



普通高等教育“十一五”国家级规划教材

SPSS在教育统计中的应用

(第2版)

Applications of SPSS to
Education Data Analysis

杨晓明 编著



高等教育出版社
HIGHER EDUCATION PRESS

普通高等教育“十一五”国家级规划教材

SPSS 在教育统计中的应用

SPSS zai Jiaoyu Tongji zhong de Yingyong

(第2版)

杨晓明 编著



高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

内容提要

本书是普通高等教育“十一五”国家级规划教材。

本书将教育科研中经常遇到的数据统计问题归纳成几种典型问题,并通过数理统计方法讲解解决这些问题的思路,最后依托 SPSS 16.0 软件演示完成实际问题的求解步骤。

本书希望帮助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,方便、准确地分析教育数据,促进其开展教育研究工作。

本书既可作为教育类专业本科生、研究生的统计课程教材,也可作为广大教育科研工作者、教育统计人员的参考书。

图书在版编目(CIP)数据

SPSS 在教育统计中的应用/杨晓明编著.--2 版.

--北京:高等教育出版社,2012.5

ISBN 978-7-04-034161-4

I. ①S… II. ①杨… III. ①教育统计-统计分析-软件包,SPSS-高等学校-教材 IV. ①G40-051

中国版本图书馆 CIP 数据核字(2012)第 055268 号

策划编辑 耿芳	责任编辑 耿芳	封面设计 于文燕	版式设计 马敬茹
责任校对 杨凤玲	责任印制 朱学忠		

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 100120
印 刷 保定市中国画美凯印刷有限公司
开 本 787mm×1092mm 1/16
印 张 24
字 数 550 千字
购书热线 010-58581118
咨询电话 400-810-0598

网 址	http://www.hep.edu.cn http://www.hep.com.cn
网上订购	http://www.landaco.com http://www.landaco.com.cn
版 次	2004 年 5 月第 1 版 2012 年 5 月第 2 版
印 次	2012 年 5 月第 1 次印刷
定 价	37.70 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 34161-00

前 言

从内容上看,本书第2版与第1版相比有了较大变化,一是对大部分案例做了调整,新版案例基本上取自作者课题研究中的实例。二是丰富了统计方法,增加了统计表、制表、统计图绘制等内容,特别是在回归分析中,增加了有关定性因变量回归分析,例如逻辑回归分析、最优尺度回归;还增加了对应分析、两阶段聚类分析、可靠性分析、典型相关分析和偏最小二乘回归等社会科学研究中重要的分析方法;删除了非参数检验。三是根据读者建议,加强了对软件输出结果的详细解释。

本书特色是采用“典型问题→统计方法→案例分析”模式,以教育科学研究中的典型问题、典型统计方法为导向,侧重讲解应用 SPSS 解决这些典型问题的方法。本书的创新之处在于其具有工具型教科书的特点,即学生在研究中遇到相关问题时,只要参照本书案例给出的问题解决方法及程序,就能很好地解决。本书叙述简明、扼要、易懂,使用者仅需要关心求解问题的统计方法,在 SPSS 软件中的操作步骤,以及掌握计算结果的解释,而不需要了解具体理论和运算过程。

本书共有 15 章。第 1 章为 SPSS 与教育统计,第 2 章为数据文件的处理,第 3 章为数据清理与基本统计分析,第 4 章为统计表与制表,第 5 章为统计图的绘制,第 6 章为均值检验,第 7 章为定量变量相关分析,第 8 章为定性变量相关分析,第 9 章为定量因变量回归分析,第 10 章为定性因变量回归分析,第 11 章为聚类分析,第 12 章为因子分析与主成分分析,第 13 章为可靠性分析,第 14 章为典型相关分析与偏最小二乘回归,第 15 章为综合问题。

第 1 章是全书的概括,也是本书的主体架构,包括 SPSS 介绍、统计学基本概念、教育研究中的基本模式、教育研究中的基本问题及其类型、解决各类问题的统计方法。第 2~14 章分别介绍了各类问题及其解决方法,包括问题、统计方法、案例、SPSS 实现步骤、结果分析等内容。第 15 章是对教育研究模式、研究问题类型、解决问题的统计分析方法的综合和总结,它将全书内容整合贯通。

本书各章编写模式为:首先提出能够解决的问题类型和要求条件;其次给出具体问题类型的典型案例;再次在 SPSS 上进行解决案例的操作;最后对各种结果进行解释、分析,回答案例中提出的问题。

本书总策划包括杨晓明、丛彦、王亚莉,杨晓明担任主编,丛彦、王亚莉、王立敏、刘晶担任副主编。各章分工如下:第 1、11、12、14、15 章由杨晓明执笔;第 2 章由刘晶、苗丽执笔;第 3 章由刘晶执笔;第 4 章由毛建军执笔;第 5 章由马红权、夏建刚执笔;第 6 章由王立敏、孙美丽执笔;第 7 章由李晓梅、王立敏执笔;第 8、9、10 章由杨晓明、王立敏执笔;第 13 章由杜秀梅执笔。本书力图做到方便、简明、实用,易于查找,工具性强。由于作者水平所限,不足之处在所难免,敬请读者批评指正。

最后,感谢中国人民大学王琪延教授在审稿中提出的修改建议。

作者联系方式:yangxiaoming508@126.com。

杨晓明

2012年1月于北京

目 录

第 1 章 SPSS 与教育统计	1	2.3.4 变量计算	34
1.1 SPSS 概述	1	2.3.5 产生计数变量	38
1.1.1 启动 SPSS	2	2.3.6 离散化变量	40
1.1.2 SPSS 的数据窗口	2	2.3.7 变量的自动赋值	45
1.1.3 SPSS 结果输出窗口	4	2.3.8 变量定义信息的查询	45
1.1.4 退出 SPSS	5	2.3.9 变量集的定义和使用	47
1.2 统计学基本概念	5	2.4 数据文件的拆分与合并	49
1.2.1 数据采集	6	2.4.1 数据文件的拆分	49
1.2.2 数据类型	6	2.4.2 数据文件的合并	50
1.2.3 数据采集方法	8	2.5 数据文件的存储和读取	57
1.3 教育研究中的基本问题	8	2.5.1 数据文件的存储	57
1.3.1 研究问题和变量	8	2.5.2 其他类型数据文件的读入	59
1.3.2 研究问题的类型	9	习题	69
1.4 解决问题的统计方法	10	第 3 章 数据清理与基本统计分析	71
习题	12	3.1 数据清理	71
第 2 章 数据文件的建立与处理	13	3.1.1 奇异数据的检查与清理	71
2.1 定义数据的结构	13	3.1.2 数据项(变量)之间的逻辑检查与 清理	76
2.1.1 定义数据项的变量名	14	3.2 常用统计量	77
2.1.2 定义变量类型、变量长度、列显示 宽度	15	3.2.1 刻画集中趋势的描述统计量	77
2.1.3 变量标签	17	3.2.2 刻画离散程度的描述统计量	78
2.1.4 变量值标签	17	3.2.3 刻画分布形态的描述统计量	80
2.1.5 缺失值的定义方式	18	3.2.4 常用统计量的统计分析过程	80
2.1.6 变量显示的对齐方式	19	3.3 频数分析	82
2.1.7 变量的测试尺度	19	3.3.1 案例及数据文件	83
2.2 数据的输入与编辑	19	3.3.2 统计分析过程	83
2.2.1 输入数据	19	3.3.3 结果分析	86
2.2.2 编辑数据	20	3.4 描述分析	87
2.3 变量的处理	30	3.4.1 案例及数据文件	87
2.3.1 增加和删除一个变量	30	3.4.2 统计分析过程	87
2.3.2 指定加权变量	32	3.4.3 结果分析	88
2.3.3 变量值的重新编码	32	3.5 探索分析	88

3.5.1 案例及数据文件	88	5.4.2 生成饼图的过程	139
3.5.2 统计分析过程	88	5.4.3 饼图的编辑	141
3.5.3 结果分析	91	5.5 散点图/圆点图	142
3.6 交叉列联表分析	96	5.5.1 案例及数据文件	142
3.6.1 案例及数据文件	97	5.5.2 生成散点图的过程	143
3.6.2 交叉制表过程	97	5.5.3 散点图的编辑	144
3.6.3 结果分析	99	5.6 直方图	147
3.7 比率分析	101	5.6.1 案例及数据文件	147
3.7.1 案例及数据文件	101	5.6.2 生成直方图的过程	147
3.7.2 统计分析过程	101	5.6.3 直方图的编辑	150
3.7.3 结果分析	103	5.7 高低图	150
习题	104	5.7.1 案例及数据文件	151
第4章 统计表与制表	106	5.7.2 生成高低图的过程	151
4.1 统计表	106	5.7.3 高低图的编辑	151
4.1.1 例题及数据文件	107	5.8 箱图	154
4.1.2 结果分析	109	5.8.1 案例及数据文件	154
4.1.3 分层分类汇总统计	111	5.8.2 生成箱图的过程	154
4.2 制表	113	5.8.3 箱图的编辑	157
4.2.1 案例及数据文件	113	5.9 双轴图	157
4.2.2 统计分析过程	113	5.9.1 案例及数据文件	157
4.2.3 结果分析	120	5.9.2 生成 y 轴图的过程	157
习题	123	5.9.3 双轴图的编辑	159
第5章 统计图的绘制	124	习题	161
5.1 条形图	125	第6章 均值检验	162
5.1.1 案例及数据文件	125	6.1 配对样本 t 检验	162
5.1.2 生成条形图的过程	125	6.1.1 案例及数据文件	163
5.1.3 条形图的编辑	128	6.1.2 统计分析过程	163
5.2 线图	131	6.1.3 结果分析	165
5.2.1 案例及数据文件	131	6.2 独立样本 t 检验	166
5.2.2 生成线图的过程	131	6.2.1 案例及数据文件	166
5.2.3 线图的编辑	134	6.2.2 统计分析过程	166
5.3 面积图	135	6.2.3 结果分析	168
5.3.1 案例及数据文件	135	6.3 单因素方差分析	169
5.3.2 生成面积图的过程	135	6.3.1 案例及数据文件	170
5.3.3 面积图的编辑	137	6.3.2 统计分析过程	171
5.4 饼图	139	6.3.3 结果分析	173
5.4.1 案例及数据文件	139	习题	176

第7章 定量变量相关分析	178	9.1.4 多重共线性的诊断及处理方法	230
7.1 连续变量的相关分析	179	9.1.5 多重共线性案例	231
7.1.1 案例及数据文件	180	9.1.6 统计分析过程	231
7.1.2 统计分析过程	180	9.1.7 结果分析	233
7.1.3 结果分析	183	9.2 曲线拟合	237
7.2 等级变量相关分析	184	9.2.1 案例及数据文件	237
7.2.1 案例及数据文件	185	9.2.2 统计分析过程	237
7.2.2 统计分析过程	185	9.2.3 结果分析	240
7.2.3 结果分析	187	9.2.4 时间序列的曲线拟合	241
7.3 偏相关分析	188	9.3 含虚拟自变量的回归分析	246
7.3.1 案例及数据文件	189	9.3.1 案例及数据文件	246
7.3.2 统计分析过程	190	9.3.2 统计分析过程	247
7.3.3 结果分析	191	9.3.3 结果分析	249
习题	192	9.4 非线性回归模型	251
第8章 定性变量相关分析	195	9.4.1 案例及数据文件	252
8.1 一般卡方检验	197	9.4.2 统计分析过程	252
8.1.1 案例及数据文件	197	9.4.3 结果分析	254
8.1.2 统计分析过程	198	习题	255
8.1.3 结果分析	202	第10章 定性因变量回归分析	257
8.2 配对卡方的一致性检验	205	10.1 二值 Logistic 回归模型	257
8.2.1 案例及数据文件	205	10.1.1 Logistic 模型	257
8.2.2 统计分析过程	206	10.1.2 案例及数据文件	258
8.2.3 结果分析	207	10.1.3 统计分析过程	259
8.3 分层卡方检验	208	10.1.4 结果分析	260
8.3.1 案例及数据文件	208	10.2 多值 Logistic 回归模型	263
8.3.2 统计分析过程	209	10.2.1 案例及数据文件	263
8.3.3 结果分析	209	10.2.2 统计分析过程	263
8.4 对应分析	211	10.2.3 结果分析	266
8.4.1 案例及数据文件	211	10.3 最优尺度回归模型	269
8.4.2 统计分析过程	212	10.3.1 案例及数据文件	269
8.4.3 结果分析	215	10.3.2 统计分析过程	270
习题	218	10.3.3 结果分析	272
第9章 定量因变量回归分析	220	习题	275
9.1 多元线性回归分析	221	第11章 聚类分析	276
9.1.1 案例及数据文件	222	11.1 层次聚类分析	276
9.1.2 统计分析过程	222	11.1.1 案例及数据文件	277
9.1.3 结果分析	225	11.1.2 统计分析过程	277

11.1.3 结果分析	282	13.2.2 统计分析过程	321
11.1.4 案例及数据文件 2	286	13.2.3 结果分析	324
11.1.5 统计分析过程 2	286	习题	328
11.1.6 结果分析 2	287	第 14 章 典型相关分析与偏最小二乘	
11.2 快速聚类分析	289	回归	330
11.2.1 基本概念与方法	289	14.1 典型相关分析	330
11.2.2 案例及数据文件	290	14.1.1 案例及数据文件	333
11.2.3 统计分析过程	290	14.1.2 统计分析过程	334
11.2.4 结果分析	293	14.1.3 结果分析	336
11.3 两步聚类法	296	14.2 偏最小二乘回归	344
11.3.1 案例及数据文件	297	14.2.1 案例及数据文件	345
11.3.2 统计分析过程	297	14.2.2 统计分析过程	345
11.3.3 结果分析	300	14.2.3 结果分析	348
11.4 聚类方法的选择	305	14.2.4 注释	352
习题	306	习题	352
第 12 章 因子分析与主成分分析	307	第 15 章 综合问题	354
12.1 基本原理	307	15.1 研究问题及研究假设	354
12.2 案例分析	310	15.2 研究设计	355
12.2.1 案例及数据文件	310	15.3 统计验证	357
12.2.2 统计分析过程	310	15.3.1 基本统计	357
12.2.3 结果分析	313	15.3.2 可靠性	360
习题	317	15.3.3 差异检验	361
第 13 章 可靠性分析	319	15.3.4 相关分析	362
13.1 基本理论	319	15.3.5 主成分因子分析	363
13.1.1 可靠性分析的数据假设	319	15.3.6 聚类分析	363
13.1.2 信度估算方法	319	15.3.7 回归分析	363
13.1.3 信度系数	320	习题	371
13.2 案例分析	321	参考文献	372
13.2.1 案例及数据文件	321		

第 1 章 SPSS 与教育统计

1.1 SPSS 概 述

SPSS(Statistical Package for the Social Science, 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 公司总部推出了世界上第一个统计分析软件微型机版本 SPSS/PC+,开创了 SPSS 微型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中第一的地位。1994 至 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 公司由原来只从事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,成为走在最新流行的“数据仓库”和“数据挖掘”领域前沿的一家综合性统计软件公司。伴随着 SPSS 产品服务领域的扩大和服务深度的增加,SPSS 的全称变为“Statistical Product and Service Solutions”,意为“统计产品与服务解决方案”,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,是世界上应用最广泛的统计软件。在国际学术界有条不成文的规定,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是 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 最突出的特点就是操作界面极为简单易懂,输出结果界面美观。它使用 Windows 的窗口方式展示出各种管理和分析数据方法的功能,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,了解统计分析原理,就可以使用该软件进行特定的科研工作,是非专业统计人员的首选统计软件。在众多用户对国际常用统计软件 SAS、BMDP、GLIM、GENSTAT、EPILOG、Minitab 的总体印象分的统计结果中,SPSS 各项功能均获得最高分。SPSS 采用类似于 Excel 表格的方式输入与管理数据,数据接口较为通用,能方便地从其他数据库中读入数据。其统计过程包括了常用的、较为成熟的统计过程,完全可以满足非统计专业人士的工作需要。对于熟悉老版本编程运行方式的用户,SPSS 还特别设计了语法生成窗口,用户只需在菜单中选好各个选项,然后单击“粘贴”按钮就可以自动生成标准的 SPSS 程序,极大地方便了中高级用户。

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、非参数检验、线性模型、对数线性模型、聚类分析、因子分析、生存分析、时间序列分析、多重响应等几大类,每类中又分为几个统计过程,如回归分析中又包括线性回归分析、曲线估计、Logistic 回归等几个统计过程,并且在每个过程中又允许用户选择不同的方法及参数。SPSS 中还有专门的制表和绘图系统,可以根据数据产生各类统计表和绘制各种图形。

SPSS 之所以命名为社会学统计软件包,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所使用,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,并且已经被应用于经济学、生物学、教育学、心理学、医学等学科以及体育、工业、农业、林业、商业和金融等各个领域。

到目前为止,SPSS 已具有适合在 DOS、Windows、UNIX、Macintosh、OS/2 等多种操作系统中使用的产品,国内常用的是适用于 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,在国内外非常流行,目前最新版本已升级至 SPSS V20。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;另外一种单机版本,即 SPSS for Windows 标准版。版本的每次升级都对前一版本或从功能上进行了改进,或添加了分析方法模块。在本书完稿时,SPSS 公司已在 2010 年 9 月被 IBM 公司收购。本书将以 SPSS 16.0 for Windows 标准版为例进行介绍,并在本书中简称为 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“程序”菜单的“SPSS Inc”子菜单中找到“SPSS Statistics 16.0”菜单项并单击,即可启动 SPSS,如图 1.1 所示。

1.1.2 SPSS 的数据窗口

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。

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、窗口切换选项卡和状态栏组成,如图 1.2 所示。

(1) 标题栏是窗口中最上面的 1 行,栏中显示编辑的数据文件名,例如在图 1.2 所示窗口中所编辑的数据文件为“EG11-1 北京市中职教育.sav”。

(2) 菜单栏是窗口中的第 2 行。栏中列出了 SPSS 的菜单命令,每个菜单都提供一组功能。其中 File 是对 SPSS 文件的操作菜单;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,主要统计分析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;Add-ons 是程序加载项目菜单;Window 是 SPSS 各窗口切换菜单;Help 是 SPSS 帮助菜单。

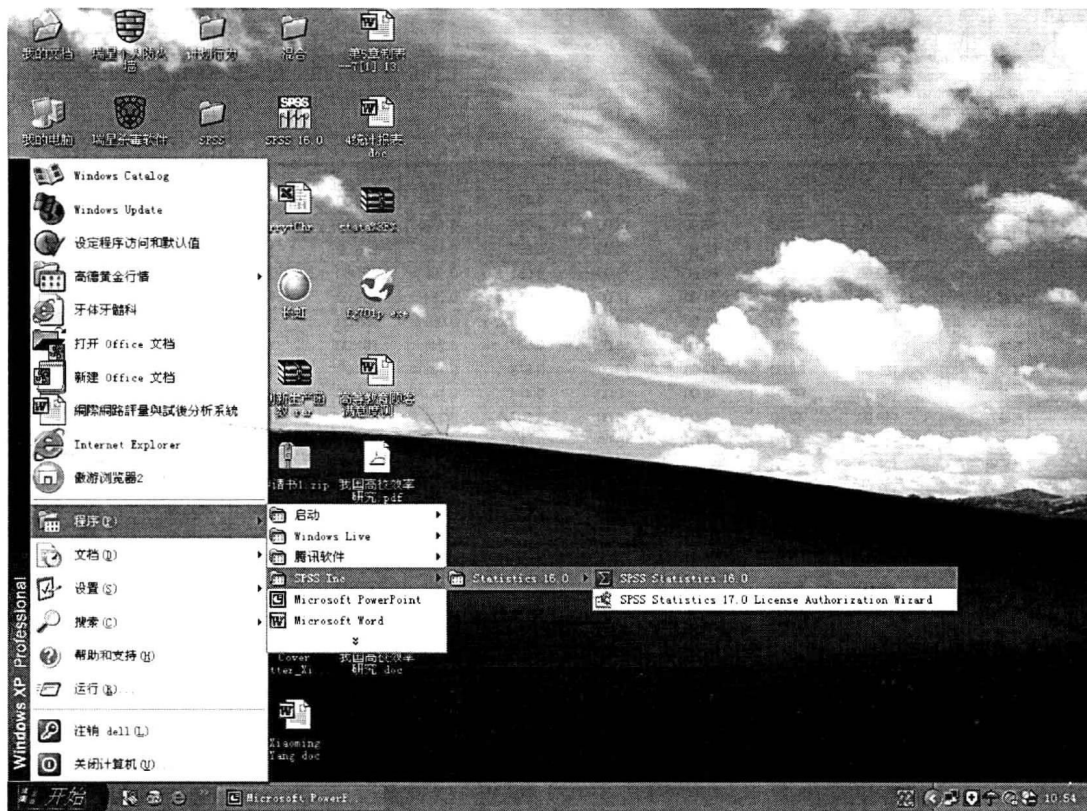


图 1.1 SPSS 启动菜单

(3) 工具栏是窗口中的第 3 行, 栏中列出了类似于 Word 软件中的一些工具栏按钮的快捷图标, 例如“保存”、“粘贴”、“打印”、“返回”等。

(4) 编辑栏是窗口中的第 4 行, 在栏中可以输入数据。

(5) 变量名栏是窗口中的第 5 行, 栏中列出了该编辑文件中所含有的变量名。SPSS 自动命名变量名为 var0001、var0002 等。本窗口中有 11 个变量: qx, x1, x2, x3, ..., x9, zh。

(6) 内容区是窗口中第 5 行下面的部分, 其中列出了各个样品的各个变量取值。SPSS 中的每一行表示一个样品。内容区的最左边是行的标号。

(7) 内容区下方有两个选项卡: “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了。

(8) 状态栏在最下面, 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。

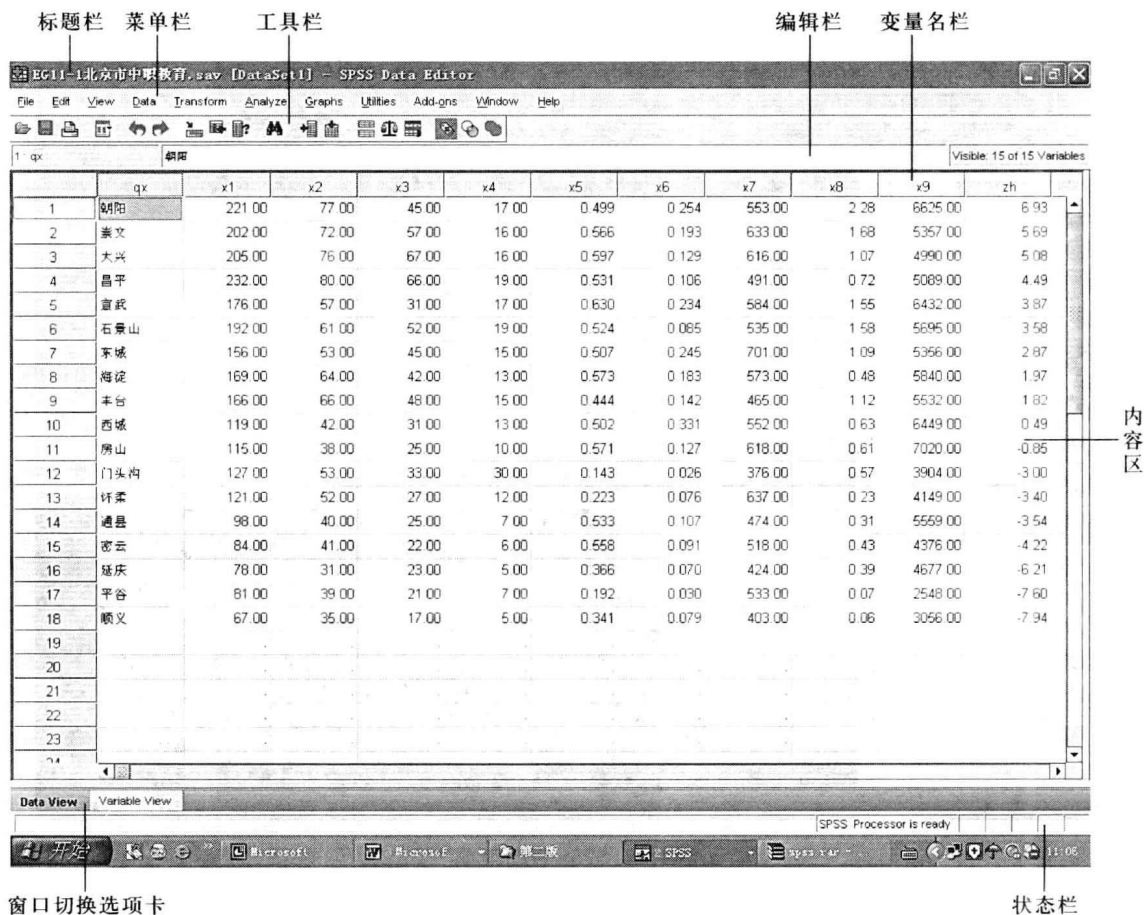


图 1.2 SPSS 数据编辑窗口

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,并打开结果输出窗口,其中存放了数据统计的结果。

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。用户可以将此窗口中的内容以结果文件 .spo 的形式保存起来。

其中结果输出窗口分成左右两个部分,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。右边部分是各种分析的具体结果,称为详解输出区,如图 1.3 所示。

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,单击即可选中某个表格。如果要对选中的表格进行编辑,可双击该表格,当表格四周出现黑色斜线边框时,即可对表格内的数据进行修改。

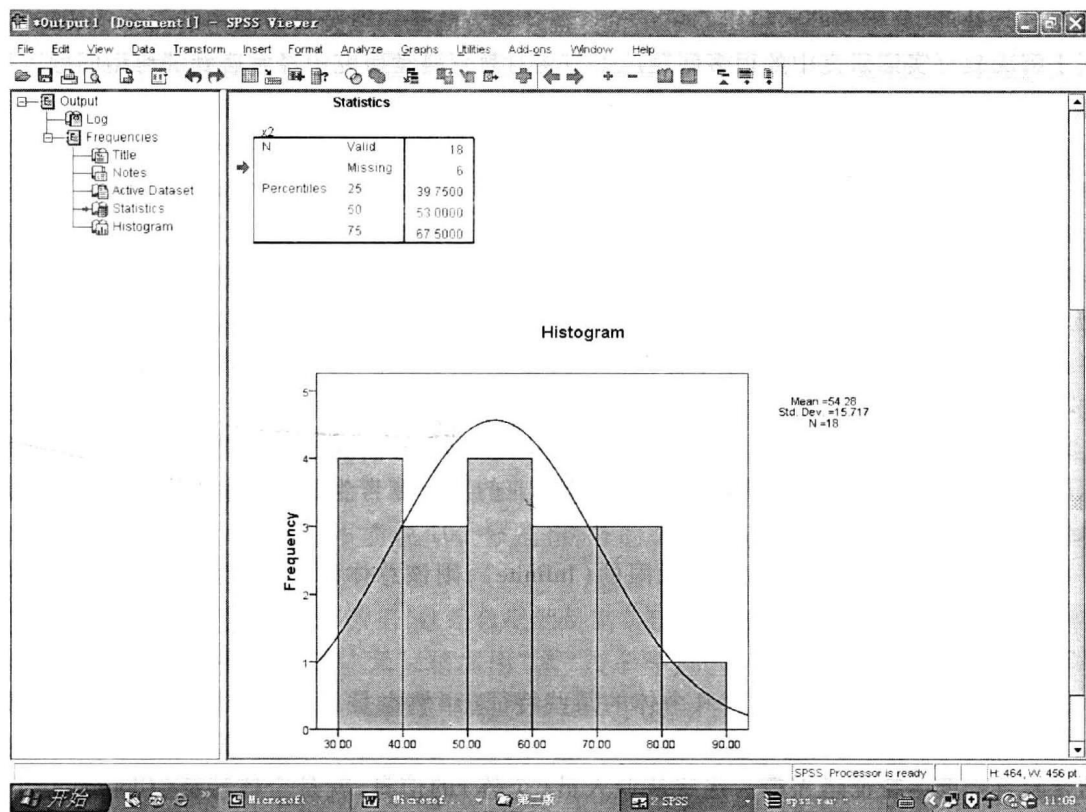


图 1.3 SPSS 结果输出窗口

1.1.4 退出 SPSS

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

1.2 统计学基本概念

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,进行实证研究的过程是: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个重要工具,可以用于解决教育实证研究中的很多问题。本节将对教育科学研究中经常碰到的典型问题进行归类,然后说明采用何种统计方法可以解决此类问题,以及在 SPSS 统计软件中怎样来实现。

1.2.1 数据采集

1. 总体

所研究对象的全体称为总体或母体(Population 或 Universe)。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

组成总体的元素(Element)称为个体(Individual 或 Case)。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。

在一个总体中,若个体的数目是无限的(Infinite),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

无论研究任何总体,必然会对其个体的某些特征(当然也是总体的特征)感兴趣。例如,对于“北京从业人员”这个总体,人们会对从业人员的年收入、从业人员对有关政策的态度、从业人员的受教育程度等特征感兴趣。这些从业人员(个体)的信息,就是个体特征(Characteristics)。研究一个总体,其实就是研究总体中的个体的特征(信息)。表示个体特征(信息)的量,称为指标或变量(Variable)。

要联合考察总体或个体的若干个特征,这时所面对的就是指标向量,或称为变量向量。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。当联合考察这两个特征时,所面对的就是指标向量,或变量向量(学生考分,学生身高)。

4. 指标值(变量值)或数据

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)。

1.2.2 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也是不同的。按照对客观事物测度的程度或精确水平,采用的计量尺度由低级到高级、由粗略到精确,可以将数据分为 4 个层次:定类、定序、定距和定比。

(1) 定类数据:也称为名义级(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。例如,性别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。不论采用上述哪种编码,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:也称为序次级(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,例如,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。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:也称为间距级的数据,它是具有一定单位的实际测量值。两个不同的变量值是存在差异的,可以进行减法运算,也可以进行加法运算。它们都有实际意义。但是,在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。例如,考试成绩,不能说100分是50分的2倍。该级别的变量值只能用数字来表示。

(4) 定比数据:也称为比率数据,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(例如收入用“元”、长度用“米”表示等)的实际测量值。这类数据不仅可以进行加减运算,例如求两个数据的差值,而且可以进行乘除运算。此时,要求其0值不是人为设定的。只有当变量的0值不是人为设定时,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。等级越高,应用范围越广泛;等级越低,应用范围越受限。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。定性数据也称为属性数据或类别数据,它说明的是研究对象的属性特征,不能用数值来表现,主要包括定类数据和定序数据;定量数据也称为数量数据,它说明的是研究对象的数量特征,能用数值来表现,主要包括定距数据和定比数据。对于这两类不同类型的数据,统计学上一般采用不同的方法来处理和分析。对于属性数据通常用各组的频数或频率来进行描述分析,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。

如果一个指标(或变量)取值为定性的数据,则称此指标(或变量)为定性指标(或属性指标、定类指标)。定性数据一般包括如下类别。

- ① 原发性类别数据(自然分类):例如性别数据。
- ② 操纵性类别数据:以人为操作的手段所获取的分类性数据,例如大学的类型。
- ③ 虚拟化类别数据:由其他类型的数据转换成类别形式的数据,例如由连续变项转换来的类别变项。

如果一个指标(或变量)取值为定量的数据,则称此指标(或变量)为定量指标(或数量

指标)。

1.2.3 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,这时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。获得数据的方法有以下两种。

(1) 普查(Census):普查是针对有限总体而言的。普查,即普遍调查,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:也称为部分调查,它是在总体中选择一部分个体进行调查,根据局部数据来了解总体情况。总体中的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(调查那些在个体上反映的、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

实验数据是指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。

1.3 教育研究中的基本问题

1.3.1 研究问题和变量

1. 研究问题

研究问题是针对两个或两个以上变量之间关系的一种陈述。几乎所有的研究都会涉及两个以上的变量。

2. 研究假设

研究假设是对变量之间关系进行预测的描述。研究问题与研究假设类似,只是它们陈述问题的形式不同。研究问题不进行详细的预测,而是通过提问方式进行简要描述。对于任何研究问题,包括两个或多个变量之间是否有相关的问题,都可以用简短的语句进行描述。例如,研究问题“教育能够促进社会进步吗?”涉及两个变量——教育和社会进步。针对该问题所做的研究假设应该是:“教育可以促进社会进步”。

3. 研究设计

研究设计就是对研究问题所提出的研究假设进行验证的步骤和方法,本书方法特指统计方法。