Detailed Syllabus
Lecture-wise Breakup

Course Code	21M21CS123	**Semester :Even**	**Semester Session : EVEN 2021**
Course Name	Essential Statistics for Data Science		
Credits		**Contact Hours**	3

Faculty (Names)	**Coordinator(s)**	Dr. Megha Rathi
	Teacher(s) (Alphabetically)	Dr. Anuja Arora Dr. Megha Rathi

COURSE OUTCOMES		**COGNITIVE LEVELS**
CO1	Understand the basic principle of probability and statistics and its need in the context of data science	Understand Level (Level 3)
CO2	Develop own statistical analyses and implement them with advanced statistical programming tools	Apply Level (Level 3)
CO3	Compare the performance of multiple statistical methods and models and articulate the limitations and abuses of formal inference and modeling.	Analyze Level (Level 4)
CO4	Evaluatestatistical techniques for constructing learning models and can use different measures of model fit and performance to assess models.	Evaluate Level (Level 5)
CO5	Create real world applications using statistical or data modeling techniques and test hypothesis.	Create Level (Level 5)

Module No.	**Title of the Module**	**Topics in the Module**	**No. of Lectures for the module**
1.	Probability and Statistical inference	Modern Statistics, Statistics and Engineering, Probability, Conditional Probability,The Axioms of Probability, Sampling, Randomness, Intro to Statistics: Mean and Variance, Covariance, Types of Convergence.	6
2.	Statistical Methods in Data Science	Data Distribution (Bernoulli, Uniform,Binomial,Normal, Poisson), Mathematical Statistics, Inferential Statistics, Descriptive Statistics, Random Variable,Gauss-Markov theorem, F-distribution	7
3	Hypothesis Testing	Hypothesis Testing, Difference of Means, Significance Level and P-Value, Z-test, ANOVA,T-Test, Redundancy Test, Chi-Square & F-test, Type-I and Type-II errors	7
4.	Data Modeling	Cross validation, Monte Carlo methods, Cluster analysis, Time Series Modeling	6
5.	Introduction to Bayesian Modeling	Bayes' Theorem, Conditional Statements, Bayesian Thinking: priors, posteriors, and Maximum Likelihood Estimation, Bayesian inference	5
6.	Correlation & Regression Models	Correlation Analysis, Linear regression methods, Ridge regression, LASSO Regression , Logistic regression	6
7.	Ensemble Learning	Bagging & Boosting, Random Forest, Adaboost	5
		Total number of Lectures	**42**
Evaluation Criteria			

Components	**Maximum Marks**
T1	20
T2	20
End Semester Examination	35
TA	25 (**Attendance and Tut Performance (10), Quiz/ Mini-Project/Assignment (15)**)
Total	**100**

Recommended Reading material: Author(s), Title, Edition, Publisher, Year of Publication etc. (Text books, Reference Books, Journals, Reports, Websites etc. in the IEEE format)

1.	Arnold, T., Kane, M., & Lewis, B. W. (2019). A computational approach to statistical learning. CRC Press.
2.	James, G., Witten, D., Hastie, T., &Tibshirani, R. (2013). Statistical learning. In An Introduction to Statistical Learning (pp. 15-57). Springer, New York, NY.
3.	Gutierrez, D. D. (2015). Machine learning and data science: an introduction to statistical learning methods with R. Technics Publications.
4.	Lomax, R. G., &Hahs-Vaughn, D. L. (2013). An introduction to statistical concepts. Routledge.
5.	Grus, J. (2019). Data science from scratch: first principles with python. O'Reilly Media.
6.	Chatfield, C., & Xing, H. (2019). The analysis of time series: an introduction with R. CRC press.
7.	Afifi, A., May, S., Donatello, R., & Clark, V. A. (2019). Practical multivariate analysis. CRC Press.
8.	Zumel, N., & Mount, J. (2014). Practical data science with R. Manning Publications Co..
9.	Saltz, J. S., & Stanton, J. M. (2017). An introduction to data science. SAGE Publications.