


21世纪高等学校规划教材 | 财经管理与应用



SPSS多元统计分析 方法及应用

朱星宇 陈勇强 主编

清华大学出版社



21世纪高等学校规划教材 | 财经管理与应用



SPSS多元统计分析 方法及应用

朱星宇 陈勇强 生编

清华大学出版社
北京

内 容 简 介

本书在阐述了 SPSS 基本功能的基础上,着重对多元统计分析的各个方法,针对目前部分统计教材以及 SPSS 丛书存在的问题,以数据分析应用需求为主线,对假设检验、方差分析、非参数检验、回归分析、聚类分析、判别分析、主成分分析、因子分析、对应分析、时间序列分析、信度分析、联合分析、生存分析、神经网络分析和结构方程模型 15 类方法,按照实际数据分析步骤从基本原理到软件操作进行了深入浅出的论述。本书基于 SPSS 17.0 版本,并在 SPSS 17.0 软件操作后附以独立案例进行分析。本书以自然科学和社会科学各领域研究人员为主要对象,同时也可供相关专业本科生、研究生、专业统计分析人员以及管理决策者进行学习或参考。

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。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

SPSS 多元统计分析方法及应用/朱星宇,陈勇强主编. —北京:清华大学出版社,2011.8
(21 世纪高等学校规划教材·财经管理与应用)

ISBN 978-7-302-25417-1

I. ①S… II. ①朱… ②陈… III. ①统计分析—软件包,SPSS IV. ①C819

中国版本图书馆 CIP 数据核字(2011)第 077091 号

责任编辑:高买花 薛 阳

责任校对:李建庄

责任印制:何 芊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62795954,jsjic@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015,zhiliang@tup.tsinghua.edu.cn

印 刷 者:北京密云胶印厂

装 订 者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:185×260 印 张:30.5 字 数:761 千字

版 次:2011 年 8 月第 1 版 印 次:2011 年 8 月第 1 次印刷

印 数:1~3000

定 价:46.00 元

产品编号:040216-01

出版说明

随着我国改革开放的进一步深化,高等教育也得到了快速发展,各地高校紧密结合地方经济建设发展需要,科学运用市场调节机制,加大了使用信息科学等现代科学技术提升、改造传统学科专业的投入力度,通过教育改革合理调整和配置了教育资源,优化了传统学科专业,积极为地方经济建设输送人才,为我国经济社会的快速、健康和可持续发展以及高等教育自身的改革发展做出了巨大贡献。但是,高等教育质量还需要进一步提高以适应经济社会发展的需要,不少高校的专业设置和结构不尽合理,教师队伍整体素质亟待提高,人才培养模式、教学内容和方法需要进一步转变,学生的实践能力和创新精神亟待加强。

教育部一直十分重视高等教育质量工作。2007年1月,教育部下发了《关于实施高等学校本科教学质量与教学改革工程的意见》,计划实施“高等学校本科教学质量与教学改革工程”(简称“质量工程”),通过专业结构调整、课程教材建设、实践教学改革、教学团队建设等多项内容,进一步深化高等学校教学改革,提高人才培养的能力和水平,更好地满足经济社会发展对高素质人才的需要。在贯彻和落实教育部“质量工程”的过程中,各地高校发挥师资力量强、办学经验丰富、教学资源充裕等优势,对其特色专业及特色课程(群)加以规划、整理和总结,更新教学内容、改革课程体系,建设了一大批内容新、体系新、方法新、手段新的特色课程。在此基础上,经教育部相关教学指导委员会专家的指导和建议,清华大学出版社在多个领域精选各高校的特色课程,分别规划出版系列教材,以配合“质量工程”的实施,满足各高校教学质量和教学改革的需要。

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21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”。推出的特色精品教材包括:

- (1) 21世纪高等学校规划教材·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) 21世纪高等学校规划教材·计算机科学与技术——高等学校计算机相关专业的教材。
- (3) 21世纪高等学校规划教材·电子信息——高等学校电子信息相关专业的教材。
- (4) 21世纪高等学校规划教材·软件工程——高等学校软件工程相关专业的教材。
- (5) 21世纪高等学校规划教材·信息管理与信息系统。
- (6) 21世纪高等学校规划教材·财经管理与应用。
- (7) 21世纪高等学校规划教材·电子商务。
- (8) 21世纪高等学校规划教材·物联网。

清华大学出版社经过二十多年的努力,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

联系人:魏江江

E-mail: weijj@tup.tsinghua.edu.cn

随着多元统计分析在自然科学各学科以及社会科学各个领域的不断发展,统计产品与服务解决方案(Statistical Product and Service Solutions, SPSS)由于其全面的功能以及灵活的配置方案,在多个行业和领域也得到越来越广泛的应用。在这种背景下,近年来高校各专业对统计学的要求和应用日益增加,对相关教材也提出了较高的要求,因此编写一部能够适应读者需求、并且能够得到读者认可的教材,是本书立项的初衷。

目前关于统计专业的教材大部分往往只注重原理而缺乏实现工具,而关于 SPSS 的丛书,多数只注重操作,却忽视对原理的讲解,因此往往使读者不能从根本上有效地利用软件对数据进行合理分析,而本书以数据分析的应用需求为主线,对多元统计分析的各种方法按照实际数据分析步骤从原理到通过 SPSS 17.0 软件操作进行深入浅出的论述,同时配合案例分析,使读者能够较快领会方法要点,明确适用条件,掌握方法的实现操作。

本书在阐述了 SPSS 基本功能的基础上,着重对多元统计分析的各个方法(包含假设检验、方差分析、非参数检验、回归分析、聚类分析、判别分析、主成分分析、因子分析、对应分析、时间序列分析、信度分析、联合分析、生存分析、神经网络分析和结构方程模型 15 类方法,各章内容示意图如图 0-1 所示),从基本原理到软件操作进行详细描述,并根据方法之间的联系分别将聚类分析与判别分析,主成分分析与因子分析放在同一章进行论述,并在 SPSS 软件操作后附以独立案例进行分析。

作为本书的编者之一,天津大学陈勇强教授曾经主持过国家自然科学基金项目“面向过程的工程项目交易方式选择及其实证研究”(71072156)并正在主持国家自然科学基金项目“工程项目交易方式及其效率衡量体系研究”(70772057),以及其他多项部委纵向课题,在研究过程中积累了丰富的经验,本书很多内容均为研究过程中的一些心得。

本书在成书过程中,得到了天津大学张宁(第 2、3、4、10、11 章)、焦俊双(第 1、6、7、9 章)、张雯(第 5、8、12 章)、张扬冰(第 13、14 章)等的协助,他们在相应章节案例收集、书稿整理、图表制作和排版等方面做了大量辛苦而细致的工作。本书参考了当前 SPSS 相关的书籍和文献,在此对这些作者一并表示感谢。同时还要感谢清华大学出版社相关工作人员,在他们的帮助下,本书才得以顺利出版。

本书以自然科学和社会科学各领域研究人员为主要对象,同时也可供相关专业本科生、研究生、专业统计分析人员以及管理决策者进行学习或参考。本书内容如图 0-1 所示。

本书受到国家自然科学基金(70772057)及(71072156)的资助。



图 0-1 本书内容示意图

本书难免存在不足之处,但可以肯定的是只有广大读者在使用过程中提出宝贵意见和建议,本书在再版时才能够更上一层楼,请随时和我们联系,可发信至 zhuxingyu@tju.edu.cn。

朱星宇 于波兰
陈勇强 于美国斯坦福大学
2011年5月

第 1 章 SPSS 综述	1
1.1 SPSS 17.0 概述	1
1.1.1 SPSS 17.0 特点	1
1.1.2 SPSS 各版本特性比较	2
1.2 SPSS 数据的管理	6
1.2.1 定义变量属性	7
1.2.2 个案标识	9
1.2.3 数据的排序	13
1.2.4 数据的转置	15
1.2.5 数据的重组	15
1.2.6 数据文件的合并	22
1.2.7 数据文件的拆分	25
1.2.8 数据的分类汇总	25
1.3 SPSS 数据的预处理	26
1.3.1 SPSS 表达式与函数	27
1.3.2 变量计算	28
1.3.3 选择个案	29
1.3.4 个案计数与加权	31
1.3.5 个案排序	32
1.3.6 数据的重新编码	34
1.3.7 SPSS 其他功能	37
1.4 基本统计分析	39
1.4.1 基本描述统计量的定义	39
1.4.2 频数分析	41
1.4.3 描述性分析	43
1.4.4 探索性分析	44
1.4.5 比率分析	45
1.4.6 P-P 图	46

1.4.7	Q-Q 图	48
1.4.8	基本统计分析实例	48
1.5	本章小结	54
	思考题	54
第 2 章	假设检验	55
2.1	常用分布及参数估计	55
2.1.1	几种与多元正态分布有关的概率分布	55
2.1.2	参数估计	59
2.1.3	正态分布的大样本推断	67
2.1.4	样本容量的确定	68
2.2	假设检验的一般问题	68
2.2.1	假设检验的概念	69
2.2.2	假设检验的基本思想	69
2.2.3	显著性水平及两类错误	69
2.2.4	假设检验的步骤	70
2.3	正态总体参数的假设检验	70
2.3.1	正态总体均值和方差的假设检验	71
2.3.2	总体比率的假设检验	72
2.4	假设检验的 SPSS 操作	72
2.4.1	单样本的 T 检验	72
2.4.2	两独立样本的 T 检验	73
2.4.3	两配对样本的 T 检验	75
2.5	假设检验实例	76
2.6	本章小结	77
	思考题	77
第 3 章	方差分析	78
3.1	方差分析的基本原理	78
3.2	单因素方差分析	79
3.2.1	数据结构与线性模型	79
3.2.2	平方和分解与自由度	80
3.2.3	显著性检验	81
3.2.4	多重比较	82
3.2.5	单因素方差分析的 SPSS 操作	83
3.2.6	单因素方差分析实例	86
3.3	多因素方差分析	89
3.3.1	多因素方差分析的分类	89
3.3.2	无交互作用的多因素方差分析	90

3.3.3	有交互作用的多因素方差分析	92
3.3.4	多因素方差分析的 SPSS 操作	93
3.3.5	多因素方差分析实例	98
3.4	重复测量方差分析	101
3.4.1	重复测量方差分析的基本原理	101
3.4.2	重复测量方差分析的 SPSS 操作	102
3.4.3	重复测量方差分析实例	103
3.5	协方差分析	106
3.5.1	协方差分析的基本原理	106
3.5.2	协方差分析的 SPSS 操作	107
3.5.3	协方差分析实例	108
3.6	本章小结	111
	思考题	111
第 4 章	非参数检验	112
4.1	单样本非参数检验	113
4.1.1	卡方检验	113
4.1.2	二项分布检验	113
4.1.3	游程检验	114
4.1.4	单样本 K-S 检验	115
4.2	两独立样本非参数检验	116
4.2.1	曼-惠特尼 U 检验	116
4.2.2	Moses 极端反应检验	117
4.2.3	K-S Z 检验	117
4.2.4	Wald-Wolfowitz 游程检验	118
4.3	多独立样本非参数检验	118
4.3.1	中位数检验	119
4.3.2	Kruskal-Wallis 检验	119
4.3.3	Jonckheere-Terpstra 检验	120
4.4	两相关样本非参数检验	121
4.4.1	McNemar 变化显著性检验	121
4.4.2	符号检验	122
4.4.3	Wilcoxon 符号秩检验	122
4.5	多相关样本非参数检验	123
4.5.1	Friedman 双向评秩方差检验	123
4.5.2	Kendall W 协同系数检验	124
4.5.3	Cochran Q 检验	125
4.6	非参数检验的 SPSS 操作	126
4.6.1	卡方检验	126

4.6.2	二项分布检验	128
4.6.3	游程检验	129
4.6.4	单样本 K-S 检验	129
4.6.5	两独立样本非参数检验	130
4.6.6	多独立样本非参数检验	131
4.6.7	两相关样本非参数检验	132
4.6.8	多相关样本非参数检验	132
4.7	非参数检验实例	133
4.8	本章小结	135
	思考题	135
第 5 章	回归分析	136
5.1	回归分析的概念和方法	136
5.1.1	概述	136
5.1.2	回归分析的研究范围	137
5.1.3	实际问题建立回归模型的过程	138
5.2	线性回归分析	140
5.2.1	一元线性回归	140
5.2.2	多元线性回归	143
5.2.3	回归诊断	147
5.2.4	多元线性回归的有偏估计	152
5.2.5	线性回归 SPSS 操作全过程	154
5.2.6	权重估计 SPSS 操作全过程	159
5.2.7	两阶最小二乘法 SPSS 操作全过程	159
5.3	非线性回归分析	160
5.3.1	可化为线性回归的曲线回归分析	160
5.3.2	曲线估计 SPSS 操作全过程	162
5.3.3	多项式回归分析	163
5.3.4	部分最小平方回归 SPSS 操作全过程	165
5.3.5	非线性回归分析	167
5.3.6	非线性回归 SPSS 操作全过程	168
5.4	Logistic 回归分析	171
5.4.1	自变量中含有定性变量的回归模型	171
5.4.2	处理定性变量的最优尺度回归 SPSS 操作全过程	173
5.4.3	逻辑回归模型	179
5.4.4	二元逻辑回归 SPSS 操作全过程	183
5.4.5	多项逻辑回归 SPSS 操作全过程	186
5.4.6	概率回归分析 SPSS 操作全过程	192
5.4.7	有序回归分析 SPSS 操作全过程	194

5.5 回归分析实例	196
5.5.1 线性回归实例	196
5.5.2 非线性回归实例	201
5.5.3 逻辑回归实例	203
5.6 本章小结	207
思考题	208
第 6 章 聚类分析与判别分析	209
6.1 聚类分析和判别分析的基本原理	209
6.2 相似性度量	210
6.2.1 区间变量	210
6.2.2 二值变量	212
6.2.3 定序变量	213
6.3 聚类分析方法	214
6.3.1 系统聚类法	214
6.3.2 逐步聚类法	217
6.3.3 二阶聚类法	218
6.4 聚类分析的 SPSS 操作	218
6.4.1 系统聚类	218
6.4.2 K-均值聚类	220
6.4.3 二阶聚类	222
6.5 判别分析方法	224
6.5.1 距离判别	225
6.5.2 Bayes 判别	226
6.5.3 Fisher 判别	227
6.5.4 判别分析步骤	228
6.6 判别分析的 SPSS 操作	230
6.7 聚类分析和判别分析实例	233
6.7.1 聚类分析实例	234
6.7.2 判别分析实例	235
6.8 本章小结	240
思考题	240
第 7 章 主成分分析与因子分析	241
7.1 主成分分析与因子分析的基本思想	241
7.2 主成分分析的模型与方法	242
7.2.1 主成分分析的代数模型与几何意义	242
7.2.2 总体的主成分	243
7.2.3 样本的主成分	247

7.3	主成分分析的 SPSS 操作	248
7.4	因子分析的模型与方法	251
7.4.1	正交因子模型	251
7.4.2	相关性分析	253
7.4.3	因子的提取	254
7.4.4	因子旋转	256
7.4.5	因子得分	257
7.5	因子分析的 SPSS 操作	258
7.6	主成分分析和因子分析实例	262
7.6.1	主成分分析实例	262
7.6.2	因子分析实例	264
7.7	本章小结	268
	思考题	268
第 8 章	对应分析	269
8.1	列联表与列联表分析	269
8.1.1	列联表	269
8.1.2	列联表分析	270
8.2	简单对应分析的基本原理	271
8.2.1	行轮廓与列轮廓	271
8.2.2	总惯量	272
8.2.3	行列轮廓的坐标	273
8.2.4	对应分析图	274
8.2.5	简单对应分析的步骤	275
8.2.6	简单对应分析的逻辑框图	275
8.3	简单对应分析的 SPSS 操作	276
8.4	多重对应分析及其 SPSS 操作	279
8.4.1	多重对应分析	279
8.4.2	多重对应分析的基本操作	280
8.5	对应分析实例	286
8.6	本章小结	290
	思考题	290
第 9 章	时间序列分析	291
9.1	时间序列的相关概念以及时间序列分析步骤	291
9.1.1	时间序列与统计学其他分析方法的关系	291
9.1.2	时间序列的相关概念	292
9.1.3	时间序列分析原理与分类	294
9.1.4	时间序列分析一般步骤	294

9.1.5	SPSS 时间序列分析	295
9.2	时间序列的数据准备与检验	295
9.2.1	时间序列的数据准备	296
9.2.2	时间序列的数据检验	296
9.2.3	时间序列的数据图形化检验	297
9.2.4	时间序列的数据统计量检验	303
9.3	时间序列的数据预处理	304
9.3.1	时间序列缺失数据的处理	304
9.3.2	时间序列数据的变换处理	305
9.4	时间序列的确定性分析	308
9.4.1	非平稳时间序列的组成要素	308
9.4.2	平滑法	309
9.4.3	趋势分析法	313
9.4.4	季节性分解法	313
9.5	时间序列的随机性分析	316
9.5.1	适用于平稳性序列的随机性时间序列模型	316
9.5.2	适用于非平稳性序列的随机性时间序列模型	318
9.5.3	时间序列随机性分析步骤	318
9.5.4	ARIMA 模型的参数设置	321
9.6	时间序列模型的 SPSS 操作	322
9.7	SPSS 时间序列的案例分析	328
9.8	本章小结	334
	思考题	334
第 10 章	信度分析	335
10.1	信度的基本原理	335
10.1.1	信度的统计学原理	335
10.1.2	信度影响因素	336
10.1.3	信度评价指标	337
10.2	信度分析及其基本方法	338
10.2.1	信度分析	338
10.2.2	信度分析的基本方法	339
10.3	信度分析的 SPSS 操作	343
10.4	信度分析实例	346
10.4.1	α 信度系数法分析	346
10.4.2	折半信度系数法分析	347
10.5	本章小结	348
	思考题	348

第 11 章 联合分析	349
11.1 联合分析的基本原理	349
11.2 联合分析的步骤	350
11.2.1 属性和属性水平的确定	351
11.2.2 受测设计	352
11.2.3 受测体的评价	354
11.2.4 效用值的估计	355
11.2.5 效用值的聚集	357
11.3 联合分析的 SPSS 操作	359
11.3.1 生成正交设计	359
11.3.2 显示设计	360
11.3.3 运行联合分析	361
11.4 联合分析实例	364
11.5 本章小结	370
思考题	370
第 12 章 生存分析	371
12.1 生存分析的基本概念和内容	371
12.1.1 生存分析的定义	371
12.1.2 生存分析的基本概念	372
12.1.3 生存分析的基本内容和方法	374
12.1.4 SPSS 中的生存分析过程	375
12.2 寿命表分析	376
12.2.1 寿命表分析的基本原理及步骤	376
12.2.2 SPSS 中的寿命表分析过程	378
12.3 Kaplan-Meier 分析	379
12.3.1 Kaplan-Meier 分析的基本原理及步骤	379
12.3.2 SPSS 中的 Kaplan-Meier 分析过程	381
12.4 Cox 回归模型分析	383
12.4.1 Cox 回归模型的基本形式和原理	384
12.4.2 SPSS 中的 Cox 回归模型分析过程	386
12.4.3 依时协变量 Cox 回归模型的基本原理	389
12.4.4 SPSS 中的依时协变量 Cox 回归模型分析过程	390
12.5 生存分析实例	391
12.5.1 寿命表分析实例	391
12.5.2 Kaplan-Meier 分析实例	395
12.5.3 Cox 回归模型分析实例	398
12.6 本章小结	403

思考题	404
第 13 章 神经网络分析	405
13.1 神经网络的发展历史以及神经网络相关概念	405
13.1.1 时间序列的发展历史简介	405
13.1.2 生物神经元模型	406
13.1.3 人工神经元模型	406
13.1.4 SPSS 神经网络分析	406
13.2 多层感知器模型	407
13.2.1 感知器神经元模型	407
13.2.2 感知器的网络结构	408
13.2.3 感知器神经网络的学习规则	408
13.2.4 感知器神经网络的训练	409
13.3 径向基函数模型	409
13.3.1 径向基函数神经网络结构	409
13.3.2 径向基函数的学习算法	409
13.4 神经网络的 SPSS 操作	410
13.4.1 变量设置	410
13.4.2 分区设置	412
13.4.3 体系结构设置	413
13.4.4 培训的设置	415
13.4.5 输出的设置	417
13.4.6 保存的设置	418
13.4.7 导出的设置	419
13.4.8 选项的设置	420
13.5 SPSS 神经网络的案例分析	421
13.5.1 数据准备	421
13.5.2 数据分析	422
13.5.3 过程摘要	422
13.6 本章小结	427
思考题	427
第 14 章 结构方程模型	428
14.1 结构方程模型概述	428
14.1.1 结构方程模型方法与统计学其他分析方法的关系	428
14.1.2 模型方程模型相关概念	429
14.1.3 结构方程模型原理与基本假定	430
14.1.4 结构方程模型特性	431
14.1.5 结构方程模型方法一般步骤	432

14.2	结构方程模型设定和识别	433
14.2.1	结构方程模型设定	433
14.2.2	结构方程模型识别	435
14.2.3	Amos 模型设定操作	436
14.3	结构方程模型数据准备	441
14.3.1	缺失数据的处理	441
14.3.2	数据的信度与效度	442
14.3.3	数据文件导入	442
14.4	结构方程模型参数估计	443
14.4.1	参数估计常用方法	443
14.4.2	Amos 参数估计操作	444
14.5	结构方程模型评价与修正	447
14.5.1	参数检验	447
14.5.2	模型整体拟合评价	448
14.5.3	模型限制修正	449
14.5.4	模型扩展修正	449
14.6	结构方程模型解释	451
14.6.1	相关关系	452
14.6.2	因果关系	453
14.7	本章小结	454
	思考题	455
	附录 SPSS 函数名及其含义	456
	参考文献	468