

Probability and Introduction to Statistics

ALY6010

Tom Breur

Week 1, 27-OCT-2020

Agenda

- Introductions
- Administrative notes
- Suggestions on how to succeed
- Preparation for Week #2
- Discussion board: Lipozene case

Introductions: Tom Breur



$$\lim_{x \rightarrow \infty} \frac{\pi(x)}{x/\ln(x)} = 1$$
$$\sum_{k=0}^n \binom{n}{k} = 2^n$$
$$\frac{1}{c} \int f(z) dz = 2\pi i (N - P)$$



WHERE I'VE STUDIED



WHERE I'VE LIVED



WHERE I'VE WORKED

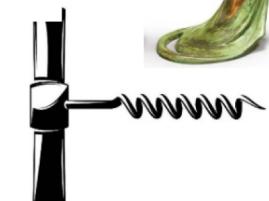


CENGAGE
Learning®

Boston
Scientific



WHAT I DO FOR FUN



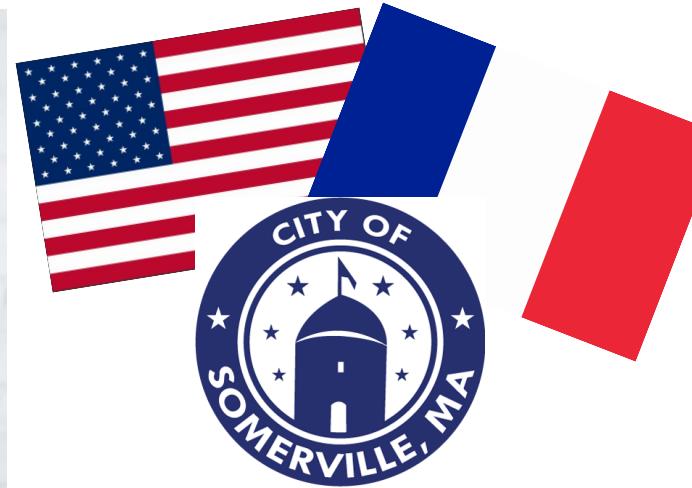
Meet your TA for ALY6010 CRN 71709

Catherine Richard

Email: richard.ca@northeastern.edu

You can reach me via :

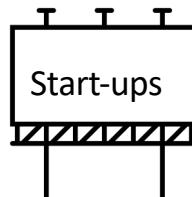
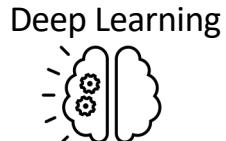
- Email
- Post on Canvas
- WhatsApp (781-526-6300)



Office Hours : Fridays 3pm

Tentative: I will send email survey & modify if certain times/days work better for students

I've worked at and am interested in :



Northeastern University
College of Professional Studies

MPS Analytics, Statistical Modeling Concentration
I'll graduate at the end of next quarter

Introductions: Discussion

- Some “suggestions” that you might include:
 - Background (undergraduate major, work experience, etc.)
 - Why are you interested in taking this Statistics class?
 - What do you hope to learn in this class?
 - One fun fact! Fun facts make you memorable, and help others connect with you
- Preferably ***something nobody in class knows, yet***



Introductions

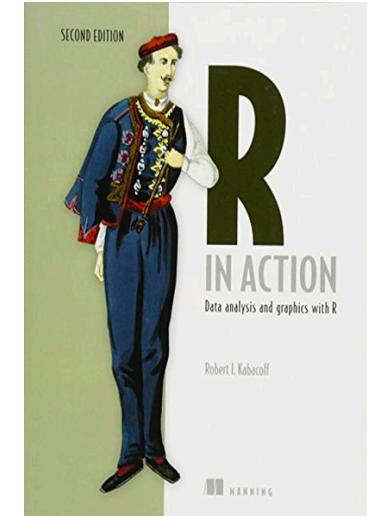
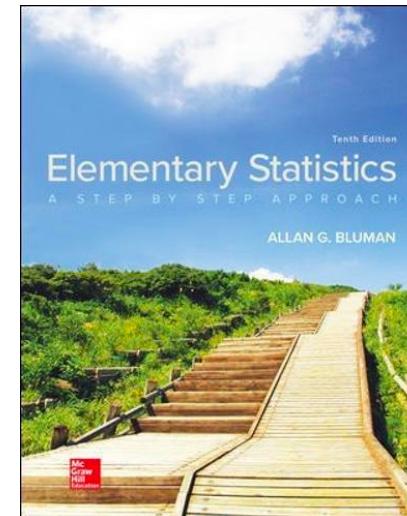
Welcome!

- Introductions:
please take a moment to introduce yourself to your classmates on the Discussion Forum:
 - Discussions/Introduce yourself!/Introduce Yourself to the Class/ reply to this thread, please
 - Building and nurturing a professional network is an extremely valuable asset
- Adding a picture or video to your profile helps others relate to you
- Point to links or references to your accomplishments, professional, and private life

Administrative notes

Prerequisites

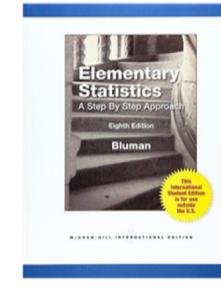
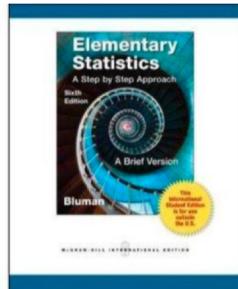
- This class assumes (!) that you have successfully completed ALY 6000, *and have mastered its content*
 - To get the most out of this class, many students who are new to R and hypothesis testing, would do well to periodically review the materials from ALY 6000, as there will be many references to it
- Software and textbooks can be found in the syllabus



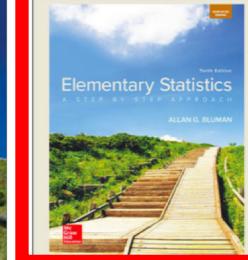
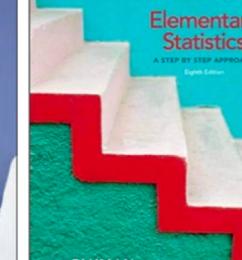
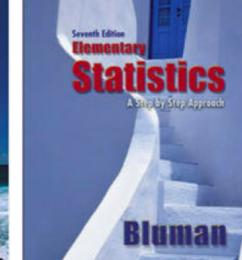
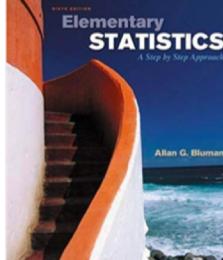
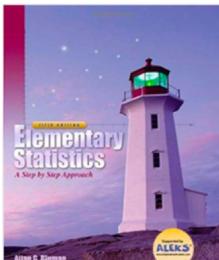
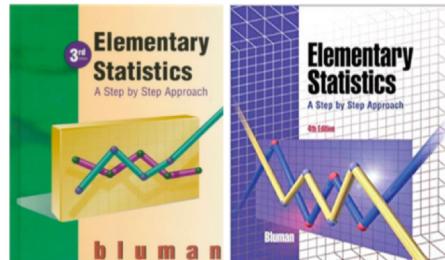
Textbook Bluman

- Older versions of the textbook books may be used, and will generally be effective; however, the “connection” to page references will obviously be lost

Ed.: 3 4 5 6 7 8 9 10



Version in use now



Academic integrity

- All students are required to take the academic integrity quiz, this is a **requirement** to enter this course and pass this class
 - *Minimum* required score is the perfect score of 100
- In case of doubt –*any* doubt at all– I pass cases on to OSCCR for their evaluation
 - This is *not* a punitive measure: sometimes “reeducation” is required, and NEU has an office that is specialized in evaluating, and possibly dissolving these cases

Important notes on grading

- Only teacher, and the teacher *alone* grades all work
- Your TA is *in no way at all* involved in the grading of your work
 - Therefore, any and all questions and remarks you share with your TA have no impact whatsoever on grading
 - Your TA will help you succeed, and provide substantive feedback and guidance, but cannot provide *any* answers
- Important: non-compliance with requirements leads to progressive deductions of grades
- Historically, late submissions and failure to follow APA formatting are the top two causes for grade deduction

Which questions should I direct at teacher personally?

- Please *only* email teacher with questions that need to be treated confidential, or require privacy
- All other substantive questions almost certainly can be posted on the Discussion Forum (Q&A), so *that your peers can also benefit from my response*
 - This way, nobody gets an unfair advantage to succeed
- Do not email me with questions about your grades unless:
 - you have a substantive argument why I should revise
 - you have evidence I have overlooked some of your contribution(s) in my grading
- All other “private correspondence” goes at the expense of time teacher can allocate across class. All of whom have submitted timely, and have worked hard to perform well

Suggestions on how to succeed (in this course)

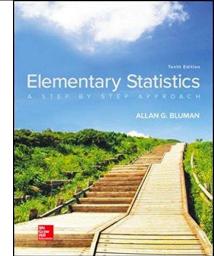
Lessons from the trenches
(AKA “best practices”)

Planning / time management

- The course ALY 6010 is fast paced, so it is imperative you do not allow yourself to “fall behind”
 - Experience shows that students that *do* fall behind, struggle to “catch up”, and tend to get far poorer grades
- The course is divided in many small(-ish) components to help you pace appropriately
 - In practice, you probably need to work on this course every single day to “keep up”, if only for a few hours

NB.: By faculty policy (!), there are ***no*** “do-overs” or opportunities for “make up work” in case of missed submission deadlines or disappointing grades

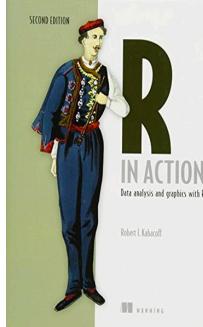
Recommended study method (1)



Bluman (10th Ed.)

- Step 1: review and internalize the chapter “Objectives”
 - For Chapter 6 these are on p. 311, **1 - 7**
 - For Chapter 7 these are on p. 369, **1 - 6**, etc.
- Step 2: read the chapter, until you find the answer/solution to objective **1**
- Step 3: **highlight the text** that provides the answer to **1**
- Step 4: go back to the chapter “Objectives”, find **2**
- Step 5: continue reading the chapter, until you find the answer/solution to objective **2**
- Step 6: **highlight the text** that provides the answer to **2**
- Rinse and repeat

Recommended study method (2)

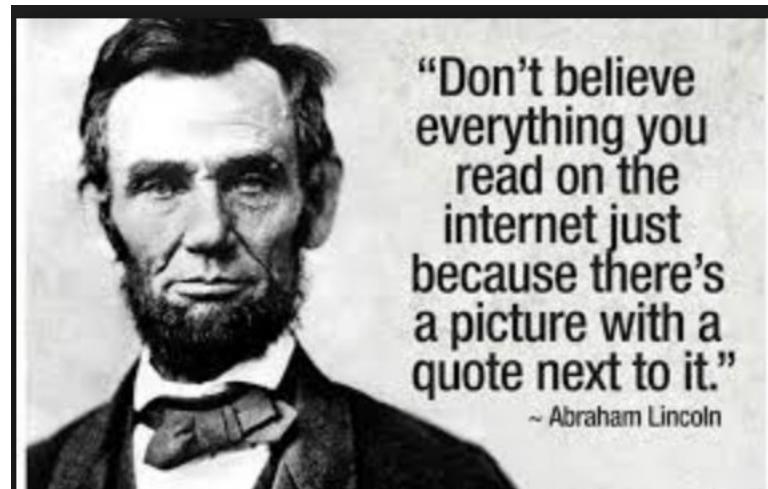


Kabacoff / R Studio

- Practice, practice, practice
 - Programming skills are best acquired by running through exercises, *repeatedly*, and writing lots of code
 - Sometimes slightly tweaking your scripts can be (very) helpful to explore the program's behavior
 - Keep the table with “Help” functions on p. 11 close by!
- Beyond the basic steps of *operating* R, there are a ton of resources online you can tap into, like:
YouTube, (free) online training resources, R community,
<https://www.r-exercises.com/>, etc.

Quoting internet sources

- Be aware that a “Google search” favors hits with high SEO, not necessarily most reliable and credible results
 - Companies that invest the most in SEO or SEM are not always putting in commensurate effort to improve their substantive content
- Always, *always* triangulate claims you read on the web



Preparation week #2

Requirements

- Academic Integrity quiz
- Discussion board:
 - Post contributions on successive (distinct!) days
 - *Minimum* of three posts, but this need not limit you
 - *First* post your primary contribution, *only then* will get access to other peoples' contributions
- Quizzes:
 - Blu 6-1 to 6-4 quizzes
 - M1 R quiz
- Reading preparation week 2 (Chapter 7 Bluman)

Discussion board

- You *first* (!) need to post an original contribution (“primary post”) first, with a minimum (!) of 250 words
- This post needs to contain an academic reference to a reliable (!) and relevant source
 - The reference needs to be set in APA standard
- A minimum (!) of two responses are required, each 80+ words, and posted on distinct, successive dates
- All contributions need to be substantive
 - For clarity: “I agree”, “I like your post”, etc. do not count as substantive replies. *Instead* reason why you agree or disagree, and refer to outside sources to justify your position
- Referring to other sources or posts, or previous classes, earns “brownie points” towards top grades (100 points) for integrative learning

Lipozene ad “pointers”

- Make yourself at least somewhat familiar with the domain of clinical trials
 - Some of the evidence in this ad relies on (implicit!) conventions that apply to clinical trials *in general* (not just studies behind Lipozene)
- Drawing evidence from research is common, just like in this ad
 - But “buywer beware” that the frame or context in which this evidence gets presented is one of the angles you need to consider
- If you find there are flaws in the reasoning that is used in this ad, then how could those concerns be addressed?

Four sampling methods (1)

- Bluman pp 11-16

Most common sampling techniques:

- Random sampling
- Stratified sampling
- Cluster sampling
- Systematic sampling

Read up (outside Bluman) ***before*** recommending a sampling approach,
and expect questions to justify your recommendation

Four sampling methods (2)

- Random sampling
 - Choose elements “at random” – everyone has the same chance of being selected
- Stratified sampling
 - Choose elements “at random” within strata, e.g. choose exactly 50% men & 50% women
- Cluster sampling
 - Sample elements randomly at a higher level of aggregation, e.g. pick households at random, and then survey exhaustively within each household (finer grain)
- Systematic sampling
 - Choose set sequence intervals, like every 100th element. In that case, for instance, select a 1% sample by choosing all account numbers whose CustomerID ends in e.g. *87