# Mangalore University

# Syllabus of

# STATISTICS

**A Major Subject for B.Sc. Degree Programme**

**(CBCS)**

**(With effect from 2024-25 onwards)**

# Mangalore University

## Revised syllabus of STATISTICS, a major subject for B.Sc. Degree Program

### PREAMBLE

Mangalore University has adapted NEP in 2021-22 and it has been three years. Now as per the direction of the Government the revision of syllabus of Statistics, a major subject for the three years B.Sc. degree program is undertaken for the students studying 2024-25 onwards. Statistics is vigorously gaining importance and recognition in today's society. For a long time, Statistics was identified solely with the displays of data and charts pertaining to the economic, demographic, and political situations prevailing in a country. However, gigantic advances during the twentieth century have enabled statistics to grow and assume its present importance as a discipline of data-based reasoning. Statistics, as a scientific discipline, stretch well beyond the confines of data display. It deals with collecting informative data, analysis interpreting these data, and drawing conclusions about a phenomenon under study. The scope of this subject naturally extends to all processes of acquiring knowledge that involve fact finding through collection and examination of data. These days Data Science emerges as a prominent and much used subject whose foundation is Statistics and it is advanced with algorithmic approach. In this programme, a foundation is given. Also, Statistics plays vital role in Big-data and Analytics, which is now an emerging field in all spheres of human activity. There is a great demand for Data Science and data analytics.

Accordingly as per CBCS Semester scheme the board designed the curriculum of B.Sc. Statistics keeping in view that the curriculum of B.Sc. Statistics should be flexible enough to prepare graduates to either be functioning Statisticians or to go on to post-graduation studies

## The following changes were made in the curriculum.

1) The Programme of study consists of revised Core Courses and Elective courses.
2) Core courses will consist of compulsory/major courses which are compulsorily studied by a student in all the six semesters. Elective courses are opted in III to IV Semesters.
3) Statistical computing is taught using R and other software in a computer lab.
4) As any career in Statistics after B. Sc. requires working with computer systems, students learn to solve the problems using computers. Introduction of computer based computation using EXCEL and R programming.
5) Description of various multivariate data analysis methods will helps them in data analysis using computers. This will help them in career and any data analysis.
6) JAMOVI, a software equivalent to SPSS is introduced.

## ELIGIBILITY

1. Only those candidates who have passed Pre-University Course in Science or an equivalent course with Mathematics as one of the subject are eligible to take Statistics as one of the major subjects in B.Sc. Course.
2. Any student taking Statistics as one of the major subjects in the B.Sc. course shall take Mathematics as another major subject.

### Programme Outcomes (POs)

By the end of the program the students will be able to:

1. **Comprehensive Understanding of Statistics**
   Acquire a structured and thorough understanding of core statistical areas, including Probability, Inference, Sampling, and Regression, along with their real-world applications.
2. **Application of Statistical Methods**
   Apply appropriate statistical techniques and methodologies to solve practical problems across various domains.
3. **Data Analysis and Software Proficiency**
   Develop proficiency in analyzing data using statistical software such as R, JAMOVI, Power BI, and Excel, and effectively communicate the results.
4. **Critical Interpretation and Reasoning**
   Interpret and critically evaluate statistical data, reports, and visualizations to draw meaningful conclusions.

5. **Statistical Experimentation and Research**
   Plan and conduct experiments or surveys, analyze outcomes, and support evidence-based decision-making for individuals, organizations, and policymakers.
6. **Career Readiness and Professional Awareness**
   Understand career prospects in government, private, and research sectors, and prepare for professional roles requiring statistical competence.
7. **Computational and Modelling Skills**
   Use statistical modelling and computing skills to explore, interpret, and solve real-world problems effectively.
8. **Transferable and Collaborative Skills**
   Demonstrate essential skills such as analytical thinking, problem-solving, independent research, communication, and teamwork for academic and professional success.

**Scheme of Instruction/Examination:**

The board recommends that the existing scheme of instruction and examination in theory and practical may be continued. The theory question paper in all semesters shall cover all the topics in the pertaining syllabus with proportional weightage to the number of hours of

Instruction prescribed.

Two questions from each unit with internal choice with or without sub-questions.

**Practical Examination:**

1) Students will have to answer any four questions out of 8 questions. All questions carry equal marks.

2) Practical paper in each semester carries 50 marks and the split up of the practical marks are as given below.

   a) Three hour examination with two examiners which includes at least one external examiner carries 30 marks.

   b) Ten marks for class records.

   c) Five marks each for the preparatory practical examination using calculator and for

   Statistical computing using R software. In the semesters where R language is not discussed, ten marks to be allotted for the preparatory practical examination.

# MANGALORE UNIVERSITY CHOICE BASED CREDIT SYSTEM (STATISTICS) SCHEME

| | Course Code | Particulars | Instruction hours/ week | Duration of the exam (hrs) | Max. Marks | | | Credits |
|---|---|---|---|---|---|---|---|---|
| | | | | | IA | Exam | Total | |
| **I SEMESTER** | | | | | | | | |
| Group I Core Subject | BSCSTCS 101 | Descriptive Statistics | 4 | 3 | 20 | 80 | 100 | 3 |
| | BSCSTPS 101 | Statistics Practical -I | 4 | 3 | 10 | 40 | 50 | 2 |
| Total number of Credits for Core Subject in I Semester:05 | | | | | | | | |
| **II SEMESTER** | | | | | | | | |
| Group I Core Subject | BSCSTCS 201 | Probability Theory and Probability Distributions | 4 | 3 | 20 | 80 | 100 | 3 |
| | BSCSTPS 201 | Statistics Practical-II | 4 | 3 | 10 | 40 | 50 | 2 |
| Total number of Credits for Core Subject in II Semester:05 | | | | | | | | |
| **In any one of the semesters III/IV** | | | | | | | | |
| Group II Discipline elective | BSCSTES 301 | Applied Statistics | 2 | 2 | 10 | 40 | 50 | 02 |
| Total number of Credits for elective course: 02 | | | | | | | | |
| **III SEMESTER** | | | | | | | | |
| Group I Core Subject | BSCSTCS 301 | Sampling distributions and Sampling Theory | 4 | 3 | 20 | 80 | 100 | 3 |
| | BSCSTPS 301 | Statistics Practical-III | 4 | 3 | 10 | 40 | 50 | 2 |
| Total number of Credits for Core Subject in III Semester : 05 | | | | | | | | |
| **IV SEMESTER** | | | | | | | | |
| Group I Core Subject | BSCSTCS 401 | Statistical Inference –I | 4 | 3 | 20 | 80 | 100 | 3 |
| | BSCSTPS 401 | Statistics Practical-IV | 4 | 3 | 10 | 40 | 50 | 2 |
| Total number of Credits for Core Subject in IV Semester:05 | | | | | | | | |
| **In any one of the semesters IV/V/VI** | | | | | | | | |
| Compulsory Skill/Practical | BSCSTIS 401 | Data Analytics | 2 | Viva-voce | 10 | 40 | 50 | 2 |
| Total number of Credits for Compulsory Skill/Practical : 02 | | | | | | | | |
| **V SEMESTER** | | | | | | | | |
| Group I Core Subject | BSCSTCS 501 | Design and Analysis of Experiments | 3 | 3 | 20 | 80 | 100 | 3 |
| | BSCSTCS 502 | Regression Analysis and Statistical Inference II | 3 | 3 | 20 | 80 | 100 | 3 |
| | BSCSTPS 501 | Practical based on BSCSTCS501 & BSCSTCS02 | 4 | 3 | 10 | 40 | 50 | 2 |
| Total number of Credits for Core Subject in V Semester: 08 | | | | | | | | |

| VI SEMESTER | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Group I Core Subject | BSCSTCS 601 | Linear Programming and Statistical Quality Control | 3 | 3 | 20 | 80 | 100 | 3 |
| | BSCSTCS 602 | Introduction to Data Science | 3 | 3 | 20 | 80 | 100 | 3 |
| | BSCSTPS 601 | Practical based on BSCSTCS601& BSCSTCS602 | 4 | 3 | 10 | 40 | 50 | 2 |
| Total number of Credits for Core Subject in VI Semester: 08 | | | | | | | | |
| Total number of Credits for Core Subject in I-VI Semesters: 36, discipline elective: 2, Compulsory skill /Practical : 2 | | | | | | | | |

# I SEMESTER

# COURSE - BSCSTCS101: DESCRIPTIVE STATISTICS

## Course Objectives:

This course aims to provide a foundational understanding of descriptive statistics, enabling students to summarize, present, and interpret quantitative and qualitative data. It also introduces concepts of bivariate analysis, categorical data analysis, and the use of R software for statistical computing and visualization.

## Course Outcomes:

By the end of the course, students will be able to:

1. Understand the basic concepts of statistics, types of data, scales of measurement, and various sampling methods used in data collection.
2. Summarize univariate data using measures of central tendency, dispersion, partition values, and explore the shape of data using moments, skewness, and kurtosis.
3. Analyze bivariate data using scatter plots, compute Pearson and Spearman correlation coefficients, and fit polynomial and exponential curves using least squares method.
4. Interpret and analyze categorical data using contingency tables and compute association measures like odds ratio, Yule's and Pearson's coefficients.
5. Use R software for statistical computing, create various data visualizations, and compute descriptive statistics for univariate and bivariate datasets.
6. Perform statistical summaries and graphical presentations using R for effective communication and exploration of data patterns.

**Unit–1: Introduction to Statistics** **10 Hours**

Statistics: Definition and scope. Data: quantitative and qualitative, cross-sectional and time series, discrete and continuous. Scales of measurement: nominal, ordinal, interval and ratio. Presentation of data: tabular and graphical. Frequency distributions, cumulative frequency distributions and their graphical representations. Stem and leaf displays. Concepts of population and sample. Different methods of data collection:- SRS, Stratified, Systematic and Cluster sampling methods: definitions only.

**Unit–2:Univariate Data Analysis** **18 Hours**

Concept of measures of central tendency: Mean, weighted mean, trimmed mean, Median, Mode, Geometric and harmonic means, properties, merits and limitations, relation between these measures. Partition Values: Quartiles, Deciles and Percentiles.

Measures of Dispersion: Range, Quartile deviation, Mean deviation, Variance, Standard deviation and their relative measures. Moments, Skewness and Kurtosis. Outliers, Box Plot.

**Unit –3: Bivariate Data Analysis** **18 Hours**

Bivariate Data, Correlation: Scatter plot, Karl Pearson's correlation coefficient, Spearman's rank correlation coefficient, properties and results. Curve fitting, concept of errors, principle of least squares, fitting of polynomial and exponential curves. Analysis of Categorical Data: Contingency table, independence and association of attributes, measures of association- odds ratio, Pearson's and Yule's measure.

**Unit -4: Statistical computing (R Software)** **10 Hours**

Introduction to R, R as a calculator, statistical software and a programming language, R preliminaries, getting help, data inputting methods(direct and importing from other spread sheet applications like Excel), data accessing, and indexing, packages, Graphics in R, built in functions, saving, storing and retrieving work. Descriptive statistics: diagrammatic representation of univariate and bivariate data (box plots, stem and leaf diagrams, bar plots, pie diagram), measures of central tendency (mean, median and mode), partition values, measures of dispersion (range, standard deviation, mean deviation and inter quartile range), summaries of a numerical data, skewness and kurtosis.

**References:**

1. Agresti, A.(2010).Analysis of Ordinal Categorical Data,2nd Edition,Wiley.
2. Anderson T.W. and Jeremy D. Finn (1996).The New Statistical Analysis of Data, Springer
3. Gupta, S. C. (2018). Fundamentals of Statistics, Himalaya Publishing House, 7th Edition.
4. Gupta S.C. and V.K. Kapoor (2020). Fundamentals of Mathematical Statistics, Sultan Chand and Co. 12th Edition.
5. Hogg, R. V. McKean J. W. and Craig, A. T. (2012). Introduction to Mathematical Statistics, Pearson 7th Edition.

6. Johnson, R.A. and Bhattacharyya, G.K. (2006). Statistics: Principles and methods. 5th Edition, John Wiley & Sons, New York.
7. Medhi, J.(2005).Statistical Methods, New Age International.
8. Ross, S.M. (2014). Introduction to Probability and Statistics for Engineers and Scientists, 5th Edition, Academic Press.
9. Tukey, J.W.(1977). Exploratory Data Analysis, Addison-Wesley Publishing Co.
10. Sudha G. Purohit, Sharad D. Gore, Shailaja R Deshmukh,(2009). Statistics Using R, Narosa Publishing House.
11. Emmanuel Paradis (2005). R for Beginners (available at https://cran.rproject.org/doc/contrib/Paradisrdebuts_en.pdf)

## BSCSTPS101: Practical Content based on Descriptive Statistics:

(Carrying-out the practical using calculator)

1. Presentation of data by frequency tables, diagrams and graphs, stem and leaf chart.
2. Arithmetic Mean (AM), geometric mean, harmonic mean, weighted AM, trimmed mean, corrected mean.
3. Median, mode and partition values.
4. Quartile deviation, mean deviation- absolute measures and relative measures, Box plot.
5. Standard deviation, coefficient of variation and, combined variance.
6. Problems on moments, skewness and kurtosis.
7. Fitting of curves by least squares method.
8. Product moment correlation coefficient and rank correlation.
9. Problems on Association of attributes.
10. Statistical Computing using R software- Descriptive Statistics.

### II SEMESTER

## COURSE - BSCSTCS201: PROBABILITY THEORY AND PROBABILITY DISTRIBUTIONS

### Course Objectives:

This course aims to build a strong foundation in probability theory and random variables, including their properties, functions, and expectations. It introduces key discrete and continuous probability distributions with their properties, relationships, and applications in statistical modelling.

### Course Outcomes:

By the end of the course, students will be able to:

1. Understand basic probability concepts including sample space, events, classical and axiomatic definitions, conditional probability, independence, and Bayes' theorem with applications.

2. Define and analyze random variables (discrete and continuous), derive their distribution functions, and compute expectation, variance, and covariance.
3. Evaluate and interpret moments, skewness, kurtosis, moment generating functions (MGF), and cumulant generating functions (CGF), and apply Tchebycheff's inequality.
4. Understand and apply standard discrete probability distributions like Bernoulli, Binomial, Poisson, Geometric, and Hypergeometric, including their properties, mean, variance, and recurrence relations.
5. Analyze continuous probability distributions such as Uniform, Exponential, Gamma, Beta, Cauchy, and Normal distributions, and use their properties in solving real-life problems.

**UNIT I - Probability** **10 Hours**

Random Experiments, Sample space, Elementary events and compound events, Algebra of events, Classical definition of probability and its limitations, relative frequency approach. Axioms of probability, Deduction of classical definition from Axiomatic definition and other results. Addition theorem, Conditional Probability and Independence, Multiplication theorem, Bayes' Theorem (with proof) and its applications.

**UNIT II - Random Variables** **15 Hours**

Random variables (discrete and continuous), properties, probability mass function, probability density function their properties, probability distribution function - properties. Joint density functions, marginal density functions and conditional density functions. Expectation of random variables - rules of expectation, addition and multiplication theorems of expectation, variance, covariance and correlation. Mean and covariance of linear combination of random variables, moments, measures of location and dispersion - skewness and kurtosis of a random variable, moment generating function (MGF), and its properties, cumulant generating function(CGF), cumulants. Chebyshev's inequality.

**UNIT III – Standard Discrete Probability distributions** **15 Hours**

Bernoulli, Binomial, Poisson, discrete uniform, Geometric, Negative Binomial, Hyper geometric distributions- definition through probability mass function (p.d.f.) - Mean, Variance, MGF, CGF, Recurrence relation between moments, properties, inter relationships.

**UNIT IV – Standard Continuous Probability distributions** **16 Hours**

Uniform, exponential, gamma (one parameter and two parameters), beta (first kind and second kind), Cauchy, normal- definition through p.d.f., mean, variance, MGF, CGF, moments and properties.

**References:**


1. Dudewitz. E. J. and Mishra.S.N.(1998).Modern Mathematical Statistics. John Wiley.
2. Goon A.M., Gupta M. K. ,Das Gupta.B. (1991), Fundamentals of Statistics, Vol. I, World Press, Calcutta.
3. Hogg R, V., Mckean J.W, and Craig, A.T (2019). Introduction to mathematical Statistics, 8th Edition, Pearson Education, New Delhi.
4. Hogg, R.V., Tanis, E.A. and Rao J.M.(2009). Probability and Statistical Inference, Seventh Edition, Pearson Education, New Delhi.
5. Mood, A.M., Graybill, F.A. and Boes, D.C. (2007). Introduction to the Theory of Statistics, 3rd Edition. (Reprint), Tata McGraw-Hill Pub.Co. Ltd.
6. Ross, S. (2002), A First Course in Probability, Prentice Hall.


## BSCSTPS 201: Practical Content based on Probability Theory:

### List of Practical Assignments

### (Carrying-out the practical using calculators and R Software)

1. Probability- Addition theorem, multiplication theorem and conditional probability.
2. Probability- Bayes' theorem.
3. Random variables (Univariate), pmf, pdf and Distribution functions, mean and variance.
4. Bivariate Probability Distributions - Conditional Mean, Conditional Variance, Correlation and Chebyshev's inequality.
5. Exercise on Binomial distribution.
6. Exercise on Poisson distribution.
7. Exercise on Geometric, Negative Binomial and, Hyper geometric distribution.
8. Exercise on normal distribution.
9. Exercise on exponential distribution and Weibull distribution.

# ELECTIVE COURSE BSCSTES 301: Applied Statistics

| **Course Objectives:** |
|---|
| This course aims to provide a foundational understanding of key applied statistical areas such as Economic Statistics and Vital Statistics. It equips students with the ability to construct and interpret index numbers, analyze time series data, and measure vital demographic indicators such as mortality, fertility, and population growth using real-world data. |

## Course Outcomes:

By the end of the course, students will be able to:

1. Explain the concept, types, and construction methods of index numbers and their applications in measuring changes in price and quantity over time.
2. Construct Consumer Price Index Numbers and interpret their role in analyzing cost-of-living variations and making adjustments.
3. Analyze time series data by identifying its components and applying techniques such as semi-averages, moving averages, and the method of least squares to study long-term trends.
4. Describe various sources of demographic data and evaluate the limitations and potential errors associated with such data.
5. Calculate and interpret key mortality indicators including crude death rate, infant mortality rate, maternal mortality rate, and standardized death rate.
6. Assess fertility and reproduction using measures such as crude birth rate, total fertility rate, and reproduction rates (gross and net), and compare their effectiveness in demographic analysis.

| Unit 1: Economic Statistics | 12 hours |
|---|---|
| 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.<br>Time Series Analysis: Components of time series, Measurement of trend by method of semi averages method of moving averages and method of least squares | |
| Unit2: Vital Statistics | 12 hour |
| Sources of demographic data, errors in data. Measurement of mortality: crude death rate, age 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, Gross reproduction rate, Net reproduction rate. Merits and demerits of each measure of fertility, comparative study of these measures of fertility, Crude growth rates. | |

References


1. J. Medhi (1992) Statistical Methods. New Age International (P) Ltd. New Delhi.
2. M.N.Das (1993) Statistical Methods and Concepts. Wiley Eastern Ltd.
3. Irwin Miller, John E Freund and Richard A Johnson(1992) Probability and Statistics for Engineers. Prentice Hal l of India New Delhi.
4. D.C. Montgomery (1996) Introduction to Statistical Quality Control.
5. Cochran, W G.(1984)Sampling Techniques, Wiley Eastern, New Delhi.
6. . Mukhopadhyay P.(2011): Applied Statistics, 2nded. Revised reprint, Books and Allied
7. Kendall M.G.(1976):Time Series, Charles Griffin.
8. Chatfield C.(1980):The Analysis of Time Series–An Introduction, Chapman & Hall.


**III Semester**

**Course - BSCSTCS 301: Sampling distributions and Sampling theory**

| **Course Objectives:** |
|---|
| This course aims to develop a foundational understanding of sampling theory and sampling distributions. It focuses on principles and techniques of probability sampling methods such as simple random, stratified, and systematic sampling, and equips students to analyze survey data efficiently. The course also introduces important theoretical distributions and their applications in statistical inference, including the central limit theorem and transformations of random variables. |

## Course Outcomes:

By the end of the course, students will be able to:

1. Understand the objectives, principles, and elements of sampling design and differentiate between sampling and complete enumeration.
2. of population parameters along with their standard errors.
3. Use stratified and systematic sampling methods, perform optimal allocation, and evaluate the gain in precision over simple random sampling.
4. Compare different Apply probability sampling techniques such as simple random sampling, and derive estimates sampling methods under various conditions including linear trend and cost constraints.
5. Derive and understand the properties of sampling distributions such as t, Chi-square, and F distributions, including their interrelationships.
6. Apply the central limit theorem and weak law of large numbers to analyze the behavior of sample statistics and support statistical inference.

| | |
|---|---|
| **Unit 1: Introduction to sampling theory** | 10 Hrs |
| Objectives and principles of sampling theory; Concept of population and sample; Elements of sampling design, Estimators, 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 |
| Simple random sampling with and without replacement, definition, and procedure of selecting asample, estimates of population mean, total and proportion, variances and standard error (SE) of these estimates, estimates of their variances related proofs, sample size determination. | |
| **Unit 3: Stratified sampling and systematic sampling** | 16 Hrs |
| 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 allocationsare used, estimation of gain in precision, post stratification and its performance.<br><br>Systematic Sampling: Linear systematic sampling technique; estimates of population mean andtotal, variances of these estimates (for N=n x k).<br>Comparison of systematic sampling with SRS and stratified sampling in the presence of linear trend and corrections. | |
| **Unit 4: Transformation of random variables and Sampling Distributions** | 15 Hrs |
| Functions of random vectors, joint density function of one-to-one transformed random vector in case absolutely continuous random vector. Definition and derivation of students' t, Chi-square and F- distributions – their properties, mean and variance. Distribution of sample mean, sample variance under normality assumption, Distribution of $\frac{ns^2}{\sigma^2}$ using the independence of $\bar{x}$ and $s^2$ when sampling from normal population. Interrelationship between the distributions, Distribution of minimum and maximum.<br><br>Limit Theorem: Sequence of random variables, convergence in probability: basic results (without proof), Weak law of large numbers, central limit theorem for i. i. d. random variables and its application. | |

**References:**


1. Cochran, W. G. (2007): Sampling Techniques, Third Edition, Wiley India Pvt. Ltd., New Delhi.
2. Changbao Wu and Mary E. Thompson (2020): Sampling Theory and Practice, Springer Nature Switzerland.
3. Raghunath Arnab (2017): Survey Sampling Theory and applications (2017), Elsevier
4. Des Raj and Chandhok P. (1998): Sample Survey Theory, Narosa Publishing House.
5. Goon A.M., Gupta M.K. and Dasgupta B. (2001): Fundamentals of Statistics (Vol.2), World Press
6. Murthy, M. N. (1967): Sampling Theory and Methods, Statistical Publishing Society, Kolkata.
7. Mukhopadhyay P (2008): Theory and methods of survey sampling. Prentice-Hall of India, New Delhi
8. Mukhopadhyay, P. (1998): Theory and Methods of Survey Sampling. Prentice Hall
9. Singh, D. and Chaudhary, F. S. (1986): Theory and Analysis of Sample Survey Designs, Wiley Eastern Ltd., New Delhi.
10. Sukhatme, P.V., Sukhatme, B. V.(1984): Sampling theory of Surveys with Applications, Indian Society of Agricultural Statistics, New Delhi.
11. Sampath S. (2005): Sampling Theory and Methods, Second edition, Narosa, New Delhi.
12. Rohatgi, V. K. (2002). An Introduction to Probability theory and Mathematical Statistics. Wiley Eastern Limited.


**BSCSTPS 301: Statistics Practical III (Using EXCEL)**

**List of Practical assignments**

1. Drawing of random sample under SRSWOR from a given population and estimation of the mean and total and the standard error of the estimator.
2. Drawing of random sample under SRSWR from a given population and estimation of the mean and total and the standard error of the estimator.
3. Estimation of the proportion, total and the standard errors of the estimators based on a random sample under SRSWR and SRSWOR
4. Estimation of the mean, total and the standard error of the estimator under stratified random sampling.
5. Exercise on allocation of samples in Stratified sampling. (Proportional Allocation)
6. Exercise on allocation of samples in Stratified sampling. (Neyman Allocation)
7. Systematic sampling.
8. Exercise on Central limit theorems.

**FOURTH SEMESTER**

**Course BSCSTCS 401: Statistical Inference – I**

## Course Objectives:

This course is designed to introduce students to the theoretical foundations of statistical inference, focusing on point estimation, hypothesis testing, interval estimation, and likelihood-based tests. It equips students with various methods of estimation and testing under parametric assumptions, and emphasizes their application in solving real-world statistical problems using classical approaches.

## Course Outcomes:

By the end of the course, students will be able to:

1. Understand the properties of estimators such as unbiasedness, consistency, sufficiency, and efficiency, and apply methods like maximum likelihood and moments for estimation.
2. Apply the concepts of statistical hypotheses, type I and II errors, significance level, p-value, and power of a test in the framework of hypothesis testing.
3. Use Neyman-Pearson Lemma and construct uniformly most powerful tests for testing simple hypotheses with examples.
4. Construct confidence intervals for means, variances, proportions, and their differences using pivotal quantity and large sample approximations.
5. Derive and apply likelihood ratio tests for parameters of normal distributions including paired t-test and correlation significance test.
6. Apply chi-square tests for independence and goodness of fit, and understand the use of large sample tests in various testing situations.

**Unit I - Point Estimation** **16** Hrs

Estimator and estimate, Unbiasedness, and consistency of estimators, Sufficient condition for consistency, Relative efficiency, Sufficient statistic. Factorization theorem and its applications, Fisher information function, Cramer Rao inequality (statement only), Minimum variance unbiased estimator, minimum variance bound estimator.

Maximum likelihood and moment methods of estimation-properties of these methods (without proof)

**Unit II - Testing of Hypotheses** **14 Hrs**

Statistical Hypotheses -Null and alternative, Simple and composite hypotheses. Critical region. Concepts of type I and type II errors, level of significance and p-value, power of the test. Power function -power curve. Most powerful - test, and best critical region. Statement of Neyman and

Pearson Lemma and its use. UMP tests with examples. Relationship between testing of hypothesis and interval estimation.

**UNIT III - Interval estimation** **12 Hrs**

Confidence coefficient, confidence interval using Pivotal Quantity method with examples. Confidence interval for mean, difference between means, variance and ratio of variances under normality. Large sample confidence interval for mean, difference of means. Proportion and difference between proportions. Lower bounds based on UMP tests.

**UNIT IV - Likelihood ratio tests (LRT**) **14 Hrs**

Derivation of tests for normal distribution only with testing for mean and variance - one sample and two sample tests - two sided alternative, Properties of LRT (without proof). Paired t test. Test for significance of correlation coefficient. Large Sample test. Chi-square test of independence and goodness of fit.

**References**


1. Gupta, S. C., & Kapoor, V. K. (2002). *Fundamental of Mathematical Statistics.* Sultan Chand & sons.
2. Hogg, R. V., & Craig, A. T. (1995). *Introduction to Mathematical Statistics.* 5/e, Prentice Hall, New Jersy, USA.
3. John freund, E. (2001). *Mathematical statistics.* Prentice hall of India, New Delhi.
4. Mood, A. M., Graybill, A., & Boes, C. (2001). *Introduction to the theory of Statistics.* Tata McGraw-Hill Publishing company Limited.
5. Robert Kabacoff, I. (2015). *R in Action - Data Analysis and Graphics with R, second edition.* dreamtech Press.
6. Ross, S. M. (2003). *Introduction to Probability Models.* 10e, Academic Press, UK.
7. Sudha Purohit, G., Sharad Gore, D., & Shailaja Deshmukh, R. (2008). *Statistics Using R.* Narosa Publishing House.


**BSCSTPS 401:Statistics Practical – IV (using R)**

**List of Practical Assignments**

1) Point Estimation -1

2) Point Estimation -2

3) Confidence intervals-1

4) Confidence intervals-2

5) Exercise on size and power of the test

6) ‘t’ test for testing mean, difference of means, Paired t test.

7) Chi square test for testing variance of a normal population, testing goodness of fit, and testing the independence of attributes.
8) F test for testing the ratio of variances of normal population.
9) Large sample tests-1
10) Large sample tests-2

## To be given in Semester IV/ V/ VI
## Compulsory Skill/Practical
## Course: BSCSTIS 401- Data Analytics

### Course Objective:

This course aims to provide students with foundational knowledge and practical skills in data analytics, simulation techniques, and data handling. Emphasis is placed on understanding the data analysis process, performing exploratory data analysis (EDA), applying simulation models, and exploring various sources of open datasets including government portals, research repositories, and industry platforms. The course integrates theory with hands-on experience using real-world datasets to extract patterns and generate short analytical reports.

### Course Outcomes:

Upon successful completion of this course, students will be able to:

1. Understand the fundamentals of data analytics and distinguish between descriptive, diagnostic, predictive, and prescriptive methods.
2. Collect, preprocess, and transform data for analysis using standard data preparation techniques.
3. Perform exploratory data analysis (EDA) on univariate, bivariate, and multivariate datasets using visual and statistical tools.
4. Apply simulation techniques, including random number generation and Monte Carlo methods, to analyze uncertainty and model real-world processes.
5. Identify and access open data sources relevant to research and public policy, and understand their structure and applications.
6. Perform essential data wrangling and exploratory analysis on real-world datasets to extract meaningful insights and communicate findings effectively.

## Unit 1: Data Analytics and Exploratory Analysis

Introduction to Data Analytics: Scope, importance, and applications. Descriptive Analytics, Diagnostic Analytics, Predictive Analytics, Prescriptive Analytics. Data collection methods, Data cleaning and preprocessing, Data reduction and transformation techniques.

Exploratory Data Analysis (EDA): Univariate Analysis, Frequency distributions, histograms, box plots. Bivariate Analysis: Correlation, cross-tabulations.

Introduction to Data Modeling - Working with real-world datasets, Applying analytics techniques to identify patterns and insights, Presenting findings using appropriate visualizations and summaries. [12 hrs]

## Unit 2: Data Handling

Introduction to Open Data – Concept of open data and its growing significance in academic research, public policy, and data-driven decision-making. Sources of real datasets, overview of dataset formats (CSV, JSON, XML, etc.). Essential techniques for importing, cleaning, and preparing data for analysis.

The practical component focuses on hands-on tasks such as accessing open datasets, performing basic data wrangling, univariate and bivariate analysis, and summarizing key findings. Generate short analytical reports, connecting theoretical knowledge with real-world applications. [10 hrs]

## Unit 3: Simulation Techniques and Applications

Fundamentals of Simulation, Random Number Generation, Monte Carlo Methods: Principles and use cases. Applications of Simulation in Data Analysis, Overview of Simulation Tools and Software. Case Studies and Practical Exercises [8 hrs]

**References:**


1. Peter Bruce, and Andrew Bruce (2017): Practical Statistics for Data Scientists: 50 Essential Concepts, O'Reilly Media
2. Averill M. Law (2015) Simulation Modeling and Analysis fifth edition, McGraw-Hill.
3. Foster Provost and Tom Fawcett (2013) – Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, O'Reilly Media.
4. Sheldon M. Ross (2006), Simulation, Elsevior Academic press
5. G. Jay Kerns (2010) Introduction to Probability and Statistics Using R, GNU Free
6. Bradley C. Boehmke (2016), Data Wrangling with R, Springer

7. Andy Field, Jeremy Miles & Zoë Field (2012) Discovering Statistics Using R, Sage publications
8. Anil Maheshwari (2015): Data Analytics Made Accessible, Amazon.in

**FIFTH SEMESTER**

**BSCSTCS 501: Design and Analysis of Experiments**

## Course Objectives:

This course aims to introduce students to the fundamental principles and techniques involved in designing and analyzing statistical experiments. It covers standard experimental designs, factorial experiments, and confounding techniques to ensure efficient and unbiased analysis of experimental data. Emphasis is placed on ANOVA techniques, fixed effect models, and interpretation of experimental outcomes.

## Course Outcomes:

By the end of the course, students will be able to:

1. Understand the principles of experimental design and perform analysis of variance (ANOVA) for one-way and two-way classified data using fixed effect models.
2. Apply and analyze standard experimental designs such as CRD, RBD, and LSD, and estimate missing observations in RBD and LSD layouts.
3. Compare the relative efficiencies of CRD, RBD, and LSD, and perform multiple comparison tests such as Tukey's method and critical difference.
4. Understand and analyze factorial experiments, interpret main and interaction effects in $2^2$ and $2^3$ designs, and use Yates' algorithm to compute effects.
5. Explain the concept of confounding, differentiate between complete and partial confounding, and analyze confounded $2^3$ factorial experiments in RBD.
6. Develop the ability to design suitable experiments, analyze the data effectively, and draw valid conclusions for real-life experimental situations.

**Unit I : Concepts of Experiment: Design and Analysis 14 Hrs**

Experiments and their designs, principles of designs of experiments, experimental error and interpretation of data contrasts and analysis of variance, Fixed and random effects models, Analysis of One-way and two-way classified data without interaction.

**UNIT II : Complete Block Design** **12 Hrs**

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. Multiple comparison tests: Tukey's method, Critical difference.

**UNIT III : Factorial Experiment** **12 Hrs**

Basic concepts – main and interaction effects, and orthogonal contrasts in $2^2$ and $2^3$ factorial experiments, Yates' method of computing factorial effects total, Analysis of $2^2$ and $2^3$ factorial experiments in RBD.

**UNIT IV : CONFOUNDING** **10 Hrs**

Need for confounding, Types of confounding - Complete and partial, Confounding in a $2^3$ factorial experiment in RBD and its analysis.

**References:**


1. Goon, A. M., Gupta, M. K., & Das Gupta, B. (1991). Fundamentals of Statistics, Vol-I,World Press, Calcutta.
2. Montgomery.D.C.(2014): Design and Analysis of Experiments , Wiley. New York.
3. Joshi. D. D. (1987): Linear Estimation and Design of Experiments, New Age International (P) Limited, New Delhi.
4. Cochran. G and G. M. Cox, G. M. (1992): Experimental Designs, John Wiley and Sons, New York.
5. Mukhopadhyaya, P. (2015). Applied Statistics, Books and Allied Pvt. Ltd.Kolkata.
6. Giri N C And Das M.N (1979) Design and Analysis of Experiments, Wiley Easter.


**BSCSTPS 501: Statistics Practical – V Part I**

**(Using calculator/ EXCEL)**

**List of Practical Assignments**

1. ANOVA for one-way classified data.
2. ANOVA for two-way classified data.
3. Analysis of CRD.
4. Analysis of RBD.
5. Analysis of LSD.
6. Missing plot techniques in RBD and LSD.

7. Analysis of $2^2$ factorial experiment using RBD layout.
8. Analysis of $2^3$ factorial experiment using RBD layout.
9. Analysis of $2^2$ factorial experiment using RBD layout (Complete confounding).
10. Analysis of $2^3$ factorial experiment using RBD layout (Partial confounding).

**FIFTH SEMESTER**

**BSCSTCS 502: Regression analysis and Statistical inference II**

## Course Objectives:

This course aims to provide students with a strong theoretical and practical understanding of linear regression analysis, sequential testing procedures, and nonparametric inference. It covers simple and multiple regression models, diagnostic techniques, and the application of distribution-free tests for various data situations.

## Course Outcomes:

By the end of the course, students will be able to:

1. Formulate and analyze simple linear regression models, estimate parameters using least squares, and perform diagnostic checks using residual analysis.
2. Apply multiple linear regression models, estimate regression coefficients, test their significance, and interpret the overall fit using appropriate statistical tests.
3. Understand the concept and advantages of sequential testing and apply Wald's Sequential Probability Ratio Test (SPRT) to common distributions.
4. Differentiate between parametric and nonparametric tests, and apply appropriate one-sample nonparametric methods such as the sign test and Wilcoxon signed-rank test.
5. Use two-sample nonparametric tests like the Wilcoxon-Mann-Whitney U test, Kolmogorov-Smirnov test, and median test for comparing distributions.
6. Apply several-sample nonparametric methods including Friedman's and Kruskal-Wallis tests to analyze data without distributional assumptions.

**Unit 1: Simple linear regression** **13 Hrs**

Simple linear regression model, assumptions, Least squares estimation, estimation of intercepts & slope along with standard error. Residual vector and proprieties, estimation of error variance. Test on regression coefficients, prediction, standard error of prediction, coefficient of determination. Diagnostic checks:- residual analysis for testing deviation from normality, homoscedasticity, outliers.

**Unit 2: Multiple Linear Regression** **13 Hrs**

Multiple Linear Regression Model, Assumptions, Gauss- Markov Theorem (Without proof), Least square estimation, variance-covariance of least squares estimators, estimation of error variance, least square estimation with restriction on parameters, Tests on linear restriction on

the parameters, testing the signification of each regressor and testing for overall signification of the model.

**Unit 3: Sequential testing** **10 Hrs**

Need for sequential test. Wald's sequential probability ratio test (SPRT) Strength of sequential tests. Wald's SPRT applied to Bernoulli, Poisson, exponential and normal distributions. Expressions to constants to be given without proof.

**Unit 4: Nonparametric tests.** **12 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.

**References**

1) Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). Introduction to Linear RegressionAnalysis, Wiley.
2) Weisberg, S. (2005). Applied Liner Regression, Wiley.
3) Domodar .N Guajarati (2017): Basic Econometrics,5th Edition.
4) George Casella, Roger L. Berger (2020): Statistical Inference, 2nd ed., Thomson Learning.
5) Rohatagi, V.K.: (2010): Statistical Inference, Wiley Eastern, New Delhi.
6) Hogg Mckean and Craig (2009): Introduction to Mathematical Statistics, 6th edition, Pearson PrenticeHall.

**BSCSTPS 501: Statistics Practical – V Part II**
**(Using calculator/ EXCEL /R)**

**List of Practical Assignments**

1. Simple Linear Regression-I
2. Simple Linear Regression-II
3. Multiple Regression-I
4. Multiple Regression-II
5. Residual Analysis
6. Construction of SPRT for Bernoulli and Poisson distributions.
7. Construction of SPRT for Normal and Exponential distributions.
8. One sample nonparametric tests: Kolmogorov-Smirnov test, sign test, Wilcoxon signed rank test.
9. Two sample nonparametric tests: Median test, Mann-Whitney U test, Run test.
10. Several sample nonparametric tests: Kruskal -Wallis test, Friedman's test.

**SEP VI Semester:**

**BSCSTCS 601: Linear Programming and Statistical Quality Control**

## Course Objectives:

This course aims to introduce students to fundamental techniques in Operations Research, especially linear programming and transportation problems, and to provide a comprehensive understanding of Statistical Quality Control (SQC). The course equips students with methods to model and solve optimization problems and apply statistical tools to monitor and improve quality in manufacturing and service processes.

## Course Outcomes:

By the end of the course, students will be able to:

1. Understand the phases of Operations Research and formulate real-life problems as Linear Programming Problems (LPP) for optimization.
2. Solve LPPs using graphical and simplex methods, including handling special cases using Charne's Big M method and understand the concept of duality.
3. Formulate and solve transportation problems using various methods such as North-West Corner Rule, Least Cost, and Vogel's Approximation Method.
4. Analyze the quality of products or processes using concepts of variation, control limits, and selection of rational subgroups in process control.
5. Construct and interpret control charts for variables ($\bar{X}$-R and $\bar{X}$-S charts) and understand the use of revised control charts in quality monitoring.
6. Develop attribute control charts (p, np, c, and u charts), analyze patterns, and make decisions based on the interpretation of control chart signals.

| **Unit 1: Operations Research and Linear Programming Problem** | 12 hours |
|---|---|
| Operations Research (OR): Origin, definition, phases of OR- types of models.<br>Linear Programming Problem (LPP): General model, formulation, graphical solution, and standard form of LPP. Simplex algorithm (without proof), Charne's big M method - indication of unique solution, multiple solution, unbounded solution, no solution and degeneracy, dual LPP and its properties. | |
| **Unit 2: Transportation problems** | 12hours |
| The transportation problem – Mathematical formulation, feasible solution, basic feasible solution, optimal solutions, degenerate basic feasible solution. A necessary and sufficient condition for the existence of a feasible solution. Finding initial basic feasible solution – North-West corner rule, Least cost, Vogel's approximation method. Optimum solution- u-v method. | |
| **Unit 3: Statistical Quality Control** | 10 hours |
| Meaning of quality, quality characteristics - variables and attributes. Causes of variation - assignable and chance causes, process control and product control. Control limits, | |

| | |
|---|---|
| specification limits, Natural tolerance limits, action limits, warning limits, Probability limits. General theory of control charts, Criteria for lack of Control, Selection of rational subgroups. | |
| **Unit 4: Control charts** | 14 hours |
| Control chart for variables: $\bar{X}$ -R charts, $\bar{X}$ - S charts, control limits with and without standard values. Revised control charts. Construction and working of these charts, Control chart for attributes: Need for attribute charts: p, np, C and U charts, analysis and interpretation. | |

**References:**


1. Kalavathy, S. K.(2013). Operations Research. Vikas Publishing House Pvt. Ltd.
2. Kanthiswarup, Manmohan, & Guptha, P. K. (2002). Operations Research. Sulthan Chand & Co.
3. Sharma, S. D. (2009). Operations Research. Kedar Nath Ram Nath.
4. Taha, H. A. (2002). Operations Research, An Introduction. Mac Millan.
5. Douglas Montgomery, C. (2001). Introduction to Statistical Quality Control. Wiely & Sons.
6. Grant, E. L., & Richard, S. L. (2001). Statistical Quality Control. Tata McGraw Ltd.
7. Mitra, A. (2001). Fundamentals of Quality Control and Improvement. Pearson education, Asia.
8. Mukhopadhyaya, P. (1999). Applied Statistics. Books and Allied Pvt. Ltd.


## BSCSTPS 601: Statistics Practical VI Part I

### (Using Calculator)

### List of Practical Assignments

1. Formulation of linear programming problem and its solution for two variables.
2. Simplex method of solving linear programming problems.
3. Charne's Big M method of solving linear programming problems.
4. Transportation problem – finding initial b.f.s.
5. Transportation problem – finding Optimal solution.
6. Exercise on $\bar{X}$ – R chart.
7. Exercise on $\bar{X}$ – S chart.
8. Control chart for attributes- p chart and np chart.
9. Control chart for attributes c chart and U chart.

**Semester VI**

**BSCSTCS 602: Introduction to Data Science**

## Course Objectives:

This course aims to introduce students to the foundational concepts of data science and data analytics, including understanding different data types, data cleaning, data preparation and visualization, business intelligence tools like Power BI, and basic machine learning methods. It prepares students to analyze, visualize, and extract insights from data using statistical and computational tools.

## Course Outcomes:

By the end of the course, students will be able to:

1. Understand the classification of data (structured, semi-structured, unstructured) and apply concepts like mean vector, covariance matrix, and multivariate normal distribution.
2. Perform data preparation tasks such as cleaning, handling missing values, removing outliers, combining datasets, and transforming data.
3. Apply various data visualization techniques including histograms, box plots, heatmaps, and use ggplot2 to extract patterns and insights.
4. Use Power BI to import, clean, transform, and visualize data using DAX functions and create interactive dashboards with various chart types.
5. Understand basic machine learning concepts including supervised (logistic regression) and unsupervised (k-means, hierarchical clustering) learning and evaluate clustering results.
6. Use **Jamovi** software to perform basic statistical analyses and support decision-making through user-friendly analytical tools.

**Unit 1** 10 Hrs

Understanding data: Introduction – Types of data: numeric – categorical – graphical – high dimensional data. Classification of digital data: Structured, Semi-Structured and Un-Structured - Examples. Introduction to Data Science.

Multivariate continuous data, mean vector, variance covariance matrix. Dependence methods and interdependence methods. Multivariate normal distribution and its properties (without proof).

**Unit 2** **12 Hrs**

Data preparation - data cleaning, methodologies and its validation.

Cleansing- Handling missing values- omitting, replacement with central tendency, identifying outliers- boxplot, removing outliers -IQR method & data Combining-merging and stacking of data frames, Data transformation method.

Data visualization: Univariate and multivariate analysis, visualization techniques including scatter plots, histograms, box plots, strip plots, count plots, and heatmaps, identifying patterns and trends in data, correlation analysis, insights and extraction using ggplot2.

**Unit 3** **12 Hrs**

1. Introduction to Power BI

Introduction to Power BI, Power BI interface and components, connecting to data sources, importing data from Excel and CSV files, data transformation using Power Query, creating conditional columns.

DAX functions (SUM, SUMX, CALCULATE, CROSSFILTER, EOMONTH), ADDCOLUMNS, IF, visualizations including cards, tree maps, line charts, pie charts, donut charts, filled maps, shape maps, bar charts, clustered bar charts, stacked column charts, Q&A visualizations, decomposition trees, narratives, error bars, dashboard creation, interactive slicers and tooltips.

**Unit 4:** **14 Hrs**

Machine learning Methods: Labelled data, supervised learning and unsupervised learning and its purpose. Classification and Prediction - Basic Concepts of Classification and Prediction.

Supervised learning: Logistic regression – logistic function/sigmoid function, logit model, logistic model as a prediction and classification.

Cluster Analysis: Basic concepts and methods – k-means clustering, hierarchical methods – Evaluation of Clustering.

JAMOVI and its application.

**References:**


1. Richard Arnold Johnson and Dean W. Wichern (2015): Applied Multivariate Statistical Analysis, Prentice Hall, 6th edition
2. Jiawei Han, Micheline Kamber (2002): Data Mining-Concepts and Techniques, Morgan Kaufman Publishers, U.S.A
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.).* Springer.
4. Zheng, A., & Casari, A. (2018): *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists.* O'Reilly Media.
5. Bruce, P., & Bruce, A. (2017): *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python.* O'Reilly Media.
6. Knaflic, C. N. (2015): *Storytelling with Data: A Data Visualization Guide for Business Professionals.* Wiley.

7. Powell, B. (2018): *Mastering Microsoft Power BI: Expert Techniques for Effective Data Analytics and Business Intelligence.* Packt Publishing.

**BSCSTPS 601: Statistics Practical VI Part II**

**(Using R Power BI, JAMOVI)**

**List of Practical Assignments**

1. Finding mean vector, variance covariance matrix and correlation matrix of multivariate data.
2. Data Cleaning.
3. Data Visualization using ggplot2.
4. Power BI-1.
5. Power BI-2.
6. Logistic regression.
7. Cluster analysis.
8. JAMOVI- general application.
9. JAMOVI-application to machine learning.

---

**(ಪ್ರೊ. ಟಿ. ಪಿ. ಎಂ. ಪಕ್ಕಳ)**
**ಅಧ್ಯಕ್ಷರು,**
**ಪದವಿ ಮಟ್ಟದ ಸಂಖ್ಯಾಶಾಸ್ತ್ರ ಅಧ್ಯಯನ ಮಂಡಳಿ**