

 *Social Science Research Methodology Series*

社会科学
研究方法
丛书

追踪数据 分析方法 及其应用

刘红云 张雷 著

Longitudinal Data Analysis
Methods and Its Applications

Hongyun Liu & Lei Chang

教育科学出版社
Educational Science Publishing House

刘红云 张雷 著

Longitudinal Data Analysis
Methods and Its Applications

追踪数据分析方法
及其应用

教育科学出版社

· 北京 ·

策划编辑 韦 禾
责任编辑 葛 都
版式设计 尹明好
责任校对 张 珍
责任印制 叶小峰

图书在版编目 (CIP) 数据

追踪数据分析方法及其应用/刘红云, 张雷著. —北京: 教育科学出版社, 2005.5

(社会科学研究方法丛书/张雷, 侯杰泰主编)

ISBN 7-5041-3210-1

I. 追... II. ①刘...②张... III. 统计分析(数学)
IV. 0212.1

中国版本图书馆 CIP 数据核字 (2005) 第 022327 号

出版发行 教育科学出版社

社 址 北京·朝阳区安慧北里安园甲 9 号

邮 编 100101

传 真 010-64891796

市场部电话 010-64989009

编辑部电话 010-64989228

网 址 <http://www.esph.com.cn>

经 销 各地新华书店

印 刷 涿州市星河印刷有限公司

开 本 787 毫米×960 毫米 1/16

印 张 22.5

字 数 350 千

定 价 34.00 元 (含光盘一张)

版 次 2005 年 5 月第 1 版

印 次 2005 年 5 月第 1 次印刷

印 数 00 001-2 000 册

如有印装质量问题, 请到所购图书销售部门联系调换。

作者简介



刘红云, 1996年毕业于山西大学数学系, 获硕士学位; 2000年毕业于北京师范大学心理学院, 获博士学位。2003年—2004年在香港中文大学教育心理系作博士后研究员一年, 现为北京师范大学心理学院讲师。主要研究方向为心理统计和测量、教育评价。



张雷, 先后在河北大学及美国南加州大学获得学士、硕士和博士学位, 曾执教于美国南加州大学和中佛罗里达大学, 现为香港中文大学教育心理系教授。主要研究方向为心理统计与测量及儿童社会性发展, 在学术期刊发表论文70余篇。在*International Journal of Behavioral Development* (美国)和*Journal of Psychology in Chinese Societies*(香港)任副主编, 并在*Educational and Psychological Measurement*(美国)和*Journal of Child Psychology and Psychiatry*(英国)任编委。

总 序

对现代科学尤其是社会科学而言，研究方法的发展在很大程度上能够起到推动整个学科发展的作用，研究方法的落后必然会限制学科的发展。在我国现阶段，同学科专业的建设及发展相比，社会科学研究方法的建设 and 普及却显得相对薄弱，很多学科的研究方法，尤其是量化方法，还远远落后于世界先进水平，这无形中妨碍了有关学科专业的进一步发展。

社会科学研究方法的落后，首先表现在缺乏一大批热心于研究方法的学者，多数优秀学者往往致力于专业研究，而较少顾及到研究方法的研究。其次表现为此领域相关书籍的匮乏，特别是介绍、讲解现代研究方法的专著不仅数量有限，而且内容陈旧，更没有一套可以系统介绍社会科学研究方法的教材，很多学科苦于找不到合适的方法课用书。而在先进国家，每一种主要的统计分析方法在每一学科中都可找到多种专著及更多的普及读物，以供学者、学生选用。针对这样的情况，我们决定编辑这套社会科学研究方法丛书。

在我国学科建设飞速发展的 20 年里，社会科学研究方法在西方也经历了一场空前的革命，20 世纪 70 年代 EM (Expectation Maximization) 算法的出现和计算机技术的发展，带来了新一代的统计和测量理论及方法。传统的回归及方差分析和经典测量理论，虽然仍广泛应用，但已不再是主要的、更不是惟一的研究方法系统，也无法应付由新一代研究方法带动下的学科专业发展。

新一代统计分析方法最突出的发展是结构方程的发展和应用。到了 20 世纪 80 年代，结构方程这套新的数据分析系统已经成熟，目前正为广大社会科学研究人员所接受，并成为各类社会科学学科研究生的必修课。结构方程的路径分析思想拓展了社会科学研究的思路，很多学科的专业课题已由过去的只研究单变量转变成研究多变量，由分析主效应到同时分析交互效应，由对单指标和直接观测变量进行研究到对多指标和潜变量进行研究。与此同时，结构方程的技术日趋专业化、深入化和复杂化，不少学者以此为研究专业，并且也有专门的学术期刊专注于结构方程及其相关技术的发展和應用。

统计分析方法的另一大突破性发展是多层分析的理论和方法。多层和嵌套分析的思想由来已久，但直到 20 世纪 90 年代才发展成一套完整而系统的理论和方法，并逐渐推广和为研究人员所接受。多层分析技术系统地解决了困扰社会科学半个多世纪的生态谬误 (Ecological Fallacy) 问题。在教育、管理、经济以及社会学、心理学等领域的研究中，取样往往呈嵌套结构，例如学生嵌套于学校，学校又嵌套于社区，这种嵌套型的样本用传统的 OLS 回归方法分析会导致估计误差。多层分析方法不仅可以减少这种统计误差，而且可以避免由人为选择分析单位而可能出现的错误。在多层分析中各层样本均可作为分析单位，而且还可以研究他们之间的交互作用，从而拓宽了各专业的研究范围，深化了各专业的研究思路。目前多层分析方法日趋成熟，并在新一代统计分析方法中处于前沿位置。

新一代统计方法的发展还体现在对追踪数据和发展模型的处理上。这一进展的主要特点是借助和延伸已有的方法，建构一套处理重复数据的具体步骤。目前最为广泛接受的是基于结构方程的追踪数据分析方法和基于多层分析技术的发展模型方法。这些方法通过对结构方程与多层分析的巧妙应用，而各自形成了一套独立的系统。

此外，在上述统计方法系统发展的同时，很多专门的数据处理技术也迅速

改进。分析类别变量 (categorical variable) 的方法就是其中一个例子。传统的统计方法多建立在对连续变量进行分析的基础上, 而对建立在分类资料基础上的社会科学问题则不能很好地加以解决, 分析类别变量的方法则解决了这一难题。另一个困扰研究者的统计问题是如何对缺失值 (missing value) 进行处理, 新一代的基于 bootstrap 等的方法则能更有效、更完整地处理缺失值。

随着新一代统计分析理论及方法的发展、推广与应用, 测量方法也有了质的飞跃。与自然科学相比, 社会科学的主要难题在于难以直接、精确地对人类行为加以测量, 而要依靠测量理论对其进行推断。然而, 传统的经典测量理论不能同时对项目难度和考生能力进行估计, 无法应付大规模自适应考试的要求。随着计算机技术的发展, 经典测量理论在教育 and 心理测量领域的应用逐渐让位于项目反应理论 (Item Response Theory, IRT)。美国和欧洲的著名测量机构均把 IRT 应用于设计、测试集成、测试校准、建构测试题库以及其他的测验发展过程。计算机自适应测验 (computer adaptive test) 的普遍应用更是 IRT 在测量应用方面的一大贡献。

另一个新一代的测量理论和方法是概化理论, 概化理论最近的发展是同人们对考试认识的改变分不开的。越来越多的学者和教育工作者对标准化、多重选择类的考试方法提出了质疑, 表现性评价 (performance assessment) 的思想则逐渐生根。然而, 随之出现的问题是由这种直接操作的测试方法所产生的多种误差无法用传统测量理论来解决, 于是概化理论在这种新环境中的作用逐渐被人所认识并得到发展。此外, 人力资源测评、管理咨询等也推动了概化理论的发展和应用。

这套丛书将对上面提到的新一代数据分析和测量评估的理论及方法陆续进行介绍, 旨在推动我国社会科学研究方法的发展, 改变该领域发展滞后的现状。每一本专著的作者都是该领域颇有建树的专家, 在写作风格上既强调数理及专业技术方面的严谨性, 也兼顾语言和介绍方式上的通俗化, 从而适合具有不同数学背景读者的需要。丛书的另一个特点是对方法应用性的强调, 旨在满足大多数社会科学专业研究人员对应用研究方法的需要。为了加强其应用性, 丛书要求作者从实际操作入手, 以具体研究为例, 准确清楚地介绍各研究方法的操作步骤。丛书的第三个特点是重点突出, 一般只抓住某一方法和理论的中心, 尽量不涉及偏难及纯技术的, 尤其是在理论上尚有争议或技术上不成熟的问题,

从而使初学者能够掌握到要领而又不致过于肤浅，或者感到信息量太大而难于接受、消化。最后，这套丛书的内容均是作者结合自己的研究实践亲笔撰写而就，并非翻译、编纂之举。

这套丛书可以作为教育、心理以及其他社会科学学科的研究生和部分本科生的教材，也可以作为从事社会科学研究的人员的参考书、工具书。我们希望这套丛书的出版能起到抛砖引玉的作用，唤起更多热心于方法研究的学者多出书、出好书，多参与我国社会科学研究方法和理论的建设工作，把我国的社会科学研究方法推向一个新起点。

张雷 侯杰泰
2002年12月于香港

序 言

在社会、经济、管理等研究领域,有时需要对事物的变化或发展规律进行研究,或者对变量之间因果关系假设的合理性进行验证,这些都是横断研究不好解决的问题。近年来,随着一系列多元统计方法和计算机软件的迅速发展,处理追踪数据的分析方法也得到了很大的改进,其中最具代表性的是多层线性模型和结构方程模型在处理追踪数据时的应用。这一系列新型的处理追踪数据的方法,不论在追踪研究设计方面,还是在结果解释方面都具有传统分析技术无法比拟的优势。本书力求从应用的角度,讲述追踪数据常用的分析方法及其新进展,并对不同方法之间的优劣进行比较,希望能够对不同学科领域的量化研究起到积极的推动作用。

本书共分三部分。第一部分介绍重复测量方差分析和多元方差分析在追踪数据处理中的应用;第二部分介绍多层线性模型在追踪数据处理中的应用;第三部分介绍潜变量增长曲线模型在追踪数据处理中的应用。

在本书的写作过程中,张咏梅、骆方和朱军梅都提出了许多宝贵的意见,并给予了许多帮助。教育科学出版社的韦禾和葛都两位编辑为本书的出版付出了辛勤的劳动,在此一并表示感谢。

本书的撰写得到香港研究资助局(RGC)授予第二作者的资助(项目号:4339/01H)和香港中文大学博士后研究基金的资助,特此鸣谢。

本书虽经反复校阅,但难免存在一些错误和不当之处,敬请广大读者指正。

刘红云 张雷

2004年12月

目 录

序言

引言

1. 追踪研究特点及研究问题 (1)
2. 追踪研究设计的类型 (2)
3. 追踪研究中数据的缺失 (7)
4. 追踪研究的效度问题 (9)
5. 追踪研究数据分析 (11)
6. 本书内容安排 (12)

第一部分 方差分析在追踪研究中的应用

- 第一章 追踪数据的一元方差分析 (14)
 - 第一节 追踪设计的常见类型 (14)
 1. 追踪设计的常见类型 (14)
 2. 追踪测量应该注意的问题 (17)
 - 第二节 追踪数据的一元方差分析 (18)

1. 追踪数据一元方差分析基本原理	(18)
2. 追踪数据方差分析的正交对照	(22)
第三节 追踪数据一元方差分析的假设	(25)
1. 追踪数据方差分析的基本假设	(25)
2. 对偏离球形假设的校正	(29)
3. 追踪数据方差分析的事后检验	(31)
第四节 追踪数据一元方差分析的效应模型	(34)
1. 一组被试、一个变量、多个时间点的测量	(34)
2. 多组被试、一个变量、多个时间点的测量	(35)
3. 一组被试、多个变量、多个时间点的测量	(37)
第五节 追踪数据一元方差分析应用	(39)
1. 一组被试、一个变量、多个时间点的测量	(39)
2. 多组被试、一个变量、多个时间点的测量	(45)
3. 一组被试、多个变量、多个时间点的测量	(51)
4. 多组被试、多个变量、多个时间点的测量	(59)
第二章 追踪数据的多元方差分析	(66)
第一节 追踪数据多元方差分析基础	(67)
1. 追踪数据的多元方差分析	(67)
2. 多元方差分析的检验统计量	(69)
第二节 多元方差分析的假设	(74)
1. 多元方差分析的基本假设	(74)
2. 检验多元方差分析假设条件的一般步骤	(79)
第三节 多元方差分析应用举例	(80)
1. GLM 的应用	(81)
2. MANOVA 的应用	(101)
小结	(112)

第二部分 多层线性模型在追踪研究中的应用

第一节 追踪数据变化趋势的初步分析	(118)
1. 追踪数据文件的建立	(118)

2. 个体随时间变化的初步描述	(120)
第二节 追踪研究中的多层次模型简介	(124)
1. 追踪研究中的多层线性模型	(125)
2. 多层线性模型的参数估计	(127)
3. 多层线性模型的假设检验	(130)
第三节 多层线性模型应用	(131)
1. 模型定义过程	(131)
2. 主要输出结果	(139)
第四节 多层线性模型的推广应用	(146)
1. 测量次数相同, 时间间隔不同	(147)
2. 测量次数不同, 时间间隔也不同	(149)
3. 非线性增长模型	(151)
4. 含有随时间变化的预测变量模型	(154)
第五节 多元多层分析模型在追踪研究中的应用	(158)
1. 多元发展的多层次模型	(159)
2. 多元增长模型应用举例	(160)
第六节 多层线性模型中误差的协方差结构	(174)
1. 多层线性模型中误差之间关系的“标准”假设	(174)
2. 多层线性模型与重复测量的方差分析	(177)
3. 多层线性模型与多元方差分析	(178)
4. 几种常见的误差协方差矩阵结构	(179)
5. 正确选择误差协方差矩阵的重要性	(181)
小结	(181)

第三部分 潜变量增长曲线模型在追踪研究中的应用

第一节 潜变量增长曲线模型简介	(186)
1. 只有两个测量时间点的两因子 LGM	(186)
2. 定义增长曲线类型的 LGM	(190)
3. 不定义曲线类型的两因子 LGM	(192)
4. 单因子潜变量增长曲线模型	(193)

第二节 潜变量增长曲线模型应用	(194)
1. 潜变量线性增长模型	(195)
2. 潜变量二次增长曲线模型	(200)
3. 不定义曲线类型的潜变量增长曲线模型	(202)
第三节 潜变量增长曲线模型的多样本比较	(203)
1. 多样本比较通常关注的问题	(204)
2. 潜变量增长曲线模型多样本比较的一般方法	(205)
3. 另一种检验多样本增长趋势是否相同的多样本比较的方法	(211)
第四节 多元潜变量增长曲线模型	(214)
1. 一阶相关潜变量增长曲线模型	(215)
2. 高阶潜变量增长曲线模型	(219)
第五节 潜变量增长曲线模型在群组序列设计中的应用	(226)
1. 群组序列设计的 LGM 模型	(227)
2. 不定义增长曲线类型的群组序列设计的 LGM	(233)
第六节 含有缺失数据的潜变量增长曲线模型	(235)
1. 潜变量增长曲线模型处理数据缺失值的方法	(236)
2. 多组比较方法的推广	(241)
第七节 多层次潜变量增长曲线模型	(243)
1. 全息极大似然估计	(244)
2. 多水平潜变量增长曲线模型	(248)
第八节 潜变量增长曲线模型和重复测量的方差分析	(254)
1. 追踪数据的 MANOVA 和 LGM	(255)
2. 包含有预测变量和结果变量的潜变量增长曲线模型	(260)
小结	(271)
附录	(274)
参考文献	(338)

引 言

追踪研究是社会学、心理学、教育学、经济学等领域普遍采用的一种方法，这种研究设计为了解事物的发生、发展和变化提供了有效的途径。本书主要介绍几种用于追踪研究数据处理的统计方法，并将这几种方法在应用上的优缺点进行比较。

1. 追踪研究特点及研究问题

追踪研究在社会科学研究中处于特殊的地位，这种研究主要用来分析一段时间或某几个时间点个体的增长趋势和个体之间的差异。追踪研究与横断研究相比，最大的优点是追踪研究设计可以合理地推论变量之间存在的因果关系。从方法论的角度讲，要想得出变量之间的因果关系，原因变量和结果变量之间至少要满足下列三个条件：(1)假设存在因果关系的原因变量和结果变量之间是相关的；(2)从时间上来讲，原因变量在前，结果变量在后；(3)在所考虑的模型中，其他原因变量对结果变量的影响能够被控制或排除。可见从方法论的角度来讲，横断研究永远不可能满足上述的第二个条件，所以要从横断研究数据本身探索变量之间的因果关系，几乎是不可能的。正是由于追踪研究有这一显著优点，所以在社会科学研究中，多用追踪研究探讨数据之间的因果关系和分析事物的增长规律。

追踪数据指的是一被试群体在一个或多个变量上、多个时间点的测量结果。假如一组追踪数据中有 N 个个体，所关心的变量有 M 个，测量时间点为 T 个，那么追踪数据与横断数据显著的区别是追踪数据有多个时间点，即 $T > 1$ ，而在横断研究中 $T = 1$ 。追踪研究中第 i 个个体，在第 j 个变量上的第 t 次的测量可表示为 Y_{ijt} ($i = 1, 2, \dots, N; j = 1, 2, \dots, M; t = 1, 2, \dots, T$)，与横断研究

相比多了一个时间的维度。

追踪数据的收集与横断研究相比所花费的时间和费用都要大得多，这主要是由于在追踪研究中不仅要测量多次，而且必须针对相同的被试。对相同的被试进行长时间的追踪，不论是在时间上还是实验组织上的投入都要比横断研究设计大得多。但是追踪研究与横断研究相比有其独特的不可替代的优点：(1) 可以就某一现象随时间的变化进行分析；(2) 可以对变量之间的因果关系做出合理的假设。

就追踪研究所关心的问题来讲主要有两个方面：(1) 描述个体 (Intraindividual) 发展趋势以及个体之间 (Interindividual) 发展趋势的差异；(2) 对被试的发展趋势及其原因进行解释，这一预测变量可以是不稳定的随时间变化的因素也可以是固定的个体特征因素。

2. 追踪研究设计的类型

从广义的角度来讲，追踪研究设计可以分为下列五类：(1) 同时性横断研究设计 (Simultaneous Cross-sectional Study Design)；(2) 趋势研究 (Trend Studies)；(3) 时间序列研究 (Time Series Studies)；(4) 干预 (Intervention) 研究；(5) 群组序列 (Cohort-Sequential) 设计或加速 (Accelerated) 设计。

(1) 同时性横断研究设计

同时性横断研究设计指的是对不同年龄组的样本进行同时性测量，可以用表 1 表示：

表 1 同时性横断研究设计

年龄组	样 本	时 间	观测变量
A_1	S_1	T_1	$X_1 X_2 X_3 \cdots X_M$
A_2	S_2	T_1	$X_1 X_2 X_3 \cdots X_M$
.	.	.	.
.	.	.	.
.	.	.	.
A_G	S_G	T_1	$X_1 X_2 X_3 \cdots X_M$

从表 1 中可以看出，在这一设计中共有 G 个不同的年龄群体，在同一时刻分别对这 G 个不同的年龄群体的 M 个变量进行测量。对于这一设计，我们往

往对不同年龄群体特征的“变化”感兴趣，即试图通过对不同年龄群体的测量来描述特征随年龄的变化。显然这一设计实质上并不是真正意义上的追踪研究。因为对于年龄 (Age) 和群组 (Cohort) 这两个概念用了相同的变量 (年龄) 来测量，所以该设计不能将年龄效应 (Age Effect) 和群组效应 (Cohort Effect) 区分开来，即无法了解不同年龄群体特征的差异究竟是由于年龄的变化而导致的，还是由于所选取群体的差异而造成的。同时，由于这一设计是对不同年龄群体的测量，因此永远不可能回答个体内发展 (Intra-individual Change) 趋势的问题。但是这一设计也有其明显的优点，即可以在很短的时间内 (一次测量) 对不同年龄的特征进行了解。关于这一设计更深的了解可以参考 Keeves (1997)、Rudinger 和 Wood (1990) 的著作。

(2) 趋势研究

趋势研究又称为重复横断研究 (Repeated Cross-sectional Studies)，指的是进行 T 次测量，但每次测量都是对同一个年龄群体中抽取的不同被试进行，可以用表 2 表示：

表 2 趋势研究设计

年龄组	样 本	时 间	观测变量
A_1	S_1	T_1	$X_1 X_2 X_3 \cdots X_M$
A_1	S_2	T_2	$X_1 X_2 X_3 \cdots X_M$
.	.	.	.
.	.	.	.
.	.	.	.
A_1	S_G	T_t	$X_1 X_2 X_3 \cdots X_M$

从表 2 中可以看出，一个相同的年龄群体分别在不同时刻 T_t 次测量，但是每次测量所用的样本都不相同。如预回答“是否教育质量逐年提高”这一问题，通常会对相同年龄组 (如初中毕业生) 的样本进行连续几年的测量，以考查学生能力的变化。这一设计由于每次测量采用的样本不同，因此实质上也不是真正意义上的追踪研究，与同时性横断研究类似，此类设计也不能将年龄效应 (Age Effect) 和群组效应 (Cohort Effect) 区分开来，也不可能回答个体内发展 (Intra-individual Change) 趋势的问题。

(3) 时间序列研究

时间序列研究是指对同一组样本进行多次测量，可以用表 3 表示：

表 3 时间序列研究设计

年龄组	样 本	时 间	观测变量
A_1	S_1	T_1	$X_1 X_2 X_3 \cdots X_M$
A_2	S_1	T_2	$X_1 X_2 X_3 \cdots X_M$
.	.	.	.
.	.	.	.
.	.	.	.
A_G	S_1	T_i	$X_1 X_2 X_3 \cdots X_M$

在时间序列研究中，相同的被试在多个连续时间点被多次观测，因此这一设计可以分析个体内发展 (Intra-individual Change) 的问题，同时可以就被试间 (Inter-individual) 变化的差异进行分析。对于时间序列分析，在社会学和心理学研究中，常用专门小组研究 (Panel Study) 这一术语。这两个术语的区别在于，时间序列设计往往指的是比较多的时间点的追踪研究设计，也就是说对于同一被试往往进行多次 (10 次、20 次等) 的测量，并且在时间序列设计中，不同被试的观测时间点往往没有相同的要求，测量之间的时间间隔也没有相等的限制；而专门小组研究往往指的是测试时间点比较少少的情况，如一般只有 3 次、5 次等少数几次的测量，另外作为社会科学研究领域常用的研究设计，通常情况要求不同被试有相同的测量次数，不同被试测量之间的时间间隔也尽量相同，但随着统计技术的发展，这些限定条件并不是专门小组研究设计必需的。关于这一设计更详细的讨论参见 Finkel (1995) 和 Keeves (1997) 的著作。

从上面关于时间序列设计或 Panel 设计的叙述可以看出，这一设计可以很好地回答个体发展趋势以及个体之间发展趋势差异的问题。如果这一设计采用的是预期的追踪研究 (Prospective Longitudinal Studies)，即被观测的个体是真正从 T_1 时刻开始观测，被追踪到 T_i 时刻，那么此研究设计就拥有比横断研究更明显的优势，尤其是其具有较高的内部效度 (Internal Validity)。从方法学的角度讲，这一研究设计具有其他追踪研究设计无法比拟的优点，但是往往所需时间长，费用大。