



国家社科基金重大项目

GUO JIA SHE KE JI JIN ZHONG DA XIANG MU

现代统计研究丛著 主编 贺铿

现代统计信息技术

田艳 高集荣 编著

XIAN DAI TONG JI XIN XI JI SHU

 中国统计出版社
China Statistics Press



国家社科基金重大项目

GUO JIA SHE KE JI JIN ZHONG DA XIANG MU

现代统计研究丛著 主编 贺铿

现代统计信息技术

田艳 高集荣 编著

XIAN DAI TONG JI XIN XI JI SHU

 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

现代统计信息技术/贺铿主编. 田艳, 高集荣编著.

—北京: 中国统计出版社, 2008. 9

(现代统计研究丛著)

ISBN 978-7-5037-5522-4

I. 现…

II. 贺… ①田… ②高…

III. 统计—信息技术

IV. C81

中国版本图书馆 CIP 数据核字(2008)第 125747 号

现代统计信息技术

主 编/贺 铿

作 者/田 艳 高集荣

责任编辑/姚 立

封面设计/杨燕超

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 57 号 邮政编码/100826

办公地址/北京市丰台区西三环南路甲 6 号

网 址/www.stats.gov.cn/tjshujia

电 话/邮购(010)63376907 书店(010)68783172

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/700×1000 mm 1/16

字 数/415 千字

印 张/27.25

印 数/1—5000 册

版 别/2008 年 12 月第 1 版

版 次/2008 年 12 月第 1 次印刷

书 号/ISBN 978-7-5037-5522-4/C·2162

定 价/60.00 元

中国统计版图书, 版权所有。侵权必究。

中国统计版图书, 如有印装错误, 本社发行部负责调换。

《现代统计研究丛著》编委会

主任委员：贺 铿

委 员：（按姓氏笔划排列）

文兼武 王振龙 冯士雍 田 艳

刘延年 邱 东 肖红叶 吴喜之

陈希孺 杨学义 林贤郁 胡 健

袁 卫 康 君 蒋 萍

总 序

《现代统计研究丛著》(以下简称为《丛著》)是国家社科基金重大课题的研究成果。

2004年,“现代统计研究”在全国哲学社会科学规划办立项以后,于当年9月10~12日召开专家座谈会,研讨了课题研究思路,课题组进行了初步分工。2005年2月23日又在北京召开课题组成员会,认真讨论了研究大纲。2006年5月6~9日在西安召开了课题结项与《现代统计研究丛著》出版座谈会,讨论并通过了各本著作的基本内容,决定了关于出版和组建《丛著》编委会的原则。编委会由课题组成员和课题承担单位西安统计研究院的主要负责人组成,主任委员由课题组的组长担任。《丛著》编辑出版中的重大问题由编委会集体研究决定。2007年12月22日在北京召开《丛著》主要作者座谈会,确定了各本专著的书名、署名及总序等问题。

“现代统计研究”课题立项时确定的目标是梳理统计学学术思想,阐释统计学的本质及学科边界,促进统计学的发展和应用。

究竟何谓统计学,历来学者立说纷纭。早在1869年,现代统计学的主要奠基人之一的A·Quêtelet在第七次国际统计大会上报告,关于统计学的定义已有180种之多。近几十年来,我国还经历了“两门统计学”和“大统计思想”的大讨论。这充分说明,为了促进统计学更好发展,对统计学的学术思想和研究范围进行梳理十分必要。

“现代统计研究”课题组成员对现代统计学的本质意义有大体一致的认识。经过研讨,其基本共识是:统计学是一门方法论科学,包

括工作性方法、技术性方法和工具性方法。工作性方法系指统计工作的制度、规范;技术性方法包括统计调查方法和统计分析方法;工具性方法在现代统计学科体系中系指与统计相关的现代信息技术,缩写为SIT。

最终形成的研究成果分为三类:第一类,阐释统计学的性质。着重从现代统计学发展的历史轨迹、学科体系和教育思想等方面进行梳理;第二类,介绍统计学的基本内容、前沿问题和发展趋势。主要介绍与社会经济问题相关的调查分析方法以及国民经济核算理论、指数理论、投入产出分析方法、经济计量学方法和统计信息技术(SIT)等内容;第三类,介绍国内外统计工作制度及规范。《丛著》共有8本10册专著。

《丛著》主编主要把握指导思想和研究路径。每本专著的基本内容由主编提出,经课题组成员集体研讨,作者决定。在具体学术观点上求同存异。学术观点表达及论证方法由作者(有多个作者的由第一作者)负责。我们认为,学术研究不可强求完全一致,各抒己见有利于学术发展。同时,我们在《丛著》中表达的观点在统计学界是否认同,也不是一厢情愿可以决定的。只有坚持“百花齐放,百家争鸣”的方针,才有可能真正推动现代统计学向前发展。

《丛著》付梓得到了多方面的关心和支持。在此我们对中国统计出版社、西安财经学院、九三学社机关和全国哲学社会科学规划领导小组办公室的领导及相关工作人员表示衷心感谢!

由于时间仓促和我本人水平所限,《丛著》难免存在缺点和错误。我们真诚地期待着读者的批评。

贺 铿

2008年3月

序

《现代统计信息技术》是 2004 年立项的国家社科基金重大课题——“现代统计研究”的成果之一，由课题承担单位西安财经学院田艳教授负责完成。

信息技术，简言之是生产和使用信息产品的技术。信息、材料和能源被认为是当今社会赖以生存的三大基础资源，信息产品在现代社会中具有十分重要的地位。现代信息技术以网络、通信和计算机技术为基础，包括信息采集技术、信息传输技术、信息存储技术和信息加工应用技术等。许多人认为，统计信息是社会信息的主体（一般认为，信息可分为自然信息、社会信息和人本身的信息三类），统计信息产品对认识社会、管理社会有极其重要的作用。随着现代信息技术的发展，统计工作现代化水平越来越高。现代信息技术与现代统计工作的融合逐渐形成了一门新的学科——统计信息技术（SIT, Statistical Information Technology）。《现代统计信息技术》对这个新学科的形成过程、基本内容和发展趋势作了系统研究和探索。我认为，本书不仅会对统计工作的现代化、信息化大有益处，而且会对 SIT 这个新的学科的建立和发展起着拓荒者的作用。

统计信息技术（SIT）这个名称是 2004 年 10 月在贵阳召开的“现代统计研究”课题组专家座谈会上提出来的。会上，专家们对 SIT 的内

涵、研究范围进行了热烈讨论。我们认为,SIT 是统计工作方法和统计分析方法与信息技术的融合。SIT 的发展与统计信息产品的生产过程——数据采集、整理、传输和存储以及对统计数据开发利用等统计工作有紧密联系,因此,是属于统计学、统计工作和信息技术的交叉学科。

《现代统计信息技术》的作者是信息技术方面的专家,对统计工作也有一定了解。本书依据我国统计信息化建设的实际水平撰写,反映了统计信息技术的发展状况,是一本对统计实际工作者、统计理论工作者和高等院校统计学专业的学生均有参考意义的好书。但是,对统计信息的采集技术和统计数据的加工与分析技术还有待在实践中进一步发展和完善。我衷心希望统计界的同仁们携手合作,努力为发展 SIT 这个新兴学科、为大统计学科的发展贡献力量!

贺 铨

2008 年 8 月于北京

前 言

本书所描述的统计信息技术是现代统计方法与现代信息技术的交叉融合,是在统计工作中收集、整理、储存、传输和开发应用统计信息所使用的相关信息技术,涉及的主要内容如下:

1. 统计网络直报技术和电子调查技术:统计网络直报技术是指通过统计网络直报平台,完成各种统计报表,完成集成数据采集、数据上报、数据挖掘、数据分析、图表分析的一体化工作;电子调查技术是指利用电话、计算机以及网络(包括 Internet)完成统计调查任务的相关技术。

2. 统计元数据库技术:元数据就是“数据的数据”,即关于数据的内容、覆盖范围、数据所有者、数据的提供方式、质量和其它特性的信息。统计元数据库是一个涵盖国家统计制度、报表、指标、填报目录、分组、分类标准和制度方法文件等内容的数据库。

3. 统计数据库技术:统计数据库是存储、管理、分析统计数据的数据库系统,是统计数据采集和统计数据传送的归宿,是统计数据存储和统计数据库管理的根本手段,是提供数据分析和数据服务的基础。

4. 统计数据仓库技术:统计数据仓库是基于现有的统计数据库上进行的建设开发,它着眼于有效抽取、综合、集成和挖掘已有数据资源,提供有价值的信息,为政府和社会服务。

5. 数据挖掘技术:数据挖掘技术是计算机技术、人工智能技术和统计技术等构成的一种新技术。数据挖掘源于统计分析,又不同于统计分析,是统计分析方法的扩展和延伸。

6. 网络及信息安全技术:统计信息化使传统统计工作转向网络方式,但网络的开放性、病毒侵袭、网络欺诈、信息污染、黑客攻击等问题也对统计信息安全形成严重威胁,因此网络及信息安全技术成为了统计信息技术的重要内容。

7. 统计电子政务:统计电子政务的建设是对传统统计工作的改造,无论是统计工作观念上、技术上、方法上还是流程上的改造都需要通过长期的一个一个信息化工程积累完成。深入进行统计信息资源的整合,完成网络与计算机等现代信息技术与统计管理和服务职能的无缝集成,优化统计工作流程、通过网络完成大量频繁的统计行政管理和日常事务,从而打破时间、空间及部门分割的制约,全方位地为社会及自身提供规范、高效、优质、透明、符合国际惯例的一体化管理和服务已成为现代统计的终极目标。

8. 统计信息系统设计与开发技术:现代社会经济与科学技术发展迅速,统计工作面临内外部环境复杂多变,因此,统计信息系统开发必须十分重视内、外部环境及发展趋势,合理规划系统目标、规模、功能和实施步骤,有效利用先进信息技术方法和现代科学管理成果,建设与我国经济社会发展环境相适应的统计信息系统。

本书系统地描述了统计信息技术涉及的主要内容,明确统计信息技术与信息技术的相关关系,提出我国必须进一步进行统计信息资源的整合,完成现代信息技术与统计管理和服务职能的无缝集成。本书希望能为各级统计部门的统计应用人员和统计信息系统开发人员提供参考,同时也为更多关注和研究统计信息技术的人员提供参考。

本书的第1章、第2章、第3章、第4章、第8章由西安财经学院田

艳教授撰写,第5章、第6章、第7章、第9章由中山大学高集荣副教授撰写。在撰写过程中得到西安财经学院、国家统计局计算中心及中国统计学会信息化分会许多专家的指导和帮助,中国统计学会信息化分会副会长胡帆就全书的体系框架结构以及相关内容提出了有益的建议和意见,特别是对第3章中的统计数据库建设部分及第8章中的统计信息资源整合部分给予了具体指导,胡帆研究员还对全书进行了审校,在此表示深深敬意和感谢。

由于信息技术的飞速发展以及现代统计发展的一些新要求,统计信息技术尚需深入研究的问题还有很多,如统计元数据库体系开发技术、统计数据质量控制技术、统计数据库实体的用户管理技术、信息安全新技术(如防水墙、等级保护、信息安全审计、网闸)在统计信息系统中的应用等。同时,对于新型的技术应用我们还需要研究新型的组织保障技术。这些都有待研究,融合到统计信息技术体系中。我们相信随着人们对统计信息技术的持续关注以及不断深入研究,统计信息技术体系将会不断完善。本书只是现代统计信息技术体系研究的一个初步成果,在此恳请同行专家批评指正。我们期待有更多的专家、学者关注统计信息技术的研究,共同完善统计信息技术体系并为我国现代统计建设提供有力的理论和技术支持。

作 者

2008年8月

目 录

第 1 章 概述	(1)
1.1 统计信息资源概述	(2)
1.2 统计信息概述	(4)
1.3 统计信息系统概述	(8)
1.4 现代统计信息技术	(13)
第 2 章 统计数据采集与整理技术	(30)
2.1 统计数据采集概述	(30)
2.2 统计调查方法	(36)
2.3 现代统计数据采集技术	(44)
2.4 统计数据的整理技术	(56)
2.5 数据采集处理平台简介	(61)
附录一 国家统计局 SARP2000 统计应用系统介绍	(62)
附录二 计算机辅助电话调查系统(FJS-CATI)功能简介	(64)
第 3 章 统计数据库技术	(67)
3.1 数据库技术概述	(67)
3.2 统计数据库概念	(87)
3.3 统计数据库技术	(94)
3.4 统计数据中心技术	(107)

第 4 章 数据仓库技术	(111)
4.1 数据仓库概述	(111)
4.2 数据仓库中的数据组织	(114)
4.3 数据仓库的体系结构	(117)
4.4 数据仓库的建立	(126)
4.5 统计数据仓库建设	(141)
附录 数据仓库主流厂商及产品	(146)
第 5 章 数据挖掘技术	(149)
5.1 数据挖掘的概述	(149)
5.2 数据挖掘的常用技术	(156)
5.3 数据挖掘的基本步骤	(161)
5.4 数据挖掘与数据分析技术	(167)
5.5 数据挖掘的现状与面临的问题	(171)
第 6 章 计算机网络技术	(176)
6.1 计算机网络的形成与发展	(176)
6.2 计算机网络的定义、功能及结构	(183)
6.3 计算机网络分类	(186)
6.4 计算机网络拓扑结构	(193)
6.5 计算机网络参考模型	(196)
6.6 网络组建技术	(203)
6.7 网络统计环境建设	(210)
第 7 章 网络信息安全技术	(214)
7.1 网络信息安全基本概念	(214)
7.2 信息加密技术	(224)
7.3 常用的网络安全技术	(228)
7.4 VPN 技术	(247)
7.5 计算机病毒防范技术	(250)

7.6 网络信息安全管理技术	(261)
附录 网络安全应用案例	(269)
第8章 电子政务与统计	(274)
8.1 电子政务概述	(274)
8.2 电子政务框架结构	(283)
8.3 电子政务的相关技术	(290)
8.4 统计信息化与统计电子政务	(297)
附录一 中共中央办公厅、国务院办公厅关于转发《国家信息化 领导小组关于我国电子政务建设指导意见》的通知 ...	(311)
附录二 中共中央办公厅、国务院办公厅关于印发《2006~2020 年国家信息化发展战略》的通知	(316)
附录三 国家统计局关于印发《“十一五”国家统计信息化建设 规划纲要》的通知	(329)
第9章 统计信息系统开发技术	(337)
9.1 统计信息系统开发涉及的基本问题	(337)
9.2 统计信息系统开发技术概述	(342)
9.3 统计信息系统开发策略	(352)
9.4 统计信息系统规划	(364)
9.5 统计信息系统分析及相关技术	(367)
9.6 统计信息系统设计及相关技术	(382)
9.7 统计信息系统构建、实现和运行支持相关技术	(403)
9.8 统计信息系统建设人员培训	(413)
参考文献	(417)

第 1 章

概 述

统计科学在 20 世纪的发展对社会经济发展与科技进步产生了巨大影响。它不仅研究抽象的数量特征,研究数据的收集、整理、分析和表示方法,同时也研究各个不同应用领域的具体数量关系,揭示各个应用领域的特殊规律。统计工作的目的在于获取、识别、处理、转移、传递、分析和应用数量化的信息,为认识事物、掌握规律、预测决策、科学研究和各项管理提供信息支持,统计工作的成果表现为统计产品,即以数据和图表为载体的能够描述客观现象及特定系统的运行状况的信息产品。数量信息是统计的本质和核心,统计的一切活动都是围绕信息而展开的,因此我们说统计科学的本质决定了它从诞生开始就无法与信息技术分离,今天信息技术飞速发展并以惊人的速度改变着整个社会的生活和生产方式,包括统计信息生产方式和统计工作方式,信息技术已经渗透到统计工作的各个环节,统计信息系统应用正在逐步成熟,统计信息化建设在不断完善,统计科学的发展以及统计信息化过程中现代信息技术对传统统计技术进行了极大补充,对传统工作方式也造成很大冲击,这也导致我们需要重新认识和研究许多新的问题、解决新的矛盾和冲突。

当今时代,信息技术的迅速发展和渗透,要求我们对统计方法做出新的拓展,也要求变革现行的传统统计工作模式,未来的统计必然是走统计信息系统和网络统计之路,我们认为统计信息系统将成为现代统计的基本模式,现代统计信息技术必然成为现代统计科学的重要分支,因此应该从系统化、体系化、规范化的角度研究现代统计信息技术。

1.1 统计信息资源概述

1.1.1 信息资源

对于信息资源的概念,自 20 世纪 80 年代始,国内外信息专家开始研究并给出了不同的定义,比较典型的观点有:

美国里克斯(Betty R. Ricks)和高(Kay F. Cow)在《信息资源管理》中指出,信息资源包括所有与信息的创造、采集、存储、检索、分配、利用、维护和控制有关的系统、程序、人力资源、组织结构、设备、用品和设置。

美国的霍顿(Forest W. Horten)认为,信息资源在英语中有单复数之分,其概念也有所不同。单数的信息资源(resource)指信息内容本身,复数的信息资源(resources)指各种信息工具,包括信息设备、信息用品、信息设施、信息工作者及其信息处理工具。

德国的斯特洛特·加龙省曼(K. A. Stroetmann)认为信息资源包括信息内容、信息系统和信息基础结构三部分:信息内容包括产生于信息服务或从外部信息源获取的信息,也包括与内容活动有关的理论和方法论信息、管理和操作信息与决策相关的信息,还包括与外部活动有关的交易信息、用户信息和市场信息;信息系统包括系统目标、操作人员、信息内容、硬件、内部规则等;信息基础设施是指一个组织的信息基础结构,它由各种可共享的数据库、计算机硬件设备、数据库管理系统和其他软件、局域网等构成。

我国图书馆与情报学家、信息资源管理专家孟广均教授在 1991 年提出,信息资源包括所有的记录、文件、设施、设备、人员、供给、系统和收集、存储、处理、传递信息所需的其他机器。

武汉大学信息资源研究中心研究员、武汉大学信息管理学院教授查先进认为,可从狭义和广义两个角度来阐述信息资源的概念。从狭义角度来说,信息资源是指人类社会经济活动中经过加工处理有序化并大量累计后的有用信息的集合,如科技信息、政策法规信息、社会发展信息、市场信息、金融信息等,都是信息资源的重要构成要素。从广义角度来看,信息资源是信息和它的生产者及信息技术集合。也就是说,信息资源由三部分构成:一是人类社会经济活动中经过加工处理有序化并大量积累后的有用信息的集合;二是为某种目

的而生产有用信息的信息生产者的集合；三是加工、处理和传递有用信息的信息技术的集合。

著名经济学家、我国数量经济学、信息经济学的创始人乌家培先生也从狭义与广义两个角度来看待信息资源，狭义的信息资源仅指信息内容本身，广义的理解是除信息内容本身外，还包括与其紧密相连的信息设备、信息人员、信息系统、信息网络等。

综合国内外的研究成果，把信息资源分为狭义的信息资源与广义的信息资源更具有代表性。

狭义的信息资源观点把信息资源等同于知识、资料 and 消息，即只是指信息本身的集合，无论信息资源是以声音、图形、图像等形式表达出来的，还是以文献、实物、数据库等载体记录下来的，其信息内容都是一样的，都是要经过加工处理的、对决策者有用的数据。准确地说，狭义的信息资源仅仅指信息内容，指信息本身或信息的集合。

广义的信息资源观点认为信息资源是一个贯穿于人类社会信息活动中从事信息生产、分配、交换、流通、消费的全过程的多要素集合，包括信息处理的对象——信息（数据），信息处理的设备——计算机等工具，信息处理的技术——网络、通信和计算机技术等信息技术手段，信息劳动者——信息专业人员。如信息生产人员、信息管理人员、信息分析人员、信息服务人员、信息传递人员等。

狭义的信息资源观点突出了信息本身这一信息资源的核心和实质。信息资源是一种经济资源，因为其中蕴含着的信息具有十分重要的经济功能，而信息生产者、信息技术与设备等信息处理要素只不过是信息资源开发利用的必要条件。

广义的信息资源观点把信息处理的各要素都纳入信息资源的范畴，更有利于全面、系统地把握信息资源的内涵。信息是构成信息资源的根本要素。人们开发利用信息资源的目的，就是为了充分发挥信息的效用，实现信息的价值。信息效用的发挥和信息价值的实现都是有条件的，信息的收集、处理、存储、传递和应用等都必须采用特定的技术手段即信息技术才能得以实施，信息的有效运动过程必须有特定的专业人员即信息人员才能对其加以控制和协调。信息、信息技术与设备和信息人员构成了完整的信息资源概念体系。