

《复杂数据统计方法》

图书基本信息

书名：《复杂数据统计方法》

13位ISBN编号：9787300163994

10位ISBN编号：7300163998

出版时间：2012-10

出版社：中国人民大学出版社

作者：吴喜之

页数：228

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《复杂数据统计方法》

内容概要

《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。

《复杂数据统计方法》

作者简介

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。

书籍目录

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)
 - 2.2.6 boosting回归
 - 2.2.7 bagging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
 - 3.1.1 logistic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)
 - 3.2.2 adaboost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
 - 4.4.1cox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
 - 7.3varx模型与状态空间模型
 - 7.3.1varx模型拟合
 - 7.3.2状态空间模型拟合
 - 7.3.3模型比较和预测
- 附录练习：熟练使用r软件
- 参考文献

《复杂数据统计方法》

编辑推荐

《复杂数据统计方法——基于R的应用》由吴喜之编著，本书首先通过一些简单的统计和数学内容介绍R软件的基本知识，然后介绍数据分析的一些基本逻辑和常识。本书的主体则是根据不同数据形式介绍相应的方法。本书以数据为主导，各章都是完全独立的。有一些统计基本知识的读者可以选读本书的任何一个完整的部分。虽然本书介绍的方法涉及应用统计的各个方面，但不可能介绍所有的数学和统计细节，否则将会是一部巨型的百科全书。笔者尽量用文字和少量数学公式对各种方法的原理予以直观介绍，并且引导读者做进一步的阅读。

《复杂数据统计方法》

精彩短评

- 1、买回来有一段时间，最初只是拿部分内容作为参考，最近细读了本书，深深感受到作者扎实的统计学功底，里面还介绍到一些机器学习的方法，可以全面的了解如何用R来处理各种统计问题，物超所值。
- 2、总体评价如标题，感觉很不错。缺点是纸质质量较差
- 3、需要理论基础先行的实战类好书
- 4、我看的是新版~
- 5、吴喜之老师的书信得过
- 6、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。
- 7、吴老师的书没得说对从事复杂数据的童鞋很有帮助 值得读
- 8、书的内容不错，但纸张很薄
- 9、太棒！统计学习必备！
- 10、不容易上手，很多没读懂。
- 11、学习R软件的很好的书~
- 12、买了看了一点感觉还行，还没有深入看，不好评论
- 13、此书绝不适用于统计和R软件初学者，有点小难，但书本身很好
- 14、被恶劣的排版和低劣的纸张毁掉的一本好书
- 15、吴老师的书，既有洞见又有理论，挺不错的：)
- 16、难得的教学思想，这是一本所有中国的教材都应当学习的样本。只有这样编写有实践价值的教材作为补充，大学教育才不会沦为有名无实混日子的地方，中国经济的转型才有坚实的基础。没有真才实干的人才，经济转型发展就是本末倒置，是一句空话。
- 17、:C819/6843
- 18、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。
- 19、大帅的书
- 20、不适合初学者，想知道原理、为什么、怎么来的以及各种判断规则还是先看别的书。如果要编程做project，这本书极有用。
- 21、复杂数据
- 22、内容很多，但是不够下详细，幸好我都学过
- 23、为什么书中好多的东西操作不了呢
- 24、有些代码有错误，希望吴老师能更正一下啊
- 25、刚刚200页。刚拿到书，只粗粗地翻了一下。内容主要是讲R的各个包在使用，如果你对具体的统计方法都不了解，看这个也许意义不大。如果是不会用R来解决实际问题，则用处应该很大。个人感觉对于R的菜鸟来说，本书后面的10页纸的附录简直太好了，这比啃厚厚的R程序设计书要舒服多了。
- 26、或许是中文书介绍机器学习和统计方法的使用方面那些书中最实用的，简洁且直接了当，重要的统计和机器学习基本思想叙述的明了。写作风格很好（不喜欢有些“书”没实质内容的捣浆糊）；强调了计算和计算技术的作用，对R的许多前沿技术的软件包都有举例。可能是篇幅限制，原理部分的理论分量偏少了一些。
- 27、挺经典的工具书，值得拥有。
- 28、亲爱的的年糕赠书~
- 29、吴喜之的书，精辟
- 30、书写得一般 就是翻译一下国外的文章
- 31、这本书极好极实用主义，虽主要面向商科，但完全可胜任甚至替代医科研究生的高阶统计教材。优点：用R x 4，案例实用好上手 x 3，有码有解说，为实用主义用户作了惨无人道的辩护，方法覆盖全面。缺点：代码断行简直是流氓，后几箱解说薄弱。R/Python+LaTeX才是科研统计教育的正途。
- 32、这本书，值得买、看、反复读
- 33、视野和一般的统计教程不同，不局限于统计的理论体系，而是problem-solving。实用性、学术性都很强。

《复杂数据统计方法》

34、还没仔细读过

35、吴老写书的一贯风格：举重若轻。涉及知识面广，但点到即止，统计忍者可用来自测知识广度。时间序列一章有点多余。可视化和平行及分布式计算对于复杂数据分析也重要，该书应该可以更好。

36、绝对推荐的好书 吴喜之出的几本书都特别赞啊~

37、喜欢这本书。原因：1、实用，现实数据一般都是复杂的，这本书里面的方法大多源于现实。2、书中包含了很多统计思想，不像好多教科书只告诉你套用什么公式，填鸭似的。

38、非常好的书，值得深读

39、从方法的角度介绍了统计知识，沟通了理论与工具，是学习R不可多得的好书

40、很实用，十个字的评价以下得不到积分。

41、非常好的一本书从问题入手，解决实际问题

42、真清楚救命啊吴老师是男神

43、囫囵吞枣的读了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

44、!!!

45、吴老师的书就是不一样，凝练

46、有点统计基础再看会非常好

47、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。

48、非常好的一本书，内容详尽，不是按照一般的数学模型来组织书本内容，而是根据数据情况比较不同模型的性能。非常喜欢，要花时间好好研读。

49、R入门 算法入门都很好的书。

50、复杂数据统计方法：基于R的应用

51、最近在啃的书。我不需要太多数学，告诉我怎么用，为什么就好。救命稻草

52、认真读书的人都能看出来，这是一本用心的教材。薄薄一本，足以反应作者几十载的深厚功力。建议入门级选手和非统计专业的购买前要三思，这是一本专业性很强的教材，需要相当扎实的基本功。

很久没看到国人写出的这么出色的教材了。鉴于排版和个别编码错误，打四星！

53、内容不错，难度事宜！

54、补条目

55、编书的角度特别好 从数据的形式入手对应不同的实际问题

然而很多地方没有对应的理论介绍 看着并不踏实

56、前两章节还是不错的。算是技术类工具书。

57、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。

58、学习R语言买的，还算不错吧

59、内容很好 讲得不错 值得

60、不仅介绍了如何处理复杂数据，而且还是基于R的，喜欢~

61、好几种数据放在一起讲！

62、正如书中所说，如果配上数学推导，就是一本统计百科全书。吴老对R在中国的运用和推广作出了巨大贡献。

63、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

64、著名教授的新作，我听过吴老师的课，幽默风趣，该书方法和数据都有，方便演练

65、【4.5】我觉得第一次感觉到了深入浅出的讲解。

66、简单撸了一遍，落后你们三年。。。

67、书是好书，可惜印刷质量不佳啊。

68、太好了！简直想给6个星！薄薄一本书介绍了现代统计，数据挖掘，机器学习等的核心内容，没有一句废话，含金量绝对相当于几千页的英文原版书。而且给出了代码，实践才是最好的学习方法。

69、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！

70、很实用 喜欢吴喜之

71、很接地气的一本书，间接明了的告诉你遇到什么样的case该上什么样的工具，看的畅快淋漓。同

《复杂数据统计方法》

事这本书也是一本很好的 R 教程，用附录的示例在 R 上过一遍，掌握的就七七八八了。给同行推荐。
72、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。

《复杂数据统计方法》

精彩书评

1、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。作者先从一个例子开始，公路一氧化碳数据，先用普通最小二乘，然后用逐步回归，一步一步的介绍自己的思路，我觉得这是非常难得的，很少有作者在书中介绍自己的思路的，仿佛结果是“显然的”。做完回归后，作者发现虽然残差的Shapiro-wilk检验不能拒绝残差来自正态总体，但是从QQ图上看，残差很难说是正态的。（不知道什么是Shapiro-wilk检验，QQ图的赶紧查其他书了）。通过研究变量关系发现，CO和traffic可以认为有线性关系，CO和Hour有类似sin曲线的关系，CO和Wind关系更加复杂。然后作者联想到可以通过谐波分析，用有穷Fourier级数来代表时间序列并且把traffic和wind的二次和三次项都包含到模型中（这当然是一个偏大的模型）。通过一步步的分析，最终得到一个“合适”的模型。后面都是介绍当某些假设不满足时（我刚看，还没看完呢），我们可以用某某方法来解决，也是一步一步的，发现问题，给出方法。我自己觉得，对我自己来说，知道什么时候用什么方法，我就觉得可以了，如果不满足，可以找找资料，看为什么这些方法是对的，更进一步，问问自己，这些方法是如何想到的，弄清楚后，会成大牛吧？

2、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。

章节试读

1、《复杂数据统计方法》的笔记-第16页

要用人类有限的脑细胞来应对无穷的宇宙，大多数情况下的结果是束手无策。

2、《复杂数据统计方法》的笔记-第16页

一个现象有多种假设来解释，每一种解释都不是真理，只不过是人们根据自己的标准从不同角度对客观世界的猜想。

《复杂数据统计方法》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu000.com