Biomedical Informatics 713 / Genetics 212

Computational Statistics for Biomedical Sciences

Meeting Information

Location: Countway Library 403.
Time: Thursdays, 1-4pm (Lecture 1-3, computer lab 3-4)
Dates: September 9, 16, 23, 30; October 7, 14, 21, 28; November 4, 18; December 2.
(Note that there is no class on September 2)

Instructor

Peter J. Park
HMS Ctr for Biomedical Informatics/CHB/BWH
Email: peter_park@harvard.edu
Countway Library Rm 409, 617-432-7373
Cell: 703-244-8577
Directions: see http://compbio.med.harvard.edu

Teaching Assistants

L. Joe Luquette, email: lovelace_luquette@hms.harvard.edu
Ruibin Xi, email: ruibinxi@gmail.com
HMS Center for Biomedical Informatics
Office: Countway Library 4th floor

Course Description

This course will provide a practical introduction to analysis of biological and biomedical data. Basic statistical techniques will be covered, including descriptive statistics, elements of probability, hypothesis testing, nonparametric methods, correlation analysis, and linear regression. Emphasis will be on how to choose appropriate statistical tests and how to assess statistical significance. To visualize data and carry out statistical testing, students will learn R, a powerful programming language for statistical computing and graphics. No previous knowledge in statistics or programming is required, although those with no programming experience will be expected to devote a significant amount of extra time. The class will be a combination of lecture and computer lab. This course is geared toward graduate students, but postdoctoral fellows and others are welcomed as space allows.

Enrollment

Limited to 26; may be increased if space is available.

References

Statistics, An Introduction using R by Michael Crawley
Introductory Statistics with R by Peter Dalgaard
Principles of Biostatistics by Marcello Pagano and Kimberlee Gauvreau

Course Content (subject to change)

- Lecture 1: Probability and distributions: random variables, binomial and normal distributions
- Lecture 2: Hypothesis testing: sampling distribution, p-values, confidence intervals
- Lecture 3: Comparison of means: t-test, analysis of variance, type I and type II errors, power
- Lecture 4: Non-parametric methods: Wilcoxon test, parametric vs non-parametric tests
- Lecture 5: Analysis of proportions: sampling distribution of proportion, confidence intervals
- Lecture 6: Contingency tables: Chi-square test, Fisher's exact test, odds ratio, multiple 2x2 tables
- Lecture 7: Correlation: Pearson correlation coefficient, Spearman correlation coefficient, p-value
- Lecture 8: Simple linear regression: linear model, inference on regression coefficients, model evaluations
- Lecture 9: Multiple regression: logistic regression, least-squares, indicator variables, model selection.
- Lecture 10: Survival analysis: Kaplan-Meier curve, log-rank test, model selection
- Lecture 11: Cluster analysis (unsupervised learning): distance measure, dendrogram, principal components
- Lecture 12: Classification (supervised learning): predictive models, cross-validation, advanced machine learning methods
- Programming (computer lab sessions): installation and editing, basic syntax, vectors and matrices, functions, loop structure, advanced data structure, graphics