



(第2版)

社会学教材教参方法系列

SM

分层线性模型: Hierarchical Linear Models:

应用与数据分析方法

Applications and Data Analysis Methods(Second Edition)

[美] 斯蒂芬·W. 劳登布什 (Stephen W. Raudenbush)
安东尼·S. 布里克 (Anthony S. Bryk)

/著

郭志刚 等/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

分层线性模型： 应用与数据分析方法 (第2版)

**Hierarchical Linear Models:
Applications and Data Analysis Methods(Second Edition)**

[美] 斯蒂芬·W. 劳登布什 (Stephen W. Raudenbush)
安东尼·S. 布里克 (Anthony S. Bryk)

/著

郭志刚 等/译



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

图书在版编目(CIP)数据

分层线性模型：应用与数据分析方法：第2版／

(美) 斯蒂芬·W. 劳登布什 (Stephen W. Raudenbush) ,
(美) 安东尼·S. 布里克 (Anthony S. Bryk) 著；郭志刚
等译. -- 北京：社会科学文献出版社，2016.9

(社会学教材教参方法系列)

书名原文：Hierarchical Linear Models；

Applications and Data Analysis Methods

(Second Edition)

ISBN 978-7-5097-9219-3

I. ①分… II. ①斯… ②安… ③郭… III. ①分层模
型-线性模型-应用-统计分析 IV. ①C813

中国版本图书馆 CIP 数据核字 (2016) 第 118990 号

· 社会学教材教参方法系列 ·

分层线性模型：应用与数据分析方法（第2版）

著 者 / [美] 斯蒂芬·W. 劳登布什 (Stephen W. Raudenbush)

[美] 安东尼·S. 布里克 (Anthony S. Bryk)

译 者 / 郭志刚 等

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·社会学编辑部 (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：30.75 字 数：545 千字

版 次 / 2016 年 9 月第 2 版 2016 年 9 月第 1 次印刷

书 号 / ISBN 978-7-5097-9219-3

著作权合同

/ 图字 01-2015-4379 号

登 记 号

定 价 / 79.00 元

本书如有印装质量问题，请与读者服务中心 (010-59367028) 联系

 版权所有 翻印必究

Hierarchical Linear Models: Applications and Data Analysis Methods (Second Edition)

Copyright © 2002 by Sage Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher. Printed in the United States of America.

本书由 Sage Publications, Inc. 于美国印刷并授权社会科学文献出版社印行中文版

中文版译者序

《分层线性模型——应用与数据分析方法》(第2版)的中文版终于与读者见面了。

分层线性模型是1990年代在国际上形成并正在迅速推广应用的新统计分析技术。由美国芝加哥大学的布里克教授和密歇根州立大学的劳登布什教授于1992年合著的《分层线性模型——应用与数据分析方法》是这一统计分析方法的代表作之一。

我是在1990年代后期才知道这种分析方法的。2000年,在美国执教的朋友王丰博士送给我此书。在研读过程中,我深为其方法论上的推进和其广泛的应用性所吸引。

此书既是一本专著,也可以作为教材,因为作者实际上是按教材体例写的。该书首先从方法论角度展开,指出研究实践中最常用的常规回归模型在处理多层次关系时的无效性。然后介绍了分层线性模型的原理,讨论了该模型在多种典型情况下的应用,并附以许多例题的分析示范,十分解渴。显然,此书的宗旨并不囿于证明一种新方法,而是考虑了能够为更多的研究人员和学生所理解,以促进该方法的普及应用。事实上,该书作者同时也是同名的专门统计软件HLM的研制者。

我于2001年开始组织一些教师和博士研究生一起分工翻译该书。2002年,劳登布什教授与布里克教授又出版了该书的第2版。其内容大大扩展,从原来的10章扩展为14章,从原来的260页扩展为480页,以充分反映10年来这一方法取得的新进展。我们随即决定中止第1版的翻译工作,改为翻译第2版。本来我们翻译此书的目的只是团结一批有志于量化研究方法的青年学者一起追踪国际学术前沿,但经过2年时间的反复校正与修改,最终我们形成了格式统一、译法一致、图表俱全的译稿。

北京大学社会学系马戎教授等同仁见到这个译本后,一致鼓励我们应当争

取正式出版,并且还为此事多方联系。但是出于各种原因,正式出版之事一再受挫。最后,在社会科学文献出版社谢寿光社长的鼎力支持下,这本译著才得以正式出版。

本书的翻译采用团队工作形式,由郭志刚(第1、2、8、10、12章)、郑真真(第11章)、陈卫(第5章)、周皓(第6、7、14章)、李强(第3、4章)、葛建军(第13章)、张磊(第9章)共同承担,最后由郭志刚负责全书的统稿和校对工作。我的研究生巫锡炜、赵联飞、李睿、王军协助进行最后的清样校对工作。

感谢杨桂凤编辑高度认真负责的精神,她的辛勤努力使这一中文版增强了文本的一致性,并提高了语言上的易读性。

由于译者水平有限,对此学术前沿巨著的理解和翻译难免有不当之处,恳请读者指教。

郭志刚

于海淀蓝旗营

献给我们的爱妻 Stella Raudenbush 和 Sharon Greenberg

致谢（英文版第2版）

自本书第1版出版以来的十年中，分层模型的有关方法得到了长足的发展，并且在各领域中得到了广泛的运用。因此可以说，本书第2版一方面是原作者之间长期大量合作的成果，另一方面也是与更多的其他同事合作与讨论的成果。需要感谢太多的人，但在此处不能一一列出。无论如何，其中一些人是必须提到的。正是因为有了他们，本书才最终得以付梓。

与 Darrell Bock、Yuk Fai Cheong、Sema Kalaian、Rafa Kasim、Xiaofeng Liu 和 Yasuo Miyazaki 等人在方法论上的讨论，不断地挑战着我们的想法。Yeow Meng Thum 的工作启发了本书第6章和第11章中有关多元分析的应用。Mike Seltzer 对第13章的贝叶斯方法提出了极为重要且非常有用的批评，并慷慨地允许将其研究成果作为最后一个例子加入该章。Meng-Li Yang 和 Matheos Yosef 在发展“分层一般化线性模型”（第10章）中所用的最大似然估计方法方面做了基础工作。Young-Yun Shin 仔细阅读了本书的初稿，并提出了许多建设性意见。Guang-lei Hong 对初稿提出的批评意见促使第12章交互分类模型的形成。作为应用程序员和长期的朋友，Richard Congdon 的工作体现在本书的每一章中。Stuart Leppescu 也为第2版的数据处理和新的分析工作提供了协助。

芝加哥社区人类发展项目（the Project on Human Development in Chicago Neighborhoods, PHDCN）中的同事，包括 Felton Earls、Rob Sampson 和 Christopher Johnson，都对本书第2版有着重要影响，这体现在第10章和第11章中关于社区效应的例子上。的确，我们还要感谢麦克阿瑟基金会（the MacArthur Foundation）、国家司法研究所（the National Institute of Justice）和国家精神卫生研究所（the National Institute of Mental Health）对芝加哥社区人类发展项目的资助，该项目也支持了本版新章节中关键性方法论方面的工作。

我们特别感谢 Pamela Gardner，她帮助检查、编辑并录入了本书的全稿。

她的高效率和幽默感是保证整个工作顺利进行的基础。

匿名评论者对这一版的新章节提出了许多有益的建议。作为 Sage 出版社的方法论编辑，C. Deborah Laughton 对本书的出版给予了令人钦佩的耐心和精神上的支持；我们还要再次感谢系列丛书的主编 Jan de Leeuw 的鼓励。

丛书主编对分层线性模型的介绍

在社会科学中，数据结构经常在以下意义上是分层的：我们有描述个体的变量，但是个体又组成较大的群体，每一群体由一定数量的个体组成。对于较大群体还有一系列变量来描述。

首推的例子也许是教育。学生组成班级，既有变量描述学生，又有变量描述班级。班级变量也许是学生变量的汇总指标，诸如学生数量或社会经济状况的平均值。但是班级变量也可以是教师（如果本班只有一个教师），或者是本班的教室（如果本班总是聚会于同一教室）。此外，在这个例子中，层次结构还会相当自然地延伸。班级还会进一步组成学校，学校构成校区，等等。我们也可以有描述学校的变量和描述校区的变量（如教学方式、学校建筑、邻里状况等）。

一旦我们发现这个分层数据结构的例子，我们就会看到更多的例子。它们自然地发生于地理和（区域）经济中。在某种意义上，社会学的基本问题就是将个人属性与所寓于其中的群体和结构的属性联系起来。以同一种方式，经济学的问题是将微观层次与宏观层次联系起来。此外，许多重复测量也是分层的。如果我们不断追踪调查一些个体，那么对任一个体的观察构成一组测量，按照同一种方式，一个学校或班级也构成一组测量。当每个调查员调查一组对象时，调查员便处于较高层次。只要对这些分层结构多加思索便会不可避免地导致一个结论，即使不能说大多数，至少也能说很多社会科学数据都有这种嵌套的或分层的结构。

在意识到分层数据的重要意义之后，下一步是考虑如何在统计技术上体现对分层结构的分析。有两种做法已经遭到否定。第一种做法是将所有高层变量分解（disaggregate）到个体水平。比如将教师、班级和学校的特征全都赋予学生个人，然后在个体层次进行分析。这一问题的问题是，要是我们知道一些学生是同一班级的，那么我们也就知道他们在班级变量上取相同的值，于是，我

们便不能采用经典统计技术的基本假定，即各观测之间相互独立。另一种做法是先将个体水平的变量汇总到较高层次，然后在较高层次进行分析。比如我们将学生特征汇总到班级，然后再对班级进行分析，分析时也许还需要按规模对班级加权。这种做法的主要问题是，我们抛弃了所有的组内信息，它也许占到最初分析时信息总量的 80% 或 90%。其结果是，汇总变量之间的关系总是显得较为密切，然而这常常与直接分析未曾汇总的变量所反映的情况大相径庭。于是，我们不但浪费了信息，而且要是我们企图在个体层次解释汇总分析结果，还将导致对结果的曲解。所以，汇总方法和分解方法都不能令人满意。

如果仅限于用常规线性模型进行分析，我们知道其基本假定有线性、正态性、方差齐性、独立性。我们希望保留前两项，但需要对后两项（尤其是关于独立性的假定）加以修改。修改的主要原因在于：由于同组的个体之间比异组的个体之间更为接近或相似，所以一个学生与其他班级的学生可能是相独立的，而与同一班级的学生在许多变量上相同。那么，这些变量并不需要观察，这意味着它们从线性模型中消失，进入误差部分，并导致残差之间出现相关。这一思路可以用方差成分模型来做公式表达。残差之间可分为组成分和个别成分两种。个别成分之间相互独立，组与组之间也是独立的，而组成分与组特征完全相关（即组内不独立）。一些组可能比另一些组有更强的方差齐性，即这些组的组内成分方差有所不同。

对这一思路的公式表达也可采取稍微不同的形式。假定每一个组都有不同的回归模型，在简单回归的情况下，每个组都有自己的截距和斜率。由于各组都是通过抽样得到的，因此我们假定这是以组为单位构成的总体的组截距和组斜率的随机样本。这定义了一套随机系数的回归模型。要是我们仅仅假定它们的截距是随机的，而令所有的斜率相同，其实我们面对的正是前面提到过的方差成分研究的情况。要是允许斜率也可以随机变化，这套模型的情况便更为复杂，残差的协方差将依赖于个体层次自变量的取值。

在随机系数的回归模型中，也仍然不能与较高层次的（比如描述班级或学校的）变量相联系。为了达到这一目的，我们需要多层模型，其中以组为单位的模型还是线性模型。于是我们假定，学业能力测试值这一学生变量的斜率与班级规模或教师特征等班级变量之间呈线性关系。所以，在每一层次都有一整套反映分层结构的模型，使各个层次的变量联系起来。

直到大约 10 年前，拟合这样的模型才在技术上成为可能。大约与此同时，

Aitkin 和 Longford、Goldstein 及其合作者，以及劳登布什和布里克纷纷研发出了有关分析技术及其相应的计算软件。其中，由布里克和劳登布什研发的 HLM 软件对用户最友好，在技术上也最完善，并且他们还及时发表了一系列既有说服力又有意思的范例。在本书中，布里克和劳登布什详细地描述了这种模型、其计算方法，以及有关程序和范例。我认为，现在来对这一技术方法提供一个完整的介绍是既重要又及时的。分层线性模型（hierarchical linear models），或称多层模型（multilevel models），当然不能解决社会科学中所有的数据分析问题，因为它们还是建立在线性关系和正态分布假设基础之上的模型，并且它们所研究的仍是相对简单的回归结构，只是其中一些变量依赖于其他变量。然而，它们在技术层面已经使汇总和分解的做法“向前迈了一大步”，这主要是因为它们在统计原理上是正确的，并且避免了信息的浪费。

我认为，这本书的主要贡献在于通过一系列例题的分析示范为读者提供了清晰的概念。本书对各种不同层次的模型做了清楚的分类，同时又未完全割断其间的联系。读者可以分别把握某一层次中可能的机制，然后再将不同层次的模型联系起来进行分析。未来，这些技术会对教育研究以及地理学、社会学和经济学的研究产生重要影响，直到它们也遇到其自然限制。为了避免这些限制，这些模型还将扩展（并且已经扩展）到更多层次，扩展到多元数据，扩展到通径分析、潜在变量、名义因变量以及一般化线性模型，等等。社会统计学家将有能力建立更为广泛的模型，并且可以在更多的模型中进行选择。如果他们能够建立起必要的前提信息条件，作为从这一模型各类型中进行合理选择的依据，那么有望获得更强的能力和精度。请把这一思想记在心底，因为你将使用这本书来探索这一类新兴而又令人兴奋的技术方法。

目 录

致谢(英文版第2版)	1
丛书主编对分层线性模型的介绍	1
丛书主编对第2版的介绍	1

第一部分 原理

1 导言	3
分层数据结构:一个常见现象	3
分层数据分析中持续存在的两难问题	4
分层模型统计理论的发展简史	5
分层线性模型的早期应用	6
个体效应的改进估计	6
对层次之间效应的建模	7
分解方差协方差成分	9
本书第1版问世以来的新发展	9
结果变量范围的扩展	9
与交互分类数据结构的结合	11
多元模型	11
潜在变量模型	12
贝叶斯推断	12
本书的框架结构	13

2 分层线性模型的原理	15
初步知识	15
对某一学校的社会经济状况与成绩关系的研究	15
对两个学校的社会经济状况与成绩关系的研究	17
对 J 个学校的社会经济状况与成绩关系的研究	18
一般模型及其简单子模型	22
带随机效应的单因素方差分析	22
将平均数作为结果的回归模型	23
带随机效应的单因素协方差分析	24
随机系数回归模型	25
将截距和斜率作为结果的回归模型	27
非随机变化斜率模型	27
本节概要	28
基本分层线性模型的推广	28
多元 X 和多元 W	28
对层 -1 和层 -2 上的误差结构的推广	30
超出基本的两层分层线性模型的扩展	30
选择 X 和 W 的定位(对中)	30
X 变量的定位	31
W 变量的定位	33
本章术语及注释概要	34
简单的两层模型	34
注释与术语概括	34
一些定义	35
子模型的类型	35
3 分层线性模型估计及假设检验的原理	37
估计理论	37
固定效应的估计	38
随机层 -1 系数的估计	45
方差协方差成分的估计	50

假设检验	55
固定效应的假设检验	56
随机层 -1 系数的假设检验	60
方差协方差成分的假设检验	62
本章术语概要	63
4 示例	66
引言	66
单因素方差分析	67
模型	68
结果	68
以均值作为结果的回归	70
模型	70
结果	71
随机系数模型	73
模型	73
结果	75
以截距和斜率作为结果的模型	78
模型	78
结果	79
估计一个特定单位的层 -1 系数	83
最小二乘法	83
无条件收缩	84
条件收缩	87
区间估计的比较	90
需要注意的问题	90
本章术语概要	91

第二部分 基本应用

5 组织研究中的应用	95
组织效应研究的背景	95

建立模型	96
个人层次模型(层-1)	96
组织层次模型(层-2)	97
例1:通过随机截距模型对组织共同效应建模	98
一个简单的随机截距模型	98
例子:考察学校对教师效率的影响	99
与传统的教师层次和学校层次分析结果的比较	102
包括层-1协变量的随机截距模型	106
例子:写作的项目效果评估	107
与传统的学生层次和班级层次分析结果的比较	108
例2:通过以截距和斜率为因变量的模型来解释组织的不同效应	112
过去在建立以斜率为因变量的回归模型时所遇到的困难	112
例子:公立学校和天主教学校成绩的社会分布	113
层-1既有随机斜率又有固定斜率的应用实例	123
专题	124
层-1方差异质性情况下的应用	124
例子:对数学成绩的层-1残差方差的类别效应建模	125
层-1存在异质性情况下的数据分析建议	127
组织效应研究中层-1自变量的对中	127
层-1固定系数的估计	128
分离个人层次效应和构成效应	131
对层-1协变量调整后的层-2效应估计	134
估计层-1系数的方差	135
估计层-1随机系数	140
使用方差统计量的削减比例	141
估计个别组织的效应	143
具体组织的效应的概念化	143
常用的学校业绩估计	144
经验贝叶斯估计的使用	144
对业绩指标进行有效推断所面临的威胁	146
设计两层组织效应研究时对功效的考虑	149

6 个体变化研究中的应用	151
个体变化研究中的背景问题	151
建立模型	152
重复观察模型(层-1)	153
个人层次模型(层-2)	153
线性增长模型	154
例子:教导对认知发展的作用	155
二次增长模型	160
例子:母亲的语言能力对儿童词汇量的影响	160
其他形式的增长模型	167
在层-1 误差结构更为复杂时的情况	167
分段线性增长模型	168
随时间变化的协变量	170
个体变化研究中层-1 自变量的对中	171
线性增长模型中截距的定义	171
在高阶多项式模型中其他增长参数的定义	172
在研究随时间变化的协变量时的可能偏差	173
增长参数的方差估计	174
比较分层模型、多元重复测量模型和结构方程模型	176
多元重复测量模型	176
结构方程模型(SEM)	177
例1:观察数据是平衡的	179
例2:完整数据是平衡的	179
例3:完整数据是不平衡的	187
层-1 中缺失观测值的影响	189
利用分层模型来预测未来情况	191
增长与变化的研究设计中有关功效的考虑	193
7 HLM 在元分析和其他层-1 方差已知情况下的运用	195
引言	195
元分析数据的分层结构	196