社会学教材教参方法系列

（第2版）

SM

分类数据分析的统计方法

Statistical Methods for Categorical Data Analysis(Second Edition)

〔美〕丹尼尔·A.鲍威斯（Daniel A. Powers）
谢　宇　/著

任　强　巫锡炜　穆　峥　赖　庆/译

社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

社会学教材教参方法系列

SM

分类数据分析的统计方法

（第2版）

Statistical Methods for Categorical Data Analysis (Second Edition)

〔美〕丹尼尔·A.鲍威斯（Daniel A. Powers）/著
谢　宇

任　强　巫锡炜　穆　峥　赖　庆/译

社会科学文献出版社
SOCIAL SCIENCES ACADEMIC PRESS (CHINA)

图书在版编目(CIP)数据

分类数据分析的统计方法 / (美)丹尼尔·A. 鲍威斯(Daniel A. Powers),(美)谢宇著;任强等译. --2版. --北京:社会科学文献出版社,2018.2

(社会学教材教参方法系列)

书名原文:Statistical Methods for Categorical Data Analysis (Second Edition)

ISBN 978-7-5201-1721-0

Ⅰ.①分… Ⅱ.①丹… ②谢… ③任… Ⅲ.①统计数据-统计分析 Ⅳ.①O212

中国版本图书馆CIP数据核字(2017)第267850号

社会学教材教参方法系列

分类数据分析的统计方法(第2版)

著　　者 / [美]丹尼尔·A. 鲍威斯(Daniel A. Powers)　[美]谢　宇

译　　者 / 任　强　巫锡炜　穆　峥　赖　庆

出 版 人 / 谢寿光

项目统筹 / 杨桂凤

责任编辑 / 杨桂凤

出　　版 / 社会科学文献出版社·社会学出版中心(010)59367159

地址:北京市北三环中路甲29号院华龙大厦　邮编:100029

网址:www.ssap.com.cn

发　　行 / 市场营销中心(010)59367081　59367018

印　　装 / 北京季蜂印刷有限公司

规　　格 / 开　本:787mm×1092mm　1/16

印　张:21　字　数:375千字

版　　次 / 2018年2月第1版　2018年2月第1次印刷

书　　号 / ISBN 978-7-5201-1721-0

著作权合同登记号 / 图字01-2017-4963号

定　　价 / 59.00元

本书如有印装质量问题,请与读者服务中心(010-59367028)联系

修订译本说明

《分类数据分析的统计方法》（第2版）一书自2009年7月出版以来，受到广大学生和教师的高度好评。此书不仅是北京大学－密歇根大学学院暑期量化课程的教学参考书，而且是许多高校讲授社会科学量化分析方面课程的教材或参考书。我们经常收到学员和教师对此书的反馈建议和评价。作为译者，我们感谢读者对本书翻译质量的肯定，同时也感谢读者针对本书一些翻译细节提出的很好的建议。这些建议加上我们在教学过程中学生提出的问题，累积起来也有不少。为此，我们在进一步理解原书内容的同时，在有机会的时候也向原书作者谢宇教授和Dan当面请教。借此书翻译修订的机会，我们对译著做了以下几方面的修正和补充。

（1）纠正了文中的一些错别字；进一步区分公式中符号的正斜体，并予以正确标注。

（2）统一了书中的个别关键词。

（3）重新翻译和修改了一些不太通顺或不太符合中文表达习惯的语句，尽可能减少语言表达上存在的明显的翻译痕迹。

（4）按照英文书籍的传统格式，我们在中文译本“主题索引”的基础上制作了中文版“索引”，在内容和格式上基本与原著的索引保持一致。

此修订译本，可能依然存在对原著的理解不足和翻译错误，或者错别字，欢迎读者批评、指正。

译者

2018年1月12日

献给我们的父母

Dick 和 Janet Powers

Liangyao Xie 和 Huazhen Zhao

中文版序

《分类数据分析的统计方法》（第2版）的中文版终于和读者见面了，我感到非常高兴。

《分类数据分析的统计方法》是我和Daniel Powers合著的，也是我的第一本书。第一版于2000年由美国的学术出版社（Academic Press）出版，第二版于2008年由英国的翡翠出版社（Emerald Group）出版。很荣幸的是，我们能在2009年英文第2版刚刚出版后不久就见到由社会科学文献出版社出版发行的中文版。

《分类数据分析的统计方法》是为社会科学——特别是社会学——做定量研究的学者和学生专门写作的教材和参考书。本书介绍、探讨了许多社会科学定量研究中实际碰到的统计方法问题。这些方法是社会科学定量研究人员都应该掌握的基本功，也是我对自己的所有学生都要求其学会的。可惜的是，一些国内的学者还认为本书包括的内容“太复杂”了。他们应该知道，社会现象本身要更复杂得多。再复杂的统计方法都是建立在我们对更复杂的社会现象做大量简化的基础之上的。虽然统计方法最终不可能让我们完美地了解社会现象，但不同的统计方法可以更好地适用于不同的社会科学研究应用之中。换句话说，统计方法虽然不能给我们十全十美的答案，但适当的统计方法相比不适当的统计方法会给我们更可信、更有学术意义的答案。所以，一个社会科学定量研究做得好的学者应该掌握各种不同的统计方法，才能做到对症下药。我希望本书中文版的出版有利于提高国内社会科学定量研究的水平。

本书的特点是着重于对有关统计方法的理解，而不是对这些统计方法的理论

证明。为了方便读者，我们另外通过互联网提供了用不同统计软件（包括 aML、GAUSS、LEM、LIMDEP、R、SAS、STATA、TDA、WinBUGS 和 OpenBUGS）编写的例题程序。网址是：http：//www. powers - xie. com；或通过我的个人主页 http：//www - personal. umich. edu/ ~ yuxie/链接；或通过 Daniel Powers 的个人主页 http：//webspace. utexas. edu/dpowers/www 链接。

本书的最初来源是我和 Daniel Powers 的教学讲义。我们也用本书作为教材教过许多学生，学生对本书有过很多好的建议。我们再次感谢他们。

翻译本书的主要负责人是北京大学的任强老师，他已和我认识多年。他在第 1 版出版不久就想翻译本书。当时我认为时机还没有成熟。2008 年我们出第 2 版时，他正好在我这里做访问学者，给翻译本书提供了很好的机会。另外三位学生巫锡炜、穆峥、赖庆也积极参加了翻译工作，做了重要的贡献。在此，我向他们四位表示感谢。

最后，我想感谢社会科学文献出版社的谢寿光社长和杨桂凤编辑。如果没有他们的积极支持和辛苦工作，就没有本书的中文版。多谢了！

谢　宇

前　言

在本书中，我们试图对社会科学研究中的分类数据分析以及应用的方法和模型做一个全面的介绍。本书主要面向研究生和社会科学应用研究的学者，同时也可以作为参考工具书。

一个区别于其他教科书主题的特点是我们有明确的目标，即整合转换方法（transformational approach）和潜在变量方法（latent variable approach），它们是处理分类数据分析的两种完全不同但相互补充的方法。在人口学和生物统计领域，处理分类数据分析的统计或转换方法是研究者最为常见的方法，而潜在变量方法则被经济学家经常使用。第1章将会讨论这两种方法。

我们假定读者已经具备初步的知识，例如，掌握了应用回归课程的知识，但不需要高级数理统计知识。尽管一些技术细节不可避免，但是我们借助了大量实例来帮助大家理解本书。一些读者可能会略过书中的技术部分，但这样也不会失去本书的很多精华。

为了充分利用互联网技术，我们为本书设置了网站（http://webspace.utexas.edu/dpowers/www/）。[①] 网站包含了书中用不同软件处理所讨论例子的数据集和程序编码，这些软件包括 GLIM（Numberical Algorithms Group Ltd.，1986），LIMDEP（Greene，2007），SAS（SAS Institute，2004），Stata（Stata，2007），TDA（Rohwer & Pötter，2000），以及 R（R Development Core Team，2006）。为

① 此主页被链接在 YuXie.com 和 Powers-Xie.com 上。

了描述估计的细节和介绍几个标准统计软件包不能估计模型的特殊程序，网站提供了一些 GLIM 的宏命令和 GAUSS（Aptech Systems，1997）与 R 子程序，如 aML（Lillard & Panis，2003）。当获得新的程序时，我们会继续更新网站内容。

◎ 第2版新增内容

我们已经更新了每一章的内容，并新增了一章关于二分类变量的多层模型（第5章）。第5章详细介绍了边际最大似然估计和现代贝叶斯估计方法（Bayesian estimation methods）。我们也针对纵贯数据分析的 Rasch 模型和随机系数模型进行了讨论，重新组织了事件史模型这一章（第6章），扩展了离散时间模型和 Cox 回归模型。对次序因变量模型（第7章）和名义因变量模型（第8章）这两章也进行了更新。

◎ 本教材在分类数据模型课程中的使用

本书适合于为期一个学期的分类数据建模课程。第1章和第2章是一般性介绍与课程基础。我们的观点是，无论数据类型如何，回归类建模方法都是一个合适的分析方法。第3章介绍并详细讨论了针对二分类数据的回归模型。第4章深入讲解了分析列联表的模型。第5章讨论了针对二分类数据的多层/分层模型。第6章介绍了事件史分析技术。第7章和第8章回顾了针对次序和非次序分类因变量的模型。这部分内容与第4章的列联表方法和第3章介绍的潜在变量分析框架是有关联的。

◎ 致谢

在本书写作的各个阶段，我们从下列学者的鼓励以及与他们的联系中获益匪浅：Paul Allison、Mark Becker、John Fox、Richard Gonzalez、Leo Goodman、David Grusky、Robert Hauser、Michael Hout、Kenneth Land、Scott Long、Charles Manski、Robert Mare、Bill Mason、Susan Murphy、Trond Peterson、Thomas Pullum、Adrian Raftery、Steve Raudenbush、Arthur Sakamoto、Herbert Smith、Michael Sobel、Chris Winship、Raymond Wong、Larry Wu 和 Kazuo Yamaguchi。此外，我们对许多学习这门统计课程的研究生表示感谢，是他们激励我们写这本书的。

资助丹尼尔·A. 鲍威斯的奥斯汀得克萨斯大学的主任基金、资助谢宇的国家自然科学基金的青年学者基金和密歇根大学基金对本书的研究提供了部分资助。

我们也要感谢外部评审对早期初稿提出的宝贵意见，以及 Pam Bennett、John Fox、Kimberly Goyette 和 James Raymo 对书稿最后版本的仔细校对和在第 1 版中对实例的编程工作。感谢 Meichu D. Chen 和许多研究生，他们指出了第 1 版中的一些错误。特别感谢 Cathy（Hui）Liu 对第 2 版新内容的仔细阅读。我们也要感谢 Cindy Glovinsky 卓越的编辑工作。我们将对书中仍然存在的错误负责。

最后，我们感谢学术出版社（Academic Press）和 Elsevier 的编辑 J. Scott Bentley 提出这个项目，并努力使第 1 版面世，同时促使我们完成第 2 版的编写工作。感谢 EmeraldInsight 的 Rachel Brown 女士对出版第 2 版的帮助，也要感谢 Macmillan 对编排此书的帮助。

丹尼尔·A. 鲍威斯

谢 宇

目录
CONTENTS

图目录

表目录