

普通高等教育“十二五”规划教材



21世纪社会学规划教材

社会学系列

SPSS基础分析教程

Introduction to
Basic Analysis
Using SPSS 22

刘爱玉 等编著



北京大学出版社
PEKING UNIVERSITY PRESS

SPSS基础分析教程

Introduction to
Basic Analysis
Using SPSS 22

刘爱玉 等编著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

SPSS 基础分析教程/刘爱玉等编著. —北京:北京大学出版社,2014.9

(21 世纪社会学规划教材·社会学系列)

ISBN 978-7-301-24698-6

I. ①S… II. ①刘… 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(2014)第 198847 号

书 名: SPSS 基础分析教程

著作责任者: 刘爱玉 等编著

责任编辑: 陈相宜

标准书号: ISBN 978-7-301-24698-6/C·1037

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> 新浪官方微博: @北京大学出版社

电子信箱: ss@pup.pku.edu.cn

电 话: 邮购部 62752015 发行部 62750672 编辑部 62753121

出版部 62754962

印 刷 者: 北京鑫海金澳胶印有限公司

经 销 者: 新华书店

787mm×1092mm 16 开本 25 印张 436 千字

2014 年 9 月第 1 版 2014 年 9 月第 1 次印刷

定 价: 55.00 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

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 总部。1984 年 SPSS 总部首先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,Microsoft 推出 Windows 后,SPSS 迅速向 Windows 移植。90 年代以来,由于 MS Windows 的普及,SPSS 公司于 1992、1993、1994 年相继推出了在 Windows 操作系统下的 5.0、6.0、6.1 版本。随着 Windows 95 与 Windows NT3.5 以上版本的 32 位操作系统的出现,1997 年以来又相继推出了 8.0 至 18.0 的各种版本。2009 年 IBM 公司收购了 SPSS 公司,将其更名为 IBM SPSS,然后推出了 IBM SPSS 19 到 IBM SPSS 22 的各种版本。本书使用的是正是 IBM SPSS 22 版本。

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 的功能主要有: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,每类中又分好几个统计过程,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,可以根据数据绘制各种图形。

SPSS 现已推广到各种操作系统上,它和 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、新闻传播学、市场营销、卫生统计等众多领域。世界上许多有影响的报刊杂志纷纷就 SPSS 的自动统计绘图、数据的深入分析、使用方便、功能齐全等方面给予了高度的评价与称赞。

SPSS 已越来越成为一种从事社会科学研究或调查,甚至是从事自然科学研究的必备工具。

本书主要介绍的是如何运用 SPSS 对通过问卷或其他手段收集的定量资料进行描述和分析。主要内容安排如下:

(1) SPSS 与社会科学研究。第一章主要介绍量化社会科学研究的一般步骤及 SPSS 在量化分析中的作用。

(2) 数据分析前的准备。主要内容为第二章 SPSS 基本知识、第三章数据录入、第四章 SPSS 数据文件的编辑与管理、第七章数据变换。这些章节将通过案例介绍如何打开文件、如何运行统计程序、检查结果、录入数据以及为进一步分析进行数据转换,另外还简要介绍了如何运用这些简单的操作来制作排行榜。

(3) 数据描述。主要内容为第五章用 Frequencies 做数据汇总、第六章描述性统计、第十五章 SPSS 统计图形。统计分析的第一步是对数据进行展示,如多少人对工作感到满意? 他们的性别、年龄如何? 可以对数据进行核查,纠正如出现年龄大于 150 岁的人,每月收入超过几百万等一些看来明显的错误。数据分析的一个重要功能是,用图示的方式来进行展示,在本书中,我们将用专门的一章来介绍一些主要的图形如何制作。

(4) 假设检验。很多情况下,我们不仅需要描述所调查群体描述性的一般性特征,还需要知道这些数据能否推论到更大的群体,对于药物使用的效果,我们不仅希望知道对于被研究对象(样本)的适用性,还需要知道对所有对象是否适用,能否推论到其他入。大学毕业的人会比其他人更快乐吗? 我们可以对数据进行检验。第十章子总体均值比较与检验、第十四章非参数检验将主要介绍假设检验的操作与过程。

(5) 相关分析。有时我们的研究关心以下一些问题,诸如广告费用与销售额之间是一种什么样的关系? 可以由出生率预知预期寿命吗? 什么因素是对

收入高低的较好预测?本书的第八章交互表与关联分析、第九章引进其他变量后的交互分析、第十一章方差分析、第十二章相关分析、第十三章线性回归分析,将通过实例来介绍不同变量之间的关系。

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,譬如,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。好的数据分析是常识、技术、好奇心等的综合,关键在于如何提出好的问题及对其进行最好的解释。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 年起为北京大学社会学系本科生开设的数据分析课程教育以及 2007 年上海世纪出版集团出版的《SPSS 基础教程》一书。十多年的课堂教育以及与学生的互动,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编写教材的一些疏漏在教学过程中不断被发现,重新编写一本新的教材,便是我们的选择。

本书各章的作者为:

刘爱玉:前言、第一章、第三章、第七章、第八章、第九章、第十一章、第十三章

刘继伟:第六章

庄家炽:第十章

付伟:第十二章

田志鹏:第十五章、附录(Stata 简介及命令)

刘爱玉、付伟:第二章

刘爱玉、刘继伟:第四章

刘爱玉、庄家炽:第五章

刘爱玉、田志鹏:第十四章

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	(13)
五、SPSS 和资料的整理与分析	(14)
思考和练习	(16)
第二章 SPSS 基本知识	(17)
一、SPSS 的安装	(17)
二、SPSS 的启动和退出	(23)
三、SPSS 主要窗口介绍	(25)
四、SPSS 的求助系统	(30)
思考和练习	(35)
第三章 数据录入	(36)
一、问卷设计及编码	(36)
二、数据录入方法	(39)
思考和练习	(64)
第四章 SPSS 数据文件的编辑与管理	(66)
一、增加新的变量(Insert Variable)	(66)

二、增加新的个案 (Insert Cases)	(67)
三、删除变量或个案	(67)
四、查找个案 (Go To Case)	(68)
五、数据的排序 (Sort Cases)	(68)
六、数据的行列对换 (Transpose)	(70)
七、数据文件的合并 (Merge Files)	(72)
八、数据文件的分类汇总 (Aggregate)	(75)
九、数据的选择 (Select Cases)	(81)
十、对变量进行加权 (Weight Cases)	(85)
十一、Data 中的其他菜单命令	(87)
思考和练习	(88)
第五章 用 Frequencies 做数据汇总	(90)
一、如何得到一张频次表	(90)
二、频次统计及统计量的计算	(93)
三、条形图、饼图和直方图	(102)
思考和练习	(105)
第六章 描述性统计——Descriptives 与 Explore 的应用	(108)
一、Descriptives 过程的应用	(108)
二、Explore 过程的应用	(112)
思考和练习	(123)
第七章 数据变换	(125)
一、用 Recode 对数据重新编码	(125)
二、用 Compute 命令创建新变量	(132)
三、用 Count 命令计算有相同情况的次数	(143)
四、变量值秩次确定	(148)
五、自动对变量重新赋值	(151)
六、缺失值的替代	(152)
思考和练习	(153)

第八章 交互表与关联分析 (Crosstabs 的应用)	(155)
一、交互表的一般形式及其特点	(155)
二、如何获得交互表	(156)
三、期望值、百分比和残差	(161)
四、定类-定类型变量的交互分析	(168)
五、两个定类变量之间相关的强度	(171)
六、两个定序变量之间相关的强度	(176)
七、关于 Kappa 复选框的含义与操作	(178)
八、关于 Risk 框的含义及其操作	(180)
思考和练习	(182)
第九章 引进其他变量后的交互分析	(184)
一、因果分析	(184)
二、阐明分析	(191)
三、条件分析	(197)
思考和练习	(201)
第十章 子总体均值比较与检验 (Means 和 T-Test 的应用)	(202)
一、子总体均值的比较 (Means 过程)	(203)
二、单样本 T 检验 (One-Sample T Test)	(206)
三、独立样本 T 检验 (Independent-Samples T Test)	(211)
四、配对样本 T 检验 (Paired-Samples T Test)	(217)
思考和练习	(220)
第十一章 方差分析	(221)
一、方差分析的原理	(221)
二、用 SPSS 进行单因素方差分析	(223)
三、双因素方差分析	(231)
思考和练习	(242)
第十二章 相关分析 (Correlate 过程分析)	(244)
一、双变量相关分析 (Bivariate Correlations)	(245)
二、偏相关分析 (Partial Correlations)	(250)

三、距离分析(Distances)	(254)
思考和练习	(264)
第十三章 线性回归分析	(265)
一、一元线性回归模型及其对变量的要求	(265)
二、在对话框中做一元线性回归模型	(267)
三、多元线性回归分析	(282)
思考和练习	(288)
第十四章 非参数检验	(291)
一、Chi-square Test: 卡方检验	(293)
二、Binomial Test: 二项分布检验	(299)
三、Runs Test: 游程检验	(301)
四、One-Sample K-S Test: 单样本 K-S 检验	(304)
五、Two-Independent Samples Tests: 双样本 K-S 检验	(307)
六、K Independent Samples Tests: 多个独立样本的秩和检验	(315)
七、Two Related Samples Tests: 两个相关样本的检验	(318)
八、K Related Samples Tests: 多个样本的秩和检验	(324)
思考和练习	(327)
第十五章 SPSS 统计图形	(331)
一、图形建构器	(332)
二、图形画板模板选择程序	(366)
三、比较子组制图程序	(370)
四、回归变量图	(374)
思考和练习	(378)
附录一 Stata 软件简介	(379)
一、Stata 软件基本界面	(379)
二、Stata 命令语句	(381)
三、Stata 统计分析示例	(382)
附录二 本书各章操作及例题对应 Stata 命令示范	(384)

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。为了达到上述目的可以有多种方法,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支柱之上。社会科学调查研究的历程,大致可分为六个步骤:问题的提出、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。整个过程可以用一个简洁的图形来表达(见图 1-1)。

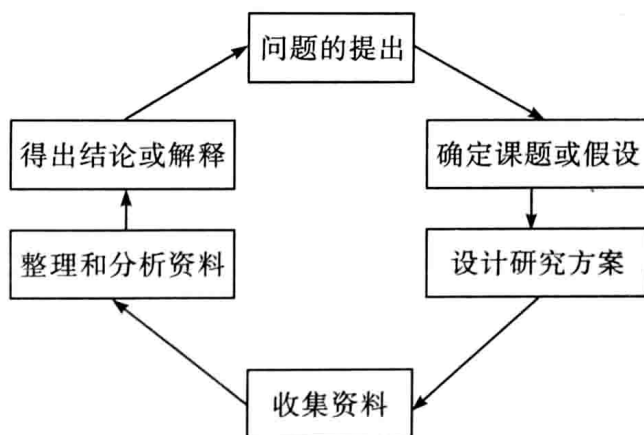


图 1-1 社会科学调查研究的一般步骤

一、问题的提出

所谓问题,是指在每一项科学研究开始时,必定存在的令研究者不解的一个

或一系列事实。

所要研究的问题必须是:(1)可以由科学研究来解答,(2)具有一定的研究价值。^①

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

(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。

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能观察到事物的存在,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

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

^② 罗伯特·K.默顿著:《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。

象加上一个“为什么”。

(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。通过这个方式,可以区分有科学价值的提问与没有科学价值的提问。确定问题、观点和发现的相对重要性是每一门科学的基本活动的内在要求。作为学术或整个科学的基础,它只是把知识作为自身追求的目标。一个社会学问题的提出,既可能着眼于其结果的社会价值,也要注意它对一定社会学观点的重要性。阐述这些也就是我们通常强调的研究的理论意义与实际意义。

社会学研究的理论意义与实际意义可能十分一致,也可能有所偏重。

(3) 阐明疑问。

发问必须再进一步指明所观察到的证据,这离不开对经验资料的考察,由此问题才被加以研究,并取得较好的成效。

二、提出研究假设

确定研究的问题以后,需要根据自己对这个问题的想法提出假设。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。社会学研究的社会现象,一般称为变量(variables),假设就是假定某一种现象的变化(即A变量)与另一种现象的变化(即B变量)具有某种关系。社会学家古德认为,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,至于进行研究是否需要研究假设,在社会学中有不同的认识。不过在量化研究中,提出假设是成功

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

进行研究的非常重要的一个环节。

举例说:我们确定的课题是关于员工工作满意度的问题,那么,通过探索性研究,发现工作满意度因人而异。具体来讲,收入和福利待遇高的员工,工作满意度高;或者对工作本身感兴趣者,工作满意度高。我们不仅关心讨论孤立的社会现象,如员工工作满意度的一般状况、收入状况、文化程度等,同时还要讨论现象之间的联系,如收入水平与工作满意度之间的联系,性别与工作满意度之间的联系等。

社会现象之间的联系可以用两种模型来表述:

第一种:差异式——只是说明现象之间存在关系,如收入水平不同的员工,工作满意度也不同。

第二种:函数式——说明现象之间不仅存在联系,而且还能够说明变化的方向和趋势。如收入水平越高的员工,工作满意度越高。

上述关于现象之间联系的表述,只是初步设想,或者称为假设,最终能否确立,需要通过实践的检验。

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。下面我们将对其中的一些方面展开分析。

(一) 研究类型选择

社会研究有不同的类型,研究方案的设计应当根据不同的类型进行设计。如根据研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,其主要任务是为了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。其中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,是很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

解释性研究可以解答“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7月31日零点等)对整个人口进行调查,以了解全国人口的总量和分布状况等。不过许多研究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,只要时间不是很长都可视为横剖研究。

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。

(二) 研究方法选择

社会研究的研究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

^① 艾尔·巴比:《社会研究方法》,北京:华夏出版社2000年版。

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的各方面信息,以便对调查对象作深入的解剖分析。

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,对里面所发生的现象和情况进行多方面的长期观察。

2. 统计调查

统计调查是目前采用非常广泛的社会经济调查方法,调查者常常使用标准化的问卷、表格,通过直接询问或者由调查对象本人填写的方式收集资料,并采用统计或其他定量分析方法分析资料。这种方法很适于对集体态度、行为倾向、社会舆论等的研究,如市场调查、民意测验等。

3. 实验法

实验法是在精心设计并高度控制的条件下,通过操纵某些因素来探讨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,来看这些条件改变的后果与影响。在操作上,首先需要将研究对象随机地区分为实验组(experimental group)和控制组(control group),以排除其他因素的影响。所谓实验组是指施以实验刺激,即受到自变量影响的那些人,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,表示 A 对 B 有影响,若相同,则表示 A 对 B 无影响。可以通过统计方法对收集到的 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研究。不过在社会研究中,由于环境难于控制,加上一些伦理限制,采用实验室的实验很困难,因而一般采用实地实验的方式,即在实际场所进行实验。与实验室实验相比,实地实验对环境和变量的控制均不太严格。

4. 文献研究

文献研究主要是利用第二手资料进行研究,是一种间接研究法。根据研究的具体方法和所使用文献的不同,文献研究又可以分为内容分析、二次分析与现存统计资料分析三类。

内容分析是一种对文献内容进行客观、系统和定量描述的分析技术。二次分析是利用其他研究者为了其他目的所收集到的原始资料所进行的新的分析,目的是从中获得新的发现。现存统计资料主要指来源于官方的各种统计资料,例如《中国统计年鉴》《中国人口统计年鉴》等统计资料,现存统计资料分析就是针对这些资料进行的分析。

(三) 资料收集方法选择

社会研究中资料收集的方法很多,主要有问卷法、访谈法、观察法、量表测量法、统计报表法、文献法等类型。

选择什么样的方式收集资料,这与研究方法有非常密切的关系,研究法或研究方式规定了社会研究的基本途径和具体的操作方式,每一种研究法由相应的资料收集方法和资料分析方法,以及特定的方法论视角区分开来。研究法与资料收集、分析方法的关系可以概括如表 1-1:

表 1-1 各研究法的资料收集与分析方法^①

研究法	资料收集方法	资料分析方法	研究性质
调查法	统计报表 自填式问卷	统计分析 数量分析	定量
实验法	自填式问卷 结构式访谈 结构式观察 量表测量	统计分析 数理分析	定量
实地研究	无结构观察 自由式访谈	定性分析	定性
文献法	官方统计资料 他人原始数据 文字声像文献 历史文献	统计分析 定性分析	定量/定性

^① 风笑天:《社会学研究方法》,北京:中国人民大学出版社 2001 年版。根据其第 8 页表 1-1 改制。