### Wharton Pre-Baccalaureate Program STAT 0001 Introduction to Statistics and Data Science Prof. Gupta

Email: <a href="mailto:shugupta@wharton.upenn.edu">shugupta@wharton.upenn.edu</a>

Class meets: TBD Office Hours: Online. TBD Office Zoom Link: TBD

#### Content

In this course, we will learn introductory statistics using R with a focus on the application of statistical thinking to business problems. We will learn basic statistical concepts such as mean, variance, quantiles and hypothesis testing, and basic R programming for data management and analysis. We will work with traditional R's dataframe structure as well as the modern tibbles structure.

### Course Textbook, Readings, Homework Software

- 1. Course Textbook: Introduction to Modern Statistics by Mine Çetinkaya-Rundel and Johanna Hardin
  - a. A pdf copy of the book has been uploaded on canvas here: https://canvas.upenn.edu/courses/1665731/files/folder/text\_book
- 2. R Textbook: <u>https://r4ds.had.co.nz</u>

#### **Course Structure and Flow**

We have both synchronous and asynchronous components:

Course Content: Chapters 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 from the prescribed text

#### Web- components

There are 2 web-components.

- 1. <u>Canvas</u> is my main communication device with you. Every relevant information could be found on Canvas.
  - On Canvas for STAT 0001, browse left-hand side menu.
    - o Announcements
    - Modules
    - Zoom (Click this to see upcoming Live sessions)
    - Class Recordings (this will be populated after our live sessions)
    - Quizzes
    - Grades
  - Under Modules you'll find:
    - Syllabus
    - Piazza Instructions (i.e. how to access homework) and Answers to Homework
    - Class Recordings (Each live session will be recorded and posted on Canvas for later review)
    - Lecture Notes (PPTs of the text chapters)
    - o Newspaper articles you have to read before lectures

- 2. <u>Piazza</u> will be the site where you receive and complete homework.
  - Instructions on how to sign up are on the Canvas site. Use your Penn email to sign up so we can give you your grade. (Not your gmail, not your other personal email)

### Grading:

- Midterm online (35%)
- Final online (not comprehensive) (35%)
- Weekly HW (Non graded)
- Class Presentation (25%)
- Class participation/attendance + answering questions on piazza (5%)

Your letter grade is NOT based on a curve, it's based on attainment.

90s => A range (A-, A) 80s => B range (B-, B, B+): divided 1/3 of the grade up to 3 decimals 70s => C range (C-, C, C+): divided 1/3 of the grade up to 3 decimals 63.333-69.999 => D+ and D <63.333 => F

For proper Zoom etiquette, prepare to have your camera on. Be up and ready by the start of class.

- **Special needs/arrangements**: Reasonable accommodation will be made for those with special needs. Requests for academic accommodations need to be made during the first two weeks of the semester, except under unusual circumstances. Students must register with Student Disabilities Services (SDS) for disability verification and reasonable academic accommodations. More at: <u>https://www.vpul.upenn.edu/lrc/sds/Weingarten</u>
- Learning Resources Center: The Office of Learning Resources offers individualized instruction and workshops to guide Penn students towards more effective academic study skills, such as time/project management, academic reading and writing, exam preparation, test-taking strategies, and study strategies. More at: <a href="https://www.vpul.upenn.edu/lrc">www.vpul.upenn.edu/lrc</a>
- **Counseling & Psychological Services**: Students with concerns of a personal, emotional, social, or educational nature may visit CAPS for help and guidance free of charge. More at: <u>http://www.vpul.upenn.edu/caps</u>

# A. Late/missed work

No late homeworks are accepted. No make-up for homeworks or mini-tests will be provided because of the flexibility already embedded in their grading. Make-up exams for the midterm and the final are provided for those with legitimate health concerns.

# **B.** Integrity.

Penn has a clear Code of Academic Integrity that we follow and cherish. Distant learning does not validate unethical behavior, and it is your responsibility to familiarize yourself with the code, and to follow it to the letter:

http://www.upenn.edu/provost/PennBook/academic\_integrity\_code\_of

In particular, you are expected to maintain a professional, respectful environment, be courteous to me and others. Plagiarism of any sort is unforgivable, even at the first offence. You will get an F in this class.

Class schedule		
	Text	Your to-do list for each day
Class 1		-Review probability. Teach how to install R, R Studio and the <i>swirl</i> package
Class 2		-Lesson 1 from <i>swirl</i> this was in the way of getting started with R.
Class 3		-Review conditional probability and simulation in R
Class 4		-Review Exploratory Data Analysis. Upload excel data in R and data visualization
Class 5		Exploratory Data Analysis and plot the graphical summaries in R
Class 6		-Finish Exploratory Data Analysis and started with Random variables. Went over problems involving Exploratory Data Analysis In class we'll solve problems, and discuss media articles
Class 7		Review session for midterm and review HW questions
Class 8		-In class midterm
Class 9		<ul> <li>-Expected value and variance of a linear combination of random variables.</li> <li>-Discussed the concept of IID Random variables</li> <li>-Review details of Binomial Distribution</li> </ul>
Class 10		-Review the details of Normal + Continuous Distn and in class problem solving
Class 11		-Class presentations based on the reading exercises assigned on canvas.