BMB 961-301

Gaps, Missteps, & Errors in Statistical Data Analysis

Welcome!

Lecture 1: Introduction and Overview

- Introductions
- Scope & topics
- Website
- Communication
- Activities
- Schedule
- Wrap-up

Introductions

- **Arjun Krishnan** [arjun@msu.edu | @compbiologist | thekrishnanlab.org]
- Assistant Professor
  - Dept. Computational Mathematics, Science, and Engineering
  - Dept. Biochemistry and Molecular Biology
- Research Interests:
  - Computational genomics, Biomedical data science, Statistical modeling, Graph theory, and Machine learning

What’s this course about?

This is an advanced short (1-credit) course designed to:

- Discuss common misunderstandings & typical errors in the practice of statistical data analysis.
- Provide a mental toolkit for critical thinking and enquiry of analytical methods and results.

Prerequisites

We will assume:

1) Familiarity with basic statistics & probability

2) Ability to do basic data wrangling, analysis, & visualization using R or Python.

What’s this course about?

Surveyed biostatisticians regarding questionable requests they receive. Most common:

- Altering some data to support hypothesis
- Interpreting findings on basis of expectation
- Not reporting missing data
- Ignoring violations of assumptions

[These requests are reported by younger statisticians.]

Survey of trainees:

- Pressured by a PI or collaborator to produce “positive” data
- Pressure to publish influences the way they report data.

Ann Intern Med. 2018;169(8):554-558

Clinical Cancer Res. 2018;2(14)

Topic 1: Statistical hypothesis testing

Lectures 2 & 3

- P-value & P-hacking
- Multiple hypothesis correction
- Estimation of error & uncertainty

Topic 2: Experimental design

Lectures 4 & 5

- Statistical power / underpowered statistics
- Sample size calculation
- Pseudoreplication
- Confounding variables & batch effects

Topic 3: Unknown variables, Cognitive biases, & Base rate

Lectures 6 & 7

- Circular analysis
- Regression to the mean & stopping rules
- Confirmation & survivorship bias
- Permutation test

Topic 4: Descriptive statistics, Modeling, Visualization

Lectures 8 & 9

- Describing different distributions
- Continuity errors & model abuse
- Visualization challenges

Topic 5: Reproducibility

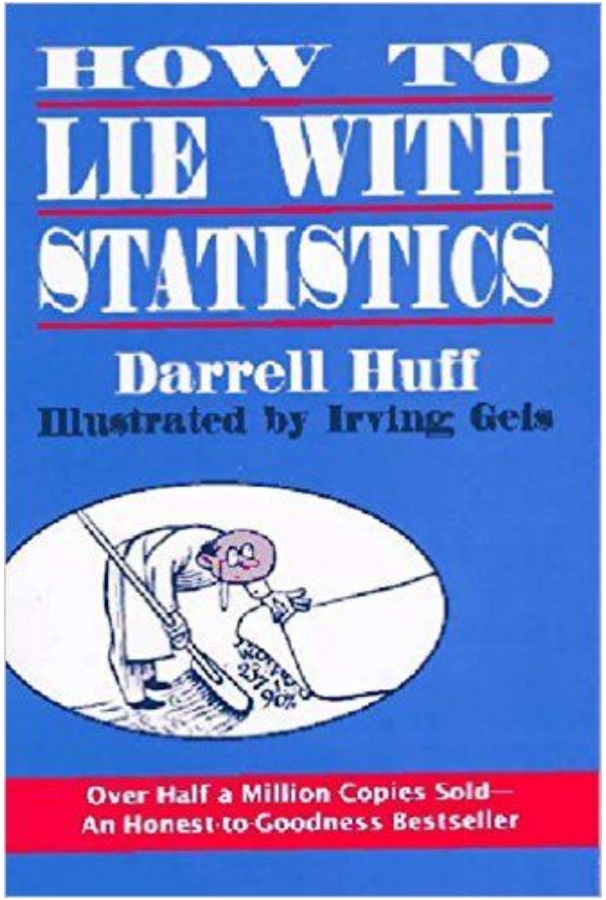
Lectures 10 & 11

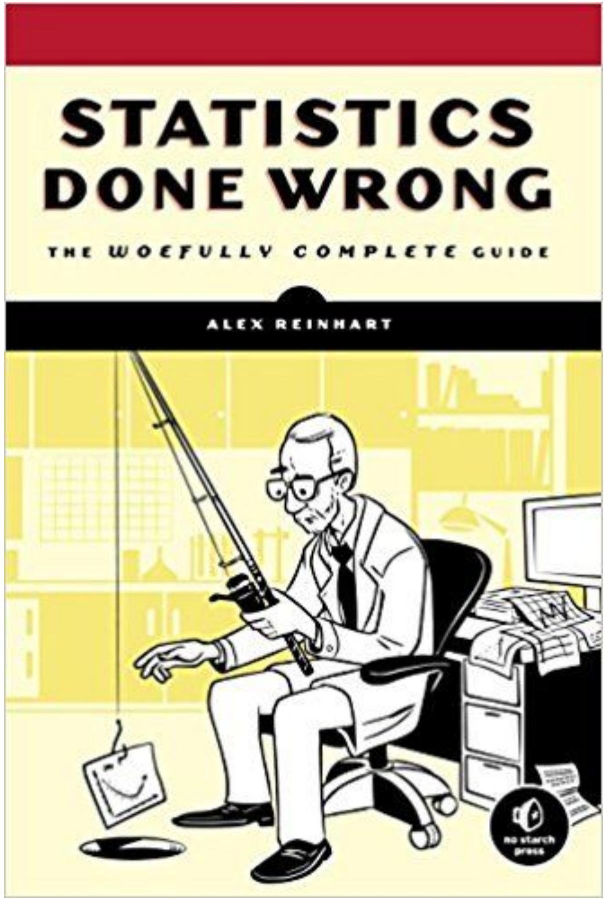
- Researcher degrees of freedom
- Pre-registration of methods & cutoffs
- Data sharing / Hiding data
- Reproducible research

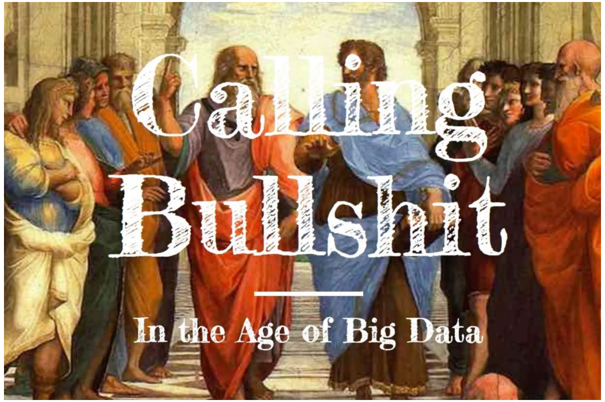
Lecture 12: Roundup

- Recap
- Difference in significance & Significant differences

Resources


HOW TO
LIE WITH
STATISTICS
Darrell Huff
Illustrated by Irving Geis
Over Half a Million Copies Sold—
An Honest-to-Goodness Bestseller





Original research articles

Reviews

Blog posts

Podcasts

Course website

bit.ly/bmb961-nov18

- Contact information
- Course outline and materials →
- Schedule, location, calendar, and office hours
- Website and communication
- Course activities
- Grading information
- Attendance, conduct, honesty, and accommodations

- Lecture slides
- Learning materials
- Assignments
- Notes

Communication

bmb961-statgaps-nov18.slack.com

- The primary mode of communication in this course (including major announcements) will be the course Slack account.
- All of you should have invitations to join this account in your MSU email.

#announcements

#slides-materials

#assignments

#articles-tutorials

#papers

#random

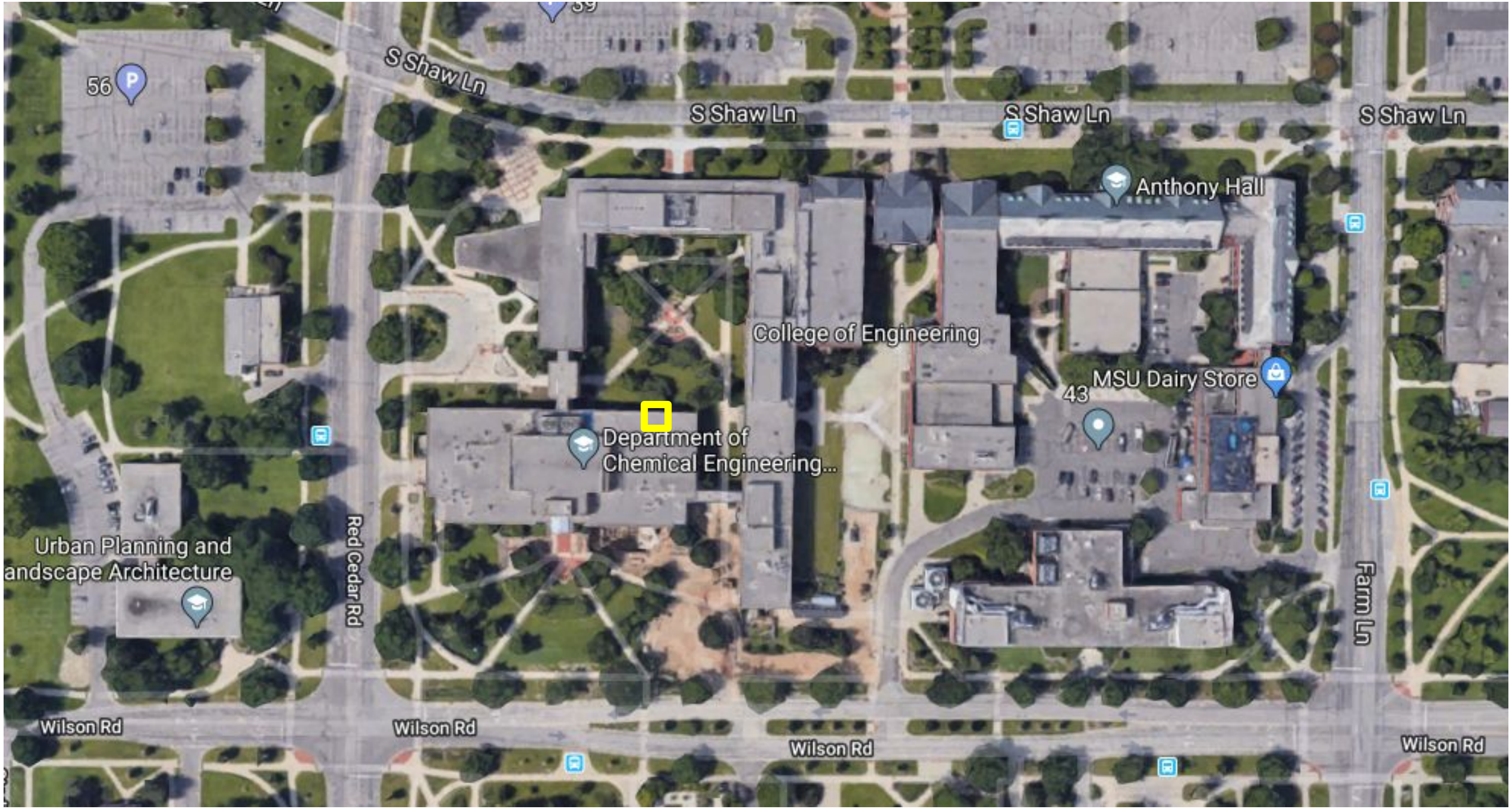
Office hours

bit.ly/bmb961-nov18-incoming

- Select convenient office hours
    - Will give preference to enrolled students
    - Happy to chat in-person but, many times, just messaging on Slack with your questions/concerns might work as well.
    - Happy to coordinate if you can't make it during this window for some reason. Again, just send message me on Slack.

Course Survey: bit.ly/bmb961_nov18_survey

My office: 2507H Engineering Building (2nd floor)



Course activities

- Assignments: ~50%
- Class participation: ~25%
- Final exam: ~25%

Assignments

- For each topic, you will be assigned a reading material after the topic's 1st class (Wed) that you are required to read. Along with this, you might be given a data analysis assignment that you have to complete.
- Submit your assignment ***before*** the topic's 2nd class (Mon).

Class participation

- Do the assignments and additional readings.
- Show up to class.
- Work in groups during in-class discussion sessions.
- No one will have the perfect background.
  - Ask questions about statistical, computational, or biological concepts.
- Contribute the material in-class and on slack.
- Correct me when I am wrong.

Final exam

- A major goal of this course is to prepare your ability to perform and critique statistical data analysis and to present your ideas and results effectively.
- The final exam will test this goal.

What you need to do before the next class

- Join slack and look out for messages on all channels: bmb961-statgaps-nov18.slack.com
- Read the course website: bit.ly/bmb961-nov18
- Fill out the incoming survey: bit.ly/bmb961-nov18-incoming