# ESSENTIALS OF DATA SCIENCE WITH R SOFTWARE - 1: PROBABILITY AND STATISTICAL INFERENCE

**PROF. SHALABH**

Department of Mathematics and Statistics
IIT Kanpur

**TYPE OF COURSE :** Rerun | Elective | UG/PG
**COURSE DURATION** : 12 Weeks (24 Jan' 22 - 15 Apr' 22)
**EXAM DATE** : 23 Apr 2022

**PRE-REQUISITES :** Introduction to R Course" is 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. The main objective of statistics is to work on a small sample of data but provide conclusions for the whole population. Such results cannot be obtained without learning the concepts and tools of theory of probability and statistical inference. With the advent of data science, it has become important to learn those tools from computational and data based aspects. Without learning the basic fundamentals of probability theory and statistical inference, it is difficult to implement them correctly on the data and draw correct statistical conclusions.

**ABOUT INSTRUCTOR :**
Dr. 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, basic calculations with R Software and probability theory

**Week** 2**:** Probability theory and random variables

**Week** 3**:** Random variables and Discrete probability distributions

**Week** 4**:** Continuous probability distributions

**Week** 5**:** Sampling distributions and Functions of random variables

**Week** 6**:** Convergence of random variables, Central limit theorems and Law of large numbers

**Week** 7**:** Statistical inference and point estimation

**Week** 8**:** Methods of point estimation of parameters

**Week** 9**:** Point and confidence interval estimation

**Week** 10**:** Confidence interval estimation and test of hypothesis

**Week** 11**:** Test of hypothesis

**Week** 12**:** Test of hypothesis for attributes and other tests