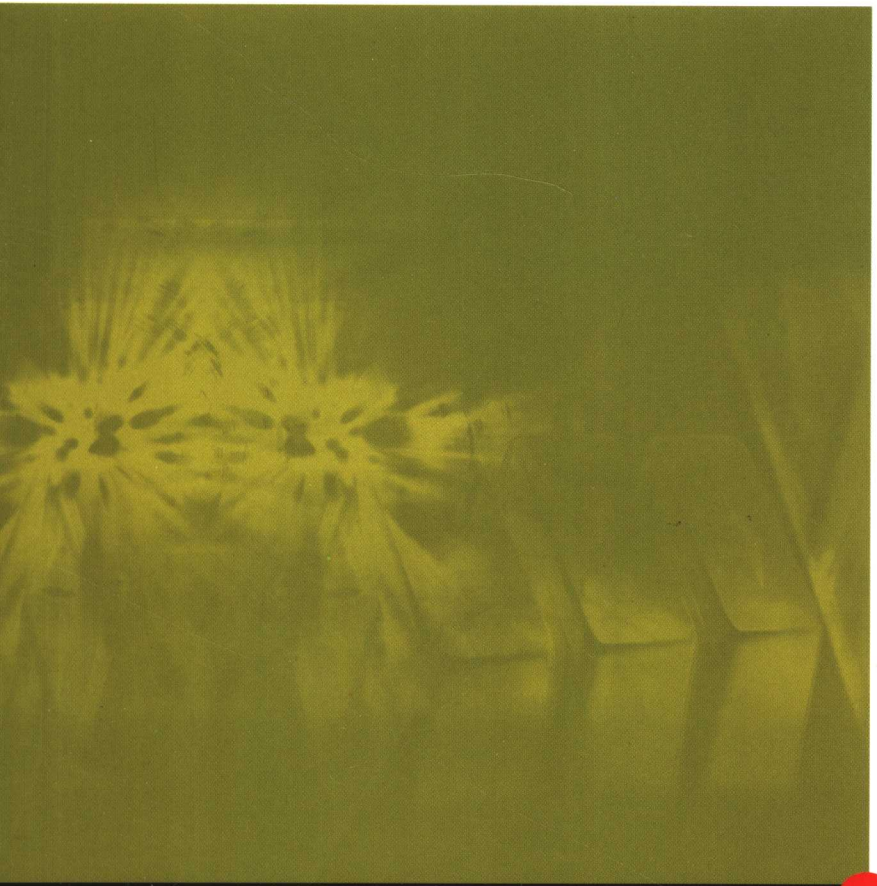
马克威软件系列丛书

马克威软件与当代数据分析

MARKWAY



主 编 黄 晖
主 审 王吉利
魏振军

中国统计出版社
China Statistics Press

马克威软件与当代数据分析

马克威软件系列丛书

马克威软件与当代数据分析

MARKWAY

主编 黄 晖

主审 王吉利 魏振军

中国统计出版社
China Statistics Press

(京)新登字041号

图书在版编目(CIP)数据

马克威软件与当代数据分析/黄晖主编.
—北京:中国统计出版社,2006.1
ISBN 7-5037-4848-6

Ⅰ.马…
Ⅱ.黄…
Ⅲ.统计分析—应用软件
Ⅳ.C819

中国版本图书馆CIP数据核字(2005)第158969号

作　　者/黄　晖
责任编辑/吕　军
封面设计/艺编广告
出版发行/中国统计出版社
通信地址/北京市西城区月坛南街75号　邮政编码/100826
办公地址/北京市丰台区西三环南路甲6号
电　　话/(010)63459084,63266600-22500(发行部)
印　　刷/河北天普润印刷厂
经　　销/新华书店
开　　本/787×1092mm　1/18
字　　数/560千字
印　　张/33.125
印　　数/1-5000册
版　　别/2006年5月第1版
版　　次/2006年5月第1次印刷
书　　号/ISBN 7-5037-4848-6/C·2097
定　　价/48.00元

《马克威软件与当代数据分析》编委会

序

中国的统计工作正在受到国内外越来越多人的关注。统计工作的性质也正在经历巨大的转变，由简单的数据采集和报表上传逐步向数据分析和决策支持发展。然而，长期以来我们却缺乏一个真正属于自己的统计分析工具，以及基于这种工具的教材。上海天律信息技术有限公司研制出了我国第一套大型统计分析软件“马克威分析系统”，填补了这项空白。现在，由马克威软件的创始人、留美博士黄晖教授编写的教材《马克威软件与当代数据分析》也将问世。这是我国统计学界一件令人高兴的事情，对我国统计工作将起到积极的推动作用。

《马克威软件与当代数据分析》一书全面介绍了当今数据挖掘和数据分析领域的方法和原理。它将理论、方法、应用案例和软件操作溶为一体，为数据分析工作者提供了极为实用的一站式导航。该书分为五个部分，总共二十九章。该书以软件介绍、基础统计、高级统计、数据挖掘、统计制图为主线，逐步介绍了现代数据分析的各个方面，条理清晰，陈述明了，是数据分析工作者的极好工具。

概括而论，《马克威软件与当代数据分析》一书有以下特点：

第一，该书涵盖了现代数据分析从数据采集和预处理到数据建模和结果展示的全过程，系统而全面地介绍了各种算法模块，尤其是对最新的理论和方法作了介绍。在数据挖掘方面，该书不仅讲解了神经网络、决策树、关联规则、模糊聚类、粗糙集、孤立点分析等比较经典的方法，还详细介绍了 RBF 神经网络、贝叶斯网络、支持向量机等最新方法。在统计分析方面，除了经典的回归分析、主成分分析、因子分析、生存分析、聚类分析、方差分析和时间序列分析以外，还讲解

了新兴的协整分析、联立方程、向量自回归等方法，并给出了详尽的实例，以帮助读者进行理解和应用。

第二，该书将数据分析知识和软件使用知识有机结合，理论讲解和具体案例有机结合，使用户不仅学会操作软件，而且掌握其背后蕴涵的统计学和数据挖掘知识。作者在指导读者熟悉软件操作的同时，结合相关的理论知识，对具体的数据分析案例进行了深入细致的讲解。通过本书，读者可以掌握马克威软件的使用，又可以获得相关统计分析和数据挖掘方面的理论知识，同时可以深入了解到数据分析和建模的方法及其应用。

第三，软件的优势有助于对教材的理解。以往我们依靠国外的软件从事数据分析工作，外文的操作界面和编程使得我们的使用极为不便。马克威分析系统是一套完全自主知识产权的、全中文界面的民族产品，具有全中文的友好界面、符合中国人的使用习惯和操作方式，具有直观、简单等特点。《马克威软件与当代数据分析》一书根据软件的这些特点，致力于将复杂难懂的统计和挖掘方法简易化和大众化，使读者能快速高效地进入数据分析的王国。

简言之，数据处理与数据分析是信息时代许多机构的核心工作之一，在各行各业几乎所有的科学决策都离不开大规模的数据处理和分析。目前在国内的零售、保险、银行、通信、离散制造、政府、医疗、分销、流程制造、教育等行业，其应用在迅猛增长。因此基于统计和挖掘的数据分析技术已经成为商业智能和企业决策支持的灵魂，马克威分析系统的广泛使用及其教材的出版，对于促进中国数据分析和商务智能方面的发展将会产生重要的作用和影响。

沈青华

2006 年 4 月于北京

前　言

《马克威软件与当代数据分析》一书终于问世了。这是中国人第一次以自己的软件为依托来编写数据分析方面的教材。对于正在中国迅速崛起的数据分析行业来说，我们希望这本书能起到及时雨的作用。

数据分析作为一个新的知识领域正在美欧等发达国家快速发展。它的实质是开辟人们获得知识的新途径。在人类认知真理的过程中，经历了绝对真理、相对真理和概率真理的变化。当我们认知物质世界时，研究对象相对静态，比较易于控制，所以物理学可以得出不易改变的定律。当我们认知生物世界时，生命在不停地变化着，传统的科学试验方法仍然可以使用，但由于涉及的变量越来越多，所以遇到的难题也越来越多，如癌症、艾滋病等等。当我们研究社会世界时，遇到难以控制的变量则更多，因而更加难以用传统的科学试验方法来获得知识。所以，基于统计学和数据挖掘的研究方法成为主要手段。我们关于人的社会行为、经济行为的知识也就变成一种概率知识。这就是为什么近年来获得诺贝尔奖的经济学家们都只能依靠统计学和数据分析的方法来寻找规律。

从20世纪90年代起，在美欧等国家开始流行各种分析师类的职业，如业务分析师(Business Analyst)、系统分析师(Systems Analyst)、数据分析师(Data Analyst)等等。其中，数据分析师直到现在仍然兴盛不衰，而且大有扩展之势。一个简单的原因就是社会越发达，人们对数据的依赖就越多。无论是政府决策还是企业运营，无论是科学研究还是舆论宣传，都离不开数据这个基础。对数据的依赖必然导致对数据分析工具的需求。没有科学的分析方法，没有好用的分析工具，数据只能是负担和累赘。把数据变成知识，变成思想和决策，这正是数据分析的作用和价值。

《马克威软件与当代数据分析》一书正是基于以上理念编写而成。本书是

以马克威分析系统第3.0版相配套的，它的特点在于将数据分析和软件使用紧密结合，以案例分析和软件操作为主线，将数据分析的各个方面连贯起来，便于学习、便于理解、便于操作。在内容方面，本书将统计分析和数据挖掘融为一体，统一在数据分析的旗帜下，为真正从事数据分析的工作者提供一站式服务。同时，本书还吸纳了计量经济学方面的许多特殊算法，如联立方程、向量自回归、协整分析等等。在挖掘方面，本书收入了支持向量机、RBF神经网络和贝叶斯网络等前沿算法。

马克威分析系统学员版可以从以下网站下载：上海天律信息技术有限公司 http://www.tenly.com/，中国统计教育培训网 http://edu.stats.gov.cn/。

本书成立了编委会，编委们来自国家统计局教育中心、计算中心、出版社、研究所、中国人民大学、台湾辅仁大学、美国雪城大学等单位，在编写和出版的过程中得到了他们多方面的帮助，在此一并表示感谢，并期待他们一如既往的支持和鼓励。参加本书的编写人员包括：郭辉明编写第二、三章，孙慧编写第四、五章，林跃跃编写第六、七章，姜勤德编写第八、九章。解放军信息工程大学的魏振军教授通读全书，对本书提供了细致全面的核对，为软件的修改屡次奔波于上海、北京等地，提出了许多宝贵的建议。国家统计局原总工程师沈青华对此书的出版给予了热情支持，并亲自为本书作序。

最后要十分感谢原国家统计局统计教育中心的王吉利主任，是他远见卓识的倡导国产统计分析软件的应用对于全面提高我国统计分析质量起到的重要作用，也是他坚持不懈的努力促成了本书的编写和出版，并主持、参加了全书的审定工作，我为在中国遇见这样的领导而感到欣慰。

鉴于时间仓促等诸多原因，本书难免有疏漏与错误之处，恳请读者指正！

黄　晖

2006年4月于上海

目 录

第一部分 认识马克威分析系统

第二部分 基础统计

第三部分 高级统计

第四部分 数据挖掘

第五部分 统计制图与电子表格