Biology 723: Scientific Computing for Biologists, Fall 2014

Instructor: Paul M. Magwene
Office: FFSC 4103
Phone: 613-8159
Email: paul.magwene@duke.edu

TA: Yuantong Ding
Office: TBA
Phone: 919-457-7658
Email: yuantong.ding@duke.edu

Description

The focus of this course is statistical computing for the biological sciences with an emphasis on common multivariate statistical methods and techniques for exploratory data analysis. A major goal of the course is to help graduate students in the biological sciences develop practical insights into methods that they are likely to encounter in their own research, and the potential advantages and pitfalls that come with their use.

Prerequisites

Enrollment is limited to graduate students or undergraduates with permission of instructor. No previous programming experience is required, but familiarity with basic statistical concepts (equivalent of STA 213) is assumed.

Grading

Grading is based on weekly homework assignments. These homework assignments will typically consist of statistical problem solving exercises and/or programming tasks.

Office Hours

Yuantong will hold office hours on Wednesdays from 2-4pm (location TBA).

Course Website

https://github.com/pmagwene/Bio723/wiki

Texts

[1] Wickens, T. D. 1995. The geometry of multivariate statistics. Lawrence Earlbaum Associates, New Jersey.

[2] A. B. Downey. 2014. Think Python: How to think like a computer scientist. Available from Greentea Press under an open source license.

[3] Matloff, N. 2011. The Art of R Programming. No Starch Press, San Francisco.

Other Recommended Texts

[6] Hamilton, A. G. 1989. Linear algebra: An introduction with concurrent examples. Cambridge University Press, Cambridge.

A well organized and readable introduction to linear algebra. This assumes no previous familiarity with lineage algebra. You'll get maximum benefit from this text if you work through the short exercises that accompany each chapter.

[7] Krzanowski, W. J. 2000. Principles of multivariate analysis. Oxford Univ. Press, New York.

I would have made this a required text but it's become unreasonably expensive, even for used copies. Nonetheless, if you plan on having just on book on multivariate statistics on your bookshelf this is the one I'd recommend.

[8] Sokal, R. R. and F. J. Rohlf. 1995. Biometry. W. H. Freeman, New York.

Another good text to have on your bookshelf. A readable and well organized basic statistics book with examples drawn from the biological literature.

Syllabus

Date	*Topic*
August 26	Introduction; Getting Acquainted with Python and Unix
September 2	Data as Vectors; Uni- and bivariate visualizations
September 9	Descriptive statistics as matrix operations; Visualizing multivariate data
September 16	Multiple regression and introduction to biplots; Regression
September 23	Non-linear regression models
September 30	Eigenvectors and Eigenvalues; Principal Components Analysis
October 7	Singular Value Decompisition and Biplots [RESCHEDULE MAKEUP CLASS]
October 14	Fall Break
October 21	Discriminant analysis and Canonical Variate Analysis
October 28	Analyses based on Similarity/Distance I; Hierarchical and K-means clustering
November 4	Analyses based on Similarity/Distance II; Multidimensional scaling
November 11	Randomization and Monte Carlo Methods; Jackknife, Bootstrap, Permutation
November 18	Building Bioinformatics Pipelines I; Pipes, redirection, subprocesses
November 25	Building Bioinformatics Pipelines II; Putting the concepts to work
	Graduate Classes End November 25