



中国科学院教材建设专家委员会规划教材
全国高等院校医学实验教学规划教材

医药数理统计上机与实践教程

主 编 宁 刚 王志超



科学出版社

中国科学院教材建设专家委员会规划教材

全国高等院校医学实验教学规划教材

医药数理统计上机与实践教程

主 编 宁 刚 王志超

副主编 涂昱超 郭志浩 董健卫 闫志来

编 委 (以姓氏汉语拼音排序)

崔慧萍 董健卫 郭志浩 胡炯健

霍 平 孟 盼 宁 刚 唐小娅

田冬梅 涂昱超 王秀凤 王志超

闫志来

科 学 出 版 社

北 京

内 容 简 介

本书是根据医药院校学生的知识结构特点,结合编者多年的医药数理统计实验教学经验编写而成。本书介绍了基础的 SPSS 统计分析及部分高级的多元统计内容。在上机案例的选取上,我们力求贴近医药类专业知识的背景,同时与《医药数理统计》理论教材相匹配。在章节内容编排上,先介绍基本原理,然后提出案例,上机实验,得到输出结果,最后给出合理解释。这样方便学生阅读,能快速掌握相关知识。

本书可供高等医药院校本科学生使用,也可供参加数学建模比赛的学生参考。各校可根据专业、学时的不同选择其中部分作为实验教学内容。

图书在版编目(CIP)数据

医药数理统计上机与实践教程 / 宁刚, 王志超主编. —北京: 科学出版社, 2016.6

中国科学院教材建设专家委员会规划教材·全国高等院校医学实验教学规划教材

ISBN 978-7-03-048889-3

I. ①医… II. ①宁… ②王… III. ①医用数学-数理统计-医学院校-教材 IV. ①R311

中国版本图书馆 CIP 数据核字(2016)第 136595 号

责任编辑: 赵炜炜 胡治国 / 责任校对: 刘亚琦

责任印制: 赵 博 / 封面设计: 陈 敬

版权所有, 违者必究。未经本社许可, 数字图书馆不得使用

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

大厂书文印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 6 月第 一 版 开本: 720×1000 1/16

2016 年 6 月第 一 次印刷 印张: 11

字数: 212 000

定价: 32.00 元

(如有印装质量问题, 我社负责调换)

前 言

本书是在统计软件 SPSS 19.0 环境下,编者根据多年的医药院校数理统计上机教学中的一些体会和经验编写的,可作为《医药数理统计》课程的配套实验教材。

本书包括基础的统计分析有:描述性统计、正态性检验、 t 检验、方差分析、总体分布的 χ^2 检验、交叉列联表分析、回归分析、综合实验,以供本科生数理统计实验教学使用。本教材还增加了一些高级的多元统计内容,如生存分析、聚类分析、判别分析、因子分析和主成分分析,以供学有余力的同学自学和有志于参加数学建模比赛的学生参考。希望每一位同学,在上机实验前详细、认真地阅读实验指导,以便能更快、更好地掌握 SPSS 处理和分析数据的强有力的统计工具。相信通过同学们的努力,一定能熟练使用这一有用的软件。事实证明,电脑高手不是老师完全能教出来的,是玩出来的。如果你想成为一位高手的话,那就赶紧动手大胆地使用它吧!

值得一提的是,统计软件包的使用虽然给我们的学习带来极大的便利,但是仍然不能忽视统计理论知识的学习。有一项最新研究结果显示,能够熟练地操作计算机的学生在学习方面不一定优秀,拥有计算机的学生学习的东西往往较少。因此,希望同学们要充分重视统计理论知识的学习,因为没有对统计基本原理和统计方法的熟练掌握,就根本谈不上正确使用统计软件,甚至还可能误用、滥用和错用统计方法。教学实践证实,那些统计概念不清的同学在综合运用软件时会感到困惑和不知所措,例如,不能合理设置变量,不能正确选择统计方法,对输出结果也不能正确解释等。

本书由广东药科大学、广州中医药大学的老师合力完成。其中闫志来老师来自广州中医药大学,其余老师均来自广东药科大学。教材编写分工如下:宁刚老师负责统稿工作及编写第三章;崔慧萍老师负责编写第一章;田冬梅老师负责编写第二章;郭志浩老师负责编写第四章;涂昱超、孟盼老师负责编写第五章;闫志来老师负责编写第六章;霍平老师负责编写第七章;王志超、胡炯健老师负责编写第八章;董健卫老师

负责编写第九章；唐小娅老师负责编写第十章；王秀凤老师负责编写第十一章；王志超老师负责编写第十二章。

本书在编写时得到上级领导的支持和不少同事的热情帮助，并参考了许多同行的书籍，在此谨致以最诚挚的感谢。限于编者的水平，书中难免有不足之处，恳请读者提出宝贵意见，以便改进。

编 者

2016年3月

目 录

第一章 SPSS 概述和实验指导使用说明	1
第二章 频数分析	11
第三章 正态性检验	21
第四章 t 检验	26
第一节 单样本 t 检验	26
第二节 配对样本 t 检验	30
第三节 独立样本 t 检验	35
第五章 方差分析	42
第一节 方差分析的概念及原理	42
第二节 单因素方差分析	44
第三节 多因素方差分析	55
第六章 总体分布的 χ^2 检验	66
第七章 交叉列联表分析	74
第八章 回归分析	85
第一节 一元线性回归分析	85
第二节 一元拟线性回归分析	90
第三节 多元线性回归分析	93
第四节 二元逻辑回归	99
第九章 生存分析	111
第十章 聚类分析	126
第十一章 因子分析	139
第一节 基本原理	139
第二节 基本分析过程	140
第三节 案例分析	145
第十二章 综合实验	157
参考文献	169



第一章 SPSS 概述和实验指导使用说明

SPSS(statistical package for social science 的缩写)社会科学统计软件包, 是世界上最优秀的统计分析软件包之一。1968 年, 三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。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 science solutions, 即统计产品与服务解决方案。

SPSS 已广泛应用于自然科学和社会科学的各个领域, 只要有需要对各种数据如数值型、字符型、逻辑型等进行统计分析的地方, 就有 SPSS 的用武之地。近年来该软件被广泛应用于我国政府部门、医疗卫生、体育、经济等领域的信息管理和决策分析工作。同时, SPSS 软件的使用业已成为许多高校学生的必备技能。

SPSS 19.0 版本在以往版本的基础上, 大大加强了编制报告能力、统计分析功能和编程扩展能力, 同时还彻底地解决了对中文字符的兼容性问题。本书主要以 SPSS 19.0 版本为例, 通过实例演示常用的统计方法, 如频数分析表分析, 正态性检验, 配对样本 t 检验, 独立样本 t 检验, 方差分析, 总体分布的 χ^2 检验, 交叉列联表分析, 生存分析, 聚类分析与判别分析, 因子分析和主成分分析, 回归分析及综合实验等内容, 使读者掌握基本的 SPSS 统计软件在实际应用中的功能及方法。

一、SPSS 19.0 的特点

常见的统计软件有 SAS, SPSS, MINITAB, EXCEL 等。这些统计软件各有所侧重, SPSS 以其界面友好、功能强大、图形精美和易学易用等优点受到了社会各界统计分析人员的喜爱。它包括了几乎全部尖端的统计分析方法, 具备完善的数据定义、操作管理和开放的数据接口以及灵活而美观的统计图表制作。SPSS 19.0 加入了一些新的特性和功能, 包括广义线性混合模型, 一个统计网页入口 portal, 自动线性模型, 增加了一些直复营销功能, 并跟 IBM 协作和部署服务系统进行整合。改进了一些统计模块的功能, 改善自动操作, 新增了分析和报告, 改善的结构和技术及交互式模型查看器。具体细节读者可参看 SPSS 专业指导书。

1. 硬件环境要求 SPSS 19.0 对计算机硬件的要求不高, 但它运算设计的数据量比较多, 因此 SPSS 19.0 软件对计算机硬件的基本要求如下: 1GHz 以上的 Intel 或 AMD 处理器, 至少 512M 的内存, 官方推荐 1G 内存及以上, 至少 800M 的硬盘剩余空间, 光盘驱动器(这是光盘安装的要求), Windows2000/XP/7/Vista 兼容的图形适配卡及以上, 支持 TCP/IP 网络协议的网络适配卡, 主要用于软件更新及各类相关服务。

2. 软件环境要求 SPSS 19.0 要求的操作系统是 Windows XP(32 位)、Vista(32 位或 64 位)或 Windows 7(32 位或 64 位), IE7.0 以上版本, Adobe Reader 阅读器。

二、SPSS 19.0 的安装、卸载、启动和退出

SPSS 19.0 的安装、卸载、启动与退出和一般的 Windows 应用软件基本一样，非常简便。

(一) SPSS 19.0 的安装与卸载

1. SPSS 19.0 的安装 打开安装包，按照安装提示，即可完成软件的安装。
2. SPSS 19.0 的卸载 在 Windows 桌面，选择“开始”→“设置”→“控制面板”→“添加或删除硬件”，在程序列表中选择 IBM SPSS Statistics 19.0，单击“卸载”，删除 SPSS 软件。

(二) SPSS 19.0 的运行模式

SPSS 19.0 主要有三种运行模式：

1. 批处理模式 这种模式把已经编写好的程序存为一个文件，提交给 Windows 桌面上“开始”→“SPSS for Windows”→“Production Mode Facility”程序运行。
 2. 完全窗口菜单运行模式 这种模式通过选择窗口菜单和对话框完成各种操作。用户无须学会编程，简单易用。
 3. 程序运行模式 这种模式是在语句窗口中直接运行编写好的程序或者在脚本窗口中运行脚本程序的一种运行方式。这种模式要求掌握 SPSS 的语句或脚本语言。
- 本实验指导书为初学者提供入门实验教程，采用“完全窗口菜单运行模式”。

(三) SPSS 19.0 的启动与退出

1. SPSS 19.0 的启动 在 Windows 桌面，选择“开始”→“程序”→“IBM SPSS Statistics”→“IBM SPSS Statistics19”，即开始运行 SPSS 19.0。打开 SPSS 19.0 之后，弹出 SPSS 的文件对话框，有六个选项，如图 1-1 所示。
2. SPSS 19.0 的退出 SPSS 19.0 的退出方法与其他 Windows 应用程序相同，有三种常用的退出方法：
 - (1) 选择菜单中的“退出”命令，即可退出。
 - (2) 直接单击 SPSS 窗口右上角的，即可退出。
 - (3) 对着 SPSS 窗口左上角的窗口控制菜单图标双击，即可关闭 SPSS 窗口。

三、SPSS 19.0 相关设置

在完成 SPSS 19.0 的安装之后，可以通过“选项”对话框设置系统的默认值和初始状态。从 SPSS 19.0 主菜单栏的“编辑”下拉菜单“选项”，进入“选项”对话框。共有 11 个参数选项卡，分别是：常规、查看器、数据、货币、输出标签、图表、枢轴表、文件位置、脚本、多重归因和语法编辑器，如图 1-2 所示。

四、SPSS 19.0 主要窗口介绍

SPSS 软件运行过程中会出现多个界面，各个界面用处不同。其中，最主要的

界面有四个：数据编辑窗口、结果输出窗口、语句窗口和脚本编辑窗口。

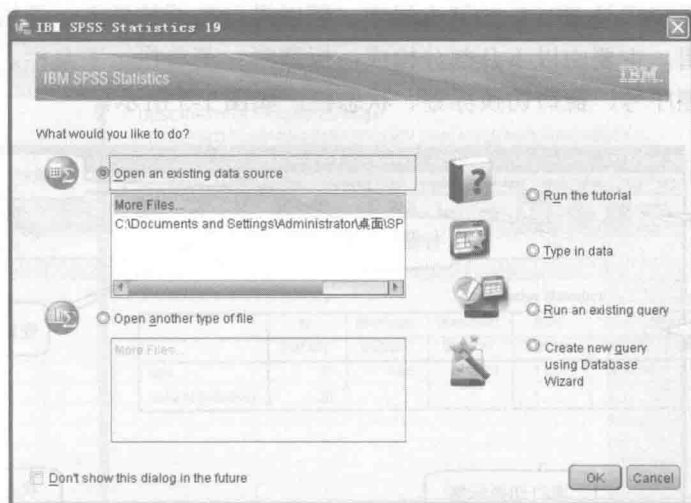


图 1-1 SPSS 文件对话框

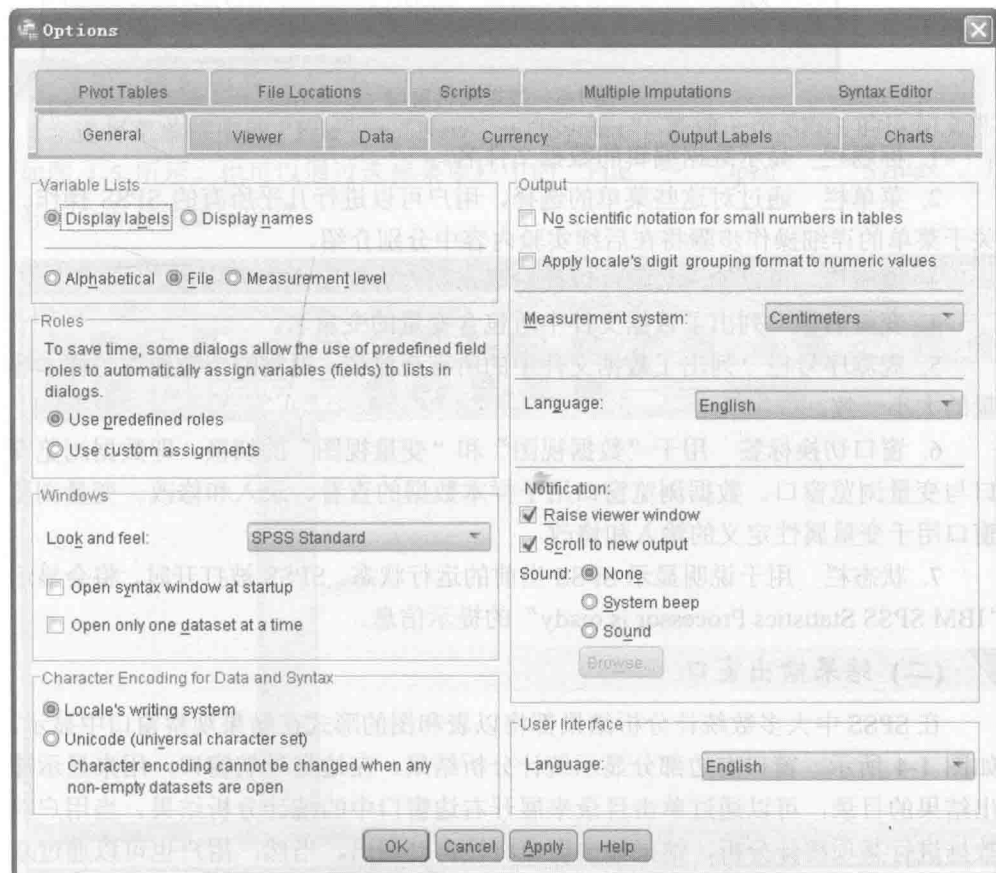


图 1-2 常规选项卡

(一) 数据编辑窗口

数据编辑窗口是 SPSS 的基本界面, 可以进行数据的录入、编辑及变量属性的定义和编辑, 主要由以下几部分构成: 标题栏、菜单栏、工具栏、编辑栏、变量名栏、观测序号、窗口切换标签、状态栏。如图 1-3 所示。

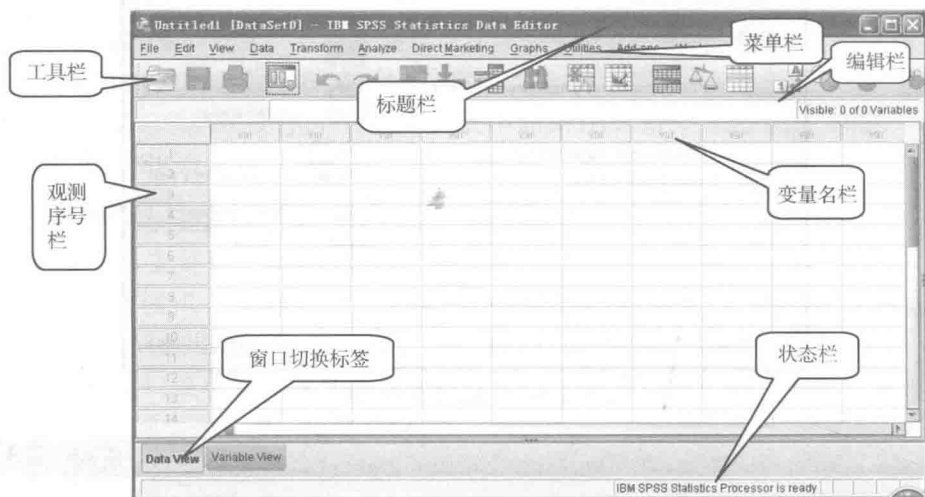


图 1-3 数据编辑窗口

1. 标题栏 显示数据编辑的数据文件名。
2. 菜单栏 通过对这些菜单的选择, 用户可以进行几乎所有的 SPSS 操作。关于菜单的详细操作步骤将在后续实验内容中分别介绍。
3. 编辑栏 可以输入数据, 以使它显示在内容区指定的方格里。
4. 变量名栏 列出了数据文件中所包含变量的变量名。
5. 观测序号栏 列出了数据文件中的所有观测值。观测的个数通常与样本容量的大小一致。
6. 窗口切换标签 用于“数据视图”和“变量视图”的切换。即数据浏览窗口与变量浏览窗口。数据浏览窗口用于样本数据的查看、录入和修改。变量浏览窗口用于变量属性定义的输入和修改。
7. 状态栏 用于说明显示 SPSS 当前的运行状态。SPSS 被打开时, 将会显示“IBM SPSS Statistics Processor is ready”的提示信息。

(二) 结果输出窗口

在 SPSS 中大多数统计分析结果都将以表和图的形式在结果观察窗口中显示, 如图 1-4 所示。窗口右边部分显示统计分析结果, 左边是导航窗口, 用来显示输出结果的目录, 可以通过单击目录来展开右边窗口中的统计分析结果。当用户对数据进行某项统计分析, 结果输出窗口将被自动调出。当然, 用户也可以通过双击后缀名为 .spo 的 SPSS 输出结果文件来打开该窗口。

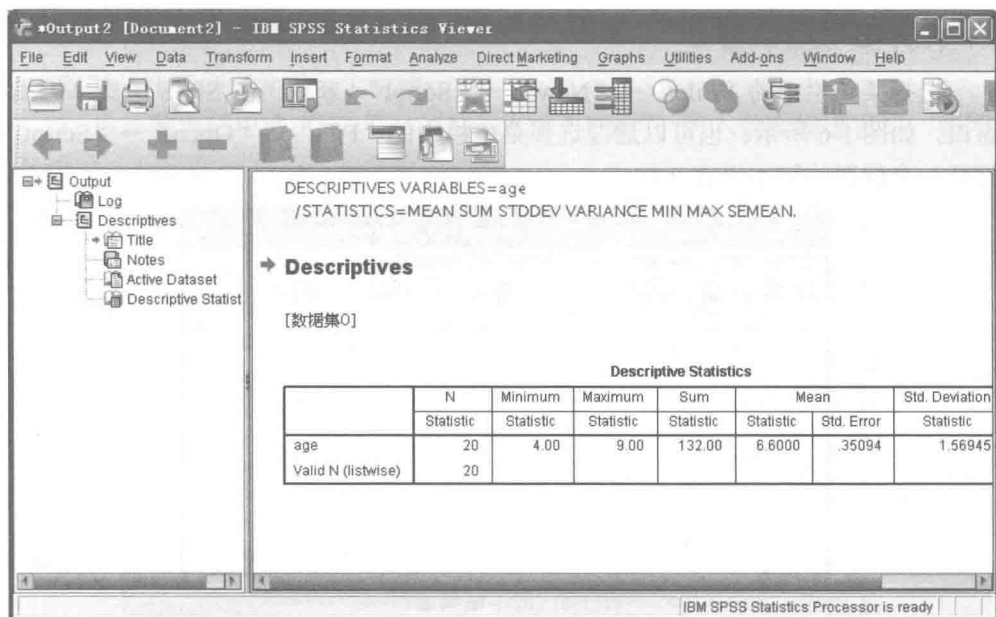


图 1-4 结果输出窗口

(三) 语句窗口

选择菜单栏中的“File”→“New”→“Syntax”，新建一个 SPSS 的语句文件，如图 1-5 所示。也可以通过选择菜单栏中的“File”→“Open”→“Syntax”，打开一个保存的语句文件。



图 1-5 语句窗口

(四) 脚本编辑窗口

选择菜单栏中的“File”→“New”→“Script”，新建一个 SPSS 的脚本编辑窗口，如图 1-6 所示。也可以通过选择菜单栏中的“File”→“Open”→“Script”，打开一个保存的脚本语言文件。

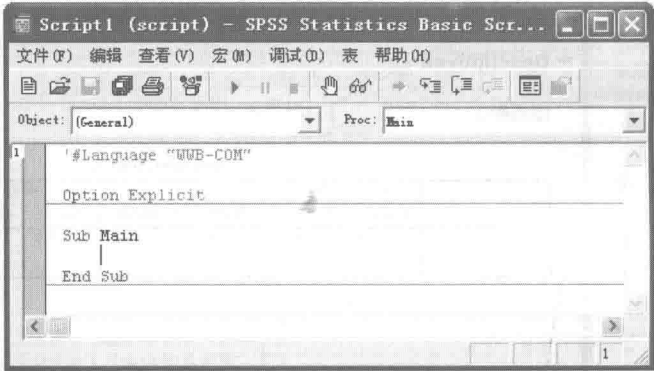


图 1-6 脚本编辑窗口

五、数据文件的管理

SPSS 数据文件是一种结构性数据文件，由数据的结构和数据的内容两部分构成，也可以说由变量和观测两部分构成。SPSS 中的变量共有 10 个属性，分别是变量名(Name)、变量类型(Type)、长度(Width)、小数点位置(Decimals)、变量名标签(Label)、变量名值标签(Value)、缺失值(Missing)、数据列的显示宽度(Columns)、对齐方式(Align)和度量尺度(Measure)。定义一个变量至少要定义它的两个属性，即变量名和变量类型，其他的属性可以暂时采用系统默认值，待以后分析过程中如果有需要再对其进行设置。在 SPSS 数据编辑窗口单击“Variable View”(变量视图)标签，进入变量视图界面，即可对变量的各个属性进行设置，如图 1-7 所示。

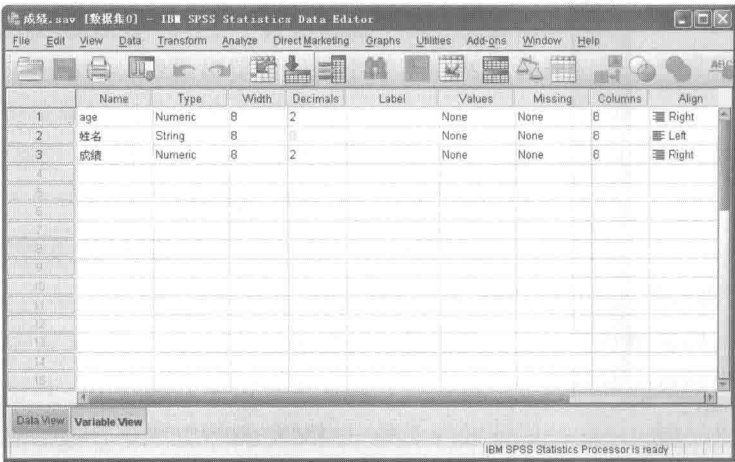


图 1-7 变量视图

(一) 创建一个数据文件

数据文件的创建分成三个步骤:

1. 选择菜单栏中的“File”→“New”→“Data”，进入数据编辑窗口，如图 1-8 所示。

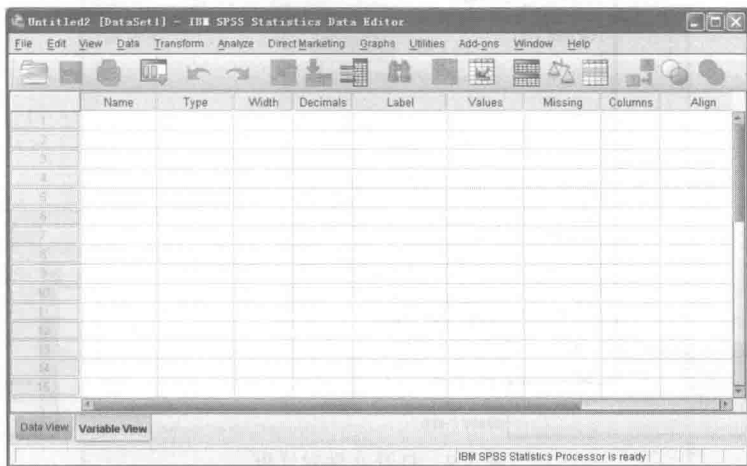


图 1-8 数据编辑窗口

2. 单击左下角“Variable View”标签进入变量视图(也叫变量窗口)界面, 根据试验的设计定义变量名和变量类型。设置变量名时, 可以是不区分大小写的中英文的多个字符, 但不能是 SPSS 的保留字。常用的 SPSS 的保留字有 ALL、AND、NOT、EQ、GE、GT、LE、LT、NE、TO、WITH 及常用的函数符号等。

3. 变量定义完成之后, 单击左下角“Data View”标签进入数据视图界面, 将每个具体的变量值录入数据库单元格内。

(二) 读取外部数据

SPSS 可以很容易地读取 Excel 数据。选择菜单栏中的“File”→“Open”→“Data”打开数据对话框, 在文件类型下拉列表中选择数据文件, 如图 1-9 所示。选定要打开的 Excel 文件, 单击“Open”, 调出打开 Excel 数据源对话框, 可在工作表的下拉列表中选择被读取数据所在的 Excel 工作表, 或者在范围输入框中限制被读取数据在 Excel 工作表中的位置。

(三) 数据编辑

在 SPSS 中, 对数据进行基本编辑操作的功能集中在“Edit”和“Data”菜单中, 读者可在后续章节中看到详细介绍。

(四) 数据的保存

SPSS 数据录入编辑整理完成后应该及时保存, 以防数据丢失。保存数据文件可以单击菜单栏中的“File”→“Save”或“File”→“Save as”。在数据保存对话框中根据不同要求进行 SPSS 数据保存, 如图 1-10 所示。

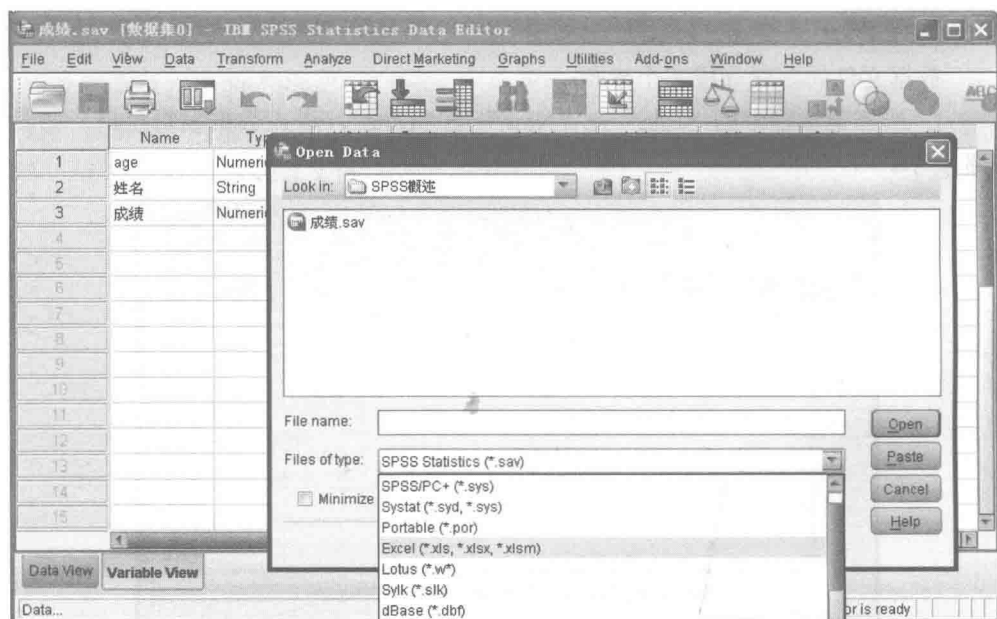


图 1-9 打开文件对话框

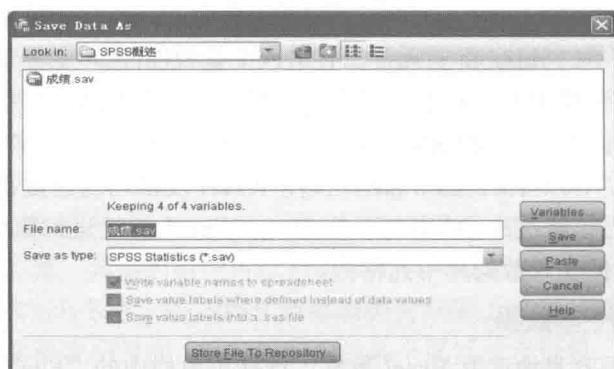


图 1-10 数据保存对话框

(五) 数据的整理

在 SPSS 中, 数据整理的功能主要集中在“Data”和“Transform”两个菜单中。

1. 数据排序(sort cases) 对数据按照某一个或多个变量的大小进行排序, 选择在菜单栏中的“Data”→“Sort Cases”打开排序对话框, 如图 1-11 所示。

2. 抽样(select cases) SPSS 可以通过 Select Cases 命令实现样本筛选的功



图 1-11 数据排序对话框

能。选择在菜单栏中的“Data”→“Select Cases”打开选择个案对话框，如图 1-12 所示。

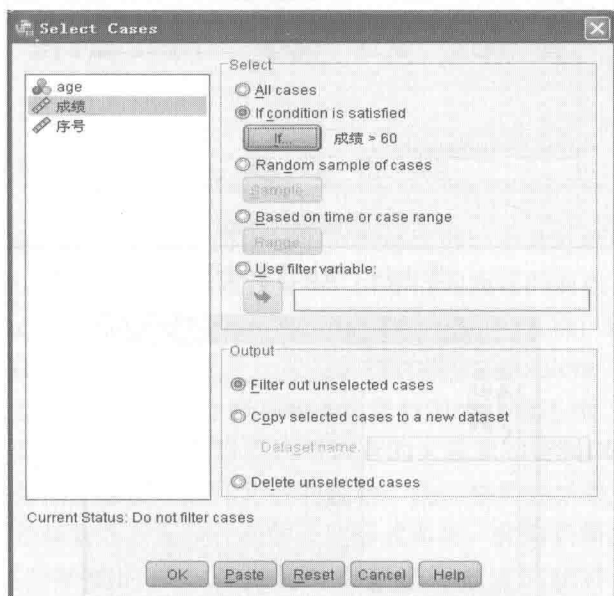


图 1-12 选择个案对话框

3. 增加个案的数据合并(merge files) 将新数据文件中的观测合并到原数据文件中，选择在菜单栏中的“Data”→“Merge Files”→“Add Cases”选择需要增加的数据文件，如图 1-13 所示。

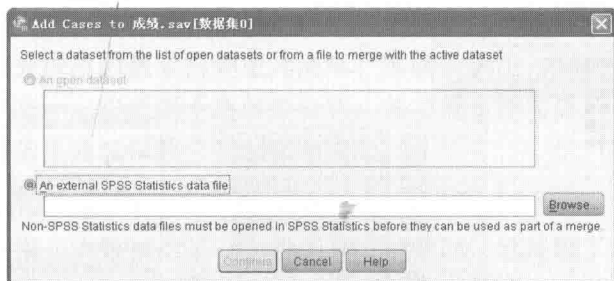


图 1-13 增加个案对话框

4. 增加变量的数据合并(merge files) 增加变量是指把两个或多个数据文件实现横向对接。选择在菜单栏中的“Data”→“Merge Files”→“Add Variables”选择需要合并的数据文件，如图 1-14 所示。

5. 数据拆分(split file) 对数据文件中的数据按照一定要求进行分组，选择在菜单栏中的“Data”→“Split File”打开数据分组对话框，如图 1-15 所示。

6. 计算新变量 在数据的统计分析中，有时需要对数据文件中的变量加工产生新的变量。选择在菜单栏中的“Transform”→“Compute Variable”打开计算

变量对话框，如图 1-16 所示。

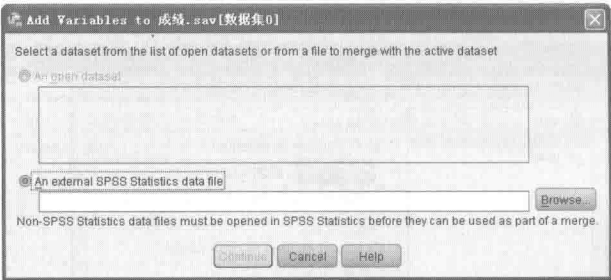


图 1-14 增加变量对话框

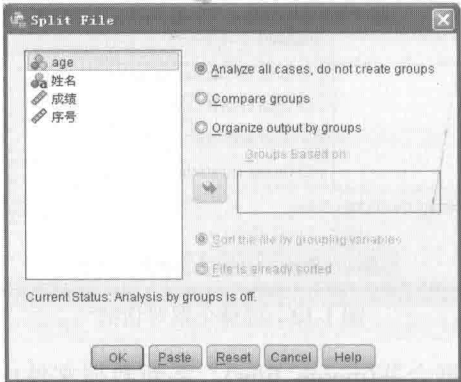


图 1-15 数据分组对话框

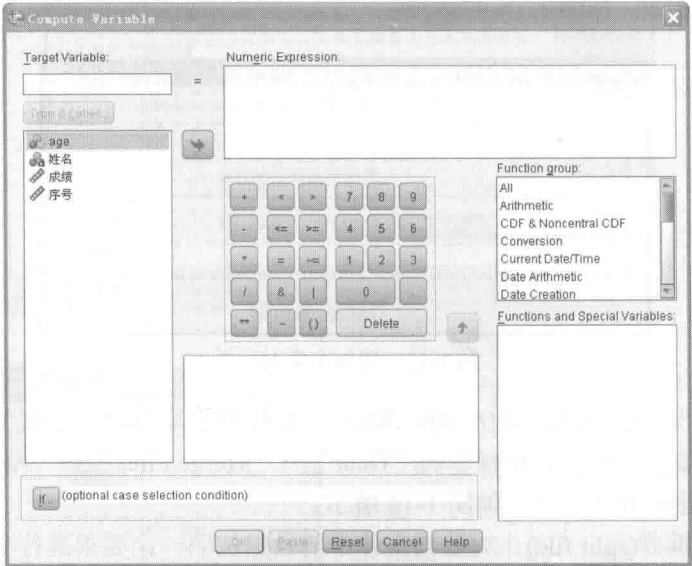


图 1-16 计算变量对话框



第二章 频数分析

一、基本原理

描述性统计分析是统计分析的第一步,做好这第一步是进行统计推断的先决条件。SPSS 的许多模块可进行描述性分析,但专门为此目的而设计的模块则集中在 Descriptive Statistics 菜单中,最常用的是最前面的四个过程: Frequencies 过程是产生频数表; Descriptives 过程是一般性的统计描述; Explore 过程用于探索性分析; Crosstabs 过程则是进行计数资料和等级资料的统计描述和一般的统计检验。

统计整理是根据统计研究的目的,对调查的大量原始资料(即初始资料),进行分类和汇总,得出能够反映总体特征的过程。统计整理的结果为统计表和统计图。统计表主要是频数表,而统计图的表现形式多样,此章只简单进行介绍。

在进行数据分析的时候,首先要对数据进行描述性统计分析(descriptive analysis),从而发现其内在的规律,再进一步选择分析的方法。描述性统计分析要对调查的总体的数据做统计性描述,主要涵盖数据的频数分析、数据的集中趋势分析、数据的离散程度分析、数据的分布及基本的统计图形。

在本章中,我们主要介绍频数分析。对于一组数据,考虑不同的数值出现的频数,或者数据落入指定区域内的频数,则可以了解数据的分布状况。通过频数分析,用户可以得到描述性统计结果,还能了解变量取值的分布情况,能使总体数据的分布频数分析出的数字得到更为清晰、准确的输出。频数分析是统计分析中最常用的功能之一,它适用于离散型资料;其功能是描述离散型变量的分布特征。频数分析通常是通过频数分布表来完成的。

(一) 描述集中趋势的统计量

集中趋势是指一组数据向某一中心值靠近的倾向,它反映了一组数据中心点的位置所在。集中趋势测度就是找到数据水平的代表值或中心值。集中趋势主要依赖各种平均指标来进行反映。

1. 均值 均值又称为算术平均数,是集中趋势测定中最重要的一种。其定义为: 设 $x_1, x_2, \dots, x_n$ 是取自某总体的一个样本,它的均值为:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2-1)$$

均值有四个重要性质: ①各变量值与平均数离差之和等于零; ②各个变量值与平均数离差平方和为最小值; ③常数的均值是其本身; ④对于任何两个变量 x 和 y , 它们代数 and 的均值就等于两个变量的均值的代数和。