BMI206: Statistical Methods for Bioinformatics

Instructor: Katie Pollard
TA: Svetlana Lyalina

http://docpollard.org/bmi206
Fall 2016

Course Objectives

- Goal: understand statistical methods that are used in bioinformatics research

- You will learn both ideas/concepts and how to implement methods in code.

- By the end of the course, you should be able to make good decisions about how to statistically analyze bioinformatics data.

Pre-Requisites

BMI206 is intended for students with a strong undergraduate background in statistics.

1. Students should be comfortable with topics in David Quigley's online biostatistics course:

https://courses.ucsf.edu/course/view.php?id=3223

2. Some fundamental concepts will be reviewed in the first week of class.

3. Students should know how to program in at least one language. We will primarily use R.

Class Schedule

- Monday and Wednesday 9:30-11:00

- Friday 1:10-4:00

All classes are held in Room BH-215

Participation and attendance are critical
(absences should be communicated)

Bring laptops to all classes

Format - weekly modules

- **Monday** (lecture and discussion)

  - 45 minutes: Introduction to a new bioinformatics topic, including biological questions, data types, and methodology

  - 45 minutes: Small group discussion of journal article on the week's topic

- **Wednesday** (lecture): In depth exploration of statistical methods with code examples

- **Friday** (computer lab): Implementation of the week's methods to analyze a data set

Module Topics

- Metagenomics / generalized linear models

- Gene regulation (RNA-seq) / multiple hypothesis testing

- Epigenomics (ChIP-seq) / machine learning

- Population genetics / multivariate statistics (e.g., PCA)

- Biological networks / graphical models

- Chromatin structure / stochastic models (e.g., HMMs)

- Protein structure / bayesian statistics

Communication

- Katie Pollard: kpollard@gladstone.ucsf.edu
- Svetlana Lyalina: svetlana.lyalina@ucsf.edu
  - Office hours by request
  - Laptop setup session Monday 9/26 @ 11am
- Course website:

  http://docpollard.org/bmi206/

Journal Articles

• For each unit, a recent paper will be assigned as background reading and to demonstrate how the concepts from lecture are used in practice.

• Papers will be discussed in groups on Mondays.

• Lectures and labs will often refer to the assigned papers.

• Papers will form the basis for the final project.

• If you want additional reading on any topic, I can provide a longer list of references.

Labs

- Fridays will be computer lab sessions.

• You will need to install various software packages on your laptop. Most labs will be in R. TA can help as needed.

• To start, please install R (r-project.org).

– No formal groups, but OK to collaborate or ask each other questions

– We will discuss our findings after each exercise

– The final portion of each lab will be an open problem to solve on your own. **A one paragraph summary of your solution plus your code will be due the following Wednesday by email to** kpollard@gladstone.ucsf.edu.

Project Objectives

Goals of the Project:

• Learn to critically read bioinformatics papers from a statistical perspective
• Obtain primary data from a publication
• Practice selecting appropriate analysis methods
• Practice making figures and other data displays
• Compare and evaluate different bioinformatics and statistical approaches to
answering a scientific question
• Evaluate sensitivity of results to analysis choices

Project Deadlines

Timeline

- Sep 23: Introduction to project

- Sep 26: Paper preferences due to TA

- Oct 10: Acquire / summarize primary data [report due]

- Oct 17: Preliminary analysis plan [report due]

- Nov 7-9: Oral presentations of reanalyses

- Nov 28 - Dec 2: Oral presentations of novel analyses

- Dec 12: Project write-ups [individual report due]

Grading

- Participation (lecture, discussions, lab) = 25%

- Lab assignments (homework) = 25%

- Project = 50%