



全国统计教材编审委员会“十一五”规划教材

统计数据处理概论

★ 刘全 编著



中国统计出版社
China Statistics Press



全国统计教材编审委员会“十一五”规划教材

统计数据处理的概论

★ 刘全 编著



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

统计数据处理概论/刘全编著.

—北京:中国统计出版社,2008.7

ISBN 978-7-5037-5513-2

I. 统…

II. 刘…

III. 统计数据—数据处理—高等学校—教材

IV. 0212

中国版本图书馆 CIP 数据核字(2008)第 105400 号

统计数据处理概论

作 者/刘 全

责任编辑/陈悟朝 谢蕾蕾

装帧设计/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 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/787×1092mm 1/18

字 数/270 千字

印 张/15.75

印 数/1—2000 册

版 别/2008 年 7 月第 1 版

版 次/2008 年 7 月第 1 次印刷

书 号/ISBN 978-7-5037-5513-2 /O·68

定 价/30.00 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。

出版说明

“十一五”时期是继续深化教育改革,加强素质教育,努力建设有利于创新型科技人才成长的教育体系的关键时期。为了更好地培育统计创新型科技人才,适应统计教育的新形势,全国统计教材编审委员会制定了《“十一五”全国统计教材建设规划》(以下简称规划)。规划坚持“以人为本”的科学发展观,坚持统计教育与实践相结合,坚持统计教育同国际接轨,坚持培养创新型的统计人才的指导思想,编写符合国民经济发展需要和统计事业发展需要的统计教材。

这批教材是在深入分析统计教育形势和统计教材建设发展状况,总结多年来统计教材建设经验的基础上,本着以建设本科统计教材为主的方针,积极探索研究生层次的统计教材,力争使规划统计教材的编写做到层次分明,有针对性和实用性。建设精品教材,是编委会自成立以来就孜孜以求的目标。考虑到统计教材建设的实际情况,“十一五”期间,本科教材主要以修订为主,对以往规划统计教材中使用面广,得到广大教师和学生普遍认可的教材组织了修订。修订后的教材,淘汰了过时的内容和例子,增加了计算机操作和大量的案例,编写手法也做了一定的调整,在实用性、可操作性等方面有了较大的改进。

近年来,我国现代化建设快速发展,高等教育规模持续扩大,尤其是研究生教育规模的扩大,使得高等学校研究生统计教学工作面临着许多新情况、新问题,任务艰巨。因此,必须坚持科学发展观,在规模持续发展的同时,把握

高研究生统计教学质量放在突出的位置,培养全面发展的创新型的统计人才。教材是统计教学的载体,建设高质量的研究生层次的统计教材是统计教育发展的需要。因此,编委会在“十一五”期间对研究生的统计基础课教材做了些有益的探索。根据《规划》的要求,这批教材主要采取招标和邀请的方式组织有关院校的专家、学者编写。

值得特别提出的是,在这批教材中,有《非参数统计》、《概率论与数理统计》、《经济计量学教程》、《医学统计》、《应用时间序列分析》、《多元统计分析》、《统计学》7种教材入选国家教育部组织编写的“普通高等教育‘十一五’国家级规划教材”,更加充实和完善了“十一五”期间统计教材的建设。

为了便于教学和学习,这批教材里面包含了与之相配套的《学习指导与习题》,使得这批教材在编辑出版上形成了比较完整的体系。我们相信,这批教材的出版和发行,对于推动我国统计教育改革,加快我国统计教材体系和教材内容更新、改造的步伐,打造精品教材,都将起到积极的促进作用。

限于水平和经验,这批教材的编审、出版工作还会有缺点和不足,诚恳欢迎教材的使用单位、广大教师 and 同学们提出批评和建议。

全国统计教材编审委员会

2006年6月

前言

统计数据处理概论是研究社会经济统计工作中计算机统计数据处理原理、方法和技术的一门方法论科学,是现代统计学的基础内容。本书作为全国统计教材编审委员会“十一五”规划教学用书,曾受到重点推荐。

本书可作为本科院校和职业技术学院财经类专业的必修课和其他相关专业的选修课教学用书。也可以作为经济管理人员的培训教材以及各类成人教育的学习参考用书,又可以作为统计专业人员必备的工作指南。

本书力图充分借鉴国内外先进的研究成果,系统总结多年来在我国社会经济统计工作中统计数据处理的理论、方法和技术,籍以充实本书内容,完善体系结构。本书以社会经济统计工作中计算机统计数据处理的主要过程为主线,全面系统深入地阐述了计算机统计数据处理的理论、方法、系统设计及其实现技术。取材上力图反映当前社会经济统计工作中统计数据处理技术的发展水平和发展趋势,内容涵盖了社会经济统计工作中计算机统计数据处理的主要过程,主要包括:统计数据处理的基本概念,统计数据的组织及管理,统计数据的搜集、录入、审核、汇总、制图、分析等数据处理过程的基本概念、方法、技术及软件,网络统计调查系统及开发,网络统计报表系统及开发,统计数据库及数据仓库,统计数据挖掘,通用统计报表汇总系统软件 SARP,统计信息系统建设及管理。

为突出反映计算机统计数据处理的最新进展,本书重点介绍了网络统计调查、网络统计报表系统、统计数据库

及数据仓库、统计数据挖掘等方面的内容,并详细探讨了网络统计调查系统、网络统计报表系统的具体开发技术,给出了开发实例。这对读者学习和掌握网络统计调查系统、网络统计报表系统的编程有很好的指导作用。

本书主要由成都信息工程学院统计系刘全教授编著,成都信息工程学院统计系高燕同志负责了网络统计调查、网络统计报表系统程序开发技术的具体研究及编写工作。

本书在编写过程中得到了国家统计局教育中心,中国统计出版社,成都信息工程学院领导和同志们的指导及大力支持,在此一并表示诚挚的谢意。

统计数据处理概论是一个庞大复杂的体系,其理论和实践都在日新月异地发展,尽管作者以高度的敬业精神为本书的撰写付出了很大努力,但限于水平和时间,书中不足之处仍然在所难免,恳请读者批评指正。

刘 全

2008年5月于蓉城

第一章 统计数据处理概述	1
§ 1-1 数据、网络与现代统计	1
§ 1-2 计算机统计数据处理	3
§ 1-3 统计数据处理软件	7
第二章 统计数据的组织及管理	11
§ 2-1 统计数据的组织	11
§ 2-2 统计数据的管理	16
§ 2-3 统计数据的标准化处理	20
第三章 统计数据的搜集	22
§ 3-1 统计数据搜集的基本概念	22
§ 3-2 计算机数据搜集的主要方法	24
§ 3-3 统计数据搜集软件	26
第四章 网络统计调查系统及开发	28
§ 4-1 网络统计调查的基本概念	28
§ 4-2 网络统计调查软件的功能设计	32
§ 4-3 网络统计调查软件的开发	35
第五章 统计数据的录入	51
§ 5-1 数据录入的基本概念	51
§ 5-2 数据录入的基本过程	52
§ 5-3 数据录入过程的基本原理和方法	54
§ 5-4 数据录入软件的功能设计	57
第六章 统计数据的审核	61
§ 6-1 数据审核的基本概念	61
§ 6-2 数据审核的基本原理及方法	62
§ 6-3 数据审核的基本过程	65
§ 6-4 数据审核软件的功能设计	67
第七章 统计数据的汇总	69
§ 7-1 数据汇总的基本概念	69
§ 7-2 数据汇总的基本原理和方法	74

§ 7-3 数据汇总系统软件	82
第八章 网络统计报表系统及开发	88
§ 8-1 网络统计报表系统的基本概念	88
§ 8-2 网络统计报表系统的功能设计	92
§ 8-3 常见网络直报系统介绍	95
§ 8-4 网络统计报表系统的开发	99
第九章 统计数据的制图	105
§ 9-1 统计制图的基本概念	105
§ 9-2 统计制图的基本过程	109
§ 9-3 统计制图的基本原理和方法	111
§ 9-4 统计制图软件	115
第十章 统计数据的分析	117
§ 10-1 统计分析的基本概念	117
§ 10-2 统计分析的基本过程	119
§ 10-3 统计分析过程的基本原理及技术 要求	120
§ 10-4 统计分析软件	122
第十一章 统计数据库及数据仓库	127
§ 11-1 数据库的基本概念	127
§ 11-2 统计数据库	130
§ 11-3 统计数据仓库	135
§ 11-4 在线事务处理 OLTP 和在线分析处理 OLAP	141
第十二章 统计数据挖掘	144
§ 12-1 数据挖掘的基本概念	144
§ 12-2 数据挖掘的基本过程	147
§ 12-3 数据挖掘的基本方法和技术	149
§ 12-4 数据挖掘软件	151
第十三章 通用统计报表汇总系统 SARP	153
§ 13-1 SARP 系统概述	153
§ 13-2 系统基本约定及用户环境描述	156

目 录

§ 13-3	《SARP》系统加载	163
§ 13-4	专业字典、词典	164
§ 13-5	数据表描述及初始化	172
§ 13-6	数据录入	174
§ 13-7	数据检索	178
§ 13-8	计算及审核	179
§ 13-9	汇总	184
§ 13-10	制表描述及操作	207
§ 13-11	数据备份	221
第十四章	统计信息系统建设及管理	224
§ 14-1	我国统计信息化系统的建设与 发展	224
§ 14-2	我国统计信息系统建设的任务及 规划	225
§ 14-3	统计信息系统开发的方法与组织	229
§ 14-4	统计信息系统开发过程的实施	233
§ 14-5	统计信息系统建设的评价与运行 管理	238
参考文献		242

第一章

统计数据处理概述

统计学是一门跨社会科学、自然科学的交叉的科学,其学科性质决定了统计学与信息技术结合的重要性的对现代信息技术的依赖性。统计数据处理概论是研究社会经济统计工作中计算机统计数据处理原理、方法和技术的一门方法论科学,是现代统计学的基础内容。

§ 1-1 数据、网络与现代统计

一、数据与统计信息

数据是指用于记载或表达信息,并可以鉴别的物理符号。一般把它看成是事物属性的反映。数据的表现形式可以是语言、文字、符号、图形、声、光、磁、波等。信息则是数据接受者对数据加工后得到的有意义的东西,它对接受者的行为能产生影响,对决策具有价值。换言之,数据是信息的载体,信息则是数据的本质属性,数据和信息密不可分。正是由于数据和信息的密不可分性,所以“数据处理”人们也称为“信息处理”。

统计信息是社会经济信息的主体,它所反映的是社会、经济、资源、环境、人口、教育等方面的状况、过程和规律。统计信息的内容既可以是统计对象的状态、规模、水平等原始信息,也可以是其结构、特征、行为、功能等再生信息;既可以是对历史、现状的描述,也可以是对其未来发展的预测。统计信息是国家基本的信息资源,是国家进行宏观管理和科学决策的依据,是社会公众、国际社会了解国情国力、了解数字化中国的主要来源。

人类社会正处在信息时代,信息是促进社会、经济、科技、文化交流与发展的基本要素。如何利用现代化的手段对人们在社会实践中获取到的大量统计

数据资料进行加工、处理,从而提取出有用的统计信息,为促进国民经济又好又快发展提供可靠的依据,是开发和利用统计信息资源的目的所在。

二、网络与现代统计

计算机和现代通信技术极大地促进了统计体制、制度和方法的改革和发展。在网络时代,统计学被赋予了全新的含义,正成为一门新的、发展了的统计学。21世纪的统计学将是统计学与网络技术紧密结合的科学。

1. 网络改造着传统统计,提高了统计工作的质量

网络正改变着统计工作的方式。当前,统计数据的搜集、整理、分析、保存、传送和发布等都依赖于网络平台进行,这极大地提高了统计工作的质量。

通过计算机网络这个计算机与现代通信技术的平台,可以极大地提高统计信息的搜集、保存、整理、分析及传输的能力,为真正实现及时、高效、便捷、更好地提供各种高质量的统计产品提供了可能。

计算机网络技术可以大大提高统计信息产品应用的市场效率,有力地促进统计信息产品市场化水平。把数据用户与生产者直接相连,使统计信息能发送到世界各地的用户,并借助计算机辅助功能的作用,可以将以往那种单纯、枯燥难懂的表格、数字扩展为二维、三维图像和声音等多媒体表现形式,丰富了统计信息产品的表现形式。统计依托网络、计算机技术,使其表现形式艺术化、载体多样化,从而促进了统计信息产品价值和使用价值的提升,推动了统计信息产品社会化进程。

2. 网络丰富着统计的内涵,促进了统计的发展

网络在改变人们的生产方式同时,还造就了全新的统计客体——虚拟企业,促进了新的经济形态的产生——网络经济。所以,网络催生了许多新的统计客体,促使统计科学对原有客体重新分类,对研究范围、方法和技术的不断扩展等,极大地丰富了统计的内涵。可以说,没有现代信息技术就没有现代统计。

网络统计工作:是指运用计算机网络对社会经济现象的数量方面进行搜集、整理和分析的工作过程。

网络统计资料:是指以光盘形式或网上公布形式反映社会经济现象特征和规律的各项数字资料。

网络统计科学:是指如何利用计算机网络技术对社会经济现象的数量方面进行研究和如何对网络经济现象数量方面进行统计的方法论学科。

3. 网络统计

现代统计学的一个重要标志就是网络统计的产生与发展。广义理解,网络统计既包括传统统计工作的网络化和技术现代化,利用互联网、计算机和网络技术来对社会经济现象的数量方面和数量特征进行统计调查、整理、分析和信息输出、发布等一系列统计活动;同时也包括对网络经济现象数量方面和数量特征进行统计。

计算机和网络通信技术,极大地丰富了统计学的学科内容 and 应用范围,促进了统计技术的快速发展。计算机统计数据处理正是在此基础上发展起来的一门方法论科学。

§ 1-2 计算机统计数据处理

计算机的应用极大地提高了人们处理数据的能力。开展社会经济统计工作,其主要任务就是对大量的反映社会经济现象数量方面的数据进行处理,揭示其内在规律。因此,只有提高统计人员的计算机数据处理能力,才能最大限度地利用统计信息资源,迎接日益增长的信息挑战。

一、计算机统计数据处理的概念

统计学包含了两大方面内涵丰富的统计技术:一是获取信息的技术,即为了经济有效地获取数据资料,研究如何科学地进行观察、调查或实验;二是提炼信息的技术,即如何运用所获取的数据资料,用统计分析方法对实际问题的规律性及其因果关系进行科学的推断。以上两方面的内容都涉及到大量的数据处理问题,即统计数据处理技术问题。

计算机统计数据处理就是在统计学基本原理和方法指导下,借助计算机及网络通讯技术手段,研究如何客观、全面、迅速地获取、提炼统计信息,高效率地解决统计工作中数据处理问题的一门综合性技术。

“数据处理”本身是一个很古老而新鲜的话题,是伴随人类文明始终的一项活动。如“结绳记事”就是一项最原始的“数据处理”活动。初期“统计活动”早在几千年前就有了,但“统计学”的诞生却只有三百多年的历史,更主要的是计算机及网络技术的发明才仅仅几十年。所以,真正高效率的数据处理是在电子计算机诞生后的近几十年才出现的。

统计数据处理是统计工作中最基础性,也是最重要的任务。当今社会,统计数据处理技术日新月异,统计数据处理技术发展水平的高低决定着统计现代化的程度。计算机统计数据处理在现代统计中发挥越来越重要的作用。

二、统计数据处理的基本过程

统计资料是指反映大量社会经济现象的数量特征和规律性的数字资料。统计工作是搜集、整理和分析统计资料并进行推论的工作。可见,统计工作基本任务就是对统计资料的处理。

1. 一般过程

根据统计学理论,我们知道统计数据处理过程一般分为:统计资料搜集、统计资料整理和统计资料分析及预测等三个阶段。其中,统计资料搜集的目的是获取与研究目的相关的大量原始、基层数据;整理阶段的处理对象是大量统计个体,是对个体的逐项处理;而分析研究阶段的数据处理对象是统计总体,是在汇总基础上对总体特征及规律的研究。

2. 计算机统计数据处理过程

计算机数据处理过程是以计算机为中心的数据处理过程。它和手工、机械数据处理方式一样,也包括三个阶段,即:数据搜集、数据加工、数据输出。



图 1.2.1 计算机统计数据处理过程

(1)数据处理的第一阶段就是数据的搜集。搜集是指按一定的目的要求获取原始统计资料的过程。数据处理质量在很大程度上取决于搜集到的原始数据的完整性和真实性。统计数据搜集的手段很多,它既可以手工方式进行,也可借助互联网和计算机以自动方式进行。统计数据搜集方式有:统计报表制度、各种普查、抽样调查、典型调查、重点调查等。

数据搜集核心是:数据的识别、采集和组织,即有目标的选择那些能反映事物运动过程属性的数据,将其及时、完整搜集起来并按一定层次加以组织。只有这样才能清晰地反映事物活动全过程。

(2)数据加工是数据处理的核⼼,一般包括数据编码、录入、审核、分组、汇总、计算、存贮、传输等内容。

在计算机中,广义的数据是指能被计算机处理的各种对象。为了使计算机能识别不同的数据并加以迅速处理,应对数据进行编码,也就是通过一定的数字、符号来表达所得到的客观事实。

数据资料经编码分类后,由录入人员将其输入计算机。在数据组织中,很重要的工作是将数据输入到计算机,并按照规范的要求组成各种形式的数据库,再由数据库管理系统对其进行管理。

数据审核是数据搜集的重要环节。所谓数据审核,就是对全部输入数据,设想其可能出现的差错,并设法对其进行检验。对查出的有错误的数据,考虑改正和补救的方法等。

统计汇总就是按照数据的特征(数字或字母)或按照数据的某种属性,将原始数据分成若干类型,并得出合计结果的处理过程。统计资料的汇总是统计资料整理的中心内容。

数据处理过程必然要涉及到大量数据的存储问题。计算机为存贮数据提供了各种存贮设备,如硬盘、软盘、光盘、磁带、闪存等。

计算机进行数据传输可以分为机内传输和机外传输。机内传输主要是通过计算机的硬件和软件进行传输,如从光盘传到内存,从磁带传到内存再传到磁盘等;此外,还有一种传输是脱离主机直接在外部设备之间交换数据,这称为脱机传输。机外传输主要是通过通讯线路(共用或专用线路)进行的,一般称为联机系统。在联机系统中数据发生一般是随机的,需要及时进行处理和输出,这种方式也称为链接传输。

(3)数据输出是将加工后所得到的信息,通过各种方式输出,以满足统计业务工作或计算机管理工作的需要。输出方式很多,如打印、显示、图样、声音等。

现代化管理要求对统计数据进行快速、适时、有效地处理,并且可以在部门、地区之间,甚至在国家之间进行快速数据通讯,从而能及时为国民经济和领导部门的决策服务。这些要求,必须使用快捷、安全、准确的输出系统才能得以实现。

目前,统计数据处理一般都是指计算机统计数据处理。如下图 1.2.2 是一个统计报表汇总系统的数据处理过程:

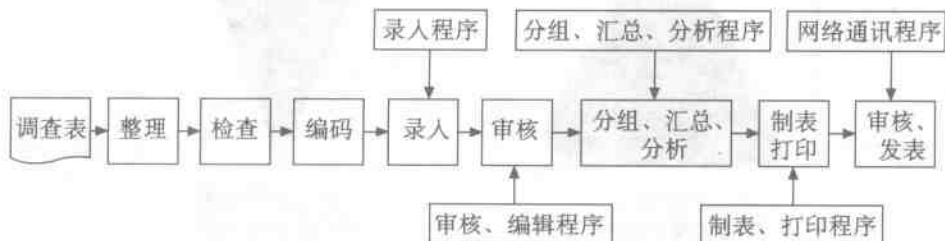


图 1.2.2 统计报表汇总系统的数据处理过程

其中：“录入”是对大量采集的数据进行整理并按一定的格式输入（或光电扫描、语音录入、磁介质数据格式转换等）到磁介质（硬盘、软盘、磁带、光盘等）中去；“审核”是对录入（或传输入）的数据进行正确性判断，保证得到正确的数据；“分组”是对录入的数据按照统计的要求，对数据进行分类、合并等操作；“制表”是对分组汇总后的数据按照统计表的格式加上标题、题注、表侧、表顶、表体、脚注等内容。由图可见，以上各环节都是在相应的计算机程序下完成的。

3. 计算机统计数据处理的特点

一般而言，统计数据整理属于“海量数据处理”范畴，其特点是：“量大、高效、质高、数据利用率高”。

“量大”，统计需处理的数据一般都是海量数据，同时，其数据处理需动用的人力、物力也很大，数据处理的程序也较复杂；

“高效”，统计数据处理一般都有严格的时间限制，必须在规定的时间内完成任务。同时，由于处理的数据量大，还必须考虑到高效的数据录入，高速的数据计算，计算机内、外部存储设备等的有效利用情况等；

“质高”，统计数据处理必须做到科学、准确、及时，特别是反映社会经济现象的重要统计数据；

“数据利用率高”，统计数据的处理还必须考虑数据的共享和进一步深加工的需要，使处理后的数据能够适应不同层次、不同用户的需要，以利于信息资源的有效利用。

计算机统计数据处理，为“获取信息资源的最大化”提供了可能。图 1.2.3 是手工、计算机数据处理对信息增减的对比示意图。

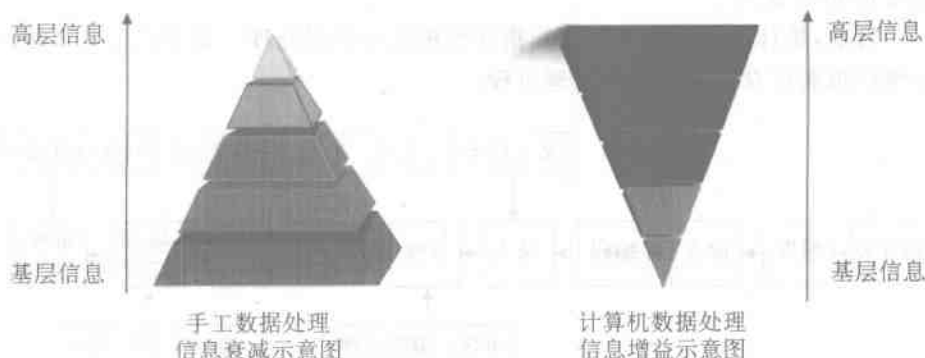


图 1.2.3 统计数据处理的信息增减示意图

利用计算机,可以高效、广泛搜集到各种统计数据,并对大量采集到的数据进行整理,按一定的格式存储起来以利于数据共享;可以在数据加工阶段发挥计算机数据处理的优势,对原始统计资料进行各种分组、集成和计算,尽可能多的挖掘出各种有用信息;可以在数据提供阶段根据用户的不同需要而输出各种结果,满足个性化的需求。与传统手工统计数据处理方式相比较,计算机统计数据处理方式在获取统计信息资源最大化方面有着明显的优势。

4. 计算机统计数据处理的方式

计算机统计数据处理方式通常有集中批处理和联机实时处理系统。

(1)集中批处理是统计部门广泛使用的一种处理方式。集中批处理是把一批统计数据处理任务,包括统计数据及任务要求一起交给计算中心,统计数据处理专业人员按系统具备的数据处理能力,对这批数据处理任务的性质和特点做出合理的安排,使计算机系统用最短的时间保质、保量地完成这批任务。在集中批处理进行的过程中,不需要其他用户干预。由于批处理是把作业汇集在一起进行处理,因此,处理效率比较高。

采用集中批处理方式,数据源就在机房或可将数据源汇集后在机房处理,且处理结果不必要实时返回,这种集中批处理方式也是通常所指的机房处理方式。

(2)实时处理是在数据产生的同时,立即进行处理,并在规定的时间内返回处理的结果。实时处理已广泛应用在网上直报、网络调查等统计工作中,也广泛应用于飞机票、车船票预约系统和银行业务、交通控制、过程控制、生产管理、仓库管理等领域。

采用实时处理,数据源一般是分散的,而且处理结果实时响应。这时系统必然是多终端的联机系统,终端分布在各数据源,通过通信线路连接终端和主机,这就是通常所指的联机处理方式。一般的实时事物处理系统和经营管理系统都采用这一方式。应该指出,在联机终端上也可以提交批处理作业,而且联机系统的基础是计算机及通信网络。

§ 1-3 统计数据处理软件

在统计数据处理中,对数据的处理必须要按照一定的统计规则、模式和相应的统计方法进行。统计数据处理软件就是随着计算机技术的发展应运而生的一种计算机应用软件,它是计算机技术与统计技术相结合的产物。