

# 探索性数据分析与非参数统计平滑方法—课程 简介

张伟平

2/14/2012

课程编号: G025200307

英文名称: Exploratory Data Analysis and Nonparametric Smoothing

学分: 2

讲授学时: 40

周学时: 3

**课程综述** 统计概念和准则已经渗透到科学的各个方面: 设计试验, 分析数据, 检验假设, 报告结果以及解释研究资料. 现代科技的飞速发展使得数据的规模越来越大, 具有一定的统计能力就显得越来越重要. 如今的统计数据分析师越来越多的面临如何收集,分析和解释复杂的大规模数据.

本课程将介绍在使用统计工具分析小规模数据集到大规模数据集时的核心概念和方法. 我们将首先介绍数据是什么? 怎么描述它以及应用中的一些表示数据的可视化方法. 重心将放在可视化方法上, 内容包括经典的描述性统计方法, 各种图示方法,非参数平滑方法等.

## 学习目标和目的

本课程的基本目标是培养学生掌握基本的探索性数据分析方法, 特别是有关于复杂数据集的分析方法, 为学生的未来工作打下坚实的基础.

特定的目标包括:

- 培养统计思维
- 描述总结数据
- 可视化表达数据
- 针对特定的问题能够选择合适的统计方法
- 对实际数据集能够使用统计软件实施一系列分析工作

## 教材和参考书

本课程没有明确的教材,但是可以参考如下一些著作. 每次课程的讲义和相关资料将提供在课程主页上.

1. David C. Hoaglin, Frederick Mosteller and John W. Tukey, Understanding Robust and Exploratory Data Analysis, John Wiley & Sons, 1982. 探索性数据分析, 中国统计出版社.

2. Wendy L. Martinez, Angel R. Martinez, and Jeffrey L. Solka, Exploratory Data Analysis with MATLAB, Chapman& Hall, 2011.

3. Antony Unwin, Martin Theus, and Heike Hoffman, Graphics of Large Datasets: Visualizing a Million, Springer-Verlag, New York, 2006.

3.<http://www.statsoft.com/textbook/stathome.html>

4. <http://www.stat.berkeley.edu/stark/SticiGui/Text/toc.htm>

5. <http://www4.stat.ncsu.edu/hzhang/st790A.html>

## 成绩评定

本课程将采取授课, project报告和讨论方式进行, 预计全部课程将有3-5个project报告, 最终的成绩将以此评定.