CENTRAL UNIVERSITY OF TAMIL NADU
THIRUVARUR – 610005
(Established by an Act of Parliament, 2009)

SCHOOL OF MATHEMATICS AND COMPUTER SCIENCES
DEPARTMENT OF STATISTICS AND APPLIED MATHEMATICS

Certificate Course
Title : Statistical Methods for Data Science

Programme Structure

No.	Course Code	Course Title	CBCS Nature	Course Type	Credit	Hours / week	Marks	
							Int	Ext
1.	SAMCC01	Statistical Methods for Data Science	CC	Theory, Lab & Project	4	5	40	60
		Total			4	5	100	
	Duration: 60 hours (12 weeks @ 5 hours per week) Total Credits: 60							

Eligibility: Candidates with any undergraduate degree

Fee: Rs.1000 per student

Mode: Both theory and practical classes will be conducted in online through Google platform

Tools: Excel/ SPSS/R studio

- Each candidate shall attend both theory and practical classes
- Each candidate shall undertake one mini /major project work, as part of the program, in order to learn applications of statistical theory and methods to real-world problems.
- Lectures will be supplemented by case studies, tutorials, home assignments, seminars and group discussions

Evaluation Criteria: Assignments, tests and Monitored Exams.

Maximum Marks: 100 Marks-40 marks based on Assignments(5 Marks), tests(5 Marks), practical(10 Marks) and Projects(20 Marks). 60 Marks based on theory examination

As per the guidelines of National Higher Education Qualifications Framework (NHEQF)

Award:

- Candidates who passed the examinations conducted for theory and practical and submission of project report will be eligible for award **Certificate in Statistical Methods for Data Science**.

Program Objectives (PO)

After successful completion of the program, the students will

PO1	develop skills and knowledge that prepare them for jobs
PO2	be able to transfer skills like critical thinking and a decision-career launch program
PO3	acquire knowledge of the research in Statistics and Data Science relating to planning for national development and social welfare
PO4	undertake collection of information, investigations, projects, and operational research for purposes of planning and improvement of the efficiency of management and production in real-time scenario"
PO5	be equipped with basic practical skills in SPSS/R programming with sufficient theory to understand the statistical principles involved to develop in them the power of systematic thinking, practical approach, and exposition

Graduate Attribute for Certificate Programme in Statistical Methods for Data Science

(i) Knowledge for jobs
(ii) Knowledge of research in Statistics and Data Science
(iii) Intensive training in theory and practical in Statistics for data analysis
(iv) Skill of interpreting the results of statistical data analysis
(v) Practical skills in usage of Excel/SPSS/R programming

Program Specific Outcomes (PSO)

On successful completion of the program, the students will be able to

PSO1	undertake scientific study / survey for collection, tabulation and analyze information
PSO2	carry out research projects
PSO3	use various modules of SPSS software /R programming for conducting statistical data analysis relating to planning for national development and social welfare
PSO4	Transfer knowledge acquired from theory and principles to real-world problems systematically
PSO5	Interpret the results of statistical data analysis

PSO to PO Mapping

	PSO1	PSO2	PSO3	PSO4	PSO5
PO1	3	3	3	3	3
PO2	2	3	2	3	3
PO3	2	3	3	2	1
PO4	1	3	2	2	2
PO5	2	2	3	3	2

Program Outcomes

- **Application of knowledge and skills**: Apply the acquired operational or technical and theoretical knowledge, and a range of cognitive and practical skills to select and use basic methods, tools, materials, and information to generate solutions to specific problems relating to the chosen fields of learning.
- It offers intensive instruction in both theoretical concepts and practical applications of statistics, the based on data analytics fostering a problem-solving mindset essential for success.
- Able to develop research and statistical skills, a graduate striving for scholarly excellence, an early-career researcher enhancing your communication and analytical abilities, or a seasoned professional in academia or industry aiming to publish impactful research, our course is tailored to meet the goals.
- Educators, non-profit staff, and individuals pursuing careers in public policy, healthcare, or social sciences also benefit from our practical insights and expert guidance in research paper writing and statistical analysis.

Certificate Course

Syllabus

SAMCC01 **STATISTICAL METHODS FOR DATA SCIENCE** **4 Credits**

a. Course Outcome (CO)

On successful completion of the course, the student will be able to

	Course Outcome	Level
CO 1	Recall data presentation, central tendency and dispersion.	Remember
CO 2	Explain bi-variate data, correlation and regression analysis.	Understand
CO 3	Carry out testing of hypotheses for parameters.	Apply
CO 4	Learn to calculate probability of an event.	Skill
CO 5	Decide appropriate sample survey design.	Analyze

b. Mapping of Program Specific Outcomes with Course Outcomes

	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6
CO1	3	2	1	2	1	2
CO2	1	3	2	1	2	3
CO3	2	3	2	3	3	1
CO4	3	3	2	1	3	2
CO5	3	2	3	3	3	3

Syllabus

Unit I: Introduction to Statistics, population and sample, representation of data using frequency distribution and diagrams, measures of central tendency, measures of dispersion, moments, skewness, and kurtosis.

Unit 2: Probability distributions – binomial, Poisson and normal. Exact sampling distributions based on normal population. Bivariate data visualization, Scatter diagram, Karl-Pearson Correlation Coefficient, Coefficient of Determination, Spearman's Rank Correlation. Simple linear regression model and its fitting.

Unit 3: Testing of Hypotheses: null and alternative hypotheses and their formulation, level of significance, type I and type II errors, their probabilities, and critical region.

Unit 4: Testing of hypotheses: Tests based on normal, S*tudent's t* and *F* distributions. Tests based on *chi*-square distribution - test for independence of attributes and goodness-of-fit..

Unit 5 : Sampling and Non-Sampling errors. Types of sampling: Probability and Non-probability Sampling. Simple Random Sampling with and without replacement, Nonparametric tests

Note: Practical exercises and Project work will be based on the contents Unit 1 to Unit 5 of the above syllabus

References

1. R. V. Hogg, A. T. Craig and J. W. McKean(2019), , Introduction to Mathematical Statistics, 8th Edition, Pearson Education, New Delhi .
2. A. M. Goon, M. K. Gupta and B. Dasgupta(2016), Fundamentals of Statistics (Vol-I and II), World Press, Kolkata
3. Gupta, S. C., and Kapoor, V. K. (2020). Fundamentals of mathematical statistics. 12th edition, Sultan Chand and Sons,New Delhi.
4. J.K.Sharma(2014), Business Statistics,4th edition, Pearson education,New Delhi
5. https://online.stat.psu.edu/statprogram/reviews/statistical-concepts