ESSENTIALS OF DATA SCIENCE WITH R SOFTWARE - 2: SAMPLING THEORY AND LINEAR REGRESSION ANALYSIS

PROF. SHALABH

Department of Mathematics and Statistics
IIT Kanpur

PRE-REQUISITES : "Introduction to R Course" and "Essentials of Data Science With R Software – 1 - Probability and Statistical Inference" are preferred. Mathematics background up to class 12 is needed. Some minor statistics background is desirable.

INTENDED AUDIENCE : UG students of Science and Engineering. Students of humanities with basic mathematical and statistical background can also do it. Working professionals in analytics can also do it.

INDUSTRIES APPLICABLE TO : All industries having R & D set up will use this course.

COURSE OUTLINE :

Any data analysis is incomplete without statistics. After getting the data, the statistical tools aims to extract the information hidden inside the data. Sampling theory and regression analysis are two important tools among others which play a fundamental role in extracting such information. The role of such classical topics of statistics are to be understood in the context of data science. Such topics have fundamental applicability in data science and are to be understood from computational aspects through software. The introductory tools of sampling theory and regression analysis are detailed in this course. How to use them with the popular free R statistical software R and what are the interpretations of the outcome is the objective of the course to be taught.

ABOUT INSTRUCTOR :

Prof. Shalabh is a Professor of Statistics at IIT Kanpur. His research areas of interest are linear models, regression analysis and econometrics. He has more than 23 years of experience in teaching and research. He has developed several web based and MOOC courses in NPTELincluding on regression analysis and has conducted several workshops on statistics for teachers, researchers and practitioners. He has received several national and international awards and fellowships. He has authored more than 75 research papers in national and international journals. He has written four books and one of the book on linear models is co- authored with Prof. C.R. Rao.

COURSE PLAN :

Week 1: Introduction to data science and Calculations with R Software

Week 2: Basic Fundamentals of Sampling

Week 3: Simple Random Sampling

Week 4: Simple Random Sampling with R

Week 5: Stratified Random Sampling

Week 6: Stratified Random Sampling with R

Week 7: Bootstrap Methodology with R

Week 8: Introduction to Linear Models and Regression and Simple linear regression Analysis

Week 9: Simple Linear Regression Analysis with R

Week 10: Multiple Linear Regression Analysis

Week 11: Multiple Linear Regression Analysis with R

Week 12: Variable Selection using LASSO Regression