Econ 2300 - Statistics for Economists
Spring 2023

Course intructor: Karun Adusumilli

Email: akarun@sas.upenn.edu

Office: Perelman building, 631

Office hours: Tuesday 4-6pm on Zoom

Recitation Instructors: 1. Mahdi Shahrabi, Office hours: TBA

2. David Mao, Office hours: TBA

3. Olivia Gilbert, Office hours: TBA

Course description: This course will teach you how to learn from data and understand uncertainty using the ideas of probability theory and statistics. After completing this course, you will be able to carry out statistical analyses yourself on the computer package R.

Prerequisites: The prerequisite for this course is multivariate calculus (Math 104 followed by 114 or 115). To do well in this course you will need to be comfortable with algebra, manipulating sums, differentiation and partial differentiation, solving unconstrained optimization problems, and integration.

Required Text: There are two main textbooks for this course:

1. 'Introductory Statistics', Sheldon M. Ross, 4th edition, Elsevier. This book can be accessed online from the UPenn library. While I suggest that you complete the assigned readings, my lecture slides, which will be posted online at the start of each week, are the final authority on course material. In particular, you are not responsible for material in the textbook unless it is also covered in lecture, but you are responsible for material from the lectures even if it is not covered in the textbook.

2. 'Hands-on programming with R', Garrett Grolemund. This textbook teaches you the basics of programming in R. You can freely access it online here: https://rstudio-education.github.io/hopr/. We will also cover some additional programming material from 'R for data science', Hadley Wickham & Garrett Grolemund. This can also be freely accessed from here: https://r4ds.had.co.nz/. If you prefer, you can order printed versions of these books, but the online ones are more up-to-date.

Required Software: We will use the statistical package R via a front-end called RStudio throughout the course. Both R and RStudio are free and open source. Installation instructions appear on the last page of this syllabus. You will be taught to use R through a series of tutorials. Additional R resources are listed on the last page of this syllabus. Make sure

you download and start using R as the exams will involve coding and running data analysis on R.

Departmental Course Policies: All Economics Department course policies are in force in Econ 2300 even if not explicitly listed on this syllabus.

Academic Integrity: All suspected violations of the code of academic integrity as set forth in the Pennbook will be reported to the Office of Student Conduct. Confirmed violations will result in a failing grade for the course.

Ed Discussion: We will be using an online discussion forum called Ed Discussion, accessible via Canvas, for all written communication in this course. We will use Ed Discussion to make course announcements, answer questions about course material and respond to private messages from individual students regarding personal issues. By asking your question and getting an answer on Ed Discussion, you create a positive externality: other students benefit from your questions and you benefit from theirs. You can even post anonymously if you like. Another benefit is that you can even post code snippets. The instructor and RIs will actively moderate Ed Discussion both to answer questions and approve (or correct) answers written by your fellow-students. All written communication for Econ 2300 should be directed to Ed Discussion, not to the instructors' personal email accounts.

Homework: I will post homework assignments on Canvas on Thursday each week. Homeworks will be collected and graded. The deadline for submitting the homework is Thursday, 12 noon ET, the week after it was posted. Group work is encouraged, but you will have to submit your own answers.

Grading and Assignments

Grades for this course will be determined based on 10 homeworks, and two mid-terms. Specifically,

$$\text{Overall score} = (40 \times \text{Homeworks}) + (25 \times \text{Mid-term I})$$
$$+ (35 \times \text{Mid-term II}).$$

Course Curve: Following department guidelines, I typically try to target an average GPA in the range between 3.0 and 3.2, or slightly above a B average. In a nutshell, I will give about 30 percent As and A-s, 40-50 percent Bs and 20 to 30 percent Cs. If necessary, I will curve overall course scores (not individual assignments) so that they approximately fall into these ranges. I reserve grades below a C-minus for those cases in which a student fails

to attain a minimum level of basic competence in statistics, an absolute rather than relative standard. The grade boundaries are: A-range = 90-100, B-range = 80-89, C-range = 70-79, D-range = 60-69. (In this case, the top two points of each range would be a "plus" and the bottom two points a "minus.")

Homeworks: There will be 10 homeworks, roughly one every week. We will post homework assignments on Canvas on Thursday each week. The deadline for submitting the homework is Sunday, 11:59pm ET, the week after it was posted. Group work is encouraged, but you will have to submit your own answers. When calculating your homework average, I will drop your lowest score and weight the remaining homeworks evenly.

Exams: There will be two mid-terms: one on March 2 and the other on April 25. The mid-terms are take-home exams of 24hrs duration. You are free to use any resource, including the textbook, course notes, recordings, and even the internet to help you with your exam. However, you are not allowed to discuss the exam with any 'person': this implies talking about the exam with your fellow students is not allowed, nor is posting on online discussion forums. In particular, you are not to email, text, call, chat, or talk to anyone about the exam except with me and your TAs. There will be no makeup midterm: if you miss the midterm, your second mid-term will be worth 60% to compensate; bear in mind that this will typically be much harder than the first mid-term. Sudden emergencies, of course, will be discussed and determined by the undergraduate chair.

Note also that both the mid-terms will involve coding in R. R is tightly integrated into the course material. There is no set percentage of marks devoted to R: it may very well be that it will not be possible to answer any question without learning R.

Submitting homeworks and exams: You should upload you solutions to Canvas as **PDF** files. It is fine to take pictures of your answers and upload them, but you should make sure that you convert all images into pdfs and collate them in the right order in which the questions were asked. You should also ensure that the pdfs are of good quality, and all your answers are legible. Failure to follow all these steps could result in your submission being declined.

It is also important that you check that the solutions are properly uploaded on Canvas by the specified deadline. Changes after the deadline will not be accepted.

Regrade Requests: Exam regrade requests must be made in writing within 10 days of receiving your graded exam. As we re-grade the entire exam, your score could rise or fall.

You may not discuss your answers with an RI or the instructor before submitting a regrade request.

Installing R and RStudio

If you have not downloaded R before. Go to: https://www.r-project.org/ and download the relevant files. The installation process varies with Windows and Mac.

You also need to download and install RStudio from: https://rstudio.com/products/rstudio/. Download the desktop version, which is free.

Updating R. If you already have R, it is important that you update to the latest version. Go to the terminal in RStudio and type `R.Version()`. If the R version is not 4.2.2 - "Innocent and trusting", you need to update it.

If you are using a Mac, simply download the latest version of R from https://www.r-project.org/. RStudio will then automatically open the latest version when you open it. There are two different versions for Intel and Apple ARM (the M1 chip) processors. Make sure to download the right one.

If you are using Windows, first download the package `installr` by typing:
`install.packages("installr")`. Then type: `library(installr)` followed by `updateR()`. This will then download and install the latest R version.

R Notebooks. R Notebook or R markdown files (with file extension .rmd) are a productive notebook interface to weave together text and code. We will be using R Notebooks to write our code and also convert it into a readable format. To open a new R Notebook from RStudio go to: R Studio > File > New File > R Notebook. Many of the lectures are associated with an R Notebook. Download these from Canvas to your computer. Clicking on them will automatically open in R Studio.

Your RI will go over the basics of R Notebooks in the second week.

Installing Packages. Throughout this course, we will be working with various R packages. Packages are collections of functions and datasets created by the community to perform various tasks such as data processing, machine learning etc. To install any package, you should type: `install.packages("*Name of Package*")`.

Note that the fact you installed a package just means its downloaded onto your computer. To use the package, you need to load it into memory. You can do this by typing: `library(*Name of Package*)`. R only keeps the packages in memory as you need them. So each time you open a new R session, you need to load them again. This is actually a feature since it frees up computer memory from being filled with packages you do not need.

A few important packages we will be using are:

tidyverse: This is actually a collection of R packages for data science. Within this collection, the most relevant ones for us are `dplyr` for data processing and `ggplot2` for data visualization. Note that typing `library(tidyverse)` will automatically load both (but make sure to install it first using `install.packages("tidyverse")`).

knitr: This is a package for compiling R Notebooks into nice html, tex or pdf files. You will need this to compile your homeworks and projects into PDF files for submission.

Make sure to install these two already. We will be downloading many more packages as we delve into the material. Sometimes you may notice that I have loaded a package, e.g., `library(leaps)`, but you have not installed it. If so, you should first install it by typing `install.packages("leaps")` into the console. If you do not do so, you will get an error message.

Further help. Your RI will go over these instructions again in recitations. If you have trouble, ask your RI or the instructor for help in office hours. In the addition to the required reading mentioned previously, here are links to some additional free resources to help you learn R:

- http://cran.r-project.org/other-docs.html
- http://www.twotorials.com/
- http://www.r-bloggers.com/google-developers-r-programming-video-lectures/
- http://cran.r-project.org/doc/contrib/Farnsworth-EconometricsInR.pdf
- http://www.ats.ucla.edu/stat/R/

Table 1. Schedule (Tentative)

Day	Date	Lecture	HW	Exam
Thu	Jan 12	1. Introductory Statistics I		
Tue	Jan 17	2. Introductory Statistics II		
Thu	Jan 19	3. Sample Statistics I; Intro to R	HW1	
Tue	Jan 24	4. Data wrangling in R		
Thu	Jan 26	5. Data visualization	HW2	
Tue	Jan 31	6. Basic Probability I		
Thu	Feb 2	7. Basic Probability III	HW3	
Tue	Feb 7	8. Basic Probability III		
Thu	Feb 9	9. Discrete Random Variables I	HW4	
Tue	Feb 14	10. Discrete Random Variables II		
Thu	Feb 16	11. Discrete Random Variables III	HW5	
Tue	Feb 21	12. Discrete Random Variables IV		
Thu	Feb 23	13. Continuous Random Variables I		
Tue	Feb 28	14. Continuous Random Variables II		
Thu	Mar 2	No class: Mid-term I		Mid-term I
Tue	Mar 7	No class: Spring break		
Thu	Mar 9	No class: Spring break		
Tue	Mar 14	15. Sampling distributions I		
Thu	Mar 16	15. Sampling Distributions II	HW6	
Tue	Mar 21	16. Confidence Intervals I		
Thu	Mar 23	17. Confidence Intervals II; RCTs	HW7	
Tue	Mar 28	18. Hypothesis testing I		
Thu	Mar 30	19. Hypothesis testing II	HW8	
Tue	Apr 4	20. Regression I		
Thu	Apr 6	21. Regression II: Coding session	HW9	
Tue	Apr 11	22. Regression III		
Thu	Apr 13	23. Regression IV	HW10	
Tue	Apr 18	24. Regression V		
Thu	Apr 20	25. Regression VI		
Tue	Apr 25	No class: Mid-term II		Mid-term II