

● 统计数据分析与应用丛书

基于信息技术

的统计信息系统

薛薇 著

● 统计数据分析与应用丛书

C819

54

2007

基于信息技术

的统计信息系统

薛薇 著

图书在版编目 (CIP) 数据

基于信息技术的统计信息系统/薛薇著.

北京:中国人民大学出版社,2007

(统计数据分析与应用丛书)

ISBN 978-7-300-07899-1

I. 基…

II. 薛…

III. 统计-管理信息系统

IV. C819

中国版本图书馆 CIP 数据核字 (2007) 第 021916 号

统计数据分析与应用丛书

基于信息技术的统计信息系统

薛 薇 著

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511398 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com (人大教研网)		
经 销	新华书店		
印 刷	北京鑫丰华彩印有限公司		
规 格	170 mm×228 mm 16 开本	版 次	2007 年 3 月第 1 版
印 张	22.25 插页 1	印 次	2007 年 3 月第 1 次印刷
字 数	357 000	定 价	32.00 元

版权所有 侵权必究 印装差错 负责调换

总序

Prologue

《管子》中有一句话：不明于计数，而欲举大事，犹无舟楫而欲经于水险也。意思是说在不清楚具体数据的情况下想做大事，就如同没有桨的船只航行在激流险滩之中。

对于国家政府来说，所谓大事就是引导社会经济沿着正确的方向健康和谐地发展；对于企事业单位来说，所谓大事就是在经营管理中做出科学有效的决策从而在激烈的市场竞争中生存发展；对于科学研究来说，所谓大事就是透过事物外在的表象深入探索其内在的规律性；对于个人生活来说，所谓大事就是在人生成长和个人理财的重要关口做出明智合理的抉择。

这些都离不开数据，离不开数据分析，离不开统计应用。通过数据处理进行科学的定量分析是成大事者的基本方法和首要条件。

当前，中国面临着国际化大背景下转轨的经济体制和转型的社会环境的综合考验，瞬息万变的信息时代对国家、企事业单位和个人都提出了严峻的挑战，而作为“信息密集”的统计应用也将迎来快速发展的机遇期。

这是我们撰写这套“统计数据分析与应用丛书”的初衷，希望它能够从事统计应用的实际工作者提供及时有效的帮助，也能够为有志于进行数据分析的在校本科生和研究生打开一扇统计应用的大门。



《论语》中有一句话：工欲善其事，必先利其器。意思是说做事情要取得较好的成就，应当首先利用先进的工具或手段。

突飞猛进的信息技术已经将统计应用引领到一个崭新的水平，并渗透到统计数据处理的各个环节，同时深刻影响着统计工作的全貌。计算机化和网络化是统计应用的必然趋势，它使得统计应用摆脱了复杂公式和计算的羁绊，可以为更多人所理解和使用，所以毫不夸张地说，现代信息技术是解决统计实际问题、掌握统计分析算法、建立统计应用系统的必要条件。

这是我们撰写这套丛书的另外一个初衷，就是让更多的统计应用读者能够使用计算机等信息技术实现统计数据分析与应用的目标，同时也可以让更多的计算机应用读者能够了解一些统计应用的特征和方法。同时，近年来在西方发达国家，信息技术人才和统计应用人才一直排名在就业需求榜的前列，我们认为具备统计知识和计算机知识的复合型应用人才在未来将具有巨大的发展前景和明显的从业优势。

在这套丛书的编写过程中，我们注重引进当前统计应用和计算机应用的前沿技术和理论方法，结合在统计应用、科研与教学中的丰富案例和实际经验，着力突出以下特点：

1. 针对性。针对不同的读者群，由浅入深地展开统计应用的论述，读者也可以按照《基于 EXCEL 的统计应用》、《基于 SPSS 的数据分析》和《基于信息技术的统计信息系统》所构成的应用体系，不断提高统计应用能力和统计理论水平。

2. 可操作性。以计算机和网络等信息技术为基础，实现统计应用的各种需求，读者可以按照丛书指引的策略和具体的方法解决自己工作生活中的数据处理问题。

3. 通俗性。以案例说明原理，以应用解释算法，以发展脉络说明理论形成，努力体现深入浅出的结构安排和文字风格。

不断改革前进是我们这个时代的主旋律，从目前国内相关论著和教材情况看，丛书的上述特点也是许多统计应用学者和同仁所追求的共同方向，在此特别感谢中国人民大学统计学院领导的鼎力支持，感谢中国人民大学出版社各位编辑所付出的辛勤劳动。

统计作为数据处理的方法论，具有广泛的应用领域，而它根本的生命力也在于应用。我们很想为相关读者奉献一套具有一定理论高度，且具备一定



指导性和实战性的统计应用书籍，它应该以统计应用案例为主线，以计算机技术为实现工具，可以使普通经营管理人员、基层科研人员、高层决策者、一般数据处理工作者和高年级的高校学生们从中获益。

薛薇

2006年10月

于中国人民大学统计学院

前言

Preface

当今社会信息资源的意义毋庸置疑，统计作为信息密集型行业在政府部门、企事业单位和科研机构中发挥着至关重要的作用，而飞速发展的信息技术则将统计应用推进到一个崭新的时代。

在统计信息化总体方针的指导下，如何结合统计应用的特征、问题和需求，建立高效、完善、实用、易用的统计信息系统既是一个紧迫的任务，也是一个长期的目标。

根据近年来国内外统计信息系统的发展历程和实践经验，在理论方法和实现技术两个方面进行认真总结和全面分析，对于我国统计应用的健康发展和统计信息系统的有效开发具有关键作用。

从宏观视角分析，统计不仅仅是一个行业，它在不断超越分析算法的局限，着力打造一种全程化的数据处理方法；计算机和网络等信息技术也不仅仅是一个手段，它在不断超越机器设备的外壳，着力打造一种自动化的数据处理系统，而统计信息系统则将二者有机地包容在一起。

长期以来，计算机应用与统计应用相关性较强，但融合交流相对较弱；计算机科学与统计学交叉性颇多，但学习借鉴相对不足，这主要是由于跨专业的复合性人才相对较少，单向的研究成果和论述虽然并不缺乏，但



是往往让统计专业人员在理解计算机与网络技术上感到吃力，而计算机专业人员在理解统计理论和概念上也感到困难。本书希望在统计应用和计算机应用的结合面上深入一步，以交织的视野审视领域中共同的经验和成果，在它们之间搭起一座桥梁，通过一个双方均可接受的理论框架，为相关人员的实际工作、技术开发和学术研究提供有价值的参考。

本书尽力突出以下特点：

1. 通俗性。尽量避免使用专业性术语，以通俗易懂的人物和实例讲解概念，叙述理论。
2. 实用性。既关注理论方法的阐述，也重视应用技术的使用分析，同时强调二者的可实现性。
3. 综合性。将统计应用与信息技术有机结合，将理论与实践有机结合，既努力体现全面性，又避免繁杂琐碎。
4. 前沿性。利用最新理论和先进技术分析统计信息系统建设中的紧迫问题。

本书从统计应用入手，以信息技术和系统理论为基础，以数据库和数据仓库建设为核心，以统计数据处理过程为线索，以统计分析方法和数据挖掘技术为手段，以丰富的实际案例为素材安排全书结构，力争以深入浅出的方式和简明流畅的语言，为相关人员提供一个可读性较强的技术读本。

本书适合各级统计部门的统计应用人员和计算机统计信息系统开发应用人员使用，同时也可作为统计学专业高年级本科生和研究生的教材。

由于相关资料较少和作者水平所限，书中不妥之处敬请各位读者批评指正。

作者

统计系列书目

一、21 世纪统计学系列教材

书名	著者	出版时间	定价	ISBN 号	备注
《统计学》(第三版)	贾俊平 何晓群 金勇进	2007 年	33.00 元	7-300-07751-X	“十一五”国家级规划教材 教育部推荐教材
《〈统计学〉(第三版)》 学习指导书	贾俊平	2007 年	15.00 元	978-7-300-07901-1	
《统计学》教学案例与 教学项目汇编	贾俊平	2004 年	25.00 元	7-300-05444-7	
描述统计	贾俊平	2003 年	13.00 元	7-300-04576-6	
应用回归分析	何晓群 刘文卿	2001 年	19.00 元	7-300-03757-7	“十一五”国家级规划教材
应用时间序列分析	王 燕	2005 年	23.00 元 (含光盘)	7-300-06654-2	
应用随机过程	张 波	2001 年	14.00 元	7-300-03769-0	
多元统计分析	何晓群	2004 年	28.00 元	7-300-05402-1	
非参数统计	王 星	2005 年	30.00 元	7-300-06269-5	
抽样技术	金勇进 蒋 妍 李序颖	2002 年	28.00 元	7-300-04079-9	“十一五”国家级规划教材 预计 2007 年出第二版,并配学习指导书
市场调查方法与技术	简 明 金勇进 蒋 妍	2004 年	25.00 元	7-300-05443-9	
经济社会统计	高敏雪 李静萍	2003 年	35.00 元 (含光盘)	7-300-04598-7	
国民经济核算原理与 中国实践	高敏雪	2006 年	32.00 元	7-300-07139-2	“十一五”国家级规划教材
《国民经济核算原理 与中国实践》习题集	高敏雪	2006 年	15.00 元	7-300-07280-1	“十一五”国家级规划教材
寿险精算学	王晓军	2005 年	23.00 元	7-300-06454-X	
非寿险精算	肖争艳	2006 年	22.00 元	7-300-07422-7	
生存分析	彭 非 王 伟	2004 年	38.00 元	7-300-05956-2	
统计分析与 SPSS 的 应用	薛 薇	2001 年	20.00 元	7-300-03907-3	

二、统计学经典译丛

书名	著者	出版时间	定价	ISBN 号	备注
工程统计学(第 3 版)	道 格 拉 斯·C· 蒙哥马利	2005 年	59.00 元	7-300-06366-7	

书名	著者	出版时间	定价	ISBN 号	备注
商务统计学(第4版)	肯·布莱克	2006年	99.00元 (附光盘)	7-300-06867-7	
行为统计学基础(第9版)	理杰德·P·鲁尼恩				预计2007年4月出版
用数字证明:法律和诉讼中的实证方法	J. B. 温斯坦				预计2007年出版
律师统计学	迈克尔·O·芬克尔斯坦等				预计2007年出版

三、教育部高校工商管理类教学指导委员会双语教学推荐教材

书名	著者	出版时间	定价	ISBN 号	备注
商务统计学(第4版)	戴维·M·莱文	2006年	36元	7-300-06124-9	

四、统计数据分析与应用丛书

书名	著者	出版时间	定价	ISBN 号	备注
基于 EXCEL 的统计应用	薛薇 陈欢歌	2006年	29.00元	7-300-07660-2	本丛书是一套以统计应用案例为主线,以计算机技术为实现工具的统计应用书籍,适合普通经营管理人员、基层科研人员、高层决策者、一般数据处理工作者和高校学生们学习使用。
基于 SPSS 的数据分析	薛薇	2006年	35.00元	7-300-07660-2	
基于信息技术的统计信息系统	薛薇	2007年			

五、21 世纪保险精算系列教材

书名	著者	备注
利息理论	孟生旺等	本丛书共10本,5本为保险精算专业基础核心课程的教材,5本为配套的学习辅导书,含有大量习题,主要用于保险精算专业的学生学习使用,同时也可作为参加北美精算师、英国精算师及中国精算师考试的考生学习使用。本系列教材预计2007年全部出齐。
《利息理论》学习辅导书	黄向阳等	
风险理论	肖争艳	
《风险理论》学习辅导书	肖争艳	
寿险精算学	王燕等	
《寿险精算学》学习辅导书	王燕等	
非寿险精算学	孟生旺	
《非寿险精算学》学习辅导书	孟生旺	
精算中常用统计方法	黄向阳等	
《精算中常用统计方法》学习辅导书	黄向阳等	

出教材学术精品 育人文社科英才
中国人民大学出版社读者信息反馈表

尊敬的读者：

感谢您购买和使用中国人民大学出版社的图书，我们希望通过这张小小的反馈卡来获得您更多的建议和意见，以改进我们的工作，加强我们双方的沟通和联系。我们期待着能为更多的读者提供更多的好书。

请您填妥下表后，寄回或传真回复我们，对我们的支持我们不胜感激！

1. 您是从何种途径得知本书的：

☐书店 ☐网上 ☐报刊杂志 ☐朋友推荐

2. 您为什么决定购买本书：

☐工作需要 ☐学习参考 ☐对本书主题感兴趣

☐随便翻翻

3. 您对本书内容的评价是：

☐很好 ☐好 ☐一般 ☐差 ☐很差

4. 您在阅读本书的过程中有没有发现明显的专业及编校错误，如果有，它们是：

5. 您对哪些专业的图书信息比较感兴趣：

6. 如果方便，请提供您的个人信息，以便于我们和您联系（您的个人资料我们将严格保密）：

您供职的单位：

您教授的课程（教师填写）：

您的通信地址：

您的电子邮箱：

请联系我们：

电话：82501868-551, 541 62514162

传真：62514775

E-mail: rdcbsjg@crup.com.cn

通讯地址：北京市海淀区中关村大街甲59号文化大厦15层 100872

中国人民大学出版社工商管理出版事业部

内容简介

● 统计信息系统的建设是一个涉及计算机与统计理论、技术和应用实践的系统工程，如何从理论上把握系统全局，从技术上定位系统风格，从实践中获取建设经验，本书将为广大读者提供有效的帮助。

● 本书从生动的统计应用人物和案例出发，利用有效的信息技术，深入浅出地阐述了统计信息系统的实践经验、构造方法和开发理论；从统计应用人员的知识背景入手，注重通俗性、实用性、综合性和前沿性，在统计信息系统的理论、方法、技术和实践等方面都有引人入胜的分析与深入浅出的论述。

● 本书适合各种统计应用人员和数据分析人员使用，可作为计算机统计信息系统开发人员的技术读本，同时也可作为统计学专业高年级本科生和研究生的统计信息系统教材。

作者简介

● 薛薇，副教授，经济学博士，现任教于中国人民大学统计学院。主要从事统计信息系统和统计数据库的研究与开发，统计分析方法在经济与社会、管理领域的应用等。主持教育部规划项目4项，各部委及企事业委托项目20余项。代表作：《SPSS统计分析方法及应用》、《数据处理与统计分析软件》。

策划编辑 陈永凤
责任编辑 肖爱娣 陈永凤
版式设计 赵星华
封面设计 许 广

CONTENTS 目录

第 1 章 统计应用概述	1
1.1 统计应用的产生与发展	1
1.1.1 自发统计应用	2
1.1.2 描述统计应用	2
1.1.3 推断统计应用	7
1.1.4 现代统计应用	13
1.2 统计的含义	22
1.2.1 对统计学的理解	22
1.2.2 对统计数据的理解	26
1.3 统计应用领域	30
1.3.1 个人统计应用	30
1.3.2 企事业单位的统计应用	31
1.3.3 自然科学中的统计应用	33
1.3.4 人文社会科学中的统计应用	35
第 2 章 信息技术概述	39
2.1 计算机技术	40
2.1.1 什么是计算机	40
2.1.2 计算机的种类	41
2.1.3 计算机的发展	41
2.1.4 计算机的硬件	42
2.1.5 计算机的软件	47
2.1.6 计算机的应用领域	53
2.2 计算机网络技术	56



2.2.1	计算机网络的概念	56
2.2.2	计算机网络的分类	57
2.2.3	计算机网络的通信协议	60
2.3	计算机网络的工作模式	62
2.3.1	主机/终端模式	63
2.3.2	客户机/服务器模式	63
2.3.3	浏览器/服务器模式	65
2.3.4	对等模式	68
第3章	统计信息系统概述	69
3.1	数据与信息	69
3.1.1	对数据的理解	69
3.1.2	对信息的理解	70
3.1.3	信息的特征	72
3.1.4	信息的作用	73
3.2	系统与信息系统	75
3.2.1	系统的概念与性质	76
3.2.2	信息系统的概念和基本结构	78
3.2.3	信息系统的功能和层次	80
3.2.4	计算机信息系统开发的复杂性	85
3.2.5	计算机信息系统的经典开发方法	87
3.2.6	计算机信息系统的开发实践	89
3.3	统计信息系统	91
3.3.1	统计数据与统计信息系统	91
3.3.2	统计信息系统的组织	96
3.3.3	统计信息系统的资源	97
3.3.4	统计信息系统的结构	101
3.3.5	统计信息系统的特征	102
3.3.6	统计信息系统的作用	104
第4章	统计数据库	109
4.1	数据库的产生和基本概念	110
4.1.1	数据库的产生	110



4.1.2	数据库的相关概念	113
4.1.3	关系数据库	116
4.1.4	数据库应用系统和数据库的设计	117
4.2	统计数据库的概念与发展	123
4.2.1	从功能角度定义统计数据库	123
4.2.2	从处理对象角度定义统计数据库	125
4.3	统计台账的数据库存储	126
4.4	统计指标的数据库存储	129
4.4.1	统计指标的概念	130
4.4.2	统计指标的特征	131
4.4.3	统计指标的生成来源	133
4.4.4	统计指标的数据库存储模式	137
4.5	统计表的数据库存储	143
4.5.1	统计表的结构分析	143
4.5.2	统计表的数据存储模式	148
第5章	统计应用中的数据仓库技术	155
5.1	数据仓库的产生背景	156
5.1.1	信息孤岛与数据仓库	156
5.1.2	联机分析处理与数据仓库	160
5.2	什么是数据仓库	163
5.2.1	数据仓库的概念	163
5.2.2	数据仓库的特征	164
5.2.3	数据仓库的基本结构	165
5.3	数据仓库的关键技术	168
5.3.1	数据获取技术	168
5.3.2	多维数据的存储处理技术	170
5.3.3	元数据	180
5.3.4	数据集市	181
5.4	数据仓库建设	183
5.4.1	数据仓库建设的主要步骤	183
5.4.2	数据仓库建设的开发策略	185



5.5	数据仓库与统计数据库系统	186
5.5.1	从相关概念看数据仓库与统计数据库系统的 共通性	186
5.5.2	从信息化进程看数据仓库与统计数据库系统的 共通性	187
5.5.3	从数据环境建设看数据仓库与统计数据库系统的 共通性	189
5.6	数据仓库与统计数据库系统发展模式的探讨	191
5.6.1	信息资源管理	191
5.6.2	应用开发平台设计	193
5.6.3	XML 的启示	197
第 6 章	统计数据采集和处理技术	199
6.1	统计数据采集技术	199
6.1.1	数据的来源	199
6.1.2	计算机辅助电话调查系统	201
6.1.3	遥感技术	204
6.1.4	OCR 技术	206
6.2	统计数据汇总技术	207
6.2.1	统计台账的汇总技术	208
6.2.2	统计指标的汇总技术	215
6.3	地理信息系统和可视化技术	221
6.3.1	地理信息系统	221
6.3.2	可视化技术	226
第 7 章	统计信息系统中的数据分析技术	235
7.1	数据挖掘产生的背景	236
7.1.1	数据挖掘产生的背景：数据分析的实际需求	236
7.1.2	数据挖掘产生的背景：理论应用实践	238
7.2	什么是数据挖掘	241
7.2.1	数据挖掘的概念	241
7.2.2	数据挖掘能做什么	243
7.2.3	数据挖掘得到的知识形式	246
7.3	数据挖掘方法论	249