清华大学大数据方向硕士学位

公共必修课课程

数据分析学（I）

Data Analytics (I)

开课单位：数据分析学（I）课程组

授课教师：黎波、张楠、郑路、庞珣、苏毓淞、罗昊、王程韡 （暂定）

教学目的：
本课是针对社会科学和管理类研究学开设的数据分析基本课程。通过本课的学习，学生将对（大）数据分析的价值、意义和基本原理建立清晰和比较全面的认识，掌握有关数据发掘、处理、建模和解释的基本原理和方法，了解和熟悉数据分析在社会科学研究、商业分析和公共管理等领域的实际案例。

Data Analytics (I) is a graduate level course mainly designed for students with social sciences and management background. The objective of the course to give students a broad overview of the basic principles and applications of data analytics. Students will also be familiar with the various aspects of data analytics such as exploring, managing, modeling and interpreting data. Students' learning will also be enhanced by their exposure to real life applications of data analytics in social science research, business analysis and public management.

主要教材：
<Data Mining and Business Analytics with R> by Johannes Ledolter, 2013, Wiley
<An Introduction to Statistical Learning: with Applications in R> by Gareth James et al, 2013, Springer
<Analyzing Social Networks> by Stephen Borgatti et al, 2013, SAGE
<Multilevel and Longitudinal Modeling using Stata> by Sophia Rabe-Hesketh and Anders Skrondal, 2008, Stata

教学软件：R, Stata, UCINET

教学内容：

一、统计分析
1. 数据分析简介
2. 概率论基础
3. 数理统计基础
4. R 软件简介，使用 R 进行探索性数据分析

5. 线性回归模型
6. 多层次、纵贯性数据分析 (Multilevel and Longitudinal Modeling)
7. 非参数回归

二、机器学习
8. 正则化监督学习（Supervised learning with regularization）
9. 在抽样统计学(Resampling methods)
10. 树状模型方法、支持向量机（Tree-based methods, Support vector machines）
11. 非监督学习：聚类、降维（Unsupervised learning: clustering, dimension reduction）

三、综合应用
12. 文本挖掘和情感分析（Text Mining and Sentiment Analysis）
13. 社会网络分析（Social Network Analysis）
14. 政策信息学简介（Policy Informatics）

成绩构成：
平时作业   20%
期中考试   20%
期末考试   30%
学期论文   30%