

Biology 723: Scientific Computing for Biologists, Fall 2013

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

TA: Colin Maxwell **Office:** FFSC 3135B
Phone: 668-3755 **Email:** colin.maxwell@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.

Course Website

<https://github.com/pmagwene/Bio723>

Texts

- [1] Wickens, T. D. 1995. The geometry of multivariate statistics. Lawrence Earlbaum Associates, New Jersey.
- [2] Matloff, N. 2011. The Art of R Programming. No Starch Press, San Francisco.
- [3] A. B. Downey, J. Elkner and C. Meyers. How to think like a computer scientist: learning with Python. Available in [HTML](#) and [PDF](#) form under an open source license.

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 one 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

<i>Date</i>	<i>Topic</i>
August 27	Introduction; Getting Acquainted with R
September 3	Data as Vectors: Geometry of Correlation and Regression; Visualizing bivariate data in R
September 10	Descriptive statistics as matrix operations; Visualizing and working with multi-variate data in R
September 17	Multiple regression and introduction to biplots; Regression in R
September 24	Non-linear regression models
October 1	Eigenvectors and Eigenvalues; Principal Components Analysis
October 8	Singular Value Decomposition, Biplots, and Correspondence Analysis
October 15	Discriminant analysis and Canonical Variate Analysis
October 22	FALL BREAK
October 29	Analyses based on Similarity/Distance I; Hierarchical and K-means clustering
November 5	Analyses based on Similarity/Distance II; Multidimensional scaling
November 12	Randomization and Monte Carlo Methods; Jackknife, Bootstrap, Permutation
November 19	Building Bioinformatics Pipelines I; Pipes, redirection, subprocesses
November 26	Building Bioinformatics Pipelines II; Putting the concepts to work

GRADUATE CLASSES END NOVEMBER 26