2017 清华大学统计学暑期学校

7 月 3 日-14 日 · 北京

课程安排（以最终发布为准）

7 月 3 日-6 日		
09:00-11:00	刘京辰	Latent Variable Models in Cognitive Assessment
14:00-16:00	马　平	Statistics Methods in Big Data Analytics
7 月 10 日-13 日		
09:00-11:00	刘传海	Computational Statistics
14:00-16:00	陈　伟	Statistical and Computational Methods in the Analysis of Genotype and Sequence Data

课程概要

Latent Variable Models in Cognitive Assessment

Latent variable models are prevalent in many studies. We consider the context of cognitive assessment that has applications in many disciplines including education, psychology /psychiatry, political sciences, marketing, etc. For instance, in educational measurement, students' solutions to test problems are observed to measure their skill levels; in psychiatric assessment, patients' responses to diagnostic questions are observed to assess the presence or absence of mental health disorders; in political sciences, politicians' voting behavior reflects their political views; in marketing analysis, consumers' purchase history reflects their preferences. A common feature in these studies is that the observed human behaviors are driven by their latent attributes that are often unobservable. Latent variable models can be employed in these contexts to describe the relationship between the observed behavior, which is often in the form of responses to items, and the underlying attributes. I will talk about several linear and nonlinear latent variable models, their inference, computational

tools, and the associated software packages.

Statistics Methods in Big Data Analytics

The rapid advance in science and technology in the past decade brings an extraordinary amount of data that were inaccessible just a decade ago, offering researchers an unprecedented opportunity to tackle much larger and more complex research challenges. The opportunity, however, has not yet been fully utilized, because effective and efficient statistical and computing tools for analyzing super-large dataset are still lacking. One major challenge is that the advance of computing technologies still lags far behind the exponential growth of the database. 'One option is to invent algorithms that make better use of a fixed amount of computing power'. In this course, I present some statistical methods developed for achieving such a goal.

Computational Statistics

This course focuses on basics on three topics: Computational Statistics, Big Data Computing with Spark (and SupR), and Deep Learning. Computational statistics covers popular iterative statistical algorithms such as EM and Markov chain Monte Carlo. For computing with big data, we introduce distributed and parallel computing with SupR/Spark. The discussion on Deep Learning focuses on statistical consideration and development of statistical methods toward automated data analysis.

Statistical and Computational Methods in the Analysis of Genotype and Sequence Data

This short course will cover a series of special topics on statistical methods for the analysis of genotype and next-generation sequencing data. The topics will include but are not limited to: genotype calling and haplotype inference in population and family-based sequencing studies; genome-wide association analysis of quantitative and time-to-event traits; rare variant association tests for sequencing data; introduction to the analysis of single cell sequencing data. The emphasis of this course is to introduce fundamental concepts and methods on each topic and to illustrate how these methods facilitate biomedical research.