Course title: Statistics for Data Science			
Course code:	**No. of credits:** 4	**L-T-P: 30-15-00**	**Learning hours: 60**
			L: Lectures**; T:** Tutorials**; P:** Practicals
Pre-requisite course code and title (if any): None			
Department: Natural and Applied Sciences			
Course coordinator:		**Course instructor:**	
Contact details:			
Course type: Core		**Course offered in:** Semester 1	

Course Description

The course has been designed and intended to help students to understand the fundamentals of statistics for data analysis. Students will be learning about data collection and how to extract information from data. The course introduces the concepts of correlation, regression, and hypothesis testing.

Course objectives

- Become familiar with the basic concepts in statistics and exploratory data analysis.
- Introduce descriptive statistics and tools of data visualization.
- Differentiate between descriptive and inferential statistics.
- Understand the concept of data sampling, parameter estimation and hypothesis testing.
- Become aware of a wide range of applications of statistics and data analysis in decision making.
- Develop technical skills to use statistical tools and software in data analysis.

Course content

	Topic	L	T	P
	Module 1: Introduction			
1	Mathematical models – deterministic and stochastic; generation of environmental data; stochastic processes in environment; the nature of random variables; populations and samples; parameters and statistics.	1		
	Module 2: Review of Basic Concepts			
2	Measurement theory, levels of measurement; statistical descriptors of data – numerical and graphical; Chebyshev's theorem; measurement uncertainty – accuracy, precision, and bias. Probability theory: probability concepts; probability distribution functions and their applications – discrete and continuous distributions.	8	2	
	Module 3: Data sampling			
3	Types of sampling designs –probability and non-probability sampling; sampling theory, sampling distributions; parameter estimation, point and interval estimates; confidence interval estimation of – means, differences of means, proportions, difference of proportions, variances, ratio of variances sample size determination for different sampling designs	12	6	
	Module 4: Tests of hypothesis			
4	Hypothesis testing – parametric and non-parametric tests (concerning means, differences of means, proportions, difference of proportions, variances, ratio of variances)	12	5	
	Module 5: Correlation and simple regression analysis			

5	Correlation analysis: graphical analysis, bivariate correlation, covariance, correlation coefficient, distribution of correlation coefficient and its statistical significance. Simple regression analysis: assumptions and definitions, principle of least squares, regression parameters their distribution and statistical significance, applications in process description and prediction	12	2	
	Total	45	15	

Evaluation criteria

- Minor Test 1: Written test [at the end of teaching of modules 1 and 2] -- 15%
- Minor Test 2: Written test [at the end of teaching of module 3 and 4] -- 15%
- Major Test: Written test [at the end of the semester, full syllabus] -- 50%
- Assignment: 20%

Learning outcomes

After completing this course, the students will be able to

- explain how data is collected, organized, and stored. [Module 1 and 2; Minor Test 1]
- build, and prepare data for use with a variety of statistical methods and models. [Module 3; Minor Test 1]
- analyse the relation between variables by using correlation and regression methods and illustrate significance of hypothesis testing [Module 3 and 4; Minor Test 2]
- extract information and draw scientific inference from data to solve real world problems. [All Modules; Major Test]

Pedagogical approach

- The course will be delivered through lectures that will focus on developing a clear foundation of various statistical tools and techniques.
- The course will also focus on classroom discussions and practical to give students the analytical knowledge to work with data and understand the problems.

Reading resources

Ayyub, B.M., McCuen, R.H. (2016). *Probability, Statistics, and Reliability for Engineers and Scientists*. CRC press.

Bruce, P., Bruce, A., Gedeck, P. (2020). *Practical Statistics for Data Science*. O'Reilly Media, Inc.

Chan, S.H. (2021). *Introduction to Probability for Data Science*. Michigan Publishing.

Johnson, R.A. (2009). *Miller & Freund's Probability and Statistics for Engineers*. PHI Learning Pvt Ltd., New Delhi.

Kottegoda, N.T., Rosso, R. (2008). *Applied Statistics for Civil and Environmental Engineers*. McGraw-Hill, International Edition.

Matloff, N. (2019). *Probability and Statistics for Data Science*. Chapman and Hall/CRC.

Miller, J.D. (2017). *Statistics for Data Science*. Packt Publishing.

Moore, D.S., McCabe, G.P. and Craig, B.A. (2009). *Introduction to the Practice of Statistics*. W.H. Freeman and Co., New York.

Montgomery, D., Runger, G.C. (2003). *Applied Statistics and Probability for Engineers*. John Wiley.

Walpole, R.E., Myers, S.L., Myers, R.H. and Ye, K.E. (2011). *Probability and Statistics for Engineers and Scientists*. Pearson Education, Prentice Hall.

Suggested Readings

Gupta, S.C., Kapoor, V.K. (2014). *Fundamentals of Mathematical Statistics*. Khanna Publications.

Helsel D.R. and Hirsch R.M. (1997) *Statistical Methods in Water Resources*, Elsevier Science Ltd., UK.

Hoshmand A.R. (1997). *Statistical Methods for Environmental and Agricultural Sciences*, CRC Press.

Ross, S.M. (2014). *Probability and Statistics for Engineers and Scientists*. Academic Press.

Shaefer, S.J. and Theodore, L. (2007). *Probability and Statistics Applications for Environmental Science*. CRC Press, Boca Raton.

Soong, T.T. (2004). *Fundamentals of Probability and Statistics for Engineers.* John Wiley & Sons Ltd.

Wilks, D.S (1995). *Statistical Methods in Atmospheric Sciences: An Introduction*. Academic Press, Inc.

Student responsibilities

The students are required to come prepared with readings that are suggested during the class. They are also expected to participate and further strengthen their understanding of concepts through the practical.

- **Reviewer 1 – Dr. Sumanth Chinthala,** Assistant Professor, Department of Civil Engineering, National Institute of Technology, Warangal
- **Reviewer 2 – Dr. Krishan Kumar,** Professor, School of Environmental Sciences, JNU