

临床统计方法及SPSS应用

主编 孙海双



科学出版社

www.sciencep.com

临床统计方法及 SPSS 应用

主 编 孙海双

科学出版社

北 京

内 容 简 介

SPSS 是世界著名的统计分析软件之一。本书通过医学实例,以 SPSS 13.0 中英对照形式,重点介绍了 SPSS 在各种统计分析方法中的应用及操作,内容主要包括统计学理论、统计方法选择、软件操作、结果分析和界面说明等方面。通过统计学理论、统计方法选择和结果分析,指导读者掌握并正确运用医学统计方法;通过软件操作和界面说明,使读者充分了解 SPSS 软件并运用 SPSS 软件开展医学科研实践。

本书适用于医疗卫生系统工作人员、医学院校的师生以及相关领域的研究生。

图书在版编目(CIP)数据

临床统计方法及 SPSS 应用 / 孙海双主编. —北京:科学出版社,2009
ISBN 978-7-03-023881-8

I. 临… II. 孙… III. 医学统计-统计分析-软件包,SPSS IV. R195.1-39

中国版本图书馆 CIP 数据核字(2008)第 001443 号

策划编辑:沈红芬 黄 敏 / 责任编辑:黄相刚 / 责任校对:钟 洋
责任印制:刘士平 / 封面设计:黄 超

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>

骏 杰 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2009 年 2 月第 一 版 开本:787×1092 1/16

2009 年 2 月第一次印刷 印张:16 1/2

印数:1—2000 字数:386 000

定价:48.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

《临床统计方法及 SPSS 应用》编写人员

主 编 孙海双

编 者 (按姓氏汉语拼音排序)

胡庆成 李晓莉 孙海双

王颖芳 夏晓燕 杨晋华

前 言

本书以临床医学统计理论和 SPSS 统计软件操作为主线,通过向读者介绍如何选择统计方法、正确使用 SPSS 统计软件、对输出结果合理解释、SPSS 界面说明等四个方面,指导读者如何掌握并运用统计理论和方法,正确使用 SPSS 统计软件处理临床医学数据,解决医学问题,实现了统计学理论、计算机软件应用和解决临床医学问题三方面的交叉与融合。

全书共分十四章,其中第九章至第十三章为高级统计分析方法部分,其余为基础统计分析方法部分。在内容选择上,一是突出临床医学统计方法的基本理论和方法,使读者理解统计理论的内涵,能正确、合理地选择统计方法;二是突出应用统计软件导出的结果解释以及软件界面说明,使读者能够正确、合理使用统计软件。

本书编写方式以临床医学统计方法为纲目,不同于 SPSS 专业人士编写的、以 SPSS 菜单界面为纲目的风格。为方便读者对英文版和汉化版的不同需求,采用中文汉化版介绍为主,关键步骤采用中、英两种版本对照的形式介绍 SPSS 的应用。本书适用于 SPSS 10.0~13.0 的各种 SPSS 版本。

本书介绍的解决临床医学问题的统计方法和软件操作,也适用于生物统计等其他学科。编写中兼顾临床工作者非统计学和计算机专业人士的特点,突出编写内容的易懂性、实用性和可操作性,可以使读者能学以致用,举一反三。本书不仅可用于指导广大医务工作者进行科研活动,也可作为医学院校的本科生和研究生的学习教材和参考书。

由于编者学识、水平所限,书中不妥之处恳请读者和同仁不吝批评指正。

孙海双

2008 年 7 月

目 录

第一章 数据文件的建立与整理	(1)
第一节 SPSS 概述	(1)
第二节 数据文件的建立与编辑	(5)
第三节 数据转换	(30)
第二章 定量资料统计描述	(44)
第一节 概述	(44)
第二节 定量资料统计描述	(47)
第三章 正态性检验	(56)
第一节 概述	(56)
第二节 正态性检验	(57)
第四章 t 检验	(64)
第一节 概述	(64)
第二节 单样本均数 t 检验	(68)
第三节 配对资料 t 检验	(70)
第四节 两组独立样本资料 t 检验	(73)
第五章 方差分析	(76)
第一节 概述	(76)
第二节 完全随机设计资料方差分析	(76)
第三节 方差分析均数间两两比较	(80)
第四节 随机区组设计方差分析	(82)
第五节 交叉设计方差分析	(85)
第六节 析因设计方差分析	(88)
第七节 重复测量设计方差分析	(90)
第六章 分类资料统计描述及 χ^2 检验	(97)
第一节 概述	(97)
第二节 四格表资料 χ^2 检验	(101)
第三节 配对资料 χ^2 检验	(106)
第四节 行列表资料 χ^2 检验	(107)
第五节 分层资料 χ^2 检验	(111)
第六节 一致性检验	(113)
第七章 秩和检验	(116)
第一节 概述	(116)
第二节 配对资料秩和检验	(117)

第三节	两独立样本资料秩和检验	(120)
第四节	多个独立样本资料秩和检验	(123)
第五节	随机区组资料秩和检验	(125)
第八章	相关与回归	(128)
第一节	概述	(128)
第二节	简单直线相关	(129)
第三节	简单直线回归	(133)
第四节	秩相关	(139)
第五节	多重线性相关	(141)
第六节	偏相关	(143)
第七节	多重线性回归	(145)
第八节	逐步回归	(148)
第九章	协方差分析	(152)
第一节	概述	(152)
第二节	完全随机设计资料协方差分析	(153)
第三节	随机区组设计资料协方差分析	(160)
第十章	Logistic 回归分析	(167)
第一节	概述	(167)
第二节	非条件 Logistic 回归分析	(168)
第三节	条件 Logistic 回归分析	(179)
第十一章	生存分析	(183)
第一节	概述	(183)
第二节	生存函数的估计和检验	(184)
第三节	Cox 比例风险模型	(197)
第十二章	聚类分析	(205)
第一节	概述	(205)
第二节	按指标聚类分析(R 型聚类)	(207)
第三节	按样品聚类分析(Q 型聚类)	(216)
第四节	K-means 聚类分析	(220)
第十三章	主成分分析与因子分析	(226)
第一节	概述	(226)
第二节	主成分分析与因子分析	(227)
第十四章	统计图	(238)
第一节	条形图	(238)
第二节	饼图	(245)
第三节	线图	(247)
第四节	直方图	(249)
第五节	散点图	(251)
主要参考文献		(254)

第一章 数据文件的建立与整理

第一节 SPSS 概述

一、概 述

SPSS(statistical package for the social science)即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,美国斯坦福大学的三位学生开发了最早的 SPSS 统计软件系统,同时成立 SPSS 公司,最初主要应用于企事业单位。1984年,SPSS 公司首先推出了世界第一个应用于微机版本的统计分析软件 SPSS/PC+,其最初几个版本是基于 DOS 环境(SPSS for DOS)。20世纪90年代以来,由于 MS Windows 的普及,SPSS 公司从1992年开始相继推出了 Windows 操作系统的 5.0 及以上版本。与其他版本相比,SPSS 13.0 for Windows 的用户界面更为友好,使用更方便,统计功能更强大,成为当今世界上最新、最流行、最受欢迎的统计软件包之一,它不但应用于社会科学,也被广泛应用于自然科学、技术科学等领域。它的用户分布于通信、医疗、银行、证券、保险、制造、商业、市场研究、科研教育等行业,是世界上应用最广泛的专业统计软件之一。本书以 SPSS13.0(中文版)和 13.0(英文版)为例介绍该软件的具体使用方法。

二、SPSS for Windows 的特点

SPSS for Windows 版本软件具有以下特点:

1. SPSS for Windows 的大多数操作是通过“菜单”、“图标按钮”、“对话框”来完成的,操作简单、使用方便。
2. SPSS for Windows 的命令语句、子命令及选择项绝大部分由“对话框”的操作完成,无需记忆大量的命令、过程、选择项。
3. SPSS for Windows 与其他软件有数据转换接口,其他软件生成的数据文件包括数据库生成的 DBF 文件等;同理,SPSS 的数据文件也可以方便地转换成其他数据文件。
4. SPSS for Windows 与其他 Windows 软件兼容,统计结果可以复制至 Word 等,以便进一步予以编辑。

三、SPSS for Windows 的运行环境

SPSS for Windows 的使用环境要求如下:

1. 硬件环境 SPSS for Windows 的硬件环境要求取决于选择哪些类分析模块以及所选用的版本号。一般应留有至少 10M 以上的自由空间。否则,存取速度会明显下降,甚至

会出现出错的信息。在运行该软件时,计算机选型应在 PC486/50MHz。此外,应有 4M 以上的内存空间,在进行多因素分析时,应有 4M 以上的内存空间。

2. 软件环境 Windows95/98/2000/XP 操作系统。在英文 Windows 环境下运行故障率较少。

四、SPSS for Windows 的安装

SPSS for Windows 的安装按如下步骤进行:

1. 启动计算机,将 SPSS13.0 光盘插入光驱。
2. 在“我的电脑”中找到 SPSS 文件夹,单击“setup.exe”,启动安装程序。
3. 根据安装程序的提示向导,依次进行安装,并输入软件序列号码、用户姓名和单位名称。
4. 退出安装程序。
5. 系统在 Windows 程序管理器窗口中建立一个 SPSS 程序组。
6. 用户可以在桌面上建立一个 SPSS 的快捷方式,以方便使用。

五、SPSS for Windows 的界面

1. SPSS for Windows 的启动 在 Windows 的程序管理器中或桌面的快捷方式上,双击 SPSS for Windows 的图标,即可启动 SPSS。SPSS 启动时,出现 SPSS 导航对话框,见图 1-1。导航对话框提示你希望做什么? 在单选框组中有“运行教程”、“输入数据”、“运行现有

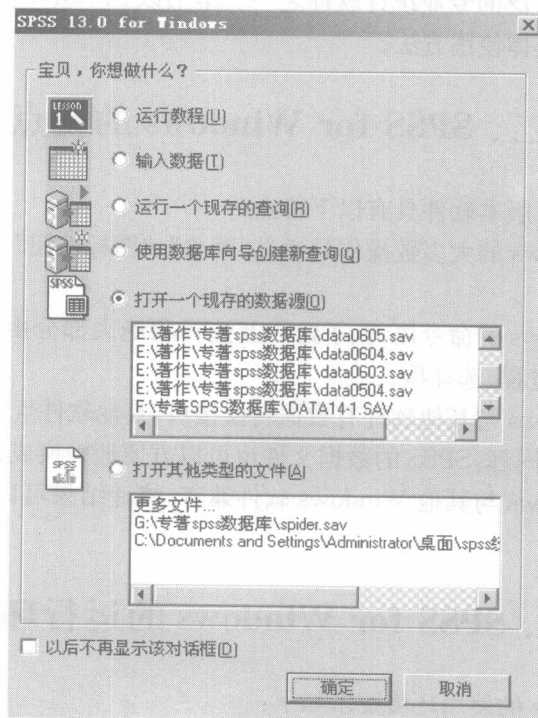


图 1-1 SPSS 导航对话框

查询”、“使用数据库向导创建新查询”、“打开现有数据库”、“打开其他文件类型”选项。如果用户选择“以后不再显示此对话框”选项,可直接进入 SPSS 主画面。

2. SPSS 的主窗口 SPSS 的主窗口名为数据编辑框 SPSS Data Editor,见图 1-2(英文版)和图 1-3(汉化版)。位于窗口顶端的标题栏中,为活动窗口,用户可对之进行操作。SPSS 的主窗口的默认标题名称为 Untitled SPSS Data Editor(无标题数据编辑框),当调用其他已经命名的数据文件或将新文件进行存储后(后缀为 .sav),标题名称即显示数据文件的名称。主窗口是一种典型的电子表格形式,用户可通过定义变量名、格式化数据类型后输入原始数据,并可根据需要对数据进行增删、剪贴、修改和存储等操作。

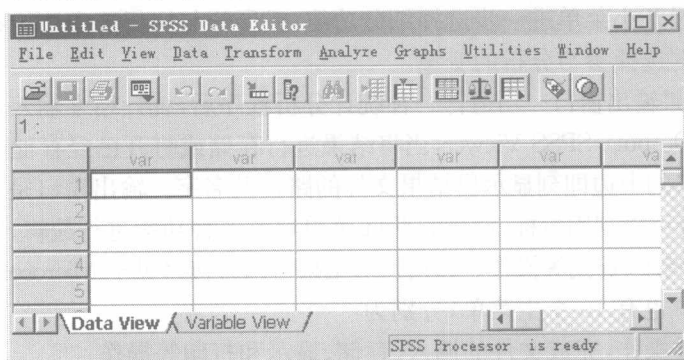


图 1-2 SPSS 的主窗口(英文版)

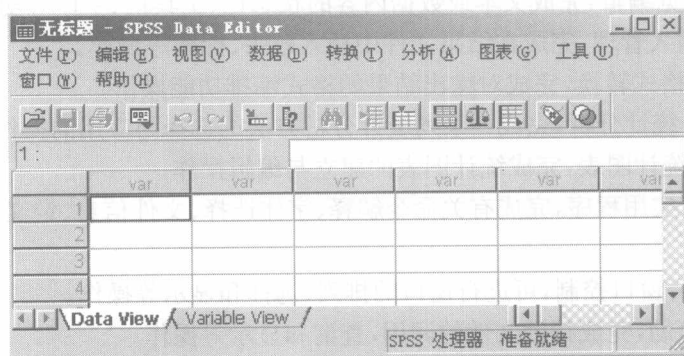


图 1-3 SPSS 的主窗口(汉化版)

SPSS 的主窗口有两个界面:一个是 Data View(数据视图)界面,另一个是 Variable View(变量视图)界面。Variable View 界面是用来定义变量的格式,Data View 界面用于输入、编辑数据。SPSS 的主窗口的底部为系统状态栏,显示系统当前的工作状态。

3. SPSS 的菜单 SPSS for Windows 启动后即在屏幕上显示主画面,主画面的最上行是由 10 个菜单项组成的主菜单,这些菜单项包括:

- (1) File:文件操作,完成文件的调入、存储、显示和打印等操作。
- (2) Edit:文件编辑,完成文本或数据内容的选择、拷贝、剪贴、寻找和替换等操作。
- (3) View:浏览编辑,完成文本或数据内容的状态栏、工具栏、字体、网格线和数值标签等功能的操作。
- (4) Data:数据管理,完成数据变量名称和格式的定义,数据资料的选择、排序、加权,数

据文件的转换、连接和汇总等操作。

- (5) Transform: 数据转换, 完成数值的计算、重新编码和缺失值替代等操作。
- (6) Analyze: 统计分析, 完成一系列统计分析方法的选择与应用。
- (7) Graphs: 统计图表, 完成统计图表的建立与编辑。
- (8) Utilities: 实用程序, 有关命令解释、字体选择、文件信息、定义输出标题和窗口设计等。
- (9) Windows: 窗口控制, 可进行窗口的排列、选择和显示等操作。
- (10) Help: 帮助, 帮助文件的调用、查询和显示等。

每个主菜单都包括几个子菜单, 分别执行不同的功能。从菜单中选择要执行的功能, 单击鼠标左键可以展开小菜单进一步选择细分功能, 或展开与执行选中功能有关的对话框, 用户可以通过对话框选择参与分析的变量。

4. SPSS 的结果输出窗口 当运行一种统计分析程序后, 统计结果输入输出窗口, 软件默认的标题名称为 Output1-SPSS Viewer, 当将结果进行存储或打开已经存储的结果文件时(后缀名为 .spo), 窗口的上边即刻显示出结果文件的路径及名字。输出窗口是一个文本窗口, 寻找该窗口中的内容可以利用鼠标、键盘的 Edit 菜单项的各种功能进行编辑, 即可以进行选择、删除、移动、拷贝、修改、查找等操作。该窗口的结果可以直接拷贝至 Microsoft 文件中。

结果输出窗口也有 10 个主菜单, 分别为

- (1) File: 文件操作, 完成文件的调入、存储、显示和打印等操作。
- (2) Edit: 文件编辑, 完成文本或数据内容的选择、拷贝、剪贴、寻找和替换等操作。
- (3) View: 浏览编辑, 完成文本或数据内容的状态栏、工具栏、字体等功能的操作。
- (4) Insert: 插入管理, 完成在结果文件中插入题头、标题等操作。
- (5) Format: 格式转换, 完成对输出结果的格式编排功能操作。
- (6) Analyze: 统计分析, 完成一系列统计分析方法的选择与应用的操作。
- (7) Graphs: 统计图表, 完成统计图表的建立与编辑操作。
- (8) Utilities: 实用程序, 完成有关命令解释、字体选择、文件信息、定义输出标题和窗口设计等的操作。
- (9) Windows: 窗口控制, 可进行窗口的排列、选择和显示等操作。
- (10) Help: 帮助, 完成帮助文件的调用, 查询和显示等操作。

5. SPSS 的退出 选择 File 菜单中的 Exit 项或者右上角的关闭按钮, 并回答系统提问(是否需要存储原始数据、计算结果或统计图等), 即可退出 SPSS。

6. SPSS 的帮助系统 SPSS 的帮助系统有两种形式:

- (1) 主窗口的 Help 菜单: 在软件运行的任何时候, 单击 Help 菜单相关的子菜单, 可得到所需的各种帮助。
- (2) 各种对话框中的 Help 按钮: 在具体操作过程中, 当弹出某一对话框时, 一般总有 Help 按钮, 单击该按钮, 用户可得到这一对话框选项的详细帮助。

7. SPSS 的系统参数设置 SPSS 的系统参数设置窗口如图 1-4 所示。系统初始状态和系统默认值的设置是通过 Options 对话框完成的, Options 功能项在 Edit 菜单中。参数与状态设置生效的时间不同, 有的在确认(按 OK 按钮)后立即生效, 有的则要在下次启动 SPSS 时才有效。但无论何时生效, 只要生效, 设定的状态或参数即代替了原来系统给定的默认值。

鼠标单击主菜单的 Edit 菜单项展开下拉菜单, 在下拉菜单中选择最后一项; Options,

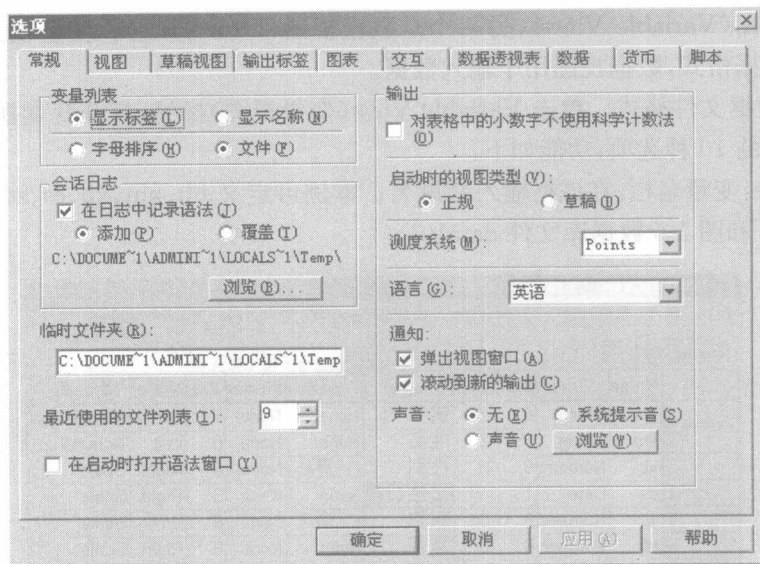


图 1-4 系统参数设置选项窗口

打开相应的对话框,在 Options 对话框中进行系统状态、参数的设置。

第二节 数据文件的建立与编辑

数据管理是 SPSS 的重要组成部分,也是对数据进行统计分析的基础。在对数据分析之前必须先建立数据文件,将收集到的各种信息、数据输入计算机中。SPSS 具有建立数据文件的功能,在 SPSS 中建立数据文件分两步:一是在变量视图(Variable View)中建立数据文件的格式(定义变量名、类型、宽度等),二是在数据视图(Data View)中,向建立好格式的数据文件中输入数据。现以某单位部分职工体检资料为例,说明建立数据文件的基本方法,见表 1-1。

表 1-1 某单位部分职工体检资料

编号(bh)	姓名(xm)	性别(xb)	出生年月(cs)	职称(zc)	婚否(hf)	身高(cm)	体重(kg)
0101	李莉娟	女	01/12/78	助工	F	165.5	48.0
0102	王万宏	男	12/23/60	高工	T	180.3	49.8
0103	张华卫	男	07/01/70	工程师	T	172.4	64.6
0104	赵斌	男	11/05/54	高工	T	178.2	51.6
0105	梁萍	女	03/12/79	助工	F	169.2	58.8
0201	王兰香	女	11/23/78	工人	F	169.2	80.2
0202	黄丽丽	女	05/12/67	工程师	T	158.2	60.2
0203	王永歌	男	06/29/72	助工	F	176.0	74.4
0204	许艳艳	女	02/28/64	高工	F	161.1	62.8
0205	李建辉	男	04/12/68	工程师	T	174.3	57.8

一、建立数据文件

启动 SPSS 软件并显示数据编辑窗口(DATA EDITOR)。数据编辑窗口有两个标签,

一个是变量视图(Variable View),另一个是数据视图(Data View)。变量视图用于定义和编辑变量的数据格式,数据视图用于输入数据。

1. 定义数据文件格式 单击 Variable View(变量视图)标签,打开变量视图窗口,系统出现定义变量的 10 种选项,功能如下:

(1) Name:变量名栏,在该栏输入变量名。本例可定义 bh、xm、xb、cs、zc、hf、cm、kg 等变量,见图 1-5 和图 1-6 数据库文件 data101。

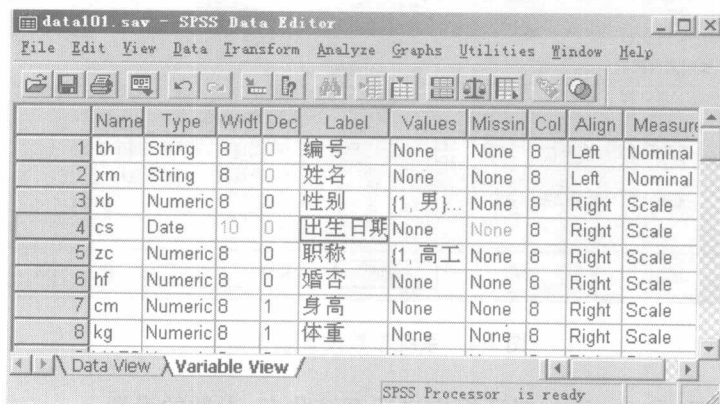


图 1-5 变量视图(英文版)

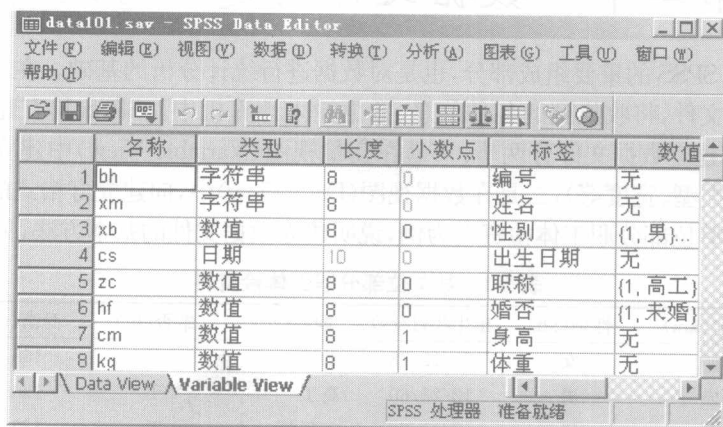


图 1-6 变量视图(汉化版)

为变量命名时,应该遵循如下原则:

1) 变量名可由字符组成,也可使用汉字,但由于汉字在变量输入与处理方面有诸多不便,一般不建议使用汉字。

2) 首字符应该是英文字母或汉字,其后可为字母或数字及除了“?”、“!”和“*”以外的字符。但圆点“.”不能出现在变量名尾。

3) 变量名不能使用 SPSS 的关键字(保留字),如 ALL、AND、OR、NOT、EQ、GE、LE、LT、NE、TO、WITH 及一些常用的符号等。

4) 系统中不区分变量名中的大小写字符。例如,ABCD 与 abcd 被认为是同一变量。

(2) Type: 变量类型, 默认变量类型为 Numeric, 单击 Numeric 右侧的小按钮, 可打开变量类型对话框, 改变变量类型。变量类型有 8 种, 见图 1-7。

1) Numeric: 标准数值型变量, 系统默认为数值型, 总长度为 8, 小数位数为 2。系统的默认长度可以从 Edit 菜单中的 Options 命令中重新设置。

2) Comma: 带逗点的数值型变量, 默认总长度为 8, 小数位数为 2。其值在显示时, 整数部分自右向左每三位用一个逗点作分隔符, 圆点做小数点。如 12 345.00, 输入时逗点可不输入。

3) Dot: 带圆点的数值型变量, 默认总长度为 8, 小数位数为 2。显示时与 Comma 相反, 其值在显示时, 整数部分自右向左每三位用一个圆点作分隔符, 逗点做小数点。

4) Scientific Notation: 科学记数法, 默认总长度为 8, 小数位数为 2。对于很大或很小的数据用此方法表示, 指数的字母可以用 E, 也可用 D, 也可省略, 如 12345 可输入为 1.2345E4、12345、1.2345D4、1.2345E+4、1.2345+4, 但显示值为 1.2345E+04。

5) Date: 日期型变量, 有 27 种表示方法, 见表 1-2 和图 1-8。

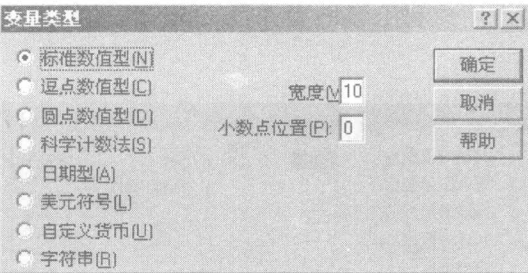


图 1-7 数值型变量类型对话框

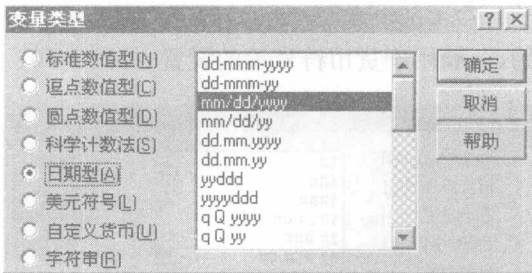


图 1-8 日期型变量类型对话框

表 1-2 日期型变量格式

格式	说明	格式	说明
dd-mmm-yyyy	日日-月月-年年年年	MONDAY, TUESDAY...	星期几
dd-mmm-yy	日日-月月-年年	MON, TUE, WED.....	星期几的缩写
mm/dd/yyyy	月月/日日/年年年年	JANUARY, FEB...	月份
mm/dd/yy	月月/日日/年年	JAN, FEB, MAR	月份缩写
dd. mm. yy	日日. 月月. 年年	dd-mmm-yyyy hh:mm	日日-月月-年年年年 时时: 分分
yyddd	年年日数	dd-mmm-yyyy hh:mm:ss	日日-月月-年年年年 时时: 分分: 秒秒
yyyyddd	年年年年日数	dd-mmm-yyyy hh:mm:ss. ss	日日-月月-年年年年 时时: 分分: 秒秒. 百分秒
qQyyyy	季度 Q 年年年年	hh:mm	时时: 分分
qQyy	季度 Q 年年	hh:mm:ss	时时: 分分: 秒秒
mmmyyyy	月份年年年年	hh:mm:ss. ss	时时: 分分: 秒秒. 百分秒
mmyy	月份年年	ddd: hh:mm	日数: 时时: 分分
wwWKyyyy	周数 WK 年年年年	ddd hh:mm:ss	日数: 时时: 分分: 秒秒
wwWKyy	周数 WK 年年	ddd hh:mm:ss. ss	日数: 时时: 分分: 秒秒. 百分秒

6) Dollar: 货币型变量, 默认总长度为 8, 小数位数为 2, 其值在显示时有效数字前有 “\$”, 用逗号做分隔符。输入时可不带 “\$”, 系统自动加上。如输入 12 345. 67 系统自动显示: \$ 12 345. 67。有 12 种表示方法, 见表 1-3 和图 1-9。

表 1-3 货币型变量格式

格式	总长度	小数位数	格式	总长度	小数位数
\$ #	2	0	\$ # # # # , # # #	8	0
\$ # #	3	0	\$ # # # # , # # # . # #	11	2
\$ # # #	4	0	\$ # # # # , # # # , # # #	12	0
\$ # # # . # #	7	2	\$ # # # # , # # # . # # # . # #	15	2
\$ # , # # #	6	0	\$ # # # # , # # # , # # # , # # #	16	0
\$ # , # # # . # #	9	2	\$ # # # # , # # # , # # # , # # # . # #	19	2

7) Custom currency: 自定义型变量, 是一种由用户利用 Edit 菜单的 Options 功能定义的, 一般用于货币符号等的设置, 见图 1-10。

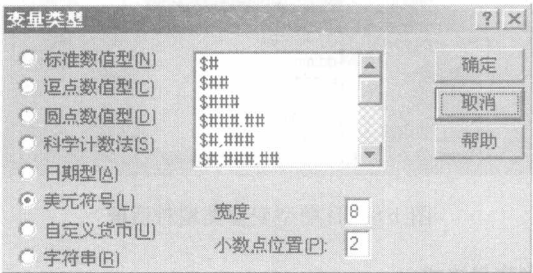


图 1-9 美元货币型变量类型对话框

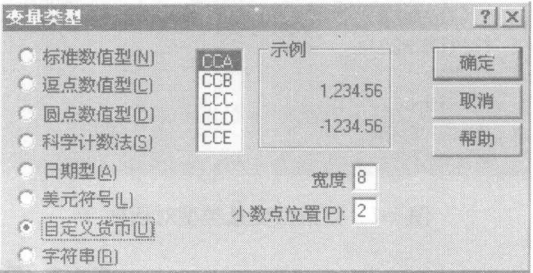


图 1-10 自定义货币型变量类型对话框

其中: CCA、CCB、CCC、CCD、CCD、CCE 是用户自己定义的 5 种自定义格式, 如定义 CCA 的格式为 ¥ 1234. 56 RMB 等。

可在 Edit 菜单的 Options 命令中, 打开 Currency(货币)选项卡, 进行设置, 见图 1-11。其中 All Values 用于设置首(Prefix)尾(Suffix)字符, Negative Value 栏用于设置负数的首(Prefix)尾(Suffix)字符, 系统默认负数的首字符是 “-”。Decimal Separator 栏用于设置小数点的符号, 默认为圆点(Period), 也可定义为逗号(Comma)。

8) String: 字符型变量。由任何可以显示的字符组成, 可以是汉字或字符, 作为常量时应用单引号 ‘ ’、双引号 “ ” 括起。注意应为英文引号。

为便于数据统计, 一般变量类型应定义为数值型。

(3) Width: 宽度。

(4) Decimals: 小数位数。

根据每个变量数据的大小(最大数)及保留小数点的位数, 定义变量的总宽度, 小数点位数, 总宽度包括小数点位数。但不包括小数点本身。如 12 345. 00。宽度定义为 7 位, 2 位小数。本例中, 定义性别为字符型、宽度为 1 (也可定义为数值型、小数位为 0)。

(5) Label: 变量标签。为了便于标示变量, 对变量的含义进行进一步说明, 常常需要用

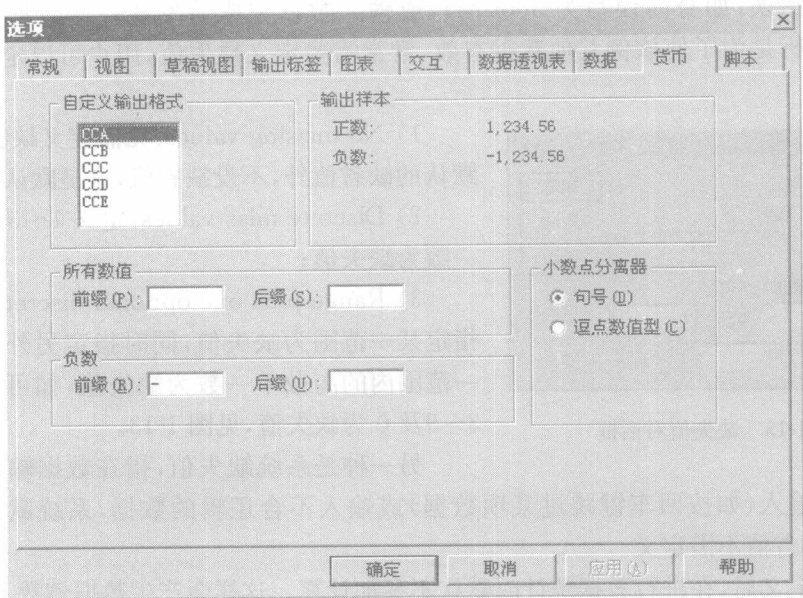


图 1-11 选项对话框

汉字表示,变量的内容,如 xb 变量的标签为“性别”。给变量加上标签后,在数据窗口鼠标指向变量时,变量名下会显示标签。

(6) Values: 数值标签。对 XB(性别)及 ZC(职称)变量,对变量的可能取值进行进一步的说明,通常仅对分类变量的取值指定值标签。当变量值是有限数据时,对这些数据输入时尽量用代码输入,以加快输入速度、方便数据处理。如性别中“男”可输入“1”(或 M)，“女”可输入“2”(或 F),职称也与此类似。结果输出时为了便于识别,可以使用数值标签,定义输入值的含义。默认没有数值标签 (None),要改变,可单击 None 后的小按钮,弹出 Value Label(数值标签)对话框,在上面的“Value”栏中输入变量值,如“1”,在下面的 Value 栏中输入标签如“男”,单击“Add”按钮,同理可输入其他数值说明。XB 中“1”为“男”,“2”为“女”。这样凡是在结果中性别是“1”的地方都会用“男”代替,性别是 2 的地方都会用“女”代替,见图 1-12。

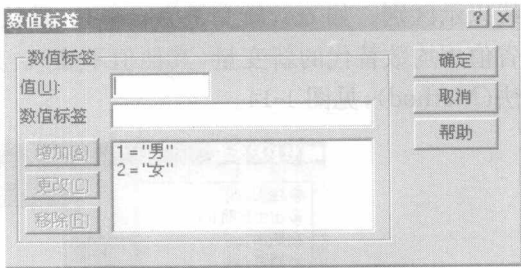


图 1-12 数值标签对话框

如要在数据视图中显示变量标签,则可执行“View 菜单 | Value label”命令,如对性别变量值标签定义完毕后,则输入 1 时将显示“男”,2 时显示“女”,也可通过单击“值”标签右侧的按钮,在弹出的下拉列表框中进行选择其他值,用于对输入值的修改。但变量中存放的数值仍然是其原值,这里是 1 和 2。

(7) Missing: 缺失值,缺失值有两种类型。一种是用户定义的缺失值。由于调查数据资料时(如测量人群生长发育情况),某项数据(如身高)没有调查,数据收集错误(调查表

上填写数据错误,如身高误写为 300CM),或输入数据完毕后发现一些数据不符合逻辑等。这样分析是为了使该例其他数据有效,常常需要定义缺失值,用户可以定义以下 3 种缺失值。

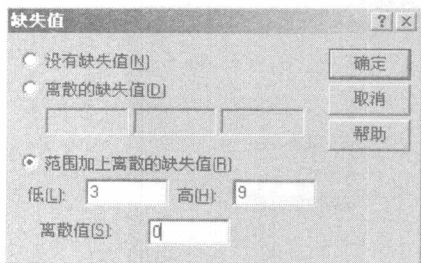


图 1-13 缺失值对话框

1) No missing values:无需定义缺省值,即除了默认的缺省值外,不设缺省值,这是默认方式。

2) Discrete miss values:定义 1~3 个离散的单一数为缺失值;

3) Range plus one optional discrete miss:定义指定某一范围为缺失值,同时指定另外一个不在这一范围内的离散单一数为缺失值,如可定义性别中 3~9 及 0 为缺失值,见图 1-13。

另一种是系统缺失值,指在数据输入时某项数据由于没有输入(如按回车键跳过某项数据)或输入不合逻辑的数据,系统默认缺省值为“.”。系统缺省值不需定义。

缺失值定义后,在进行数据统计时默认不参加计算。这样会产生数据例数不一的情况,对数值型及日期型数据,为保持数据完整性,系统提供了 5 种不同替代缺省值的方法。

选择主菜单的“Transform(数据转换)⇒ Replace Missing Values...(替换缺省值)”命令,打开替换缺省值对话框。在左侧变量列表框中选择要转换缺省值的变量,方法与 Windows 中选择文件方法类似,要选择一个变量单击某变量,要选择多个连续变量,可单击第一个变量,然后按 Shift 键后单击最后一个变量,或用鼠标拖动的方法选择;要选择多个不连续变量,可单击第一个变量,然后按 Ctrl 键后单击其他要选择的变量。变量选择完后,单击对话框中间的右向箭头按钮。将选择变量选择到 New variables(新变量)框中,变量名后加下划线以示区别。如 cm,新变量为 cm_1。默认为使用均数替代缺省值,cm_1 变量中为一将缺省值用均数替代的新变量(其他值不变)。其他替代缺省值(Replace Missing Values)的方法(Method),见图 1-14。

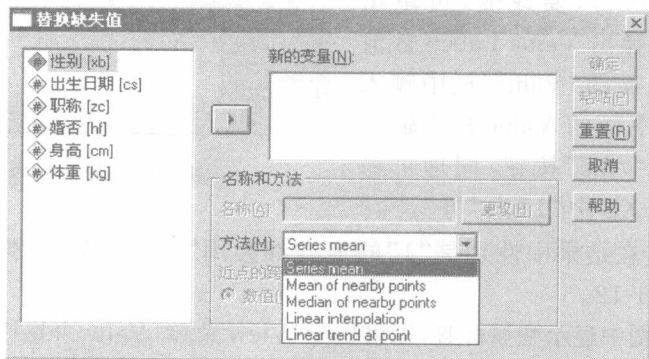


图 1-14 替代缺失值对话框

4) Series mean:将缺省值替代为均数。

5) Mean of nearby points:用邻近点有效数值的均数替换缺省值。

6) Median of nearby points:用邻近点有效数值的中位数替代缺省值。