

DA 241: Statistical Foundations for Data Science

Basic Information

- Instructor: Rhythm Grover
- Office: 2nd floor, CCC Building
- Email: rhythmgrover@iitg.ac.in
- Webpage: <https://rhythmgrover.github.io>
- Teaching Assistant: To be decided

Course Logistics

- Schedule: This course has three lectures a week scheduled in Slot B.
 - 9:00 am - 9:55 am Monday,
 - 10:00 am - 10:55 am Tuesday,
 - 11:00 am - 11:55 am Wednesday
- Venue: 5103, Core 5.
- Attendance is compulsory, however attendance will be taken on random days.

Course Description

This course is an introduction to statistical ideas and tools, underlying the foundations of data science. The course is broadly divided into 5 modules:

- Module 1: Descriptive Statistics
- Module 2: Probability & Random variables
- Module 3: Estimation & Inference
- Module 4: Statistical Modeling
- Module 5: Statistical Computing

Course Syllabus

Elements of descriptive statistics, averages, dispersion, skewness, quantiles; graphical displays, pie charts, bar charts, histograms, scatter plots, box plots, steam and leaf plots.

Probability spaces, conditional probability, independence; Random variables, distribution functions, probability mass and density functions, functions of random variables, standard univariate discrete and continuous distributions; Mathematical expectations, moments, moment generating functions, inequalities; Multidimensional random variables, joint, marginal and conditional distributions, conditional expectations, independence, covariance, correlation, standard multivariate distributions, functions of multidimensional random variables; Forms of convergence, law of large numbers, central limit theorem.

Sampling distributions; Point estimation - estimators, minimum variance unbiased estimation, maximum likelihood estimation, method of moments estimation, Cramer -Rao inequality, consistency; Interval estimation; Testing of hypotheses - tests and critical regions, Neymann-Pearson lemma, uniformly most powerful tests, likelihood ratio tests.

Linear regression, ANOVA, discriminant analysis.

Computing techniques, cross-validation, bootstrap re-sampling .

Course Evaluation

There will 5 surprise quizzes and 5 assignments, a mid-semester examination and an end-semester examination with the following weightage:

- Quizzes: 20%
- Assignments: 20%
- Mid semester exam: 30%
- End semester exam: 30%

Honor Code

Be honest and transparent with your exams, quizzes, and assignments. Any form of cheating is unacceptable and will lead to disciplinary actions.

Course materials and references

- We will maintain a course webpage: <https://rhythmgrover.github.io/DA241.html>
- We will add references on the webpage as we move along.