

● 高等学校教材

# SPSS

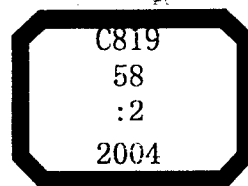
## 统计分析 高级教程

主 编 张文彤

副主编 董 伟



高等教育出版社



# SPSS 统计分析高级教程

主 编 张文彤  
副 主 编 董 伟

高等教育出版社

## 内容简介

SPSS 是最为优秀的统计软件之一,深受各行业用户的青睐。为满足广大读者学习和掌握高级统计分析方法的需求,本书以 SPSS 12.0 为准,详细介绍了各种多变量统计模型、多元统计分析模型的方法原理和软件实现技术,其内容涵盖了各种有广泛应用、经典或现代的模型和方法。全书共分 20 章,作者在书中结合了自身多年的统计分析实践和 SPSS 行业应用经验,既有深入浅出的理论介绍,又有针对实际问题的解决办法,更侧重于对统计新方法、新观点的讲解。在保证统计理论严谨的同时,又避免了传统教材过于学术化的缺陷,充分注重了文字的浅显易懂,使本书更加易学易用,是一本不可多得的使用 SPSS 进行高级统计分析的教材。

本书适合于已具备统计分析基础知识的读者阅读,可用作各专业研究生的统计学教材和参考书,也可供各行业中希望深入学习和应用高级统计分析方法的读者以及资深统计分析师作为参考书或工作手册使用。

## 图书在版编目(CIP)数据

SPSS 统计分析高级教程/张文彤主编. —北京:高等教育出版社,2004.9(2006重印)  
ISBN 7-04-015864-7

I. S... II. 张... III. 统计分析-软件包,SPSS-研究生-教材 IV. C819

中国版本图书馆 CIP 数据核字(2004)第 087685 号

策划编辑 耿芳 责任编辑 俞丽莎 市场策划 韩飞 封面设计 于文燕  
版式设计 史新薇 封面校对 金辉 责任印制 韩刚

出版发行 高等教育出版社  
社址 北京市西城区德外大街 4 号  
邮政编码 100011  
总机 010-58581000

经销 蓝色畅想图书发行有限公司  
印刷 北京市鑫霸印务有限公司

开本 787×1092 1/16  
印张 27.75  
字数 660 000

购书热线 010-58581118  
免费咨询 800-810-0598  
网址 <http://www.hep.edu.cn>  
<http://www.hep.com.cn>  
网上订购 <http://www.landracom.com>  
<http://www.landracom.com.cn>  
畅想教育 <http://www.widedu.com>

版次 2004 年 9 月第 1 版  
印次 2006 年 12 月第 3 次印刷  
定价 36.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究  
物料号 15864-00

# SPSS 统计分析高级教程参编人员

主 编 张文彤(复旦大学)  
副 主 编 董 伟(博塔(中国)有限公司)  
编 者 (以姓氏笔画为序)  
叶 勇(重庆工商大学)  
邝春伟(华东师范大学)  
刘晓云(复旦大学)  
汪 涛(复旦大学)  
宋艳艳(上海第二医科大学)  
周 琴(复旦大学)  
钟云飞(博塔(中国)有限公司)  
祝迎春(北京大学)

# 序 言

知识经济时代,数据成为宝贵的经济资源。在国外,电信、医疗、银行、证券、保险、制造、商业、科研、教育等行业已广泛采用先进的统计分析技术从数据中提取有价值的信息和知识。在国内,随着市场的蓬勃发展,了解成熟的统计分析产品,借鉴成功的统计分析行业应用经验,运用科学的统计分析方法,从数据中总结、归纳有用的知识,并将知识用于市场营销、运营决策和信用风险管理等领域,帮助企业、事业单位降低消耗、增加效益,从而提高整体运行效率,已成为各行业中具有远见的人士普遍关注的问题。

SPSS 软件是全球专业统计分析软件的领导者,一直致力于帮助企业、事业单位提高科学运用统计分析方法的能力,20 世纪 80 年代就已经被许多学者引入中国市场。它包含了丰富的统计分析算法,而且在使用层面上更多地考虑了客户在整个统计分析过程中的应用感受,其简洁的界面、完善的数据准备功能和杰出的图表输出能力使得 SPSS 软件在全球有超过 25 万家的机构用户,并成为国内的主流统计分析软件。随着该产品在各行业应用的深入以及 SPSS 培训和认证的广泛开展,目前国内已涌现出一大批应用 SPSS 的专家。近两年,国内统计分析市场如火如荼,SPSS 在产品技术上也不断推陈出新,继 SPSS 英文版在国内成功应用之后,SPSS 公司在今年首次推出了简体中文版,该产品更加符合中国人的使用习惯,一经推出便受到各行业人士的喜爱。

为了让中国的 SPSS 软件爱好者更好地使用软件,我们在经过两年的筹备后向市场推出了 SPSS 统计分析大型丛书。该丛书是一套全面了解、认识和应用 SPSS 最新统计分析软件、掌握统计分析方法的专业书籍;以统计分析在国内的应用现状为切入点,本着学以致用原则,在介绍统计理论及 SPSS 软件功能模块的同时,更侧重于统计分析在各项工作中的实际应用,引导读者不仅掌握 SPSS 软件及技术原理,而且学会运用统计方法解决工作和学习中的实际问题。

该丛书由业内权威专家主笔编写,资料引用详实可靠,实例剖析切中肯綮,不仅融合了行业专家在统计应用领域多年的研究成果,而且还融入了很多 SPSS 软件新老行业用户的实际应用经验。丛书总结了 SPSS 软件在各行业的实践应用状况,并综合 SPSS 最新行业应用方案,使各行业读者能通过学习提高 SPSS 软件的运用能力,解决工作中的实际问题。在丛书编写过程中,SPSS 公司和博塔(中国)有限公司(SPSS 中国地区分销商)的技术专家还及时提供了国际最新的行业发展信息和 SPSS 最新产品和技术信息,并结合 SPSS 的全球应用状况提出了宝贵意见。

丛书将分期分批出版相应的分册,其中首批面世的为通用教材《SPSS 统计分析基础教程》、《SPSS 统计分析高级教程》,均由复旦大学张文彤老师主编,全国多所高校的统计教师和统计专业人士参编。张文彤老师长期以来一直致力于积极推进统计分析工具在国内的普及应用,他在 2002 年编著的《SPSS11 统计分析教程》基础篇和高级篇因内容翔实、风格独特,受到了广大读者的热烈欢迎,并被多所高校列为本科生或研究生教材,其中基础篇一书已通过教育部评审,成为 2003—2004 年度教育部研究生工作办公室推荐的“研究生教学用书”。他不仅拥有丰富的教学经验,而且熟谙统计分析产品,本次全新编写的这两册教材分别针对不同读者群,由浅入深、结合实际应用全面介绍了 SPSS 产品和应用。全书实例引用突出,分析讲解透彻,读者可由本书管窥

全套丛书“应用为本”的特色。

显然,上述两本书还不能完全覆盖 SPSS 的所有应用领域,因此,本套书从现在还有数本分册正在组织编写中,包括全新的行业应用分册。这里我们也热忱邀请各行各业的 SPSS 资深用户,以及各高校的统计教师加入到本套丛书的编写工作中来,以共同推动我国各行业统计应用水平的迅速提高。

希望本套丛书能够让读者更清晰地了解统计分析,从而进一步促进统计分析在国内的普及。为便于读者交流和使用本套丛书,这里特公布相关网址如下:

SPSS 最新版本的全模块试用版下载:[www.spssbj.com.cn](http://www.spssbj.com.cn)

丛书相关案例数据下载:[www.spssbj.com.cn](http://www.spssbj.com.cn)、[www.MedStatStar.com](http://www.MedStatStar.com)

读者答疑、经验交流:[www.spssclub.com](http://www.spssclub.com)、[www.StatStar.com](http://www.StatStar.com)

博塔(中国)有限公司 SPSS 丛书编委会

# 前 言

笔者前作《SPSS 11 统计分析教程》(基础篇)和《SPSS 11 统计分析教程》(高级篇)自 2002 年面世以来,因其内容翔实、风格独特,受到了广大读者的热烈欢迎,这从读者用 E-mail 反馈信息的数量即可看出,还有数位读者详细指出了书中的用字错误,可见其阅读的详细程度,在此笔者对读者深表谢意。同时,数所高校均将其作为本科生或者研究生教材,这无疑都是对前作质量的充分肯定。

但是,再优秀的作品也有其生命周期,随着时间的推移,上述著作的不足之处也逐渐显现出来。对于高级篇而言,作为统计软件高级用户的参考书,其主线应当是统计理论,软件仅仅是各种方法的实现工具。而前作由于被定位为字典式的参考书,在内容上不得不同时照顾到统计软件操作界面的解释和统计方法的讲解,在这一点上就显得主线不够突出。而且许多读者也希望能够对一些复杂的方法进行深入讲解,并给出更为丰富的分析实例,这些都促使我们重新编写相应的高级教程。

在近两年的素材积累和编写工作后,这本全新的 SPSS 统计分析高级教程终于脱稿出版。本书主要针对 SPSS 的中、高级用户,定位为应用统计专业的研究生教材和其他专业的统计分析参考书。它以 SPSS 12.0 的功能为准,以统计理论为主线,详细介绍了 SPSS 中的各种多变量统计模型和多元统计分析方法。在保持全书简明易懂风格的基础上,对统计理论作了详细的讲解。全书内容共分四大部分:第一部分讲解了一般线性模型和混合线性模型,并重点对前者中的方差分析模型进行了介绍;第二部分则在此基础上进一步介绍了回归模型,包括对连续因变量建模的线性回归模型、线性回归的衍生模型、通径分析模型和非线性回归模型,以及对分类因变量建模的 Logistic 模型族和 Probit 模型;第三部分系统介绍了因子分析、判别分析、对应分析、多维尺度分析等多元统计方法的原理及其在 SPSS 中的实现;第四部分则对信度分析、生存分析、缺失值分析方法等较难归类,但又比较重要的统计分析方法进行了讲解。各章后均提供了参考文献和思考练习题,书后附录则以流程图的方式提纲挈领地给出了统计方法的分类体系,便于读者理解。另外,为便于读者自行对比分析结果,书中大部分表为 SPSS 自动生成的。因此,大部分表及表题为英文。

需要指出的是,作为本套丛书的通用统计教材,本高级教程严格遵循了统计理论这一主线,在统计方法的纳入上是有所选择的。方法体系中比较特殊的时间序列模型并未纳入本书范畴,另有分册专门介绍;而对于联合分析、多维偏好分析、离散选择分析等在方法原理上并无特别之处,行业应用特点明显的模型,则将被放在相应的行业应用分册中讲解,本书不再专门介绍。

除作为各专业研究生的统计教材和参考书外,本书还适用于各行业中希望深入学习和应用高级统计分析方法的读者。我们希望广大读者能一如既往地踊跃提出自己使用中的宝贵意见和建议,使得本书再版的时候能够更上一层楼,更完美地满足大家的学习和工作需求。

张文彤

2004 年 7 月于复旦公共卫生学院

# 目 录

## 第一部分 一般线性与混合线性模型

<b>第 1 章 方差分析模型</b> .....	3	2.1.1 完全随机设计 .....	29
1.1 模型简介 .....	3	2.1.2 配伍组设计 .....	29
1.1.1 模型入门 .....	3	2.1.3 交叉设计 .....	30
1.1.2 常用术语 .....	5	2.1.4 拉丁方设计 .....	32
1.1.3 方差分析模型的适用 条件 .....	7	2.2 考虑交互作用的实验设计方案 .....	34
1.2 简单分析实例 .....	8	2.2.1 析因设计 .....	34
1.2.1 模型表达式 .....	8	2.2.2 正交设计 .....	37
1.2.2 初步分析结果 .....	8	2.2.3 均匀设计 .....	39
1.2.3 模型参数的估计值 .....	10	2.3 误差项变动的特殊实验设计 方案 .....	40
1.2.4 两两比较 .....	12	2.3.1 嵌套设计 .....	40
1.2.5 其他常用选项 .....	13	2.3.2 重复测量设计 .....	42
1.3 两因素方差分析模型 .....	14	2.3.3 裂区设计 .....	42
1.3.1 分析实例 .....	14	2.4 协方差分析 .....	43
1.3.2 边际均数与轮廓图 .....	17	2.4.1 协方差分析的必要性 .....	43
1.3.3 拟和劣度检验 .....	19	2.4.2 平行性假定的检验 .....	45
1.4 因素各水平间的精细比较 .....	20	2.4.3 计算和检验修正均数 .....	46
1.4.1 POSTHOC 子句 .....	20	思考与练习 .....	48
1.4.2 EMMEANS 子句 .....	20	参考文献 .....	49
1.4.3 LMATRIX 和 KMATRIX 子句 .....	22	<b>第 3 章 多元方差分析与重复测量方差分析</b> .....	50
1.4.4 CONTRAST 子句 .....	24	3.1 多元方差分析 .....	50
1.5 随机因素的方差分析模型 .....	24	3.1.1 模型简介 .....	50
1.6 其他问题 .....	25	3.1.2 分析实例 .....	51
1.6.1 自定义效应检验使用的 误差项 .....	25	3.1.3 检验统计量的计算 .....	53
1.6.2 4 类方差分解方法 .....	27	3.1.4 对引例的进一步分析 .....	56
思考与练习 .....	27	3.2 重复测量资料的方差分析 .....	57
参考文献 .....	27	3.2.1 模型简介 .....	57
<b>第 2 章 常用实验设计分析方法</b> .....	28	3.2.2 分析实例 .....	59
2.1 仅研究主效应的实验设计方案 .....	28	思考与练习 .....	65
		参考文献 .....	65



<b>第4章 混合线性模型入门</b> .....	66	4.3.2 拟合混合线性模型的基本 结构 .....	79
4.1 模型简介 .....	66	4.3.3 考虑重复测量间的 相关性 .....	82
4.1.1 问题的提出 .....	66	4.3.4 更改对测量间相关性的 假定 .....	84
4.1.2 模型入门 .....	67	4.3.5 模型中可用的相关阵 种类 .....	85
4.2 层次聚集性数据分析实例 .....	69	4.4 模型总结 .....	86
4.2.1 拟合混合线性模型的 基本结构 .....	70	4.4.1 混合效应模型的用途 .....	86
4.2.2 在固定效应中加入 自变量 .....	72	4.4.2 混合效应模型与一般线性 模型的联系 .....	87
4.2.3 在随机效应中加入 自变量 .....	74	思考与练习 .....	87
4.2.4 更多解释变量的引入 .....	75	参考文献 .....	88
4.2.5 其他常用选项 .....	77		
4.3 重复测量数据分析实例 .....	77		
4.3.1 对数据的初步分析 .....	78		

## 第二部分 回归模型

<b>第5章 多重线性回归模型</b> .....	91	5.6 本章小结 .....	115
5.1 模型简介 .....	91	5.6.1 回归模型的建立步骤 .....	115
5.2 简单分析实例 .....	92	5.6.2 多重线性回归模型结果解释 时应注意的问题 .....	115
5.2.1 对数据的初步分析 .....	92	思考与练习 .....	116
5.2.2 回归模型的假设检验 .....	94	参考文献 .....	117
5.2.3 偏回归系数的假设检验 .....	95	<b>第6章 线性回归的衍生模型</b> .....	118
5.2.4 标准化偏回归系数 .....	96	6.1 非直线趋势的处理——曲线 直线化 .....	118
5.2.5 衡量多元线性回归模型 优劣的标准 .....	96	6.1.1 方法简介 .....	118
5.3 回归预测与残差分析 .....	99	6.1.2 使用 Linear 过程进行 分析 .....	118
5.3.1 回归预测与区间估计 .....	99	6.1.3 使用曲线拟合过程分析 .....	120
5.3.2 残差分析与模型适用条件的 检验 .....	101	6.2 方差不齐的处理——加权最小 二乘法 .....	122
5.4 逐步回归 .....	106	6.2.1 方法简介 .....	122
5.4.1 筛选自变量的基本原则 .....	106	6.2.2 使用 Linear 过程进行 分析 .....	123
5.4.2 常用的逐步回归方法 .....	107	6.2.3 使用 WLS 过程分析 .....	125
5.4.3 分析实例 .....	108	6.3 共线性的处理——岭回归 .....	127
5.5 模型的进一步诊断与修正 .....	111	6.3.1 方法简介 .....	127
5.5.1 强影响点的识别与处理 .....	111		
5.5.2 多重共线性的识别与 处理 .....	113		

6.3.2 分析实例 .....	128	8.5.1 参数初始值的设定 .....	161
6.4 分类变量的数值化——最优尺度		8.5.2 模型的拟合方法 .....	161
回归 .....	130	思考与练习 .....	162
6.4.1 方法简介 .....	130	参考文献 .....	162
6.4.2 分析实例 .....	131	<b>第9章 二分类 Logistic 回归模型</b> .....	163
6.4.3 最优尺度方法的应用注意		9.1 模型简介 .....	163
事项 .....	135	9.1.1 模型入门 .....	163
思考与练习 .....	136	9.1.2 一些基本概念 .....	165
参考文献 .....	136	9.2 简单分析实例 .....	167
<b>第7章 路径分析入门</b> .....	137	9.3 分类自变量的定义与比较方法 .....	171
7.1 两阶段最小二乘法 .....	137	9.3.1 使用哑变量的必要性 .....	171
7.1.1 模型简介 .....	137	9.3.2 SPSS 中预设的哑变量编码	
7.1.2 使用 Linear 过程进行		方式 .....	173
分析 .....	138	9.3.3 设置哑变量时要注意的	
7.1.3 使用 2SLS 过程进行		问题 .....	175
分析 .....	140	9.4 自变量的筛选方法与逐步回归 .....	176
7.2 路径分析入门 .....	142	9.4.1 模型中的假设检验方法 .....	176
7.2.1 模型简介 .....	142	9.4.2 自变量的筛选方法 .....	177
7.2.2 分析实例 .....	145	9.4.3 分析实例 .....	177
思考与练习 .....	148	9.5 模型拟合效果与拟合优度检验 .....	180
参考文献 .....	148	9.5.1 模型效果的判断指标 .....	181
<b>第8章 非线性回归模型</b> .....	149	9.5.2 拟合优度检验 .....	184
8.1 模型简介 .....	149	9.6 模型的诊断与修正 .....	186
8.1.1 问题的提出 .....	149	9.6.1 残差分析 .....	187
8.1.2 模型入门 .....	149	9.6.2 多重共线性的识别及其对回归	
8.2 简单分析实例 .....	150	系数的影响及处理办法 .....	187
8.2.1 软件操作与界面说明 .....	150	思考与练习 .....	187
8.2.2 基本分析结果 .....	151	参考文献 .....	188
8.2.3 模型的进一步分析 .....	153	<b>第10章 多分类、配对 Logistic 回归与</b>	
8.3 自定义损失函数——最小一乘法		<b>Probit 回归</b> .....	189
实例 .....	153	10.1 有序多分类 Logistic 回归模型 .....	189
8.3.1 分析实例 .....	154	10.1.1 模型简介 .....	189
8.3.2 结果解释 .....	156	10.1.2 分析实例 .....	190
8.4 分段回归模型的拟合 .....	157	10.1.3 模型适用条件的检验 .....	193
8.4.1 分析实例 .....	157	10.2 无序多分类 Logistic 回归模型 .....	195
8.4.2 结果解释 .....	158	10.2.1 模型简介 .....	195
8.4.3 模型的进一步分析 .....	160	10.2.2 分析实例 .....	196
8.5 其他需要注意的问题 .....	161	10.3 1:1 配对 Logistic 回归 .....	199

10.3.1 模型简介 .....	199	比较 .....	204
10.3.2 分析实例 .....	200	10.4.3 实例二——计算 LD50 .....	207
10.4 Probit 回归模型 .....	203	思考与练习 .....	209
10.4.1 模型简介 .....	203	参考文献 .....	209
10.4.2 实例一——与 Logistic 模型 .....			

### 第三部分 多元统计分析方法

<b>第 11 章 主成分分析与因子分析</b> .....	213	12.4 两步聚类法简介 .....	252
11.1 主成分分析 .....	213	12.4.1 方法原理 .....	252
11.1.1 模型入门 .....	213	12.4.2 分析实例 .....	253
11.1.2 简单分析实例 .....	216	12.5 本章方法小结 .....	258
11.1.3 对主成分分析的进一步 说明 .....	218	思考与练习 .....	260
11.2 因子分析 .....	218	参考文献 .....	260
11.2.1 模型入门 .....	218	<b>第 13 章 判别分析</b> .....	261
11.2.2 简单分析实例 .....	220	13.1 模型简介 .....	261
11.3 因子分析的进一步讨论 .....	226	13.1.1 典型判别分析的基本 原理 .....	261
11.3.1 不同的因子分析法 .....	226	13.1.2 判别分析的适用条件和 违背条件时的处理 方法 .....	262
11.3.2 相关阵和协方差 .....	227	13.1.3 判别效果的评价 .....	263
11.3.3 确定公因子数量 .....	227	13.1.4 判别分析的一般步骤 .....	264
11.4 因子分析综合案例 .....	227	13.2 简单分析实例 .....	265
11.5 主成分分析和因子分析的 比较 .....	233	13.2.1 软件操作与界面说明 .....	265
思考与练习 .....	234	13.2.2 基本分析结果 .....	266
参考文献 .....	234	13.2.3 判别结果的图形化 展示 .....	269
<b>第 12 章 聚类分析</b> .....	235	13.2.4 判别效果的验证 .....	271
12.1 模型简介 .....	235	13.2.5 适用条件的判断方法 .....	272
12.1.1 问题的提出 .....	235	13.3 贝叶斯判别分析 .....	273
12.1.2 聚类分析入门 .....	236	13.3.1 方法原理 .....	274
12.1.3 聚类分析的方法体系 .....	237	13.3.2 软件实现 .....	275
12.2 层次聚类法 .....	238	13.4 对判别分析的进一步讨论 .....	276
12.2.1 方法原理 .....	238	13.4.1 逐步判别分析 .....	276
12.2.2 分析实例 .....	239	13.4.2 判别分析和因子分析的 相似性和差异 .....	276
12.2.3 对层次聚类法的进一步 讨论 .....	246	13.4.3 二类判别和多重回归的 等价性 .....	276
12.3 K-均值聚类法 .....	247		
12.3.1 方法原理 .....	247		
12.3.2 分析实例 .....	248		

思考与练习 .....	277	15.4.2 分析实例 .....	304
参考文献 .....	277	15.5 对应分析中的其他问题 .....	308
<b>第14章 典型相关分析</b> .....	<b>278</b>	15.5.1 对应分析结果的正确	
14.1 方法介绍 .....	278	解释 .....	308
14.1.1 典型相关分析的基本		15.5.2 罕见类别和相似类别的	
思想 .....	278	处理 .....	308
14.1.2 典型相关分析的数学		15.5.3 有序类别的处理 .....	309
描述 .....	279	15.6 本章方法小结 .....	309
14.2 分析实例 .....	280	15.6.1 对应分析与其他分析方法的	
14.2.1 两组变量间的相关		关系 .....	309
系数 .....	281	15.6.2 对应分析的优势与	
14.2.2 典型相关系数及显著性		劣势 .....	311
检验 .....	282	思考与练习 .....	311
14.2.3 典型变量的系数 .....	283	参考文献 .....	312
14.2.4 典型结构分析 .....	284	<b>第16章 多维尺度分析</b> .....	<b>313</b>
14.2.5 典型冗余分析 .....	286	16.1 古典 MDS 模型 .....	313
14.3 小结 .....	288	16.1.1 方法原理 .....	313
14.3.1 典型相关分析的应用 .....	288	16.1.2 分析实例 .....	314
14.3.2 典型相关分析和因子		16.1.3 距离的计算方式 .....	320
分析 .....	288	16.2 非度量 MDS 模型 .....	321
思考与练习 .....	289	16.2.1 数据测量尺度的设定 .....	321
参考文献 .....	289	16.2.2 方法原理 .....	322
<b>第15章 对应分析</b> .....	<b>290</b>	16.2.3 分析实例 .....	323
15.1 模型简介 .....	290	16.3 考虑个体差异的 MDS 模型 .....	325
15.1.1 问题的提出 .....	290	16.3.1 方法原理 .....	325
15.1.2 模型入门 .....	290	16.3.2 分析实例 .....	326
15.1.3 SPSS 中的相应功能 .....	291	16.3.3 空间定位图的含义	
15.2 简单分析实例 .....	291	解释 .....	329
15.2.1 对数据的初步分析 .....	292	16.4 基于最优尺度变换的 MDS	
15.2.2 正式分析 .....	293	模型 .....	331
15.2.3 对引例的进一步分析 .....	297	16.4.1 方法简介 .....	331
15.3 基于均数的对应分析 .....	299	16.4.2 分析实例 .....	334
15.3.1 方法原理 .....	300	16.5 本章方法小结 .....	337
15.3.2 分析实例 .....	301	思考与练习 .....	337
15.4 多重对应分析 .....	304	参考文献 .....	337
15.4.1 方法原理 .....	304		

## 第四部分 其他统计分析方法

### 第 17 章 对数线性模型与 Poisson

#### 回归 ..... 341

#### 17.1 对数线性模型简介 ..... 341

##### 17.1.1 问题的提出 ..... 341

##### 17.1.2 模型入门 ..... 341

##### 17.1.3 SPSS 的相应功能 ..... 342

#### 17.2 一般对数线性模型分析实例 ..... 343

##### 17.2.1 对数据的初步分析 ..... 343

##### 17.2.2 正式分析 ..... 344

##### 17.2.3 对引例的进一步分析 ..... 347

#### 17.3 因果关系明确时的对数线性模型 ..... 349

#### 17.4 对数线性模型的选择 ..... 352

##### 17.4.1 模型的选择策略 ..... 352

##### 17.4.2 分析实例 ..... 352

#### 17.5 对数线性模型与其他模型的关系 ..... 359

##### 17.5.1 对数线性模型与方差分析模型的关系 ..... 359

##### 17.5.2 对数线性模型与 Logistic 回归的关系 ..... 359

#### 17.6 Poisson 回归模型 ..... 359

##### 17.6.1 模型简介 ..... 359

##### 17.6.2 分析实例 ..... 360

#### 思考与练习 ..... 361

#### 参考文献 ..... 362

### 第 18 章 信度分析 ..... 363

#### 18.1 信度理论入门 ..... 363

##### 18.1.1 真分数测量理论 ..... 363

##### 18.1.2 信度与效度 ..... 364

##### 18.1.3 内在信度与外在信度 ..... 364

##### 18.1.4 信度的判断标准 ..... 364

#### 18.2 简单分析实例 ..... 365

##### 18.2.1 Alpha 信度系数 ..... 365

##### 18.2.2 对各题目的深入分析 ..... 366

#### 18.2.3 对真分数理论假设的

#### 考察 ..... 368

#### 18.3 其余常用的信度系数 ..... 369

##### 18.3.1 重测信度 ..... 369

##### 18.3.2 折半信度 ..... 370

##### 18.3.3 Guttman 系数 ..... 371

##### 18.3.4 平行模型的信度系数 ..... 371

##### 18.3.5 严格平行模型的信度系数 ..... 372

##### 18.3.6 评分者信度 ..... 373

##### 18.3.7 信度系数总结 ..... 374

#### 18.4 信度理论进阶 ..... 375

##### 18.4.1 真分数测量理论的缺陷 ..... 375

##### 18.4.2 概化理论入门 ..... 375

##### 18.4.3 SPSS 中相应的分析功能 ..... 377

#### 思考与练习 ..... 378

#### 参考文献 ..... 378

### 第 19 章 生存分析 ..... 379

#### 19.1 生存分析简介 ..... 379

##### 19.1.1 生存分析简史 ..... 379

##### 19.1.2 生存分析中的基本概念 ..... 380

##### 19.1.3 生存分析的基本步骤 ..... 383

##### 19.1.4 SPSS 与生存分析 ..... 383

#### 19.2 生存函数的估计和检验 ..... 384

##### 19.2.1 生存函数的基本估计方法 ..... 384

##### 19.2.2 Kaplan - Meier 法 ..... 385

##### 19.2.3 寿命表法 ..... 391

##### 19.2.4 Kaplan - Meier 法和寿命表法比较 ..... 394

#### 19.3 Cox 回归模型 ..... 395

##### 19.3.1 Cox 模型入门 ..... 395

##### 19.3.2 分析实例 ..... 396

19.3.3 比例风险性的图形 验证 .....	399	<b>第 20 章 缺失值分析入门</b> .....	408
19.4 含时间依存性变量的 Cox 模型 .....	400	20.1 缺失值理论简介 .....	408
19.4.1 时依协变量的种类 .....	400	20.1.1 数据的缺失机制 .....	408
19.4.2 用时依模型验证比例 风险性 .....	401	20.1.2 SPSS 中对缺失值的处理 方法 .....	409
19.4.3 用时依模型评价处理因素的 影响 .....	402	20.2 对缺失情况的基本分析 .....	410
19.4.4 用时依模型评价重复测量 因子的影响 .....	403	20.2.1 缺失值数据的生成 .....	410
19.5 关于 Cox 模型的一些高级 话题 .....	404	20.2.2 对缺失模式的分析 .....	411
19.5.1 生存分析中的分层 变量 .....	404	20.2.3 缺失情况的统计 描述 .....	414
19.5.2 用 Cox 回归过程拟合配伍 Logistic 回归 .....	405	20.3 缺失值填充技术 .....	415
19.5.3 竞争风险的 Cox 模型 .....	406	20.3.1 列表输出 .....	416
思考与练习 .....	407	20.3.2 使用回归算法进行 填充 .....	417
参考文献 .....	407	20.3.3 使用 EM 算法进行 填充 .....	419
		20.3.4 多重填充技术简介 .....	421
		思考与练习 .....	422
		参考文献 .....	422
		<b>附录</b> .....	424

第一部分

一般线性与混合线性模型





# 第1章 方差分析模型

在本系列丛书的基础教程中,大家已经详细学习了 SPSS 软件的基本操作、图表绘制方法、统计描述技术和单因素统计分析方法。但是,在许多实际问题中,仅仅依靠统计描述或者简单的统计推断方法是不够的,现实世界中变量间的联系错综复杂,往往要同时考虑多个因素的作用,并为之建立多变量模型。而本章将要介绍的方差分析模型就是多变量模型中最为基础和常用的一种。

## 1.1 模型简介

在许多情况下,都需要同时研究多个因素对因变量的影响情况,比如要研究性别对身高的影响,显然就要考虑到年龄、遗传、营养状况等因素的作用。这时单因素分析方法是无能为力的,而以方差分析为代表的多因素分析方法可以在控制其他因素影响的同时研究两者之间的关系,因此,分析的效率更高,适用的范围更广。

同时,许多时候各自变量之间还会存在交互作用,如研究催化剂对化学反应的催化能力,如果该催化剂只在某个温度范围内效果最佳,则只单独研究该催化剂的催化作用是没有实际意义的,此时这种交互作用也成为了研究的重点,即必须要研究在什么温度条件下该催化剂的催化能力最佳。对交互作用的分析也是方差分析模型的特长。

### 1.1.1 模型入门

在基础教程中,已经学习了单因素方差分析方法,并从中了解到方差分析的基本思想是变异分解即根据资料类型以及研究目的,将样本的总变异分解为若干个部分,除有一部分代表随机误差的作用外,其余每个部分的变异分别代表了某个影响因素的作用(或交互作用),通过比较可能由某因素所致的变异与随机误差的大小,借助  $F$  分布做出推断,即可了解该因素对结果变量的影响是否存在。在多因素方差分析模型中,其方法原理没有任何的变化,只是模型中考虑的因素更多而已。

#### 1. 单因素方差分析模型的结构

为了让读者能够对方差分析模型有更为清楚的了解,下面以一个虚拟的例子来引入模型的基本结构,假设现在希望比较三种职业的月收入有无差异,这三类职业分别是医生、律师和软件工程师。那么最简单的做法就是在这三类人群中都进行随机抽样,各自得到一组受访者,收集他们的月收入状况,然后进行检验。则在此问题中,每一位受访者月收入的平均估计值  $Y_{ij}$  可以被表达为如下形式: