



新世纪高校经济学管理学核心课教辅用书

统计应用软件

—— EXCEL 和 SAS

朱建中 邵建利 编著

 上海财经大学出版社

新世纪高校经济学管理学核心课教辅用书

统计应用软件

—— EXCEL 和 SAS

朱建中 邵建利 编著

上海财经大学出版社

图书在版编目(CIP)数据

统计应用软件——EXCEL和SAS/朱建中,邵建利编著. —上海:
上海财经大学出版社,2002.3
新世纪高校经济学管理学核心课教辅用书

ISBN 7-81049-715-4/TP·13

I. 统… II. ①朱… ②邵… III. ①表处理软件, EXCEL-应用-统计
分析②统计分析-应用软件, SAS IV. C812

中国版本图书馆 CIP 数据核字(2002)第 011866 号

责任编辑 何苏湘
封面设计 优典工作室

TONGJI YINGYONG RUANJIAN

统计应用软件

——EXCEL和SAS

朱建中 邵建利 编著

上海财经大学出版社出版发行
(上海市武东路321号乙 邮编200434)

网 址: <http://www.sufep.com>

电子邮件: webmaster@sufep.com

全国新华书店经销

上海第二教育学院印刷厂印刷

上海市印刷七厂一分厂装订

2002年4月第1版 2002年4月第1次印刷

890mm×1240mm 1/32 8.875印张 255千字

印数:0 001—4 000 定价:17.00元

前 言

现代化社会评价一个企业是否一流,常从企业在尖端生产技术、先进生产设备和专业优秀人才等方面拥有程度上加以判别。进入 21 世纪后,随着经营管理方法中信息技术的不断融入,企业有否强力有效的信息处理分析决策系统,也成了企业现代化水准的又一评判标志。

在信息处理分析决策系统的众多构成要素中,最为重要的是 DC (Data & Computer)。进一步讲,要使信息处理分析决策系统正常运行,既能充分挖掘数据的能量,又能极大程度地发挥计算机工作的威力,其核心方法,当论专事信息数据采集、处理和分析的现代统计应用方法莫属了。

无可辩驳的事实是,现代统计应用方法已经被越来越多地运用于社会、经济、科技等一切领域,而且开始发挥巨大而不可估量的影响和作用。然而,统计学在普通人面前却仍似云中之山岳、海上之蜃楼,会产生朦胧不知所以然的感觉。何种原因使得统计学令人生畏? 何种原因造成统计学不能“利其器”? 究其根本,烦琐的演算操作程式,堆砌组合的大量数学公式,可能是制约统计学普及应用的一大障碍。

有鉴于此,本书试图绕开数学公式甚至不用复杂的编程,来强调统计学方法和计算机的应用结合,以求解决演算过程之艰难;强调适应普通大众基本的计算机应用能力,以求解决普及推广之困惑;强调既附和

“阳春白雪”又顾及“下里巴人”的品味,以求解决应用者能各取所需的问题。基于上述三个方面的思考,总结笔者多年来通过统计学教学中计算机应用获得的实践经验,精心汇集制作了这套应用计算机统计软件进行数据处理和分析的程序,将其提供给所有兴趣的朋友们。不要求你的统计学功底如何,不在乎你的计算机水平怎样,相信在这本书中你能找到兴趣所在。其中 EXCEL 统计数据处理分析程序可供非专业统计人员使用,而 SAS 统计数据处理分析程序则可供专业统计人员使用。

本书的苦心在于让普通人了解统计学究竟是什么,知道怎样去做统计数据的处理和分析,而不必涉及那些深奥的数学公式和经历枯燥的演算过程。应用者如果借助这些小小的程序,通过工作、学习、生活中的实际事例去演练统计数据的处理分析技术,相信会慢慢地为统计学方法的强大生命力所折服,也会因利用计算机作数据处理分析的无穷魅力而着迷。

正如极少数的人才能深谙电子原理的奥秘,而绝大多数的普通人都能使用电子产品这个简单道理一样。统计学理论和方法的研究毕竟是少数人的事,然而让大多数人能够使用先进的统计方法进行统计数据的处理分析却是当今世界的一件大事。如果企业的经理会用,办公室的文员会用,甚至连小学生也在用;大家能学,大家能用,用得方便,用得见效,统计数据应用的技术就会得到极大程度的普及,而统计的作用和地位也将随之得到提高。此外,类似于少量开发专业型的照相机,大量开发普及型的“傻瓜”照相机的策略一样,注意开发出界面友好、简单易用的“傻瓜”型统计数据处理分析应用软件可能是一件现实工作中更受欢迎的事情。本书介绍 EXCEL 活动电子工作表的函数公式组合技术、SAS 软件的菜单驱动应用就是这样的一种尝试。当然,我们的努力仅仅是一种探索,“抛砖引玉”,但愿能引来更多更好而且应用范围更加广泛的应用软件的问世。

本书第一部分“统计数据处理分析软件 EXCEL”由朱建中编写(包括全部 EXCEL 计算程序的设计),第二部分“统计分析系统 SAS”

由邵建利编写(包括全部 SAS 计算程序的设计),前言和引言部分由朱建中撰写,最后由朱建中对本书进行统稿。本书在编著的过程中,得到上海财经大学统计学系徐国祥教授、刘汉良教授和其他同仁的热情支持和帮助,SAS 统计数据处理分析部分获得上海宝信软件股份有限公司嵇晓先生、SAS 软件(上海)有限公司曹建静先生等 SAS 应用研究同行们的支持,谨此一并致以谢意。同时还要感谢上海财经大学出版社何苏湘女士的鼎力相助,使本书得以尽快与读者见面。

由于编著者水平有限,书中的错误与缺点难免,恳请读者和专家斧正。

编著者
2002 年 2 月

目 录

前 言 (1)

引言 统计数据处理方法及其计算机应用 (1)

第一节 统计数据处理分析方法 (1)

第二节 统计数据处理分析方法的应用软件 (4)

第一部分 统计数据处理分析软件 EXCEL

第一章 EXCEL 的统计数据处理分析功能概述 (11)

第一节 EXCEL 数据分析工具库宏程序的开启和操作 (11)

第二节 EXCEL 的统计函数及其使用方法 (14)

第二章 描述统计的 EXCEL 应用 (27)

第一节 统计数据的分组和整理 (27)

第二节 统计数据分布特征值的计算 (30)

第三章 概率基础	(33)
第一节 常用概率计算表和概率曲线图的编制	(33)
第二节 全概率公式和贝叶斯公式的应用	(50)
第四章 参数估计和假设检验	(55)
第一节 参数估计	(55)
第二节 假设检验	(61)
第五章 回归和相关分析	(68)
第一节 EXCEL 一元线性回归分析	(68)
第二节 相关分析	(74)
第六章 非参数统计	(79)
第一节 非参数 χ^2 检验	(79)
第二节 列联表独立性非参数检验与正态分布拟合优度检验	(83)
第三节 符号检验和均匀分布检验	(92)
第七章 时间数列分析	(102)
第一节 时间数列数据处理技术	(102)
第二节 时间数列数据的季节变动分析	(115)
第八章 统计指数	(118)
第一节 综合指数的编制和分析	(118)
第二节 多因素指数体系分析	(127)

第二部分 统计分析系统 SAS

第九章 统计分析系统 SAS 总论	(135)
第一节 概述	(135)
第二节 SAS 软件的操作	(138)
第十章 描述性统计分析	(160)
第一节 图形	(161)
第二节 描述性统计量的计算	(173)
第十一章 置信估计与假设检验	(178)
第一节 置信区间	(178)
第二节 假设检验	(184)
第三节 方差分析	(194)
第十二章 非参数统计	(202)
第一节 χ^2 检验	(202)
第二节 威尔科克森带符号的等级检验	(208)
第三节 曼-惠特尼 U 检验	(210)
第四节 多个样本的检验	(211)
第十三章 相关与回归	(215)
第一节 相关系数计算	(215)
第二节 简单线性回归	(220)
第三节 多元线性回归模型	(226)
第四节 非线性回归模型	(229)

目次 (CONTENTS) (1)

第十四章 时间数列	(234)
第一节 SAS/ETS 模块	(234)
第二节 时间数列预测系统	(235)
第三节 时间数列预测实例	(237)
附录一 SAS 软件编程的有关内容	(246)
附录二 SAS/BASE、STAT 中的几个主要 PROC	(267)
参考文献	(271)

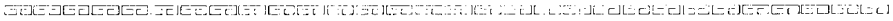
统计数据处理方法及其计算机应用

第一节 统计数据处理分析方法

统计数据处理分析,是指根据掌握的统计数据资料,运用统计学数据处理分析方法和计算机应用技术,采取定性定量相结合手段,对被研究对象的数量表现进行技术处理和分析研究的统计活动,它也是在统计设计、统计调查和统计整理基础上进行的最后阶段的统计工作。通过统计数据处理分析能够帮助人们正确地认识并掌握客观事物,能够揭示被研究事物数量表现的本质及其发展趋势,能够给决策者提供有效的决策依据。统计数据处理分析与统计工作前三个阶段一样,除了具有客观性、变异性和通用性等显著特点外,在计算方法和实施操作处理的技术方面,还具有创造性、灵活性、概括性等特点。

一、客观性

统计数据处理分析方法面对的是各种各样的统计数据,统计数据的全面、准确和完整是其真实性表现的生命线。因此,应该站在总体的角度认真审核所获得的统计数据,使其保持客观真实、全面完整。在处理和析这些数据时必须自始至终地维护客观性的原则,保证处理结果不至于失真。



二、变异性

统计数据处理分析的对象一般比较复杂,应该根据其变异性特点开展分析活动。首先,要处理的数据是大量的,数据之间的关系又比较错综复杂,只有抓住统计数据变异的特征加以处理分析,才能够深刻揭露事物的本质关系。另外,统计数据的处理分析过程通常比较复杂,统计分析的方法有简有繁,而且可以从各个角度、各个层面进行分析,只有扣紧分析对象基本的变异性表现进行静态、动态比较分析,才能够以简明扼要的方法有效地反映事物的状况和趋势。此外,统计数据的处理分析结果有时不便立即作出评价,需要运用多种统计分析方法加以比较验证,通常可以使用多个分析方法、多种统计参数进行反复比较,从多个分析结果的区别比较中选择并确定出一个最为有效的结论。

三、通用性

统计数据的处理分析方法是一种通用的方法,具有广泛应用领域。它可以作为统计、金融、贸易、企业管理等经济问题分析的工具,也可以在自然科学研究方面乃至社会文化生活活动的分析方面发挥积极作用。以往运用传统的统计分析方法进行手工处理,由于计算公式及其运算过于复杂,大大限制了其应用的范围。但是,巧妙地应用计算机统计分析软件可以避免许多手工或计算器的反复运算,甚至变繁琐耗时的计算过程为举手之劳。因此,采用计算机统计分析软件进行数据处理分析是普及应用统计分析方法的重要途径,也是体现统计数据处理方法通用性特点的基本条件之一。

四、创造性

统计数据处理分析不仅仅是一项周密细致的数量关系测定活动,尤为重要的是,它可以在认识不断深化的基础上创造数据挖掘的分析成果。虽然从统计活动的设计、调查、整理等环节看,时刻都要遵循客观真实的原则,然而作为形成数据分析成果的最终阶段,统计数据的处

理分析应该充分体现创造性,即在对数据深入开发挖掘的基础上创造出一个个既符合事物现实特征又能揭示发展规律的有用的评价结论或决策方案。同时,在计算机统计分析软件的应用上也应该充分体现创造性,针对处理分析对象的数据特点创造出一个个实用的处理程序,成为可以举一反三而得到扩展应用的有效分析工具。

五、灵活性

进行统计数据处理分析,在确定分析目标、取用对象数据资料以及选择计算机数据处理软件等方面应该体现较大的灵活性。从分析目标的要求出发,既能进行一般意义的常规性分析评价,也能在某一个专题上作出深刻的研究解析;既可以根据静态的空间分布数据资料进行处理分析,也可以根据动态的时间序列数据资料进行处理分析;既可以采用比较复杂的计算机数据处理软件进行计算处理,也可以采用相对比较简单计算机数据处理软件进行计算处理。

六、概括性

统计数据处理分析活动是建立在统计设计、统计调查和统计整理基础上的,而且,作为其处理分析依据的数据资料通常来源于直接的统计调查和整理,或者来源于间接的统计调查和整理,因此,统计数据的处理分析活动能概括所有统计工作的内容,也能综合地反映整个统计工作的成果。此外,从统计数据处理分析活动的计算技术角度看,无论是空间分布数据资料的处理结果,还是时间序列数据资料的处理结果,通常都可以用少数的统计量加以概括说明,使人们能够一目了然地认清事物的本质特征。

第二节 统计数据处理分析方法的应用软件

一、统计数据处理分析应用的计算机软件简介

统计数据处理分析方法的应用软件,根据其是否依附或独立于计算机的某种操作程序的不同情形,大致可以分为附属型统计应用分析软件和独立型统计应用分析软件两大类。附属型统计应用分析软件,是指依附于某种计算机操作程序的数据处理分析软件,如 OFFICE 等办公自动化应用软件中的 EXCEL 和 LOTUS 1-2-3 等数据处理分析的电子表格应用软件。独立型统计应用分析软件,是指能独立安装在 DOS 或 WINDOWS 操作环境下的数据处理分析应用软件,如社会科学方面的统计应用软件 SPSS(Statistical Program for Social Sciences)、统计分析系统应用软件 SAS(Statistical Analysis System)、生物医学统计应用软件 BMDP(Biomedical Computer Program)、数据处理组合应用软件 IMPS(美国普查局开发的 CNTRY、CONCOR 和 CENTS 三个软件的组合)等。

下面我们将选择中文 EXCEL 2000 作为附属型统计应用分析软件的代表和选择英文版 SAS 6.12 作为独立型统计应用分析软件的代表进行一般的介绍,并在书中以后的章节里结合统计数据处理分析的实际例子来进行更加具体的讲解,进行操作使用的练习和解析。

(一)统计数据处理分析的附属型软件 EXCEL 2000

EXCEL 2000 中文版是美国 Microsoft 公司开发的电子表格软件。它不仅具有强大的电子表格处理功能,而且附带有内容丰富的统计数据处理功能(统计函数和统计数据分析宏)。EXCEL 2000 中文版的强大功能主要表现在以下几个方面:具有非常友好的用户界面,可以直接通过各种菜单的选项进行某种特定程式的计算机数据处理;其工作表简单明了、易学易懂,特别是通过单元格地址的引用,可以把表格中所有数据之间的关系表达出来;EXCEL 2000 中文版大约有 300 多个工

作表函数,在工作表中引入EXCEL 2000的公式和函数后,数据的计算处理功能将得到大的扩展和增强,基本上可以满足有关数学、财务、统计等方面的数据计算处理要求;EXCEL 2000具有自动绘制数据统计图和绘制其他普通图形的功能,能够不加定制方便地为你绘制各种形式的平面图形;更具有特色的是,EXCEL 2000还带有一个几乎囊括所有统计数据处理分析方法,包括描述统计和推断统计在内的分析工具宏,通过菜单管理形式直接进行各种统计数据的处理分析,能用最直观的方式给出分析结果的统计量;此外,通过EXCEL 2000的宏扩展功能和自嵌的VBA(Visual Basic for Application)指令等的运用,普通人员也能够掌握有关宏的编写技术和宏的使用方法,即学会如何使用EXCEL的Visual Basic(简称VB)工具进行编程,这使得非专业(指非计算机或相关专业)人员能够轻松地根据自己的需要作编程条件下的统计数据处理分析,大大提高非专业人员统计数据处理分析的水平。

由于EXCEL 2000的统计函数和统计数据分析宏附带在OFFICE 2000办公自动化应用软件中,不需要单独安装,因此,所有安装有OFFICE 2000的正在使用的计算机都能够直接调用EXCEL 2000。只要懂得使用EXCEL 2000,再稍加一些统计学方面的知识培训,任何人员都能够胜任使用EXCEL 2000进行统计数据处理分析工作,使自己的工作成果增添上定量分析的光彩。

(二)统计数据处理分析的独立型软件 SAS

SAS,即统计分析系统(Statistical Analysis System)的英文缩写。最早的SAS系统于1966年由美国北卡罗莱纳州(North Carolina)州立大学所研制。1976年成立商业性质的美国SAS公司,主要开展对SAS系统的维护、开发、销售和培训等。1985年SAS公司首先推出了可以在个人电脑上使用的SAS/PC(6.02)版本。以后版本更新很快,功能也不断增加。目前已推出适用于多种操作系统的最新版本SAS/PC(6.12和8.20)版本。

SAS系统是处于世界领先地位的计算机信息处理系统,它具有完备的数据存取、管理、分析和显示功能。在数据处理和统计分析的领域

内,SAS系统已经被大多数专业人员所运用,目前已经成为国际上广泛使用的标准统计应用软件系统。

SAS系统的最大特点在于,它把数据管理和数据分析有机地结合为一体。具体地说,它具有以下几方面的主要特点:

1. 使用灵活方便、数据处理功能齐全

SAS系统为所有需要进行数据处理分析的非计算机专业人员提供了一种易学易用、完整可靠的软件系统。其使用操作简单方便,用户只需把要解决的问题,用SAS语言(近乎自然英语的非过程语言)表达出来,组成SAS程序,提交给SAS系统就可以解决你提出的问题。执行的情况和输出结果都可以在屏幕上显示出来。

2. SAS语言是编程能力强且简洁易学的非过程语言

SAS语言是SAS系统的基础,是用户与系统对话的语言。SAS系统综合了超过150个函数和大量的有关编程语句,形成了一套可用于各种各样数据处理分析的程序语言编制规则。SAS语言是一种非过程语言(第四代语言)。其特点是用户不必告诉SAS“怎样做”,只需告诉它你要“做什么”就行了。

3. SAS系统把数据处理与统计分析融为一体

SAS程序的结构由两个基本步骤任意组合而成。DATA步用于加工处理数据;PROC步用于分析数据和编写分析报告。SAS系统克服了通常软件或偏重于数据管理功能或偏重于统计分析的弊病,把数据管理功能与统计分析功能有机地结合在一起,还能够对所存贮的数据连续地进行各种统计分析。某个过程产生的输出数据,可以作为另一过程的输入数据(必要时用DATA步进行处理)做进一步地分析。

4. 适用性强、应用面广

SAS系统适用于任何经验的人员(初学者或有经验用户);SAS系统可满足从生产第一线到管理部门各方面人员对信息的需求(SAS提供多种界面供不同人员使用)。SAS系统可以访问各种数据资源。包括各种数据库生成的数据文件(如DBF文件),其他软件(如SPSS)生成的数据文件,高级语言生成的数据文件等。SAS系统可用来解决自

然科学和社会科学各个领域中的各种问题。

SAS 统计数据处理分析软件的主要功能有：数据管理，统计图形和表格绘制，推断统计，统计检验，相关回归分析，方差分析，聚类分析，因子分析，主成分分析和时间数列分析等。

因此，SAS 统计数据处理分析系统较适用于专业统计内容的高级分析对象。

二、统计数据处理分析软件的基本使用方式

计算机是根据软件所能识别的一条条命令进行工作的。在统计数据处理分析的过程中只有输入正确的运行命令后，计算机才能按照有关的指令去完成统计数据处理分析的计算过程，从而得到符合要求的处理分析结论。为了保证程序的运行有序和准确无误，同时也为了能尽量采用简单易用的操作方法输入运行命令，必须视处理分析对象的具体情况合理地选择指令的输入方式。通常统计数据处理分析软件的指令输入方式有以下几种形式：

1. 直接对话方式

某些统计数据处理分析软件通常备有一套自身特有的操作命令，用户可以选择其中适宜的操作命令，逐条输入，一步步地完成有关的数据处理分析步骤，这就是人机直接对话的运行命令输入方式。人机直接对话方式的基本特点是用户逐条输入标准命令后由计算机逐条判别，如输入的命令不规范或错误则由计算机显示出错信息，等待用户更正，如正确则执行该命令并显示出该步骤的处理分析结果，同时转入到下一个步骤，如此重复，直到完成全部处理分析。

2. 批命令复合方式

根据统计数据处理分析的过程要求，将运行软件所认定的一批标准命令复合后预先输入到一个程序文件中，然后再调用该程序文件，通过一次性启动，由计算机自动逐条顺序地执行该程序文件中的各条命令，自动地完成统计数据处理分析的全部过程。

3. 菜单对话方式