



高等学校教材

SPSS

在教育统计中的应用

主编 杨晓明



高等教育出版社

高等学校教材

SPSS 在教育统计中的应用

主 编 杨晓明
副主编 丛 彦 王亚利
孙美丽 夏建刚

高等教育出版社

内容提要

SPSS 是软件英文名称的首字母缩写, 全称为 Statistical Package for the Social Sciences, 即“社会科学统计软件包”, 是世界上应用最广泛的专业统计软件之一。目前国内针对有关 SPSS 在教育领域应用的专门论著还较少, 编写本书的主要目的是弥补这部分内容, 为广大教育科学研究工作者进行数据分析提供一个参考工具。

本书的主要思路是: 将教育科研中经常遇到的数据统计问题, 归纳分成几类典型的教育问题, 并说明解决这些问题的数理统计方法, 最后以典型的教育科研的实际问题作为案例, 说明如何应用 SPSS 来解决这些问题。本书希望帮助那些对数学知识了解不太深的非数学工作者, 特别是教育科学工作者, 学习运用现代统计技术。使用者仅需要关心某个问题应该采用何种统计方法, 并掌握对计算结果的解释, 而不需要了解其具体运算过程, 更加方便准确地分析教育相关数据, 促进其教育研究工作的开展。

本书所配光盘为教材中引用的例题, 学习者可直接在 SPSS 软件环境下操作练习。

本书可作为教育技术及相关专业的本科生及专科生教材, 也可作为广大教育科研工作者、教育统计人员和其他科研工作者的参考书。

图书在版编目(CIP)数据

SPSS 在教育统计中的应用/杨晓明主编. —北京: 高等教育出版社, 2004.4
ISBN 7-04-014643-6

I. S... II. 杨... III. 教育统计 - 统计分析 - 软件包, SPSS - 高等学校 - 教材 IV. G40-051

中国版本图书馆 CIP 数据核字(2004)第 027059 号

策划编辑 耿 芳 责任编辑 耿 芳 市场策划 韩 飞
封面设计 于文燕 责任印制 孔 源

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 - 82028899

购书热线 010 - 64054588
免费咨询 800 - 810 - 0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>

经 销 新华书店北京发行所
印 刷 北京铭成印刷有限公司

开 本 787 × 1092 1/16
印 张 21.25
字 数 450 000

版 次 2004 年 5 月第 1 版
印 次 2004 年 5 月第 1 次印刷
定 价 26.00 元(含光盘)

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

前 言

SPSS 全称为 Statistical Package for the Social Sciences, 即“社会科学统计软件包”。目前, 世界上最著名的数据分析软件有 SAS 和 SPSS。SAS 由于是为专业统计分析人员设计的, 具有功能强大、灵活多样的特点, 为专业人士所喜爱; 而 SPSS 是为广大的非专业人士设计, 它操作简便, 好学易懂, 简单实用, 因而很受非专业人士的青睐。1988 年, 此软件开始在高校中推广使用, 从此成为国内教育科研人员最常用的工具。

本书的主要思路是: 将教育科研中经常遇到的数据统计问题, 归纳分成几类典型的教育科研需解决的问题, 并说明解决各类问题的数理统计方法, 最后以典型的教育科研需解决的实际问题作为案例, 说明如何应用 SPSS 来解决这些问题。本书希望帮助那些对数学知识了解不太深的非数学工作者, 特别是教育科学工作者, 学习运用现代统计技术, 使用者仅需要关心某个问题应该采用何种统计方法, 并掌握对计算结果的解释, 而不需要了解其具体运算过程, 更加方便准确地分析教育相关数据, 促进其教育研究工作的开展。

本书章节内容安排为: (1) 本章方法可以解决哪一类问题; (2) 基本概念与方法: 本章中每一种方法解决问题的具体类型和要求条件; (3) 例题与数据文件: 给出具体问题类型的典型实际案例及数据存储格式; (4) 统计分析过程: 说明在 SPSS 上如何进行操作, 解决实际案例中所提问题; (5) 结果和讨论: 对 SPSS 给出的各种结果进行解释、分析, 回答典型实际案例中提出的问题。

全书分为教育科研中的常见问题分析、SPSS 基本操作、常用描述性统计方法、统计报表、T 检验、方差分析、相关分析、回归分析、非参数检验、卡方检验、聚类分析、因子分析与主成分方法、SPSS 统计作图等十三章内容。附有 40 个例题及其数据文件。

本书由杨晓明担任主编, 丛彦、王亚利、孙美丽、夏建刚担任副主编。各章编写分工为: 第 1 章由杨晓明、丛彦、王亚利编写; 第 2、3 章由苗丽、杨晓明编写; 第 4 章由孙美丽、石鸿编写; 第 5、6 章由孙美丽、杨晓明编写; 第 7、8 章由杨晓明、李晓梅编写; 第 9、10、11、12 章由杨晓明编写; 第 13 章由夏建刚、杨晓明编写。另外高等教育研究所高峰、张秀云、洪歌、刘颖、陈锦红等同学参与了本书的资料搜集、文字录入工作。

本书欲向教育科学工作者及其他社会科学工作者介绍一个便利的研究工具, 从写作结构、体例及案例选择上借鉴了多种相关书籍, 力图做到方便、简明、实用。由于水平所限, 不足之处在所难免, 请读者批评指正。作者联系方式: yangxiaoming508@126.com。

杨晓明

于北京科技大学高等教育研究所

2003 年 12 月

目 录

第 1 章 SPSS 与教育统计	(1)	2.3.2 观察量的增加和删除	(18)
1.1 SPSS 概述	(1)	2.3.3 数据的排序	(18)
1.1.1 启动 SPSS	(2)	2.3.4 数据的行列互换	(19)
1.1.2 SPSS 的数据编辑窗口	(3)	2.3.5 选择观察量的子集	(19)
1.1.3 SPSS 结果输出窗口	(4)	2.3.6 数据的分类汇总	(21)
1.1.4 退出 SPSS	(5)	2.3.7 缺失值的替代	(23)
1.2 教育统计中的问题	(5)	2.3.8 数据次序的确定	(24)
1.2.1 数据采集	(5)	2.3.9 数据的剪切、复制和粘贴	(25)
1.2.2 教育统计的常见问题类型	(7)	2.4 处理变量	(27)
1.3 SPSS 如何解决教育统计中的 问题	(8)	2.4.1 增加和删除一个变量	(27)
习题一	(9)	2.4.2 指定加权变量	(28)
第 2 章 数据文件的处理	(10)	2.4.3 对数据进行重新编码	(29)
2.1 定义变量	(10)	2.4.4 根据已存在的变量建立 新变量	(31)
2.1.1 定义数据项的变量名 (Name)	(11)	2.4.5 产生计数变量	(34)
2.1.2 定义变量类型(Type)	(12)	2.4.6 产生分组变量	(37)
2.1.3 变量长度(Width)	(13)	2.4.7 变量的自动赋值	(38)
2.1.4 变量小数位数(Decimals)	(13)	2.4.8 变量定义信息的查询	(39)
2.1.5 变量标签(Label)	(13)	2.4.9 变量集的定义和使用	(40)
2.1.6 变量值标签(Values)	(14)	2.5 数据文件的拆分与合并	(42)
2.1.7 缺失值的定义方式(Missing)	(14)	2.5.1 数据文件的拆分	(42)
2.1.8 变量的显示宽度(Columns)	(15)	2.5.2 数据文件的合并	(43)
2.1.9 变量显示的对齐方式(Align)	(15)	2.6 数据文件的存储和读取	(49)
2.1.10 变量的测试尺度(Measure)	(15)	2.6.1 数据文件的存储	(49)
2.2 输入数据	(15)	2.6.2 其他类型数据文件的读入	(50)
2.2.1 按变量输入数据	(15)	习题二	(59)
2.2.2 按观察量序号输入数据	(16)	第 3 章 数据清理与基本统计分析	(61)
2.2.3 按全屏幕任意单元格输入 数据	(16)	3.1 数据清理	(61)
2.3 编辑数据	(16)	3.1.1 奇异数据的检查与清理	(61)
2.3.1 数据的搜索	(16)	3.1.2 数据项(变量)之间的逻辑检 查与清理	(65)
		3.2 基本统计分析	(66)

3.2.1 频数统计	(66)	5.1.4 结果和讨论	(112)
3.2.2 描述性统计	(69)	5.2 两组独立样本的 T 检验	(113)
3.2.3 均值、均值标准误差、中 位数、众数和全距	(70)	5.2.1 基本概念与方法	(113)
3.2.4 方差和标准差	(76)	5.2.2 例题与数据文件	(113)
3.2.5 四分位数、十分位数和 百分位数	(77)	5.2.3 统计分析过程	(114)
3.2.6 峰度和偏度	(80)	5.2.4 结果和讨论	(115)
3.3 参数估计	(82)	习题五	(116)
3.3.1 总体均值(μ)与总体方差 (σ^2)的估计	(82)	第 6 章 方差分析	(118)
3.3.2 总体均值的置信区间的 计算	(83)	6.1 单因素方差分析	(118)
习题三	(85)	6.1.1 基本概念与方法	(118)
第 4 章 统计报表	(87)	6.1.2 例题与数据文件	(120)
4.1 OLAP Cubes 过程	(87)	6.1.3 统计分析过程	(121)
4.1.1 例题与数据文件	(87)	6.1.4 结果和讨论	(124)
4.1.2 统计分析过程	(88)	6.2 无重复实验的双因素方差分析	(127)
4.1.3 结果和讨论	(91)	6.2.1 基本概念与方法	(127)
4.2 Case Summaries 过程	(93)	6.2.2 例题与数据文件	(128)
4.2.1 例题与数据文件	(93)	6.2.3 统计分析过程	(129)
4.2.2 统计分析过程	(93)	6.2.4 结果和讨论	(133)
4.2.3 结果和讨论	(95)	6.3 重复实验的双因素方差分析	(138)
4.3 Report Summaries in Rows 过程	(97)	6.3.1 基本概念与方法	(138)
4.3.1 例题与数据文件	(97)	6.3.2 例题与数据文件	(140)
4.3.2 统计分析过程	(97)	6.3.3 统计分析过程	(141)
4.3.3 结果和讨论	(103)	6.3.4 结果和讨论	(143)
4.4 Report Summaries in Columns 过程	(104)	6.4 协方差分析	(145)
4.4.1 例题与数据文件	(104)	6.4.1 基本概念与方法	(145)
4.4.2 统计分析过程	(104)	6.4.2 例题与数据文件	(146)
4.4.3 结果和讨论	(107)	6.4.3 统计分析过程	(147)
习题四	(107)	6.4.4 结果和讨论	(149)
第 5 章 T 检验	(109)	习题六	(149)
5.1 配对 T 检验	(109)	第 7 章 相关分析	(151)
5.1.1 基本概念与方法	(109)	7.1 连续变量的相关分析	(151)
5.1.2 例题与数据文件	(110)	7.1.1 基本概念与方法	(151)
5.1.3 统计分析过程	(110)	7.1.2 例题与数据文件	(152)
		7.1.3 统计分析过程	(152)
		7.1.4 结果和讨论	(154)
		7.2 等级变量的相关分析	(154)
		7.2.1 基本概念与方法	(154)
		7.2.2 例题与数据文件	(155)

7.2.3 统计分析过程	(155)	9.1.2 例题与数据文件	(196)
7.2.4 结果和讨论	(156)	9.1.3 统计分析过程	(197)
7.3 偏相关分析	(157)	9.1.4 结果和讨论	(199)
7.3.1 基本概念与方法	(157)	9.2 单样本 K-S 检验	(199)
7.3.2 例题与数据文件	(158)	9.2.1 基本概念与方法	(199)
7.3.3 统计分析过程	(159)	9.2.2 例题与数据文件	(200)
7.3.4 结果和讨论	(160)	9.2.3 统计分析过程	(200)
7.4 距离相关分析	(161)	9.2.4 结果和讨论	(202)
7.4.1 基本概念与方法	(161)	9.3 两独立样本非参数检验	(202)
7.4.2 例题与数据文件	(161)	9.3.1 基本概念与方法	(202)
7.4.3 不相似性分析	(162)	9.3.2 例题与数据文件	(203)
7.4.4 相似性分析	(164)	9.3.3 统计分析过程	(204)
习题七	(166)	9.3.4 结果和讨论	(206)
第 8 章 回归分析	(167)	9.4 多独立样本非参数检验	(207)
8.1 一元线性回归分析	(168)	9.4.1 基本概念与方法	(207)
8.1.1 基本概念与方法	(168)	9.4.2 例题与数据文件	(208)
8.1.2 例题与数据文件	(169)	9.4.3 统计分析过程	(209)
8.1.3 统计分析过程	(170)	9.4.4 结果和讨论	(211)
8.1.4 结果和讨论	(175)	9.5 两配对样本非参数检验	(212)
8.2 多元线性回归分析	(179)	9.5.1 基本概念与方法	(212)
8.2.1 基本概念与方法	(179)	9.5.2 例题与数据文件	(213)
8.2.2 例题与数据文件	(180)	9.5.3 统计分析过程	(214)
8.2.3 统计分析过程	(181)	9.5.4 结果和讨论	(216)
8.2.4 结果和讨论	(183)	9.6 多配对样本非参数检验	(218)
8.3 曲线拟合	(185)	9.6.1 基本概念与方法	(218)
8.3.1 基本概念与方法	(185)	9.6.2 实例 1	(219)
8.3.2 例题与数据文件	(185)	9.6.3 实例 2	(221)
8.3.3 统计分析过程	(186)	9.6.4 实例 3	(223)
8.3.4 结果和讨论	(189)	习题九	(225)
8.4 含虚拟自变量的回归分析	(190)	第 10 章 卡方检验	(227)
8.4.1 基本概念与方法	(190)	10.1 基本概念与方法	(228)
8.4.2 例题与数据文件	(190)	10.1.1 问题的数学描述	(228)
8.4.3 统计分析过程	(192)	10.1.2 分类变量数据录入格式	(228)
8.4.4 结果和讨论	(192)	10.2 一般卡方检验	(229)
习题八	(194)	10.2.1 实例 1	(229)
第 9 章 非参数检验	(195)	10.2.2 实例 2	(235)
9.1 二项分布检验	(195)	10.3 配对卡方的一致性检验	(237)
9.1.1 基本概念与方法	(195)	10.3.1 例题与数据文件	(237)

10.3.2 统计分析过程	(238)	12.1.2 因子分析的基本步骤	(274)
10.3.3 结果和讨论	(240)	12.2 统计分析过程	(276)
10.4 分层卡方检验	(241)	12.2.1 实例1	(276)
10.4.1 例题与数据文件	(241)	12.2.2 实例2	(286)
10.4.2 统计分析过程	(241)	习题十二	(296)
10.4.3 结果和讨论	(243)	第13章 统计图形	(298)
习题十	(245)	13.1 交互式图形	(299)
第11章 聚类分析	(247)	13.1.1 交互式图形的生成	(299)
11.1 层次聚类分析	(247)	13.1.2 交互式图形的编辑	(302)
11.1.1 基本概念与方法	(247)	13.2 统计图形的生成	(308)
11.1.2 实例1	(248)	13.2.1 简单条形图	(308)
11.1.3 实例2	(258)	13.2.2 分组(类)条形图	(316)
11.2 快速聚类分析	(263)	13.2.3 堆积图(分段条形图)	(316)
11.2.1 基本概念与方法	(263)	13.3 线图	(318)
11.2.2 例题与数据文件	(264)	13.3.1 复合的线图	(318)
11.2.3 统计分析过程	(264)	13.3.2 点线图	(320)
11.2.4 结果和讨论	(266)	13.4 饼图	(322)
习题十一	(270)	13.5 直方图	(325)
第12章 因子分析与主成分分析	(272)	13.6 散点图	(327)
12.1 基本概念与方法	(273)	习题十三	(328)
12.1.1 数学模型	(273)	参考文献	(330)

第 1 章 SPSS 与教育统计

1.1 SPSS 概述

SPSS (Statistical Package for the Social Sciences, 社会科学统计软件包) 是世界著名的统计分析软件之一。1968 年, 3 位美国斯坦福大学的学生开发了最早的 SPSS 统计软件系统, 并基于这一系统于 1975 年在芝加哥合伙成立了 SPSS 公司。20 世纪 80 年代以前, SPSS 统计软件主要应用于企事业单位, 1984 年 SPSS 总部推出了世界第一个统计分析软件微机版本 SPSS/PC+, 开创了 SPSS 微机系列产品的开发方向, 从而确立了该软件在个人用户市场第一的地位。1994~1998 年间, SPSS 公司陆续购并了 SYSTAT 公司、BMDP 软件公司、Quantime 公司、ISL 公司等, 并将各公司的主要产品融合集成, 使 SPSS 公司由原来单一统计产品的开发与销售转向为企业、教育科研及政府机构提供全面信息统计决策支持服务, 成为走在了最新流行的“数据仓库”和“数据挖掘”领域前沿的一家综合统计软件公司。伴随着 SPSS 产品服务领域的扩大和服务深度的增加, SPSS 公司已决定将它的英文全称更改为 Statistical Product and Service Solutions, 意为“统计产品与服务解决方案”。目前, SPSS 公司全球约有 25 万家产品用户, 它们分布于通信、医疗、银行、证券、保险、制造、商业、市场研究、科研、教育等多个领域和行业, SPSS 是世界上应用最广泛的统计软件之一。在国际学术界有条不紊的规定, 即在国际学术交流中, 凡是用 SPSS 软件完成的计算和统计分析, 可以不必说明算法, 由此可见其影响之大和信誉之高。目前, 世界上最著名的数据分析软件有 SAS 和 SPSS。SAS 由于是为专业统计分析人员设计的, 具有功能强大、灵活多样的特点, 为专业人士所喜爱; 而 SPSS 是为广大非专业人士设计的, 它操作简便, 好学易懂, 简单实用, 因而很受非专业人士的青睐。

SPSS 使用 Windows 的窗口方式展示各种管理和分析数据的方法, 使用对话框展示出各种功能选择项, 只要掌握一定的 Windows 操作技能并了解统计分析原理, 就可以使用该软件为特定的科研工作服务。SPSS 的基本功能包括数据管理、统计分析、输出管理等。其统计分析过程包括描述性统计、均值比较、方差分析、相关分析、回归分析、对数线性模型、聚类分析、因子分析、生存分析、时间序列分析、多重响应等大类, 每类中又分好几个统计过程。如回归分析中又分线性回归分析、曲线估计、Logistic 回归等几个统计过程, 并且每个过程中又允许用户选择不同的方法及参数。SPSS 中还有专门的报表和绘图系统, 可以根据数据产生各类统计表和绘制各种图形。

SPSS 名为社会科学统计软件包,这是为了强调其社会科学应用的一面,即为非理工专业的社会科学工作者使用,而实际上它在社会科学、自然科学的各个领域都能发挥巨大的作用,并已经应用于经济学、生物学、教育学、心理学、医学以及体育、工业、农业、林业、商业和金融等各个领域。

到目前为止,SPSS 已具有适合于 DOS、Windows、UNIX、Macintosh、OS/2 等多种操作系统使用的产品,国内常用的是其适用于 DOS 和 Windows 的版本。DOS 版本现已不再使用,SPSS for Windows 界面友好,功能强大,使用方便,在国内非常流行,主要版本有 SPSS V7.0、SPSS V7.5、SPSS V8.0、SPSS V9.0、SPSS V10.0、SPSS V11.0 和 SPSS V11.5 等。SPSS V10.0 以上版本有两种结构,一种是服务器(Server)/客户机(Client)结构,由 SPSS Server 和 SPSS for Windows 两部分组成;另外一种结构是单机版本,即 SPSS for Windows 标准版。

本书以 SPSS 11.5 for Windows 标准版为例进行讲解,并在书中简称为 SPSS。

1.1.1 启动 SPSS

SPSS 安装完毕后,系统会自动在 Windows 菜单中创建快捷方式。单击 Windows 的“开始”按钮,在“所有程序”菜单项“SPSS for Windows”中找到“SPSS 11.5 for Windows”,如图 1.1 所示,单击即可启动 SPSS。

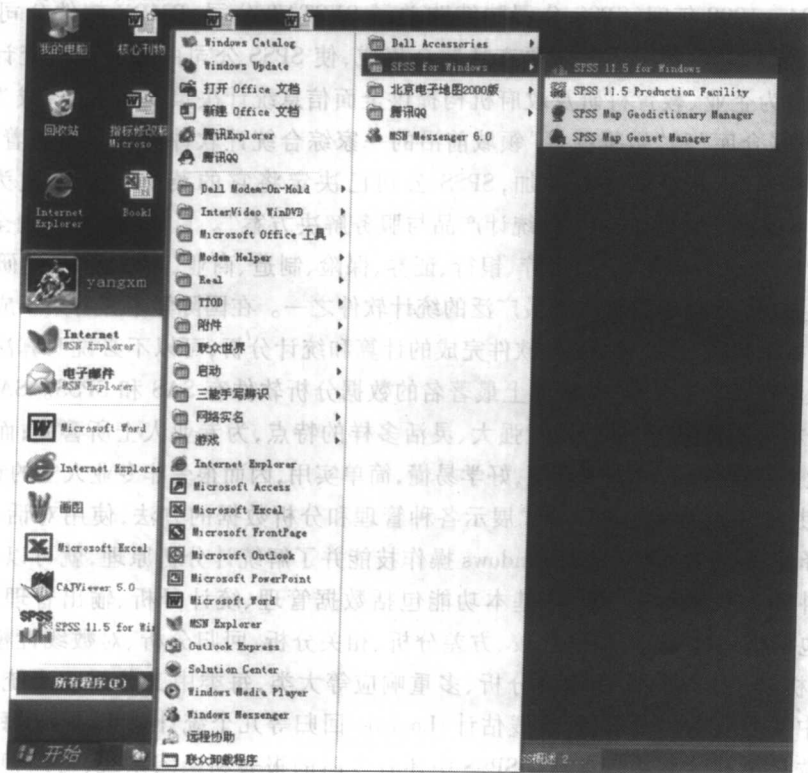


图 1.1 启动 SPSS

1.1.2 SPSS 的数据编辑窗口

SPSS 主界面主要有两个,一个是 SPSS 数据编辑窗口,另一个是 SPSS 输出窗口。本节先介绍数据编辑窗口。数据编辑窗口由标题栏、菜单栏、工具栏、编辑栏、变量名栏、内容区、窗口切换标签页和状态栏组成,如图 1.2 所示。

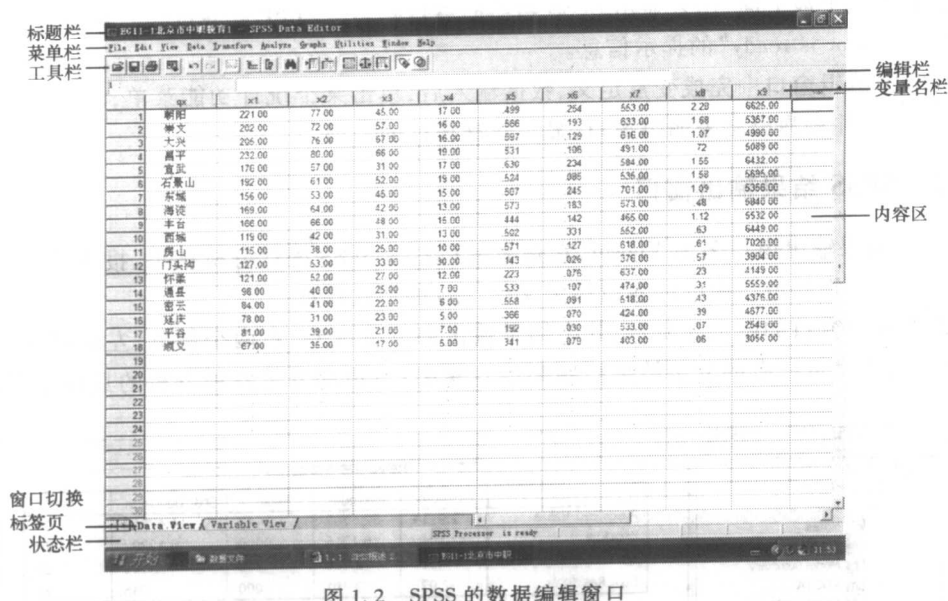


图 1.2 SPSS 的数据编辑窗口

标题栏中显示编辑的数据文件名,本图中所编辑的数据文件为“EG11-1 北京市中职教育1”。

菜单栏中列出了 SPSS 的命令菜单,每个菜单对应一组相应的功能。其中“File”是对 SPSS 文件的操作;“Edit”是 SPSS 文件的编辑菜单;“View”是用户界面设置菜单;“Data”是数据文件的建立和编辑菜单;“Transform”是数据基本处理菜单;“Analyze”是统计分析菜单,主要统计功能都集中在该菜单中;“Graphs”是统计图形菜单;“Utilities”是相关应用和设置菜单;“Window”是 SPSS 各窗口切换菜单;“Help”是 SPSS 帮助菜单。

工具栏中列出了一些常用命令的快捷图标。

编辑栏中可以输入数据。

变量名栏中列出了该编辑文件中所含有的变量名。SPSS 将变量自动命名为 var0001、var0002 等。图 1.2 中有 10 个变量:qx、x1、x2、x3、…、x9。

内容区列出了各个样品在变量中的取值。SPSS 中每一行表示一个样品。内容区的最左边是行的标号。

窗口切换标签页位于窗口的下方,图 1.2 中有两个,分别为“Data View”(数据视图)和“Variable View”(变量视图)。这两种视图提供了一种类似于电子表格的方法,用以产生和编辑 SPSS 数据文件中的变量和数据。“Data View”对应的表格用于查看、录入和修改数据;“Variable View”对应的表格用于输入和修改变量的定义,这样使用者就可以非常方便地进行变量类型的定义和数据的输入。

状态栏位于最下面,用来显示 SPSS 当前的运行状态。当 SPSS 等待用户操作时,会出现“SPSS Processor is ready”的提示信息。

在数据编辑窗口中完成变量定义、数据输入后,单击某个统计功能菜单,SPSS 会自动完成统计分析,并将弹出结果输出窗口,显示数据统计的结果。

1.1.3 SPSS 结果输出窗口

SPSS 结果输出窗口名为 Viewer,它是显示和管理 SPSS 统计分析结果、报表及图形的窗口。读者可以将此窗口中的内容以结果文件 .spo 的形式保存。

结果输出窗口如图 1.3 所示,其中结果输出部分分成左、右两个部分,左边部分是索引输出区,用于显示已有的分析结果标题和内容索引;右边部分是各项分析的具体结果,称为

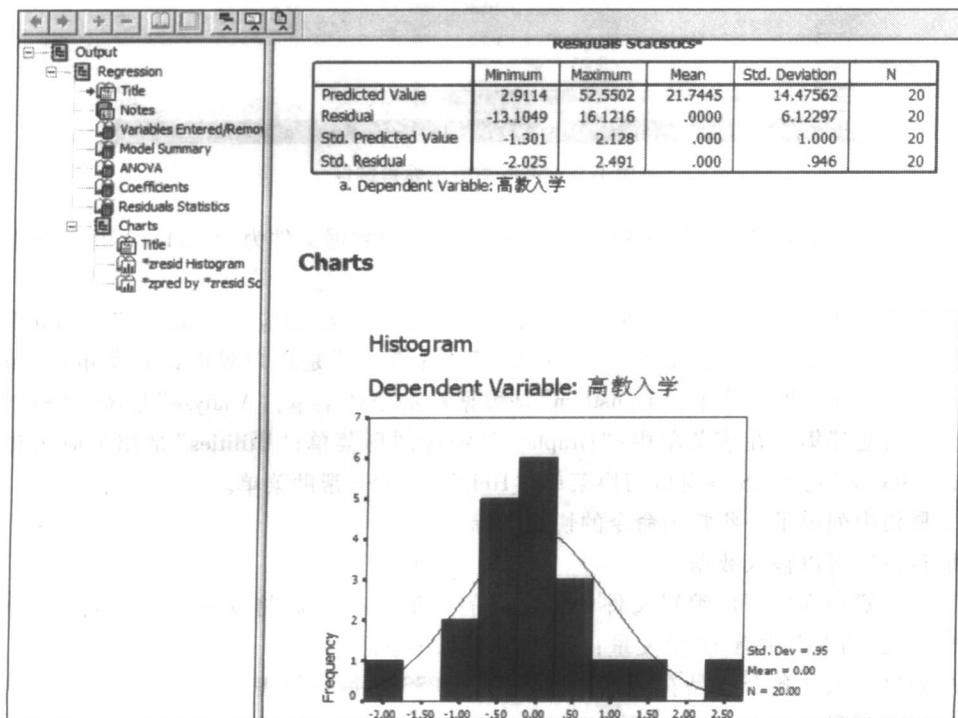


图 1.3 SPSS 结果输出窗口

详解输出区。

可以对详解输出区中的表格进行编辑等操作。当某个表格处于选中状态时,相应表格的四周会出现黑色单线边框,可以通过鼠标单击来选中某个表格。如果要对选中表格进行编辑,可双击该表格,当表格四周出现黑色斜线边框时,即可对表格内数据进行修改。

1.1.4 退出 SPSS

选择数据编辑窗口的“File”菜单中“Exit”命令,或单击标题栏上的关闭按钮,即可退出 SPSS。

1.2 教育统计中的问题

统计学是一门搜集、整理、描述、显示和分析统计数据的科学,其目的是探索数据内在的数量规律性,也称为“数据的科学”。教育统计是统计学在教育科学中的应用,是研究教育统计数据的科学。

实证研究方法是教育科学研究中极为重要也是非常普遍的研究方法,其过程是:根据研究目的,设计调查问卷,利用特定调查手段,采集所需数据,利用定量分析方法对获得的数据进行处理、分析、判断,得出相关结论。统计学方法是定量分析方法中的一种重要手段,它可以解决教育实证研究中的很多种问题。本节主要将教育科学研究中经常碰到的典型问题,进行归类,然后给出可以解决此类问题可采用的统计方法,说明在 SPSS 统计软件中如何实现。

1.2.1 数据采集

1. 基本概念

(1) 总体(Population Or Universe)

所研究对象的全体称为总体(或母体)。例如,某专业同一年的考生可以构成一个总体,所有的大学生、所有的中学生、某中学所有 18 岁的人等都可以分别构成一个总体。

(2) 个体(Individual Or Case)

组成总体的元素(Element)称为个体。在一个总体中,若个体的数目是有限的(Finite),则该总体称为有限总体,如某专业的同一年考生、所有的大学生等都分别是有限总体的例子;若个体的数目是无限的(Infinte),则该总体称为无限总体。

教育研究中的总体一般都是有限总体。

(3) 指标或变量(Variable)

研究任何总体,必然要对其个体的某些特征(当然也是总体的特征)感兴趣。例如,对“北京从业人员”这个总体,会有兴趣于研究从业人员的年收入、从业人员对有关政策的态

度、从业人员的受教育程度等。这些从业人员(个体)的信息,就是所说的个体特征(Characteristics)。研究总体,其实就是研究总体中的个体的这些特征(信息)。表征个体特征(信息)的量,称为指标或变量。

有时,若要联合考察总体或个体的若干个特征,所面对的就是指标向量或称变量向量。例如,当考察“学生”这个总体时,可能会对“学生的考分”、“学生的身高”等特征感兴趣。当要想联合考察这两个特征时,所要面对的就是指标向量或变量向量(学生的考分、学生的身高)。

(4) 指标值(变量值)或数据(Data)

在研究一个总体时,所要研究的每个特征(指标或变量)在每个个体上,都有一个反映该特征的具体描述(可以是数字,也可以是文字),这些特征的具体描述被称为指标值(变量值)或数据。例如,80分、90分等是变量“考生考分”的变量值;男或女是变量“考生性别”的变量值。这些具体的数值,也称为数据(Data)。

2. 数据类型

统计数据是对客观现象进行计量的结果,在搜集数据之前,要先对现象进行计量或测度,不同事物使用的计量或测度的程度是不同的。按照对客观事物测度的程度或精确水平不同,采用的计量尺度由低级到高级、由粗略到精确,数据可分为4个层次:定类、定序、定距和定比。

(1) 定类数据

也称名义级(Nominal)数据,是数据的最低级。它仅仅是一种标志,用以区分变量的不同值,但没有序次关系。比如,“性别”便是一个名义测度等级的变量,可以用数码1表示男性,数码2表示女性;或者反过来编码,数码1表示女性,数码2表示男性;也可以用其他任意两个不同的数码或者文字代表类别。不论上述哪一种编码,变量所包含的信息都没有任何损失。该级别的变量值可以用数字来表示,也可以用字母来表示。

(2) 定序数据

也称序次级(Ordinal)数据,是数据的中间级。例如,受教育程度就是序次测度等级的数据。可以采用数字编码表示不同类别,比如,文盲半文盲=1,小学=2,初中=3,高中=4,大学=5,硕士研究生=6,博士研究生=7。当然,也可以倒过来用7~1来表示不同的等级,信息一点也没有损失。各编码的序值代表了受教育程度的高低差异,虽然其不能准确描述这一差别的大小,但是可以确定其顺序。该级别的变量值可以用数字来表示,也可以用字母来表示。

(3) 定距数据

也称间距级数据,它是具有一定单位的实际测量值。通过它可以知道两个不同的变量值存在差异,可以做减法运算,也可以做加法运算,它们都有实际意义。但是,间距级测度的变量值之间不能进行乘、除计算。这是因为这一测度等级的变量所取的0值,不是物理意义上的绝对0。该级别的变量值只能用数字来表示。

(4) 定比数据

也称比率数据,这是数据的最高等级。比率级是数据最高级的测度等级。它是具有一定单位(例如元、米等)的实际测量值。这类数据不仅可以做加、减运算(例如求两个数据的差值),而且可以做乘、除运算。此时,要求其0值不是人为制定的。例如,0收入。只有当变量的0值不是人为制定时,其任意两个取值的比率(分母不为0)才能有确定的意义。

一般说来,数据的等级越高,应用范围越广泛;等级越低,应用范围越受限。不同测度级别的数据,应用范围不同。等级高的数据,可以兼有等级低的数据的功能;而等级低的数据,不能兼有等级高的数据的功能。统计学上也将这4类数据分为两种类型:定性数据和定量数据。定性数据也称品质数据,它说明的是研究对象的品质特征,不能用数值来表现,主要代表定类数据和定序数据;定量数据也称数量数据,它说明的是研究对象的数量特征,能用数值来表现,主要代表定距数据和定比数据。对于这两类不同类型的数据,采用不同的统计方法来处理和分析。对品质数据通常计算出各组的频数或频率;而数量数据则可以用均值或其他更为复杂的统计方法进行分析。

3. 数据采集方法

(1) 观察数据

凡是在获得数据的过程中,不对被调查对象数据产生的条件施加任何控制,所得到的数据称为观察数据。观察数据通常是(但不一定是)众多因素共同作用的结果。例如,调查农民年收入情况、受教育情况、教育费支出情况的数据,都只能是观察数据。获得数据的方法主要有两种。

① 普查

普查是针对有限总体而言的。普查就是收集有限总体中每个个体的有关指标的指标值(或变量值)。

② 抽样调查

也称部分调查,它是在总体中选择一部分个体进行调查,从所了解的局部数据来了解总体情况。总体中,部分个体所组成的集合称为样本(Sample)集合。抽样调查有两层含义:一是抽样(抽取部分个体);二是调查(调查那些反映在个体上的、所要研究的特征的数据)。调查样本个体(抽样所得到的个体)所得到的数据,称为样本数据。

(2) 实验数据

在获得数据的过程中,对数据产生的条件实施了控制而得到的数据。实验数据通常是(但不一定是)单一因素作用的结果,获得数据的方法是实验方法。

1.2.2 教育统计的常见问题类型

1. 问题类型

(1) 对采集数据的一般性统计,如频数、频率、均值、方差等

例如,抽样调查某地区家庭义务教育支出,其中问卷调查项目有家庭人口、父母受教育年限、子女人数、上学人数、家庭人均收入、家庭人均支出、教育支出、少数民族比例。要对整个抽样做统计,说明此地区上述指标情况,就要做一般性统计。

(2) 两个总体之间某类特征数据的差异显著性

例如,研究我国重点与非重点两类大学毕业生收入有无差异问题。

(3) 多个总体之间某类特征数据的差异显著性

例如,研究具有博士学位、硕士学位和学士学位毕业生期望收入有无差异问题。

(4) 一个或多个因素对结果影响的显著性

例如,不同性别、不同地区、不同家庭背景的学生接受高等教育情况有无差异;教学手段与课外科研活动是否对学生学习成绩有影响。

(5) 两个特征变量数据的相关性(相关程度)大小

例如,个人受教育年限与个人收入关系密切程度。

(6) 一变量与另一变量或多个变量之间的近似函数关系

例如,一个地区人均教育支出与人均国内生产总值近似函数关系。

(7) 某变量是否服从特定分布

例如,某校学生月生活费支出是否服从正态分布。

(8) 如何将多个研究对象进行分类(聚类)

例如,将我国 31 个省市按人均教育经费多少分成五大类。

(9) 如何将多个指标描述的对象简化成少量指标描述

例如,影响小学辍学率的因素有很多,比如人均国内生产总值、人均教育经费、农民人均收入、当地文盲率等十几个因素,能否简化成几个综合因素(因子)。

(10) 如何将多个用不同量纲指标描述的研究对象进行综合排序

例如,衡量一个地区教育现代化水平有多个指标,而且这些指标量纲都不一样,现有几个地区,按教育现代化水平排序,如何进行。

2. 数据类型

按照统计学处理问题的方法分类,不同的数据类型有不同的统计分析方法。一般分为定性数据和定量数据。定量数据中又分为服从或近似服从正态分布数据、非正态分布数据,如图 1.4 所示。

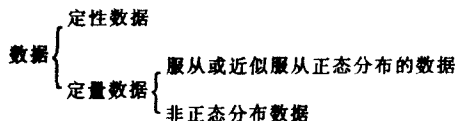


图 1.4 数据类型的分类

1.3 SPSS 如何解决教育统计中的问题

本书的框架基本是按照解决上述十种类型问题展开的,SPSS 在教育统计中的应用就是

如何利用 SPSS 软件解决这十种问题。解决问题的统计方法对应本书章节如下表所示。

表 解决问题的统计方法

问 题 类 型	解 决 方 法	数据类型	
		定性数据	定量数据
			服从或近似 服从正态分布 非正态分布
1	第 3 章 数据清理与基本统计分析 基本统计分析	第 3 章 数据清理与基本统计分析 基本统计分析	第 3 章 数据清理与基本统计分析 基本统计分析
2	第 10 章 卡方检验 一般卡方检验	第 5 章 T 检验 配对 T 检验 两组独立样本的 T 检验	第 9 章 非参数检验 两独立样本非参数检验 两配对样本非参数检验
3	第 10 章 卡方检验	第 6 章 方差分析	第 9 章 非参数检验 多独立样本非参数检验 多配对样本非参数检验
4		第 6 章 方差分析	第 9 章 非参数检验 多独立样本非参数检验 多配对样本非参数检验
5	第 10 章 卡方检验	第 7 章 相关分析	第 7 章 相关分析
6		第 8 章 回归分析	第 8 章 回归分析
7			第 9 章 非参数检验 单样本 K-S 检验
8	第 11 章 聚类分析	第 11 章 聚类分析	第 11 章 聚类分析
9	第 12 章 因子分析与主成分分析	第 12 章 因子分析与主成分分析	第 12 章 因子分析与主成分分析
10	第 12 章 因子分析与主成分分析	第 12 章 因子分析与主成分分析	第 12 章 因子分析与主成分分析

习 题 一

1. SPSS 的数据类型有哪些？有何特征？
2. 在研究学习中碰到哪些需要解决的问题？这些问题是否可以归类到本章中所提的问题类别？