

第2版

Advanced Medical Statistics

现代医学统计学

主编 方积乾 陆盈



现代医学统计学

Advanced Medical Statistics

第2版

主编

方积乾 陆 盈

副主编

金 华 田 鲁

人民卫生出版社

图书在版编目 (CIP) 数据

现代医学统计学/方积乾, 陆盈主编. —2 版.
—北京: 人民卫生出版社, 2015
ISBN 978-7-117-20285-5

I. ①现… II. ①方…②陆… III. ①医学统计-统计学 IV. ①R195.1

中国版本图书馆 CIP 数据核字(2015)第 035064 号

人卫社官网	www.pmph.com	出版物查询, 在线购书
人卫医学网	www.ipmph.com	医学考试辅导, 医学数据库服务, 医学教育资源, 大众健康资讯

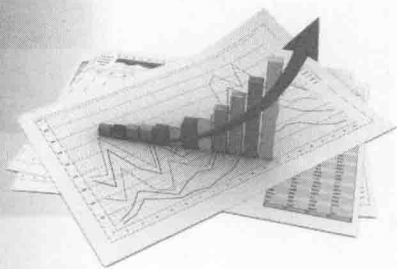
版权所有, 侵权必究!

现代医学统计学
第 2 版

主 编: 方积乾 陆 盈
出版发行: 人民卫生出版社 (中继线 010-59780011)
地 址: 北京市朝阳区潘家园南里 19 号
邮 编: 100021
E - mail: pmph@pmph.com
购书热线: 010-59787592 010-59787584 010-65264830
印 刷: 北京铭成印刷有限公司
经 销: 新华书店
开 本: 787×1092 1/16 印张: 60 插页: 2
字 数: 1460 千字
版 次: 2002 年 4 月第 1 版 2015 年 6 月第 2 版
2015 年 6 月第 2 版第 1 次印刷 (总第 3 次印刷)
标准书号: ISBN 978-7-117-20285-5/R · 20286
定 价: 160.00 元
打击盗版举报电话: 010-59787491 E-mail: WQ@pmph.com
(凡属印装质量问题请与本社市场营销中心联系退换)

集學衍之精華
作科學的貢獻

蔣希環誌



主编及副主编简介

方 积乾, 2009 年获国家教学名师奖, 2010 年获中华预防医学突出贡献奖。曾任中国卫生信息学会副会长, 国际生物统计学会中国组负责人。1961 年获复旦大学数学学士学位, 1982—1985 年在加利福尼亚大学伯克利分校获生物统计学博士。1985 年在北京医科大学由讲师直接提升为教授。1991 年起任中山大学公共卫生学院医学统计系主任、教授、博士生导师。曾在英国肯特大学、澳大利亚国立大学讲学, 1993—2009 年任香港中文大学兼职教授。曾主持国家和省部级课题 19 项; 获省部级科研奖励 14 项、教学奖励 2 项。主编专著 6 部和国家规划教材 7 种, 其中, 专著《Advanced Medical Statistics》和研究生教材《Medical Statistics and Computer Experiments》以中英文同时出版; 主编本科生教材《卫生统计学》第 5、6、7 版, 第 5 版获全国医学教材一等奖。他主持的《医学统计学》课程, 2008 年被评为国家级精品课程, 2010 年被评为国家级双语教学示范课程。

陆 盈, 美国斯坦福大学医学院健康研究与政策系生物统计学教授、美国联邦政府退伍军人事务部合作研究项目 Palo Alto 协调中心主任。1982 年获复旦大学数学学士学位, 1984 年获上海交通大学应用数学硕士学位, 1990 年获美国加州大学伯克利分校生物统计学博士学位。1990—1994 年任迈阿密大学医学院助理教授, 1994—2009 年在加州大学旧金山分校历任助理教授、副教授、教授, 2009 年开始任斯坦福大学生物统计学教授。2011 年当选为美国统计协会会士, 2014 年泛华统计学会会长。曾获 1990 年美国加州大学伯克利分校统计系 Evelyn Fix 纪念奖, 2003 年中国药学会基金会康辰骨质疏松科学成就奖, 2010 年美国 FDA 顾问委员会服务奖。更多的信息请见 http://med.stanford.edu/profiles/Ying_Lu/。

金 华, 现为华南师范大学数学科学学院教授、广东省现场统计学会副理事长、中国现场统计研究会理事、Statistics in Medicine、Medical Decision Making、Contemporary Clinical Trials 审稿人。1988 年获北京大学理学学士, 1991 年获北京大学概率统计专业理学硕士学位, 1999 年获中山大学流行病与卫生统计专业博士学位。在统计研究、Journal of Statistical Planning and Inference 等发表论文 50 余篇。2005 年获北京市科学技术二等奖。目前研究兴趣包括非劣假设检验、树型生存分析、临床试验设计等。

田 鲁, 斯坦福大学卫生研究与政策系副教授。1995 年获南开大学数学学士学位, 1998 年获南开大学应用数学硕士学位, 2002 年获哈佛大学生物统计学博士学位。曾获 Howard Hughes 霍华德·休斯奖学金, 以及哈佛大学 Robert B. Reed 生物统计学卓越奖。2004—2008 年任西北大学预防医学系助理教授。在统计方法研究、大型流行病学研究、随机临床试验数据管理及应用方面具有丰富经验。目前研究方向是半参数回归模型、生存分析和高维数据分析。



第2版前言

自第1版发行以来,生物统计学和生物信息学有了迅猛的发展。本次再版希望能尽量多涵盖一些重要新兴研究领域的最新进展。许多著名统计学家在各自研究领域取得了杰出的成就,我们特邀请了99位作者分别参与各章的修订增补工作,或更新原有的章节,或撰写新的篇章,对本书新版作出了极大贡献。

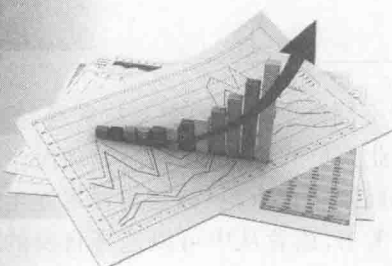
这次新版重新编辑成四部分:医学与流行病学中的统计方法;临床试验中的统计概念与方法;遗传学中的统计方法;概率统计理论与方法。第一部分增加两章新的内容:生存变量的预后生物标记物的协变量特异和协变量校正的预测方法、个体化医学,都是探讨用综合的基因组和/或生物标记物信息预测未来临床结局或治疗效果的最新统计方法。数据描述、变异建模与统计实践,Meta分析,以及流行病学监测中的捕获-再捕获方法四章也做了重大修改,以反映这个领域的最新进展。第二部分增加三章新的内容:定性评价处理作用的替代指标;临床研究中的动态试验设计;中医研究中的统计方法。它们都是介绍不同情形下临床试验设计或分析方法的最新动态。这部分章节进行了更新的包括:统计学在生物制药研发中的应用;药理学和临床前研究的统计学;毒理学的统计方法以及确证性临床试验中的统计问题。第三部分由全新的章节构成,包括:稀疏片段识别及拷贝数变异分析,转录组分析,连锁研究设计与分析的统计方法,人类种群遗传结构,以及全基因组关联研究中的数据整合方法。这部分是自第一版发行以来最热的研究领域,前两章介绍生物信息学中的新方法,后三章讨论全基因组关联研究和连锁研究设计、分析及结果解释中的重要统计问题。第四部分的许多章节也作了实质性的更新,包括:因果推断,Bayes分析,合适的无先验概率推断及其在医学统计中的应用,随机过程及其在医学中的应用,时间序列分析及其应用。这些领域中的统计方法不仅在理论上,而且在计算方法上都经历了革命性的飞速发展,我们删去了过时的内容,为读者展现了最新的研究状况。本书还去掉了第1版中的三章:证实偏倚下的诊断试验准确性评价、遗传学中的统计学、不完全数据极大似然估计的EM型算法。

尽管全书内容密切相关,但是各个章节相对完整,可以单独阅读。本书希望为读者搭建一个方便的平台,使他们对相关的话题感兴趣,并将统计方法应用于科学研究中,同时也为他们深层次的研究提供最新的参考文献。

衷心感谢读者指出第1版中的错误,我们尽量予以更正并避免新的错误。特别感谢Word Scientific Co.的编辑V. K. Sunjeed和人民卫生出版社的编辑,因为他们的耐心和帮助,第2版才能以英文和中文同时顺利面市。

主编 方积乾 陆 盈

2015年1月



第1版前言

回顾上个世纪,许多华人精英先后在西方国家学习、研究统计学与生物统计学。早期先行者中,许宝禄、蒋庆琅、李景均、刁锦寰等享誉国际。六十年代之后,大批台湾、香港和大陆的中国学生在国外获得统计学或生物统计学博士学位,并受聘为美、英、加、澳等国大学的教授和副教授,或在医院、药厂和政府机关的医学统计部门担任要职,成为这一领域的中坚力量。他们中许多人是美国统计学会和国际生物统计学会的资深会员,以及 *Annals of Statistics*, *Annals of Probability*, *J. Royal Statistical Society*, *Biometrika*, *Biometrics* 和 *Statistica Sinica* 等优秀统计学和生物统计学杂志的主编、副主编;先后有多位获得美国统计学会会长大奖 (COPSS) 等学术殊荣,其中,黎子良、范剑青参与了本书的写作。与此同时,在国内的医学院校和科研机构,一大批新型的医学统计学工作者成长起来;他们接受了统计学研究生教育,参与了大量医学统计学教学和科研,积累了丰富的实践经验。在我们共同倡议下,海内外华人医学统计学家汇集一起,按照各自的特长和兴趣,互相学习、取长补短,共同撰写了这本中文专著,命名为《现代医学统计学》。

全书共分四篇二十九章。第一篇生物医学研究中的统计学,含医学中的统计思维、医学诊断、相依资料、医学仪器的质量控制、成本-效果分析、生存质量、Meta 分析、描述与图象、时间序列和模式识别等十章,均系医学和生物学特定领域现代化过程中形成的颇具特色的统计方法学。第二篇药物开发中的统计方法,含药理学和非临床试验、制药与开发研究、毒理学和临床试验等四章,均系现代药物研究和开发中普遍推行的统计学方法。第三篇流行病学中的统计学,含遗传统计、危险度评价、传染病模型、抽样调查的特殊模型、重捕获模型、肿瘤筛查资料分析和因果推断等七章,均系近几十年发展起来的现代流行病学统计方法,正是这些崭新手段使得传统流行病学发生了质的飞跃。第四篇现代统计学方法与理论,含生存分析、纵向资料分析、非参数回归、Bayes 统计、随机过程、树结构回归、EM 算法和人工神经网络等八章,均系与医学统计密切有关的、实用性很强的基础性内容。第四篇不仅概括了各重要分支的现状与发展方向,而且也为前三篇提供了必要的理论与技术。

本书各章相对独立,每一章反映一个专门分枝的最新进展和发展趋势。为适应不同读者群的需要,各章的布局是统一的。第一节概述,以简明的实例和易懂的语言介绍该章的主要内容,包括能解决什么问题,解决问题的思路和需要什么样的资料;后继各节介绍基本概念和常用统计方法,着重实际应用、操作方法和意义的解释;重要内容的推导、证明或文献推荐与介绍放在最后。

本书的读者面十分广泛。相信,不论医学研究者还是医学统计学和数理统计学专业人员,乃至研究生们都可以各取所需、各有所获。统计学思维和技术是医学研究的得力工具,医学研究者阅读本书,不难知道有哪些现代统计方法可以为你所用、基本思路如何、需要怎样的设计和数据、应用的结果如何解释;至于相应的统计学理论和技术,不妨视而不见。本书的每一章侧重医学统计学的一个分支,医学统计学专业人员阅读本书,不但可以全面深入

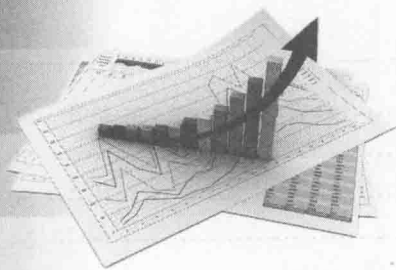
地掌握众多分支的最新成果与未来发展方向,而且还有助于不同分支之间的交叉与贯通,提升咨询服务与研究工作的水准。应用统计和理论统计学专业人员阅读本书,将会发现医学领域对统计学的期待与挑战,生动而丰富的医学问题将成为崭新统计理论、方法与技术诞生的土壤。本书是统计学与医学统计学研究生必读的专业参考书,读者从中可以选择自身的发展方向,一旦选定,有关章节又可帮助迅速入门。

本书每一章的作者都是相关领域的专家,每一章都凝聚了他们独特的学术思想、研究心得和自己的成果;他们在百忙之中精心组织素材,斟字酌句、浸透了心血。在此我们对全体作者的无私奉献深表谢意。

本书得到中国华夏出版基金和中山大学中山医学院“211”工程重点学科建设经费的资助,特此鸣谢。中山大学中山医学院医学统计学教研室的同事和研究生们、尤其是凌莉、曾芳芳、朱淑明和刘国辉老师协助了许多秘书工作,在此一并致谢。

方积乾 陆盈

2001年11月



目 录

第一篇 医学与流行病中的统计方法

第一章 医学中统计思维的历史	陈达 方积乾	3
第一节 引言		3
第二节 Laplace 及其远见		3
第三节 Louis 与数值方法		4
第四节 统计分析与实验室研究		5
第五节 现代统计学的开端		6
第六节 医学统计学的开端		7
第七节 实验中的随机化		8
第八节 第一个随机化有对照的临床试验		8
第九节 政府法规与统计学		9
第十节 尾声		10
参考文献		11
第二章 数据描述、变异建模与统计实践	杜鸿雁 谭铭 陈豪 金华	14
第一节 概述		14
第二节 描述数据的方法		15
第三节 通过模型调整因素来描述数据		19
第四节 过度离散问题		19
第五节 统计实践		22
参考文献		29
第三章 生存变量的预后生物标记物的协变量特异和协变量		
校正的预测方法	马昀蓓 周晓华 KC Gary Chan 赵星 李晓松	32
第一节 前言		32
第二节 协变量特异时依 PPV 曲线		33
第三节 协变量校正时依 PPV 曲线		36
第四节 模拟研究		38
第五节 多中心艾滋病队列研究		42
第六节 讨论		43
参考文献		48

第四章 个体化治疗的统计学方法	田鲁 赵晓光 陈新林	50
第一节 简介		50
第二节 个体化治疗的风险预测		50
第三节 个体化治疗的选择		54
第四节 讨论		60
参考文献		60
第五章 统计学方法在放射医学研究的质量控制、质量保障和质量改进中的应用	陆盈 赵守军 颜杰	64
第一节 序言		64
第二节 测量误差		65
第三节 统计过程控制图		76
第四节 一致性评价		84
第五节 校准和标准化		92
第六节 结论		96
参考文献		96
第六章 统计方法在医学图像分析中的应用	金声 赵耐青	102
第一节 概述		102
第二节 应用图像的统计特性进行采样和压缩		103
第三节 用各向异性统计扩散减少噪声		106
第四节 医学图像的分割		110
第五节 利用 3 维 Monte Carlo 模拟改进图像配准的置信区间		115
第六节 结论		117
参考文献		117
第七章 成本效果分析与循证医学	李建立 宇传华	121
第一节 前言		121
第二节 成本-效果分析的基础		123
第三节 统计方法		125
参考文献		142
第八章 与健康有关生存质量研究的统计设计与分析	郝元涛 方积乾	144
第一节 生存质量的概念及发展		144
第二节 生存质量测定量表的研制方法		146
第三节 生存质量测定量表的国际化方法		149
第四节 生存质量研究的设计与实施		152
第五节 生存质量资料的特点及相应的统计分析方法		156
第六节 心理统计的基本原理与方法		162

参考文献	172
第九章 Meta 分析	周旭毓 方积乾 宇传华 许宗利 田鲁 陆盈 175
第一节 Meta 分析概述	175
第二节 Meta 分析的常用统计方法	183
第三节 Meta 分析的 Bayes 方法	204
第四节 诊断试验的 meta 分析	213
第五节 适宜于连锁分析文献的 meta 分析方法	228
第六节 Meta 分析中的偏倚	235
参考文献	240
第十章 传染病模型	武虎林 赵守军 王显红 245
第一节 概述	245
第二节 流行病学数学统计模型	246
第三节 病毒动力学模型	254
第四节 干预和预防	257
第五节 小结	263
参考文献	263
第十一章 抽样调查中的特殊模型	高素娟 高桂明 方积乾 270
第一节 估计疾病患病率的模型	271
第二节 小区域分析的随机效应模型	274
第三节 捕获-再捕获模型	281
参考文献	286
第十二章 重复捕取(捕获-再捕获)方法在流行病 监测上的应用	赵莲菊 谢宗震 杨欣洲 289
第一节 前言	289
第二节 重复捕取的背景与动机	295
第三节 不考虑协变量的模型	298
第四节 包含协变量的模型	310
第五节 结论与讨论	317
参考文献	317
第十三章 疾病筛查资料的统计方法	柳青 321
第一节 疾病筛查的基本概念	321
第二节 筛查疾病自然史的一阶段简化模型	325
第三节 筛查疾病自然史的两阶段模型	328
第四节 筛查疾病自然史的多阶段马尔科夫链随机模型	332

第五节 筛查方案的模拟评价及优化	339
参考文献	342
第二篇 临床试验中的统计方法	
第十四章 统计学在生物制药研发中的应用	周贤忠 彭安佩 张新信 贺佳 347
第一节 医药研发	347
第二节 主要统计学原则	348
第三节 生物利用度与生物等效性	352
第四节 有重复记录的 QT/QTc 研究	357
第五节 全球性临床试验进展	367
第六节 影像诊断中的统计	379
第七节 GCP 中的统计原则	384
参考文献	388
第十五章 药理学和临床前研究的统计学	黎子良 施美琼 朱光锐 王强 396
第一节 引言	396
第二节 药物动力学和药效学	397
第三节 生物利用度和生物等效性	409
第四节 测定方法的开发与确证	412
第五节 药物开发	412
参考文献	413
第十六章 毒理学的统计方法	陈章荣 李妍 宇传华 417
第一节 动物致癌试验	417
第二节 生殖研究	427
第三节 混合试验	432
参考文献	437
第十七章 健康危险度评价中的剂量-反应模型	朱怡良 孙晓武 441
第一节 引言	441
第二节 两分类数据:致癌性	443
第三节 聚集性多项数据:发育毒性	445
第四节 纵向数据:神经行为毒性检测	448
第五节 基准剂量	454
第六节 讨论	455
参考文献	455
第十八章 确证性临床试验中的统计学基本	

问题	季渝华 许海燕 刘清 夏结来	459
第一节 引言		459
第二节 最小化可能引入的偏倚		461
第三节 第 I 类错误率的控制		467
第四节 期中分析、成组序贯设计与适应性设计		478
第五节 阳性对照试验		482
参考文献		490
第十九章 定性评价处理作用的替代指标	耿直	498
第一节 引言		498
第二节 定义和记号		499
第三节 基于因果性先验知识的替代指标准则		505
第四节 基于关联性先验知识的替代指标准则		508
第五节 讨论		510
参考文献		511
第二十章 临床研究中的动态试验设计	彭安佩 周贤忠 刘裕	513
第一节 简介		513
第二节 什么是动态设计		514
第三节 方案修订的影响		518
第四节 按照设计调整的挑战		520
第五节 回顾性调整的困难		525
第六节 结束语		527
参考文献		528
第二十一章 中医研究中的统计方法	易丹辉 李扬	533
第一节 概述		533
第二节 中医证候评价研究		534
第三节 中医症状筛选研究		540
参考文献		546

第三篇 遗传研究中的统计方法

第二十二章 稀疏片段识别及拷贝数变异分析的		
应用	郑欣格 蔡天文 李宏哲 肖云 范慧慧 平艳艳 李霞	551
第一节 背景		551
第二节 基于 SNP 芯片获取的 LogR 比值数据进行单样本最优拷贝数		
变异识别		552
第三节 基于单样本第二代测序数据的具有鲁棒性的拷贝数变异识别		555

第四节 基于群体样本对重复出现的拷贝数变异的识别	560
第五节 讨论	562
参考文献	564
第二十三章 连锁研究设计与分析的统计	
方法	李启寨 覃红 李照海 郑刚 李彩霞 567
第一节 绪言	567
第二节 连锁分析	571
第三节 讨论	580
第四节 软件	580
参考文献	580
第二十四章 用新一代测序技术的转录组分析	李婧翌 黄海艳 钱敏平 张学工 584
第一节 引言	584
第二节 序列数据产生的综述	585
第三节 讨论	592
参考文献	594
第二十五章 人类种群遗传结构	唐华 唐鲲 599
第一节 引言	599
第二节 量化群体结构	601
第三节 人群结构的经验数据	605
第四节 近期的遗传混合	607
第五节 遗传关联研究中的种群结构和混淆情况	608
第六节 结束语	610
参考文献	610
第二十六章 全基因组关联研究中的数据整合方法	孙宁 赵宏宇 伍亚舟 易东 614
第一节 引言	614
第二节 GWAS 分析中不同的数据类型	615
第三节 整合不同数据类型的统计方法	617
第四节 讨论	620
参考文献	621

第四篇 通用的概率统计理论与方法

第二十七章 统计因果推断	耿直 627
第一节 引言	627
第二节 Yule-Simpson 悖论	628

第三节	因果推断的反事实模型	629
第四节	混杂偏倚与混杂因素	632
第五节	基于因果图的混杂因素的准则	640
第六节	多混杂因素	643
第七节	纵向研究的因果推断	644
第八节	小结	645
参考文献		645
第二十八章	生存分析	林丹瑜 赵国龙 648
第一节	引言	648
第二节	基本概念	648
第三节	生存分布的估计	649
第四节	计数过程理论	650
第五节	两样本统计	651
第六节	回归模型	651
第七节	Cox 比例危险率模型	652
第八节	相乘性强度模型	653
第九节	回归模型诊断	654
第十节	Cox 模型的替代模型	655
第十一节	多元失效时间资料	656
第十二节	结语	657
参考文献		658
第二十九章	纵向数据分析的回归模型	伍鸥 田欣 余启汾 吴密霞 663
第一节	引言	663
第二节	参数、半参数方法概述	669
第三节	变系数非参数模型	676
第四节	基于分布函数的非参数模型	693
第五节	结果自适应协变量的回归方法	701
参考文献		715
第三十章	核密度估计和非参数局部多项式回归	范剑青 李润泽 颜杰 720
第一节	引言	720
第二节	密度估计	720
第三节	局部多项式拟合	727
第四节	局部似然函数和局部部分似然函数	733
第五节	非参数拟合优度检验	739
第六节	其他应用	744
参考文献		749

第三十一章	非独立数据的统计分析方法	陈峰	754
第一节	常见的非独立数据实例		755
第二节	常见的非独立数据的组内相关结构		765
第三节	传统方法及其局限性		769
第四节	非独立数据的广义估计方程		774
第五节	非独立数据的多水平模型		778
第六节	组内相关系数的区间估计及抽样分布		786
第七节	非独立试验的样本含量及成本效益		790
	参考文献		794
第三十二章	Bayes 方法	陈明辉 叶可英 曹文君 徐勇勇	796
第一节	概述		796
第二节	后验推断		797
第三节	先验的选取		804
第四节	Bayes 计算		812
第五节	Bayes 非劣效临床试验设计		817
第六节	应用与实例		821
	参考文献		835
第三十三章	合适的无先验概率推断及其在医学统计中的应用	Duncan Ermini Leaf Hyokun Yun 刘传海 李松 金华	844
第一节	引言		844
第二节	推断模型(IM)		846
第三节	条件推断模型和边际推断模型		850
第四节	受约束推断模型		853
第五节	选择推断		856
第六节	基于二项分布数据的推断		858
第七节	Meta 分析:多正态均值问题		861
	参考文献		865
第三十四章	随机过程及其在医学中的应用	李彩霞 方积乾	868
第一节	Markov 链		868
第二节	Markov 链的应用		874
第三节	一些特殊的 Markov 链		880
第四节	Markov 链在统计计算中的应用——MCMC 方法		886
第五节	基于 Markov 链的流行病传播模型		891
第六节	计数过程与生存资料回归模型		893
	参考文献		897

第三十五章	时间序列中的缺失值填补和移动	
	假日效应的调整	张晋昕 张熙 薛允莲 赵志 899
第一节	基于周期信息的时间序列缺失值填补方法	899
第二节	逐日观测序列移动假日效应的调整	909
	参考文献	917
第三十六章	树形法	张和平 宇传华 919
第一节	前言	919
第二节	CART 的基本知识	919
第三节	生存树	923
第四节	多元反应变量的分类树	925
第五节	结束语	927
	参考文献	927
第三十七章	BP 神经网络模型简介	夏结来 唐启义 930
第一节	神经网络简介	930
第二节	BP 神经网络	931
第三节	DPS 数据处理系统操作步骤	933
第四节	应用实例	934
第五节	基于遗传算法的人工神经网络	936
第六节	未来发展之预测	939
	参考文献	940