LIFS6000C Biostatistics: theory and application in life science research
Spring 2020

Instructor

Yu Hu MATH and LIFS Email: mahy@ust.hk Office: 3428
Office Hour by appointment.

Meeting Time and Venue

Lectures: Mon & Wed, 10:30am to 11:50am

Venue: CYT G002

Course description

Statistical analysis has become an essential component in life science research for analyzing experimental data, experimental design, and exploratory analysis. This course is tailored to an audience of graduate students and researchers working in biological sciences and more broadly disciplines working with experimental or simulated data. We will introduce some widely used methodologies in contemporary research, such as hypothesis testing, nonparametric approaches, cross-validation and bootstrapping. To ensure the correct use of these statistical tools and to be able to generalize to new scenarios, we will also discuss some of the fundamental theories behind these methods. The students will gain practical experience from homework problems including programing and use of packages.

Prerequisite

Since this is a PG course, we do not require by specific course codes, but multivariate calculus, linear algebra, basic probability and experience in programing (e.g. python, R, MATLAB) are required. Note that we only need the very basic aspects in these topics. To give you some examples (not a complete list): vector, matrix, eigenvalue, derivative and gradient, method of Lagrange multipliers, probability distribution function, independence, covariance, central limit theorem, linear regression. The programing requirement can also be satisfied by doing any coding tutorial available online before the course. In case of doubt on prerequisite, please check with the instructor.

No auditing policy: Please register. Commitment and participations are very important for this course.

Exclusion

None

Weekly outline

Week 1 Introduction, basic concepts
Week 2 Hypothesis testing, t test and p-values
Week 3 Analysis of variance (ANOVA)

Week 4 Two-way ANOVA and multiple testing
Week 5 Experimental design
Week 6 Model assumptions and nonparametric tests
Week 7 Linear regression
Week 8 Conditional dependence and partial correlation
Week 9 Model selection
Week 10 Logistic regression and generalized linear model
Week 11 Bootstrapping and shuffling
Week 12 Linear discriminant analysis and clustering
Week 13 Overview of advanced topics

Intended Learning Outcomes

Upon successful completion of this course, students should be able to:
1. Know how and when to apply the main statistical analysis methods introduced
2. Develop statistical thinking and understand the basic theories of the methods, potential pitfalls and limitations in applications
3. Gain practical skills and experience in applying the methods to experimental data and research

Evaluation

Credit points: 3

Assessment	Course ILOs
Homework 50%	1,2,3
In-class discussion and participation 15%	1,2
Final exam 35%	1,2

Homework includes analytical calculations, derivations, and numerical programming problems
Final exam format: 24 hour take-home exam

References

Lecture notes and slides will be made available after each class.
Optional reference textbook:
An Introduction to Statistical Analysis in Research: with Applications in the Biological and Life Sciences by Weaver et al.
Biostatistical analysis by J. Zar
Using R at the bench : step-by-step data analytics for biologists by Bremer and Doerge
Advanced Statistics in Research by L. Hatcher