UNIVERSITY OF KERALA

First Degree Programme under CBCSS

Scheme and Syllabus (Outcome Based Education) of Complementary STATISTICS for B. Sc. Mathematics Core

(with effect from 2022 Admission)

The syllabus is designed with an aim to equip the students with the major concepts and methods of Statistics along with the tools required to implement them in practical situations. The syllabus is prepared in accordance with the Outcome Based Education (OBE) paradigm. The curriculum is dispensed using a combination of classroom teaching, discussions, presentations, practicals, assignments, class tests etc. The syllabus has been designed to stimulate the interest of the students in Statistics and prepared in order to equip the students with a potential to contribute to the academic and industrial requirements of the society. Emphasis is given to understand the basic concepts and data analysis tools. There are practical sessions in each semester. Numerical problems solving using scientific calculators is also included in the End Semester Examination (ESE) of Courses in the semesters I, II, III & IV. Statistical computation with R is introduced which would help the students for analyzing data by making optimum usage of time and resources. For practical classes, there shall be one faculty member in charge of every 16 students (based on sanctioned strength), in accordance with the University regulations. There will be one ESE of 2 hours duration on practical using R in Semester IV.

It is mandatory to submit a duly certified Record book of practical sheets, consisting of printout of numerical problems, their R codes and results, for appearing for ESE of practical course. ESE of the practical course with a maximum of 60 marks will be held under the supervision of External Examiners duly appointed by the University. The External Examiner will also evaluate the Record books of practical work done at Lab for 20 marks.

Course Structure:

Semester	Course Code	Title of the course	Hours/ week		No. of credits	Total Hrs/ Semester	ESE Duration	Weightage In %	
			L	P				CE	ESE
I	ST 1131.1	Descriptive Statistics and Bivariate Analysis	2	2	2	72	3 hrs	20	80
II	ST 1231.1	Probability and Random Variables	2	2	2	72	3 hrs	20	80
III	ST 1331.1	Statistical Distributions	3	2	3	90	3 hrs	20	80
IV	ST 1431.1	Statistical Inference	3	2	3	90	3 hrs	20	80
	ST 1432.1	Practical using R			4		2 hrs	20	80

L – Lecture hour; P- Practical (Lab) hour

Semester - I

Course - I

ST 1131.1: Descriptive Statistics and Bivariate Analysis

Credits: 2

Hours/week: 4 (L-2, P-2)

Course Outcomes

On completion of the course, students will be able to:

CO.1: Explain the concepts of statistical surveys, sampling, census and various sampling methods like simple random sampling, systematic sampling, and stratified sampling.
CO.2: Design questionnaires and carry out surveys.
CO.3: Collect and present raw data using frequency tables as well as appropriate graphs.
CO.4: Summarize data using various measures of central tendency, dispersion, skewness and kurtosis.
CO.5: Explain the concepts of scatter diagram, correlation and calculate the correlation between two variables.
CO.6: Explain the concept of regression, fit various regression equations to given data sets and predict values of response variables.
CO.7: Explain various concepts associated with the two regression lines and identify the regression lines for given data sets.
CO.8: Practicals: Use R built in functions to solve numerical problems associated with topics covered in Modules I and II.

Module Outcomes

Sl. No:	**Outcomes**		**Taxonomy Level**
	On completion of each module, students should be able to:		
MODULE 1 Part B	MO 1.1	Define various scales of data	Remember
	MO 1.2	Distinguish between primary and secondary data	Understand
	MO 1.3	Articulate concepts of statistical surveys, sampling, and census	Understand
	MO 1.4	Define various methods of sampling	Remember
	MO 1.5	Design a questionnaire and carry out a simple survey	Understand
	MO 1.6	Construct various frequency tables	Create
MODULE 2	MO 2.1	Calculate the various measures of central tendency, dispersion, skewness and kurtosis.	Apply

	MO 2.2	Compare the merits and demerits of various measures of central tendency and dispersion.	Understand
	MO 2.3	Describe certain theoretical properties of the measures of central tendency, measures of dispersion and moments	Understand
	MO 2.4	Compare various data sets based on measures of central tendency, dispersion, skewness and kurtosis.	Evaluate
MODULE 3	MO 3.1	Explain concepts of scatter diagram, correlation and regression.	Understand
	MO 3.2	Apply principle of least squares to fit various curves	Apply
	MO 3.3	Fit various curves to data sets	Apply
MODULE 4	MO 4.1	Construct regression lines for data sets.	Apply
	MO 4.2	Identify regression lines	Analyze
	MO 4.3	Calculate angle between lines, point of intersection etc.	Analyze
	MO 4.4	Calculate Pearson's coefficient of correlation, Spearman's rank correlation coefficient and interpret the results.	Evaluate
	MO 4.5	Coefficient of determination and coefficient of alienation	Remember
MODULE 5 (Only for Practical Exam)	MO 5.1	Use built in R functions: (i) For representing data using diagrams and graphs. (ii) For calculating the various measures of descriptive statistics	Apply

Course Content

Module I:

Part A: **Introduction (Not for Examination Purpose):** Definition and significance of Statistics, Limitations and misuse of Statistics, Official Statistical system of India. Types of Data: Concepts of primary data and secondary data, population, and sample; Classification of data based on geographic, chronological, qualitative and quantitative characteristics.

Part B: Collection and Presentation of Data: Scales of data-nominal, ordinal, interval and ratio. Methods of collection of primary data–Preparation of questionnaires / schedules. Secondary data – major sources and limitations; Census and Sample Surveys; Methods of sampling ***(concepts only):*** Probability and non-probability sampling, simple random sampling with replacement (SRSWR) & simple random sampling without replacement (SRSWOR), Systematic sampling and Stratified sampling; sampling and non-sampling errors; Classification and tabulation - Construction of tables with one or more factors of classification, frequency distributions, relative and cumulative frequency distributions.

Module II:

Summarization of Data: Central tendency- mean, median, mode, geometric mean, harmonic mean; properties of arithmetic mean and median; Relationship between AM, GM and HM; Absolute and relative measures of dispersion: Range, quartile deviation, mean deviation and standard deviation; Properties of mean deviation, standard deviation, combined mean and combined standard deviation;

coefficient of variation; moments - raw and central moments; relationship between raw and central moments; effect of change of origin and scale; skewness, kurtosis and their measures.

Module III:
Bivariate data: Scatter diagram, fitting of curves- Principle of least squares, fitting of straight line $y = ax + b$, fitting of curves: $y = ax^2 + bx + c, a \neq 0$, $y = ab^x$, $y = ax^b$, $y = ae^{bx}$.

Module IV:
Karl Pearson's coefficient of correlation, Spearman's rank correlation coefficient, regression lines and prediction, coefficient of determination and coefficient of alienation (definition only)

Module V: (only for Practical Exam)
Basics of R (as given in Practical Sheet - 1); Practical based on Modules I & II – Data analysis: presentation of data – charts and diagrams, calculation of descriptive statistics, moments, measures of skewness and kurtosis.

References:

1. Agarwal, B.L. (2006). *Basic Statistics*. 4th Edition, New Age international (P) Ltd., New Delhi.
2. Gupta S. P. (2004). *Statistical Methods.* Sultan Chand & Sons, New Delhi.
3. Gupta, S. C., and Kapoor, V. K. (1994). *Fundamental of Mathematical Statistics.* Sultan Chand & Sons, New Delhi.
4. Kenny J. F (1947). *Mathematics of Statistics Part One*. 2nd Edition, D. Van Nostard Company, New Delhi-1.
5. Kenny J. F & Keeping E. S (1964). *Mathematics of Statistics –Part Two*. 2nd Edition, D. Van Nostard Company, New Delhi-1.
6. Mukhopadhyay, P. (1996). *Mathematical Statistics*. New Central Book Agency (P) Ltd, Calcutta.

Semester – II

Course - II

ST 1231.1: Probability and Random Variables

Credits: 2

Hours/week: 4 (L-2, P-2)

Course outcomes

On completion of the course, the students should be able to:

CO.1: Distinguish between random and non-random experiments.
CO.2: Evaluate the probabilities of events using classical, statistical and axiomatic approaches.
CO.3: Identify independent events; calculate conditional probability and application of Bayes' theorem.
CO.4: Distinguish between discrete and continuous random variables with its probability distributions.
CO.5: Assess the independence of random variables.
CO.6: Calculate moment generating function and characteristic function.
CO.7: Determine the conditional mean and variance of a random variable.
CO.8: Evaluate the correlation between two random variables.
CO.9: Practical: Use R built in functions to solve numerical problems associated with topics covered in modules III and IV of ST 1131.1 (of Semester -I)

Module Outcomes

Sl. No:	Outcomes		Taxonomy Level
	On completion of each module, students should be able to:		
MODULE 1	MO 1.1	Distinguish between Random and non-random experiments	Understand
	MO 1.2	Explain the concepts of sample space, types of events and algebra of events	Understand
	MO 1.3	Describe the probabilities of events using classical, statistical and axiomatic approaches.	Apply
	MO 1.4	Identify mutually exclusive and exhaustive events	Understand
	MO 1.5	Define equally likely events	Remember
MODULE 2	MO 2.1	Determine the conditional probability and apply multiplication theorem	Evaluate
	MO 2.2	Explain the concepts of independence of events	Analyze
	MO 2.3	Use Bayes' theorem to evaluate posterior probabilities	Apply
MODULE 3	MO 3.1	Explain the concept of random variables	Understand
	MO 3.2	Distinguish between the discrete and continuous random variables and find its probability distributions	Analyze
	MO 3.3	Evaluate marginal and conditional distributions of bivariate random variables	Evaluate
	MO 3.4	Check for the independence of random variables	Analyze
	MO 3.5	Apply the concepts of transformation of univariate random variables	Analyze
MODULE 4	MO 4.1	Explain the concepts of mathematical expectation and its properties.	Understand
	MO 4.2	Determine the mathematical expectation of a discrete and continuous random variable	Apply
	MO 4.3	Calculate the conditional mean and variance of bivariate distributions	Apply
	MO 4.4	Explain the basic concepts of moment generating function and characteristic function	Apply

	MO 4.5	Evaluate the covariance and correlation coefficient of two random variables.	Apply
MODULE 5 (only for Practical Exam)	MO 5.1	Use built in R functions to solve numerical problems corresponding to modules III and IV of ST 1131.1 of Sem-1	Apply

Course content

Module I:
Random experiments - sample space and sample point; Events-algebra of events, concepts of equally likely, mutually exclusive and exhaustive events.
Probability: Statistical regularity, classical approaches, Axiomatic approach, theorems in probability, probability space.

Module II:
Conditional probability, multiplication theorem, independence of two and three events, compound probability, Bayes' theorem and its applications.

Module III:
Random variables - discrete and continuous, probability mass function and probability density function, distribution function, joint distribution of two random variables, marginal and conditional distributions, independence, transformation of variables-one-to-one transformations - univariate case only.

Module IV:
Expectation of random variables and its properties, theorems on expectation of sums and product of independent random variables, conditional expectation, moments, moment generating function, characteristic function, their properties and uses; bivariate moments, Cauchy- Schwartz inequality and correlation coefficient.

Module V: Practical (Numerical Problems) based on Modules III & IV of ST 1131.1 (Sem.1) – Scatter diagram, curve fitting, measures of correlation, regression analysis.

References

1. Bhat B. R. (1985). *Modern Probability Theory*. New Age International (P) Ltd, New Delhi.
2. Gupta, S. C., and Kapoor,V. K. (1994). *Fundamentals of Mathematical Statistics.* Sultan Chand & Sons. New Delhi.
3. Mukhopadhyay, P. (1996). *Mathematical Statistics*. New Central Book Agency (P) Ltd, Calcutta.

4. Pitman, J. (1993). *Probability*. Narosa Publishing House, New Delhi.
5. Rao C. R. (1973). *Linear Statistical Inference and its Applications*. 2nd edition, Wiley, New York.
6. Rohatgi V. K. (1993). *An Introduction to Probability Theory and Mathematical Statistics*. Wiley Eastern, New Delhi.

Semester – III

Course - III

ST 1331.1: Statistical Distributions

Credits: 3

Hours/week: 5 (L-3, P-2)

Course Outcomes

On completion of the course, students will be able to:

CO.1: Define various discrete and continuous standard distributions and explain their theoretical properties.
CO.2: Solve numerical problems associated with discrete and continuous standard distributions.
CO.3: Fit binomial, Poisson and normal distributions to data sets and calculate theoretical frequencies.
CO.4: Explain the laws of large numbers and apply them to solve numerical problems
CO.5: Define sampling distributions (normal, chi-square, Students t and F) and solve elementary numerical problems.
CO.6: Practicals: Use built in functions of R to solve numerical problems on modules I, II & IV.

Module Outcomes

Sl. No:	Outcomes		Taxonomy Level
	On completion of each module, students should be able to:		
MODULE 1	MO 1.1	Explain discrete standard distributions and their practical applications.	Understand
	MO 1.2	Describe the theoretical properties of these distributions.	Understand
	MO 1.3	Solve numerical problems associated with these distributions.	Apply
	MO 1.4	Fit binomial and Poisson distributions to data sets and calculate theoretical frequencies.	Analyze
MODULE	MO 2.1	Define continuous standard distributions.	Understand

2	MO 2.2	Describe the theoretical properties of these distributions.	Understand
	MO 2.3	Solve numerical problems associated with these distributions.	Apply
	MO 2.4	Fit Normal distribution to data sets and calculate theoretical frequencies.	Analyze
MODULE 3	MO 3.1	Explain Chebycheff's inequality and laws of large numbers.	Understand
	MO 3.2	Derive Chebycheff's inequality and laws of large numbers.	Apply
	MO 3.3	Apply the laws of large numbers to solve numerical problems.	Analyze
MODULE 4	MO 4.1	Distinguish between parameter and statistic.	Understand
	MO 4.2	Define sampling distributions (normal, chi-square, t and F).	Remember
	MO 4.3	Derive distributions of sample mean and sample variance.	Understand
	MO 4.4	Solve numerical problems associated with these distributions using their respective table values.	Apply
	MO 4.5	State relations between the sampling distributions.	Remember
MODULE 5 (only for Practical Exam)	MO 5.1	Use built-in R functions to solve numerical problems associated with standard distributions and sampling distributions. (to the extent of the portions covered in the modules I, II and IV)	Apply

Course Content

Module I:
Standard Distributions (Discrete)- uniform, binomial, Poisson and geometric- moments, moment generating function, characteristic function, problems, additive property (binomial and Poisson), recurrence relation (binomial and Poisson), Poisson as a limiting form of binomial, memoryless property of geometric distribution; Fitting of binomial and Poisson distributions; hypergeometric distribution (definition, mean and variance only).

Module II:
Standard Distributions (Continuous)– uniform, exponential, and gamma - moment generating function, characteristic function, problems; memoryless property of exponential distribution; additive property of gamma distribution; beta distribution (I and II kinds)- moments, normal distribution- moments, moment generating function, characteristic function, problems, recurrence relation of central moments; convergence of binomial and Poisson to normal.

Module III:
Chebychev's inequality; Law of large numbers-BLLN, convergence in probability (definition only), WLLN; central limit theorem (Lindberg-Levy form) - statement and applications only.

Module IV:
Sampling distributions - Parameter and statistic, Sampling distributions- Distribution of mean of a sample taken from a normal population, chi-square - definition and properties, t and F distributions (definitions only) and statistics following these distributions, relation between normal, chi-square, t and F distributions.

Module V:
Numeric problems based on Modules I, II & IV – Discrete and continuous probability distributions and evaluation of probabilities, sampling distributions and their probability evaluation, random number generation.

References

1. Gupta S.C. and Kapoor V.K. (1980). *Fundamentals of Mathematical Statistics*. Sultan Chand and Sons, New Delhi.
2. John E. Freund (1980*). Mathematical Statistics*. Prentice Hall of India, New Delhi.
3. Medhi J. (2005). *Statistical Methods-An Introductory Text*. New Age International (P) Ltd, New Delhi.
4. Mukhopadhyay, P. (1996). *Mathematical Statistics*. New Central Book Agency (P) Ltd, Calcutta.
5. Rohatgi V. K. (1993). *An Introduction to Probability Theory & Mathematical Statistics.* Wiley Eastern, New Delhi.

Semester – IV

Course - IV

ST 1431.1: Statistical Inference

Credits: 3

Hours/week: 5 (L-3, P-2)

Course outcomes

On completion of the course, the students should be able to:

CO.1: Analyze a sample to draw valid inferences about the parameters of a statistical population.
CO.2: Explain the properties of estimators and solve numerical problems for the point and interval estimators of the parameters.
CO.3: Explain the concept of testing statistical hypotheses.
CO.4: Identify two types of errors, compute level of significance and power of a test.
CO.5: Conduct tests for hypothesis about the population mean and proportion using large samples.

CO.6: Conduct tests for hypothesis about the homogeneity and independence using chi-square statistics.

CO.7: Conduct tests for hypothesis about the mean and variance for normal population using small samples.

CO.8: Carry out and interpret ANOVA.

CO.9: Practical: Use R built-in functions to solve numerical problems associated with topics covered in various modules.

Module outcomes

Sl. No:	Outcomes		Taxonomy Level
	On completion of each module, students should be able to:		
MODULE 1	MO 1.1	Define point estimator of a parameter in a statistical population.	Remember
	MO 1.2	Illustrate whether an estimator satisfying unbiased and consistent.	Understand
	MO 1.3	Explain sufficiency and efficiency of an estimator.	Apply
	MO 1.4	Describe maximum likelihood estimator and moment estimator of a parameter.	Apply
	MO 1.5	Define confidence interval.	Remember
	MO 1.6	Construct confidence intervals for mean, variance and proportion in a population.	Apply
MODULE 2	MO 2.1	Explain the concept of statistical hypothesis.	Understand
	MO 2.2	Describe two types of errors in a statistical hypothesis.	Understand
	MO 2.3	Determine the level of significance and power of a test.	Apply
	MO 2.4	Explain Neyman- Pearson lemma.	Apply
MODULE 3	MO 3.1	Define large sample and small sample tests.	Remember
	MO 3.2	Describe the test procedure for mean and proportion (one and two sample cases) using large samples.	Apply
	MO 3.3	Examine the homogeneity and independence using chi-square tests	Apply
	MO 3.4	Explain paired t test.	Apply
	MO 3.5	Describe the test procedure for mean and variance (one and two sample cases) for normal population using small samples.	Apply
MODULE 4	MO 4.1	Explain the concept of Analysis of variance.	Understand
	MO 4.2	Explain the model and hypothesis of one way and two way classified data.	Understand
	MO 4.3	Construct ANOVA table and draw inferences from it.	Evaluate
MODULE 5 (only for Practical Exam)	MO 5.1	Use built-in R functions to solve numerical problems associated with Modules III & IV.	Apply

Course content

Module I:
Point estimation, desirable properties of estimators – unbiasedness, consistency, efficiency and sufficiency; Methods of estimation –Maximum likelihood method and method of moments; Interval estimation of mean, variance and proportion (single unknown parameter only).

Module II:
Testing of Hypothesis: statistical hypotheses, simple and composite hypotheses, two types of errors, significance level, p-value, power of a test, Neyman-Pearson lemma (statement only) and applications.

Module III:
Large sample tests – testing mean and proportion (one and two sample cases), chi-square test of goodness of fit, independence and homogeneity.
Small sample tests- Z-test for means; one sample test for mean of a normal population, equality of means of two independent normal populations, t-test for independent samples and paired samples, chi-square test for variance, F-test for equality of variances.

Module IV:
Design of Experiments- assumptions and principles, Analysis of Variance (ANOVA) of one way and two way classified data (Derivation of two– way model is not included).

Module V: Practical based on Modules III &IV – tests of hypotheses (as given in Practical Sheet – 11); one way and two way ANOVA.

References

1. Das M. N., Giri N. C. (2003). *Design and analysis of experiments*. New Age International (P) Ltd, New Delhi.
2. John E. Freund (1980). *Mathematical Statistics*. Prentice Hall of India, New Delhi.
3. Medhi J. (2005). *Statistical Methods-An Introductory Text*, New Age International (P) Ltd. New Delhi.
4. Paul G. Hoel, Sidney C. Port, Charles J. Stone (1971). *Introduction to Statistical Theory*. Universal Book stall, New Delhi.

Semester – IV

Course - V

ST 1432.1: Practical using R

Credits: 4

Any standard version of R in any operating system can be used. The Record book is mandatory to appear for the Practical examination. The Record book should contain following Practical sheets based on Module V of Courses ST1131.1 to ST 1431.1. Minimum number of questions covering all functions/methods given therein must be included in each practical sheet along with R code, their outputs, interpretation / conclusion.

Practical Sheet - 1: Data Types in R

Basics of vector, matrix and data frame, basic functions – c( ), sequence( ), scan( ), factor( ) table( ), and cut( ).

Minimum number of questions - 12

Practical Sheet - 2: Sampling and Frequency Tables.

Forming ungrouped and grouped frequency tables with raw data using table and cut functions. SRSWR and SRSWOR with sample( )

Minimum number of questions - 8

Practical Sheet - 3: Measures of Central Tendency

Descriptive measures: sum, sort, min, max, length, mean, median, mode (using sort and table), geometric mean, harmonic mean.

Minimum number of questions - 10

Practical Sheet - 4: Measures of Dispersion

Range, mean deviation, IQR, quartile deviation, sd, var, coefficient of variation, quantile, summary.

Minimum number of questions - 10

Practical Sheet - 5: Moments, Skewness and Kurtosis

Computation of raw, central moments, moment measures of skewness and kurtosis.

Minimum number of questions - 8

Practical Sheet - 6: Graphical Methods

Simple bar plot, multiple bar plot (side by side and subdivided), pie chart, histogram, scatter plot, plot function and lines function.

Minimum number of questions - 8

Practical Sheet - 7: Probability Distributions

Binomial, Poisson, normal, chi-square, t and F distributions – The **d**, **p, q** and **r** functions, the scale function, evaluation of probabilities using these functions.

Minimum number of questions - 10

Practical Sheet - 8: Fitting of Distributions

Fitting of binomial, Poisson and normal distributions.

Minimum number of questions - 3

Practical Sheet - 9: Correlation and Regression

Computation of covariance for a bivariate data using cov(), Pearson's and Spearman's correlation coefficient using cor(). Linear regression models: fitting using lm(), prediction from fitted model.

Minimum number of questions - 6

Practical Sheet - 10: Curve Fitting

Fitting of a straight line and $y = ax^2 + bx + c, a \neq 0$; $y = ae^{bx}, y = ab^x$ and $y = ax^b$, where $a, b \text{ and } c$ are real constants.

Minimum number of questions - 5

Practical Sheet - 11: Testing of Hypotheses

Testing of hypothesis: prop.test (one sample and two sample), t.test (one sample, two sample, and paired), chi squared tests (goodness of fit, and independence of attributes). F test for equality of variances.

Minimum number of questions - 8

Practical Sheet - 12: Analysis of Variance

Analysis of Variance: One way anova and two way anova with one observation per cell.

Minimum number of questions - 4

References:

1. Dalgaard, P.(2008). *Introductory Statistics with R*, Springer, New York.
2. Kerns, G J. (2010). *Introduction to Probability and Statistics using R.* ISBN-10 : 0557249791
3. Lander J. P. (2017). *R for everyone* 2/e. Addison-Wesley Professional, U. S.
4. Michael J. Crawley (2013*). The R Book*, 2/e, Wiley, New York.
5. Purohit, S. G., Deshmukh, S.R., & Gore, S. D. (2008). *Statistics using R.* Alpha Science International, United Kingdom.

Web Resources:

1. https://cran.r-project.org
2. https://cran.r-project.org/manuals.html
3. https://www.r-project.org/other-docs.html
4. https://journal.r-project.org/
5. https://www.r-bloggers.com