MS 6031 Data Analysis for Research

Instructor: R K Amit*
TA: Yash Gopal Deshpande†

Research Level Course
Semester-I: 2026-2027
Slot: QQ; Venue: CRC 302

Prerequisite: Eagerness to Learn

I shall never believe that God plays dice with the world.

Albert Einstein

Life is the art of drawing sufficient conclusions from insufficient evidence.

Samuel Butler

About the Course

In the networked world, we are inevitably part of the data economy, and making sense of data is becoming one of the most important levers of economic growth. The main learning objective of the course is to make rational inferences from data using a unique blend of ideas from multiple disciplines, such as statistics, management, economics, and computer science. The course covers core concepts of data analysis and views them through a machine learning lens. The course discusses the Bayesian, Frequentist, and Fisherian schools of inference while developing data-theoretic intuition. The course is designed for any curious student who wants to venture into modern data analysis.

Readings

- **R K Amit, Fourteen Lectures on Data Analysis**, Manuscript under preparation.
- Morris H. Degroot and Mark J. Schervish, Probability and Statistics, Addison-Wesley, Fourth Edition, 2002. Reprinted in India by Pearson.
- Fred N. Kerlinger, Foundations of Behavioral Research, Cengage Leraning. Reprinted in India by Surjeet Publications.
- Stephen M. Stigler, The Seven Pillars of Statistical Wisdom, Harvard University Press, 2016.

Grading

The grading for this course will be based on relative performance.

End-Quarter Exam: 50%

Class Discussion and Participation: 30%

Poster (Max. 4 students): 20%

*Office: DoMS 504; Phone: 4575; email: rkamit@iitm.ac.in
†ms25d200@smail.iitm.ac.in
Phones must be silent during the classes.

Code of Honor

The participants must adhere to IIT Madras and DoMS policies. The participants are expected to go through the reading material and come prepared in the class whenever it would be announced. Group work has to be done as a team work. While evaluating for the group work, anyone participant from the group can be called for randomly to answer the questions related to assigned work.

Course Schedule & Topics

Lecture	Topics
Lecture 1	**Primitives** 1. Introduction: Why Do We Need Data Analysis? 2. Knowledge, Chance, Uncertainty 3. Galton Board and the Law of Large Numbers 4. Mathematics, Statistics, and the Logic of Scientific Inquiry 5. What is Research and the Principle of Falsification? 6. Regression vs Curve Fitting 7. Correlation vs Causation

Lecture 2	**Reasoning about Uncertainty and Knowledge**
	1. Probability, Chance, Uncertainty, Randomness
	2. Types of Uncertainty: Aleatoric & Epistemic Uncertainty
	3. Schools of Statistical Inference
	  ◦ Frequentist Approach
	  ◦ Bayesian Approach
	  ◦ Fisherian Approach
	4. Hidden Information
	  ◦ Origin and Evolution of Management
	  ◦ Management Incentive Problem
	  ◦ Adverse Selection and Moral Hazard
	  ◦ Risk Preferences and Decision Making
	  ◦ Stochastic Dominance
	5. Gödel's Incompleteness Theorems
	6. De Finetti's Theorem
	7. Benford's Law

Lecture 3	**Building Blocks of Data Analysis-I** 1. Types of Data 2. Scales of Measurement 3. Cumulative Distribution Function (CDF) and Probability Distribution Function (PDF) 4. Measures of Central Tendency 5. Measures of Dispersion 6. Measures of Association 7. Some Common Distributions & its Characteristics 8. Distributions Derived from Gaussian Distribution

Lecture 4	**Building Blocks of Data Analysis-II**
	1. Why are errors Gaussian?
	2. Distributions of Sample Statistics
	3. Estimation of μ and σ
	4. Central Limit Theorem
	5. Causal Inference
	  ◦ Regression Discontinuity Design
	  ◦ Difference-in-Differences

Lecture 5	**Pillars of Statistical Wisdom** 1. Aggregation 2. Information: 3. Likelihood 4. Intercomparison 5. Regression 6. Design 7. Residual
Lecture 6	**Estimation** 1. Logic of Induction 2. Induction vs Deduction 3. Goodness Properties of Estimators: Efficiency, Consistency, Unbiasedness 4. Efficiency of Estimators: Minimum Variance Bound (MVB)
Lecture 7	**Estimation** 1. Maximum Likelihood Estimators 2. Maximum Simulated Likelihood 3. Confidence Interval: Classical Approach 4. Limitation of Inference
Lecture 8	**A Detour Through Bayesian Statistics** 1. Bayes Postulate 2. Conjugate Prior, Jeffrey's Prior & Harsanyi Prior 3. Confidence Interval: Bayesian Approach
Lecture 9	**Hypothesis Testing** 1. Hypotheses Testing: Simple and Compound 2. Types of Errors 3. Uniformly Most Powerful Test 4. The Shadow Price of Statistical Power
Lecture 10	**Hypothesis Testing** 1. The Neyman–Pearson Theory 2. Likelihood Ratio Test (LRT) 3. A Note on *P*-Value 4. Do Two Samples Have the Same Mean? 5. Comparing the Means of Paired Samples 6. Do Two Samples Have the Same Variance?
Lecture 11	**Non-parametric Statistics** 1. Tests of Fit: Sign tests, Kolmogorov-Smirnov test 2. Robust Tests for Means: Wilcoxon's U statistic
Lecture 12	**Computational Statistics** 1. Computational ideas in inference 2. The Jacknife and the Bootstrap 3. Connecting machine learning with inference: random forests and boosting 4. Boosting and game theory

