

普通高等学校教材

供临床医学·护理学·药学·医学检验·康复治疗学等医学专业使用

医学统计学与 SPSS 软件简明操作教程

*Medical Statistics
and SPSS Software Simple Operation*

主 编 童 毅

副主编 唐丽媛 曹丽琰 缪希松



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社

普通高等学校教材

供临床医学·护理学·药学·医学检验·康复治疗学等医学专业使用

医学统计学与 SPSS 软件 简明操作教程

主 编 童 毅

副主编 唐丽媛 曹丽琰 缪希松



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社

图书在版编目(CIP)数据

医学统计学与SPSS软件简明操作教程/童毅主编.

—杭州:浙江大学出版社,2013.12

ISBN 978-7-308-12663-2

I. ①医… II. ①童… III. ①医学统计—统计分析—软件包—教材 IV. ①R195.1-39

中国版本图书馆CIP数据核字(2013)第302636号

医学统计学与SPSS软件简明操作教程

主编 童 毅

责任编辑 邹小宁

文字编辑 黄丽丹

封面设计 王聪聪

出 版 浙江大学出版社

(杭州市天目山路148号 邮政编码 310007)

(网址: <http://www.zjupress.com>)

排 版 杭州教联文化发展有限公司

印 刷 浙江省良渚印刷厂

开 本 787mm×1092mm 1/16

印 张 10.75

字 数 249千

版 次 2013年12月第1版 2013年12月第1次印刷

书 号 ISBN 978-7-308-12663-2

定 价 22.00元

版权所有 翻印必究 印装差错 负责调换

本书编委名单

主 编 童 毅

副主编 唐丽媛 曹丽琰 缪希松

编 者 (以姓氏汉语拼音字头为序)

曹丽琰(昆明学院医学院)

缪希松(昆明学院医学院)

唐丽媛(昆明学院医学院)

童 毅(昆明学院医学院)

前 言

医学统计学是医学科学研究的基础课程之一,其任务是对数据的整理、分析、表达和解释。科学研究的数据量较大,使用传统的手工方法计算,费时费力,又容易出错,严重影响科学研究的工作进度。统计软件的出现是科研工作者的福音,它没有手工方法的诸多弊端,极大地提高研究效率。因此,学习医学统计学除了要掌握理论知识之外,还需要掌握统计软件的使用方法,两者相辅相成,才能熟练地对数据进行分析,为科学研究奠定基础。

本书以 SPSS 17.0 为基础来介绍,共 12 章,大致划分为 SPSS 基本操作、统计图、统计方法三个部分。其中统计方法部分包括统计描述、均数比较、非参数检验、相关与回归等,能满足读者基本的数据分析需求。在统计方法介绍的各有关章节内,第一部分主要包括统计方法的概念、基本思想、应用范围等内容;第二部分主要包括不同统计方法的变量设置、实例图解具体操作步骤、输出结果解释。以上结构安排的目的在于帮助有一定统计学理论基础的读者回忆相关内容,从而更好地理解实例、读懂输出结果。

本书参编人员为医学院校教学岗位的一线青年教师,具有丰富的教学和使用统计软件的经验,书中的数据实例都经过严格的筛选,具有典型性,有利于读者举一反三,结合自己的实际工作正确运用软件进行数据分析。

参与编写的各位老师,秉持着可贵的敬业精神和严谨的科学态度,将自己扎实的理论知识、丰富的实践技能倾注到本书中,在此对各位参编老师表达深深的敬意和感谢!

本书可单独作为统计软件培训教材,也可作为统计教学的实习教材使用,主要面对非统计专业的医学和其他相关专业科研人员使用。

由于编者能力所限,本书中难免有疏漏和不当之处,恳请读者批评指正,不胜感激!我们非常珍惜并认真听取您宝贵的意见和建议,今后将进一步修改完善。

编 者

2013 年 8 月 16 日

目 录

第一章 SPSS 软件概述	1
第一节 SPSS 软件概述	1
第二节 SPSS 的界面窗口	3
第三节 数据文件的建立	5
第二章 统计图	11
第一节 操作概述	11
第二节 条 图	14
第三节 圆 图	18
第四节 百分条图	19
第五节 线 图	21
第六节 箱 图	23
第七节 散点图	25
第三章 统计描述	28
第一节 频数分布	28
第二节 计量资料的统计描述	34
第四章 正态性检验	45
第一节 P-P 图和 Q-Q 图	45
第二节 Kolmogorov-Smirnov 检验	47
第五章 两样本均数比较的假设检验	49
第一节 概 述	49
第二节 单样本均数 t 检验	49
第三节 配对样本均数 t 检验	51
第四节 两独立样本均数 t 检验	54

第六章 多样本均数比较的假设检验	59
第一节 概 述	59
第二节 完全随机设计的方差分析	59
第三节 多个样本均数间的多重比较	63
第四节 随机区组设计的方差分析	65
第七章 行列表资料的假设检验	70
第一节 概 述	70
第二节 四格表资料的 χ^2 检验	71
第三节 四格表资料 χ^2 检验的校正公式	74
第四节 四格表的确切概率法	76
第五节 配对四格表资料的 χ^2 检验	78
第六节 行 $\times$ 列表资料的 χ^2 检验	80
第八章 基于秩次的假设检验	84
第一节 概 述	84
第二节 单样本资料的符号秩和检验	84
第三节 配对设计资料的符号秩和检验	86
第四节 完全随机设计两独立样本的秩和检验	89
第五节 完全随机设计多个独立样本的秩和检验	92
第六节 随机区组设计资料的秩和检验	98
第九章 简单线性相关与回归	101
第一节 简单线性相关	101
第二节 简单线性回归	104
第十章 多因素线性回归与逐步回归	108
第一节 多因素线性回归	108
第二节 多因素逐步回归	112
第十一章 Logistic 回归分析	117
第一节 概 述	117
第二节 二分类 Logistic 回归	119
第三节 无序多分类 Logistic 回归	126
第四节 有序 Logistic 回归	132
第五节 条件 Logistic 回归	136
第十二章 生存分析	143
第一节 概 述	143

第二节 生存率的估计与生存曲线	144
第三节 Cox 比例风险回归模型	154
参考文献	159

第一章 SPSS 软件概述

第一节 SPSS 软件概述

一、SPSS 软件概述

社会科学统计软件包(Statistical Package for the Social Science, SPSS), 是世界上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。其用户分布于通讯、医疗、银行、证券、保险、制造、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 SPSS 17.0 为基础, 以医学领域的相关资料为例, 介绍该软件常用功能的具体使用方法。

SPSS 具有以下特点。

- ◆ 工作界面操作简便, 大部分统计分析可利用鼠标, 通过选择菜单命令、设置对话框参数和点击功能按钮来实现, 而不需要用户记住大量操作命令。

- ◆ 拥有较强的数据转换能力, 可以读取 Excel、FoxPro、Lotus 等电子表格和数据库软件产生的数据文件。

- ◆ 具备强大的程序编辑和二次开发能力, 为高级用户完成复杂的统计分析任务提供方便。

- ◆ 统计图绘制和编辑功能强大, 且增强了三维统计图的绘制功能, 图形更加美观大方。

二、SPSS 17.0 的安装

SPSS 17.0 的安装方法同 Windows 操作系统下各种应用软件的安装方法相同, 用户可以按照安装向导的提示来逐步安装。SPSS 17.0 安装界面, 如图 1-1 所示。

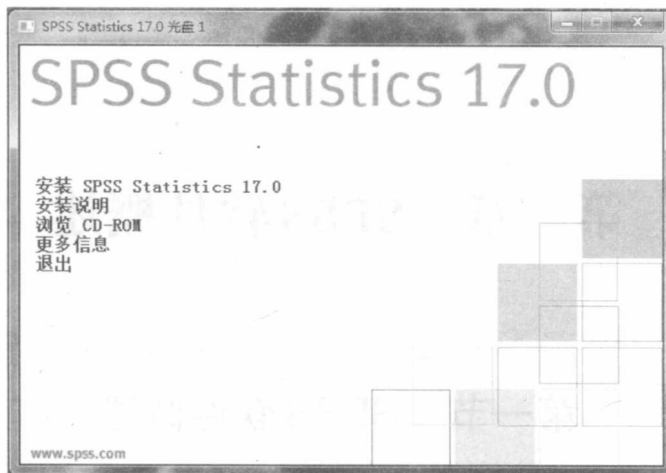


图 1-1 SPSS 17.0 安装界面

三、SPSS 的启动和退出

正确安装 SPSS 17.0 后,每次启动时,会弹出如图 1-2 所示的 SPSS 文件对话框。该对话框提供进入 SPSS 的各种方式。

Run the tutorial:运行操作指导选项,选择此项可以查看基本操作指导。

Type in data:数据窗口数据输入选项,选择此项显示数据编辑窗口,等待输入数据建立新数据集。

Run an existing query:运行一个已经存在的查询文件,选择此项用户可以选择任意一个已存在的*.spq 文件。

Create new query using Database Wizard:用数据库处理工具建立新查询文件。

Open an existing data source:打开一个已经存在的数据文件。

Open another type of file:打开其他类型的文件。

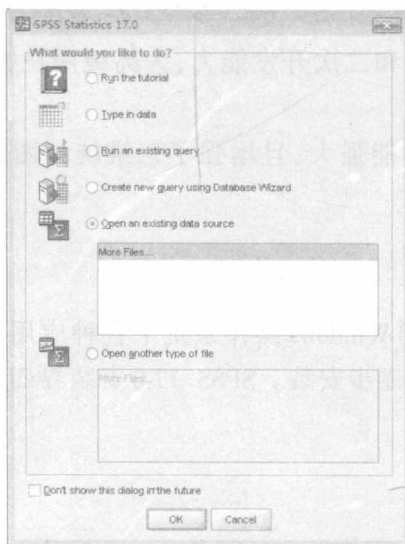


图 1-2 SPSS 文件对话框

该对话框底部有复选框 Don't show this dialog in the future,选中此复选框后,下次启动 SPSS 时将不再显示该对话框,直接进入数据编辑窗口。若直接点击 Cancel 按钮,则进入 Statistics Data Editor 数据编辑窗口,如图 1-3 所示。

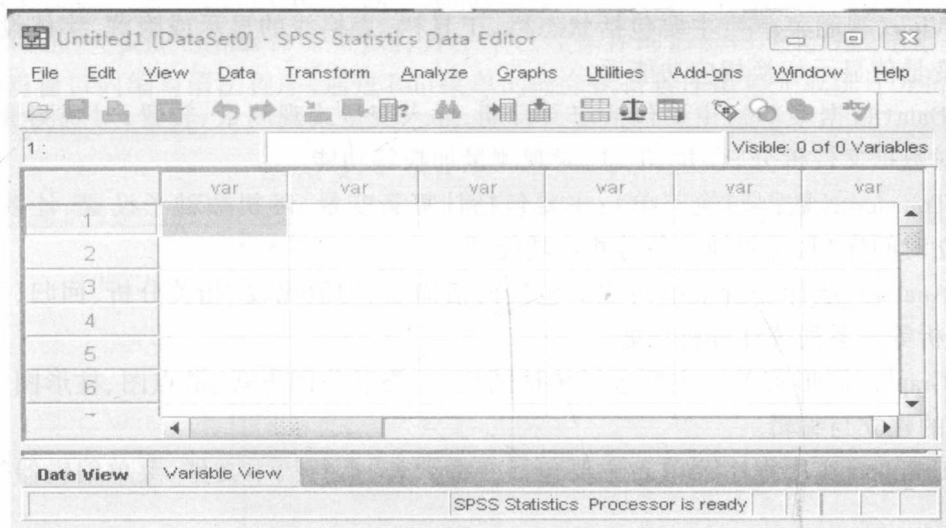


图 1-3 SPSS 数据编辑窗口

当要退出 SPSS 时,可选择下列方法中任一种。

- ◆ 打开 File 菜单,单击 Exit 选项退出 SPSS。
- ◆ 单击数据编辑器右上角的关闭按钮。
- ◆ 右击数据编辑器标题栏任何位置,从弹出的快捷菜单中选择关闭,也可以双击编辑器左上角的编辑器图标。
- ◆ 按 Alt + F4。

第二节 SPSS 的界面窗口

SPSS 的界面窗口主要有三个,分别是:数据编辑窗口(Statistics Data Editor)、结果输出窗口(Statistics Viewer)和程序编辑窗口(Statistics Syntax Editor)。

一、数据编辑窗口

在图 1-2 的 SPSS 打开对话框中,选择 Type in data 单选框,单击 OK 按钮,或者直接单击 Cancel 按钮,会直接进入数据编辑窗口,如图 1-3 所示。

数据编辑窗口是 SPSS 默认的用户启动界面,也是 SPSS 的工作界面,在这个界面中,可以建立、读取、编辑数据文件,开展相应的统计工作。

数据编辑窗口分为标题栏、主菜单栏、工具栏、状态栏、数据编辑栏和数据显示区域。SPSS 数据编辑窗口的主菜单栏集中大量的命令和功能,下面将对各个菜单项作简



要介绍。

- ◆File(文件菜单):主要包括数据文件的新建、打开、保存、退出和打印等命令。
- ◆Edit(编辑菜单):主要包括撤消、恢复、剪切、复制、粘贴、清除、查找等功能项。
- ◆View(视图菜单):主要包括状态栏、工具栏、表格线的显示或隐藏、字体设置、值标签、变量值显示切换相应功能项。
- ◆Data(数据菜单):主要包括定义日期,插入变量或观测量、转置,对观测量定位、排序,对数据文件拆分、合并、汇总,对观测量加权等功能。
- ◆Transform(数据转换菜单):主要包括计算新变量、随机数种子设置、计数、重编码、建立时间序列、重置缺失值等相应功能项。
- ◆Analyze(统计分析菜单):主要包括概括描述、均值比较、相关分析、回归、聚类与判别分析等一系列统计分析功能。
- ◆Graphs(图形菜单):主要包括统计图概述、交互作图方式、散点图、柱形图等各种统计图的建立与编辑。
- ◆Utilities(实用程序菜单):主要包括变量列表、运行稿本文件、菜单编辑器等相应功能项。
- ◆Windows(窗口控制菜单):主要包括数据窗口最小化、数据编辑窗口和 SPSS 输出窗口等的切换功能。
- ◆Help(帮助菜单):主要包括帮助主题、用户指南、SPSS 主页、统计指导和关于本软件的协议和版本信息等功能。

二、结果输出窗口

SPSS 的结果输出窗口,如图 1-4 所示。通过执行 File→New→Output 命令新建一个草稿输出窗口;执行 File→Open→Output 命令打开一个已经保存的输出文件。SPSS 的结果输出窗口除了显示统计分析结果之外,还允许用户对统计图表进一步编辑操作。

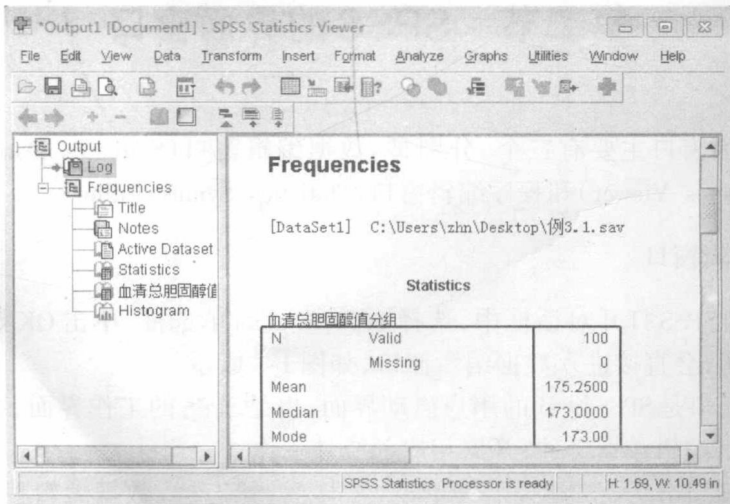


图 1-4 SPSS 结果输出窗口

三、程序编辑窗口

SPSS 在为用户提供良好的数据编辑环境和较完善的统计分析功能的同时,还提供程序编辑窗口。通过执行 File→New→Syntax 命令,可新建一个语法窗口,如图 1-5 所示;执行 File→Open→Syntax 命令可以打开一个已经保存的语法程序文件。用户可以在语法编辑窗口内编写语法程序,通过 Run 菜单的命令执行程序语句完成统计分析任务。

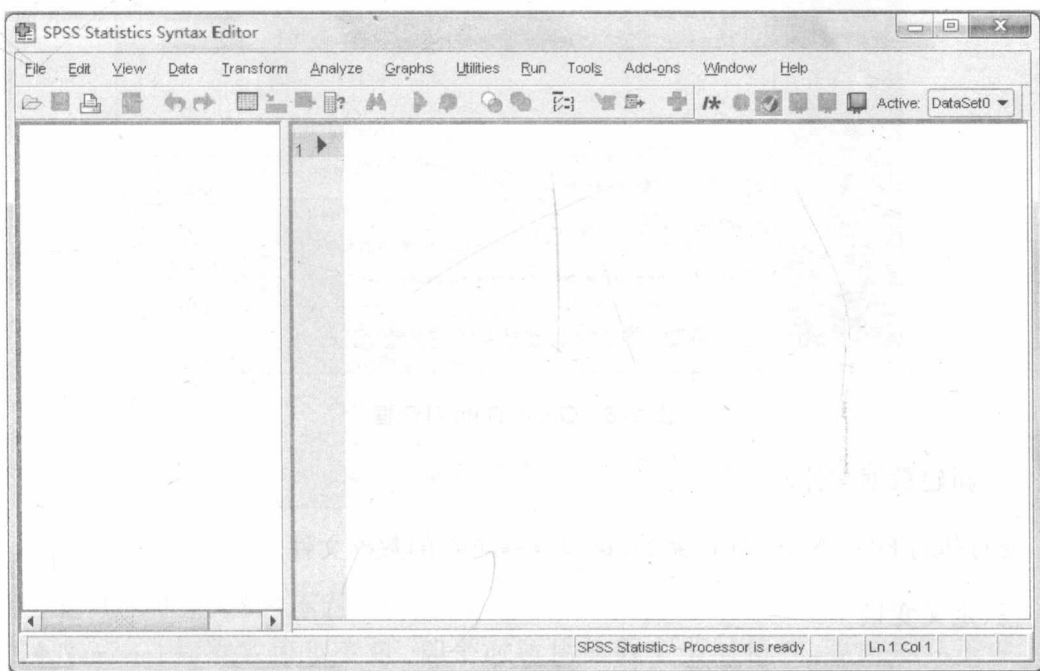


图 1-5 SPSS 程序编辑窗口

第三节 数据文件的建立

统计调查或搜集所得数据在分析之前,首先要将其输入计算机,生成相应数据文件。SPSS 数据文件的建立、编辑和整理主要是通过主菜单中的 Data(数据)菜单和 Edit(编辑)菜单相应功能项来实现的。

一、打开数据文件

通过执行 File→Open→Data 命令,弹出如图 1-6 所示的打开文件对话框,在 Files of type(文件类型)中出现 SPSS 17.0 能够读取的文件类型。能够读取的文件类型有: SPSS Statistics(*.sav)、SPSS/PC+(*.sys)、Systat(*.syd,*.sys)、Portable(*.por)、Excel(*.xls,*.xlsx,*.xlsm)、Lotus(*.w*)、Sylk(*.slk)、dBase(*.dbf)、SAS(*.sas7bdat,*.sd7,*.sd2,*.ssd01,*.xpt)、Stata(*.dta)、Text(*.txt,*.dat)。

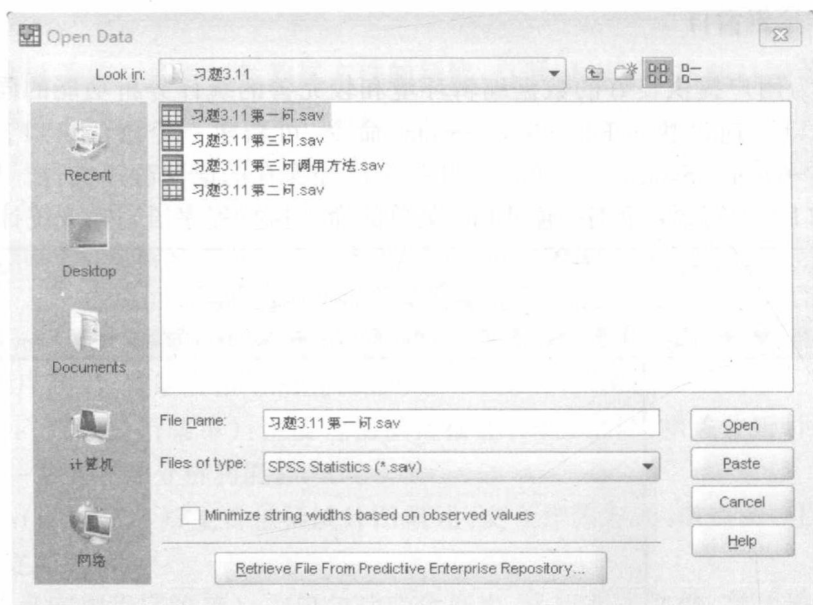


图 1-6 Open Data 对话框

二、新建数据文件

通过执行 File→New→Data 命令,可建立一个新的数据文件。

三、定义变量

在建立 SPSS 数据文件之前要定义变量,即要定义变量名、变量类型、变量长度、变量标签、缺失值、单元格宽度、对齐格式、测度水平等。

具体步骤为:单击数据编辑窗口左下角的 Variables View 选项卡,进入变量视图窗口定义变量(见图 1-7);定义变量完成之后再单击 Data View 选项卡,返回数据视图窗口输入数据。

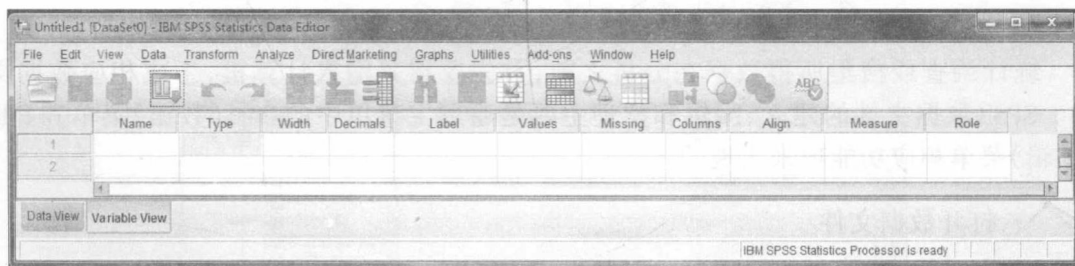


图 1-7 Variables View 窗口

1. Name(变量名)

变量名,即变量名称,定义变量的前提是为其命名。SPSS 中变量命名的具体规则为:①变量名由不多于 8 个字符组成,若定义的变量名中的字符个数大于 8 个,将会自动截去尾部而不完全显示;②首字符必须为字母,其后为字母、符号或数字,同时也可

以使用汉字作为变量名,但是如“?”、“!”、“*”以及算数运算符是不允许使用的,“.”不能作为变量名的最后一个字符;③不能使用SPSS保留字:ALL、AND、BY、EQ、GE、GT、LE、LT、NE、NOT、OR、TO、WITH;④系统不区分大小写字母。

2. Type(变量类型)

SPSS变量有3种基本类型:Numeric(数值型)、Data(日期型)和String(字符型),如图1-8所示。

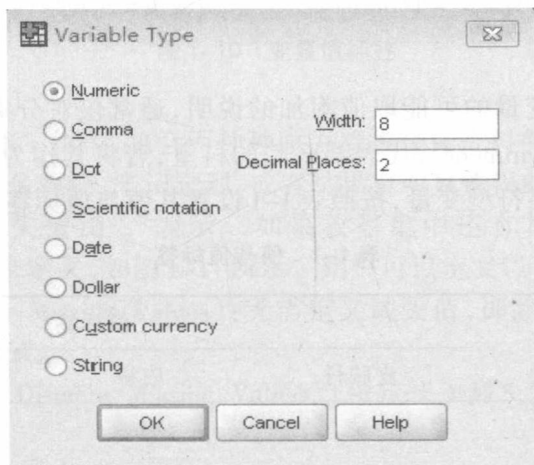


图 1-8 Variable Type 对话框

变量具体可分为8种不同类型,主要为:

◆Numeric:标准数值型变量,即平时所用的数字书写格式,系统默认变量宽度(Width)为8,小数点位数(Decimal Places)为2。

◆Comma:带逗点型的数值变量,即变量值的整数部分从右向左每三位加一个逗点作分隔符,用圆点作小数点。如1,000.00。

◆Dot:带圆点型的数值变量,即变量值的整数部分从右向左每三位加一个圆点作分隔符,用逗点作小数点。如1.000,00。

◆Scientific Notation:科学计数法型变量。如10,000的科学计数法为1.00E+04。

◆Data:日期型变量。

◆Dollar:带美元符号\$型变量。

◆Custom Currency:自定义型变量。

◆String:字符型变量,即一串字符,在输入时系统对大小写字母是严格区分的,系统默认字符长度为8。

Data(日期型变量)和String(字符型变量)是不允许参与运算的。

3. Width(变量格式宽度)

变量格式宽度,指变量列在数据窗口中占据的单元格的列宽度。在定义变量格式宽度时要满足“变量格式宽度 $\geq$ 变量宽度”且“变量格式宽度 $\geq$ 变量名长度”。如果变量宽度 $>$ 变量格式宽度,此时数据窗口中显示变量名的字符数不够,变量名被截去尾部

不能完全显示,输入的数据可能截去尾部,被截去部分用“*”代替。

4. Decimals(小数点位数)

小数点位数指输入变量的数值在数据窗口中显示时的小数点位数。

5. Label(变量标签)

变量标签是对变量名的附加说明,利用其实现对变量的意义作进一步解释和说明。给变量加了标签后,在数据窗口中,当鼠标指向变量的时候,变量名下即可显示其标签。

6. Values(值标签)

变量值标签是对变量的可能取值附加的说明,通常仅对分类变量指定值标签。例如,定义一个变量“Department”,代表某医院的科室,若将其作为分类变量统计分析,可将其定义为数值型或字符型变量,按照表 1-1 设置其值与值标签。

表 1-1 值与值标签

变量类型的值	值标签			
	皮肤科	内科	外科	...
字符型变量值	a	b	c	...
数值型变量值	1	2	3	...

使用值标签的目的是简化录入的数据,减小工作量,提高工作效率。其操作方法:

单击 Variable View 选项卡,先点击 Type,选择 Numeric,再点击 Values,在 Values Labels 对话框中,Value 栏填入数值,Label 栏填入与数值相对应的标签,单击 Add,如图 1-9 所示;单击 Data View 选项卡,在数据编辑窗口输入数值,如图 1-10 所示。

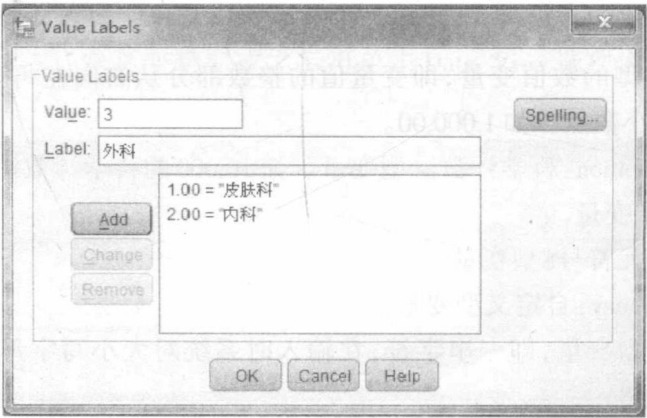


图 1-9 Value Labels 对话框

或单击 Variable View 选项卡,先点击 Type,选择 String,再单击 Data View 选项卡,返回数据编辑窗口输入标签(见图 1-10)。

4:			13:		
	Department	Value		Department	Value
1	1		1	a	
2	2		2	b	
3	3		3	c	
4			4		
5			5		

(a) 数值型变量标签

(b) 字符型变量标签

图 1-10 变量值标签

7. Missing(缺失值)

在统计调查的具体工作中,总会因种种原因造成统计资料的缺失、遗漏和差错。统计中,把在调查过程中未观测、未记录到,或者结果有明显错误的数据称为缺失值。

SPSS 中默认的缺失值用“.”表示。如果数据集中还有其他表示方法,则是用 Missing Values 对话框来定义,如图 1-11 所示。用户可以定义以下 3 种缺失值。

◆ 没有缺失值(No Missing Values):无需定义缺失值,即除了默认的缺失值外,不设缺失值,这是默认方式。

◆ 离散的缺失值(Discrete Missing Values):可定义 1 到 3 个离散的单一数值为缺失值。

◆ 范围加离散的缺失值(Range Plus One Optional Discrete Missing Value):定义某一指定范围为缺失值,同时指定一个不在这一范围内的离散单一数值为缺失值。



图 1-11 Missing Values 对话框

8. Align(单元格对齐方式)

数据窗口中,变量在单元格显示时有左对齐和右对齐之分,一般情况下,系统默认数值型变量的对齐方式为右对齐,字符型变量的对齐方式为左对齐。用户可根据具体情况在 Align 中设置对齐方式。

9. Measure(测度)

统计学中所谓的测度是指按照某种法则来给事物、事件等赋予的一定的数字或符号。测度是用来刻画事物的特征和属性的。统计学通常将测度分为三种:Scale(定比测度)、Ordinal(定序测度)和 Nominal(定类测度)。