

## 《统计计算与软件包》 2022 年春

统计计算是现代统计的重要组成部分。**统计计算是统计方法和实际计算的结合，可宽泛地描述为用计算、图形和数值方法解决统计问题。**主要有以下两个方面的内容：

- 把统计方法变成可靠、高效的算法，并编程实现。这是经典的统计计算要解决的问题，比如计算分布函数值、分位数函数值、计算线性回归中参数估计及检验、求解最大似然估计等。
- 借助于现代计算机的强大处理能力，发展新的统计方法。这是计算技术对统计学的贡献，比如利用随机模拟方法求解 Bayes 模型、Bootstrap 置信区间等。有时这个方面被称为“计算统计”(Computational Statistics)。也包括机器学习、数据挖掘等。

在具体问题中常常没有现成的方法可以套用。即使有现成的统计软件可用，我们也需要理解这些软件工作原理以避免错误使用；在遇到新问题时，需要能够修改原有代码或编写新代码，把计算工具结合在一起解决实际的数据分析问题。**要求掌握统计计算方法、理解统计计算思想。**

**R 环境是基于 S 的用来数据分析和可视化的一套软件和编程语言。**最初由新西兰 Auckland（奥克兰）大学的 Ross Ihaka 和 Robert Gentleman 及其他志愿人员开发，现在由 R 核心开发小组维护，全世界的用户贡献了数千个软件包，功能涵盖了绝大部分的经典和现代统计方法，是世界上许多顶尖的统计学家进行统计研究和发表算法的工具。

**将讲授 R 的基本使用，在算法示例和习题中使用 R 作为编程语言，并在讲到具体统计计算方法时提及 R 有关的函数。**

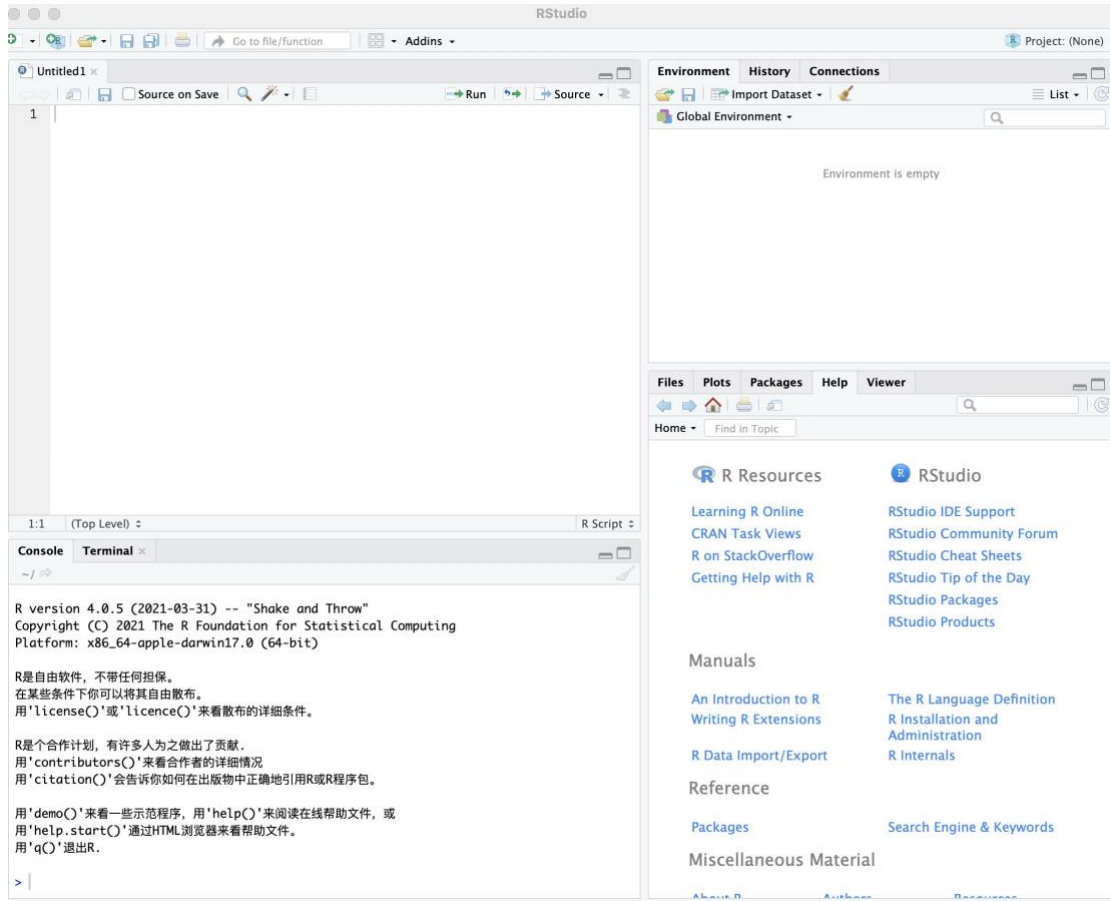
- ✧ **本课程将主要使用 R 语言来介绍现代统计计算的内容，主要包括：**R 语言简介、随机数的生成、Monte Carlo 积分、统计推断（估计、假设检验）中的 Monte Carlo 方法、Bootstrap（自助法）和 Jackknife（刀切法）、R 中的数值方法、MCMC 方法等
- ✧ **要求：**高等数学、线性代数、概率论、数理统计、（随机过程、回归分析、多元统计分析）等
- ✧ **教材：**《统计计算使用 R -Statistical Computing with R》[美]Maria L. Rizzo 著， 胡锐 李义译， 机械工业出版社， 2016.
- ✧ **参考书：**
  1. Maria L. Rizzo, Statistical Computing with R, Chapman & Hall, 2008.
  2. Geof H. Givens and Jennifer A. Hoeting, Computational Statistics (Second Edition), Wiley, 2013.
  3. 《计算统计 Computational Statistics (Second Edition) 》Geof H. Givens and Jennifer A. Hoeting 著， 周丙常 孙浩 译 西安交通大学出版社， 2017.
  4. 《统计计算》李东风， 高等教育出版社， 2017.
  5. 《R 语言实战 R in Action-Data Analysis and Graphics with R》[美] Robert I. Kabacoff 著 高涛 肖楠 陈钢 译 人民邮电出版社 2013.
  6. 《R for Beginners》、《R 导论》、《R 数据的导入与导出》
- ✧ **上课时间、地点：**周一 6-8 节， 14：05-16：30， 数院机房
- ✧ **最终成绩：**20%平时作业+10%考勤+20%中期报告+50%期末考试
- ✧ **考试时间：**随堂 2022 年 5 月 30 日
- ✧ **学习群：**QQ 群：2022 统计计算 153477128

◇ 详细内容：

1. **R 语言基础：**《R 语言实战》第 1-5 章
2. **随机数生成方法：**逆变换法、接受拒绝法、其他变换法、求和与混合变换、多元分布(多元正态、多元正态的混合分布)、随机过程(齐次 Poisson 过程、非齐次 Poisson 过程、更新过程)
3. **蒙特卡洛积分和方差缩减：**蒙特卡洛积分法(简单的蒙特卡洛估计量、方差和效率)、方差缩减、对偶变量法、控制变量法、重要抽样法、分层抽样法、分层重要抽样法
4. **统计推断中的蒙特卡洛方法：**估计中的蒙特卡洛方法(蒙特卡洛估计和标准误差均方误差估计、估计置信水平)、假设检验中的蒙特卡洛方法(经验第一类错误率、检验功效、功效比较)
5. **Bootstrap (自助法) 和 Jackknife (刀切法)：**Bootstrap (自助法)(标准误差的自助法估计、偏差的自助法估计)、Jackknife (刀切法)(偏差、标准误差的刀切法估计、刀切法失效的情况)、基于自助法的刀切法、自助法置信区间(标准正态自助法置信区间、基本自助法置信区间、百分位数自助法置信区间、自助法 t 区间、更好的自助法置信区间)、应用：交叉验证
6. **R 中的数值方法：**一维中的求根法(二分法、Brent 法、Newton 法、Fisher Scoring)、多元优化(Newton 法和 Fisher Scoring)、数值积分、极大似然问题、优化问题(一维最优化、二维最优化)、EM 算法、线性规划
7. **概率密度估计：**一元密度估计(直方图、频率多边形密度估计、平均移位直方图)、核密度估计、二元和多元密度估计(二元频率多边形、二元平均移位直方图、多维核方法)、密度估计的其他方法
8. **马尔科夫链蒙特卡洛方法：**引言(贝叶斯推断中的积分问题、马尔科夫蒙特卡洛积分法)、Metropolis-Hastings 算法(M-H 样本生成器、M 样本生成器、随机游动 M 法、独立性样本生成器)、Gibbs 样本生成器、收敛性监测(Gelman-Rubin 方法)、

⑩ R 的安装: <http://www.r-project.org> 最新版本 4.1.2

⑩ RStudio 是一款免费开源的 R 集成开发环境, 提供了各种增强功能, 比如语言高亮、自动缩进、自动补全、同时管理多个工作目录等功能。下载地址: <http://www.rstudio.com>  
安装



主菜单、工具条、

窗口: 左下角 **Console** 窗口: 执行代码, 执行结果显示

左上角 **Script** 窗口: 撰写代码, 第一次启动看不到这个区, 点击左上角[File]—[New File]—[RScript]打开

右上角: **Environment**: 用来记录当前变量的数值

**History**: 所有 Console 区内代码执行的历史记录

**Connections**: 用来方便连接外部数据库。

右下角: **Files**: 显示当前工作目录下的文件, 让使用者了解所在的工作目录

**Plots**: 画好的图将在这里显示

**Packages**: 显示已经安装好了的包, 打钩代表已经加载

**Help**: 查询函数如何使用