

多元统计分析高级教程

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 19.0 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 19 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 / 刘文彤主编. —北京: 高等教育出版社, 2014.12
ISBN 978-7-04-041360-9
I. ①刘… II. ①刘… III. ①统计分析—软件包—教材—研究生—教材 IV. ①O21
中国版本图书馆 CIP 数据核字(2014)第 242223 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 摇摇摇

出版发行 高等教育出版社	购书热线 010-58581000
社址 北京市西城区德外大街 4 号	免费咨询 800-8100022
邮政编码 100120	网 址 http://www.hep.com.cn
总机 010-64015068	邮 政 800-8100022

经销 新华书店北京发行所
印刷 北京人卫印刷厂

开 本 787mm×1092mm 1/16	版 次 第 1 版
印 张 18	印 次 第 1 次印刷
字 数 450 千字	定 价 38.00 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 954001

《多元统计分析高级教程》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 摇摇言

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、事业单位降低消耗、增加效益,从而提高整体运行效率,已成为各行业中有关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、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 100 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 2004 年编著的《SPSS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 2004—2005 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 经济的所有应用领域,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 经济资深用户,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:

经济最新版本的全模块试用版下载 : [http://www.btt.com.cn](#)

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

读者答疑、经验交流 : [http://www.btt.com.cn](#)

摇摇摇摇摇摇摇摇摇摇博塔(中国)有限公司 经济丛书编委会

前摇摇言

笔者前作《~~第1版~~统计分析报告》(基础篇)和《~~第1版~~统计分析报告》(高级篇)自~~2000~~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~~书后~~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~~第2版~~统计分析报告高级教程终于脱稿出版。本书主要针对~~第2版~~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~~第2版~~的功能为准,以统计理论为主线,详细介绍了~~第2版~~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~~逻辑回归~~模型族和~~判别分析~~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~~第2版~~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~~第2版~~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

~~2000~~年 苑月于复旦公共卫生学院

目摇摇录

第一部分摇一般线性与混合线性模型

第 1 章摇方差分析模型	猿	摇摇摇圆摇摇摇完全随机设计	圆
摇摇摇圆摇摇摇模型简介	猿	摇摇摇圆摇摇摇配伍组设计	圆
摇摇摇圆摇摇摇模型入门	猿	摇摇摇圆摇摇摇交叉设计	猿
摇摇摇圆摇摇摇常用术语	缘	摇摇摇圆摇摇摇拉丁方设计	猿
摇摇摇圆摇摇摇方差分析模型的适用 条件	苑	摇摇摇圆摇摇摇考虑交互作用的实验设计方案	猿
摇摇摇圆摇摇摇简单分析实例	愿	摇摇摇圆摇摇摇析因设计	猿
摇摇摇圆摇摇摇模型表达式	愿	摇摇摇圆摇摇摇正交设计	猿
摇摇摇圆摇摇摇初步分析结果	愿	摇摇摇圆摇摇摇均匀设计	猿
摇摇摇圆摇摇摇模型参数的估计值	愿	摇摇圆摇摇摇误差项变动的特殊实验设计 方案	愿
摇摇摇圆摇摇摇两两比较	愿	摇摇摇圆摇摇摇嵌套设计	愿
摇摇摇圆摇摇摇其他常用选项	猿	摇摇摇圆摇摇摇重复测量设计	愿
摇摇摇圆摇摇摇两因素方差分析模型	猿	摇摇摇圆摇摇摇裂区设计	愿
摇摇摇圆摇摇摇分析实例	猿	摇摇圆摇摇摇协方差分析	猿
摇摇摇圆摇摇摇边际均数与轮廓图	猿	摇摇摇圆摇摇摇协方差分析的必要性	猿
摇摇摇圆摇摇摇拟和劣度检验	猿	摇摇摇圆摇摇摇平行性假定的检验	猿
摇摇摇圆摇摇摇因素各水平间的精细比较	圆	摇摇摇圆摇摇摇计算和检验修正均数	愿
摇摇摇圆摇摇摇孕变量的悦子句	圆	摇摇思考与练习	愿
摇摇摇圆摇摇摇耘云云云云云杂子句	圆	摇摇参考文献	愿
摇摇摇圆摇摇摇粤云云云云云粤云云云云云 子句	圆	第 猿 章摇多元方差分析与重复测量方差分析	缘
摇摇摇圆摇摇摇粤云云云云云粤云云云云云 子句	圆	摇摇猿摇摇摇多元方差分析	缘
摇摇猿摇摇摇随机因素的方差分析模型	圆	摇摇猿摇摇摇模型简介	缘
摇摇猿摇摇摇其他问题	缘	摇摇猿摇摇摇分析实例	缘
摇摇猿摇摇摇自定义效应检验使用的 误差项	缘	摇摇猿摇摇摇检验统计量的计算	缘
摇摇猿摇摇摇源类方差分解方法	圆	摇摇猿摇摇摇对引例的进一步分析	缘
摇摇思考与练习	圆	摇摇猿摇摇摇重复测量资料的方差分析	缘
摇摇参考文献	圆	摇摇猿摇摇摇模型简介	缘
第 圆 章摇常用实验设计分析方法	愿	摇摇猿摇摇摇分析实例	缘
摇摇圆摇摇摇仅研究主效应的实验设计方案	愿	摇摇思考与练习	缘
		摇摇参考文献	缘

第 1 章 混合线性模型入门	远	摇摇摇摇摇摇拟合混合线性模型的基本	
摇摇摇摇模型简介	远	结构	苑
摇摇摇摇问题的提出	远	摇摇摇摇考虑重复测量间的	
摇摇摇摇模型入门	苑	相关性	愿
摇摇摇摇层次聚集性数据分析实例	苑	摇摇摇摇更改对测量间相关性的	
摇摇摇摇拟合混合线性模型的		假定	愿
基本结构	苑	摇摇摇摇模型中可用的相关阵	
摇摇摇摇在固定效应中加入		种类	愿
自变量	苑	摇摇摇摇模型总结	愿
摇摇摇摇在随机效应中加入		摇摇摇摇混合效应模型的用途	愿
自变量	苑	摇摇摇摇混合效应模型与一般线性	
摇摇摇摇更多解释变量的引入	苑	模型的联系	愿
摇摇摇摇其他常用选项	苑	摇摇思考与练习	愿
摇摇摇摇重复测量数据分析实例	苑	摇摇参考文献	愿
摇摇摇摇对数据的初步分析	苑		

第二部分 回归模型

第 2 章 多重线性回归模型	愿	摇摇摇摇本章小结	愿
摇摇摇摇模型简介	愿	摇摇摇摇回归模型的建立步骤	愿
摇摇摇摇简单分析实例	愿	摇摇摇摇多重线性回归模型结果解释	
摇摇摇摇对数据的初步分析	愿	时应注意的问题	愿
摇摇摇摇回归模型的假设检验	愿	摇摇思考与练习	愿
摇摇摇摇偏回归系数的假设检验	愿	摇摇参考文献	愿
摇摇摇摇标准化偏回归系数	愿	第 3 章 线性回归的衍生模型	愿
摇摇摇摇衡量多元线性回归模型		摇摇摇摇非直线趋势的处理——曲线	
优劣的标准	愿	直线化	愿
摇摇摇摇回归预测与残差分析	愿	摇摇摇摇方法简介	愿
摇摇摇摇回归预测与区间估计	愿	摇摇摇摇使用 蕴蕴过程进行	
摇摇摇摇残差分析与模型适用条件的		分析	愿
检验	愿	摇摇摇摇使用曲线拟合过程分析	愿
摇摇摇摇逐步回归	愿	摇摇摇摇方差不齐的处理——加权最小	
摇摇摇摇筛选自变量的基本原则	愿	二乘法	愿
摇摇摇摇常用的逐步回归方法	愿	摇摇摇摇方法简介	愿
摇摇摇摇分析实例	愿	摇摇摇摇使用 蕴蕴过程进行	
摇摇摇摇模型的进一步诊断与修正	愿	分析	愿
摇摇摇摇强影响点的识别与处理	愿	摇摇摇摇使用 蕴蕴过程分析	愿
摇摇摇摇多重共线性的识别与		摇摇摇摇共线性的处理——岭回归	愿
处理	愿	摇摇摇摇方法简介	愿

摇摇摇远源愿摇分析实例	员愿	摇摇摇愿缘愿摇参数初始值的设定	员员
摇摇摇远源愿摇分类变量的数值化——最优尺度 回归	员园	摇摇摇愿缘愿摇模型的拟合方法	员员
摇摇摇远源愿摇方法简介	员园	摇摇思考与练习	员圆
摇摇摇远源愿摇分析实例	员员	摇摇参考文献	员圆
摇摇摇远源愿摇最优尺度方法的应用注意 事项	员缘	第 愿章摇二分类 结果逻辑回归模型	员猿
摇摇思考与练习	员远	摇摇愿缘愿摇模型简介	员猿
摇摇参考文献	员远	摇摇摇摇愿缘愿摇模型入门	员猿
第 苑章摇路径分析入门	员苑	摇摇摇摇愿缘愿摇一些基本概念	员缘
摇摇苑缘愿摇两阶段最小二乘法	员苑	摇摇愿缘愿摇简单分析实例	员苑
摇摇摇摇苑缘愿摇模型简介	员苑	摇摇愿缘愿摇分类自变量的定义与比较方法	员员
摇摇摇摇苑缘愿摇使用 蕴蕴过程进行 分析	员愿	摇摇摇摇愿缘愿摇使用哑变量的必要性	员员
摇摇摇摇苑缘愿摇使用 圆蕴过程进行 分析	员园	摇摇摇摇愿缘愿摇杂务中预设的哑变量编码 方式	员猿
摇摇苑缘愿摇路径分析入门	员圆	摇摇摇摇愿缘愿摇设置哑变量时要注意的 问题	员缘
摇摇摇摇苑缘愿摇模型简介	员圆	摇摇愿缘愿摇自变量的筛选方法与逐步回归	员远
摇摇摇摇苑缘愿摇分析实例	员缘	摇摇摇摇愿缘愿摇模型中的假设检验方法	员远
摇摇思考与练习	员愿	摇摇摇摇愿缘愿摇自变量的筛选方法	员苑
摇摇参考文献	员愿	摇摇摇摇愿缘愿摇分析实例	员苑
第 愿章摇非线性回归模型	员怨	摇摇愿缘愿摇模型拟合效果与拟合优度检验	员圆
摇摇愿缘愿摇模型简介	员怨	摇摇摇摇愿缘愿摇模型效果的判断指标	员员
摇摇摇摇愿缘愿摇问题的提出	员怨	摇摇摇摇愿缘愿摇拟合优度检验	员源
摇摇摇摇愿缘愿摇模型入门	员怨	摇摇愿缘愿摇模型的诊断与修正	员远
摇摇愿缘愿摇简单分析实例	员圆	摇摇摇摇愿缘愿摇残差分析	员苑
摇摇摇摇愿缘愿摇软件操作与界面说明	员圆	摇摇摇摇愿缘愿摇多重共线性的识别及其对回归 系数的影响及处理办法	员苑
摇摇摇摇愿缘愿摇基本分析结果	员员	摇摇思考与练习	员苑
摇摇摇摇愿缘愿摇模型的进一步分析	员猿	摇摇参考文献	员愿
摇摇愿缘愿摇自定义损失函数——最小一乘法 实例	员猿	第 愿章摇多分类、配对 结果逻辑回归与 多元逻辑回归	员怨
摇摇摇摇愿缘愿摇分析实例	员源	摇摇愿缘愿摇有序多分类 结果逻辑回归模型	员怨
摇摇摇摇愿缘愿摇结果解释	员远	摇摇摇摇愿缘愿摇模型简介	员怨
摇摇愿缘愿摇分段回归模型的拟合	员苑	摇摇摇摇愿缘愿摇分析实例	员圆
摇摇摇摇愿缘愿摇分析实例	员苑	摇摇摇摇愿缘愿摇模型适用条件的检验	员猿
摇摇摇摇愿缘愿摇结果解释	员愿	摇摇愿缘愿摇无序多分类 结果逻辑回归模型	员缘
摇摇摇摇愿缘愿摇模型的进一步分析	员圆	摇摇摇摇愿缘愿摇模型简介	员缘
摇摇愿缘愿摇其他需要注意的问题	员员	摇摇摇摇愿缘愿摇分析实例	员远
		摇摇愿缘愿摇多元配对 结果逻辑回归	员怨

摇摇摇多元统计模型简介	圆源	摇摇摇多元统计模型简介	圆源
摇摇摇多元统计模型分析实例	圆源	摇摇摇多元统计模型分析实例二——计算	圆源
摇摇摇多元统计模型回归模型	圆源	摇摇摇多元统计模型回归模型	圆源
摇摇摇多元统计模型简介	圆源	摇摇摇多元统计模型简介	圆源
摇摇摇多元统计模型实例一——与	圆源	摇摇摇多元统计模型实例一——与	圆源

第三部分摇多元统计分析方法

第 员章摇主成分分析与因子分析	圆源	摇摇摇多元统计两步聚类法简介	圆源
摇摇摇多元统计主成分分析	圆源	摇摇摇多元统计聚类方法原理	圆源
摇摇摇多元统计模型入门	圆源	摇摇摇多元统计模型分析实例	圆源
摇摇摇多元统计简单分析实例	圆源	摇摇摇多元统计本章方法小结	圆源
摇摇摇多元统计对主成分分析的进一步 说明	圆源	摇摇摇多元统计思考与练习	圆源
摇摇摇多元统计因子分析	圆源	摇摇摇多元统计参考文献	圆源
摇摇摇多元统计模型入门	圆源	第 圆章摇判别分析	圆源
摇摇摇多元统计简单分析实例	圆源	摇摇摇多元统计模型简介	圆源
摇摇摇多元统计因子分析的进一步讨论	圆源	摇摇摇多元统计典型判别分析的基本 原理	圆源
摇摇摇多元统计不同的因子分析法	圆源	摇摇摇多元统计判别分析的适用条件和 违背条件时的处理方法	圆源
摇摇摇多元统计相关阵和协方差	圆源	摇摇摇多元统计判别效果的评价	圆源
摇摇摇多元统计确定公因子数量	圆源	摇摇摇多元统计判别分析的一般步骤	圆源
摇摇摇多元统计因子分析综合案例	圆源	摇摇摇多元统计简单分析实例	圆源
摇摇摇多元统计主成分分析和因子分析的 比较	圆源	摇摇摇多元统计软件操作与界面说明	圆源
摇摇摇多元统计思考与练习	圆源	摇摇摇多元统计基本分析结果	圆源
摇摇摇多元统计参考文献	圆源	摇摇摇多元统计判别结果的图形化 展示	圆源
第 圆章摇聚类分析	圆源	摇摇摇多元统计判别效果的验证	圆源
摇摇摇多元统计模型简介	圆源	摇摇摇多元统计适用条件的判断方法	圆源
摇摇摇多元统计问题的提出	圆源	摇摇摇多元统计贝叶斯判别分析	圆源
摇摇摇多元统计聚类分析入门	圆源	摇摇摇多元统计方法原理	圆源
摇摇摇多元统计聚类分析的方法体系	圆源	摇摇摇多元统计软件实现	圆源
摇摇摇多元统计层次聚类法	圆源	摇摇摇多元统计对判别分析的进一步讨论	圆源
摇摇摇多元统计方法原理	圆源	摇摇摇多元统计逐步判别分析	圆源
摇摇摇多元统计分析实例	圆源	摇摇摇多元统计判别分析和因子分析的 相似性和差异	圆源
摇摇摇多元统计对层次聚类法的进一步 讨论	圆源	摇摇摇多元统计二类判别和多重回归的 等价性	圆源
摇摇摇多元统计运用均值聚类法	圆源		
摇摇摇多元统计方法原理	圆源		
摇摇摇多元统计分析实例	圆源		

思考与练习	圆苑	摇摇摇摇员缘源摇摇分析实例	猿源
参考文献	圆苑	摇摇摇摇员缘源摇摇对应分析中的其他问题	猿愿
第 员章摇摇典型相关分析	圆愿	摇摇摇摇员缘源摇摇对应分析结果的正确解释	猿愿
摇摇摇摇员缘源摇摇方法介绍	圆愿	摇摇摇摇员缘源摇摇罕见类别和相似类别的处理	猿愿
摇摇摇摇员缘源摇摇典型相关分析的基本思想	圆愿	摇摇摇摇员缘源摇摇有序类别的处理	猿愿
摇摇摇摇员缘源摇摇典型相关分析的数学描述	圆怨	摇摇摇摇员缘源摇摇本章方法小结	猿怨
摇摇摇摇员缘源摇摇分析实例	圆园	摇摇摇摇员缘源摇摇对应分析与其他分析方法的关系	猿怨
摇摇摇摇员缘源摇摇两组变量间的相关系数	圆员	摇摇摇摇员缘源摇摇对应分析的优势与劣势	猿员
摇摇摇摇员缘源摇摇典型相关系数及显著性检验	圆圆	思考与练习	猿员
摇摇摇摇员缘源摇摇典型变量的系数	圆猿	参考文献	猿圆
摇摇摇摇员缘源摇摇典型结构分析	圆源	第 远章摇摇多维尺度分析	猿猿
摇摇摇摇员缘源摇摇典型冗余分析	圆远	摇摇摇摇员缘源摇摇古典 酝孕 模型	猿猿
摇摇摇摇员缘源摇摇小结	圆愿	摇摇摇摇员缘源摇摇方法原理	猿猿
摇摇摇摇员缘源摇摇典型相关分析的应用	圆愿	摇摇摇摇员缘源摇摇分析实例	猿源
摇摇摇摇员缘源摇摇典型相关分析和因子分析	圆愿	摇摇摇摇员缘源摇摇距离的计算方式	猿园
思考与练习	圆怨	摇摇摇摇员缘源摇摇非度量 酝孕 模型	猿员
参考文献	圆怨	摇摇摇摇员缘源摇摇数据测量尺度的设定	猿员
第 缘章摇摇对应分析	圆园	摇摇摇摇员缘源摇摇方法原理	猿圆
摇摇摇摇员缘源摇摇模型简介	圆园	摇摇摇摇员缘源摇摇分析实例	猿猿
摇摇摇摇员缘源摇摇问题的提出	圆园	摇摇摇摇员缘源摇摇考虑个体差异的 酝孕 模型	猿缘
摇摇摇摇员缘源摇摇模型入门	圆园	摇摇摇摇员缘源摇摇方法原理	猿缘
摇摇摇摇员缘源摇摇等级中的相应功能	圆员	摇摇摇摇员缘源摇摇分析实例	猿远
摇摇摇摇员缘源摇摇简单分析实例	圆员	摇摇摇摇员缘源摇摇空间定位图的含义解释	猿怨
摇摇摇摇员缘源摇摇对数据的初步分析	圆圆	摇摇摇摇员缘源摇摇基于最优尺度变换的 酝孕 模型	猿员
摇摇摇摇员缘源摇摇正式分析	圆猿	摇摇摇摇员缘源摇摇方法简介	猿员
摇摇摇摇员缘源摇摇对引例的进一步分析	圆苑	摇摇摇摇员缘源摇摇分析实例	猿源
摇摇摇摇员缘源摇摇基于均数的对应分析	圆怨	摇摇摇摇员缘源摇摇本章方法小结	猿苑
摇摇摇摇员缘源摇摇方法原理	猿园	思考与练习	猿苑
摇摇摇摇员缘源摇摇分析实例	猿员	参考文献	猿苑
摇摇摇摇员缘源摇摇多重对应分析	猿源		
摇摇摇摇员缘源摇摇方法原理	猿源		

第四部分 其他统计分析方法

第 5 章 对数线性模型与 多元回归	猿猿猿
回归	猿猿猿
对数线性模型简介	猿猿猿
对数线性模型问题的提出	猿猿猿
对数线性模型入门	猿猿猿
对数线性模型杂项的相应功能	猿猿猿
一般对数线性模型分析实例	猿猿猿
对数据的初步分析	猿猿猿
正式分析	猿猿猿
对引例的进一步分析	猿猿猿
因果关系明确时的对数线性模型	猿猿猿
对数线性模型的选择	猿猿猿
对数线性模型的选择策略	猿猿猿
对数线性模型分析实例	猿猿猿
对数线性模型与其他模型的关系	猿猿猿
对数线性模型与方差分析模型的关系	猿猿猿
对数线性模型与 多元回归的关系	猿猿猿
多元回归模型	猿猿猿
对数线性模型简介	猿猿猿
对数线性模型分析实例	猿猿猿
思考与练习	猿猿猿
参考文献	猿猿猿
第 6 章 信度分析	猿猿猿
信度理论入门	猿猿猿
真分数测量理论	猿猿猿
信度与效度	猿猿猿
内在信度与外在信度	猿猿猿
信度的判断标准	猿猿猿
简单分析实例	猿猿猿
重测信度系数	猿猿猿
对各题目的深入分析	猿猿猿

摇摇摇	愿源猿	摇摇对真分数理论假设的考察	猿愿
摇摇愿	猿猿	摇摇其余常用的信度系数	猿怨
摇摇摇	愿猿猿	摇摇重测信度	猿怨
摇摇摇	愿猿猿	摇摇折半信度	猿园
摇摇摇	愿猿猿	摇摇耶伦森系数	猿员
摇摇摇	愿猿猿	摇摇平行模型的信度系数	猿员
摇摇摇	愿猿猿	摇摇严格平行模型的信度系数	猿圆
摇摇摇	愿猿猿	摇摇评分者信度	猿猿
摇摇摇	愿猿猿	摇摇信度系数总结	猿源
摇摇愿	猿猿	摇摇信度理论进阶	猿缘
摇摇摇	愿源猿	摇摇真分数测量理论的缺陷	猿缘
摇摇摇	愿源猿	摇摇概化理论入门	猿缘
摇摇摇	愿源猿	摇摇杂项中相应的分析功能	猿苑
摇摇思考与练习			猿愿
摇摇参考文献			猿愿
第 员章摇摇生存分析			猿怨
摇摇愿	猿猿	摇摇生存分析简介	猿怨
摇摇摇	愿猿猿	摇摇生存分析简史	猿怨
摇摇摇	愿猿猿	摇摇生存分析中的基本概念	猿园
摇摇摇	愿猿猿	摇摇生存分析的基本步骤	猿猿
摇摇摇	愿猿猿	摇摇杂项与生存分析	猿猿
摇摇愿	猿猿	摇摇生存函数的估计和检验	猿源
摇摇摇	愿猿猿	摇摇生存函数的基本估计方法	猿源
摇摇摇	愿猿猿	摇摇运费表原酝法	猿缘
摇摇摇	愿猿猿	摇摇寿命表法	猿员
摇摇摇	愿猿猿	摇摇运费表原酝法和寿命表法比较	猿源
摇摇愿	猿猿	摇摇悦曾回归模型	猿缘
摇摇摇	愿猿猿	摇摇悦曾模型入门	猿缘
摇摇摇	愿猿猿	摇摇分析实例	猿远

摇摇摇摇比例风险性的图形验证	猿怨	第 圆章摇摇缺失值分析入门	源愿
摇摇摇摇含时间依存性变量的 悦着模型	源园	摇摇摇摇摇摇缺失值理论简介	源愿
摇摇摇摇摇摇时依协变量的种类	源园	摇摇摇摇摇摇数据的缺失机制	源愿
摇摇摇摇摇摇用依模型验证比例风险性	源员	摇摇摇摇摇摇杂等中对缺失值的处理方法	源怨
摇摇摇摇摇摇用依模型评价处理因素的影响	源圆	摇摇摇摇对缺失情况的基本分析	源园
摇摇摇摇摇摇用依模型评价重复测量因子的影响	源猿	摇摇摇摇摇摇缺失值数据的生成	源园
摇摇摇摇关于 悦着模型的一些高级话题	源源	摇摇摇摇摇摇对缺失模式的分析	源员
摇摇摇摇摇摇生存分析中的分层变量	源源	摇摇摇摇摇摇缺失情况的统计描述	源源
摇摇摇摇摇摇用 悦着回归过程拟合配伍结果摇摇回归	源缘	摇摇摇摇摇摇缺失值填充技术	源缘
摇摇摇摇摇摇竞争风险的 悦着模型	源远	摇摇摇摇摇摇列表输出	源远
摇摇思考与练习	源苑	摇摇摇摇摇摇使用回归算法进行填充	源苑
摇摇参考文献	源苑	摇摇摇摇摇摇使用 耘算法进行填充	源怨
		摇摇摇摇摇摇多重填充技术简介	源员
		摇摇思考与练习	源圆
		摇摇参考文献	源圆
		附录	源原

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 方差分析模型简介

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。

1.2 方差分析模型入门

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。

1.2.1 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,收集它们的月收入状况,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 $\bar{y}_j$ 可以被表达为如下形式:

再越垣

其中 $\mu_{\text{再}}$ 代表第 再 个职业组中第 再 位受访者的具体收入。显然,在此表达式中 $\mu_{\text{再}}$ 表示某一个职业组的平均收入, 再 的取值范围为 员~猿,分别代表三种职业之一,而 $\varepsilon_{\text{再}}$ 表示第 再 组的第 再 位受访者的随机误差,反映的是因各种原因导致的该受访者月收入 and 该职业平均收入间的差异。下面来看模型中对 $\varepsilon_{\text{再}}$ 的设定,模型中假设各组的 $\varepsilon_{\text{再}}$ 服从同一个正态分布,即无论 再 取值是多少, $\varepsilon_{\text{再}}$ 均服从同一个均数为 园标准差为某个定值的正态分布 晕(园, $\sigma^{\text{圆}}$)。这样一来,如果三种职业收入无差异,则它就应当等于总体均数(平均水平)再加上一个随机误差项,实际上就变成了同一个变量的分布 晕(μ , $\sigma^{\text{圆}}$)。为了能够对收入水平进行预测,人们又规定 耘再越 $\mu_{\text{再}}$,即第 再 组个体的收入估计值等于该组的平均水平,结合模型结构,这应当不难理解。实际上,如果对应样本数据,该预测值就是各组的样本均数。

为了统计推断的需要,以上模型往往被改写成如下形式:再越垣 $\alpha_{\text{再}}$ 垣 $\varepsilon_{\text{再}}$,其中 μ 表示不考虑职业时收入总的平均水平, $\alpha_{\text{再}}$ 表示职业为 再 类时的附加效应,即在 再 职业时平均收入水平的改变情况。例如 $\alpha_{\text{再}} > 0$ 表明当职业为 再 类时,平均收入要比总的平均水平高 员元。如果职业员和职业猿的平均收入不相等,则应当有 $\alpha_{\text{再}} \neq \alpha_{\text{猿}}$,反之,如果三种职业的平均收入无差异,则因为各类均不存在附加效应,应当有 $\alpha_{\text{再}} = 0$ 。因此,如果要检验职业种类是否对收入有影响,就是检验如下假设:

匀:对任意的 再 取值,都有 $\alpha_{\text{再}} = 0$;至少有一个 $\alpha_{\text{再}} \neq 0$

在基础篇中,已经学习了方差分析的基本思想是变异分解,例如在单因素方差分析中总变异被分解为如下两部分:总变异越处理因素导致的变异垣随机变异。现在对照上述模型表达式,大家就会发现实际上 $\alpha_{\text{再}}$ 就对应了所谓处理因素导致的变异,而 $\varepsilon_{\text{再}}$ 就对应了相应的随机变异。

但是 μ 、 $\alpha_{\text{再}}$ 等显然应当是一个相对的大小,例如职业员比职业猿的平均收入高 员元,则当 $\alpha_{\text{猿}}$ 为 缘元时 $\alpha_{\text{再}}$ 就应当是 员元, $\alpha_{\text{猿}}$ 为 员元时 $\alpha_{\text{再}}$ 就应当是 员元,总之加上 员元即可。为了能够在实际问题中得到对 μ 、 $\alpha_{\text{再}}$ 具体的估计值,模型拟合中又会对它们有一些附加的设定,这被称为模型拟合时的约束条件,详细介绍见后面有关章节。

圆两因素方差分析模型的结构

下面开始对单因素模型进行扩展。同样是上面的问题,有研究人员提出:性别应当也对收入水平有影响,也许正是因为 员组中男性比例要高于 圆组,才导致 员组的收入均数高于后者,因此,应考虑控制性别的作用。如果要同时考虑性别和职业对收入的影响,则建立的基本模型如下:

$$\text{再越垣}\mu\text{垣}\alpha_{\text{再}}\text{垣}\beta_{\text{再}}\text{垣}\varepsilon_{\text{再}}$$

摇摇该模型对应了如下变异分解方式:

总变异越职业导致的变异垣性别导致的变异垣随机变异

基本模型中 $\alpha_{\text{再}}$ 、 $\beta_{\text{再}}$ 分别表示职业为 再 类、性别为 再 类时的附加效应, $\varepsilon_{\text{再}}$ 仍为服从某个正态分布的随机误差变量。此时如果要说明职业种类对收入有无影响,就是检验如下假设:

匀: $\alpha_{\text{再}} = 0$;至少有一个 $\alpha_{\text{再}} \neq 0$

此时性别的影响因被包含在了 $\beta_{\text{再}}$ 中,从而不会影响到对职业的检验。

如果要说明因素 月(职业)有无影响,就是检验如下假设:

源元素(标准领域)

元素指用于测量因变量值的最小单位,比如研究在上文所提到的收入的例子中,元素就是每一位受访者。而在配伍设计等重复测量问题中,元素可能是受试者每一次具体的测量。根据具体的实验设计,一个单元格内可以有多个元素,也可以只有一个,甚至没有元素。

源均衡(标准领域)

如果在一个实验设计中任一因素各水平在所有单元格中出现的次数相同,且每个单元格内的元素数均相同,则该试验是均衡的;否则,就被称为不均衡。不均衡的实验设计在分析时较为复杂,需要对方差分析模型作特别设置才能得到正确的分析结果,详细介绍见后面有关章节。

源协变量(标准领域)

协变量指对因变量可能有影响,需要在分析时对其作用加以控制的连续性变量,实际上,可以简单地把因素和协变量分别理解为分类自变量和连续性自变量。当模型中存在协变量时,一般是通过找出它与因变量的回归关系来控制其影响,详情参见协方差分析部分。

源交互作用(标准领域)

如果一个因素的效应大小在另一个因素不同水平下明显不同,则称为两因素间存在交互作用。当存在交互作用时,单纯研究某个因素的作用是没有意义的,必须区分另一个因素的不同水平研究该因素的作用大小。

如果所有单元格内都至多只有一个元素,则交互作用无法进行分析,只能不予考虑,最典型的例子就是配伍设计的方差分析。

源固定因素(标准领域)与随机因素(标准领域)

两者都是因素的不同种类,固定因素指的是该因素在样本中所有可能的水平都出现了。换言之,该因素的所有可能水平仅此几种,针对该因素而言,从样本的分析结果中就可以得知所有水平的状况,无需进行外推。比如要研究三种促销手段的效果有无差别,所有样本只会是三种促销方式之一,不存在第 源种促销手段的问题,则此时该因素就被认为是固定因素。

和固定因素相对应的是随机因素,它指的是该因素所有可能的取值在样本中没有都出现,或不可能都出现。换言之,目前在样本中的这些水平是从总体中随机抽样而来,如果重复本研究,则可能得到的因素水平会和现在完全不同,这时,研究者显然希望得到的是一个能够“泛化”,即对所有可能出现的水平均适用的结果。例如研究广告类型和投放的城市对产品销量是否有影响,在设计中随机抽取了 源个城市进行研究,显然,研究者希望分析结果能够外推到全国的所有大、中型城市,此时就涉及将结果外推到抽样未包括的城市中的问题,在这种情况下,城市就应当是一个随机因素。又如研究什么温度下催化剂的效果最好,因经费有限,样本中只取了 猿益、源益、缘益三个水平,但是我们希望研究的是整个有效温度范围内哪个温度的效果最好,即在分析结果中能同时外推 猿益、源益这些水平的情况,此时温度也应当是随机因素。

一般来说固定因素和随机因素在分析时应分别指定,如果将随机因素按固定因素来分析,则可能得出错误的分析结果。但是,在许多时候,判断一个因素究竟是固定因素还是随机因素并不是件容易的事情。在这里需要提醒各位读者的是:区别这两者的并非是该因素本身的特性,而是我们的分析目的,假如将其看成是固定因素,则结论就不应当外推到未出现的其他水平中去,否则,就应当考虑按照随机因素来分析。