



全国统计教材编审委员会“十一五”规划教材

统计实验教材系列

多元统计分析实验



★ 张立军 任英华 编著

4-33



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

统计实验教材系列

多元统计分析实验

★ 张立军 任英华 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

多元统计分析实验/张立军,任英华编著.

—北京:中国统计出版社,2008.8

全国统计教材编审委员会“十一五”规划教材

ISBN 978-7-5037-5510-1

I. 多…

II. ①张… ②任…

III. 多元分析:统计分析—实验—高等学校—教材

IV. 0212.4—33

中国版本图书馆 CIP 数据核字(2008)第 103111 号

多元统计分析实验

作 者/张立军 任英华

责任编辑/吕 军 吕忠伟

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/220 千字

印 张/16.5

印 数/1—3000 册

版 别/2009 年 3 月第 1 版

版 次/2009 年 3 月第 1 次印刷

书 号/ISBN 978-7-5037-5510-1 /O·67

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培训体系的关键时期。为了更好地培育统计创新型科技人才,适应统计教育培训的新形势,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,把提

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》、《现代指数理论》、《现代金融投资统计分析》9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人的《学习指导与习题》,使得这批教材在编辑出版上形成了一套比较完整的体系。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、广大教师和学生提出批评和建议。

全国统计教材编审委员会
2006年6月

“统计实验教材系列”

总序

随着我国经济、社会和科技的发展,统计应用的面越来越广,如何用好统计工具,成为人才培养、科学研究和实际工作中面临的共同问题。而科学地应用统计工具,至少要解决好三个问题:一是用好统计数据,即了解所要采用的统计数据的性质和特点,做到“用数据说话”;二是用好统计方法,即能够根据统计数据的性质和特点,采用恰当的统计方法对之进行分析研究,使得“用数据说话”说得恰当而有力;三是用好统计软件,即在现代信息技术条件下,能够运用统计软件对大量数据进行处理和分析,提高分析研究的精度与效率。在现代科技条件下,上述三个方面是统计工具应用的有机组成部分,三者必须相互结合、不可分割。而统计实验教学,就是上述三者相互结合的一种新的教学模式,它针对所要研究解决的实际问题,将相应的统计数据、统计方法和统计软件相结合,在专门实验室内完成统计过程的模拟、统计知识的学习和统计技能的培养,使学习者能够在模拟实践的环境下,掌握统计工具应用技能,提高统计实践能力。

传统的统计教学与通常的社会科学类专业教学一样,基本上属于理论教学模式。我们所倡导的实验教学模式与理论教学模式有很大的不同:第一,教学场所不同,理论教学是在教室进行,实验教学是在实验室进行(在具备条件的单台计算机上亦可进行个别学习);第二,

教学工具不同,与一般教室的教学工具配给不同,实施统计实验教学的实验室至少需要配备计算机、统计软件及相关配套设施,并具备良好的网络运行环境;第三,教学方式不同,理论教学通常是教师讲、学生听、偶尔有交流,实验教学要求学生与教师一起在计算机上进行操作,学习的自主性更强;第四,教学内容不同,与理论教学比较,实验教学使学生在统计理论与方法的同时也学习了统计软件应用技术,并且能够进行更复杂的统计方法的学习和更大规模的统计数据处理的;第五,教学效率不同,由于节省了板书时间,实验教学在单位时间内比理论教学能够提供更大的信息量,讲授更多的内容;第六,教学效果不同,实验教学是在模拟实践环境和遇到实际问题时通过实验项目来完成,一般情况下比理论教学更有利于培养学生的研究能力、创新能力和实践能力。当然,我们倡导实验教学,并不是要用实验教学完全取代理论教学,而是将统计课程中能够实验化的部分(通常就是统计方法及其应用问题)予以实验化,即:将之设计成为实验项目来实施教学,它既是统计教学模式的创新,也是对统计理论教学的延伸和发展。对于大多数统计课程来说,并不是所有的教学内容都可以实验化,因而实验教学应当与理论教学相结合。

然而,目前我国高校统计教学工作中,理论讲授过多、实际操作过少、缺少实验教学环节的问题仍然比较突出,统计实验教材少,而与相关课程理论教学相配套的实验教材更少。为此,我们组织编写了本套“统计专业课程系列实验教材”,其设计和编写的基本思路是:

一、按照统计专业主干课程设计实验教材。为了更好地体现统计理论教学与实验教学相结合并便于组织教学实施,我们基本上根据现阶段统计专业的主干课程来设计实验教材的种类,使得各门专业主干课程都有实验教材与之相配套。初步计划,系列实验教材包括八种,

即:《统计学基础实验》、《数理统计学实验》、《计量经济学实验》、《多元统计分析实验》、《时间序列分析实验》、《数据挖掘实验》、《市场调查与分析实验》和《经济预测与决策实验》。

二、各种实验教材的内容由系列实验项目组成。实验项目是根据课程教学内容进行实验设计而成,一般包括三类:一是演示性实验,通常用来介绍所选用的统计软件,包括统计软件运行的硬件条件与环境;二是设计性实验,用来展示课程中各种统计方法的应用技术及其实现过程,一般按课程的主要统计方法类别分别设计实验项目,是一门课程实验项目的主要形式;三是综合性实验,属于一门课程各种统计方法综合运用的大型实验项目,主要培养学生综合运用各种统计方法分析和解决实际问题的能力。

三、实验项目的设计体现教学规律的要求。实验教材包括的各个实验项目,大多都是编著教师在长期的教学实践中经过不断探索、反复试验而总结形成的,不仅在内容和形式上具有创新性,而且力求体现教学规律的要求。一个实验项目按照教学程序大致包括以下几个部分内容:(1)实验目的,概述一个实验项目设计的意图与思路,应达到的目的与要求;(2)实验原理,简要介绍实验项目所应用的理论、方法与模型;(3)实验数据,给出实验项目要用到的数据;(4)实验过程,具体说明实验项目操作的环节、步骤及其要求,是实验项目的主体;(5)实验小结,归纳实验的要点,给出类推实验和相关提示;(6)练习实验,给出同类或相关方法实验的练习题。在课时设计上,一般一个实验项目对应2个课时,大型综合性实验项目可安排4个课时。

由此可见,本套教材是统计教学模式改革和统计实验教学建设的结果。

本套教材的编写和出版,各位编著教师付出了艰辛的努力,国家统计局统计教育中心全国统计教材编审委员会

和中国统计出版社给予了大力的支持,并得到湖南省普通
高等学校教学改革研究项目《经济管理类专业实验教学模
式的创新与应用》的支持,在此一并致以诚挚的感谢!

统计实验教学及其教材编写在我国还是一项具有探
索性的工作,加之编者的学识水平和研究能力有限,教材
中的问题和错漏在所难免,敬请各位专家和读者批评指
正,也祈盼有关教学研究和教材编写工作能够不断推进。

许涤龙

2009年3月

前言

市场经济对统计人才的需求更趋微观化、市场化,这要求学生具有较强的分析问题和解决问题的综合能力。进行实验教学,是全面完成统计教学和培养统计专业人才培养过程的重要过程,是统计理论学习的继续、补充和发展,是实现理论与实践相结合的重要环节,是培养学生实践技能和独立工作的重要手段,也是使高校培养出来的学生更好地符合社会需要的一个重要步骤。

多元统计分析是统计专业中的一门应用性、操作性极强的课程,它的产生与发展始终与社会实际紧密相联,同时,多元统计理论和应用的有机结合与统计软件的发展密切相关。在多元统计分析课程教学实践中,无论是教师还是学生常常会遇到一个实际而棘手的问题——数据处理。多元统计分析课程内容具有理论深奥、数据处理量大、计算方法复杂等特点,手工操作不但费时、费力、误差较大,而且在很多情况下是不可能做到的。很多学生对于多元统计分析理论是能够接受的,但是在解决实际问题时,却因不会使用统计软件使得学习效果不佳。故而在教学内容中有必要安排一部分课时,讲解多元统计方法在统计软件中的实现过程,使多元统计分析教学与计算机实验有机结合起来。通过上机实验,能使學生进一步理解多元统计思想、方法,进而提高学生分析与解决实际问题的能力。

由于目前国内还没有专门的《多元统计分析实验》教材,为此,我们编写了这本书。希望这本书的出版,对多元

统计分析方法的应用与推广以及统计教育的改革与发展，能起到一定的促进作用。

本书可作为高等院校统计、财经、管理等专业的本科生学习《多元统计分析》课程的教材，也可以作为相关专业的研究生和广大科技工作者学习该课程的参考用书。

本书由许涤龙教授负责提出写作构思，设计全书的主体框架，由张立军、任英华编著。全书共分为 10 个实验项目，项目 1、项目 2、项目 3、项目 10 由张立军执笔；项目 4、项目 5 由任英华执笔；项目 6、项目 7 由曹玮执笔；项目 8、项目 9 由罗珍执笔。此外，曹玮、罗珍在本书的数据资料收集与整理、SPSS 软件运行等方面做了大量的工作。

限于编者的知识和水平，书中的错误及结构安排不当之处在所难免，恳请广大读者不吝批评指正。

本书在编著过程中，参考了许多专家、学者的有关教材、专著、论文，并引用了部分资料，从中获益匪浅；同时本书的出版还得到了中国统计出版社的大力支持，在此一并表示衷心的感谢。

编者

2008 年 12 月

目 录

项目 1 SPSS13.0 入门 1

- 1.1 实验目的 1
- 1.2 实验原理 1
- 1.3 实验设备 3
- 1.4 实验过程及结果分析 4
- 1.5 实验练习 20

项目 2 聚类分析 23

- 2.1 实验目的 23
- 2.2 实验原理 23
- 2.3 实验数据 38
- 2.4 实验过程及结果分析 42
- 2.5 实验练习 61

项目 3 判别分析 65

- 3.1 实验目的 65
- 3.2 实验原理 65
- 3.3 实验数据 77
- 3.4 实验过程及结果分析 78
- 3.5 实验练习 99

项目 4 主成分分析 102

- 4.1 实验目的 102
- 4.2 实验原理 102
- 4.3 实验数据 111
- 4.4 实验过程及结果分析 111
- 4.5 实验练习 119

项目 5 因子分析 122

- 5.1 实验目的 122
- 5.2 实验原理 122
- 5.3 实验数据 129
- 5.4 实验过程及结果分析 131
- 5.5 实验练习 142

目 录

项目 6	对应分析	146
6.1	实验目的	146
6.2	实验原理	146
6.3	实验数据	152
6.4	实验过程及结果分析	153
6.5	实验练习	158
项目 7	典型相关分析	162
7.1	实验目的	162
7.2	实验原理	162
7.3	实验数据	165
7.4	实验过程及结果分析	167
7.5	实验练习	173
项目 8	Logistic 回归	177
8.1	实验目的	177
8.2	实验原理	177
8.3	实验数据	192
8.4	实验过程及结果分析	194
8.5	实验练习	207
项目 9	路径分析	210
9.1	实验目的	210
9.2	实验原理	210
9.3	实验数据	219
9.4	实验过程及结果分析	220
9.5	实验练习	229
项目 10	综合实验:上市公司经营业绩指标选择与 综合评价	234
10.1	实验目的	234
10.2	实验原理	234
10.3	实验数据	236
10.4	实验过程及结果分析	240
10.5	实验练习	246
参考文献		248

项目 1

SPSS13.0 入门

1.1 实验目的

通过本实验项目,旨在使学生理解掌握 SPSS13.0 软件包的基本功能以及相关的基本操作,具体包括:

- (1)了解 SPSS 软件包基本概况;
- (2)掌握 SPSS13.0 软件包的启动与退出、数据的输入与保存等;
- (3)掌握 SPSS13.0 软件包中有关数据文件整理的基本操作,如数据查询、数据修改、数据删除、数据的排序、选择、拆分和合并等。

1.2 实验原理

1.2.1 SPSS 软件概况

SPSS 是软件英文名称的首字母缩写,原意为 Statistical Package for the Social Sciences,即“社会科学统计软件包”。但是随着 SPSS 产品服务领域的扩大和服务深度的增加,SPSS 公司已于 2000 年正式将英文全称更改为 Statistical Product and Service Solutions,意为“统计产品与服务解决方案”,标志着 SPSS 的战略方向正在做出重大调整。

SPSS 是世界上最早的统计分析软件,由美国斯坦福大学的三位研究生于 20 世纪 60 年代末研制,同时成立了 SPSS 公司,并于 1975 年在芝加哥组建了 SPSS 总部。1984 年 SPSS 总部首先推出了世界上第一个统计分析软件微机版本 SPSS/PC+,开创了 SPSS 微机系列产品的开发方向,极大地扩充了它的应

用范围,并使其能很快地应用于自然科学、技术科学和社会科学的各个领域,世界上许多有影响的报刊杂志纷纷就 SPSS 的自动统计绘图、数据的深入分析、使用方便、功能齐全等方面给予了高度的评价与称赞。迄今 SPSS 软件已有 30 余年的成长历史。全球约有 25 万家产品用户,它们分布于通讯、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等多个领域和行业,是世界上应用最广泛的专业统计软件之一。在国际学术界有条不成文的规定,即在国际学术交流中,凡是用 SPSS 软件完成的计算和统计分析,可以不必说明算法,由此可见其影响之大和信誉之高。

1994 年至 1998 年间,SPSS 公司陆续并购了 SYSTAT 公司和 BMDP 软件公司、Quantime 公司和 ISL 公司等,并将各公司的主打产品收纳 SPSS 旗下,从而使 SPSS 公司由原来的单一统计产品开发与销售转向为企业、教育科研及政府机构提供全面信息统计决策支持服务,成为走在了最新流行的“数据仓库”和“数据挖掘”领域前沿的一家综合统计软件公司。

SPSS 最突出的特点就是操作界面极为友好,输出结果极为美观(从国外的角度看),它使用 Windows 的窗口方式展示各种管理和分析数据方法的功能,使用对话框展示出各种功能选择项,只要掌握一定的 Windows 操作技能,精通统计分析原理,就可以使用该软件为特定的科研工作服务,是非统计专业人员的首选统计软件。在众多用户对国际常用统计软件 SAS、BMDP、GLIM、GENSTAT、EPILOG、MiniTab 的总体印象分的统计中,其诸项功能均获得最高分。SPSS 采用类似 EXCEL 表格的方式输入与管理数据,数据接口较为通用,能方便地从其他数据库中读入数据。其统计过程包括了常用的、较为成熟的统计过程,完全可以满足非统计专业人员的工作需要。对于熟悉老版本编程运行方式的用户,SPSS 还特别设计了语法生成窗口,用户只需在菜单中选好各个选项,然后按“粘贴”按钮就可以自动生成标准的 SPSS 程序,极大地方便了中、高级用户。

1.2.1 SPSS 运行方式与基本操作

SPSS 是一个组合式软件包,是统计学基本理论、方法与技术在计算机上的实现。SPSS for Windows 软件有三种运行方式:完全窗口菜单管理方式、程序运行管理方式和混合运行管理方式。它的基本操作是窗口式操作,使用对话框展示各种功能选择,清晰、直观、易学。SPSS 还具有强大的图形功能,形象地显示对原始数据和分析结果的各种描述。在数据录入时有关术语如表 1-1:

表 1-1 SPSS 有关术语表

Data Editor	SPSS 术语	中文解释	统计学概念
二维表	Data file	数据文件	总体或样品数据
一行	Case	样品、个案	一个概率事件、观测量
行号	Case number	样品编号	观测量序号
一列	Variable	变量	预测对象的某个特征
列名	Variable name	变量名	特征名称
单元格中的数值	Value	变量值	某观测对象的特征值

在做统计分析前,有时要对原始的数据文件进行新的整理,包括排序、抽样、拆分和合并,这些功能都包含在 Data 菜单栏中。

1. 样品排序(Sort case):数据文件中样品的顺序是按数据输入的先后顺序排列的,Sort case 可对当前数据文件的样品重新按指定的变量或变量组升序或降序排序,这个变量称为排序变量。

2. 转置(Transpose):数据文件的结构对统计分析、统计制图时选择计算或描述的统计量有很大的影响,因此有时为了需要,要对现有的数据文件进行调整,一个常用的命令是行列转置。

3. 文件拆分(Split File):由于统计分析的需要,有时要对数据文件进行多层次的划分,分成若干个样品组,对每组进行统计分析或制作统计图形,Split File 能满足这种要求。

4. 文件合并:文件合并是指将外部文件的数据合并到当前数据文件中。

5. 选择样品(Select Case):如果用户只想对数据文件中的部分样品做统计分析,则可用 Select Case 命令进行部分样品的选择或抽样。

6. 计算(Compute):Compute 命令可对当前数据文件中已有的数值或字符串变量(老变量)进行数据更新,或者是建立新变量,变换后的每个样品的老变量或新变量的值均由用户定义的表达式进行计算。

1.3 实验设备

由于 SPSS 主要用途为大型数据库导向,它的运算一般涉及的数据量比较大,因此 SPSS13.0 官方网站对用户计算机提出了以下基本配置要求:

操作系统:Windows 98/ME/XP/2003/Vista;

CPU:Premium 133MHz;

内存:128MB;

显卡:VGA 16 位色、SVGA 32 位真彩色、硬件支持 DirectX7、硬件支持 DirectX8、硬件支持 DirectX9、硬件支持 DirectX10;

磁盘空间:20G;

光驱:无、CD-ROM、CD+R-ROM、CD+RW-COM、DVD-ROM、DVD+R-ROM、DVD+RW-ROM。

1.4 实验过程

1.4.1 SPSS13.0 for Windows 的启动与退出

1.4.1.1 SPSS13.0 for Windows 的启动

在开机启动 Windows 之后,在屏幕左下方鼠标按下的“程序”,再选择“SPSS for Windows”,最后选择“SPSS13.0 for Windows”,即开始运行 SPSS13.0 for Windows,并显示版本提示画面。

在提示画面之后出现的是 SPSS 文件对话框,对话框中共有六个选项,如图 1-1 所示:

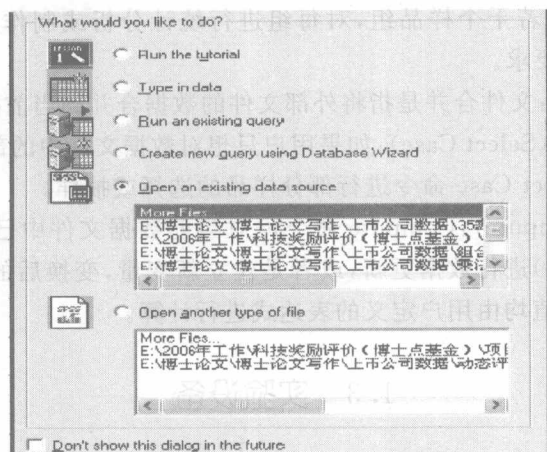


图 1-1 SPSS13.0 for Windows 对话框画面

图 1-1 所示的对话框中的“What would you like to do?”栏内共 6 个选项,选择不同的选项,将会打开不同类型的文件,除了这 6 个选项之后,在对话