EVE 231: Principles of Biological Data Analysis (Fall 2019)

Bruce Rannala

Outline

Course Objectives

Course Logistics

Course Requirements

The "Scientific Method"

- ► Observations/Intuition + Induction = Hypothesis
- ► Hypothesis + Deduction = Predictions
- ► Predictions + Data = Accept/Reject Hypothesis

The Role of Statistics

- ▶ Build a statistical model representing a hypothesis
- ▶ Infer unknown model parameters using data
- ▶ Test alternative hypotheses (models) using data
- ▶ Data exploration (induction)

Topics

- ▶ The unix operating system
- ▶ Data formats, files, incremental backup (git)
- ▶ Editing (emacs) and manipulating data (grep, etc)
- ▶ Basic programming in R, Python and unix shells
- ▶ Measurement errors (propagation in calculations)
- ▶ Probability, random variables, simulation
- ▶ Statistical models and parameters
- ▶ Parameter estimation and hypothesis tests

Office Hours:

Location: Storer Hall Rm 5339

Time: Tues/Thurs 16:30 - 17:30 or by appointment

Grading:

P/NP based on project and presentation

Project will propose an experimental design and methods for statistical analysis of results

Presentations will be 15 minutes (+ 5 mins for questions) describing project proposal

Computer running a unix operating system

Linux

- ▶ Install a debian linux distribution on a PC. I recommend installing Ubuntu. Notes on Canvas for single (linux) or dual-boot (Windows+linux).
- ▶ Install class software by running an apt installer script. Notes on Canvas.

Mac OSX

- ▶ Mac OSX is a unix operating system – access it through the terminal application (notes on Canvas)
- ▶ Install Homebrew package manager and run brew installer script. Notes on Canvas.