

普通高等教育“十三五”规划教材



21世纪社会学规划教材

社会学系列

SPSS 数据分析教程

Data Analysis
Using SPSS 24

刘爱玉 田志鹏 著

清华大学出版社
TSINGHUA UNIVERSITY PRESS



博
雅

SPSS 数据分析教程

Data Analysis
Using SPSS 24



刘爱玉 田志鹏 著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

SPSS 数据分析教程/刘爱玉,田志鹏著. —北京:北京大学出版社,2017.10

(21 世纪社会学规划教材·社会学系列)

ISBN 978-7-301-28853-5

I. ①S… II. ①刘… ②田… III. ①统计分析—软件包—高等学校—教材

IV. ①C819

中国版本图书馆 CIP 数据核字(2017)第 247670 号

书 名 SPSS 数据分析教程

SPSS SHUJU FENXI JIAOCHENG

著作责任者 刘爱玉 田志鹏 著

责任编辑 陈相宜

标准书号 ISBN 978-7-301-28853-5

出版发行 北京大学出版社

地 址 北京市海淀区成府路 205 号 100871

网 址 <http://www.pup.cn>

新浪微博 @北京大学出版社 @未名社科-北大图书

微信公众号 ss_book

电子信箱 ss@pup.pku.edu.cn

电 话 邮购部 62752015 发行部 62750672 编辑部 62753121

印 刷 者 北京鑫海金澳胶印有限公司

经 销 者 新华书店

730 毫米×980 毫米 16 开本 22 印张 357 千字

2017 年 10 月第 1 版 2017 年 10 月第 1 次印刷

定 价 56.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子信箱:fd@pup.pku.edu.cn

图书如有印装质量问题,请与出版部联系,电话:010-62756370

前 言

SPSS 是英文“Statistical Package for the Social Science”(社会科学统计软件包)的缩写。20 世纪 60 年代末,美国斯坦福大学的三位研究生研制开发了最早的统计分析软件 SPSS,同时成立了 SPSS 公司,并于 1975 年在芝加哥组建了 SPSS 公司总部。1984 年 SPSS 公司推出了世界上第一个统计分析软件微机版本 SPSS/PC+。SPSS 最初的几个版本都是在 DOS 系统下运行的,虽然功能比较强,但在用户界面、输入、输出环境等方面并不十分理想。80 年代末,微软公司推出 Windows 操作系统后,SPSS 迅速推出相应的 Windows 版本。由于 MS Windows 的普及,SPSS 公司于 1992 年、1993 年、1994 年相继推出了可在 Windows 系统下运行的 5.0、6.0、6.1 版本。1995 年微软发布了 32 位操作系统 Windows 95,此后微软又多次发布新的操作系统,包括 Windows XP、Windows 7 以及最新的 Windows 10 等多个 32 位或 64 位操作系统。与之相应,1997 年以来,SPSS 公司也在不断升级优化其统计软件,推出了 8.0 至 18.0 等版本。2009 年 IBM 公司收购了 SPSS 公司,并将其更名为 IBM SPSS Statistics(以下仍简称 SPSS),在对用户界面和一些统计模块进行了较大的升级和调整,推出了 IBM SPSS Statistics 19.0,当前最新版本为 IBM SPSS Statistics 24.0(2016 年发布)。本书的撰写即是在 Windows 10 专业版 64 位操作系统下,使用最新的 24.0 版本软件进行操作演示的,而使用其他操作系统(如 Mac OS)或使用 SPSS 旧版软件的用户也可以参考本教材学习。

随着 SPSS 版本的不断升级,其统计分析的功能也不断更新和完善,目前 SPSS 的主要功能有样本数据的描述和预处理、描述性统计、均值比较、一般线

性模型、相关分析、回归分析、对数线性模型、聚类分析、数据简化、假设检验(包括参数检验、非参数检验及其他检验)、生存分析、时间序列分析、多重响应等几大类,每类中又包括若干统计过程,比如回归分析中又分线性回归分析、曲线估计、Logistic 回归、Probit 回归、加权估计、两阶段最小二乘法、非线性回归等多个统计过程,而且每个过程中又允许用户选择不同的方法及参数。SPSS 也有专门的绘图系统,可以根据数据类型和研究需要绘制各种统计图形。

SPSS 和 SAS、BMDP 并称为国际上最有影响的三大统计软件。目前 SPSS 已被广泛应用于社会学、经济学、政治学、心理学、管理学、市场营销、医学统计等众多领域,因其全面的统计分析功能,强大的建模、评估和自动化功能,以及前沿的大数据分析功能,得到了各界的认可和好评。中英文期刊发表的很多研究论文使用 SPSS 软件作为其统计分析的工具,而全球最大的信息技术和业务解决方案公司 IBM 全资收购 SPSS 公司,也足见其在商务统计中的重要作用。

本书着重介绍如何使用 SPSS 对量化数据进行描述和分析,涉及的数据类型包括研究者或机构就某一具体问题以问卷采集的方式获得的原始数据,也包括基于已有研究或成果的汇总数据,以及由学术机构实施的大型综合社会调查数据(如“中国家庭追踪调查”数据)。

本书基本内容如下:

内容一:SPSS 与社会科学研究。本部分内容对应书中第一章,在概要性地介绍了量化社会科学研究的基本步骤后,指明了 SPSS 在研究环节中起到的重要作用。

内容二:数据分析前的准备。本部分内容对应书中第二章至第五章,包括 SPSS 基本知识、数据录入、数据文件的编辑与管理、数据变换等内容。在相应章节中,将具体演示如何打开/录入数据文件,如何运行统计程序、检查及清理数据,以及如何完成进一步分析所需的数据转换。

内容三:数据描述与展示。本部分包括书中第六章“频次表与描述统计分析”、第十四章“SPSS 统计图形”。统计分析的第一步是对数据进行描述和展示。如多少人对工作感到满意?他们的性别、年龄构成如何?基于数据描述性统计分析结果,可以进一步对数据进行核查,如果在统计结果中发现有人年龄大于 150 岁,或者有人月收入超过几百万等明显的错误,便可以进行相应的

纠正或删除。在呈现数据分析结果时,一张准确的统计图可能胜过大段的文字叙述,而且可以更为直观地呈现结果,快速传递信息,本书最后一章将专门介绍各类统计图形的制作方法。

内容四:推论统计方法。当前,几乎所有的量化社会科学研究所使用的数据均为有一定代表性的抽样调查数据,因此,在研究分析中,更为重要的是检验基于样本数据分析得到的结论能否推论到其所代表的总体。例如,通过药物临床试验证实某新型药物的效果,我们不仅希望知道其对于参与者(样本)的有效性,更重要的是认证这一新药是否对所有同类对象均显著有效,即能否推论到总体。对于此类问题,我们可以借助 SPSS 对数据进行统计检验。第九章“子总体均值比较与检验”将主要介绍假设检验的操作方法与结果分析。在相关分析的部分章节中,也会结合案例展示统计检验。

内容五:相关分析。科学研究的根本目的就在于探察现象背后客观稳定的规律,这一规律可以表述为一组共变关系,也可以更为精确地表述为因果关系。比如在市场营销中,我们会关心广告投放与产品销售额之间的关系,以便在未来决策时更接近“最佳”支出费用。又如在政策研究中,我们会关心某地政治、经济、社会等诸多因素对当地人口预期寿命的影响,进而有针对性地改进当地福利水平。诸如此类问题,都需要在对现象测量的基础上,通过操作化手段生成研究所需变量,并将问题表述为变量关系的假设,在此基础上进一步借助统计软件探索现象背后可能的规律。主要根据由低到高的变量测量水平,本书第七章和第八章介绍交互表与关联分析以及引进其他变量后的交互分析,第十章至第十三章依次为方差分析、相关分析、线性回归分析以及 Logistic 回归分析。

学习数据分析的最好方法是进行实际操作,在本书中,我们将提供一些真实的数据,并使用这些数据一步步地展示操作方法,你会得到以下问题的答案:男性比女性工作更快乐吗?三大地区的收入差距怎么样?上大学的生活成本在不同性别/家庭背景的学生之间有什么样的差异?等等。

与其他同类教材或 SPSS 的操作类书籍相比,本书作者力图以简明、扼要、直观的方式,展示 SPSS 分析的基本原理与操作,并在展示过程中将统计原理与 SPSS 数据分析相结合。好的数据分析是理论、经验、技术、好奇心等的综合。SPSS 作为一个强有力的统计分析工具,能够帮助我们更好地通过数据分

析寻求对社会现象的解释,从数据分析中得出更有说服力的结论。

本书的主要内容架构,得益于 2002 年起作者为北京大学社会学系本科生开设的数据分析课程、2007 年上海人民出版社出版的《SPSS 基础教程》以及 2014 年北京大学出版社出版的《SPSS 基础分析教程》。多年的课堂教学以及与学生的互动,使我受益良多,面对那些来自全国的优秀学子,唯有不断改进和提高自己的教育水平,才是对得起他们的唯一选择。面对 SPSS 软件版本的不断升级,以及之前教材中的一些疏漏在教学过程中不断被发现,重新编写一本教材,便是我们的选择。

本书所用数据可在北京大学出版社官网(<http://www.pup.cn>)上查阅。

目 录

第一章 SPSS 与社会科学调查研究	1
一、问题的提出	1
二、提出研究假设	3
三、设计研究方案	4
四、收集资料	14
五、SPSS 与资料的整理分析	15
思考和练习	17
第二章 SPSS 基本知识	18
一、SPSS 的启动和退出	18
二、SPSS 主要窗口介绍	22
三、SPSS 的求助系统	28
思考和练习	31
第三章 数据录入	32
一、问卷设计及编码	32
二、使用 SPSS 窗口录入数据	35
三、用 SPSS 命令程序录入数据	42
四、在 Excel 中录入数据	47
五、利用 EpiData 程序录入数据	48
思考和练习	58

第四章 SPSS 数据文件的编辑与管理	59
一、增加新的变量或个案	59
二、删除变量或个案	60
三、数据查找和替换	60
四、数据的排序	62
五、数据的行列变换	63
六、数据文件的合并	65
七、数据文件的分类汇总	68
八、选择个案	73
九、对变量进行加权	75
思考和练习	78
第五章 数据变换	79
一、用 Recode 对变量重新编码	80
二、用 Compute 命令对变量计算赋值	86
三、用 Count 命令统计同类情况的频次	95
四、生成新变量指示变量取值序次	99
五、自动对变量重新编码	102
六、对缺失值自动赋值	102
思考和练习	105
第六章 频次表与描述统计分析	106
一、生成变量频次统计表	106
二、在频次分析中完成描述统计	109
三、条形图、饼图和直方图	119
四、定距变量的描述统计与生成标准分	122
五、探索性描述分析	125
思考和练习	131
第七章 交互表与关联分析 (Crosstabs 的应用)	133
一、交互表的基本形式	133

二、输出交互表	134
三、期望值、百分比和残差	138
四、定类-定类变量的交互分析	146
五、两个定类变量之间相关的强度	148
六、两个定序变量之间相关的强度	153
七、Kappa 选项的含义与操作	155
八、Risk 选项的含义与操作	157
思考和练习	159
第八章 引入其他变量后的交互分析	162
一、因果分析：排除虚假相关	163
二、阐明分析：发现中介变量	170
三、条件分析：排除压抑变量	175
思考和练习	179
第九章 子总体均值比较与检验 (Means 和 T-Test 的应用)	181
一、子总体均值的比较 (Means 过程)	182
二、单样本 T 检验 (One-Sample T Test)	185
三、独立样本 T 检验 (Independent-Samples T Test)	190
四、配对样本 T 检验 (Paired-Samples T Test)	196
思考和练习	199
第十章 方差分析 (One-Way ANOVA)	200
一、方差分析的原理	200
二、用 SPSS 进行单因素方差分析	202
三、双因素方差分析	209
思考和练习	221
第十一章 相关分析 (Correlate 过程)	223
一、双变量相关分析 (Bivariate Correlation)	224
二、偏相关分析 (Partial Correlations)	228
三、距离分析 (Distances)	233

思考和练习	242
第十二章 线性回归分析 (Linear Regression)	243
一、一元线性回归模型及其对变量的要求	243
二、一元线性回归模型实例	245
三、多元线性回归分析	261
思考和练习	266
第十三章 Logistic 回归分析	269
一、Logistic 回归模型	270
二、Logistic 回归模型实例	272
三、在回归模型中引入交互项	282
思考和练习	286
第十四章 SPSS 统计图形	288
一、图表建构器	288
二、图形画板模板选择程序	318
三、比较子组制图程序	321
四、回归变量图	323
思考和练习	327
附录一 Stata 简介	329
一、Stata 软件基本界面	329
二、Stata 命令语句	331
三、Stata 统计分析示例	332
附录二 本书各章例题对应 Stata 命令示范	335

第一章

SPSS 与社会科学调查研究

社会研究的目的是对社会现象进行描述、解释和预测。有多种社会研究方法可以实现上述目标,而当代社会科学研究主要采用科学方法,该方法建立在逻辑和观察两大基础之上。社会科学调查研究的过程,大致可分为六个步骤:提出问题、确定课题或假设、设计研究方案、收集资料、整理和分析资料、得出结论或解释(见图 1-1)。

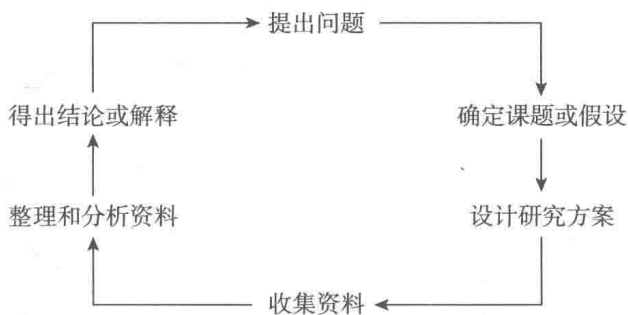


图 1-1 社会科学调查研究的一般步骤

一、问题的提出

所谓“问题”,是指在每一项科学研究开始时,研究者关心的或令其困惑不解的一个或一系列现象。而符合规范的研究的问题必须是:(1)可以由科学研

究来解答,(2)具有一定的研究价值。^① 以下将以社会学研究为例,说明提出问题的基本思考过程。

社会学家默顿(R. K. Merton)指出,提出问题不是仅仅对确定的事实或事物常规地加上一个“为什么”,社会学研究问题的提出是一项需要训练有素的想象力的困难工作。对于社会学研究而言,并不存在单一的提出问题的套路。研究问题的提出可以是对社会事实证据的怀疑,对社会生活领域中的不同社会学变量之间关系的分析,对一个曾经很有启发性但看来与所期望解释的事实有差异的社会学观点的澄清,等等。默顿指出,一个问题演化形成的过程主要由以下三个部分构成:第一,发问,陈述想解决的问题;第二,基本依据或理由,说明为什么要解答这一问题;第三,阐明疑问,即对提出的疑问找出与理由相符的可能答案。^② 以下将依次说明这三个部分的基本内容。

1. 发问

如果一种方法是直接针对未知的社会事实,它不能构成一种问题的形式,但却是问题形成的重要一步。默顿指出,这类问题的提出,典型地来源于试图“解释”那些研究者认为满意地提出的作为独特形式的社会事实。一些研究或解释可能只是单纯的事实发现,虚假的事实往往会导致虚假的问题,因此我们可以对此提出疑问。事实发现是形成问题的第一个阶段。

另一种提出疑问的方式是直接关注社会学变量之间的关系。例如,社会如何调和社会成员与社会群体之间的冲突,从而维持社会的凝聚力和权威的合法性?这类问题比确定社会事实所要求的更多,因为它不仅要问为什么一类观察到的事物存在,而且还要提出问题的线索。这类疑问由较概括的社会学变量所构成。

有些疑问的提出是关于观察到的经验概括性问题,另外一些则涉及出于或源自于某些基本观念的社会一致性问题,一些是关于社会组织形式起源的问题,另外一些是关于其结果的问题。总之,问题的提出并不是对所观察到的社会现象加上一个“为什么”。

^① 袁方主编:《社会研究方法教程》,北京:北京大学出版社1997年版,第117页。

^② 参见罗伯特·K. 默顿:《社会研究与社会政策》,林聚任等译,北京:生活·读书·新知三联书店2001年版,第19—47页。

2. 疑问的基本依据

这是指提问的原因。基本依据所表达的是解答一个问题将对其他知识或实践有何意义。通过这种方式,可以区分有科学价值的提问与没有科学价值的提问。判定不同问题、观点和发现的相对重要性是任何一项科学活动的内在要求。科学或者说全部学术性活动始终将知识作为其追求的核心目标。作为科学的社会学研究也不例外,提出一个社会学问题,既要注重其对现有社会学乃至整个社会科学知识推进的重要性,也要思考对这一问题的求解将会带来的社会价值和现实意义,即通常所强调的研究的理论意义与实际意义。

3. 阐明疑问

如果说“发问”需要对纷繁复杂的现象或事实有好奇心,“依据”强调要在知识传统中形成问题,那么,“阐明”则是要清晰明确地将问题表达出来。阐明问题是科学提问的最后一步,也是决定提问是否科学的关键一步。我们在生活和研究中会形成很多问题,当然,还有很多视而不见的问题,但这些问题或困惑只有在按照学术规范表达出来后,学术共同体的其他成员才能对其加以判定或给出研究建议,进而推动知识进步。此外,所谓“阐明”,除了“阐”,还要“明”,即表达的问题一定要明确,切不可担心犯错而提出模糊的问题。按照著名科学哲学家波普尔(K. Popper)的理解,科学进步就是“证伪”的过程,在否证中推进知识的积累。而在科学研究中,只有在清晰阐明研究问题后,才能进一步将研究问题表达为研究假设。

二、提出研究假设

假设是根据对问题的了解,假定现象与现象之间的关系,即一种现象的变化与另一种现象的变化是什么样的关系。社会学所研究的社会现象,一般被称为变量(Variables),假设就是假定某一种现象(即A变量)的变化与另一种现象(即B变量)的变化具有某种关系。社会学家古德(E. Goode)认为,假设必须满足以下几个条件:

- (1) 以明确的概念为基础;
- (2) 具有经验的统一性(能够被经验检验);
- (3) 对假设的适用范围要有所界定;

(4) 与有效的观测技术相联系;

(5) 与一般理论相关联。^①

研究假设可以通过查阅理论文献而得出,也可以通过实地考察或访谈而获得,并非任何研究都能够在研究的准备阶段提出明确的假设。而且,进行研究是否需要研究假设,在社会学界也有不同的认识。不过在量化研究中,提出假设是研究过程中必不可少的一个环节。

社会现象之间的关系可以用两种模型来表述:(1)差异式——某个变量的不同取值在另一变量或指标上是否存在显著差异,这是进一步探索变量间关系的前提,例如不同收入水平的员工在工作满意度上是否存在统计显著的差异,如果存在,或者不存在显著的差异(或变异),可以再进行进一步探讨何以如此;(2)函数式——在确定多个变量存在明显变异后,进一步假定多个变量变异共变的方向和程度,仍以收入与满意度之间的关系为例,在共变方向上,可以假定收入水平越高,员工工作满意度越高,而在共变程度上,可以进一步假设在较低收入水平上,收入提高会带来更多工作满意度的提高。

使用实证数据对研究假设进行科学的证实或证伪需要以完备的统计学知识为基础,统计软件的作用在于让统计检验更为简单易行,特别是使大样本数据统计检验成为可能。关于假设检验方法,读者可进一步参考相关统计学教材。

三、设计研究方案

研究方案设计大致包括以下九个方面的内容:阐明研究课题与研究目的、确定研究类型与研究方法、确定资料收集方法、确定研究层次和分析单位、确立概念和变量测量方法、制定问卷/观察表格与访问提纲、确定抽样方案、确定调查的场所与时间计划、研究经费与物质手段的计划与安排。下面我们将对其中一些重要的方面展开分析。

(一) 研究类型选择

社会研究有不同的类型,研究方案应当根据不同的类型进行设计。根据研

^① 参见袁方主编:《社会研究方法教程》,北京:北京大学出版社 1997 年版,第 124 页。

究的性质可以划分为理论性研究和应用性研究;按照研究目的可以划分为描述性研究、探索性研究与解释性研究;根据研究对象的时间跨度可划分为横剖研究和纵贯研究;等等。

1. 理论性研究和应用性研究

理论性研究是通过对社会现实问题的调查来发展和丰富社会理论,并提供有关社会发展一般规律的知识,其主要任务是解答社会科学领域和各个实践领域中的理论问题,探求社会现象的规律性。为了发展社会科学而去收集和概括经验材料的经验性理论研究,是十分重要的一种理论性研究类型。

应用性研究主要侧重于解答各个实际工作部门、各个社会领域中的具体问题,典型的如政策评估研究、社会指标研究、社会问题的调查与诊断等。

2. 描述性研究、探索性研究和解释性研究

描述性研究旨在了解研究对象的状况和特点,回答社会现象“是什么”和“怎么样”的问题。描述性研究可以通过提出研究假设来引导研究的展开,但更多的时候是通过研究问题来构建研究过程。

探索性研究是对一个陌生或新的现象与问题进行探究,其作用有三个:“(1)满足研究者的好奇心和对事物更加深入了解的欲望;(2)探讨对某议题进行细致研究的可行性;(3)发展后续研究中需要使用的方法。”^①

解释性研究可以回答“为什么”的问题,它能说明社会现象发生的原因,预测事物的发展后果,探讨社会现象之间的因果联系。

3. 横剖研究与纵贯研究

横剖研究是指在某一时点对调查对象进行横断面的研究,最典型的横剖研究是人口普查,选定某一时间点(如2010年7月31日零点整)对整个人口进行调查,以了解全国人口的总量和分布状况。不过许多调查很难像人口普查那样做到同时点进行,只能集中在某一时间段内展开,只要这一时间段不是很长,均可视为横剖研究。

纵贯研究是指在较长时期的不同时点收集资料,并对社会现象做纵向研究。纵贯研究关注的是社会现象的变迁,可分为趋势研究、同期群研究与追踪研究。

^① 艾尔·巴比:《社会研究方法》,邱泽奇译,北京:华夏出版社2000年版。

(二) 研究方法选择

社会研究的方法主要有以下四种类型:

1. 实地研究

实地研究有时也称“实地调查”,是社会调查研究的主要方法之一。该调查方式是研究者深入到调查现场,利用观察、访问、座谈等方法收集少数单位各方面的信息,以便对调查对象做深入的解剖分析。在某些实地研究中,调查者本人加入调查对象的团体中生活,扮演其中的一员,对其中所发生的现象和情况进行多方面的长期观察。

2. 统计调查

统计调查是目前国家机关、学术科研机构以及市场调查公司等部门广泛采用的社会经济调查方法,在实施中,通常由调查员使用标准化的问卷、表格,通过直接询问或者由受访者本人填写的方式收集资料。除全国人口普查外,统计调查一般采用随机抽样调查的方式,力求通过科学的抽样方法获取一个较大的代表性样本(样本量一般在 1000 个以上),进而采用统计或其他定量分析方法分析收集到的数据,并基于此估计总体中的状况。这种方法很适合于对集体态度、行为倾向、社会舆论等问题展开研究,如市场调查、民意测验等。

3. 实验法

实验法是在精心设计和高度控制的条件下,通过操纵某些因素来验证变量之间因果关系的方法。实验法以因果关系的假设为起点,通过有意改变某些条件,来看这些条件改变所带来的后果和影响。在设计方面,首先需要将研究对象随机地分为实验组(experimental group)和控制组(control group),以排除其他因素的影响。所谓“实验组”是指在实验中被施以实验刺激的参与者,亦即受到自变量影响的人;所谓“控制组”是指所有可控影响因素保持不变的那些人。如果改变实验组的 A 变量(条件),但不改变控制组的 A 变量(条件),然后比较两组的 B 变量在实验前后的变化,若变化的程度不同,则表示 A 对 B 有影响,若相同,则表示 A 对 B 无影响。在获取实验数据后,可进一步通过统计方法对 A 变量和 B 变量之间的关系进行假设检验和分析。

实验法特别适用于范围有限、界定明确的概念与假设,与描述相比,实验法更适合对社会现象进行解释。在社会科学中,实验法多用于小群体内部的互动