Discipline Specific Core (DSC): 04 Credits

DSC 7 (III.1) Demystifying the Power of Data: Probability and Statistics

Discipline Specific Core

1. Credit Distribution of the course

Course title & Code DSC 7 (III.1)	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Demystifying the Power of Data: Probability and Statistics	4	3	0	1	Undergraduate	NIL

2. Learning Objectives

This paper aims to provide a basic understanding of the uncertainty in decision-making, testing sample compatibility in day-to-day applications. Students will gain insights into the concepts, contexts, and the tools used to understand these uncertainties. This course will concentrate on probability, random variables, distributions, concepts of the central limit theorem, joint, marginal, and conditional distributions, and computer programs in R and its application to simple models, hypothesis testing, and interval estimation. This course will also expose the students to various real-life applications of probability and statistics.

3. Learning Outcomes

- A good understanding of application of Statistics to daily life problems.
- Able to describe probability distributions and hypothesis formation.
- Able to represent data both graphically and numerically.
- Able to statistically analyze data.
- Create projects using fundamental knowledge in order to provide a hands-on experience of the same.
- Able to work on statistical package R.
- Make data driven and statistically robust decisions in probability based real-life applications.

4. Syllabus **[45 hours]**

Unit I Probability and Random Variables - Sample spaces and events, random experiments, axioms of probability, intersections of events and multiplication and total probability rules, independence, Bayes' theorem, conditional probability, random variables, real life probability problems. **[12 hours]**

Unit II Probability Distributions - Discrete distributions (Bernoulli, binomial, and Poisson) with their properties and applications, Continuous distributions (Uniform, Normal, Exponential) with their properties and applications, concept of central limit theorem. **[12 hours]**

Unit III Curve fitting and Regression Models using R - Joint, marginal, and conditional distributions, curve fitting, regression model, test statistic and their significance. Computer program in R and its application to simple models. **[12 hours]**

Unit IV: Sampling and Testing of Hypothesis - Sampling distributions, hypothesis testing, interval estimation, p-value interpretation, Type I/II errors. **[9 hours]**

5. Practicals **[30 hours]**

- Statistical procedures and their implementation through the statistical package R
- Data import from CSV/excel, cleaning and arranging.
- Introduction to basic syntax of R for arithmetic operations, creating arrays and matrices
- Getting data into R
- Visualization and basic data analysis in R
- Case study: Problems with data
- Mini projects using real, publicly available datasets from different domains

6. Essential Readings

- Bowker, B. & Liberman, G.J. (1972). *Engineering Statistics.* Asia.
- Hogg, R.V. & Tanis, E.A. (1983). *Probability and Statistical Inference*, Macmilan.
- Johnson, N.L. & Leone, F.C. (1977). *Statistics and Experimental Design in Engineering and the Physical Sciences*, Vol. I and II, Wiley Inter Science.
- Montgomery, D.C., Fricker, R. & Rigdon, S.E. (2024). *Introduction to Probability and Statistics for Data Science: With R*. Cambridge University Press.
- Matloff, N. (2020). *Probability and Statistics for Data Science: Math + R + Data.* CRC Press.
- Chan, S.H. (2021). *Introduction to Probability for Data Science*. Michigan Publisher.

7. Suggestive Readings

- Montgomery, D.C. & Runger, G.C. (2018). *Applied Statistics and Probability for Engineers*, 7th edition. John Wiley and Sons.
- Miller, I., Freund's, J.E. & Johnson, R.A. (2017). *Probability and Statistics for Engineers*, 9th edition. Pearson.