# DS303: Statistical Foundations of Data Science

---

**Instructor:** Deepak Prajapati

**Email:** deepakprajapati@iitmandi.ac.in

**Lectures:** Monday 11:00 - 11:50 AM, Wednesday: 9:00 - 9:50 AM and Friday: 10:00 - 10:50 AM

**No. of Credits:** 3

**Prerequisite:** IC110 - Engineering Mathematics, IC252 - Data Science 2

**Books:**

1. Ross, Sheldon. A first course in probability. Pearson, 2014.
2. Grimmett, Geoffrey, and David Stirzaker. Probability and random processes. Oxford university press, 2001.
3. Bishop, Christopher M. Pattern recognition and machine learning. Springer, 2006.

**Course Objectives:**

To understand fundamental concepts of Probability and Statistics.

To understand the concepts that will help to learn more advanced topics of Data Science.

**Lectures and Discussions:**

Classes will be held via Zoom/WebEx/Google meet.

Meeting links will be sent through email.

Lecture notes will also be separately uploaded on the institute moodle.

All the course materials can be accessed from the institute moodle.

**Email Policy:**

I encourage students to reach me by email with their questions and concerns.

Please give me adequate time to answer your queries.

**Grades:**

The grade will be assigned according to performance in the following categories:

**To be announced.**

**If any student is found cheating in any standards will face consequences that contain zero marks in the exam or possible disciplinary action from relevant academic authorities.**

**Syllabus:**

**Probability:** - Sample space, Sigma field, axiomatic definition of probability, conditional probability and independence, Bayes Rule.

**Random variables:** discrete and continuous random variables, probability mass function, probability density function, some standard (important) pdfs, independence, expectation, variance, conditional distribution, conditional expectation, covariance and correlation, Functions of random variables.

**Generating function:** Probability generating function, moment generating function and characteristic functions - properties and applications.

**Convergence of random variables:** basic results, inequalities (Markov and Chebyshev), law of large numbers (weak and strong), central limit theorem.

**Random processes:** Random vectors and covariance and correlation matrix, stationarity, WSS, Autocorrelation, cross correlation, power spectral density, Ergodicity. Wiener processes, Markov processes, Poisson Process.

**Sampling methods:** Inverse transforms sampling, Rejection sampling, adaptive rejection sampling, importance sampling, Markov chains and MCMC.

**Graphical models:** ML and MAP estimation, directed and undirected models, Bayesian networks, CRF, Learning and Inference method (ML, MAP, Sampling).