

# Statistics 5790: Introduction to R

**Instructor:** Ranjan Maitra (pronounced Ron-jone Moi-tro, if you care)

**Office:** Snedecor Hall 2107      e-mail: maitra

**Course website:** This course uses Canvas. See <https://canvas.iastate.edu>

**Instructor Office hours:** Snedecor Hall 2107. TR: 11-11:50 am, also accessible online

**Teaching assistant:** Gautham Venkatsubramaniam      e-mail: gautham

**TA Office hours:** WF 12:00-1:00 pm, Snedecor Hall 1404.

**Lecture times:**

Section 1: R 12:10–2:00, Snedecor 3121

**Recommended (not required) text:** A First Course in Statistical Programming with R by W. John Braun and Duncan J. Murdoch, *Second Edition*, Cambridge University Press.      ISBN: 978-1-10757646-9.

(*I consider this text to be generally helpful, but not directly related to the class materials.*) **Lecture Recordings:** As is customary with most of my graduate classes, all lectures will be recorded for easy reference by students to material covered in the class. However, technical failures do happen once in a while, and in any case, the recordings are not a substitute for coming to class, so please attend class!

**Lecture Audit:** Audit students must be officially registered on Accessplus. They will enjoy all the rights students registered for credit but will not be evaluated or require to turn in homework or projects. “Sitting in” in the class without being registered hurts the department’s abilities to provide more TA support and is not allowed by the Registrar, and will not be allowed.

**Prerequisites:** Concurrent registration in STAT 5000, or equivalent. Solid grounding in linear algebra, calculus and introductory but graduate-level statistics. **No Exceptions.**

**Important Dates:**

October, second/third week: Midterm Exam (timed 4-hour take-home).

Final: tbd (timed 4-hour take-home).

**Course schedule:**

An introduction to the logic of programming, numerical algorithms, and graphics. The statistical package R will be used to demonstrate how data can be stored, manipulated, plotted, and analyzed using both built-in functions and user extensions. Concepts of modularization, looping, vectorization, conditional execution, and recursion will be emphasized.

**Course information and Policies:** This class follows the policies detailed at <https://celt.iastate.edu/prepare-and-teach/design-your-course/syllabus-statements/> on **Free Expression, Academic Dishonesty, Accessibility, Discrimination and Harassment, Mental Health and Well-being, Prep Week, Religious Accommodation.**

**Contact Information for Academic Issues:** If you are experiencing, or have experienced, a problem with any of the above statements, email [academicissues@iastate.edu](mailto:academicissues@iastate.edu).

**Electronic device policy:** The *unauthorized* use of electronic devices and laptops is prohibited. You are part of a learning community and these devices are disruptive to the class and instructor. If I see someone on a device I will initially make a polite reminder about this policy and then enforce it with greater firmness.

**E-mail:** Checking your e-mail on a **daily** basis is necessary as course communication will be using your University e-mail account. Occasions may occur when the Instructor may need to disseminate important course updates via e-mail.

**Assessments:**

**Final grade:** Letter grades including plus/minus will be given based on performance on the assessment categories above. The percentage distribution is as follows:

Midterm Exam	25%	Homeworks	40%	Final Exam	35%	Bonus	1%
--------------	-----	-----------	-----	------------	-----	-------	----

**Homework:** Unless otherwise announced, homework assignments are **due every Saturday by 11:59 pm**. These are to be turned in as a pdf document, only at the Canvas site by the date the homework is due. **Homeworks turned in in paper form or by e-mail can not and WILL NOT not be graded because there is no record created on the Canvas site. Homeworks not submitted on Canvas will not be graded. The TA will provide feedback on Canvas. Please contact him if you need more feedback.** I recommend that you compare your answers with the solutions that I post and ASK me if you have questions (or have better ideas at solving the problems). I encourage you to collaborate on the homework problems, but for purposes of learning, to work out your own answers independently. *Please note that you will be permitted only one attempt. Therefore, please “Save to Draft” your homework attempt, and click the “Submit” button only when you are sure that you like what is proposed to be submitted, Please again note that this submission is a one-time deal.* **Please do not send e-mails to the TA or the instructor with your Homework submissions.** *Because homework solutions will be provided soon after the submission window closes, late homework submissions are not accepted unless with prior permission from the instructor.*

**Exams:** Exams are scheduled in consultation with the class.

**The exam will be held, as is typical for this class, over a four-hour window on a weekend day, though I am also open to having it on a weeknight if so desired and voted on by the entire class.**

**Conflicts with exams:** Accommodations will only be made for students with appropriate university documentation (e.g. NCAA competition). Please provide at least **two (2) weeks notice** of potential conflicts. All other conflicts will count as a missed exam.

**Final exam:** The final exam is cumulative. Check Canvas for the date, time, and place of the final exam.

### **Additional Information:**

**Statistical Software:** This course will use the free statistical software R, available for Windows, Mac, or Linux. We will learn to use this software to assist with probability calculations, statistical inference, and regression topics.

As with any software, there is an initial frustrating learning curve. However, learning this software is important, because R is heavily used in the statistics world. Even though you might not become a statistician (this is true of Section B), it is easy to learn and should be very helpful in your learning of statistics. Because it is freely available and versatile in every sense of the word, you can use it in the future for any projects that may arise. I am available as a resource to help you with this software should any difficulties arise. Not using this software for homeworks when needed is not an option for this class, so please plan accordingly.

Important: This is a graduate class designed for graduate statistics majors in R. Stat 5790 Section 1 is restricted to graduate students in statistics for whom this is a required class in the graduate program. Stat 579 Section 2 (for non-majors) is for quantitative-minded students who desire to learn R **at the same level** as a graduate student majoring in statistics. Stat 579 (both sections) is a fast-paced 5000-level graduate statistics major class in statistical programming; graduate students majoring in disciplines outside statistics and desiring a gentler introduction should consider Stat 4800 offered TR 9:30-11 am in the Spring. *This class is designed to be fast-paced and taught at the graduate level in statistics. It will assume proficiency in the pre-requisites. Do not expect this class to go at a gentle pace.*

**Course Outcomes/Objectives:** The objective of the course is to have students use the open-source statistical software R to implement statistical methodologies and to display and perform statistical analyses of data. Upon successful completion of this class, students should, upon proper application of their faculties and knowledge, be able to display and summarize data and perform statistical operations in an efficient manner.

**How to do well in the class:** The lectures and homeworks contain examples, activities, and discussion to help you learn the material. Therefore, both attending lecture, understanding the material and doing the homework will provide you the most depth to reinforce your learning of statistics.

Working out statistics problems in R is an integral part of learning R and statistical programming and the statistical way of thinking. This means that you must put forth honest effort in your homework assignments and examples in lecture. Mastering the problem-solving skills needed for the examples, and homework will likely result in a high grade for this course.

This is a one-credit-hour class. As per the department and the university catalog, each week students should spend two 50-minute periods in lecture and six hours outside of class learning the material. Students should be prepared to put forth this kind of effort in order to master the large amount of covered material. Do not fall behind and expect to catch up later in the semester!

**Feedback:** I am always open to feedback about the syllabus, the course, and its policies. I seriously consider any feedback. If a change to the syllabus is deemed necessary, I will discuss the change in lecture and post the updated syllabus on Canvas. Accordingly this syllabus is subject to change.