

DPS<sup>©</sup> DATA PROCESSING SYSTEM  
Experimental Design, Statistical Analysis and Data Mining

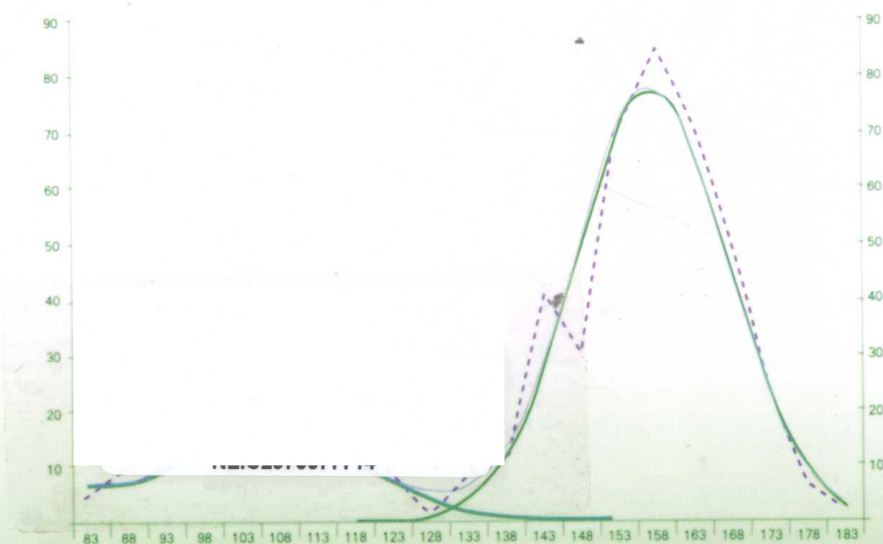
.Tang Qiyi

# DPS<sup>©</sup> 数据处理系统

实验设计、统计分析及数据挖掘

(第2版)

唐启义 著



科学出版社  
www.sciencep.com

# DPS<sup>©</sup>数据处理系统

——实验设计、统计分析及数据挖掘

(第2版)

DPS<sup>©</sup> Data Processing System

——Experimental Design, Statistical  
Analysis and Data Mining

(Second Edition)

科学出版社

北京

## 内 容 简 介

本书从应用角度简要地阐述了现代统计学 400 多种实验数据统计分析和模型模拟方法,如包括混料实验设计在内的各种实验设计、方差分析、列联表分析及非参数检验;专业统计包括量表分析、数据包络分析、随机前沿面分析、顾客满意指数模型、试验诊断 ROC 曲线、生物测定、生存分析、动植物遗传育种、品种区域试验、空间分布型、地理统计、数值生态学方法等;各种回归分析、聚类分析、主成分分析、判别分析、典型相关分析、对应分析等多元分析技术;非线性回归模型参数估计、模型模拟技术;单目标和多目标线性规划、非线性规划等运筹学方法;以及状态方程、数值分析、时间序列分析、模糊数学、BP 神经网络、径向基函数(RBF)、支持向量机(SVM)、层次分析、小波分析、灰色理论等方法。全书共 8 篇 44 章,并配以作者自主开发的具有自主知识产权的 DPS 统计软件系统光盘 1 张。

本书可供从事农林牧渔、医药卫生、气象、水文地质、工程计算、市场调研、经济等领域的广大科研、教学、管理和技术推广人员使用,也可作为有关专业大学生和研究生学习统计课程的参考书及实践工具书。

### 图书在版编目(CIP)数据

DPS 数据处理系统:实验设计、统计分析及数据挖掘/唐启义著. —2 版,  
—北京:科学出版社,2010  
ISBN 978-7-03-026443-5

I. ①D… II. ①唐… III. ①计算机应用—统计分析(数学) IV. ①  
0212.1-39

中国版本图书馆 CIP 数据核字(2010)第 010039 号

责任编辑:赵彦超 鄢德平 / 责任校对:李奕莹

责任印制:钱玉芬 / 封面设计:王浩

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

双青印刷厂印刷

科学出版社编务公司排版制作

科学出版社发行 各地新华书店经销

\*

2007 年 1 月第 一 版 开本:787×1092 1/16

2010 年 2 月第 二 版 印张:73

2010 年 2 月第一次印刷 字数:1 650 000

印数:1—3 500

定价:148.00 元(含光盘)

如有印装质量问题,我社负责调换

# 序 一

统计学有很长的历史，在工农业、科技、商业、社会科学、政府决策等领域已得到普遍的应用，已经或正在成为社会发展的必不可少的工具。统计学的普遍应用只是近 40 年的事，得益于计算机的飞速发展。毫不夸张地说，没有计算机及统计软件的支撑，统计学的发展绝大部分只能停滞在理论研究上。一个新的统计方法，如果没有统计软件的支撑，是很难普及的。统计软件成为统计学发展的重要方面。

中国学者对统计软件的关注由来已久。早在文革期间，由于地质勘探的需要，中国许多地质大队和大的油田花了大量的人力汇编应用多元统计分析的统计软件。可惜编制这些软件的力量没有能够集中、提高，绝大多数团队均自然衰退。以中国科学院前计算中心(现属中国科学院数学与系统科学研究院)的张建中、杨自强研究员为首的团队，一直在发展中国人自己的大型系统统计软件，在 20 世纪 80 年代，他们推出了统计软件包 SASD，对中国应用统计的发展起了很大作用。随着中国的改革开放，许多国际通用的大型统计软件包如 SAS, SPSS, MATLAB, R 等逐渐进入中国的统计教育和统计软件市场，中国人自己发展的各种统计软件包的国内市场却日益缩小。

中国是一个大国，需要有中国人自己发展的视窗系统，同时也需要具有自主知识产权的大型统计软件系统。这样的想法不是“闭关自守”，也不是“狭隘的民族主义”。

统计是一门大的学科，包括了众多的方法，如数据的描述性统计、单变量的统计分析(如  $t$  检验,  $F$  检验等)、多变量的统计分析、定性数据的数据分析、生存分析、试验的设计和建模、非参数统计等。一个大的统计软件包大多是由一个公司、一个研究单位或一所大学的一个大的团队来创造、发展、推销及服务。使我万分惊讶的是 DPS 统计软件是由唐启义教授个人用了 20 多年的时间，以“愚公移山”的精神创立和发展起来的。该软件涉及面广，使用方便，很受广大用户欢迎。唐启义教授坚持不懈的精神，令人感动。相信在唐启义教授的呵护下，DPS 统计软件会与时俱进，更加人性化，不仅在中国处处开花结果，而且能深入到全世界广大华语地区。



2009 年 5 月于中国珠海市

## 序 二

### ——青目睹人少，问路白云头

有所谓“文字的学问”，有所谓“数字的学问”。从科举制度迄今，文字学问的重要性不言而喻。相形之下，数字学问的普及与认知，尚有一段漫长的路。常言道：“数字会说话”，此言不假。但是当数字说话时，您听得懂、看得明白吗？统计学，作为研究数字学问的一门主要学科，在中国历史源远流长。相传在伏羲时代，人们便开始“结绳记事”（详见《周易》），《管子》问篇亦提及：“不明于数，欲举大事，如舟之无楫而欲行于大海也。”商鞅在公元前 390 年更提出：“强国需知十三数”，为统计的重要性立下经典。近来在商场上盛行的古代兵书《孙子兵法》更直言：“多算胜，少算不胜”。这些都说明统计在古代中国受重视的程度，数千年后的今天读起来依然意味深长。

计算机当属 20 世纪对人类历史推动力最大的技术。从 1946 年在美国宾夕法尼亚大学诞生第一台计算机以来，计算机在功能上以飞快的速度发展，影响领域也从开始的军事领域拓及到经济社会，以及其他各个领域。1974 年，Intel 发布了其面向个人机的八位微处理器芯片 8080。随着超大规模集成电路和微处理器技术的进步，计算机技术更从过去主要集中在大型机和小型机高端领域逐渐进入普通人的生活。80 年代初第一台个人计算机面世以来，计算机的成本快速下降。之后，随着互联网技术、多媒体技术的飞速进步，计算机技术快速发展。1993 年互联网开始商业化运行，1995 年 Windows 95 发布，日新月异的技术促进了计算机的快速发展与普及。1997 年 IBM 深蓝计算机战胜了人类国际象棋世界冠军，标志着人工智能发展日趋成熟。进入新世纪以来，计算机发展更为迅猛，其与通讯技术等相互渗透和融合，大大推动了人类社会的进步。

计算机硬件的发展为统计的发展提供了难得的历史机遇。统计学发展初期的方法虽然很好，但是由于计算工作量大，使得统计分析方法并没有在实践中有效推广开来。而电子计算机的发展促进了软件的开发与商业化，推广了统计学的应用。具体来说，计算机从两个方面促进了统计学的发展。一方面，统计计算技术使得统计学家可以在计算机上实现自己的理论，推动了统计理论的发展；另一方面，大量统计软件的普及使得统计分析广泛应用到社会的各个方面，这些实践活动又推动了统计理论的发展。总之，现代统计学已经离不开计算机的发展，计算机已成为统计学发展的基本工具。

根据 1977 年 Kohm 在美国统计学会上的介绍，当时美国有 55 个公众可用的统计软件。1981 年，Francis 的著作中介绍了当时世界各国已有的统计软件 117 个。随着统计方法的不断改进，统计软件也在不断改进中。70 年代百家争鸣以来，研发出了一批目前占据主流地位的统计分析软件，比如 SAS, MINITAB, BMDP, SPlus, SPSS 等。由于使用目的的不同，这些统计软件有着不同的使用人群。一般而言，BMDP 和 SPSS 比较受到生物医药和社会科学领域人员的喜爱，而主流的应用统计人员则更倾向于 SAS, SPlus 则

更适合于进行统计研究的科研工作者。其中 SAS 更是目前全世界最大的统计软件公司。其他享有较好声誉的统计分析软件还有 SYSTAT, DataDesk 和 JMP 等。这些软件都是商业软件, 由于这些软件比较昂贵, 在教育界未能充分普及。20 世纪 90 年代末期, 出现了类似于 S 语言的 R 软件, 该软件平台是免费的, 大大增加了其普及性, 而由于其灵活性及强大的编程实现能力, 该软件获得迅速发展和推广。

在中国统计学快速发展的过程中, 统计分析软件必不可少。虽然 80 年代以来, 各科研单位所编制的统计分析软件和程序数不胜数, 但是随着统计学的发展, 绝大多数统计软件由于各种各样的原因而渐渐淡出市场, 使得国外较规范的统计分析软件如 SAS, SPSS, BMDP, GLIM 等占据了国内统计分析软件的大部分市场。但是, 国内也产生了一批优秀的统计分析软件, 如 NOSA, SPLM, Qstat, PEMS, Markway, DPS 等, 其中 DPS 分析软件便是最为出色的统计分析软件之一。它立足国内市场, 从初期主要用于处理农林生物科研数据起步, 经过不断改进, 至今已发展成为一款综合性的大型统计软件, 尤其是它还开发了具有全球领先地位的均匀试验设计、混料试验设计、动态聚类分析几个功能模块的独创算法。如今, DPS 应用已遍及全国各地, 涉及自然科学和社会科学各个领域, 如农林牧渔、医药卫生、气象、水文地质、工程计算、市场调研、社会经济、金融等。

改革 30 年, 我国经济一天比一天繁荣, 但我们的教育与科研并没有明显的同步提升, 距离国际水平尚有一大段发展空间。许多方面尚需努力与世界接轨, 更遑论领导世界潮流。不过从另一个角度看, 我们也不该一味迷信西方理论, 而应该从我国国情出发, 建立一套适合我国学术与教育的统计软件。这正证实了《孙子兵法》(虚实篇) 所言: “致人而不致于人”。换言之, 如何自力更生, 建立一套属于中国人特质的统计分析软件, 而不将就于外国人的思维与文化, 进而进军国际市场, 影响全世界。作为国内最为优秀的统计分析软件之一, DPS 统计软件研发始于 1988 年, 经过 20 多年的努力, 发展到现在的规模, 填补了国内具有自主知识产权的大型统计软件的空白。

统计大师赵民德先生在《探索真相》一书的导读中引用了《鹿鼎记》第二十二回中韦小宝为了学“一指禅”就询于少林寺首座澄观和尚的一段对话。此话给我印象极深, 特别引用如下:

“韦小宝: ‘你说那一指禅并不难学, 可是从少林长拳练起, 一路路拳法练将下来, 练成这一指禅, 要几年功夫?’”

澄观道: ‘这在般若堂的典籍中是有得记载的。五代后晋年间, 本寺有一位法慧禅师, 生有宿慧, 入寺不过三十六年, 就练成了一指禅, 进展神速, 前无古人, 后无来者。料想他前生一定是一位武学大宗师, 许多功夫是前生带来的。其次是南宋建炎年间, 有一位灵兴禅师, 也不过花了三十九年时光。那都是天纵聪明、百年难遇的奇才, 令人好生佩服。前辈典型, 后人也只有神驰想象了。’”

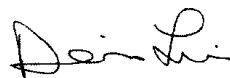
统计学之深奥, 多少人穷极一生都无法参透, 吾等凡夫俗子又如何能在有生之年得到真谛呢? 《DPS<sup>®</sup>数据处理系统——实验设计、统计分析及数据挖掘》让我们看到了一线曙光。该书内容涵盖之广, 仿佛一本统计百科全书。全书 8 篇 44 章, 包山包海, 几乎包括了所有目前统计应用的主要方向。全书由浅入深, 精辟易懂, 配合软件操作, 让一门

高深的统计学可以在短期内得以理解并熟练地应用在生活周边的事务上。那统计学的“一指禅”在一年半载内借由该书得以学到六七成功力，不亦乐乎！

浙江大学唐启义教授是个认真的学者，同时也是热情而务实的工作者。从另一个角度看，唐教授是个理论扎实的统计家，同时也是应用统计实务的执行人。更者，他是个计算机专家。通过他精心设计，妙笔生辉，该书充分反映了这些特质，也适时地满足了我国目前统计界的需求。

是以为序。

林共进



2009年春写于美国宾夕法尼亚州

## 第二版前言

《DPS<sup>®</sup>数据处理系统——实验设计、统计分析及数据挖掘》(第二版)是作者基于自主开发的 DPS 软件,以实用为目的的统计工具书。

根据读者、用户建议,这次再版除对原有的一些内容进行修改、补充外,还增加了作者在过去两年研发出来的若干新功能的介绍。全书新增了 3 章共 30 多节,并对一些章节的顺序进行了调整。新增的 3 章是:

第 11 章“混料试验设计与分析”。在第一版的均匀试验设计技术取得的重大进展的基础上,本章介绍了自主开发的、较目前已有的混料试验设计方法更有效的混料试验设计的新技术,该技术能适合各种情形的混料设计,特别是组分多、试验次数大、限制条件“苛刻”的情形。同时提供了混料试验数据分析和优化的数学模型,为工业、农业和科学试验中应用混料设计提供了新的技术。

第 12 章“数据包络分析与随机前沿面分析”。主要用来评价相同类型部门或单位(称为决策单元)中各成员之间的相对有效性。其中包括非参数的数据包络分析(data envelopment analysis, DEA)和属于参数方法的随机前沿面分析(stochastic frontier analysis, SFA),并给出了在实际工作中计算生产效率时应用较多的 Malmquist 指数。

第 16 章“序贯试验分析”。在进行下一次试验以前,已经知道前面试验的结果,这种一次接着一次的抽样试验为序贯试验。本章提供了在序贯试验中计算各接收时刻的接收概率和各时间区段的拒收概率的方法,并给出了工作特性曲线。

除这 3 章外,还在方差分析、非参数检验、多因素优化与分析、量表分析和顾客满意度指数模型、多变量统计检验、回归分析、多因素分析、非线性回归模型、时间序列及多指标综合评价等 10 多个章节,增加了新的统计分析功能,如试验设计和方差分析中的 Box-Behnken 设计、平衡不完全区组设计试验方差分析、增广随机区组设计及方差分析、多元方差分析的线性模型方法;回归分析中的稳健回归、积分回归和主导分析;多变量时间序列的独立分量分析;以及结合分析、展开法、非参数回归、支持向量机和投影寻踪分类等。

经过这两年多的努力,尽管 DPS 系统功能又进一步增强,但距离广大读者的要求仍差距甚大。希望广大读者一如既往地继续支持 DPS 这一国产统计分析软件,及时指出目前版本的缺点和不足,以便我们继续改进,加速开发新的数据分析技术。读者可通过访问网站 <http://www.chinadps.net> 得到技术咨询和软件升级。

本书再版得到了国家高技术研究发展计划(“863”计划)项目(2006AA10Z217)和农业公益性行业科研专项(200803003)的资助。在本书出版过程中,作者所在单位和科学出版社给予了大力支持。这里,作者特别感谢浙江大学生命科学学院冯明光教授多年来对作者工作的支持和帮助。对浙江大学原副校长程家安教授、湖南省农业厅李绍石研究员、中国水稻研究所胡国文研究员及已故的中国科学院庞雄飞院士对作者在 DPS 统计软件



开发、升级，以及本书写作过程中给予了殷切关怀和鼓励致以衷心的感谢。并对方开泰教授、王静龙教授、周仁郁教授、刘明芝教授、张承恩高级工程师、董玉恒教授及夏结来博士等给予的帮助表示感谢。浙江大学软件工程专业唐睿和外语系沈宁宁协助完成了 DPS 数据处理系统软件英文版的翻译与转换工作；研究生时嵩、彭奇、高薇、缪清玲协助完成了本书文稿的文字处理；浙江大学植物保护系、昆虫科学研究所农业部作物病虫害分子生物学重点开放实验室以及浙江省农业遥感与信息技术重点研究实验室的同事给作者的工作以很大帮助，在此一并表示感谢。

最后，对 DPS 广大用户以及关心和扶持 DPS 成长的各个领域的专家、学者表示衷心感谢。

唐启义

2009 年春节于杭州华家池

## 第一版前言

呈现在读者面前的《DPS<sup>®</sup>数据处理系统——实验设计、统计分析及数据挖掘》是作者于 2002 年出版的《实用统计分析及其 DPS 数据处理系统》的补充和发展。和原来版本一样，它既是一本以实用为目的的统计工具书，更是一份请柬。它邀请读者进入 DPS 数据处理系统漫游，寓复杂繁冗而又沉闷乏味的统计分析于轻松自如和无穷乐趣之中。

现代科学研究，不管是自然科学还是社会科学的各个领域(物理、化学、工程、医学、生物、经济、社会等)，只要存在着带有随机性的数据，就需要应用数理统计方法去有效地收集、整理和分析，以便得出正确的结论，为相关决策行动提供科学依据。

统计学的形成可以追溯到 20 世纪 20 年代。当时 Fisher 等以农业科学试验为对象，探讨试验资料的整理和统计分析方法。1922 年，Hayes 研究亲代与子代蛋白质含量的相关性，提出了相关分析的理论。翌年，Fisher 和 Maekenzie 研究肥料对马铃薯产量的影响，首次提出了方差分析和交互作用的概念。1925 年，Engedow 研究不同品种在不同地区和不同年份的产量表现时，进一步阐述了因素间的交互作用。Fisher 等正是在研究农业科学试验方法的基础上，建立起了统计学基本理论和方法。这些已被广泛地应用于科学研究、工农业生产和现代经济管理。

随着经典数理统计科学的发展，在 20 世纪四五十年代，多元统计分析在 Pearson 及 Fisher 等人开拓性工作基础之上又有了坚实的理论基础，如关于多元正态分布的理论著述、Anderson(1958)的《多元统计分析导论》等著作，标志着统计学进入到一个新的发展阶段。但是，多元统计分析的发展、完善和广泛应用，离不开计算机科学和计算数学的发展。由于多元统计分析的计算工作量十分惊人，没有计算机是不可想像的。正是由于计算机技术的飞速发展，一些独特的多元统计分析技术，如偏最小二乘法、主成分分析、典型相关(又称“典范相关”)、因子分析、对应分析、判别分析、聚类分析等，在农林、医药、气象、地质、生物、心理、工业、经济、社会等许多研究和生产领域获得了广泛应用。目前，多元统计分析已是统计学中最活跃的分支学科之一。

传统的自然科学和社会科学研究以定性描述为主。Fisher 等创立的一套经典统计方法在定量研究方面迈出了一大步。但囿于其固有缺陷，试验结果的统计推断不能对有关数量关系从机理上予以合理解释，因而其统计模型仍属经验范畴。因此，现代科学工作者们正力图使具体学科的理论及研究对象同数学有机地结合起来，以模型来描述、解释自然现象和社会、经济的运动规律，从而解决科研和生产中的实际问题。

自 20 世纪六七十年代以来，一些新的统计理论和分析技术发展起来。如 20 世纪 60 年代 Zadeh 提出的模糊集合论在各个领域的应用已十分广泛。实践证明，模糊数学方法在农业、图像识别、天气预报、地质地震、交通运输、医疗诊断、信息控制、人工智能等领域的应用已初见成效，其发展前景广阔。作为预报和控制的有力工具，研究系统变化规律的时间序列分析技术目前正被广泛用于地质、石油、气象、商业、工程控制等领

域, 在农业中已被用于作物病虫害的中长期预测预报, 并获得了可喜的成功。灰色系统是我国学者邓聚龙先生于 20 世纪 80 年代初提出的用于控制和预测的新技术, 目前亦在我国农业、社会和经济等领域广泛应用, 并已取得显著成就。

统计学的发展源远流长, 内容十分丰富。迄今为止, 国内外统计学专著和教科书已有多种版本见诸于世。如果按照一般统计学著述的内容和体裁, 我们没有必要再推出一本教科书似的理论专著, 让广大科技工作者在复杂的数学公式堆中艰难跋涉。本书的编写宗旨强调实用性。作者在内容上刻意求新, 形式上力求通俗易懂, 在取材上尽可能全面, 并注意反映国内外在试验设计和统计分析方面的最新研究进展。书中介绍的 400 多种统计分析方法, 可以满足各个领域的统计分析需要, 且都能在作者设计制作 **DPS** 计算机数据处理系统中随意调用, 自如操作。本着服务于读者、服务于用户的最高宗旨, 作者对入选的 400 多种统计分析方法分门别类, 每一类(种)统计分析技术按照方法或原理简介与操作示例两个基本部分编写, 全书共分 8 篇 41 章。同时, 作者按书中篇目类别设计 **DPS** 系统的套叠下拉式主菜单和子菜单, 将带给读者和用户极大的便利。下面分篇给予简单介绍, 以引导读者高效率地使用书中的各种分析技术和 **DPS** 数据处理系统。

第一篇主要介绍 **DPS** 系统的性能、特点和操作要领。**DPS** 数据处理系统是作者独立研究开发的大型通用多功能数据处理分析应用软件系统。该系统在 **IBM PC** 及其兼容机、中文 **Window 98/2000/XP** 下运行。因此, 在使用 **DPS** 系统之前, 建议读者首先浏览本篇内容, 熟悉该系统将使读者在数据处理和统计分析中事半功倍。

第二篇的编写充分考虑到经典试验统计的完整性和系统性, 提供了比较全面的试验设计和试验结果统计分析的内容, 如  $t$  测验、方差分析、列联表分析、非参数检验、最优回归试验设计与分析以及多元方差分析等, 尤其对方差分析给予了较为详尽的介绍。因此, 各个层次、各个领域的读者都可使用这些基本的统计学分析方法。

第三篇介绍了一些应用于某些较特殊专业领域的统计分析技术, 如心理统计中的量表分析、顾客满意指数的结构方程模型分析、试验评价的 **ROC** 曲线分析、地理统计分析, 还为育种学工作者提供了品种区域试验的统计分析技术, 为生态学工作者提供了动植物种群数量抽样、动植物种群数量空间分布研究、生物群落的消长演替过程分析、生存分析等方面的内容; 对从事生物测定研究的科技人员来说, 作者既提供了经典 **Finney** 概率分析方法, 还重点引入近几年才逐渐发展起来的更为完善的分析工具, 即时间-剂量-死亡率模型(**TDM**)分析技术。

第四篇以较大篇幅介绍多变量统计分析技术, 并注意反映目前国内外的最新研究动态, 将最新的技术提供给读者。近年来, 多变量统计分析技术已被广泛应用于各个领域的科学试验中, 因为人们越来越多地将研究对象作为一个有机结合的整体即多元素系统(亚系统)对待, 而不是分割开来作为单独事物处理。“从树木到森林”, 抓住系统运动变化的本质, 舍弃次要因素。通过简化系统的结构, 探索系统运动变化的总趋势, 为人类征服自然、改造自然服务。因此, 偏最小二乘回归分析、主成分分析、因子分析、对应分析、典型相关分析、聚类与判别分析以及马尔可夫链分析等多因素分析技术备受自然科学和社会科学各领域科研人员的重视。在 **DPS** 中还提供了我们研发的含有定性变量的

逐步回归分析技术。

第五篇和第六篇向读者介绍相当灵活的通用数学模型的模拟分析和数值计算工具。因前面介绍的分析方法和程式基本上都有完整的体系和要求,读者必须按部就班地照着去做。而利用第五、六篇的通用工具,读者将尽可能地发挥自己的专长和展现自己的学识水平,实现思维活动的飞跃。例如,模拟作物生长发育过程或经济活动的演替规律,建立结合自己专业和学科特点的数学模型,而不是套用别人的现成模型。读者可以借助于 DPS 系统强大的数学建模功能,在计算机上实现自己的思维活动,获得所想即所见、所见即所得的效果。

第七篇介绍时间序列分析技术。时间序列分析是根据系统观测得到的时间序列数据,通过曲线拟合和参数估计来建立数学模型的理论和方法,常用于国民经济宏观控制、区域综合发展规划、企业经营管理、市场潜量预测、气象预报、水文预报、地震前兆预报、农作物病虫害灾害预报、环境污染控制、生态平衡、天文学和海洋学等方面。

第八篇为读者提供一些不属于经典统计学但近年来发展迅速且应用已经相当广泛的模型分析技术,即模糊数学分析、灰色系统分析和神经网络模型技术。只要读者有兴趣,DPS 系统可以帮助读者熟悉和掌握上述各项分析技术,并应用到研究实践中去。

综上所述,本书和 DPS 数据处理系统是密不可分的整体。作为通用的数据处理和统计分析的工作平台,作者在系统的设计方面注重专用性与通用性以及“傻瓜”性和灵活性的关系,以适应不同层次读者的需要。

专用性和通用性。DPS 的有些功能是专为某些领域或完成某一类工作而专门设计的,如“生物测定”是专为从事化学或生物制剂的生物活性测定的人员而设计的;种群空间分布型及群落参数统计分析、品种区域试验等章节都是服务于某个领域的科学工作者。相对来说,它们的用途比较专一。但是,本书中绝大部分统计分析和建模技术都是通用的,例如丰富的数学函数、交互方式的矩阵运算功能以及通用数学模型的模拟分析技术。DPS 处理数学公式的功能非常独特,有利于读者发挥潜在的想象力和创造力。同时,数学计算功能,特别是矩阵交互式处理功能,在当前微机普及的情况下,将有助于高等院校数理统计课程的教学,使繁冗无味的矩阵运算得以在计算机上精确而快速地进行,并且在一定程度上提高学生的学习积极性和教学质量。

“傻瓜”性和灵活性。在 DPS 平台上进行经典统计分析的过程十分简单。它无需读者懂得数学推导,无需读者了解运算细节,只需读者将自己的实验数据输入电子表格,将它们定义成数据块(移动鼠标将待分析的数据“框”起来),然后调用分析功能(用鼠标点选菜单),按回车键就可在当前光标下面得到分析结果。虽然有些较复杂的统计分析需要读者输入参数,但系统会自动给以提示,如用户事先没有相关的经验,系统会给出缺省值,用户只管用鼠标点击“确定”按钮,就可以获得较优的分析结果;又如时间序列分析,系统以直观的图形方式让读者确定参数,只要读者稍有常识,无需牢记一大堆统计分析过程中的参数辨识准则,就可成为高水平的建模能手;再如在系统聚类分析中,聚类结果一次成“像”,给出的是非常直观的聚类谱系图等。系统的这些功能,好比一只傻瓜型照相机,使用时只需按快门。另一方面,DPS 系统在数据处理和模型模拟中又比较灵活,完全取决于读者的经验和知识水平。如数据行列的转换

功能、通用方差分析的广义线性模型的应用、自定义数学模型的拟合、逐步回归及逐步判别分析中  $F$  检验临界值的随意调整等等，使读者大有用武之地，挥洒自如。

应当指出，本书不是一本基础理论教材，而是一本不需要高深数理统计知识即可看懂的统计方法论专著。事实上，要把本书中每种统计方法的理论背景在这样一本书中完全交待清楚是不太可能的，因为有些章节本身就是一本书的内容。作者的主要目的是要把读者从繁琐的统计分析中解放出来，既要用最适合、最先进的统计分析技术为自己的研究工作服务，又不被它们复杂的数学公式和运算所困扰。如果本书及其 DPS 数据处理系统能使读者在数据处理和统计分析方面有所获益或帮助，作者的基本目的就达到了。

DPS 统计软件自 1997 年中国农业出版社第一次出版以来，已经历了 10 个不平凡的年头。在这 10 年中，国内外统计软件品种繁多，竞争十分剧烈，但由于有广大用户的大力支持，DPS 终于突破重围，在国内应用统计软件领域占有了一定地位。为此，我们对 DPS 广大用户以及关心和扶持 DPS 成长的各个领域的专家、学者表示衷心的感谢！

经过 4 年的努力，尽管 DPS 系统的功能又进一步增强，但离广大读者的要求还有较大差距。作者希望广大读者继续支持 DPS 数据处理系统这一国产统计分析软件，使之不断完善。在使用过程中如发现缺点和不足请及时反馈给作者，以便再次修订时改进。作者将不辜负广大读者的期望，努力工作，不断充实新内容。为方便广大读者、用户使用 DPS 数据处理系统，及时解答用户在应用中的技术问题，加快更新 DPS 软件步伐，使之日臻完善，特设立了为 DPS 统计软件提供技术支持的网站 <http://www.chinadps.net> 和 <http://www.statforum.com>。读者可通过访问该网站及时得到 DPS 使用技术咨询和软件的升级服务。

本书的出版得到国家自然科学基金(30370914)和国家“863”计划项目(2002AA243041)的资助。在本书修订和出版过程中，作者所在单位和科学出版社给予了大力支持；浙江大学原副校长程家安教授、湖南省农业厅李绍石研究员、中国水稻研究所胡国文研究员及已故的中国科学院庞雄飞院士对作者在 DPS 统计软件开发、升级，以及本书写作过程中给予了殷切关怀和鼓励；香港浸会大学方开泰教授、华东师范大学王静龙教授、成都中医药大学周仁郁教授、湖南中医学院刘明芝教授以及第四军医大学夏结来博士等给予大力帮助；研究生孙传恒、唐洁和沈爱华协助本书文稿的文字处理；浙江大学昆虫科学研究所的各位同事给作者工作以很大帮助，作者在此一并表示感谢。

作者

2006 年 3 月于杭州

# 目 录

序一	
序二	
第二版前言	
第一版前言	

## 第一篇 DPS<sup>®</sup>数据处理系统

第 1 章 DPS 系统简介	3
1.1 系统功能简介	3
1.2 DPS 系统的不同版本	5
1.3 系统运行环境与安装、使用	5
1.4 DPS 的基本操作	9
1.5 文本数值转换及字符串数值转换	11
1.6 数据行列转换及行列重排	12
1.7 分类变量的取值和编码	12
1.8 数据统计分析及其建模基本步骤	14
1.9 DPS 系统函数应用	16
参考文献	18
第 2 章 DPS 数据处理基础	19
2.1 数据基本参数计算	19
2.2 常用统计分布及 DPS 统计函数	23
2.3 正态性检验及参考值范围	28
2.4 Trimmed 及 Winsorized 均值	30
2.5 二项分布和 Poisson 分布的置信区间	31
2.6 混合分布参数估计	34
2.7 Pearson-III 型分布	39
2.8 异常值检验	41
2.9 图表处理	44
参考文献	47

## 第二篇 试验统计分析

第 3 章 一组样本和两组样本统计检验	51
3.1 显著性检验基本原理	51
3.2 平均数和总体差异检验	53
3.3 总体均值样本量估计	54

3.4	样本率和总体率的比较 .....	55
3.5	Poisson 分布的均数和总体比较 .....	57
3.6	两组样本均值差异 $t$ 检验 .....	59
3.7	小样本均值差异 Fisher 非参数检验 .....	62
3.8	Bonferroni 检验 .....	63
3.9	两组样本率差别检验 .....	65
3.10	两总体检验样本含量及功效估计 .....	69
3.11	概率模型拟合优度检验 .....	72
	参考文献 .....	74
<b>第 4 章</b>	<b>方差分析 .....</b>	<b>75</b>
4.1	方差分析基本原理和步骤 .....	75
4.2	单因素完全随机设计 .....	82
4.3	单因素随机区组设计 .....	87
4.4	系统分组(巢式)设计 .....	89
4.5	二因素(组内无重复)完全随机设计 .....	91
4.6	二因素完全随机设计 .....	93
4.7	二因素随机区组设计 .....	96
4.8	平衡不完全区组设计试验 .....	99
4.9	多因素试验设计 .....	102
4.10	裂区试验设计 .....	105
4.11	重复测量资料方差分析 .....	115
4.12	拉丁方设计 .....	122
4.13	随机区组实验的协方差分析 .....	123
	参考文献 .....	128
<b>第 5 章</b>	<b>一般线性模型 .....</b>	<b>130</b>
5.1	线性模型基本原理 .....	130
5.2	GLM 模型用户操作界面 .....	134
5.3	GLM 模型输出结果分析 .....	136
5.4	一般方差分析的 GLM 模型 .....	138
5.5	混合效应模型方差分析 .....	139
5.6	系统分组(或嵌套)设计 .....	141
5.7	裂区试验统计分析 .....	143
5.8	协方差分析 .....	145
5.9	数量化方法 I .....	147
	参考文献 .....	150
<b>第 6 章</b>	<b>分类数据列联表分析 .....</b>	<b>151</b>
6.1	列联表分析及卡方检验概述 .....	151

6.2	列联表的生成与分析	153
6.3	四格表分析	156
6.4	多层 $2 \times 2$ 表 Mental-Haenszel 检验	160
6.5	$R \times C$ 列联表卡方检验	162
6.6	单向有序 $R \times C$ 表统计检验	164
6.7	双向有序且属性不同的 $R \times C$ 表统计检验	166
6.8	McNemar 检验及 Kappa 检验	171
6.9	$2 \times C$ 表和多层 $2 \times C$ 表	174
6.10	配对病例——对照列联表分析	177
6.11	重复测定资料似然比卡方检验	181
	参考文献	183
<b>第 7 章</b>	<b>分类数据模型分析</b>	<b>184</b>
7.1	Logistic 回归	184
7.2	条件 Logistic 回归	187
7.3	多分类无序反应变量 Logistic 回归	190
7.4	多分类有序反应变量 Logistic 回归	192
7.5	Poisson 回归	194
7.6	对数线性模型	197
	参考文献	209
<b>第 8 章</b>	<b>非参数检验</b>	<b>210</b>
8.1	两样本配对符号检验	211
8.2	两样本配对 Wilcoxon 符号-秩检验	211
8.3	两样本 Wilcoxon 检验	213
8.4	Kruskal Wallis 检验	215
8.5	中位数检验	218
8.6	Jonckheere-Terpstra 检验	218
8.7	Friedman 检验	220
8.8	Kendall 协同系数检验	223
8.9	Cochran 检验	224
8.10	非参数回归分析	225
	参考文献	226
<b>第 9 章</b>	<b>圆形分布资料统计分析</b>	<b>227</b>
9.1	平均角及其假设检验	227
9.2	两个或多个样本平均角的比较	230
9.3	多个样本平均角的比较	233
9.4	圆-圆相关	234
9.5	圆-线相关	236



参考文献 .....237

**第 10 章 多因素优化设计与分析** ..... 238

10.1 正交试验统计分析 .....238

10.2 二次正交回归组合(中心复合)设计 .....246

10.3 Box-Behnken 设计 .....249

10.4 均匀试验设计 .....250

10.5 二次饱和 D-最优设计 .....257

10.6 二次正交旋转及二次通用组合实验统计分析 .....258

10.7 二次多项式回归模型 .....262

10.8 “3414” 测土配方施肥实验统计分析 .....269

参考文献 .....272

**第 11 章 混料试验设计与分析** ..... 273

11.1 单纯形格子设计 .....273

11.2 单纯形重心设计 .....274

11.3 有下界约束条件限制的混料试验设计 .....275

11.4 单纯形格子设计和单纯形重心设计操作示例 .....276

11.5 基于均匀设计表的混料试验设计 .....277

11.6 有上下限条件约束的混料试验设计 .....279

11.7 具附加线性约束的混料试验设计 .....283

11.8 混料试验分析的数学模型 .....284

11.9 混料试验数据的回归分析 .....294

11.10 偏最小二乘回归分析 .....297

参考文献 .....300

**第三篇 专业试验统计**

**第 12 章 数据包络分析和随机前沿面分析** ..... 303

12.1 生产效率分析基本原理 .....304

12.2 数据包络分析 CCR 模型 .....307

12.3 数据包络分析扩展模型 .....314

12.4 考虑价格因素时的 DEA 模型 .....321

12.5 面板数据的 Malmquist 指数 .....323

12.6 随机前沿面模型 .....328

12.7 面板数据随机前沿面模型 .....335

参考文献 .....341

**第 13 章 量表分析和顾客满意指数模型** ..... 342

13.1 项目分析 .....342

13.2 量表可信度分析 .....345