

复杂数据统计方法



[复杂数据统计方法 下载链接1](#)

著者:吴喜之

出版者:中国人民大学出版社

出版时间:2012-10

装帧:

isbn:9787300163994

《复杂数据统计方法——基于r的应用》用自由的R软件分析30多个可以从国外网站下载的真实数据，包括横截面数据、纵向数据和时间序列数据，通过这些数据介绍了几乎

所有经典方法及最新的机器学习方法。

《复杂数据统计方法——基于r的应用》特点：(1)以数据为导向；(2)介绍最新的方法(附有传统方法回顾)；(3)提供r软件入门及全部例子计算的r代码及数据的网址；(4)各章独立。

《复杂数据统计方法——基于r的应用》的读者对象包括统计学、应用统计学、经济学、数学、应用数学、精算、环境、计量经济学、生物医学等专业的本科、硕士及博士生，各领域的教师 and 实际工作者。

作者介绍:

吴喜之，北京大学数学力学系本科，美国北卡罗来纳大学统计博士。中国人民大学统计学院教授，博士生导师。曾在美国加利福尼亚大学、美国北卡罗来纳大学、南开大学、中国人民大学、北京大学等多所著名学府执教。

目录: 第1章引言

1.1作为科学的统计

1.2数据分析的实践

1.3数据的形式以及可能用到的模型

1.3.1横截面数据：因变量为实轴上的数量变量

1.3.2横截面数据：因变量为分类(定性)变量或者频数

1.3.3纵向数据，多水平数据，面板数据，重复观测数据

1.3.4多元数据各变量之间的关系：多元分析

1.3.5路径模型／结构方程模型

1.3.6多元时间序列数据

1.4 r软件入门

1.4.1简介

1.4.2动手

第2章横截面数据：因变量为实数轴上的数量变量

2.1简单回归回顾

2.2简单线性模型不易处理的横截面数据

2.2.1标准线性回归中的指数变换

2.2.2生存分析数据的cox回归模型

2.2.3数据出现多重共线性情况：岭回归，lasso回归，适应性lasso回归，偏最小二乘回归

2.2.4无法做任何假定的数据：机器学习回归方法

2.2.5决策树回归(回归树)

2.2.6boosting回归

2.2.7bagging回归

2.2.8随机森林回归

2.2.9人工神经网络回归

2.2.10支持向量机回归

2.2.11几种回归方法五折交叉验证结果

2.2.12方法的稳定性及过拟合

第3章横截面数据：因变量为分类变量及因变量为频数(计数)变量的情况

3.1经典logistic回归，probit回归和仅适用于数量自变量的判别分析回顾

3.1.1logistic回归和probit回归

3.1.2经典判别分析

3.2因变量为分类变量，自变量含有分类变量：机器学习分类方法

3.2.1决策树分类(分类树)

3.2.2adaboost分类

- 3.2.3 bagging 分类
- 3.2.4 随机森林 分类
- 3.2.5 支持向量机 分类
- 3.2.6 最近邻方法 分类
- 3.2.7 分类方法五折交叉验证结果
- 3.3 因变量为频数(计数)的情况
 - 3.3.1 经典的poisson对数线性模型回顾
 - 3.3.2 使用poisson对数线性模型时的散布问题
 - 3.3.3 零膨胀计数数据的poisson回归
 - 3.3.4 使用机器学习的算法模型拟合计数数据
 - 3.3.5 多项logit模型及多项分布对数线性模型回顾
- 第4章 纵向数据(多水平数据, 面板数据)
 - 4.1 纵向数据: 线性随机效应混合模型
 - 4.2 纵向数据: 广义线性随机效应混合模型
 - 4.3 纵向数据: 决策树及随机效应模型
 - 4.4 纵向数据: 纵向生存数据
 - 4.4.1 cox随机效应混合模型
 - 4.4.2 分步联合建模
 - 4.5 计量经济学家的视角: 面板数据
- 第5章 多元分析(不区分因变量及自变量)
 - 5.1 实数轴上的数据: 经典多元分析内容回顾
 - 5.1.1 主成分分析及因子分析
 - 5.1.2 分层聚类及k均值聚类
 - 5.1.3 典型相关分析
 - 5.1.4 对应分析
 - 5.2 非经典多元数据分析: 可视化
 - 5.2.1 主成分分析
 - 5.2.2 对应分析
 - 5.2.3 多重对应分析
 - 5.2.4 多重因子分析
 - 5.2.5 分层多重因子分析
 - 5.2.6 基于主成分分析的聚类
 - 5.3 多元数据的关联规则分析
- 第6章 路径建模(结构方程建模)数据的pls分析
 - 6.1 路径模型概述
 - 6.1.1 路径模型
 - 6.1.2 路径模型的两种主要方法
 - 6.2 pls方法: 顾客满意度的例子
 - 6.3 协方差方法简介
 - 6.4 结构方程模型的一些问题
- 第7章 多元时间序列数据
 - 7.1 时间序列的基本概念及单变量时间序列方法回顾
 - 7.1.1 时间序列的一些定义和基本概念
 - 7.1.2 常用的一元时间序列方法
 - 7.2 单位根及协整检验
 - 7.2.1 概述
 - 7.2.2 单位根检验
 - 7.2.3 协整检验
 - 7.3 varx模型与状态空间模型
 - 7.3.1 varx模型拟合
 - 7.3.2 状态空间模型拟合
 - 7.3.3 模型的比较和预测
- 附录 练习: 熟练使用r软件
- 参考文献

• • • • • ([收起](#))

[复杂数据统计方法_下载链接1](#)

标签

统计

R

数据分析

统计学

数据挖掘

Statistics

数学

吴喜之

评论

简单撸了一遍，落后你们三年。。。

虽然书不厚但内容丰富、思路清晰。但绝对不适合初学者入门，有一定难度。

好书，但是这本书不是侧重于数学部分而是侧重于数据实战的部分。数学很简单，可以参照《统计学习》这本书对照着学习。

:C819/6843

给统计学专业学生读的一本书 强调了编程在于统计学习中的重要性
可贵的是结合了很多实际例子并结合例子来讲述统计学算法

统计学家写的统计书对于心理学读者来说完全就是灾难啊！

亲爱的的年糕赠书~

真清楚救命啊吴老师是男神

虽然门槛比较高，但是绝对讲解清晰，语言简洁。遇到统计分析基础不扎实的地方，通过书里的描述也可以自己查找根源。是很好的书。

不仅是R的应用，也讲了很多统计的思想方法，适合有一定统计基础的读者。

数学概念和R软件的应用结合起来建模解决实际问题，抽象概念好理解很多。数据分析领域很多优秀教材都是近五年出版的，感慨生正逢时何其幸。薄薄一本册子内容丰富，概念和案例极其紧凑，把复杂的概念用简洁的方式呈现并讲清楚，时年72岁依旧不断学习在一线岗位任教写书，值得敬重。

补条目

【4.5】我觉得第一次感觉到了深入浅出的讲解。

囫圇吞枣的读了1遍多 感觉书的结构和角度特别好
但问题是很难在很少使用数学公式的情况下把模型说清楚 如果没说清楚模型
即便有R的代码和例子 用起来也还是不踏实+有困惑
这书应该在看完一本简明的理论书之后再看 能够辅助相关方向的R操作
直接看难免不得要领

不容易上手，很多没读懂。

!!!

需要理论基础先行的实战类好书

绝对推荐的好书 吴喜之出的几本书都特别赞啊~

很不错的一本书，介绍的算法和模型很全面，都是以实际案例入手。很有借鉴意义。

里面有很多复杂数据的方法，而且对于计算机学习和许多boosted方法都有介绍，毕竟是一本12年的书，很好很好

[复杂数据统计方法_下载链接1](#)

书评

这是一本200+页薄书，但是介绍的内容却很多，很多书上一句话的东西，可能别的书

上要用几页介绍。所以读该书需要参考其他书，遇到不懂的就去查。
顾名思义，该书介绍的复杂数据统计方法，作者当然假设你会“简单”数据的处理。最好会一些R语言的知识。会读的稍微轻松些。 作者...

不适合非统计学专业人士。
例子举了很多，也很好，但分析太少了，全是数学模型公式。连要用分析达到什么目的都没有讲清楚
作者的意图非常好，看到前言部分令人热血沸腾，但事与愿违，我未能从中有收获。

[复杂数据统计方法_下载链接1](#)