课程详述

COURSE SPECIFICATION

以下课程信息可能根据实际授课需要或在课程检讨之后产生变动。如对课程有任何疑问，请联系授课教师。

The course information as follows may be subject to change, either during the session because of unforeseen circumstances, or following review of the course at the end of the session. Queries about the course should be directed to the course instructor.

1.	课程名称 Course Title	统计计算与软件 **Statistical Computation and Software**
2.	授课院系 **Originating Department**	数学系 Department of Mathematics
3.	课程编号 **Course Code**	MA308
4.	课程学分 **Credit Value**	3
5.	课程类别 **Course Type**	专业选修课 Major Elective Courses
6.	授课学期 **Semester**	秋季 Fall
7.	授课语言 **Teaching Language**	中英双语 English & Chinese
8.	授课教师、所属学系、联系方式（如属团队授课，请列明其他授课教师） **Instructor(s), Affiliation& Contact （For team teaching, please list all instructors）**	李曾　Zeng Li 数学系 Department of Mathematics lizeng124@gmail.com
9.	实验员/助教、所属学系、联系方式 **Tutor/TA(s), Contact**	待公布 To be announced （请保留相应选项 Please only keep the relevant information）
10.	选课人数限额(可不填) **Maximum Enrolment （Optional）**	

11.	授课方式 **Delivery Method**	讲授 Lectures	习题/辅导/讨论 Tutorials	实验/实习 Lab/Practical	其它(请具体注明) Other（Please specify）	总学时 Total
	学时数 **Credit Hours**	48				48

12.	先修课程、其它学习要求 **Pre-requisites or Other Academic Requirements**	概率论与数理统计 （MA212）或者 数理统计（MA204） Probability and Statistics (MA212) or Mathematical Statistics (MA204)
13.	后续课程、其它学习规划 **Courses for which this course is a pre-requisite**	
14.	其它要求修读本课程的学系 **Cross-listing Dept.**	

教学大纲及教学日历 SYLLABUS

15. **教学目标 Course Objectives**

本课程通过实际案例引导学生利用统计软件 R 来解决实际问题，借助数据模拟使学生加深对常用分布函数的理解，达到让学生学会利用统计方法，结合统计软件，解决实际问题的目的。

This course aims to guide undergraduate students to carry out statistical analysis in real data problems by utilizing the statistical software R. It helps students deepen their understanding of commonly used distribution functions and reach the goal of solving practical problems using statistical methods and software.

16. **预达学习成果 Learning Outcomes**

学生通过该课程的学习应该学会以下技能：

1. 使用 R 计算常见分布的经验分布函数，分布函数以及生成服从相应分布的随机数；

2. 能够通过 R 的数据模拟实现矩估计和中心极限定理， 借此加深对中心极限定理的理解；

3. 学会用 R 实现对样本均值、方差和分布的假设检验，并用 R 做回归分析，方差分析；

4. 学会用 R 做聚类分析，判别分析，主成分分析和因子分析。

Upon successful completion of the course, students should be able to:
1. use R to calculate cumulative distribution functions and quantiles for the commonly used distributions and generate corresponding random variables;
2. realize some stochastic simulation experiments in probability including central limit theorem;
3. understand testing hypotheses, regression analysis, ANOVA by the software R;
4. apply R to cluster analysis, discriminant analysis, principal components analysis and factor analysis.

17. **课程内容及教学日历 （如授课语言以英文为主，则课程内容介绍可以用英文；如团队教学或模块教学，教学日历须注明主讲人）**
Course Contents (in Parts/Chapters/Sections/Weeks. Please notify name of instructor for course section(s), if this is a team teaching or module course.)

1. R 基础 (6 hours)
   1.1. 基本数据结构与常见函数
   1.2. 数据导入和导出
   1.3. 常见数据清理和预处理方法
   1.4. 描述性统计量及图表的 R 实现

2. 常用分布函数、分位数算法与 R 实现 (3 hours)
   2.1. 标准正态分布的分布函数和分位数的计算
   2.2. Beta 分布、T 分布、F 分布、二项分布的分布函数
   2.3. 卡方分布，泊松分布的函数和分位数的计算

3. 常用随机变量产生算法与 R 实现 (4 hours)
   3.1. 连续随机变量的分布及相应随机数生成，包括均匀分布，正态分布，指数分布，卡方分布，t 分布，柯西分布
   3.2. Weibull 分布的直接抽样法
   3.3. 对数正态分布的变换抽样法
   3.4. 离散随机变量分布及相应随机数生成，包括二项分布，泊松分布，几何分布，负二项分布

4. 概率中的随机模拟实现 (3 hours)
   4.1. 投骰子问题理论计算与模拟实验

4.2. 参数矩估计的理论计算与模拟实验
4.3. 中心极限定理实验

5. 假设检验 (4 hours)
    5.1. 正态总体均值和方差的假设检验
    5.2. 两组独立样本 Wilcoxon 秩和检验
    5.3. 分布的假设检验

6. 回归分析 (6 hours)
    6.1. 简单线性回归模型和相关统计推断及 R 实现
    6.2. 多元线性回归模型和相关统计推断及 R 实现
    6.3. 线性回归模型诊断及 R 实现
    6.4. 广义线性回归模型和相关统计推断及 R 实现

7. 非参数估计方法及 R 实现 (4 hours)
    7.1. 非参数密度函数估计方法及 R 实现
    7.2. 非参数回归分析及 R 实现

8. 方差分析 (4 hours)
    8.1. 单因素方差分析及 R 实现
    8.2. 两因素方差分析及 R 实现

9. 聚类分析 (4 hours)
    9.1. 层次聚类分析及 R 实现
    9.2. 快速聚类分析及 R 实现

10. 判别分析 (4 hours)
    10.1. 线性判别分析及 R 实现
    10.2. 二次判别分析及 R 实现

11. 主成分分析理论及 R 实现 (3 hours)

12. 因子分析理论及 R 实现 (3 hours)

1. R basics (6 hours)
    1.1. Data structure and basic R functions
    1.2. Data import and export
    1.3. Data clean and pre-processing
    1.4. Descriptive statistics and exploratory data analysis in R

2. Common distribution functions and R realization (3 hours)
    2.1. Standard normal distribution function and quantile
    2.2. Beta, T, F, Binomial distribution function and quantile
    2.3. Chi square, Poisson distribution functions and quantile

3. Generation of common random variables with R (4 hours)
    3.1. Continuous random variables and generation with R, including uniform, exponential, chi-square, t and Cauchy distribution
    3.2. Random sampling of Weibull distribution
    3.3. Random sampling of lognormal distribution
    3.4. Discrete random variables, distribution and sampling with R, including Binomial, Poisson, Geometry, Negative Binomial distribution.

4. Simulation experiments in probability (3 hours)
    4.1. Problems in rolling dices and simulation with R
    4.2. Moment estimation and simulation with R
    4.3. Central Limit Theorem and simulation with R

5. Hypothesis Testing (4 hours)
    5.1. Testing mean and variance of Gaussian sample with R
    5.2. Two sample Wilcoxon rank test with R

   5.3. Test of distribution with R

6. Regression Analysis (6 hours)
   6.1. Simple linear regression and inference with R
   6.2. Multiple linear regression and inference with R
   6.3. Model diagnostics with R
   6.4. Generalized linear regression and inference with R

7. Nonparametric methods with R (4 hours)
   7.1. Nonparametric density estimation with R
   7.2. Nonparametric regression models and realization with R

8. Analysis of Variance (4 hours)
   8.1. One-way ANOVA and MANOVA with R
   8.2. Two-way ANOVA and MANOVA with R

9. Clustering (4 hours)
   9.1. Hierarchical clustering methods with R
   9.2. Fast clustering methods with R

10. Discriminant analysis (4 hours)
    10.1. LDA with R
    10.2. QDA with R

11. Principle Component Analysis with R (3 hours)

12. Factor Analysis with R (3 hours)

教材及其它参考资料 Textbook and Supplementary Readings

1. Hothorn, Torsten, and Brian S. Everitt. *A handbook of statistical analyses using R*. Chapman and Hall/CRC, 2014.
2. Kuhn, Max, and Kjell Johnson. *Applied predictive modeling*. Vol. 26. New York: Springer, 2013.
3. Prabhanjan N. Tattar, Suresh Ramaiah and B.G. Manjunath. *A course in statistics with R.* John Wiley & Sons, 2016.
4. 高惠璇(1995). 统计计算. 北京大学出版社.

课程评估 ASSESSMENT

19.	评估形式 Type of Assessment	评估时间 Time	占考试总成绩百分比 % of final score	违纪处罚 Penalty	备注 Notes
	出勤 Attendance				
	课堂表现 Class Performance				
	小测验 Quiz				
	课程项目 Projects		20		
	平时作业 Assignments		35		
	期中考试 Mid-Term Test		20		
	期末考试 Final Exam				
	期末报告 Final Presentation		25		

其它（可根据需要改写以上评估方式） **Others (The above may be modified as necessary)**		5		

20. 记分方式 **GRADING SYSTEM**

☑ A. 十三级等级制 **Letter Grading**
☐ B. 二级记分制（通过/不通过） **Pass/Fail Grading**

课程审批 REVIEW AND APPROVAL

21. 本课程设置已经过以下责任人/委员会审议通过
This Course has been approved by the following person or committee of authority