LIFS6000C Biostatistics: theory and application in life science research
Spring 2021

Instructor

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

Meeting Time and Venue

Lectures: Mon & Wed, 9am to 10:20am

Venue: (up to Feb 17) Zoom meeting link from Canvas

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
Week 12,13 Additional 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 derivations, proofs, and numerical (programming) problems
Final exam format: 24 hour, open book, take-home exam

We will use Canvas to distribute and collect the homework and exam. We will also send course anoucements via Canvas. So please turn on notification or check regularly.
When you have a question, you are encouraged to post it under Discussion on Canvas, so others can see it and may also help to answer. The instructor will also check and answer the Discussion regularly.

Important: You can and are encouraged to discuss and collaborate with classmates on homework, as long as each person submit his/her own answers (copy-and-paste is not allowed).
For final exam, however, you are not allowed to disucss with anyone.

References

Lecture slides will be made available after the class.
Optional reference textbook (reserved at HKUST-lib):
Biostatistical analysis by J. Zar
An Introduction to Statistical Analysis in Research: with Applications in the Biological and Life Sciences by Weaver et al.