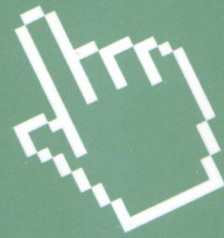




统计软件应用与方法系列丛书

SPSS统计分析



程琮 主编



中国统计出版社
China Statistics Press

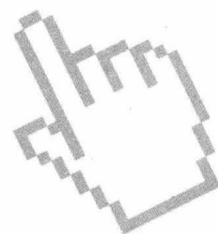


统计软件应用与方法系

SPSS统计分析



程琮 主编



 中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

SPSS 统计分析/程琮主编. —北京:中国统计出版社,2016.9

(统计软件应用与方法系列丛书)

ISBN 978-7-5037-7965-7

I. ①S… II. ①程… III. ①统计分析—软件包—教材 IV. ①C819

中国版本图书馆 CIP 数据核字(2016)第 216032 号

SPSS 统计分析

作 者/程 琮

责任编辑/梁 超

封面设计/黄俊杰 张 冰

出版发行/中国统计出版社

电 话/邮购(010)63376909 书店(010)68783171

网 址/<http://www.zgtjcs.com>

印 刷/河北鑫宏源印刷包装有限公司

经 销/新华书店

开 本/880×1230mm 1/16

字 数/700千字

印 张/23.25

印 数/1-2000册

版 别/2016年9月第1版

版 次/2016年9月第1次印刷

定 价/49.00元

版权所有。未经许可,本书的任何部分不得以任何方式在世界任何地区
以任何文字翻印、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

《SPSS 统计分析》参编人员名单

主 编 程 琮

副主编 刘一志 王素珍 范 华

编 委 程 琮 刘一志 王素珍

范 华 邓通洋 桑金玲

张显鹏 孙校男 吴多文

王 芳 王 佳 邓 阳

王伟伟 刘亚奇

序 言

SPSS 的英文全称为“Statistical Product and Service Solutions”，一般译为“统计产品及服务解决方案”，功能强大，界面友好，应用广泛，涉及社会科学和自然科学的各个学科领域。SPSS 几乎每年更新一个版本，内容不断改进、丰富和完善。

本书主要突出实用性及可操作性。实用性是指以统计学的应用为主，而不去探讨统计学深奥的理论。作为一般科研人员，只需要掌握最基本的统计学知识，即可以使用本书。本书以科研工作的常用统计方法为主线，对统计学基本知识作了简要阐述，对应用实例的结果进行详细分析和注释。可操作性是指对选取的每一种统计学方法在 SPSS 中的实现及其操作步骤都作详细地介绍，并有操作的图形界面。本书精选出 SPSS 最为常用和重要的统计分析方法，每种分析方法都有特定的输入数据方法。本书的一大特色是每个例题都有输入数据模式并对其详加注释，大大提高了读者对 SPSS 统计软件的操作能力。

本书的读者主要是公共卫生及预防医学专业高年级本科生；各专业硕士研究生、博士研究生；中级以上医学科研人员；生物学及农学领域的科研人员；以及其他相关学科的科研人员。

在本书编写过程中，全体编委认真努力工作，顺利完成编书任务。刘一志副教授对全书所有内容，进行了全面的校对核查，并对所有实例的 SPSS 程序进行了运行及核对，付出了艰辛的劳动，花费了大量的时间、精力和心血。在此表示最衷心和最诚挚的感谢。还有许多老师、研究生及科研工作者，对本书的编写给予了很多关心、支持和帮助，提出了许多有益的建议，对他们表示敬意和感谢。由于编者水平所限，本书错误、疏漏及不完善之处在所难免，希望广大读者提出宝贵意见和建议。

作者联系 E-mail: ccta0509@sina.com。

程 琮

2016年9月

于泰山医学院

目 录

第一章 绪 论	1
第一节 概 述	1
第二节 SPSS 软件的安装	1
第三节 SPSS 中的基本概念	2
第四节 SPSS 统计分析的步骤及方法	4
第五节 SPSS 的显示窗口	6
第六节 常用操作术语及符号意义	6
第二章 建立 SPSS 数据文件	10
第一节 概 述	10
第二节 建立 SPSS 数据文件	10
第三节 建立 SPSS 数据文件步骤详解	10
第四节 SPSS 读取 Excel 格式数据文件	14
第五节 SPSS 输出结果的编辑	15
第三章 描述性统计量	18
第一节 频数过程	18
第二节 描述过程	21
第三节 探索过程	22
第四节 比率过程	29
第五节 均数过程	33
第四章 t 检验	38
第一节 单样本 t 检验	38
第二节 独立样本 t 检验	40
第三节 配对样本 t 检验	44
第四节 两样本几何均数 t 检验	45
第五节 变量变换	50
第五章 方差分析(一)	54
第一节 概 述	54
第二节 单因素方差分析	55
第三节 双因素方差分析	63
第四节 拉丁方设计方差分析	70

第六章 方差分析(二)	76
第一节 完全随机设计的协方差分析	76
第二节 随机区组设计的协方差分析	80
第三节 交叉设计的方差分析	84
第四节 析因设计的方差分析	87
第五节 正交设计的方差分析	91
第六节 嵌套设计的方差分析	94
第七节 裂区设计的方差分析	97
第八节 重复测量设计的方差分析	101
第七章 两变量相关与回归	111
第一节 概 述	111
第二节 两变量直线相关	113
第三节 两变量直线回归	114
第四节 秩相关	117
第五节 曲线回归	121
第八章 交叉表与卡方检验	129
第一节 概 述	129
第二节 四格表卡方检验	131
第三节 配对计数资料的检验	138
第四节 行×列表卡方检验	143
第五节 关联性卡方检验	146
第六节 一致性检验	150
第七节 危险度分析	155
第八节 双向有序分类资料的线性趋势检验	174
第九节 诊断试验的评价	177
第十节 ROC 曲线	184
第九章 非参数统计	187
第一节 两个相关样本检验	187
第二节 两个独立样本检验	190
第三节 多个独立样本检验	194
第四节 多个相关样本检验	199
第五节 多样本顺序效应检验	201
第六节 非参数方法的分布检验	204
第七节 卡方拟合优度分布及二项分布检验	209
第八节 游程检验	212

第九节 极端反应检验	215
第十章 统计图	218
第一节 条形图	218
第二节 构成比图	220
第三节 线图	221
第四节 面积图	224
第五节 高低图	225
第六节 控制图	227
第七节 直方图	228
第八节 散点图	229
第九节 P-P 概率图	231
第十节 Q-Q 概率图	233
第十一节 序列图	234
第十二节 误差条图	235
第十三节 箱图	237
第十四节 金字塔图	238
第十一章 生存分析	240
第一节 概述	240
第二节 寿命表法	241
第三节 Kaplan-Meier 法生存分析	253
第四节 Cox 比例风险回归	260
第十二章 多元回归分析	266
第一节 多元线性回归分析	266
第二节 多元线性相关分析	269
第三节 逐步回归分析	272
第四节 Logistic 回归分析	274
第五节 条件 Logistic 回归分析	279
第十三章 聚类分析与判别分析	286
第一节 概述	286
第二节 快速聚类过程	287
第三节 分层聚类过程	290
第四节 判别分析	300
第十四章 因子分析与主成分分析	312
第一节 因子分析	312
第二节 实例分析	314

第三节 主成分分析 320

第十五章 效度与信度分析 327

 第一节 基本概念 327

 第二节 效度分析 330

 第三节 信度分析 333

SPSS 英汉统计学名词及术语对照表 346

参考文献 358

15.1 358

15.2 358

15.3 358

15.4 358

15.5 358

15.6 358

15.7 358

15.8 358

15.9 358

15.10 358

15.11 358

15.12 358

15.13 358

15.14 358

15.15 358

15.16 358

15.17 358

15.18 358

15.19 358

15.20 358

15.21 358

15.22 358

15.23 358

15.24 358

15.25 358

15.26 358

15.27 358

15.28 358

15.29 358

15.30 358

15.31 358

15.32 358

15.33 358

15.34 358

15.35 358

15.36 358

15.37 358

15.38 358

15.39 358

15.40 358

15.41 358

15.42 358

15.43 358

15.44 358

15.45 358

15.46 358

15.47 358

15.48 358

15.49 358

15.50 358

15.51 358

15.52 358

15.53 358

15.54 358

15.55 358

15.56 358

15.57 358

15.58 358

15.59 358

15.60 358

15.61 358

15.62 358

15.63 358

15.64 358

15.65 358

15.66 358

15.67 358

15.68 358

15.69 358

15.70 358

15.71 358

15.72 358

15.73 358

15.74 358

15.75 358

15.76 358

15.77 358

15.78 358

15.79 358

15.80 358

15.81 358

15.82 358

15.83 358

15.84 358

15.85 358

15.86 358

15.87 358

15.88 358

15.89 358

15.90 358

15.91 358

15.92 358

15.93 358

15.94 358

15.95 358

15.96 358

15.97 358

15.98 358

15.99 358

16.00 358

第一章 绪论

第一节 概述

SPSS 英文为“Statistical Package for the Social Sciences”,中文译为“社会科学统计软件包”,它是一种集成化的计算机数据处理应用软件。1968年,美国斯坦福大学 H. Nie 等三位大学生开发了最早的 SPSS 统计软件,并于 1975 年在芝加哥成立了 SPSS 公司,已有数十年的发展历史,全球约有几十万家产品用户,广泛分布于通信、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 是世界非常著名和流行的统计分析软件之一。1994~1998 年,SPSS 公司陆续购并了 SYSTAT 公司、BMDP 公司等,由原来单一统计产品开发转向为企业、教育科研及政府机构提供全面信息统计决策及支持服务。随着 SPSS 服务领域的扩大和深度的增加,SPSS 公司在 2000 年的 Windows 11.0 及以后各版本,将其英文全称更改为“Statistical Product and Service Solutions”,中文译为“统计产品与服务解决方案”,缩写仍为“SPSS”四个字母。2009 年 IBM(International Business Machines Corporation,国际商业机器公司)收购 SPSS 软件公司,2010 年以后,即从 SPSS 19.0 版本开始,软件界面出现 IBM SPSS。

1995 年以后,SPSS 公司与微软公司合作开发 SPSS 界面后,SPSS 界面由 DOS 界面变为 Windows 界面,其界面非常友好,操作以菜单和对对话框的参数设置为主,使用更加简单和方便。在 Windows 下使用 SPSS,免去了许多繁杂的编程工作,不但适合于专业统计人员,也为非专业统计人员提供了更加方便、快捷、简单的操作方法,使 SPSS 的应用群体范围更为广泛。SPSS for Windows 自 1999 年以来,平均每 1~2 年更新一个版本,版本已从 10.0 更新到 22.0 版本以上,各版本的基本功能及操作相同或相近。随着版本的不断更新和提高,SPSS 的每个新版本的功能都有一定的改进和增加。本书以 SPSS 20.0 版本为基础,介绍了 SPSS 在医学及相关科研领域中的主要及常用统计分析方法。本书的特点为:①统计学原理及方法的讲述比较简洁。②SPSS 操作步骤和结果解释较为详尽。本书的使用对象主要是硕士及博士研究生、中级及高级医学科研人员、生物学及社会学领域科研人员以及相关学科的科研人员,主要内容包括:绪论,统计描述,t 检验,方差分析,协方差分析,重复测量资料的方差分析,直线相关与回归分析, χ^2 检验,非参数统计,统计图,生存分析,多元回归及相关分析,聚类分析及判别分析,因子分析与主成分分析,效度与信度分析。

SPSS 是可以用于各个学科领域的统计软件,功能非常强大和丰富。熟练使用 SPSS,需要具备统计学基础知识,并应掌握一定的统计学英文专业词汇,同时还应具备扎实的本学科专业基础知识;此外,还应具备对 Windows 操作系统的基本操作技能,这几个方面的良好结合是学好用好 SPSS 的基础。读者或使用者通过学习本教程,打下一个较好的基础,并可以继续深入开发和应用 SPSS 的许多高级功能。本书所用实例以医学科研资料为主,少量使用了生物学及社会学方面的资料。

第二节 SPSS 软件的安装

SPSS for Windows 软件一般存储在光盘上。各版本安装后所占空间不完全相同。一般地,随着版本的提高,其安装后所占空间更多。SPSS 10.0 版软件安装后约占空间 74 兆,SPSS 13.0 版软件安装后约占空间 280 兆,SPSS 20.0 版软件安装后约占空间 665 兆。SPSS 软件的安装过程比较简单,进入安装画面后,一般按照提示,点击“Yes”或“Next”按钮,即可完成软件安装的全过程。SPSS for Windows 20.0 版安装过程的主要步骤如下。

一、准备工作

启动 Windows 7 操作系统,进入 Windows 7 桌面状态,关闭所有正在运行的程序,包括关闭所有杀毒软件及个人防火墙等软件。

二、启动 SPSS 20.0 安装程序

将 SPSS 20.0 软件光盘放入光驱中,打开 Windows 资源管理器或“我的电脑”,找到“SPSS 20.0”文件夹,双击该文件夹,进入 SPSS 20.0 软件的下一级文件夹目录。再找到子文件“Setup”,这是一个可执行的自

动安装文件。双击该子文件即启动了软件的安装程序,并显示特有的蓝色安装画面及安装进度条。

三、SPSS 20.0 软件安装过程

双击文件“Setup”,显示“Install Shield Wizard”,即安装向导对话框,并显示“Welcome”欢迎说明。

根据安装向导对话框提示,每单击一次“Next”按钮,即可进入下一个对话框及安装内容。

安装路径:系统默认为“C:\Program Files\IBM”。如果想改变安装路径,可以点击“Browse”按钮,选择其他路径。一般情况下,使用默认安装路径。

按照安装向导对话框提示,多次单击“Next”,最后单击“Finish”按钮,完成安装。

重新启动计算机并显示 Windows 桌面。

四、SPSS 快捷图标

重新启动计算机后,在“开始”菜单→“所有程序”子菜单中→显示有“IBM SPSS Statistics”项。点击该项则显示其中一个为“IBM SPSS Statistics”项,这是打开 SPSS 主窗口的快捷方式,单击该快捷方式即可打开一个 SPSS 主窗口。如果想使用方便,可将 SPSS 快捷方式复制到 Windows 桌面上。

第三节 SPSS 中的基本概念

一、资料的分类及测量水平

(一) 医学科研资料

医学科研中常见的资料或数据,通常分为以下几类。

1. 数值变量

数值变量也称为计量资料。是用一定的检测方法或检测手段来测量,并用某种度量衡单位来表达的数据类型。如身高、体重、血压、脉搏等。

2. 计数资料

计数资料也称为无序分类变量。按观察对象的属性或特征来分类,再清点各类别观察单位的个数所得到的资料。包括二项分类变量和多项分类变量。如性别的“男”与“女”、疗效的“治愈”与“未治愈”、人群的职业分类等。

3. 等级资料

等级资料也称为有序分类变量。按观察对象的属性或特征的程度或等级来分类,再清点各等级类别观察单位的个数所得到的资料,也称为等级资料。如人群的文化程度可分为文盲与半文盲、小学、中学、大学、研究生等;疗效可分为治愈、好转、显效、无效、死亡等。

4. 圆形分布资料

圆形分布资料指具有循环性或周期性的资料。常见的主要有角度资料或时间的周期性重复资料。如出生时间或死亡时间,某病在一年 12 个月中的发病时间等。

(二) 一般科研资料

SPSS 中对资料或数据的分类与医学领域中对资料的分类有所不同,尤其是名称上不同,但实质内容却是相同的。SPSS 对数据的分类、定义及测量水平如下。

1. 名义测量(Nominal Measurement)

名义测量也称为标称测量或定类数据,是各类测量水平中级别最低的一类数据类型,相当于分类资料中的无序分类变量,即计数资料。数值之间没有内在的联系,数值仅仅是作为某个变量的一种符号或标称。如男与女之间,北京与上海之间,城市与农村之间,并无次序的前后之分,也不存在等级的上下之别。

2. 次序测量(Ordinal Measurement)

次序测量也称为顺序测量或定序数据。其特点是,所有变量值可按一定标准分类排序,相当于分类变量中的有序分类资料,即等级资料。例如,病人的病情可分为轻度、中度和重度;对病人的治疗结果可分为无效、显效、好转、治愈;对某项化验结果可分为-、±、+、++、+++等。这些资料均可以分类排序,且不可随意颠倒次序。

3. 区间测量(Interval Measurement)

区间测量也称为定距数据。其特点是,除能对各种数据排序分类外,还可以测量出两种数据之间的距

离。如温度计是用来测量温度的,温度计中可有+10度,0度或-10度。但0度并不表示没有温度,即0度并不是真正意义上的“没有温度”,而只不过是所有温度值中的一个普通数值。0度是人为定义的,在摄氏与华氏温度计量中,0度的位置是不同的。

4. 比例测量(Ratio Measurement)

比例测量也称为定比数据。相当于计量资料。为测量水平中级别最高的一类数据类型。其特点是,该测量除具有次序测量和区间测量的各种特性之外,还具有一个真正意义上的零点,并以零点为标志,对两个数据进行测量与比较,从中测得两者的比例程度。例如,血压计用来测量人体的血压。它在接近零点时表示生命所存在的危险性;而处在零点时,则生命的存在有了质的变化。两物体的测量距离为零时,则表示两物体之间没有距离。比例测量资料可以用于任何数学换算与运算。这里强调,比例测量中,零点的存在意义极其重要。

以上数据的四种测量水平的级别有高低之分。由低级到高级依次排列顺序为名义测量、次序测量、区间测量和比例测量。

二、综合统计量

根据数据的测量水平,应该选择合适的统计量来反映事物的特征。综合统计量主要有以下几类:集中趋势统计量;离散趋势统计量;形状测量统计量。

(一)集中趋势统计量

1. 均数(Mean, \bar{X})

均数是指所有观察值的平均值或平均水平,也称为算术平均数,如平均血压、平均年龄、平均身高等。均数只用于区间测量数据以上级别数据的测量,如区间测量及比例测量数据。不能用于名义数据和次序数据的测量,即不能用于计数资料及等级资料。均数的大小易受极端数值大小变化的影响。均数主要用于正态分布或对称分布的数据。

2. 几何均数(Geometric Mean, G)

几何均数是一组数据乘积的开方。主要用于数据为对数正态分布,或呈等比、等倍关系的数据。

3. 中位数(Median, M)

中位数是指一组数值按增序或降序排列后,位居中间位置的数值。其特点是,观察值个数的一半落在中位数之上,观察值个数的另一半则落在中位数之下,而且中位数不易受极端数值变化的影响。中位数可用于各种分布类型数据。主要用于非正态分布的数据。

4. 众数(Mode, M_0)

众数是一组数值中频数最多处的数值。通常只用于计量资料。众数的特点是可以不存在,也可以不止一个。众数可用于各种分布类型数据。

5. 调和均数(Harmonic Mean, H)

调和均数是一组数据倒数的平均数的倒数。主要用于偏态比较严重或非常偏态的数据。

6. 调整均数(Trimmed Mean, M_{tr})

调整均数是将一组数据排序后,两端各去除 $x\%$ 的数据,再计算剩余数据的算术平均数。SPSS 默认计算 5% 调整均数。

以上六个指标主要反映了一组数据或一个群体集中趋势的综合水平。如果数据接近正态分布,则六个指标的计算结果比较接近。若数据为比较典型的正偏态分布,则六个指标的计算结果一般不会相同。计算结果由小到大依次为:调和均数、几何均数、众数、中位数、调整均数和均数。

(二)离散趋势统计量

1. 极差(Range, R)

极差也称为全距,是一组数值中最大值与最小值之差。其特点是:计算简单,使用方便,容易理解,但不太稳定,易受极端值影响。极差只反映数据两端的变异,故反映变异不全面。

2. 方差(Variance, S^2, σ^2)

方差是一组数据中每个变量值与其均数之差平方和的均值。它的特点是,能够全面反映一组数据的变异状况,指标稳定,用途广泛,在统计学中是极为重要的变异指标。但其概念不易理解,计算比较复杂,数据

的度量衡单位也与原数据不同,是原数据单位的平方。

3. 标准差(Standard Deviation, S, SD)

标准差是方差的开平方。意义与方差相同,但其数据的度量衡单位则与原数据相同。通常作为反映一组数据离散趋势的标准指标,其应用广泛,在科研中具有极为重要的意义。标准差常与均数结合表达一组数据的集中与离散趋势,即 $\bar{X} \pm S$ 。

4. 变异系数(Coefficient of Variation, CV)

变异系数主要用来反映度量衡单位不同或均数相差很大时的两组或多组数据的离散趋势。变异系数用百分数表示,本身是无量纲指标。变异系数一般不宜超过 20%~30%。变异系数过大,反映该指标不太稳定。

5. 四分位数间距(Interquartile Range, Q)

四分位数间距是指 $P_{75} \sim P_{25}$ 之间的距离,是一组数据排序后,中间 50% 变量值的极差。可用来反映各种非正态分布的离散趋势。它比极差稳定,但也只反映中间 50% 数据的极差,反映离散趋势不全面,而且计算稍显复杂,不易理解。该指标常与中位数结合表达一组数据的集中与离散趋势,即 $M \pm Q$ 。

6. 其他分位数

一组数据按升序排列后,分为相等的四个部分的数称为四分位数,分为十个部分的数称为十分位数,分为一百个相等部分的数,称为百分位数。可以根据需要,计算四分位数、十分位数、百分位数等指标。

(三) 形状测量的统计量

1. 对称分布(Symmetric Distribution)

对称分布主要是指正态分布(Normal Distribution)。在医学及生物学领域,生物体以及人体的正常生理生化指标多为正态分布。例如,人体的身高、体重、血压、脉搏等生理指标。

2. 非对称分布(Nonsymmetric Distribution)

(1) 偏态分布(Skewness): 当分布的高峰偏左侧,长尾巴拖向右侧时,称为正偏态(Positive Skewness)分布。机体的许多异常指标为正偏态分布,如尿铅含量、尿汞含量等重金属含量的分布,疾病潜伏期的分布也是正偏态分布。当分布的高峰在右侧,长尾巴在左侧时,称为负偏态(Negative Skewness)分布,该分布在医学生物学领域应用较少。

偏态分布的量度用偏度系数(Coefficient of Skewness)来表示。样本偏度系数用 g_1 表示,总体偏度系数用 γ_1 表示。理论上, $\gamma_1 = 0$ 为对称分布, $\gamma_1 > 0$ 为正偏态分布, $\gamma_1 < 0$ 为负偏态分布。偏度系数的标准误用 σ_{g_1} 表示。

(2) 峰态分布(Kurtosis)

① 正态峰(Mesokurtosis): 指标准正态分布的峰态状态。

② 尖峭峰(Leptokurtosis): 也称为尖峰态,其峰顶比正态分布更陡峭更集中。

③ 平阔峰(Platykurtosis): 也称为低峰态,其峰顶比正态峰低,峰顶更平阔且分散。

峰态分布的量度用峰度系数(Coefficient of Kurtosis)来表示。样本峰度系数用 g_2 表示,总体峰度系数用 γ_2 表示。 $\gamma_2 = 0$ 为正态峰, $\gamma_2 > 0$ 为尖峭峰, $\gamma_2 < 0$ 为平阔峰。峰度系数的标准误用 σ_{g_2} 表示。

(四) 数据特征的综合描述

1. 若数据为正态分布、近似正态分布或对称分布,一般用均数 \pm 标准差 ($\bar{X} \pm S$) 表示,用来反映一组数据的集中趋势与离散趋势的综合特征。这是科研工作或科研论文中最为常见的表达方式。

2. 若数据为典型的偏态分布,一般用中位数 \pm 四分位数间距 ($M \pm Q$) 表示,用来反映非正态分布的一组数据的集中趋势与离散趋势的综合特征。

第四节 SPSS 统计分析的步骤及方法

SPSS 统计分析步骤主要有以下几步:

- (1) 建立数据文件;
- (2) 选择统计分析过程;
- (3) 选择统计分析变量并进行参数设置;

- (4) 运行统计分析过程;
- (5) 显示统计分析过程的结果;
- (6) 对分析结果做出合理的解释与讨论。

(一) 建立数据文件

这是 SPSS 统计分析的第一步,也是关键性和至关重要的一步。建立数据文件是指按照 SPSS 统计分析过程的格式要求输入待分析的数据,并建立文件名存盘保存,以备分析使用。输入数据前,应先对数据进行整理归纳,然后仔细将数据输入电脑。在输入数据的过程中,应每隔 5~10 分钟将数据存盘一次,以免因意外或操作不当而丢失数据。SPSS 中不同的统计分析过程对输入数据的输入格式有不同的要求。此点要特别注意。如果输入数据的输入格式不对,则无法进行进一步的操作与分析。

【注意】不能按照个人意愿、个人习惯或 SPSS 不承认的格式输入数据,以免造成时间、人力和物力的浪费。

(二) 选择统计分析过程

传统意义上的统计分析方法,如 t 检验, χ^2 检验,方差分析等,在 SPSS 中称为一个过程,即“Processor”。一个过程就是指一种具体的统计分析方法。SPSS 主窗口的状态栏中,经常会显示“IBM SPSS Statistics Processor is ready”,意思是:“SPSS 统计分析过程准备就绪”,正等待使用者去运行使用。SPSS 统计分析过程相当丰富和庞大,几乎涵盖了常用统计教科书及参考书的统计方法。所以,要想学好用好并掌握 SPSS,需要使用者不断努力去钻研和开发。确定一种 SPSS 的统计分析过程,可由“Analyze”或“Graphs”菜单栏中去选择。“Analyze”是最常用的选择菜单。

(三) 选择统计分析变量

SPSS 中建立的数据文件,都是按照人为定义的变量而输入数据的。数据文件中可定义一个或多个变量。根据统计分析的需要,选择好一个过程并打开该过程,则会显示出一个对话框(Dialog Box)。该对话框在本书中称为主对话框。出现主对话框后,要求选择分析的变量。选择分析变量后,还需进一步设置对话框中的各项参数。在对话框中设置参数,有的分析过程要求比较简单,有的则比较复杂。有的参数是 SPSS 默认的,有的则必须人为设置。如果参数设置不正确,常常无法关闭或运行对话框,或无法进行对其他参数的设置,因此,也就无法正常运行一个统计分析过程。参数设置中,常常会有多个对话框嵌套在一起,使用者应尽可能理解清楚各个参数的意义。设置参数的原则应该是:在充分表达预期分析结果的前提下,能少设置则少设置。SPSS 中常用统计方法的参数设置,在一般情况下比较容易。对于多因素分析等复杂的统计分析方法,参数设置则比较复杂,对于初学者可能会有一定困难。如果不能正确设置参数,可能导致显示的分析结果无法理解或难于解释。SPSS 中的参数设置,需要使用者有一定的、甚至是相当深厚的统计学功底。

(四) 运行统计分析过程

当各个对话框中的参数设置正确无误后,通过点击“Continue”按钮并返回主对话框。最后,在主对话框中单击“OK”按钮,即开始运行一个统计分析过程。运行一个过程所花费的时间,一般可有数秒钟,数十秒钟,数分钟,个别为数小时。虽然 SPSS 软件所占硬盘空间比较庞大,但电脑的运算速度也越来越快。运行一个统计分析过程所花费的等待时间与建立数据文件时输入数据所花费的时间相比,是微不足道的。例如,按两个人每人每天工作 8 小时,两人合作每 3 秒钟输入一个数据,一天可以输入 9600 个数据。如取近似值,按两人每天可输入 1 万个数据计算,工作一个月 30 天可以输入 30 万个数据。而选择一个过程并运行这些数据的部分,大部分或全部并显示出分析结果,可能只需要数秒钟,数分钟或数小时。就普通常见的医学科研数据而言,运行一个过程的时间都是非常短暂的。因此,在运行一个过程以前,对原始数据进行归纳、整理和总结,并按照 SPSS 输入数据的格式要求,将待分析数据认真仔细地输入电脑,对每一个科研人员都是非常关键而重要的,决不可以马马虎虎。

一般来说,为了输入数据的方便和快捷,可先将数据输入 Excel 软件中,进行检查、核对、编辑,并达到 SPSS 的输入数据格式要求,保存 Excel 格式的数据文件。然后再转到 SPSS 软件中。这是一种比较省时省力的数据输入方法,也便于携带和保存。

(五) 显示统计分析过程的结果

运行一个过程结束后,会在“IBM SPSS Statistics Viewer”窗口显示该过程的分析结果。对于输出的结果,可以用多种格式的文件存盘保存以备查看,也可以打印出来进行仔细分析和判断。输出结果的显示方

式和内容是根据参数设置的不同而不同的。

(六)对分析结果作出合理的解释与讨论。

SPSS 只对输入的数据进行计算,并输出计算结果。但 SPSS 对输出的计算结果并不会分析。科研人员应根据本专业的知识和特点,对 SPSS 的计算结果进行解释和讨论,这是一项非常重要而专业化的工作,没有良好的专业知识和扎实的统计学功底,就无法对结果进行分析和讨论。

第五节 SPSS 的显示窗口

SPSS 主要显示以下三种窗口:

(1)IBM SPSS Statistics Data Editor:SPSS 数据编辑窗口,简称 SPSS Data Editor 窗口或 Data Editor 窗口。

(2)IBM SPSS Statistics Syntax Editor:SPSS 语句编辑窗口,简称 SPSS Syntax Editor 窗口或 Syntax Editor 窗口。

(3)IBM SPSS Statistics Viewer:SPSS 结果输出窗口,简称 SPSS Viewer 窗口或 Viewer 窗口。

一、IBM SPSS Statistics Data Editor 窗口

当双击 SPSS 图标时,SPSS 默认显示的主窗口即为该窗口,称之为“SPSS 数据编辑窗口”。在此窗口中的主要工作是建立数据文件,并可对数据文件进行各种编辑工作,包括定义变量名、定义变量的格式、定义输入数据的格式,对数据进行剪切、复制、粘贴,修改数据并可保存已建立的数据文件。也可打开新的数据文件或打开已建立的数据文件。建立的数据文件保存格式为:“*.sav”。其中,“*”号代表任意命名的文件名,“.sav”表示数据文件的扩展名。在“SPSS Data Editor”窗口中,可以通过打开和使用菜单、子菜单,使用对话框并设置参数和命令,对数据进行各种统计分析。主要的统计分析过程均可以在此窗口中完成。

二、IBM SPSS Statistics Syntax Editor 窗口

该窗口称为“SPSS 语句编辑窗口”。主要用来将 SPSS 提供的各种语句组织起来编写 SPSS 的统计分析程序,这主要用于专业统计人员或计算机编程人员,本书不涉及此部分内容,故不对此窗口做更进一步的介绍。

三、IBM SPSS Statistics Viewer 窗口

该窗口称为“SPSS 结果输出窗口”,用来显示统计分析结果。只要运行某个统计分析过程,就会显示该窗口。所有的显示结果均可用某种格式的文件形式加以保存,以备今后查阅分析之用。分析结果的文件名默认的保存格式为:“*.spv”其中,“*”代表输出结果的文件名,“.spv”为输出结果文件名的扩展名。其他常见的保存格式或文件类型有:Word 格式、excel 格式、html 格式、text 格式、power point 格式等。为了使用及携带方便,一般常将 SPSS 结果转换为 Word 格式或 excel 格式文件保存。

第六节 常用操作术语及符号意义

本书的全部操作过程及分析结果均用文字、符号、表格或图形界面加以描述或说明。描述尽量简洁,以方便读者使用。

一、主窗口

指 SPSS 默认的数据编辑窗口,即 SPSS Data Editor 窗口。双击 SPSS 图标即可打开该窗口。

二、菜单栏

指 SPSS 主窗口中的各菜单栏。主要有下列菜单:File 菜单;Edit 菜单;View 菜单;Data 菜单;Transform 菜单;Analyze 菜单;Direct Marketing 菜单;Graphs 菜单;Utilities 菜单;Add-ons 菜单;Window 菜单;Help 菜单。见图 1-1。



图 1-1 SPSS 主窗口菜单栏

三、常用工具栏

指 SPSS 主窗口打开后,系统默认的工具栏,共有 19 个常用工具及其相应的图标(见图 1-2)。图 1-2 只显示了 15 个图标。



图 1-2 SPSS 主窗口常用工具栏

四、变量列

在 SPSS 主窗口中,系统默认显示的窗口是“Data View”标签状态下的窗口。变量未加定义时,系统默认以“var”作为变量名并在整个窗口中共显示 9 列。还可通过窗口下的横向滚动条或左右横向箭头来调整变量列的显示状况。见图 1-3。

五、观察单位显示行

也称“Case”显示行。SPSS 中自动为其编号为 1,2,3,……。在主窗口中默认显示的行共 23 行。可用窗口右边的上下滚动条或上下箭头来调整“Case”行的显示状况。见图 1-3。

六、“Data View”标签

也称为数据查看标签,在主窗口下方。在该标签状态下,可以对数据进行各种编辑操作。见图 1-4。

	var	var	var	var
1				
2				
3				

图 1-3 主窗口变量列及行

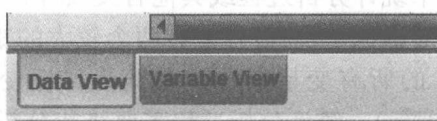


图 1-4 SPSS 主窗口下方的两个标签

七、“Variable View”标签

也称为变量查看标签,在“Data View”标签的右边。单击该标签则进入该标签状态。在此状态下,可以对变量进行各种定义。见图 1-5。

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1										
2										
3										

图 1-5 SPSS 主窗口下方 Variable View 标签窗口

定义变量属性的各种类型的标签条或称为按钮,显示在主窗口常用工具栏的下方。由左至右依次排列,其意义如下。

(一)Name:定义变量名称

一般用英文小写字母表示,也可以用英文小写字母加上数字来表示。也可以用中文汉字表示,但使用有时不太方便。

(二)Type:定义变量类型

可以定义 8 种变量类型。常用者只有两种,即“Numeric”型,为数值型变量;“String”型,为字符型变量。最常用为“Numeric”数值型变量。

(三)Width:定义输入测量值或字符的个数

系统默认为 8 位,包括小数点在内。如可输入数据 12345.67 共 8 位。也可以根据需要进行调整到所需的字符个数。

(四)Decimals:定义输入数据的小数位

系统默认为 2 位小数。可以通过单击单元格中的上下小箭头改变小数位数。

(五)Label:定义变量名称的附加说明

如果变量名用 x1, x2 来表示,并不能说明该变量的真实意义。在 Label 中可给以说明。如定义变量名为 x1,在 Label 中说明为“身高”,x2 可以说明为“体重”。

(六)Values:定义变量测量值的说明

例如定义变量名为“sex”,在“Value”标签条或按钮下的单元格中打开一个对话框并给以说明:“0”为女性,“1”为男性。

(七)Missing:定义缺失值情况

系统默认为“None”，即没有缺失值。当出现变量中有缺失值情况时，系统自动将该缺失值剔除。不对该缺失值进行运算。

(八)Columns:表示变量“var”条的宽度

系统默认为8位数字宽度。可以通过单击“Columns”条下相应单元格中的上下箭头来调整宽度。最小宽度只能为1位数的宽度。

(九)Align:输入的测量值的对齐方式

系统默认为右对齐(Right)。还可选择左对齐(Left)或居中(Center)对齐方式。

(十)Measure:测量值类型

系统默认为比例尺度即“Scale”型，还可选择次序尺度(Ordinal)型，也称为等级尺度，或选择名义尺度(Nominal)型，也称为分类变量型。

八、主对话框

指运行一个统计分析过程或其他有关项目时，显示的第一个对话框。该对话框主要进行相关内容的参数设置。在主对话框中，一般有左右两个较大的矩形框。左框称为源变量框(Source Variables)，显示已建立的数据文件的所有变量名称。右框为目标变量框(Target Variables)。统计分析过程需要将待分析的变量从左框选入到右框中。选择方法是：先单击选定左框中1个或多个变量，再单击左右框之间的方向箭头，左框中的变量即被选入到右框中。如果要取消右框中已选入的变量，可以先单击选定右框中变量，再单击左右框之间的方向箭头，则右框中的选定变量又重新回到左框。主对话框中间的方向箭头是可以向左右方向变换的。另外，左框中的变量还可以选入到右侧或下方的其他矩形框或文本框中，操作与上述相同或相似。例如，描述(Descriptives)过程主对话框，见图1-6。

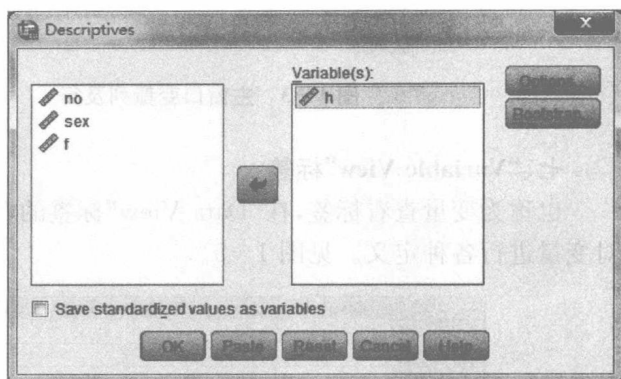


图 1-6 描述(Descriptives)过程主对话框

九、按钮

在主对话框或其他对话框中，常有许多矩形按钮。按钮上标有英文名称，显示该按钮的意义，见图1-7。

描述过程主对话框常用按钮意义如下。

(一)“Continue”按钮

单击主对话框中某个按钮就可以打开一个新的对话框。新的对话框中又有一些设置项目和按钮；其中，一般都有一个标有英文字母“Continue”的按钮。设置好参数后，只有单击“Continue”按钮才能返回主对话框。如果“Continue”按钮呈现灰色，说明尚未设置好参数，此时，无法返回主对话框，必须重新设置参数。主对话框中的其他按钮的操作与上述相同。

(二)“OK”按钮

在主对话框中有一个“OK”按钮。单击“OK”按钮表示运行某个统计分析过程。当主对话框中所有相关参数及按钮都设置完成后，才能最后单击“OK”按钮。一旦单击“OK”按钮，则开始运行某个统计分析过程，并显示分析过程的结果。如果“OK”按钮显示为灰色，则说明某些参数尚未按要求设置好，无法运行该统计分析过程，必须重新设置参数。

(三)“Paste”按钮

该按钮用来组织语句编写程序。

(四)“Reset”按钮

重新设置参数按钮。只要单击该按钮，以前设置的参数全部作废，必须重新设置参数才能运行某个统

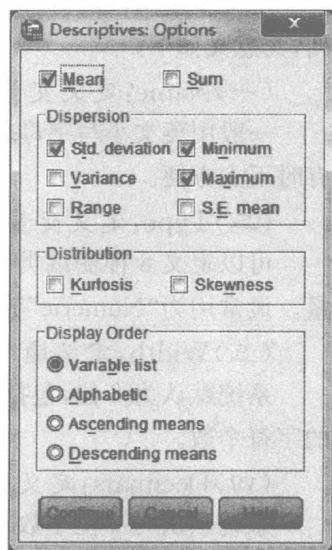


图 1-7 描述过程 Options 对话框