

可下载教学资源

<http://www.tup.tsinghua.edu.cn>



高等学校教材
计算机应用

SPSS 统计分析 实例精选

蔡建琼 于惠芳 朱志洪 等 编著
阮桂海 审定

清华大学出版社



SPSS 20.0 统计分析

中国统计出版社
中国统计出版社

SPSS 统计分析 实例精选

中国统计出版社
中国统计出版社

中国统计出版社

高等学校教材
计算机应用

SPSS 统计分析 实例精选

蔡建琼 于惠芳 朱志洪 等 编著
阮桂海 审定

清华大学出版社
北京

内 容 简 介

本书是以 SPSS 全套英文文献作为指南,按照 SPSS 最新版本 13.0 的新功能、新特色编写而成的(输出风格截然不同于 SPSS 12.0 以前的版本)。

本书的前 17 章囊括了各种基本统计分析法。第 18 章后面各章涵盖了常用的专业统计和高级统计知识,可满足本科生、研究生、科研工作者进行各种定量分析的需求。有些章节,如数据快速输入(DATA LIST 法)、复方差分析、结合分析和对应分析等知识,是本书的特色之一,也是当今社会调查和市场调研的精锐武器。

本书力求写成国内 SPSS 分析应用的精品教材,可作为全国高校经济学、统计学、市场营销学、医学、心理学、人文社会学、管理学、运筹学专业计算机统计分析的教材,同时,本书也是统计师、科研人员、行政管理人员以及广大自然科学工作者进行课题研究及定量分析的首选参考书。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

SPSS 统计分析实例精选/蔡建琼,于惠芳,朱志洪等编著. —北京:清华大学出版社, 2006.3

(高等学校教材·计算机应用)

ISBN 7-302-12434-5

I. S... II. ①蔡... ②于... ③朱... III. 统计分析—软件包, SPSS—高等学校—教学参考资料
IV. C819

中国版本图书馆 CIP 数据核字(2006)第 005685 号

出版者: 清华大学出版社
http://www.tup.com.cn
社总机: 010-62770175
地址: 北京清华大学学研大厦
邮编: 100084
客户服务: 010-62776969

责任编辑：闫红梅

印 装 者: 清华大学印刷厂

发 行 者：新华书店总店北京发行所

开 本: 185×260 印张: 28 字数: 698 千字

版次: 2006 年 3 月第 1 版 2006 年 3 月第 1 次印刷

书 号: ISBN 7-302-12434-5/TP · 7973

印 数: 1 ~ 3000

定 价：38.00 元

改革开放以来,特别是党的十五大以来,我国教育事业取得了举世瞩目的辉煌成就,高等教育实现了历史性的跨越,已由精英教育阶段进入国际公认的大众化教育阶段。在质量不断提高的基础上,高等教育规模取得如此快速的发展,创造了世界教育发展史上的奇迹。当前,教育工作既面临着千载难逢的良好机遇,同时也面临着前所未有的严峻挑战。社会不断增长的高等教育需求同教育供给特别是优质教育供给不足的矛盾,是现阶段教育发展面临的基本矛盾。

教育部一直十分重视高等教育质量工作。2001年8月,教育部下发了《关于加强高等学校本科教学工作,提高教学质量的若干意见》,提出了十二条加强本科教学工作提高教学质量的措施和意见。2003年6月和2004年2月,教育部分别下发了《关于启动高等学校教学质量与教学改革工程精品课程建设工作的通知》和《教育部实施精品课程建设提高高校教学质量和人才培养质量》文件,指出“高等学校教学质量和教学改革工程”是教育部正在制订的《2003—2007年教育振兴行动计划》的重要组成部分,精品课程建设是“质量工程”的重要内容之一。教育部计划用五年时间(2003—2007年)建设1500门国家级精品课程,利用现代化的教育信息技术手段将精品课程的相关内容上网并免费开放,以实现优质教学资源共享,提高高等学校教学质量和人才培养质量。

为了深入贯彻落实教育部《关于加强高等学校本科教学工作,提高教学质量的若干意见》精神,紧密配合教育部已经启动的“高等学校教学质量与教学改革工程精品课程建设工作”,在有关专家、教授的倡议和有关部门的大力支持下,我们组织并成立了“清华大学出版社教材编审委员会”(以下简称“编委会”),旨在配合教育部制定精品课程教材的出版规划,讨论并实施精品课程教材的编写与出版工作。“编委会”成员皆来自全国各类高等学校教学与科研第一线的骨干教师,其中许多教师为各校相关院、系主管教学的院长或系主任。

按照教育部的要求,“编委会”一致认为,精品课程的建设工作从开始就要坚持高标准、严要求,处于一个比较高的起点上;精品课程教材应该能够反映各高校教学改革与课程建设的需要,要有特色风格、有创新性(新体系、新内容、新手段、新思路,教材的内容体系有较高的科学创新、技术创新和理念创新的含量)、先进性(对原有的学科体系有实质性的改革和发展、顺应并符合新世纪教学发展的规律、代表并引领课程发展的趋势和方向)、示范性(教材所体现的课程体系具有较广泛的辐射性和示范性)和一定的前瞻

性。教材由个人申报或各校推荐(通过所在高校的“编委会”成员推荐),经“编委会”认真评审,最后由清华大学出版社审定出版。

目前,针对计算机类和电子信息类相关专业成立了两个“编委会”,即“清华大学出版社计算机教材编审委员会”和“清华大学出版社电子信息教材编审委员会”。首批推出的特色精品教材包括:

- (1) 高等学校教材·计算机应用——高等学校各类专业,特别是非计算机专业的计算机应用类教材。
- (2) 高等学校教材·计算机科学与技术——高等学校计算机相关专业的教材。
- (3) 高等学校教材·电子信息——高等学校电子信息相关专业的教材。
- (4) 高等学校教材·软件工程——高等学校软件工程相关专业的教材。
- (5) 高等学校教材·信息管理与信息系统。

清华大学出版社经过近 20 年的努力,在教材尤其是计算机和电子信息类专业教材出版方面树立了权威品牌,为我国的高等教育事业做出了重要贡献。清华版教材经过 20 多年的精雕细刻,形成了技术准确、内容严谨的独特风格,这种风格将延续并反映在特色精品教材的建设中。

清华大学出版社教材编审委员会
E-mail: dingl@tup.tsinghua.edu.cn

前言

高等学校教材·计算机应用

SPSS(Statistical Package for Social Science)称为社会科学统计包,这是 20 世纪 60 年代刚推出 SPSS 时的说法。如今在医学、心理学、社会学、市场学、经济学乃至自然科学领域都得到广泛的普及和应用。

SPSS 及 SAS(Statistical Analysis System)等统计软件并称为国际上最有知名度的几大统计软件。SPSS 在国内之所以比 SAS 普及,是因为统计方法简单,统计结果科学清晰。国内众多的同仁越发体会到在当今市场经济的时代,若不学 SPSS 和 SAS,则很难有所作为。

笔者于 1987 年开始着手开发研究 SAS 和 SPSS 统计软件的分析应用,一边研发一边教学,并于 1991 年在经济日报出版社率先出版了《微电脑的使用与软件应用大全——SPSS/PC⁺ 4.0 高级教程》系列教材,为国内的 SPSS 爱好者尽了“投石问路”的绵薄之力。

本书内容突出“创新实用”。全书共分 23 章,每一章后面都有习题。为了方便读者,附录 A 还特地配上了习题答案。读者可通过出版社的网址 www.tup.com.cn 下载实习正文中提及的部分程序数据(但有版权,请勿扩散或作为他用),有的数据虽然没有提供,但可采用已给的相关数据进行变通和实习。本书固然注重版本更新,但更注重内容与所有版本之间的通用。

书中对每一种数据,既介绍对话框的统计法又通过 Paste 按钮粘贴出相应的一条命令。命令是奉献给熟悉 SPSS 的老读者,便于他们开发研究;对话框统计法对于不太了解 SPSS 功能的新读者,力求起到聪明人读“傻瓜书”那样的作用。根据笔者的体会,单击对话框固然省事,但是学习 SPSS 和 SAS 必须懂一些命令,不懂命令则掌握不到真正的知识内涵。本书虽然是按 SPSS 的新版本 13.0 写的,但对各种 SPSS 老版本都适用,因此本书非常适合于各个层面的读者。

在西方发达国家,SAS 和 SPSS 是大学生的必修课和毕业后谋生的“摇钱树”。在我国,SPSS 和 SAS 也已逐渐成为统计师和文理科大学生的“看家本领”。但是限于篇幅,有些内容未能入选,为此每年五一国际劳动节、国庆节和暑假笔者将分别举办一次 SAS 和 SPSS 应用研讨,以便传授新知识、解决新问题。若有兴趣请联系 010-62767251 电话或 13693258566,也可发 E-mail(rgh114716@sina.com 或 rghai@sohu.com)联系。

参加编写本系列教材的成员有蔡建琼、朱志洪、周爽、吴宝科、朱志海、于惠芳、阮开良、

阮开林、朱艳香、阮开顺、朱秀萍、朱星萍、顾林枝、崔博、阮潮海、阮潮霞、蔡文英、蔡建平、蔡建瓴、樊爱萍、蔡建洪、蔡建华、蔡福金、曲庆云、严康敏、孟峥、张才彬、刘云、刘萍、陈丹、陈耀辉、杨武栓、杨靖、罗安娜、郭崇德、赵晓梅、秦小峰、郝正元、谢力丹、綦淑娟、李欣、阮桂海等。

在编著过程中北京大学社会学系杨善华教授提供了全套 SPSS 英文文献作为理论指南,教育部高等学校文科计算机基础教学指导委员会副主任卢湘鸿教授及有关专家给予充分的肯定和宝贵的指导,在此一并表示由衷的感谢!

北京大学社会学系阮桂海

2005 年 9 月

rgl14716@sina.com 或 rghai@sohu.com

第 1 章 SPSS 的关键命令	1
1.1 SPSS 自身的程序结构(即过程)	2
1.2 SPSS 的程序(命令)行	3
1.2.1 一个简单的 SPSS 程序例子	3
1.2.2 SPSS 的操作命令	5
1.2.3 SPSS 的数据编码	6
1.2.4 另一个简单的 SPSS 编程	8
习题 1	9
第 2 章 采用命令法快速输入数据	10
2.1 SPSS 的简单命令	10
2.2 对问卷中数据的编码	11
2.3 定义 SPSS 变量的属性	13
2.4 SPSS 数据的固定格式	14
2.5 SPSS 数据的自由格式	14
2.6 建立 SPSS 的命令文件	14
2.7 SPSS 命令文件的 4 种格式	15
2.8 编程后立即进行统计分析	16
2.9 调用老命令文件(程序)进行统计分析	16
2.10 正确编辑 SPSS 原始数据小结	17
2.11 统计分析的最佳方案	17
习题 2	18
第 3 章 数据备份与分组	20
3.1 创建新变量	20
3.1.1 通过 Compute 对话框创建新变量	20
3.1.2 采用 COMPUTE 命令创建新变量	24

3.1.3	数学函数	25
3.1.4	缺失值函数	25
3.2	采用 IF 命令进行条件变换及逻辑校验	26
3.2.1	IF 命令格式	26
3.2.2	关系表达式中的关系符	26
3.2.3	逻辑表达式中的运算符	27
3.2.4	两种缺失值的逻辑函数	27
3.2.5	缺失值的处理	28
3.2.6	IF 命令的对话框	28
3.3	数据重新分组	30
3.3.1	单击对话框重新编码数据的实例	30
3.3.2	RECODE 命令的格式	34
3.3.3	RECODE 在命令文件中的位置	35
3.4	计算有相同情况的次数	35
3.5	SPSS 的运算次序	36
3.6	运算符的用法	36
习题 3		37
第 4 章	图文并茂的 SPSS 图形	39
4.1	数据要求	39
4.2	绘制条形图	44
4.3	在 Graphs 中绘制“一图多变量”的条形图	46
4.4	在 Graphs 中进一步绘制多变量的条形图	48
4.5	直方图的描述	49
4.6	图文并茂的圆形图	51
4.6.1	对话框法	51
4.6.2	命令法	53
习题 4		53
第 5 章	SPSS 统计分析的预备知识	55
5.1	总体与样本	55
5.2	参数与统计量	56
5.3	正态分布	56
5.4	经验规则的应用	57
5.5	正态分布的假设检验	58
5.5.1	检验正态性的重要统计量	58
5.5.2	检验正态性的其他图示法	60
5.6	显著性水平与 α 值	62
习题 5		62

第 6 章 数据的排序、行列互换及抽样和加权	64
6.1 SPSS 的 Data Editor 窗口	64
6.2 几十个 Cases 的数据输入法	64
6.3 成百上千个 Cases 的数据输入法	66
6.4 个案排序	67
6.5 数据的行列互换	69
6.6 限选个案 Cases 进行局部统计	70
6.6.1 限选某个子总体进行统计	70
6.6.2 提取随机样本进行统计	72
6.6.3 限选前 n 个 Cases	77
6.7 个案加权	77
6.8 Data 中的其他菜单命令	79
习题 6	79
第 7 章 用 Frequencies 和 Tables 描述频次	81
7.1 应用 Frequencies 过程的预备知识	86
7.2 绘制单变量的频率表	88
7.3 进一步通过 TABLE 过程绘制多变量的频率表	89
7.4 更深入的分析	90
7.5 Paste 回来的程序	91
习题 7	92
第 8 章 多选项的统计技术	95
8.1 多选项的计算机编码	95
8.2 多选项中二分法与分类法的区别	96
8.2.1 多选项分类法	96
8.2.2 多选项二分法	97
8.3 多选项的数据及程序	97
8.4 多选项频次分布表及分析	100
8.5 多选项的交叉汇总表及分析	101
8.5.1 交叉汇总表的统计法	101
8.5.2 输出的交叉汇总表	102
8.5.3 交叉汇总表分析	102
习题 8	103
第 9 章 采用 Crosstabs 过程对双变量交叉汇总	105
9.1 双变量交叉汇总法	105
9.1.1 统计方法	105

9.1.2 分析方法	107
9.2 次序—次序(定序—定序)变量用 Gamma 或 D 系数测量	107
9.3 比例—比例(定比—定比)变量用皮尔逊 CORR 系数测量	110
9.4 标称—区间(定类—定距)变量用 Eta 系数测量	112
9.5 标称—次序(定类—定序)变量用 λ 系数测量	113
9.6 标称—标称(定类—定类)变量用 λ 系数测量	115
9.7 Crosstabs 对话框的应用	117
9.8 CROSSTABS 过程的命令总表	118
习题 9	120
第 10 章 两个子总体均值的比较	123
10.1 描述子总体均值的差异(MEANS 过程的应用)	123
10.1.1 寻找相关方面	123
10.1.2 论证所用的数据及其程序	124
10.1.3 高考成绩与父亲职业的关系	126
10.1.4 不同地区考生的成绩分析	128
10.1.5 采用对话框进行统计	132
10.1.6 MEANS 过程的命令、子命令的功能及用法	133
10.2 一个样本 T-TEST	135
10.2.1 一个样本 T-TEST 的程序及数据	136
10.2.2 一个样本 T-TEST 的对话框	138
10.2.3 一个样本 T-TEST 的假设检验	139
10.2.4 一个样本 T-TEST 结果分析	139
10.3 独立样本 T-TEST 的应用	139
10.3.1 采用命令法进行独立样本 T 检验	140
10.3.2 采用对话框法进行独立样本 T 检验	141
10.4 配对差值 T-TEST	142
10.4.1 采用对话框进行配对差值 T-TEST	142
10.4.2 配对差值 T-TEST 的结果分析	144
10.4.3 采用命令统计法进行配对差值 T 检验	145
10.5 单因素方差分析	146
10.5.1 One-Way ANOVA 的菜单位置	146
10.5.2 采用对话框比较两组平均成绩的差异	146
10.5.3 One-Way ANOVA 的结果分析	148
习题 10	148
第 11 章 采用 Explore 探测数据的正态性	151
11.1 Explore 过程可探测数据的错误来源	151
11.2 Explore 探测的数据	152

11.3	图形显示之一:直方图	155
11.4	图形显示之二:茎叶图	155
11.5	框图	157
11.6	统计分析前对假设的检验	159
11.7	幂转换的形式	161
11.7.1	幂转换	161
11.7.2	正态性检验	161
11.8	集中趋势分布的3种较佳平稳测度	163
11.9	采用 Explore 的对话框进行数据探测	165
11.10	在 Syntax 窗口编程	168
	习题 11	169
第 12 章	过程 ANOVA 的应用	172
12.1	ANOVA 过程的应用之一:描述统计	172
12.2	ANOVA 过程的应用之二:方差分析	175
12.2.1	方差分析的假设	175
12.2.2	方差分析的结果	175
12.2.3	交互效应的检验	176
12.3	主效应的检验	177
12.4	ANOVA 过程的命令及其用法	177
12.5	采用对话框做 ANOVA(多因素方差)分析	179
	习题 12	179
第 13 章	非参数检验	181
13.1	非参数检验过程的菜单和数据文件	181
13.2	Chi-Square(卡方)单样本检验	182
13.3	Binomial(二项式)检验	183
13.4	药量的 Runs 检验	184
13.5	单样本的 Kolmogorov-Smirnov 检验	185
13.6	双样本的 Kolmogorov-Smirnov 检验	186
13.7	K 个独立样本的 Kruskal-Wallis 检验	187
13.8	成对样本的 Wilcoxon Signed Ranks 检验	188
13.9	K 对样本的 Friedman 检验	190
	习题 13	191
第 14 章	相关分析	193
14.1	双变量相关分析	193
14.1.1	数据	193
14.1.2	从相关矩阵看“SARS 疫情”与广东气温的关系	197

15.7 COXREG 过程的命令及子命令	225
习题 15	226
第 16 章 品牌研究中的结合分析	227
16.1 结合分析的要领	227
16.2 采用正交设计建立卡片	227
16.3 精品购物的一个调查例子	228
16.4 结合分析的方法	229
16.5 市场预测与决策	230
16.6 结合分析的命令	231
习题 16	232
第 17 章 多元线性回归	233
17.1 多元线性回归模型的拟合优度	234
17.2 多元回归的基本条件	236
17.3 设定异常值和影响点	239
17.3.1 从标准残差看异常值	239
17.3.2 检测 5 个最大及最小的异常值	240
17.4 数据不符合回归假设时的处理法	242
17.5 向前选择变量法	242
17.6 自后淘汰变量法	243
17.7 逐步回归法	245
17.8 多元回归中影响点的检测	246
17.9 多元共线性诊断	248
17.10 解释回归模型	251
17.11 在对话框中运行线性回归	251
习题 17	258
第 18 章 非线性回归分析	261
18.1 曲线估计	261
18.1.1 曲线估计过程的菜单位置	261
18.1.2 曲线估计的数据及其统计法	262
18.1.3 存储临时变量	267
18.2 二分的逻辑斯蒂克回归	267
18.2.1 采用对话框法进行二分逻辑斯蒂克回归	267
18.2.2 采用命令法进行二分逻辑斯蒂克回归	269
18.2.3 二分逻辑斯蒂克回归的输出与分析	269
18.2.4 三种常用的回归法	270
18.3 多分逻辑斯蒂克回归分析	270

18.4	Two-Stage Least-Squares 回归	274
18.4.1	Two-Stage Least-Squares 回归例子	274
18.4.2	结果分析	276
18.5	Nonlinear Regression 过程在人口普查与预测中的应用	276
18.5.1	从非线性回归模型解出初始值	277
18.5.2	非线性回归模型的主要统计量	278
18.5.3	人口预测	280
18.6	利用 Probit 过程分析单位概率	280
18.6.1	中度效果分析	282
18.6.2	三组预测水平值的比较	283
18.6.3	各组药效的比较	285
	习题 18	286
第 19 章	聚类分析	292
19.1	二阶聚类	292
19.1.1	二阶聚类的特点	292
19.1.2	二阶聚类的数据	293
19.1.3	二阶聚类的对话框例子	294
19.1.4	二阶聚类的结果分析	295
19.1.5	二阶聚类小结	300
19.2	谱系聚类	301
19.3	个案 Q 聚类	301
19.3.1	从冰柱图看聚类结果	303
19.3.2	平均连接法的图表	304
19.3.3	从树形图分析聚类成员	304
19.4	变量 R 聚类	305
	习题 19	307
第 20 章	判别分析	308
20.1	从抽烟与心脏病的关系入手研究	309
20.1.1	选择分析的个案	311
20.1.2	组间均值差别	312
20.1.3	Wilks 的 λ 值	312
20.1.4	相关系数的评价	312
20.1.5	判别系数的估计	313
20.1.6	判别函数的解释	314
20.1.7	Bayes 的分组规则	315
20.1.8	判别分的计算及分组	315
20.1.9	个案分组后的概括统计表	317

20.1.10	判别分的直方图(两种版本的比较)	317
20.1.11	计算错误分组的比例	318
20.1.12	不正确分组的期望比例	320
20.1.13	判别分析的其他统计量	320
20.1.14	判别函数与变量二者之间的关系	322
20.1.15	费歇尔分组函数系数	323
20.2	逐步判别法	323
20.3	其他几种选择变量的标准	324
20.4	三组判别分析	325
20.5	当违背假设时的情形	330
20.6	判别分析对话框的用法	331
20.7	判别分析的命令总表	333
	习题 20	334
第 21 章	因素分析及对应分析	336
21.1	因素分析所用的数据	336
21.2	因素分析的对话框设置法	337
21.3	输出结果及其分析	338
21.3.1	因素的初始统计量(共通性)	339
21.3.2	因素抽取	339
21.4	用主成分法抽取前 3 个因素	341
21.5	通过未加权最小二乘法抽取因素	341
21.6	提炼主成分	342
21.7	转轴前后的对比	343
21.8	对转轴效果的验证	345
21.9	因素分析	345
21.10	社会调研与市场分析中的对应分析	346
	习题 21	349
第 22 章	复方差分析	350
22.1	什么是复方差分析	350
22.2	假设与检验	351
22.2.1	假设的必要条件	351
22.2.2	检验假设	351
22.3	多因变量二水平模型的假设检验	353
22.3.1	多因变量之间相关性的检验	354
22.3.2	因变量联合分布为正态性的假设与检验	354
22.3.3	方差齐性检验	355
22.4	双因变量单因素的 Hotelling T^2 检验	356

22.4.1	单因变量的 Hotelling T^2 检验	357
22.4.2	参数估计	358
22.5	因变量是多水平(多样本)模型的假设检验	359
22.5.1	本节分析用的数据及程序	359
22.5.2	进行描述统计	361
22.5.3	进行主成分分析或 Bartlett 的球状检验	361
22.5.4	多元差异性检验	361
22.5.5	效应检验	361
22.6	用残差检验效度	362
22.6.1	观察值、期望值及其残差	362
22.6.2	参数的系数估计	363
22.6.3	均值的预测	364
22.7	最后的一点启示	365
22.8	采用 GLM 对话框进行复方差分析	365
	习题 22	366
第 23 章	反复测量方差分析	370
23.1	什么是反复测量	371
23.2	绘出图形并进行描述统计	371
23.2.1	描述性统计量	371
23.2.2	图形描述	372
23.3	均值差的分析	373
23.3.1	转换变量	374
23.3.2	正交正态差对照的图示	374
23.4	检验各种效应的差别	375
23.4.1	检验常数效应	376
23.4.2	关于方差分析表	376
23.4.3	检验考试对象效应	376
23.4.4	平均的单因变量显著性检验	378
23.4.5	选择多元或一元结果	379
23.5	选择 polynomial 等其他对照	380
23.6	二因子的设计模型	381
23.6.1	二因子模型的检验	381
23.6.2	变量的转换	382
23.7	效应的假设检验	383
23.8	增加一个对象间的因子后的交互效应	389
23.8.1	对象内与对象间的因子模型	389
23.8.2	有组间因子时的非饱和模型	390
23.8.3	有组间因子时的饱和模型	395