SCIENCE & CULTURE

科文医学文库

多变量分析

临床实用指南

Multivariable Analysis A Practical Guide for Clinicians

(美) 米歇尔·H·凯茨 (Mitchell H. Katz) 著
姚晨 刘玉秀 陈峰 顾海燕 译 苏炳华 校

中国科学技术出版社
科文(香港)出版有限公司
Science & Culture Publishing House LTD.(H.K.)

科文医学文库

多变量分析

临床实用指南

MULTIVARIABLE ANALYSIS

A PRACTICAL GUIDE FOR CLINICIANS

（美）米歇尔·H·凯茨 （Mitchell H. Katz）著

姚 晨 刘玉秀 陈 峰 顾海燕 译

苏炳华 校

中国科学技术出版社

科文（香港）出版有限公司

Science & Culture Publishing House LTD .(H.K.)

著作权合同登记：图字 01－2000－2696 号

图书在版编目（CIP）数据

多变量分析临床实用指南/（美）凯茨（Katz，M.H.）著；姚晨等译 .—北京：中国科学技术出版社，2000.9
ISBN 7－5046－2947－2

Ⅰ.多… Ⅱ.①凯…②姚… Ⅲ.多变量－统计分析－临床应用 Ⅳ.R4

中国版本图书馆 CIP 数据核字（2000）第 47637 号

中国科学技术出版社出版

新华书店北京发行所发行

经售：中国科学技术出版社

（北京海淀区白石桥路 32 号 邮政编码：100081）

北京科文剑桥图书有限公司

（北京市安定门外大街 208 号 三利大厦 电话：64203023）

北京民族印刷厂印刷

*

开本：850 毫米×1168 毫米 1/32 印张：10.75 字数：230 千字

2000 年 9 月第 1 版 2000 年 9 月第 1 次印刷

印数：1－3000 册 定价：26.00 元

译 者 序

我在解放军总医院（军医进修学院）从事临床医学研究生的医学统计学教学工作十多年，先后为硕士、博士研究生开设了多变量统计分析课程。多变量统计分析的应用，使得大量临床医学实践中所积累的资料得到进一步深入分析，特别为病因学、治疗学等方面提供了科学的分析手段，因此，深受临床医生的青睐。但在实际应用中，由于多变量分析的应用有其相应的条件、使用技巧、建模策略和结果解释方法，加之临床随访资料经常有缺失数据的情况，因此，临床医生在应用多变量分析方法解决临床实际问题时，经常会遇到诸如：如何根据资料的性质选择多变量分析方法？进行多变量分析需要多少样本量？如何对非数值指标进行数量化？为什么重要的因素不包含在多变量模型中？如何考虑交互作用？计算机输出的结果如何解释？研究论文中如何表达分析结果？……虽然不乏多元统计分析的教材，但由于这些教材着重基本概念的论述、模型的估计等，免不了有许多数学公式或推导，而对应用技巧以及实际应用中可能出现的问题论述不多或不深入。很多临床医生虽然学了多元分析，但真正处理实际问题时又觉举步维艰，无从下手。正当我为寻找有关多变量统计分析通俗易懂的教材之时，恰逢由Mitchell H. Katz撰写的《多变量分析临床实用指南》出版并在1999年德国法兰克福国际书展中展出。我非常感谢科文（香港）出版有限公司购得该书的翻译出版发行权，并使我们现在

有机会向临床医生推荐此书。作为应用多元统计分析方法的参考书，它用问答的形式深入浅出地解答了临床医生在应用多变量分析技术时所遇到的许多问题，并结合大量成功应用多元分析的范例，采用更贴近临床医生的语言，展现了多元统计分析在医学研究应用中的博大与精妙。

本书的翻译得益于有南京军区南京总医院刘玉秀副主任医师、南通医学院医学统计学教研室陈峰教授和顾海雁讲师的参加，使得全书能在较短时间内与读者见面，译文亦能保持原著风格，更适合临床医生阅读。在翻译本书过程中，我们发现原著构思精巧，内容编排紧凑，语言非常生动，例子精心选择且切合实际。某些论点的提法、难点的解释和优美的用词常令译者叹为观止，回味无穷。如果读者从本书的译文很少有这种体会，那只能怪罪于译者的翻译水平，对此我们只能表示歉意，并敬请广大读者和同事斧正。

翻译本身也是学习和思考的过程，对有些问题的认识也更深透了，获益匪浅。更重要的是，受其启迪，使我们对以前的统计学教学思路和方法作了一次深刻的反思。相信以后的医学统计学教学将是一个崭新的面貌。

最后，非常感谢我尊敬的老师，上海第二医科大学生物统计学教研室苏炳华教授在百忙之中抽空为此书进行了审校，并提出了许多宝贵意见，使本书增色许多。

姚　晨

2000 年 6 月 12 日于北京

前　言

身为研究者和研究的教授者，我为缺少有关多变量分析通俗易懂的教科书而感到不安，故而决定撰写一本这方面的统计书。现有的大多数教科书对初学研究者太难。尽管有一些基础生物统计书能为非生物统计工作者理解，但其中多未包括多变量分析。我要写出一种我第一次使用多变量模型时追寻的那样的书，也是一种能提供为临床研究者第一次实施要求用多变量分析的研究计划的那样的书。但愿本书如此。

本书容易理解，不太需要数学知识，没有推导，公式极少。重点放在模型可以干什么的概念性说明以及如何对输出结果进行解释上。采用问答式提出和解答大多数的实际问题。经验表明，许多统计书并未回答诸如“如果系数为负号说明什么”（见 9.3 节）的基本问题。我把许多摘要表格、经验方法以及分析技巧贯穿于书中。希望阅读本书后，大家能在没有生物统计学家的帮助下从事和解释多数的多变量分析。对于需要生物统计学家帮助的复杂的分析，力求提供充足的内容以说明需要做什么，尽管最后还是要人帮助。

我引入了大量的医学文献例子，以阐明本书的重点。不依靠一两种数据集而选择此种策略有几方面的原因。最主要的是，想让读者感受到我曾感受到的多变量

分析大量且变化多端应用的兴奋。其次，想让研究者领略到达到同一目标有许许多多的不同途径。那些认为某一分析方法总是比另一方法好（例如总是用后退剔除法而不用前进选择法做逐步回归）的研究者也许反对这种做法。当然，有理由偏爱一种合意的方法。若效应确切而强大，用不同的方法将得到相似的结果。事实上，某种效应如果用一种方法分析时存在，而用另外的虽然也是可接受的不同方法分析时，效应却不见了，则其可能是假象或很弱。

没有在真空中写出来的书。我对几本我已从中学到许多的统计书的作者们深感谢忱。特别是，S.Glantz 著《生物统计学初步》（第 4 版，McGraw - Hill，1997）是我学习的第一本生物统计书（当时为第 1 版），同时也是本书参考的优秀用书。S.Glantz 和 B.Slinker 著《应用回归与方差分析初步》（McGraw - Hill，1990）是第二本优秀的生物统计书。此书包括一些与本书相同的材料，但其对方差分析的解释更为详细。D.W.Hosmer 和 S.Lemeshow 著《应用 Logistic 回归》（Wiley，1989）是一经典的读本，有关于该种回归的技术比本书更加深入。A.Feinstein 著《多变量分析》（耶鲁大学出版社，1996）包括了不少与本书相同的背景，但却采用了不同的和互有补充的方式。其他重要的参考读物置为脚注。

我最感激的是我的老师、学生以及同事们。Chaya Piotrkowski教我如何使用软件包和怎样在计算机卡片上

用打孔机打孔。Daniel Singer 带我开始临床研究并向我提供 Logistic 回归的介绍。Warren Browner、Steven Cummings、Deborah Grady、Stephen Hulley 以及 Thomas Newman 教我怎样思考和教授生物统计学。Walter Hauck 帮我战胜比例风险分析。以前的学生 Alan Chan、Karla Kerlikowske 和 Anthony So 挑起我教会他们这些技术的兴趣，我在这一过程中得到更好地领会。几年来，在旧金山加利福尼亚大学从事临床研究项目的学生们通过其富有见识的问题和观察为本书做出贡献。在旧金山公共卫生部我非常有幸地与出色而富有创新的临床研究者们一起工作，他们包括 Tomas Aragon、Susan Buchbinder、Jan Gurley、Nancy Hessol、Willi McFarland、Rani Marx 和 Sandy Schwarcz。他们教会我许多东西。我要特别感谢 Eric Vittinghoff，他仔细地审阅原稿并提出许多重要改进。如果有什么错误，唯我是问。我的朋友和同事 Walter Mebane 从上大学起就耐心教我统计学和计算机。我的朋友 David French 和 Perri Klass 在整个过程给我快乐。

书中插图由生物医学艺术学会的 Ward Ruth 精心绘制。本书告成，我要对剑桥大学出版社的编辑 Jo – Ann Strangis 和其工作人员的支持表示感谢。

加利福尼亚州旧金山市

1998 年 10 月

目　录

1
绪论

1.1 为什么做多变量分析?

我们生活在一个多元化的世界里。大多数事件，不管是医学的、政治的、社会的或个体的，均有多种原因。这些原因是相互关联的。多变量分析[①]是一种用于制定不同原因对某一事件或结果相对作用大小的统计学工具。

→ 定 义
多变量分析是一种用于制定不同原因对某一事件或结果相对作用大小的统计学工具。

临床研究人员尤其需要多变量分析，因为大多数疾病都有多种病因，而且预后是由多种因素决定的。即使对那些已知是由一病原体导致的感染性疾病，也有许多因素影响受感染个体是否

①术语“多变量分析”和“多因素分析”经常可被交替使用。严格说来，多变量分析是指同时预测多个结果。因为本书涉及的技术是用多变量预测单一结果，我喜欢用常用术语“多因素分析”。

发病，这些因素包括病原体的特征（如株的毒性）、感染的途径（如呼吸道）、感染的程度（例如 Innoculum 大小）和宿主反应（如免疫防御）。

多变量分析将使我们找出危险因素的多面性本质，以及它们对结果的相对作用大小。例如；观察流行病学使我们知道有一系列危险因素与早期死亡有关，如吸烟、久坐生活方式、肥胖、高胆固醇、高血压。我并没有讲这些因素导致早期死亡，统计学本身并不能证明一个危险因素和结果之间的因果关系[①]。因果关系是建立在生物学和严格研究设计的基础上的，如随机对照试验将消除潜在偏倚的来源。

通过观察研究来鉴定与早期死亡有关的危险因素是相当重要的，因为你不

---

①虽然在本书中我交替使用了术语“有关”和“相关”。同样，我也交替使用术语“危险因素”和“自变量”以及术语“结果”和“因变量”。虽然更多使用术语“预测”是指自变量与结果的关联性，该术语意指因果关系，当我们决定一个模型预测个体结果好坏时，我喜欢保留它。

可能随机地将人分配至导致早期死亡的条件下，如吸烟、久座的生活方式或肥胖，另外，这些因素有可能同时发生；吸烟者可能运动较少而且极可能肥胖。多变量分析是怎样将这些因素的每一个独立的作用分离开来？让我们看看运动这一例子，许多研究表明，长期运动人的寿命比久坐型生活方式人的寿命长。但是，如果运动者寿命长的唯一原因是不吸烟，可能低脂肪饮食使胆固醇较低，那么运动将不能改变一个人的寿命。

Aerobics 中心的纵向研究解决了这一重要的问题[①]，它们评价了 2 534 名男性和 7 084 名女性中运动和死亡率的关系。所有参加者在 1970 ~ 1989 年间接受了基线检查，这些检查包括体检、实验室检测及踏车试验评价身体适宜运动量。男性参加者被追踪观察平均 8.4 年，女性平均 7.5 年。

---

①Blair, S.N., Kampert, J.B., Kohl, H.W., et al. "Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all - cause mortality in men and women." JAMA 1996; 276: 205 - 210