SYLLABUS
BIOL107, Data Analysis and Statistics Workshop

Contact Details

Dr Paul Miller (Instructor)
Email: pmiller@brandeis.edu
Please contact via email to set up a meeting outside of office hours.

Joshua Harrison (TA)
Email: joshharrison@brandeis.edu

Meeting Times

Classes
Monday, Wednesday, 4:00 – 5:30 PM (via Zoom)

Office Hours
Tues 9:30-10:30 am & Fri 1:30-2:30 pm (with Paul Miller via Zoom or Volen 201 as Fri option)
Tues 4:00-5:00pm and Thur 5:00-6:00 am (with Josh Harrison via Zoom)

Course Description

Learning Goals:
The interpretation of data is key to making new discoveries, making optimal decisions, and designing experiments. Students will learn skills of data analysis through hands-on, computer-based tutorials and exercises that involve experimental data from the biological sciences.

Students will gain familiarity and comfort with the MATLAB computing software, and learn a number of methods for carrying out statistical tests or otherwise analyzing data using MATLAB.

Teaching/learning strategies
Classroom time is devoted to self-guided computational tutorials using MATLAB software, following a short (typically 5-10 minute) introduction by the professor. The TA and professor will be available during class times and office hours to help. Students will be expected to share their screen as requested during class times, or to visit at least one office hour per week if they cannot make class times due to time zone limitations.
The course is designed for students to be actively participating in a hands-on manner, because it is crucial to integrate conceptual thinking with the mathematical and coding skills needed to carry out the analysis of data. The in-class labs are designed to break down, into very small parts, the concepts and skills that are required to address certain questions. The purpose of the homework is to allow students to practice what they have learned and to develop the ability to solve similar problems independently.

The team projects require students to work in groups, much as one would in a research laboratory, in order to share ideas, solutions, and constructive criticisms, and to receive feedback from other members of the group and from other groups.

Prerequisites

A knowledge of basic statistics (mean and median) is assumed. It is helpful, though not essential, for you to have taken a statistics class previously, such as AP Stats, or BIOL51A, or PSYC51A.

Credit Hours:

Success in this four-credit course is based on the expectation that students will spend a minimum of nine hours of study time per week outside of class, to complete problem sets, to work on team projects, to prepare for the next class, and to get advice during office hours.

Brandeis UNIVERSITY

Brief Summary of the Schedule

Wed Aug 26	Lab 1.1: Getting started: Matlab and Data Analysis/Statistics
Mon Aug 31	Lab 1.2: Variables, Functions, and Simple Plots
Wed Sep 02	Lab 1.3: More plotting and (dot) m-files
Wed Sep 09	***PS1.1 due.** Lab 1.4: M-file Management; Differences in Cumulative Histograms
Thu Sep 10	Lab 1.5: Did my drug really work? Statistical inference
Mon Sep 14	Lab 1.6: What was the impact of my drug on average?
Wed Sep 16	***PS1.2 due.** Lab 1.7: Basic sample descriptors; tinkering with plots
Mon Sep 21	Lab 1.8: How well did my drug do? Variance explained and discriminability
Wed Sep 23	Lab 2.1: Describing data with fits (models)
Wed Sep 30	***PS1.3 due.** Lab 2.2: Using fitting as a tool, not like a fool
Mon Oct 05	Lab 2.3: Creating robust index values
Wed Oct 07	Lab 2.4: Dealing with a truckload of data
Mon Oct 12	***PS2.1 due.** Lab 3.1: Time-series (correlation, feature detection, rates)
Wed Oct 14	Lab 3.2: Frequency and Fourier analysis
Mon Oct 19	***Team Project 1 Videos Due.** Lab 3.3: Convolution and filtering
Wed Oct 21	**Team Project 1 Q&A**
Mon Oct 26	***TP1 write-up due.** Lab 4.1: Basic image representations
Wed Oct 28	Lab 4.2: Digital image representations of the physical world
Mon Nov 02	**PS3.1 due.** Lab 4.3: Image analysis I (Regions of Interest, ROI)
Wed Nov 04	Catch-up day
Mon Nov 09	***Team Project 2 Videos Due.** Lab 4.4: Image analysis II (Functional imaging)
Wed Nov 11	**Team Project 2 Q&A**
Mon Nov 16	***TP2 write-up due.** Lab 5.1: High-dimensional space I; data points as vectors
Wed Nov 18	Lab 5.2: High-dimensional space II (clustering; dimensionality reduction)
Mon Nov 30	No Class: Work on Team Project 3.
Wed Dec 02	Lab 5.3: High-dimensional space III (interpolation and PCA)
Mon Dec 07	***Team Project 3 Videos due.**
Wed Dec 09	**Team Project 3 Q&A**
Mon Dec 14	***PS4.1 & TP3 write-up due along with any remaining material by 9pm**

Course Requirements

Brandeis UNIVERSITY

Academic Integrity

Every member of the University community is expected to maintain the highest standards of academic integrity. A student shall not submit work that is falsified or is not the result of the student's own effort. Infringement of academic honesty by a student subjects that student to serious penalties, which may include failure on the assignment, failure in the course, suspension from the University or other sanctions (see section 20 of R&R). Please consult Brandeis University Rights and Responsibilities for all policies and procedures related to academic integrity. Students may be required to submit work to TurnItIn.com software to verify originality. A student who is in doubt regarding standards of academic honesty as they apply to a specific course or assignment should consult the faculty member responsible for that course or assignment before submitting the work. Allegations of alleged academic dishonesty will be forwarded to the Department of Student Rights and Community Standards. Citation and research assistance can be found at Brandeis Library Guides - Citing Sources (https://guides.library.brandeis.edu/c.php?g=301723).

Collaboration Policy

Students are encouraged to discuss the homework sets and help each other, with some limitations:

1) Students should attempt each homework question for 20 minutes before seeking help on that portion from students or the professor.
2) Students may talk about concepts, discuss lines or fragments of code in emails, message boards, or on a whiteboard, and may debate whether various solutions are correct.
3) Students are **not** permitted to exchange complete '.m' files. Each student should write their own '.m' files themselves and understand the code and results that they submit for grading.
4) When you have consulted other students while completing your homework, **credit them** by including their names.

Assignments

Each in-class computer Lab (labeled "Lab1.1", "Lab1.2", etc.) should be completed and handed in before the beginning of the next class.

Each of the seven Problem Sets for homework (labeled "PS1.1", "PS1.2", etc.) should be completed by the beginning of class on the due dates indicated on the schedule. These due dates are typically one week after assignment of the problem set.

Each of the three Team Projects should be commenced at least two weeks before the presentation day given in the schedule. Write-ups of the team projects are due before the start of class on the due dates indicated in the schedule, typically one week after the presentation. You will be, as much as possible, in a distinct team of three for each project. While members of the team will work together, each will be assigned a primary responsibility to either: 1) complete the computer code; 2) present the project to the class orally; or 3) complete the final write-up. Each student should be assigned each of the lead roles across the three team projects.

Participation

Do attend online during the in-class hours if you can. If that is not possible for reasons of time-zone, ensure you attend at least one session of office hours per week so that the instructor or TA can check in on your progress. We can arrange the time of at least one weekly session of office hours to be between 8am and 10pm in your time zone as necessary.

All presentations will be made via Zoom, recorded, and visible on Latte via the Echo 360 link.

Notabene: Materials for each class will be uploaded on NotaBene. All students are expected to ask a question, make a comment, or answer a question, on NotaBene by the end of the day of the class (midnight, Boston time).

Evaluation

The overall course grade will consist of a weighted average of scores for in-class Labs (20%), out-of-class Problem Sets (40%), Team Projects (30%), and NotaBene questions/comments (10%).

Where there is a clear "correct" answer (e. g. the code calculates the desired quantity) credit is awarded for obtaining the correct answer. Where graphs should be plotted, full credit requires appropriate titles and labeling of axes. In team projects, engaging visual and oral presentations, as well as clear responses to questions, are needed for students to receive highest credit. NotaBene credit is awarded for any comment reflecting a thought process.

Essential Resources

Accommodations

Brandeis seeks to welcome and include all students. If you are a student who needs accommodations as outlined in an accommodations letter, I want to support you. In order to provide test accommodations, I need the letter more than 48 hours in advance. I want to provide your accommodations, but cannot do so retroactively. If you have questions about documenting a disability of requesting accommodations, please contact Student Accessibility Support (SAS https://www.brandeis.edu/accessibility/) at 781.736.3470 or access@brandeis.edu.

Course Materials

All course materials are available on the course website on LATTE.

Apps or Tools/Equipment

Students will need to download and install MATLAB software, available to Brandeis students for free via a site license. I recommend you download and install the software before the first class. You will need to use your UNET id to access the instructions found here, which describe how to get Matlab running on your computer.

Feel free to email me if you have spent more than 15 minutes on any one step in the process and it does not work! pmiller@brandeis.edu

LATTE

LATTE is the Brandeis learning management system: http://latte.brandeis.edu. Login using your UNET ID and password. All class materials are available on the site website on Latte.

Library

The Brandeis Library collections and staff offer resources and services to support Brandeis students, faculty and staff. These include workshops, consultations, collaboration, materials and instruction on emerging trends in technologies such as machine learning, emerging trends in research such as data visualization, and emerging trends in scholarship such as open access. Librarians at the Circulation Desk,

Research Help Desk, Archives & Special Collections, Sound & Image Media Studios, MakerLab, AutomationLab, and Digital Scholarship Lab are available to help you.
https://www.brandeis.edu/library/about/index.html

Student Support
Brandeis University is committed to supporting all our students so they can thrive. The following resources are available to help with the many academic and non-academic factors that contribute to student success (finances, health, food supply, housing, mental health counseling, academic advising, physical and social activities, etc.). Please explore the many links on this Support at Brandeis page (https://www.brandeis.edu/support/undergraduate-students/browse.html) to find out more about the resources that Brandeis provides to help you and your classmates to achieve success.

Teaching Continuity
If campus closes, we will continue the course via Zoom. If there is a week's break in the class for any reason, I will announce it via Latte so that you will receive an email as you do other course announcements.

Detailed Course Schedule

Wed Aug 26

Lab 1.1: Getting started: Matlab and Data Analysis/Statistics

Matlab/computer Skills:

- Using the command line to assign values to individual variables, or lists of variables stored in matrices, and to manipulate those values.
- Download and find Matlab files that can be run from the command line.

Data Analytic Skills:

- Compare two datasets using a small number of examples, or statistical measures using larger numbers of examples.
- Compare two datasets using graphical measures, such as bar graphs of mean and standard deviation, histograms, and cumulative histograms.

Mon Aug 31

Lab 1. 2: Variables, Functions, and Simple Plots

Matlab/computer Skills:

- Accessing specific entries or subsets of entries in a matrix.
- Reasonable names for variables.

- How to use functions.
- Using the command line to plot data, label axes and graphs etc.
- Words and sentences as character strings.

Data Analytic Skills:

- Mean, median, and percentiles of data sets.

Wed Sep 02

Lab 1. 3: More plotting and (dot) m-files

Matlab/computer Skills:

- Making bar graphs.
- Creating your own functions.
- Periodic arrays and the colon ':' operator.
- Saving files.

Data Analytic Skills:

- Working with histograms and cumulative histograms.
- Choosing bin sizes.

Wed Sep 09

Lab 1.4: M-file Management; Differences in Cumulative Histograms

Matlab/computer Skills:

- Mounting other drives.
- Defining the search path for Matlab to find files.

Data Analytic Skills:

- Comparing two distributions of data via cumulative histograms.
- Finding the maximum difference between two cumulative histograms.

Due:

- Problem Set 1.1.

Thu Sep 10

Lab 1.5: Did my drug really work? Statistical inference

Matlab/computer Skills:

- Loops using the `while` command.

Data Analytic Skills:

- The Kolmogorov-Smirnov test for significant differences between distributions.
- Meaning of the Null Hypothesis.
- Setting a significance level (alpha).

Brandeis UNIVERSITY

Mon Sep 14

Lab 1.6: What was the impact of my drug on average?

Matlab/computer Skills:

- Loading ascii (standard text) files.
- Sampling and random permutations.
- "Recipe" for Kolmogorov-Smirnov test.
- "Recipe" for 2-sample t-test.

Data Analytic Skills:

- The Central Limit Theorem.
- Standard error of the mean.
- Conducting a two-sample t-test.

Wed Sep 16

Lab 1.7: Basic sample descriptors; tinkering with plots

Matlab/computer Skills:

- Altering properties of figures.
- Adjusting tick marks on axes.
- "Objects" and "children" in Matlab.
- Some basic stats functions.
- Producing known distributions with `randn`.

Data Analytic Skills:

- Statistics for the population versus from our samples.
- A bit on "expected values" of sample measurements.
- Biased and unbiased estimators of variance.

Due:

- Problem Set 1.2.

Mon Sep 21

Lab 1.8: How well did my drug do? Variance explained and discriminability

Matlab Skills:

- Extracting relevant indices from an array using `find`.

Data Analytic Skills:

- Within group versus between group variance.
- The Analysis of Variance (ANOVA).
- Discriminability and Receive-Operating Characteristic (ROC).

Wed Sep 23

Lab 2.1: Describing data with fits (models)

Matlab Skills:

- Element-wise operations with the "dot" operator.
- Matlab fitting tools (using functions with options).

Data Analytic Skills:

- Scatterplots.
- Covariance and correlation.
- Linear fits, nonlinear fits, and regression.
- "r"-value versus "r^2".

Wed Sep 30

Lab 2.2: Using fitting as a tool, not like a fool

Data Analytic Skills:

- Local optimization versus global optimization.
- Problems with fitting – when it goes wrong.
- Model selection and overfitting (penalizing extra parameters)

Due:

- Problem Set 1.3.

Mon Oct 05

Lab 2.3: Creating robust index values

Data Analytic Skills:

- Constructing an index or measure with a range of -1 to 1 (or 0 to 1).
- The Kruskal-Wallis test (a non-parametric test of rank order).
- Orientation-selectivity index and adding directional vectors.

Wed Oct 07

Lab 2.4: Dealing with a truckload of data

Matlab Skills:

- String variables.
- Extended directory structure (creating and placing files in folders/subfolders).
- File paths.
- "for" loops.
- "supersubplot" code for a number of subplots in a figure.

Data Analytic Skills:

- Testing whether an index does a good job at representing differences in data or datasets

Mon Oct 12

Lab 3.1: Time-series (correlation, feature detection, rates) Matlab Skills:

Data Analytic Skills:

- Dealing with time series.

Correlations across time.

Matlab Skills:

- Plotting correlograms and autocorrelograms.
- Extracting transitions or threshold-crossings.
- Generating rates.

Due:

- Problem Set 2.1.

Wed Oct 14

Lab 3.2: Frequency and Fourier analysis

Matlab Skills:

- Placing text on graphs.
- Reading sound files into Matlab.

Data Analytic Skills:

- Sines, cosines, and Fourier transforms/Fourier series.
- Extracting frequency components and recreating a signal from them.
- Aliasing (why "extra" frequency components appear).
- Spectrograms for visualization of sound signals.

Mon Oct 19

Lab Lab 3.3: Convolution and filtering

Data Analytic Skills:

- Smoothing noisy signals
- Convolution as a way to carry out filtering.
- Edge effects.
- Filtering over a specific range of frequencies.

Due:

- Team Project 1 videos.

Wed Oct 21

Team Project 1 Q&A

Mon Oct 26

Lab 4.1: Basic image representations

Matlab Skills:

- Reading/loading images.
- Creating images. Color tables in Matlab.
- Index-based images.

Data Analytic Skills:

- RGB format vs CMYK color.
- Intensity (black-and-white) vs color images;

Due:

- Team Project 1 write-up.

Wed Oct 28

Lab 4.2: Digital image representations of the physical world

Matlab Skills:

- Combining colors on displays.
- Altering colors.

Data Analytic Skills:

- Properties of light and reflecting surfaces.
- Continuous versus discrete intensity measures.
- Rescaling intensity values in images.

Mon Nov 02

Lab 4.3: Image analysis I (Regions of Interest, ROI)

Matlab Skills:

- Mouse-click locations as inputs in Matlab.
- Image data as x-y matrix or column vector.
- Using the Matlab image toolbox.

Data Analytic Skills:

- Manually selecting regions of interest.
- Automatic selection of regions of interest.
- Identifying overlaps across images.

Due:

- Problem Set 3.1.

Wed Nov 04

Catch-up day with full assistance available

Mon Nov 09

Lab 4.4: Image analysis II (Functional imaging)

Data Analytic Skills:

- Examining temporal changes in images.
- Using false color.

Due:

- Team Project 2 Videos.

Wed Nov 11

Team Project 2 Q&A

Mon Nov 16

Lab 5.1: High-dimensional space I; data points as vectors

Matlab Skills:

-
- Matrix and vector operations in Matlab.
- 3D plots in Matlab.

Data Analytic Skills:

- Using matrix multiplication to transform data.
- High-dimensional (high-D) data.
- Intro to visualizing higher-D data.

Due:

- Team Presentation 2 write-up.

Wed Nov 18

Lab 5.2: High-dimensional space II (clustering; dimensionality reduction)

Matlab Skills:

- Using the "cell" datatype.;
- Carrying out principle component analysis in Matlab.

Data Analytic Skills:

- More high-dimensional data.
- Cluster analysis of data (K-means).
- Description of principle comoponents analysis (PCA).
- Projecting data to dimensions of interest;

Mon Nov 30

No Class: Work on Team Project 3 in your groups

Wed Dec 02

Lab 5.3: High-dimensional space III (interpolation and PCA)

Data Analytic Skills:

- Realignment of data and interpolation.
- Technical explanation of PCA.
- Overview of eigenvectors and eigenvalues (the essence of linear algebra).

Mon Dec 07

Due:

- Team Project 3 Videos.

Wed Dec 09

Team Project 3 Q&A

Mon Dec 14

Due by 9pm:

- Problem Set 4.1.
- Team Presentation 3 write-up and any remaining material for the course.