

21世纪社会学系列教材

# 社会统计 分析方法 — SPSS软件应用

郭志刚  
主编

中国人民大学出版社



社會統計  
分析方法  
— 統計軟件應用

21 世纪社会学系列教材

# 社会统计分析方法

## ——SPSS 软件应用

郭志刚 主编

中国人民大学出版社

## 图书在版编目 (CIP) 数据

社会统计分析方法——SPSS 软件应用/郭志刚主编.  
北京: 中国人民大学出版社, 1999  
21 世纪社会学系列教材

ISBN 7-300-02898-5/C·172

I. 社…

II. 郭…

III. ①社会统计-多元分析-教材②多元分析-软件包、SPSS

IV. C91-03

中国版本图书馆 CIP 数据核字 (1999) 第 68228 号

21 世纪社会学系列教材

**社会统计分析方法**

——SPSS 软件应用

郭志刚 主编

---

出版发行: 中国人民大学出版社

(北京海淀路 157 号 邮编 100080)

发行部: 62514146 门市部: 62511369

总编室: 62511242 出版部: 62511239

E-mail: rendafx@public 3. bta. net. cn

经 销: 新华书店

印 刷: 北京市鑫鑫印刷厂

---

开本: 787×980 毫米 1/16 印张: 31.5

1999 年 12 月第 1 版 1999 年 12 月第 1 次印刷

字数: 566 000 印数: 1-5 000

---

定价: 39.00 元

(随书赠送数据磁盘, 如有问题, 本社负责调换)



## 序

当代计算机技术的飞跃发展，不仅影响到自然科学和人类生产、生活方式的转变，也影响到社会科学的发展。计算机及其软件的日益普及，使得过去许多束之高阁的统计分析方法成为社会科学研究得心应手的工具。我们必须跟上这一时代发展的步伐，大力推广新兴统计方法和相应计算机软件的应用，提高我国社会科学研究的水平，更好地为我国的社会经济发展和两个文明建设服务。正是在这样一个背景之下，我们将本书呈献给读者，希望能够起到铺路搭桥的作用。

早在 1987 年，我与郝虹生、杜亚军、曲海波开始合作写作一本社会调查研究的教科书。经过一个冬夏轮回的苦干，终于在 1988 年 5 月完成了书稿，书名定为《社会调查研究的量化方法》，1989 年 3 月由中国人民大学出版社正式出版。该书受到了读者的欢迎，几次印刷都不能满足需要。在该书的基础上，中国人民大学人口研究所和人口学系开设了一系列本科生和研究生的社会调查研究课程。同时，该书在兄弟院校的教学中也被选作教材或参考书。在该书中，我负责撰写统计分析部分。在实际应用过程中，我深感到我所撰写的这一部分，有两个较大的不足。

首先，该书的统计分析部分没有与计算机软件的应用相联系，这个缺陷大大限制了该教材的实用性。在后来的实际教学中，我们进行了将统计原理的讲授与

计算机应用相联系的探索，先是选用了 SYSTAT 软件，后来又转而选用 SPSS 软件。这些教学实践取得了许多宝贵的经验。

其次，该书的统计分析部分偏重于基础性，涵盖面相对窄一些，对有的统计分析方法没有加以系统介绍。然而，在当时由于作者水平有限，各种资料也十分缺乏，要弥补这个不足是很困难的。认识到了这一不足，就是给自己提出了新的要求。在随后的几年中，我利用几次出国进修、访问、做博士后研究的机会，选修课程、收集资料，并制定了组织力量再编写一本新的统计分析教材的计划。

这本面向 21 世纪的、新的社会统计教材追求以下一些特点：

1. 强调多元分析，介绍多种分析方法，以满足研究分析中的不同要求。
2. 深入浅出，不要过多的数学证明，强调实际意义说明，并辅之以例题分析，以方便数学基础不太强的读者进行学习。也就是说，本书虽然介绍的是一些多元统计分析方法，却不想使读者感到过于深奥而望而生畏。
3. 每种方法都要与相应的计算机统计软件结合起来介绍，从原理的讨论到具体案例分析示范及统计软件使用说明，成为一体化的内容。
4. 强调指出各种方法使用中常见的理解错误和使用不当的讨论。也就是说，不但要追求统计技术的普及化，而且要避免过于简单化，并且，附加一些研究方法论方面的讨论。

根据上述设想，教材写作的工作量太大，很难在短期内由一个人独自完成，于是我采取了按照上述设想邀请多位撰稿人同时参与工作的方式来编写这本新的社会统计教材。一方面是为了尽快出版以满足教学和科研方面的急需，另一方面也是为了借助各位专家的特长。本书所有撰稿人都具有深厚的统计理论基础，并且都拥有多年统计教学和统计研究的实际经验。其中三位撰写人是我同一教研室的同事，经常一起切磋统计问题；还有三位是我的老同学，他们在美国获得博士学位，目前仍在美国从事教学和科研工作，他们虽然总是处于紧张工作之中，但对于为国内学生写作这样一本统计教材表现出了极大的热忱。编写这本教材得到了所有撰稿人的鼎力支持，才有可能在最短的时间内完成编写工作，并于 1998 年 5 月交付中国人民大学出版社。作为主编，我要感谢所有撰写者在百忙之中挤出时间积极参加这项颇费精力的工作，并按照进度表高质量地完成了初稿；还要感谢他们放弃了个人的偏好，按照我制定的写作大纲和体例方面的要求进行了写作和多次修改。呈献在读者面前的这一本教材便是这一精诚合作的最终结晶。

本书各章撰写人姓名及其工作单位如下：

第一章 导论

郭志刚

中国人民大学人口研究所

第二章	多元线性回归	高凌、郭志刚	中国人民大学人口研究所
第三章	因子分析	刘金塘	中国人民大学人口研究所
第四章	聚类分析	刘金塘	中国人民大学人口研究所
第五章	通径分析	郭志刚	中国人民大学人口研究所
第六章	logistic 回归	郭志刚	中国人民大学人口研究所
第七章	对数线性模型	郭志刚	中国人民大学人口研究所
第八章	多元方差分析	郭志刚	中国人民大学人口研究所
第九章	鉴别分析	郭志刚	中国人民大学人口研究所
第十章	典型相关分析	郭志刚	中国人民大学人口研究所
第十一章	结构方程模型	王济川 (WANG Jichuan, Ph.D.)	美国怀特州立大学医学院社区 卫生系 ( Department of Community Health, School of Medicine, Wright State University, OH, USA)
第十二章	事件史分析	梁 在 (LIANG Zai, Ph.D.)	美国纽约市立大学皇后学院社 会学系 ( Department of Sociology, Queens College, City Univer- sity of New York, NY, USA)
第十三章	使用 SPSS 软 件对事件史 原始数据进 行预处理	郭申阳 (GUO Shenyang, Ph.D.)	美国凯西·西保留地大学应用 社会科学学院 (Mandel School of Applied So- cial Sciences, Case Western Reserve University, Clevel- and, OH, USA)
第十四章	对应分析	杜 鹏	中国人民大学人口研究所

这本教材在完成过程中既有所得也有所失。在最初策划时，曾计划各种方法都应用同一套数据来进行示范，以免读者实践时在输入数据上花费太多的精力。但是，由于我事先没有做好这方面的准备，同时由于本书撰稿人很多，且天各一方，一旦开始撰写便很难再协调成所有例子都使用同一套数据，以至最终不得不

放弃了这一要求。这不能不说是本书的一个遗憾。

另外，按照原计划的体裁要求，每个例子的案例数都不超过 100，这原本也是从读者容易实验操作方面考虑的。但一些统计方法要求较大的样本规模才能取得统计显著的结果，因此撰稿人的初稿中的某些例子还是涉及了较大的数据样本。为了便利读者，主编与各撰稿人又做了多次修改，尽量压缩例题的数据量，有时不得不以牺牲统计显著为代价。尽管这样做可以减轻读者在运用这些统计方法时输入数据的负担，但仍然要花费许多精力。感谢中国人民大学出版社的支持，采取在书中附加一块数据磁盘最终解决了这一问题。

本书选择了 SPSS 软件作为计算机统计软件加以应用，主要是因为这一软件比较易学且已在国内广为流行。虽然许多专业统计人员都感到这一软件有时不能满足实践中一些特殊的需要，但由于本教材的主要目的并不在于精深、而在于普及，所以我们仍然选用 SPSS 软件，这对于初学者和一般用户是相适宜的。

本教材的大部分内容都经过了教学实践的检验。特别是第二次开设课程时，不仅本所的学生踊跃参加，还有本校社会学系、工商管理学院、国际经济系的研究生以及一些其他研究部门的人员也选修了这一课程。他们对于本书的内容和形式给予了充分肯定，并提出了许多宝贵的意见和建议，相当一部分已经得到采纳，在全书最后的统编过程中得到体现。在一定程度上，也可以说本书的最终形成是教学相长的产物。人口研究所学术委员会对于这本教材的写作及相应课程的开设始终给予鼓励和支持；本所计算机室黄可尤工程师在设备方面提供了大力支持和帮助，博士研究生陈功、邓国胜、宋健等承担了校阅工作，并就本书的内容和形式提出了很多好的建议。这些支持对于本书最后统稿工作起到十分重要的促进作用。

另外，还要特别感谢潘宇同志作为本书的责任编辑对本书的最终出版所做的努力。

由于王济川、梁在、郭申阳三位博士多年在美国进行教学科研活动，他们工作负担很重，为了能够减轻他们写作的负担，最大限度地利用已有的教学科研成果，我作为主编同意他们直接用英文写作他们所承担的各章，再由我本人和陈卫博士将这些初稿翻译成中文，然后由我完成编辑工作。王济川博士所撰写的第十一章采用原版 LISREL 软件分析结构模型。为了帮助读者应用视窗 SPSS 第 6 版中附带的 LISREL 软件，我为该章加了一个附录。并且，我还为郭申阳博士所撰写的第十三章中的程序命令添加了更为详细的说明。最后，中文稿全部经过撰写者本人的审阅和修订。

本书内容涉及大量统计学的新概念、新词汇，在翻译这些词汇中，曾参考了



不同统计学词典中的定名，有些词汇目前没有标准译法、或译法不当，则不得不由主编与各位撰写人商量确定其翻译。为了与国际统计学界保持一致，也使得读者和同行明了我们所用词汇的来源和原义，我们对于关键词汇以及尚无统一翻译准则的词汇都同时给予英文原文作为参考。本书的译法和表述法只是一家之言，这些概念和词汇的标准化还需要通过我国统计学的进一步发展得到最后的确定。

由于各撰写人写作和最后统稿的时间都很仓促，本书中恐难免存在一些缺点和不当之处，我们由衷地希望得到读者和同行的批评指正。

**郭志刚**

于北京塔院迎春园

1999年5月1日



# 目 录

<b>第一章 导论</b> .....	(1)
一、关于本书各章的简介.....	(2)
1. 变量的测度等级 .....	(2)
2. 本书所介绍的分析方法的概述 .....	(4)
3. 本书介绍的统计方法的分类框架 .....	(5)
二、在应用统计的研究中，要以研究方法论为指导.....	(8)
1. 理论、观察、统计之间的关系 .....	(8)
2. 统计研究中的常见谬误 .....	(10)
3. 社会科学的研究对象和模型 .....	(13)
4. 统计分析与理论分析 .....	(14)
5. 真理性的检验与统计检验 .....	(14)
<b>第二章 多元线性回归</b> .....	(17)
一、变量的关系和回归的任务 .....	(18)
二、一元线性回归模型 .....	(20)
1. 一元线性回归方程 .....	(20)
2. 一元回归系数的意义 .....	(22)

3. 变量变换 .....	(22)
4. 最小二乘估计的统计性质 .....	(25)
5. 模型的假设条件 .....	(25)
三、多元线性回归模型 .....	(29)
1. 多元线性回归方程 .....	(29)
2. 回归平面和回归系数的意义 .....	(30)
3. 一般回归模型 .....	(32)
4. 多元回归模型估计的统计推断 .....	(33)
四、方程的解释能力 .....	(34)
1. 确定系数 $R^2$ .....	(34)
2. 调整的决定系数 $R_{\text{adj}}^2$ .....	(36)
3. 多元相关系数 $R$ .....	(36)
4. 偏确定系数 .....	(37)
5. 偏相关系数 .....	(40)
6. 方差分析 .....	(40)
五、回归方程的检验和回归系数的推断统计 .....	(41)
1. 回归方程的显著性检验 .....	(41)
2. 回归系数的显著性检验 .....	(42)
3. 回归系数的置信区间 .....	(43)
4. 回归系数不显著的原因 .....	(44)
六、标准化回归系数 .....	(44)
七、回归预测的区间估计 .....	(45)
1. $\hat{Y}_0$ 的区间估计 .....	(46)
2. $Y_0$ 的预测 .....	(48)
八、多重共线性及其解决方法 .....	(49)
1. 多重共线性的含义及其影响 .....	(49)
2. 多重共线性的检验 .....	(51)
3. 补救多重共线性影响的办法 .....	(51)
4. 最优回归方程的选择 .....	(53)
九、虚拟变量的应用 .....	(55)
1. 虚拟变量的建立 .....	(55)
2. 虚拟变量回归系数的意义 .....	(57)
3. 采用虚拟变量的回归分析的检验 .....	(61)

4. 其他形式的分类变量编码和应用 .....	(63)
附录 关于 SPSS for Windows 进行回归分析的有关操作 .....	(68)
1. SPSS 基本操作 .....	(68)
2. 线性回归的基本操作 .....	(70)
3. 回归分析假设条件的检查 .....	(72)
4. 其他回归分析指标输出 .....	(79)
5. 要求 SPSS 输出偏回归系数的方差协方差矩阵并计算矩阵 $C$ .....	(83)
6. 将 SPSS 回归结果存成新变量 .....	(84)
<b>第三章 因子分析</b> .....	(87)
一、因子分析原理 .....	(88)
1. 因子分析模型 .....	(88)
2. 因子分析中的有关概念 .....	(90)
3. 因子分析的步骤 .....	(92)
二、求解初始因子 .....	(95)
1. 主成分分析法 .....	(95)
2. 公因子分析法 .....	(100)
3. 因子求解方法对结果的影响 .....	(101)
三、解释因子 .....	(102)
1. 正交旋转方法 .....	(104)
2. 斜交旋转方法 .....	(105)
3. 旋转方法的选择 .....	(107)
4. 解释因子 .....	(107)
四、因子值及其应用 .....	(109)
五、用 SPSS 软件进行因子分析 .....	(111)
1. Extraction 对话框 .....	(112)
2. Rotation 对话框 .....	(112)
3. Factor Scores 对话框 .....	(113)
4. Descriptive 对话框 .....	(113)
5. Options 对话框 .....	(114)
<b>第四章 聚类分析</b> .....	(117)
一、聚类分析的主要步骤 .....	(118)
1. 选择变量 .....	(118)

2. 计算相似性 .....	(118)
3. 聚类 .....	(119)
4. 聚类结果的解释和证实 .....	(119)
二、相似性测度 .....	(119)
1. 相关测度 .....	(119)
2. 距离测度 .....	(121)
3. 关联测度 .....	(122)
4. 数据的标准化问题 .....	(124)
三、聚类方法 .....	(125)
1. 层次聚类法 .....	(125)
2. 迭代聚类法 .....	(136)
3. 分类数的确定 .....	(138)
4. 聚类方法的选择 .....	(140)
四、聚类结果的解释和证实 .....	(140)
五、利用 SPSS 软件进行聚类分析 .....	(142)
1. Hierarchical Cluster .....	(142)
2. k-means Cluster .....	(143)
<b>第五章 通径分析</b> .....	(145)
一、引言 .....	(145)
二、通径模型的设置 .....	(147)
三、递归通径模型与非递归通径模型 .....	(148)
1. 递归通径模型 .....	(148)
2. 非递归通径模型 .....	(149)
3. 递归通径模型分析的假设条件 .....	(152)
四、分解简单回归系数的通径分析 .....	(152)
1. 计算一个变量对最终反应变量 (ultimate response variable) 的各种影响 .....	(153)
2. 以不同通径传递的间接影响 .....	(154)
3. 在控制某些变量的条件下的总影响的分解工作 .....	(155)
4. 标准化与非标准化的通径系数 .....	(156)
5. 用列表法报告各种影响作用分解 .....	(157)
五、分解简单相关系数的通径分析 .....	(159)
六、通径模型的调试与检验 .....	(165)

1. 通径模型的调试 .....	(166)
2. 通径模型的识别 .....	(168)
3. 对过度识别的通径模型的整体检验方法 .....	(171)
<b>第六章 logistic 回归 .....</b>	<b>(177)</b>
一、引言 .....	(177)
二、从多元线性回归到 logistic 回归 .....	(178)
1. 多元线性回归中使用定性因变量严重违反本身假设条件 .....	(178)
2. 线性概率模型及其问题 .....	(180)
三、logistic 回归模型的建立 .....	(181)
1. logistic 函数及其性质 .....	(182)
2. logistic 回归模型因变量的不同形式 .....	(185)
四、logistic 回归系数的意义 .....	(188)
1. 以 logit $p$ 方程的线性表达式来解释回归系数 .....	(188)
2. 以发生比 $\Omega$ 的指数表达式来解释回归系数 .....	(190)
3. 几种特殊情况的讨论 .....	(192)
4. 标准化的 logistic 回归系数 .....	(194)
五、logistic 模型各参数估计的评价及统计检验 .....	(194)
1. 对于整体模型的检验 .....	(194)
2. 回归系数的检验 .....	(196)
3. 系数子集的联合假设检验 .....	(197)
六、示范例题 .....	(198)
1. 对两个自变量同时纳入模型进行回归 .....	(199)
2. 两个自变量并加上交互作用项的回归 .....	(203)
3. 自变量组分步纳入的回归 .....	(205)
4. 自动选择显著自变量的分析 .....	(209)
<b>第七章 对数线性模型 .....</b>	<b>(215)</b>
一、从常规交互表分析到对数线性模型分析 .....	(216)
1. 传统交互表分析的缺点 .....	(216)
2. 对数线性模型的发展 .....	(218)
二、对数线性模型的基本原理 .....	(218)
1. 例 1 分析模型的背景情况 .....	(218)
2. 交互表单元频数的对数线性模型化 .....	(220)
3. 对数线性模型参数估计值的理解 .....	(223)

4. 对数线性模型的统计检验 .....	(226)
5. logit 模型分析 .....	(232)
三、两种不同类型的输入数据格式及本章例题数据 .....	(236)
1. 以交互单元为记录的数据及 SPSS 中要求加权的操作 .....	(236)
2. 以原始案例为记录的数据 .....	(238)
四、SPSS 对数线性模型分析各种功能的介绍 .....	(239)
1. 选用不同模型 .....	(239)
2. 设置模型中变量 .....	(240)
3. 定义模型中的项目 .....	(240)
4. 设置 Delta 值和参照组类别 .....	(242)
5. 设置输出类型 .....	(243)
6. 在分层模型分析中增设自动筛选显著效应项 .....	(244)
7. 对于 SPSS 对数线性模型各主要功能的归纳 .....	(245)
五、例 2 分析 .....	(245)
1. 对于例 2 饱和模型的分层检验及筛选出重要效应项 .....	(246)
2. 采用一般模型来取得例 2 简约模型的参数估计 .....	(250)
3. 对例 2 简约 logit 模型参数估计的分析 .....	(252)
<b>第八章 多元方差分析</b> .....	(257)
一、简介多元方差分析与一元方差分析的关系 .....	(257)
1. 从 $t$ 检验到一元方差分析 .....	(258)
2. 从一元方差分析到多元方差分析 .....	(259)
二、多元方差分析的数据要求和假设条件 .....	(261)
三、例题数据及三个分析模型 .....	(262)
四、第一模型：单因素二元模型 .....	(264)
1. SPSS 多元方差分析中单因素模型及其他检查的设置 .....	(264)
2. 第一模型分析输出的结果及讨论 .....	(266)
3. 多元方差分析假设条件的检查 .....	(268)
4. 关于第一模型分析的小结 .....	(270)
五、多元方差分析与一元方差分析区别的图示说明 .....	(270)
六、第二模型：双因素二元饱和模型 .....	(277)
1. SPSS 多元方差分析中多因素饱和模型的设置 .....	(277)
2. 第二模型分析输出的结果及讨论 .....	(278)
七、第三模型：双因素二元非饱和模型 .....	(280)

1. SPSS 多元方差分析中多因素非饱和模型的设置·····	(280)
2. 第三模型分析输出的结果及讨论·····	(281)
<b>第九章 鉴别分析</b> ·····	(285)
一、鉴别分析的假设条件和基本模型·····	(286)
1. 鉴别分析的假设条件·····	(286)
2. 鉴别分析的基本模型·····	(288)
二、关于例题·····	(290)
三、鉴别分析模型的各项参数指标及统计检验·····	(291)
1. 非标准化鉴别系数·····	(292)
2. 标准化鉴别系数·····	(293)
3. 结构系数·····	(294)
4. 分组的矩心·····	(295)
5. 鉴别力指数·····	(295)
6. 残余鉴别力·····	(297)
7. Fisher 鉴别系数·····	(298)
四、用 SPSS 进行鉴别分析的程序·····	(299)
1. SPSS 鉴别分析数据格式·····	(299)
2. 利用 SPSS 菜单定义鉴别模型及各项参数·····	(299)
3. SPSS 鉴别分析结果的输出格式·····	(301)
<b>第十章 典型相关分析</b> ·····	(307)
一、典型相关分析思路的简介·····	(308)
二、典型相关模型的基本假设和数据要求·····	(311)
三、应用 SPSS 软件进行典型相关分析·····	(313)
1. 使用 SPSS 附带的典型相关分析命令程序进行分析·····	(313)
2. 用 SPSS 中 MANOVA 菜单进行典型相关分析·····	(314)
3. 关于两种操作方法可能取得某些统计指标的不同结果的 讨论·····	(316)
四、典型相关分析的统计指标·····	(316)
1. 典型相关系数·····	(317)
2. 典型相关系数的平方·····	(317)
3. 特征值及其他有关指标·····	(319)
4. 检验典型相关系数·····	(319)
5. 典型系数·····	(322)



6. 典型负载系数 .....	(323)
7. 交叉负载系数 .....	(324)
8. 变式对本组观测变量总方差的代表比例 .....	(325)
9. 冗余指数 .....	(326)
五、关于典型冗余分析的讨论 .....	(328)
六、例题分析 .....	(332)
<b>第十一章 结构方程模型</b> .....	(339)
一、应用结构方程模型分析的五个主要步骤 .....	(340)
二、模型的设定 .....	(341)
三、模型的识别 .....	(345)
四、模型估计 .....	(348)
五、模型评价 .....	(349)
六、模型的修正 .....	(355)
七、LISREL 模型使用的示范 .....	(355)
附录 例题的小样本及有关统计软件应用的说明 .....	(372)
1. LISREL 第 8 版软件应用操作的简介 .....	(373)
2. LISREL 命令程序及有关说明 .....	(375)
3. 在 SPSS 中进行 prelis 和 lisrel 统计分析 .....	(379)
<b>第十二章 事件史分析</b> .....	(385)
一、什么是事件史分析 .....	(385)
1. 传统 OLS 回归模型的局限性 .....	(386)
2. 事件史分析概述 .....	(387)
3. 删截的类型 .....	(389)
4. 一些统计关系 .....	(390)
5. 估计生存函数的 Kaplan-Meyer 方法 .....	(392)
二、事件历史分析模型 .....	(395)
1. 离散时间 Logit 模型 .....	(395)
2. Cox 比例风险模型 .....	(396)
3. 其他一些连续时间模型 .....	(397)
4. 关于模型的选择 .....	(399)
三、事件史分析的四个例子 .....	(400)
1. 关于职位晋升的假设例子——离散时间风险模型 .....	(400)
2. 研究初育间隔的假设例子——Cox 比例风险模型 .....	(403)