced transportation problem. Degeneracy in TP and resolving it. Concept of Transshipment problem.

UNIT –IV

Formulation and description of Assignment problem and its variations. Assignment problem as special case of TP and LPP. Unbalanced assignment problem, optimal solution using Hungarian method and traveling salesman problem and its solution.
Problem of Sequencing. Optimal sequence of N jobs on two and three machines without passing.

Course Outcome:

Concepts of Sequences of Random Variables

- Understand the definition of a stochastic process and in particular a Markov process;
- Classify a stochastic process according to whether it operates in continuous or discrete time and whether it has a continuous or a discrete state space, and give examples of each type process;
- Describe a Markov chain and its transition matrix;
- Calculate the distribution of a Markov chain at a given time;
- Classify the states of a Markov chain;
- Determine the stationary distributions of a Markov chain;
- Demonstrate how a Markov chain can be simulated;
- Describe a time-inhomogeneous Markov chain and its simple applications;

SEC-1: Concepts of Sequences of Random Variables
(2 HPW with 2 Credits and 50 Marks)

UNIT-I

Stochastic process, Index set, state space, classification of stochastic process with examples, stationary process, Covariance stationary process, Martingale sequence of random variables. Applications of stochastic process through examples.

UNIT-II

Definition and examples of finite- dimensional distributions of Markov Chain, time-homogeneity, transition probability matrix, marginal distribution. Classification of states-recurrent, transient, positive recurrent and null recurrent states.

Course Outcome:

Statistics for Psychology and Education

- Explain the logic and appropriate applications of statistical analyses for univariate or bivariate research designs, problems, or hypotheses.
- Calculate the statistics necessary to solve problems (both manually and via computer), including descriptive statistics, statistical significance tests, effect sizes, and confidence intervals.
- Communicate the meaning of statistical analyses in everyday language and professional formats (e.g., graphs, tables, and words).

SEC-2: Statistics for Psychology and Education
(2 HPW with 2 Credits and 50 Marks)

UNIT-I

Introduction, scaling procedures, scaling of rankings interms of Normal probability curves.

UNIT-II

Reliability of test scores, effect of test length and different ranges on reliability of the test, Validity of test scores, comparison between reliability and validity.

Course Outcome:

Big Data Analysis

The main goal of this course Is to help students learn, understand and practice big data analysis and machine learning approaches, which includes the study of modern computing big data technologies and scaling up machine learning techniques focusing on industry applications. Mainly the course outcomes are : Conceptualization and summarization of big data and machine learning, trivial data vs big data , big data computing technologies, machine learning techniques and scaling up machine learning approaches.

The students learning outcomes are designed to specify what the students will be able to perform after completion of the course:

- Ability to identify the characteristics of datasets and compare the trivial data and big data for various applications.
- Ability to select and implement machine learning techniques and computing environment that are suitable for the applications under consideration.
- Ability to solve problems associated with batch learning and online learning, and the big data characteristics such as high dimensionality, dynamically growing data and in particular scalability issues.
- Ability to understand and apply scaling up machine learning techniques and associated computing techniques and technologies.
- Ability to recognize and implement various ways of selecting suitable model parameters for different machine learning techniques.
- Ability to integrate machine learning libraries and mathematical and statistical tools with modern technologies

SEC-3: Big Data Analysis
(2 HPW with 2 Credits and 50 Marks)

UNIT I

The Rise of Big Data: What is Big Data and why does it matter; Web Data: The original Big Data; The cross section of Big Data and the value they hold;

UNIT II

Taming Big Data: The Technologies, Process and Methods: The Evolution of Analytic Scalability, The Evolution of Analytic Process, The Evolution of Analytic Tools and Methods.

Course Outcome:

Statistical Techniques in Data Mining

- To introduce students to basic applications, concepts, and techniques of data mining.
- To develop skills for using recent data mining software (eg. R) to solve practical problems in a variety of disciplines.
- To gain experience doing independent study and research

SEC-4: Statistical Techniques in Data Mining
(2 HPW with 2 Credits and 50 Marks)

UNIT-I

Introduction: Introduction to Data mining, The nature of Data sets, Types of structure, Models and patterns, Data mining Tasks, components of data mining algorithms, The Interacting roles of Statistics and Data mining, Data mining: Dredging, snooping and fishing.

UNIT-II

Data mining: Definitions, KDD vs Data mining, DBMS vs DM, other related areas, DM Techniques, other mining problems, Issues and challenges in Data mining,

Association Rules: What is an association rule, methods to discover association rules; Apriori Algorithm, Partition Algorithm