

Modern Statistics for Modern Biology (Bios221/Stats366)
Statistics for Genetics, Immunology, Microbiology and Ecology

Susan Holmes, Statistics, Sequoia 102, susan@stat.stanford.edu
Wolfgang Huber, Sequoia 236, wolfgang.huber@embl.de

Crash course in Modern Statistics : Why?

Standard courses are based on iid Normal data, in fact:

- Much data is discrete: Counts, transitions, states.
- Need for normalization, filtering, corrections.
- Independence is not the norm, (dependent data, spatial data).
- Interface statistical modeling tools with standard databases.
- Estimation for non standard parameters: trees, graphs, images, sequences, ...

Some Aims of the Course

Learn the useful Probabilistic Tools specific to genetic/protein/image data

Discrete random variables.(Binomial, Multinomial, Poisson, Dirichlet)

Continuous Models (Gamma, Point Processes)

Monte Carlo Simulation, Markov Chains, Mixture Models. Expectation, conditional probability, variance.

Learn the statistical tools for analyzing large heterogeneous data sets

- Data Transformations, filtering, smoothing and statistical modeling, goodness of fit tests.
- Visualization and Exploratory Data Analysis, Multivariate analyses (PCA, SVD, CA, DA, Clustering, supervised learning) and interpretation.
- Bootstrap, permutation and omnibus nonparametric tests.
- Multiple Hypothesis Testing, Design of Experiments.

Learn to use R/Bioconductor for statistical analyses of genomic/proteomic/ecological data

- HTS sequence analyses, DESeq2, edgeR (RNA-seq, 16S-rRNA).
- Microbiome data and phylogenetic trees: ape, phyloseq
- Flow cytometry analyses flowviz, Mass cytof, mass spec (xcms).
- Image analyses EBImages and Network analyses igraph.

Course Grade (letter grade) computed from

- | | |
|---|------|
| • Participation in class and scribe once. | 20 % |
| • Finish 6 computer lab/quizzes. | 20 % |
| • Hand in the "3 errors" before July 12th | 10 % |
| • Midterm proposal by July 10th. | 15 % |
| • Final Paper by August 6th. | 35 % |

Course Grade (CR/NC) achieved by

- | | |
|---|------|
| • Participation in class and scribe once. | 40 % |
| • Finish 6 computer labs/quizzes. | 50 % |
| • Hand in the "3 errors" before July 12th | 10 % |

Format This is an intensive course with 24 hours of lectures and 18 hours of computer labs. Attendance to all lectures is required (send an email if you have to miss one), labs can be done in sessions or at home as long as the lab assignments are finished.

Date: June 26th- July, 14th 2017. Lectures MTWThF, 10.30-12.20, (except July 4th) LKC 120.

Labs: 4 afternoons a week: 1.30-2.45pm.

For more information see course website <http://bios221.stanford.edu/>

and canvas: <https://canvas.stanford.edu/courses/68050>