

现代实用统计 与计算机应用

XIANDAI SHIYONGTONGJI YU
JISUANJI YINGYONG

骆克任 韩效宥·著

立信会计出版社



现代实用统计与计算机应用

骆克任 韩效宥 著

立信会计出版社出版发行

(上海中山西路 2230 号)

邮政编码 200233

新华书店经销

立信会计常熟市印刷联营厂印刷

开本 850×1168 毫米 1/32 印张 10 插页 2 字数 241,000

1997 年 11 月第 1 版 1997 年 11 月第 1 次印刷

印数 1—3,000

ISBN7-5429-0527-9/F·0489

定价: 16.00 元

前 言

随着社会的迅速发展,我们已进入了信息爆炸的时代。呈现在我们面前的精彩世界,随时产生着各种各样的社会信息,这些信息不仅种类繁多、数量庞大,而且变化很快。在我们进行社会科学研究,编制社会与经济计划,进行市场调研、企业管理和营销决策时,都需要根据辩证唯物论的观点,经过定量研究大量有关的社会经济信息数据,从事物的运动中分析事物之间的关系、矛盾和发展变化,从而认识事物发展的现状和发展的趋势。从统计研究角度讲,就是在研究社会现象时,根据数理统计原理和方法,透过搜集汇总的统计数字和指标,分析、发现事物运动的本质和变化规律。现代统计研究方法为上述工作提供了强有力的科技手段。它可以帮助我们多快好省地收集各类社会和经济的统计信息,并对其进行去粗取精、去伪存真、由此及彼、由表及里的工作(例如研究方案的制定、调查的策划、问卷的设计、调查的执行、数据的录入、统计和定量分析与预测,甚至包括与撰写调查研究报告有关的诸方面的现代科学技术方法),使我们能准确地搜集和显示社会客观情况,并对发生的社会现象的总体进行概括性的推论,找出带有普遍性的社会现象规律。另外,由于信息的爆炸和社会的发展,逼迫统计分析运用相应的现代计算机统计辅助手段,来提高工作效率。为了适应上述的迫切要求,本书试图将现代实用统计分析技术及定量研究方法,与计算机的辅助统计手段溶为一体,来简明扼要和系统地加以介绍,避免读者费力地去单独寻找和学习有关的社会经济学、统计学和计算机知识。作者希望通过此书,使更多的社会研究者和

调查统计工作者的日常工作能进一步科学化和规范化,从而为我国的社会经济的持续发展作出贡献。

书中特别注意以下两个方面:一、深入浅出地介绍,增强实用性和可操作性,避免烦琐的数理推导过程,使读者在没有坚实的数理统计知识和扎实的计算机功底前,即可学会在遇到研究统计问题时,能够一步步地解决之;至于那些有兴趣学习更深奥的理论知识的人士,则可以进一步另选专门的书籍。二、书中注意介绍目前国内社会经济统计书籍里缺少的社会研究过程(例如规范的科学研究原则和步骤)及定量分析方法(例如注重讨论不可以直接计量,只能用间接方法来反映的技术。这些方法常用于社会和市场经济中的很多方面,如对大众的看法或态度等进行计量)。

根据理论联系实际的原则,作者强调理解和应用书中的这些技术,而不仅是阅读和死记硬背。书中每个部分都包括着统计知识及用于解决重要的社会和经济研究课题的方法论方面的推理和逐步分析。各部分的例题和图表,皆鼓励读者积极地思考和了解有关统计和定量研究方法问题的不同阐述过程。各部分的练习,能帮助读者实际应用讨论过和举例说明过的方法,以便进一步掌握和巩固所学知识。例题和练习使读者在有计算机或没有计算机的情况下,都能进行统计和方法论的学习。当然,没有或不熟悉计算机统计软件的部分读者,做一些有关计算机的练习,将是非常有益的。对于已经掌握了一定的计算机统计程序的读者,也会通过对现代统计方法及其两种著名的计算机统计分析软件包的学习,而更加大有可为。

本书的每个部分,包含着对统计知识和定量研究方法的某个重要内容的阐述、举例和练习。作者努力将通常包含在有关统计和定量研究方法方面的基本内容,与国际上流行使用的著名计算机统计分析软件相结合,以期各类人员都能够同时简捷地掌握统计分析和计算机的相关技术。书中介绍的社会科学统计软件包 SPSS

(Statistical Program for Social Sciences)/PC+版和 WINDOWS 版及统计分析系统软件包 SAS(Statistical Analysis System)/PC, 将支持各种统计和定量研究方法的使用。这两个软件虽然都是用于统计分析研究的著名软件,但是 SPSS 偏重于社会科学研究方面的统计分析,SAS 则在分析理论应用方面更为严谨,数据处理功能也较强。然而,用这两个软件统计分析数据,其分析理论和算法都是得到国际认可的。

通过此书,作者计划达到下列目的:

1. 组织简要的合乎逻辑的关于基础统计知识和定量研究方法的讨论。

2. 通过清楚的有意义的例子,使读者理解计算机统计程序及其如何运用。

3. 举例说明如何解释研究报告中的统计内容和其他定量分析结果。

4. 给出与章节内容直接有关的练习,使读者有机会去实际进行统计和做些定量研究。

5. 使计算机用户和非计算机用户都能从本书的课文和练习中获益。

6. 书中提供了某贫困地区的抽样调查数据集(以下简称 SQ 数据集)。该数据集与教材、例题和练习紧密配合,容易使用,并有实际意义。在社会调查课程及整个教学过程中,SQ 数据集都充当重要角色。SQ 数据集容许使用本书的教师进行讨论,并用它艺术地陈述所进行的研究。这个 SQ 数据集在全书有关基础统计及定量研究过程的讨论、例题和练习中,将作为实际有效的用具。通过附录 1 的说明,可以清楚了解该数据集的内容。重要的是 SQ 数据集为个人计算机的 SPSS 和 SAS 统计分析程序,提供了必不可少的实用数据,使计算机的用法与统计学和方法论的教学溶为一体。

使用这本书的读者,将能更深入地了解统计知识和定量研究

方法,并简捷地掌握相应的计算机应用技术。

由于下列理由,本书新的教学法与传统的方法相比,将使读者获益更多。

1. 与传统书籍中常见的冗长讨论相比,读者能更好地理解统计和定量研究方法讨论中的那些短小、简明的观点。

2. 多数读者将乐意进行结合实际统计工作和研究方法的练习。

3. 书中提供的有意义的变量及容易使用的数据集,将能引起读者更大的兴趣。

4. 没有计算机知识的读者,只要按提示的指示去做,即可容易地在 IBM PC 计算机或兼容机上运行 SPSS 或 SAS 软件包。

5. 与仅仅是阅读和死记硬背统计和计量方法的某些观点相比,完成练习中的笔头和上机作业,可以使读者学会并牢记更多的知识。

由于热心于学习社会经济统计分析方法和计算机辅助研究方法的人群在我国迅速扩大,作者相信这本书将受到欢迎和重视。当然,教授们乐于进行这门教学,主要是因为统计学和研究技术将成为分析、理解社会的有效工具。至于那些畏惧或不喜欢统计定量研究方法课程的教授也会发现,花费一些时间在这门课程上是可行的和有价值的。

章节内容的组织框架是基于作者进行多年社会科学研究和有关统计学课程教学的经验,并主要参考《BASIC SOCIAL STATISTICS AND QUANTITATIVE RESEARCH METHODS — A Computer Assisted Introduction》(James W. Grimm & Paul R. ozniak, esternKentacky University, 1990)一书而设立的。书中内容安排如下:首先在“引导”部分介绍计算机的基础知识,为学习下面各章的计算机应用内容打基础。对于基本的数学常识,作者认为每个研究人员都应该具备,所以没有必要在此书中介绍。随后书中

各章的编排是：第一至三章的论题是给读者介绍统计与定量研究的基本步骤和目的及统计描述。第四至五章讨论交叉分组列表、双变量的分析及假设的提出。第六至九章讨论各种定量研究方法，每种定量方法都有相应的统计实例的应用。第十至十六章提出了许多统计法则和必要的统计步骤。定量研究人员应通过这些法则和步骤对数据资料进行分析、理解及深入的推理，然后得出完善的解释。本书在内容上和结构上都可以满足一门既包括统计又包括定量研究方法的课程的需要。然而，它又是足够灵活的，每一个特定章节，都既可以用于统计学，又可以用于定量研究方法的课程。因为每一个章节中的内容，都自成体系，包括内容、例题和练习（这些材料既可以用于非上机的学习活动，又可以用于上机的学习活动）。作者鼓励读者用不同的方法使用这本书，把它当作有广泛适用性的书籍，而不是固定不变的蓝图。

作者衷心希望，读者通过对本书的学习能提高对必要的社会与经济研究过程的理解，了解统计技术和定量研究方法是如何共同用于一个研究过程的，且能够掌握相应的计算机的应用，并用它来协助研究的进行，从而在工作中变得如虎添翼。

骆克任

1996年1月8日

目 录

引导 计算机基础知识	1
一、计算机的基本结构	1
二、统计软件包	2
三、SPSS 与 SAS 统计软件包的功能介绍	3
四、SPSS 与 SAS 软件包的使用	13
 第一章 变量及研究过程	21
一、变量的类型	21
二、变量在研究中的应用	23
三、变量及其研究步骤	25
练习一	28
 第二章 单变量的频数分布	29
一、单变量频率分布表的构成	29
二、频数分布表及计算机软件的应用	32
练习二	37
 第三章 单变量频数分布的统计计量和描述	39
一、比例、百分比和比率	39
二、集中趋势的计量	41
三、离中趋势的计量	49
四、单变量频数分布计量及计算机软件的应用	53

五、频数分布的图形描述及计算机软件的应用	60
练习三	63
第四章 双变量的频数分布:表及其统计过程	65
一、名义双变量的分布	65
二、用名义双变量表检验假设及计算机软件的应用	68
三、有序双变量的分布	79
四、用有序双变量表检验假设及计算机软件的应用	80
练习四	97
第五章 假设的形成与检验	100
一、假设的概念	100
二、双变量假设	102
三、理论假设与工作假设	103
四、假设的检验及计算机软件的应用	104
练习五	111
第六章 计量	113
一、概念化单方面的概念	113
二、可操作化单方面的概念	114
三、构造指标	115
四、指标可靠性和有效性的判断	115
练习六	118
第七章 复杂概念的指数构成及解释	121
一、指数的构成	121
二、指数体系中的指数项的保留与去除	125
三、构造作为变量的指数及计算机软件的应用	129

四、有关计数(SCALES)的简短讨论	131
练习七	133
第八章 抽样过程	135
一、非概率抽样	135
二、概率抽样	137
练习八	145
第九章 调查研究	147
一、分析的个案	147
二、错误的推断	148
三、组织的影响	149
四、调查研究设计的类型	150
五、数据收集技术	152
六、高质量的问题和问卷的设计准则	155
七、全面回答的比率	160
八、调查研究的道德准则	162
练习九	164
第十章 推断统计的基础	168
一、正态曲线	168
二、正态曲线的应用	169
三、中心极限定理	176
四、均值置信区间的估计	178
五、比例的置信区间的估计	181
六、样本量的确定	182
七、计算机软件的应用	184
练习十	185

第十一章 推断统计:显著性检验	187
一、显著性检验的基础	187
二、单样本比例的 Z 检验	192
三、t 分布	193
四、单样本均值的 t 检验	194
五、双样本比例的 Z 检验	195
六、双样本均值的 t 检验及计算机软件的应用	196
七、前/后或成对样本的 t 检验及计算机软件的应用	201
八、方差的 F 检验	203
九、两个以上样本均值的单因子方差分析检验及计算机 软件的应用	205
练习十一	210
第十二章 精选的非参数统计	214
一、斯皮尔曼等级顺序相关	214
二、卡方拟合优度检验及计算机软件的应用	216
三、曼-惠特尼 U 检验及计算机软件的应用	219
四、威尔科克森配对正负号等级检验	221
五、克罗斯考尔-瓦里斯 H 检验及计算机软件的应用	223
练习十二	226
第十三章 双区间/比率变量的关联	229
一、线性回归	229
二、皮尔逊相关系数	232
三、r 的显著性检验	234
四、预测及其误差	236
五、计算机软件的应用	237
六、应注意的问题	240

练习十三	241
第十四章 双变量表的详述	244
一、因果关系	244
二、因果关系的判断标准	247
三、过程和结果的详述	248
四、计算机软件的应用与分析	256
练习十四	265
第十五章 多元线性回归和相关分析	268
一、介绍多元线性回归和相关分析	268
二、贝塔系数或贝塔权数	270
三、回归统计的显著性检验及计算机软件的应用 与分析	271
练习十五	283
第十六章 非线性回归分析	285
一、非线性回归模型的拟合	285
二、曲线类型的选择	289
三、计算机软件的应用	291
练习十六	293
【附录 1】 SQ 数据集介绍	295
【附录 2】 SPSS 的分类命令	296
【附录 3】 统计表格	299
3.1 Z—得分和正态曲线下的面积	299
3.2 t 的评判值	300
3.3 F 的评判值	302

3.4 卡方评判值	304
3.5 随机数表	305
【附录 4】 参考书目	307

计算机基础知识

社会的发展,导致了信息量的猛增,为了更便于研究社会现象,当今的研究人员越来越感到,以往笔头和算盘一类的统计工具已远远不能适应统计工作的需要。为了解决这个问题,越来越多的研究人员想要或正在求助计算机这个现代工具。对此情况,本书将结合介绍现代统计分析方法,引导大家掌握一些有关的计算机辅助工作方法。在讲解正文之前,我们有必要先了解一些有关的计算机的基础知识。

一、计算机的基本结构

计算机系统是由硬件和软件两部分组成,缺一不可。硬件指计算机的物理结构;软件指为了运行、管理和维修计算机所编制的各种程序和程序的有关说明资料。

硬件系统可以分为中央处理单元(CPU)、存储器、输入设备和输出设备四个部分。中央处理单元由运算器和控制器组成,它是计算机的核心。存储器是计算机的存储部分,用来存储各种数据和程序。存储器又可分为内存和外存:内存是电路板上的一种电子芯片;外存则是磁盘、光盘、磁带等。输入设备是用于输入数据和程序的,常用的有键盘、鼠标器、磁盘驱动器、光电读入器等。输出设备用于输出数据结果等,常用的有显示器、打印机、绘图仪和磁盘驱动器等。以上四部分加上电源和一些附加电路,并通过总线连接,就构成计算机硬件系统。人们通常将装在机箱中

的中央处理单元和存储器称为主机，将输入和输出设备称为外部设备。

计算机不会自己去解决问题，而是人们通过使用计算机软件来让计算机硬件去解决问题的。软件主要有系统软件、计算机语言和应用软件。系统软件用于管理和控制计算机资源，组织计算机的工作流程，如用户要通过 DOS 磁盘操作系统来使用计算机。计算机语言是表达人的意志的程序语言，计算机语言经历了机器语言、汇编语言、高级语言（如 BASIC、FORTRAN、PASCAL、COBOL、C 等等），直到现在开发的第四代程序设计语言。应用软件是已经编好的用于各种实际目的的程序。那种已形成标准化、模块化的解决不同类型问题的组合应用程序被称为软件包，如目前流行的字处理软件包 WPS、WORD，数据库管理软件包 FOXPRO 和 ORACLE，以及本书中介绍的统计软件包 SPSS 和 SAS 等等。

二、统计软件包

研究人员有时候会因为所研究问题的特殊复杂性，去自己编制一些特别的应用程序来解决问题。但在多数情况下，他要完成的统计将是一些标准的和通用性的工作，这时借助于一种权威可靠的统计软件包将是明智的。统计软件包是针对各种不同统计任务而编制的专项应用程序的集合。它为适应统计手段现代化的要求而产生。这些统计软件包的功能全面，且比较独立。统计软件包一般由数据和文件管理系统、统计命令及运行操作控制系统三部分组成。三者联系密切，但作用不同。要统计的数据、文件以及其结果，需经过数据和文件管理系统来转换，再输入和输出。统计功能要靠命令程序来组织，而数据和管理以及统计命令的执行，都要由运行操作控制系统来完成。

当今已有许多统计软件包。国外比较权威和著名的有社会科

学统计软件包 SPSS (Statistical Program for Social Sciences)、统计分析系统软件包 SAS (Statistical Analysis System)、生物医学统计软件包 BMDP (Biomedical Computer Program)、数据处理组合软件包 IMPS (美国普查局开发的 CENTRY、CONCOR 和 CENTS 三个软件的组合)等;国内有关系式数据分析系统软件包 RDAS (Relational Data Analysis System)。下面我们将选择介绍 SPSS 和 SAS 软件包(微机版),并在书中结合统计分析来讲解和使用。

三、SPSS 和 SAS 统计软件包的功能介绍

在利用计算机进行统计分析应用之前,需要根据统计的数据量(包括变量的数目、类型、个案记录的条数等)以及统计的项目来选择计算机的硬件和软件。对于数据量大、统计计算工作量大(如需要进行大矩阵运算和复杂预测)的任务,需要选择高档次中央处理单元 CPU、大内存量、具有大硬盘的计算机,这样,计算机的计算速度就快一些,并可以处理大量的数据。另外,对于数据量大的分析资料,也要选择能够处理相应数据量的计算机软件包,并且该软件包还要具有所需要的统计功能。

下面来了解一下 SPSS 和 SAS 软件包的功能,首先全面了解两个软件包各自的功能,以便今后进一步有针对性地去学习。

(一) SPSS 统计软件包的功能

经过不断改进,目前 SPSS 的微机版,不仅有 DOS 版本,还有 WINDOWS 版本。它几乎包括了所有的数据管理和统计分析功能,并且命令简短,具有帮助(HELP)功能,操作也更加方便。尤其是 WINDOWS 版本,不必记忆统计命令即可调用主要的统计功能,用户只要通过菜单来选择所用的功能,即可完成工作,因而它更适应非统计专业的统计分析工作者来使用。下面按 SPSS/PC+ 版本的模块,来简介其主要的功能:

1. 基本系统(BASE SYSTEM)

(1) 数据的定义和操作。包括数据类型与长度的定义、数据的输入输出、数据的重新编码和加权、数据样本的选择、进行算术和逻辑的计算、漏填数据的处理等;并可以对样本数据进行分类排序,读写数据文件及选择打印输出格式等。

(2) 频数分布。对给定变量进行频数分析,并可以输出表征频数的条形图、直方图和分布图,另外可以给出均值、众数、中位数、陡度和偏斜度等统计值。

(3) 描述量数。以简短列表的形式给出频数分布中所需的各种表征数。

(4) 建交叉表。完成两个或两个以上变量在分项不多的情况下的联表。

(5) 求分组样本的均数。求由一个或多个自变量分类而成的不同的分组样本的均值及标准差等。

(6) t 检验。用于检验两组或两组以上样本的均值,从而进行组间显著性测试。

(7) 绘图。绘制两个变量之间的散点图、回归直线或叠加图、多层次弧形图。

(8) 多元相关。计算多元相关系数矩阵。

(9) 单因子分析。进行单因子试验的方差分析,可以计算平方和、均方差、自由度,并进行 F 检验等。

(10) 方差分析。进行因子间的变异量分析。

(11) 非参数检验。用于在不知道变量概率分配情况下而要进行假设检验的工作。

(12) 多元回归。用于多元回归的计算。

(13) 报表输出。以报表形式输出样本的清单和多种总结表征数。

(14) 全屏幕编辑。提供一个强大的编辑系统 REVIEW,用于