



格致方法·定量研究系列

吴晓刚 主编

因子分析：统计方法与应用问题

[美] 金在温 (Jae-On Kim) 著
查尔斯·W. 米勒 (Charles W. Mueller) 著
叶华 译



革新研究理念



丰富研究工具



最权威、最前沿的定量研究方法指南

格致出版社



上海人民出版社

20

格致方法·定量研究系列 吴晓刚 主编

因子分析：统计方法与应用问题

[美] 金在温 著
查尔斯·W.米勒
叶 华 译

SAGE Publications, Inc.

格致出版社 上海人民出版社

图书在版编目(CIP)数据

因子分析:统计方法与应用问题/(美)金在温(Jae-On Kim)
(美)米勒(Mueller, C. W.)著;叶华译. —上海:格致出版社;上海
人民出版社, 2012

(格致方法·定量研究系列)

ISBN 978-7-5432-2113-0

I. ①因… II. ①金… ②米… ③叶… III. ①因子分
析 IV. ①0212.1

中国版本图书馆 CIP 数据核字(2012)第 126356 号

责任编辑 王亚丽

格致方法·定量研究系列
因子分析:统计方法与应用问题

[美]金在温 著
查尔斯·W. 米勒
叶华 译

出版 世纪出版集团 www.ewen.cc 格致出版社
www.hibooks.cn 上海人民出版社
(200001 上海福建中路193号24层)



编辑部热线 021-63914988
市场部热线 021-63914081

发行 世纪出版集团发行中心
印刷 浙江临安曙光印务有限公司
开本 920×1168 毫米 1/32
印张 4.75
字数 88,000
版次 2012年7月第1版
印次 2012年7月第1次印刷
ISBN 978-7-5432-2113-0/C·65
定价 15.00 元

出版说明

由香港科技大学社会科学部吴晓刚教授主编的“格致方法·定量研究系列”丛书,精选了世界著名的 SAGE 出版社定量社会科学研究丛书中的 35 种,翻译成中文,集结成八册,于 2011 年出版。这八册书分别是:《线性回归分析基础》、《高级回归分析》、《广义线性模型》、《纵贯数据分析》、《因果关系模型》、《社会科学中的数理基础及应用》、《数据分析方法五种》和《列表数据分析》。这套丛书自出版以来,受到广大读者特别是年轻一代社会科学工作者的欢迎,他们针对丛书的内容和翻译都提出了很多中肯的建议。我们对此表示衷心的感谢。

基于读者的热烈反馈,同时也为了向广大读者提供更多的方便和选择,我们将该丛书以单行本的形式再次出版发行。在此过程中,主编和译者对已出版的书做了必要的修订和校正,还新增加了两个品种。此外,曾东林、许多多、范新光、李忠路协助主编参加了校订。今后我们将继续与 SAGE 出版社合作,陆续推出新的品种。我们希望本丛书单行本的出版能为推动国内社会科学定量研究的教学和研究作出一点贡献。

总序

往事如烟，光阴如梭。转眼间，出国已然十年有余。1996年赴美留学，最初选择的主攻方向是比较历史社会学，研究的兴趣是中国的制度变迁问题。以我以前在国内所受的学术训练，基本是看不上定量研究的。一方面，我们倾向于研究大问题，不喜欢纠缠于细枝末节。国内一位老师的话给我的印象很深，大致是说：如果你看到一堵墙就要倒了，还用得着纠缠于那堵墙的倾斜角度究竟是几度吗？所以，很多研究都是大而化之，只要说得通即可。另一方面，国内（十年前）的统计教学，总的来说与社会研究中的实际问题是相脱节的。结果是，很多原先对定量研究感兴趣的学生在学完统计之后，依旧无从下手，逐渐失去了对定量研究的兴趣。

我所就读的美国加州大学洛杉矶分校社会学系，在定量研究方面有着系统的博士训练课程。不论研究兴趣是定量还是定性的，所有的研究生第一年的头两个学期必须修两门中级统计课，最后一个学期的系列课程则是简单介绍线性回归以外的其他统计方法，是选修课。希望进一步学习定量研

究方法的可以在第二年修读另外一个三学期的系列课程,其中头两门课叫“调查数据分析”,第三门叫“研究设计”。除此以外,还有如“定类数据分析”、“人口学方法与技术”、“事件史分析”、“多层线性模型”等专门课程供学生选修。该学校的统计系、心理系、教育系、经济系也有一批蜚声国际的学者,提供不同的、更加专业化的课程供学生选修。2001年完成博士学业之后,我又受安德鲁·梅隆基金会资助,在世界定量社会科学研究的重镇密歇根大学从事两年的博士后研究,其间旁听谢宇教授为博士生讲授的统计课程,并参与该校社会研究院(Institute for Social Research)定量社会研究方法项目的一些讨论会,受益良多。

2003年,我赴港工作,在香港科技大学社会科学部,教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来,非常重视社会科学研究方法论的训练。我开设的第一门课“社会科学里的统计学”(Statistics for Social Science)为所有研究型硕士生和博士生的必修课,而第二门课“社会科学中的定量分析”为博士生的必修课(事实上,大部分硕士生修完第一门课后都会继续选修第二门课)。我在讲授这两门课的时候,根据社会科学研究生的数理基础比较薄弱的特点,尽量避免复杂的数学公式推导,而用具体的例子,结合语言和图形,帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题上,即社会研究者主要为定量统计方法的“消费者”而非“生产者”。作为“消费者”,学完这些课程后,我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章;另一方面,也能在自己的研究中运用这些成熟的

方法论技术。

上述两门课的内容,尽管在线性回归模型的内容上有少量重复,但各有侧重。“社会科学里的统计学”(Statistics for Social Science)从介绍最基本的社会研究方法论和统计学原理开始,到多元线性回归模型结束,内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型,以及线性回归模型的假设和模型诊断。“社会科学中的定量分析”则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法,将重点放在因变量为定类数据的分析模型上,包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型,包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生,一直鼓励和支持我将两门课的讲稿结集出版,并帮助我将原来的英文课程讲稿译成了中文。但是,由于种种原因,这两本书拖了四年多还没有完成。世界著名的出版社 SAGE 的“定量社会科学研究”丛书闻名遐迩,每本书都写得通俗易懂。中山大学马骏教授向格致出版社何元龙社长推荐了这套书,当格致出版社向我提出从这套丛书中精选一批翻译,以飨中文读者时,我非常支持这个想法,因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种

语言的精准把握能力,还要有对实质内容有较深的理解能力,而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容,只有语言能力是远远不能胜任的。在短短的一年时间里,我们组织了来自中国内地及港台地区的二十几位研究生参与了这项工程,他们目前大部分是香港科技大学的硕士和博士研究生,受过严格的社会科学统计方法的训练,也有来自美国等地对定量研究感兴趣的博士研究生。他们是:

香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛,硕士研究生贺光烨、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊,应用社会经济研究中心研究员李俊秀;香港大学教育学院博士研究生洪岩璧;北京大学社会学系博士研究生李丁、赵亮员;中国人民大学人口学系讲师巫锡炜;中国台湾“中央”研究院社会学所助理研究员林宗弘;南京师范大学心理学系副教授陈陈;美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛;美国加州大学洛杉矶分校社会学系博士研究生宋曦。

关于每一位译者的学术背景,书中相关部分都有简单的介绍。尽管每本书因本身内容和译者的行文风格有所差异,校对也未免挂一漏万,术语的标准译法方面还有很大的改进空间,但所有的参与者都做了最大的努力,在繁忙的学习和研究之余,在不到一年的时间内,完成了三十五本书、超过百万字的翻译任务。李骏、叶华、张卓妮、贺光烨、宋曦、於嘉、郑冰岛和林宗弘除了承担自己的翻译任务之外,还在初稿校对方面付出了大量的劳动。香港科技大学霍英东南沙研究院的工作人员曾东林,协助我通读了全稿,在此

我也致以诚挚的谢意。有些作者,如香港科技大学黄善国教授、美国约翰·霍普金斯大学郝令昕教授,也参与了审校工作。

我们希望本丛书的出版,能为建设国内社会科学定量研究的扎实学风作出一点贡献。

吴晓刚

于香港九龙清水湾

序

本书扩展了 Jae-On Kim 和 Charles W. Mueller 的专著《因子分析导论：它是什么以及如何运用》(*Introduction to factor analysis: what it is and how to do it*)。后者着重对因子分析技术基础的介绍，关心的是“为什么要进行因子分析”这个问题，并且希望读者弄清楚运用这种方法涉及哪些假定。

对阅读使用因子分析方法的书或文章，或者用自己的数据来简单试用因子分析方法的读者来说，本书将把他们引导到另一个阶段。在书中，Kim 和 Mueller 用更为明确的数据分析例子，更为详细地介绍了因子分析的不同方法，以及它们在何种情况下最有用。验证性和探索性因子分析的差别在这里会比在《因子分析导论：它是什么以及如何运用》中讨论得更深入，同样会进行讨论的还有因子旋转的各种标准。特别值得一提的是对不同形式的斜交旋转的讨论，以及如何解释从这些分析中得到的各项系数。Kim 和 Mueller 也回答了在探索性因子分析中抽取出的因子数量的问题，讨论了在验证性因子分析中检验假设的方法。对于这些研究中抽取

出来的尺度得分,他们也讨论了分析它们时会遇到的问题。本书会提供一份术语表,同时还会解答使用因子分析技术的人经常提出的一些问题。读者可能会被其中一些问题的回答所困扰,但只要加以注意,分析者就能避免错误的推论。本书尽管有一些矩阵代数的应用,但其数理内容基本上是自明的。如果读者不能很好地理解矩阵应用,建议参考作者之前的那本专著。

在经济学中,因子分析被用于抽取一组互不相关的变量来做进一步分析,因为此时使用高度相关的变量在回归分析中可能会产生误导性的结果。政治学家比较不同国家在一系列政治和社会经济变量上的特点,尝试找出什么是区分国家的最重要特征(例如富裕程度和国家大小)^①;社会学家通过研究相互联系最多的人(而不是与其他人联系)来确定“朋友群体”。心理学家和教育学者使用这种技术,研究人们对不同的“激励”有什么感受,并把它们区分为不同的反应系列,例如语言的不同要素是相互联系的。

正如作者所指出的,本书不可能涵盖因子分析的所有方面,因为在这个领域里经常会有新的发展。然而,如果读者能够建立起系统性的知识,知道如何运用这种技术,以及我们实际上在无形中至少做了什么假定,那么本书就已经达到其目的了。

E. M. 尤斯拉纳

^① 参见 Rudolph J. Rummel, *The Dimensions of Nations* (Beverly Hills, CA: Sage Publications, 1972)。

目 录

序	1
第 1 章 导论	1
第 1 节 对因子分析基础的回顾	4
第 2 节 本书涵盖的基本方案和方法	7
第 2 章 抽取初始因子的方法	11
第 1 节 主成分、特征值和向量	14
第 2 节 公因子模型的变型	22
第 3 节 最小二乘法	23
第 4 节 基于最大似然法的解决方案	25
第 5 节 Alpha 因子分析法	29
第 6 节 映像分析	31
第 3 章 旋转的方法	33
第 1 节 图像旋转、简单结构和参照轴	36
第 2 节 正交旋转的方法：四次方最大法、最大方差法和均等变化法	40

第 3 节	斜交旋转的方法	44
第 4 节	旋转至目标矩阵	48
第 4 章	再论因子数量的问题	51
第 1 节	检验显著性	53
第 2 节	通过特征值进行设定	55
第 3 节	实质重要性的准则	57
第 4 节	碎石检验	58
第 5 节	可解释性和恒定性的原则	59
第 5 章	验证性因子分析简介	61
第 1 节	因子分析性模型被经验验证的程度	63
第 2 节	模型所暗示的经验限制的数量	67
第 3 节	另一种概念的经验证实:抽样准确性	73
第 4 节	验证性因子分析	76
第 6 章	建立因子尺度	83
第 1 节	因子尺度的不确定性	85
第 2 节	抽样变异性和模型拟合的不同标准	91
第 3 节	多个公因子和更复杂的情况	94
第 4 节	基于因子的尺度	97
第 5 节	成分得分	100

第 7 章 对常见问题的简单回答	101
第 1 节 与变量的性质以及它们的测量有关的问题	102
第 2 节 与使用相关或协方差矩阵有关的问题	106
第 3 节 与显著性检验和因子分析结果稳定性有关的问题	108
第 4 节 其他各种统计问题	109
第 5 节 与书、期刊和计算机程序有关的问题	111
注释	112
参考文献	113
术语表	120
译名对照表	125

第 **1** 章

导 论

因子分析的概念性基础简单易懂。然而,一些原因可能使得在实际运用中掌握这种方法会比较困难。首先,一般而言,理解统计估计的原则相比理解背后的概念模型,前者对数理复杂性的要求更高。第二,文献中提出了许多提取因子的方法,即使是相对简单的计算机程序,也可能在分析的每一步给出许多选项。这些复杂性会让初学者很难适应,即使专家也会感到不习惯。第三,我们平时遇到的实际研究问题几乎总是比因子分析模型所假定的情况要复杂。例如,(1)可能会有一些或者全部变量的测量层次都没有达到因子分析的要求;(2)对一些人所用的数据来说,模型的某些假定是不现实的,例如测量误差之间不相关;(3)我们可能会有一些次要的因子,对它们的识别并不是我们考虑的重点,但是它们的存在会影响我们对主要的公因子的识别。问题的关键在于,研究者最终必须自行作出一些与统计无关的决定。幸运的是,正如我们将会讲述的那样,这些困难是可以克服的。

正如我们在《因子分析导论:它是什么以及如何运用》所指出的,研究者或多或少地被迫依赖于已有的计算机程序进行实际分析,这些程序通常提供了默认的标准选项,使用者会依赖它们,直到必须进行一些改动。更重要的是,随着研

究者逐渐熟悉因子分析的多种选项,他们会发现大多数的差别在很大程度上都只是表面的。实际上,有一些共同的线索贯穿于这些差别。更重要的是,研究者会发现,对相同的数据运用不同的方法和标准,在大多数实际应用中其结果是一致的。简言之,读者没有必要立刻学会和使用所有的选项。然而,重要的是使用者要知道进行因子分析时最重要的差别和选项,并且使用者从一开始就要明白,对大多数问题我们没有绝对唯一的(或者最好的)解决方案。

本书假定读者对因子分析的概念基础有基本的认识,那些内容在上一卷书中已经介绍过。我们也希望读者明白以下两种内在的不确定性之间的区别,即从观察到的协方差结构推论潜在的因果结构的不确定性^①(逻辑问题),以及从估计的样本统计量推论总体参数的不确定性(统计问题)。尽管这两个问题在解决实际的因子分析问题时有内在的联系,但在头脑中要分清这两者之间在概念上的差别。在讨论统计方法与实际问题之前,我们有必要简单复述一下上一卷书所介绍的内容。

① 也就是变量之间有相关关系并不意味着有因果关系。——译者注