

Syllabus for MA 5750

Applied Statistics, Fall 2025

INSTRUCTOR INFORMATION

INSTRUCTOR: Rakhi Singh
CLASS MEETING: Tues @5, **Wed @11**, Thurs @9, and Fri @8 (avg. 3 classes per week)
CLASS LOCATION: KCB 505
OFFICE: KCB 642
E-MAIL: rakhi@smail.iitm.ac.in

COURSE INFORMATION

Description. The course is expected to provide an introduction to the basic concepts of probability and statistics along with a brief period invested in learning linear regression and its extensions to generalized linear models. It is intended for graduate students who have one undergraduate statistics course and who wish to review the fundamentals before taking additional statistics courses and/or the students who are wish to understand statistics for it being helpful in the industry jobs. By the end of the course, you should expect to

- appreciate and understand the role of statistics in the field of study of your interest.
- Develop an ability to apply appropriate statistical methods to summarize and analyze data for some of the more routine experimental settings.
- Make sense of data and be able to report the results in appropriate table or statistical terms for inclusion in your thesis or paper.
- Perform appropriate statistical techniques using R and interpret the results/outputs.

This course will be adapted from the online course at Penn State's webpage available here: <https://online.stat.psu.edu/statprogram/stat500>

TEXTBOOK(S)

I will primarily be referring to the following two books:

Walpole, Ronald E.. (2012). Probability and Statistics for Engineers and the Scientists 9th ed. (9th ed.). Prentice Hall.

Ott, R. L., and Longnecker, M. (2016). An Introduction to Statistical Methods and Data Analysis, 7th Edition, Cengage Learning.

You can also use the book by Sheldon M Ross (which is available in our library). While you are not required to buy these books for the class, it is advisable that you read at least one of the books cover to cover. I will also be occasionally referring to:

Agresti, A., and Kateri, M. (2021). Foundations of statistics for data scientists: with R and Python. Chapman and Hall/CRC.

Chihara, Laura M., and Tim C. Hesterberg. Mathematical statistics with resampling and R. John Wiley & Sons.

Some knowledge of the statistical software package R is required. Some parts of the homeworks will typically be in R. There are many online resources where you can learn the basics of R. For example,

- An Introduction to R (<https://cran.r-project.org/doc/manuals/R-intro.pdf>);
- R tutorial by Kelly Black (<https://www.cyclismo.org/tutorial/R/>);
- The undergraduate guide to R by Trevor Martin (<https://biostat.jhsph.edu/~ajaffe/docs/undergradguidetoR.pdf>); and
- a pointer to R bloggers (<https://www.r-bloggers.com/>).

COURSE PLAN

Time permitting, this is the plan for the semester:

Lesson 0	Overview	Week 1
Lesson 1	Revision: Collecting and Summarizing Data	Week 1
Lesson 2	Revision: Probability	Week 1
Lesson 3	Revision: Random variables and Probability Distributions	Week 2
Lesson 4	Sampling Distributions	Week 3
Lesson 5	Point Estimates and Confidence Intervals	Week 4–5
Lesson 6a	Hypothesis Testing for One-Sample Proportion	Week 6
Lesson 6b	Hypothesis Testing for One-Sample Mean	Week 7
Lesson 7	Comparing Two Population Parameters	Week 8
Lesson 8	Linear Regression Foundations	Week 9–10
Lesson 9	Introduction to ANOVA	Week 11
Lesson 10	Generalized linear models	Week 12
Lesson 11	Introduction to Nonparametric Tests and Bootstrap	Week 13
Lesson 12	Chi-Square Test for Independence	time permitting

GRADING

Your grade will be based on your performance on

- quiz 1 (12 percent),
- quiz 2 (20 percent),
- public-media based assessment (8 percent),
- final exam (60 percent),

Ungraded homeworks and their solutions will be provided.

Public-media based Assessment (PMA): In the fields of financial markets, pandemic, sports, judiciary, and policy making, you may have noticed how the same data can be interpreted differently by different people. Critical thinkers need to be able to use the tools of probability and statistical analysis to meaningfully interpret data. We will use public media-based resource selected by you with our mutual agreement. I will give representative examples of podcasts/news articles/YouTube videos/opinion polls. The PMA will involve thinking about how we communicate data to the general public and the tools used for interpreting real-life data. Detailed information on the PMA will be provided during the Week 3. The media will be decided by the end of Week 6. Depending on the strength of the class, it might be feasible to make groups of at most 2 people. The assessment will be based on a 3-minute thesis style presentation during the final two-three weeks of lecture (one of the lectures in each week.) See here for an example: <https://www.youtube.com/watch?v=zqb6Sq4Ml68> and here for tips on making it successful: https://www.youtube.com/watch?v=D_1MK3Ub9QM.

Exams: The final end-semester exam will be comprehensive. Quizzes 1 and 2 will be held in the institute-planned slots. The material covered in quizzes will remain flexible and will be conveyed in due course of time.