SPSS 10.0
CHANGYONG

SHENGWUYIXUE
TONGJISHIYONGZHIDAO

SPSS 10.0 常用

生物医学统计使用指导

主编　李湘鸣　王劲松

东南大学出版社
SOUTHEAST UNIVERSITY PRESS

扬州大学出版基金资助

SPSS 10.0 常用生物医学统计使用指导

主　编　李湘鸣　王劲松
副主编　唐　尧　孙　峰
　　孙　蓉　陈秀云

东南大学出版社

图书在版编目(CIP)数据

SPSS10.0常用生物医学统计使用指导/李湘鸣,王劲松主编.
南京:东南大学出版社,2005.3
ISBN 7-81089-869-8

Ⅰ.S... Ⅱ.①李...②王... Ⅲ.①生物统计-统计分析-软件包,SPSS-高等学校-教学参考资料②医学统计-统计分析-软件包,SPSS-高等学校-教学参考资料
Ⅳ.①Q-332②R195.1-39

中国版本图书馆CIP数据核字(2005)第013956号

东南大学出版社出版发行
(南京四牌楼2号 邮编210096)
出版人:宋增民
江苏省新华书店经销 南京京新印刷厂印刷
开本:787 mm×1092 mm 1/16 印张:10.75 字数:256千字
2005年3月第1版 2005年3月第1次印刷
印数:1~3000册 定价:18.00元

(凡因印装质量问题,可直接向发行部调换。电话:025-83795801)

前　言

美国大型统计软件SPSS 10.0具有多种统计分析功能，可满足许多行业的统计需要和要求。但是，每个人不可能也不需要将它的所有统计功能全部掌握，因此本书着眼于大多数从事生命科学学习研究的工作者所经常遇到的科研资料，就如何使用SPSS软件进行处理给予介绍，使其结果更加科学、严谨。

目前市场所流行的SPSS软件多为英文菜单操作，涉及很多术语，用普通汉化软件汉化有一定的困难，而且即使使用某些汉化软件对其进行了汉化，其中文意思也使人难以理解。本书对医学统计常用的某些英文菜单命令给予中文解释，并且给出具体的例子，使其更加简单易懂，起到举一反三的作用。

本书出发点主要以实例为主，具有以下优点：(1) 使缺乏统计知识但具一定计算机知识者，根据自己资料的性质，进行模拟操作，进而得出可靠的统计结果；(2) 可使具有一定统计知识及计算机知识者，进行模拟练习，进一步熟练掌握该软件。书中实例均来自实际的医学、生物学科学研究中的课题。本书主要供生物医学院校各专业的研究生、本科生在论文写作时进行统计数据之用，也可供青年教师、临床医生学习及统计处理科研数据使用，还可供综合性大学、农林院校及师范院校有关专业的师生学习参考。

总而言之，计算机不断换代，软件不断更新，但万变不离其宗。学习计算机软件除了需要阅读一定的参考书外，更重要的是练习。只有通过反复不断地练习、虚心与他人交流、仔细琢磨其涵义及规律性，才能更好地掌握计算机软件。通过练习来学习SPSS，这是作者的最终目的。由于SPSS是以数据库为基本框架结构，各章节之间自然会出现数据共享或统计模块共用的可能，因此本书采取由细到粗的写作原则，初学者最好能认真学习好第一、二章的内容，并抽时间学一点数据库方面的知识。

目　录

第一章　概　　述

一、运行环境

SPSS 10.0 要求以 Windows 操作系统为平台，计算机内存至少 8 M。为了充分提高其软件的性能，产生良好的输出效果，建议最好选用 PⅡ以上机型，内存 32 M 以上，显示器分辨率最好不少于 800×600 像素(pixels)，显卡最好选用 SVGA 显卡，可支持 256 色。

二、安装与启动

将有 SPSS 10.0 的软件置入光驱，进入 SPSS 10.0 文件夹，用鼠标双击 Setup 命令，便弹出一对话框，单击 Install SPSS，便开始运行 SPSS 安装程序。然后根据屏幕提示，键入公司密码，便可完成安装。安装完成后，单击 Windows 下的“开始”按钮，将光标移到“程序”选项上，便出现“SPSS for Windows”菜单，在下拉式“SPSS 10.0 for Windows”及“SPSS 10.0 Production Facility”菜单中，选中“SPSS 10.0 for Windows”便可启动 SPSS 10.0(图 1.1)。

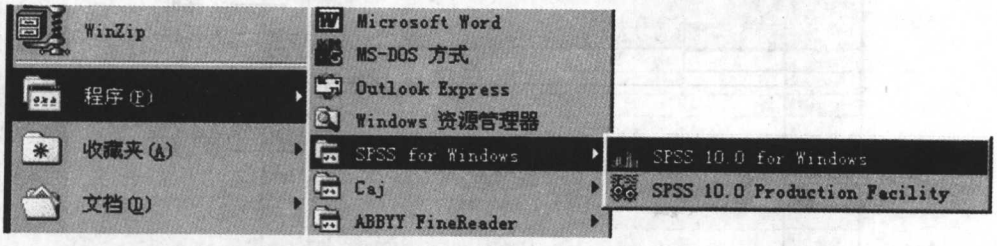

图 1.1　启动 SPSS 10.0 的方法

SPSS 10.0 启动后，产生图 1.2 对话框。常用的四个命令菜单为：Type in data(输入数据)、Open an existing data source(打开已存在的数据文件)、Run the tutorial(运行指导)和 Open another type of file (打开其他类型的文件)。用鼠标选中相应命令，便可进入有关操作。最常用的方法是点击 Cancel(取消)按钮，直接进入 SPSS 视窗(图 1.3)进行相关操作。

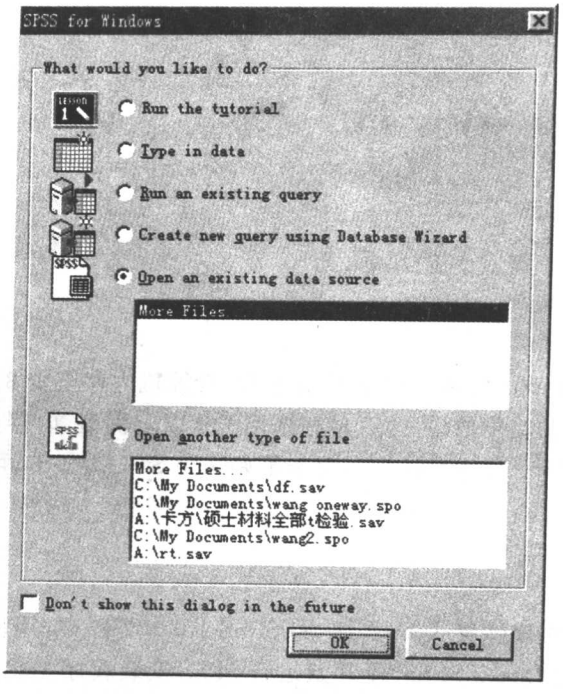

图 1.2　SPSS 10.0 启动后所弹出的对话框

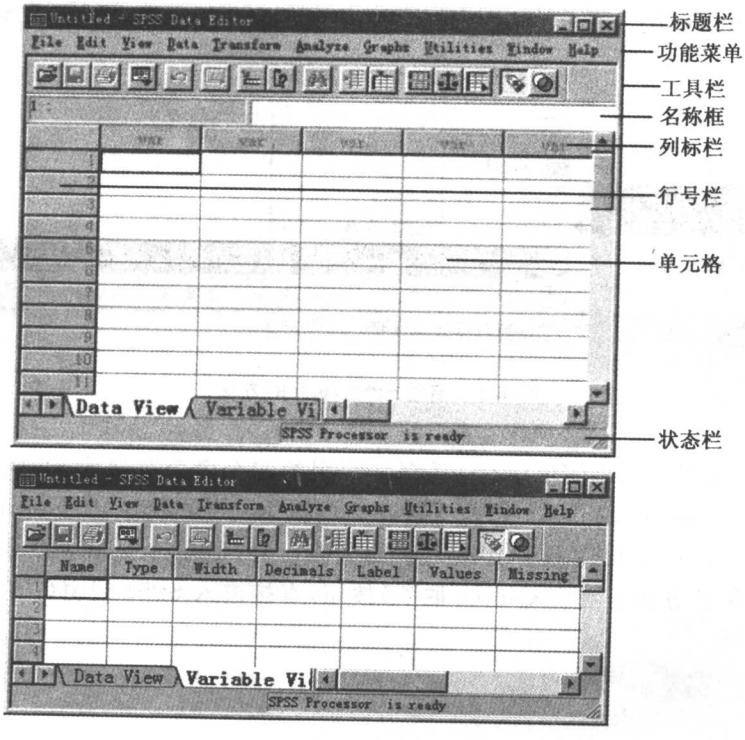

图 1.3　数据查看(Data View)和变量查看(Variable View)工作表

三、视窗介绍

SPSS 10.0 视窗分为两个工作表：数据查看(Data View)工作表和变量查看(Variable View)工作表。视窗最上方是标题栏。标题栏下方是功能菜单，分别为 File(文件)、Edit(编辑)、View(查看)、Data(数据)、Transform(转化)、Analyze(分析)、Graphs(制图)、Utilities(应用)、Window(窗口)和 Help(帮助)。点击功能菜单便可出现一个下拉式菜单，在其中可选取相应的命令。在功能菜单的下方是工具栏，当鼠标指针指向相应图标时，便会自动显示工具按钮的名称，按鼠标左键便会快速执行该命令。工具栏的下方是名称框，左边用以显示所选定单元格的名称，右方用以输入变量中数据。再下方是工作表，在工作表中可以编辑各种数据，其上方为列标栏，左边为行号栏。在列标栏中可选相应的变量名进行编辑。工作表最下方是状态栏，显示当前执行的命令或提示的信息。同时工作表旁边还提供了滚动条，翻动滚动条可以查看工作表的不同部分(图 1.3)。

数据查看(Data View)工作表主要用于查看资料中各变量的内容。变量查看(Variable View)工作表主要有以下功能：命名(Name)变量；定义变量的类型(Type)、宽度(Width)以及小数点的位数(Decimals)；由于 SPSS 规定变量名的最大长度不得超过 8 个字符(4 个汉字)，当变量名较长时可对其进行标记(Label)。关于两个工作表的详细功能请读者在下面的章节中认真体会。

第二章　数据库的结构

一、定义变量

当输入数据时，应当告诉计算机当前输入数据的名称、大小、类型及性质。现通过表 2.1的例子加以说明。

表 2.1　某医生测得病人血清胆固醇结果记录

病人姓名	生活所在地	血清胆固醇(mg/100 ml)
王水生	大城市	178
李焱	中小城市	148
金光辉	农村	157.2
梁凯歌	乡镇	150.4
↓	↓	↓

启动 SPSS 进入 SPSS 工作表，首先单击 Variable View 进入变量查看视窗(图 2.1)，将鼠标置入 Name 列中单元格内，依次输入(命名)三个变量："病人姓名"、"所在地"和"胆固醇"。此步是对所涉及的变量进行命名。命名目的只是为了在数据处理时有助于记忆，可根据资料的具体情况进行命名，以简明扼要、便于记忆为原则，亦可在符合 SPSS 规定的前提下，随心所欲。

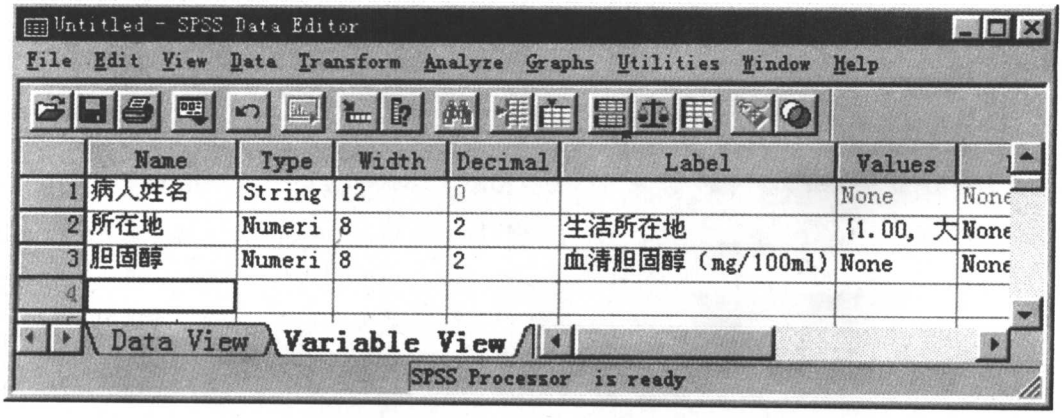

图 2.1　某医生测得病人血清胆固醇结果的变量定义方式

在变量“病人姓名”中，由于要输入每个病人的姓名，所以必须对此变量的类型(Type)进行定义。操作如下：用鼠标单击 Type 列中相应的单元格，则会弹出一对话框(图 2.2)，选中 String(字符串变量)。由于中国人的姓名多数不会超过 6 个汉字(12 字符)，可在 Characters(字符)框内键入 12，然后单击 OK 按钮，便完成对变量“病人姓名”的定义工作。SPSS 可将字符串最大宽度设定为 99 个字符。

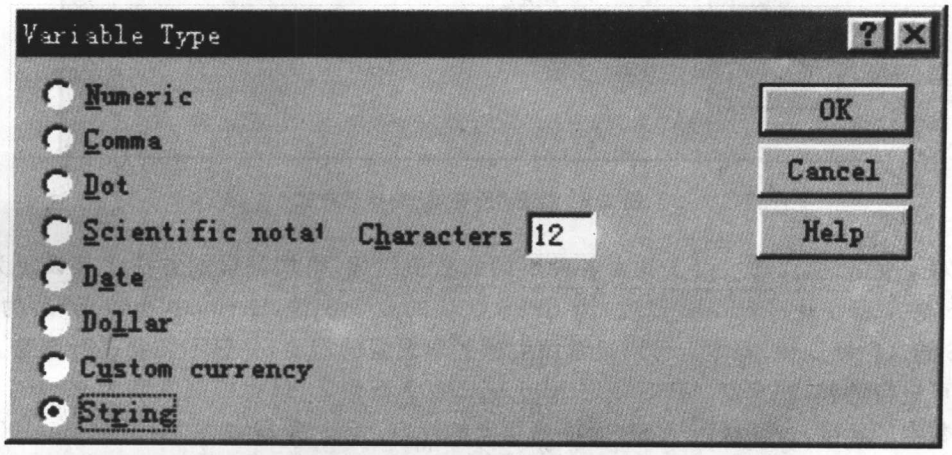

图 2.2　变量“病人姓名”的定义方式

在变量“所在地”中，设计者人为地将病人的所在地分为“大城市”、“中小城市”、“农村”和“乡镇”。如果将变量“所在地”定义为字符串变量，当输入每个病人的资料时，就会对“大城市”、“中小城市”、“农村”和“乡镇”出现重复输入，这显然较麻烦。对这种类型的变量，SPSS 提供了方便的“数值标记(Value Labels)”功能，即可用数值 1 代表“大城市”，用 2 代表“中小城市”，用 3 代表“农村”，用 4 代表“乡镇”。具体操作如下：

(1) 用鼠标单击变量“所在地”右边 Type 列中相应的单元格，则会弹出一对话框(图 2.2)，选中 Numeric(数字型变量)，然后单击 OK 按钮，将变量“所在地”定义为数字型变量(这通常是 SPSS 的默认方式，故可省略此操作)。

(2) 在其相应的 Label 列单元格中输入“生活所在地”，对变量“所在地”进行标记。由于 SPSS 规定变量名长度不得超过 8 个字符(4 个汉字)，无法建立变量“生活所在地”，为了便于以后的操作中不会忘记变量的属性，可对变量用 Label 命令进行标记。此步操作主要是为了方便记忆，对输出结果不会产生影响。如果读者对所命名变量了如指掌，可省略此步操作。

(3) 用鼠标单击 Values 列中相应的单元格，则会弹出一对话框(图 2.3)，在第一个 Value 右边的框内键入某一数值(如 1)，在第二个 Value 右边的框内输入该数值所代表的内容(如大城市)，每次按“Add”按钮，依次将所标记的代码(数值)及内容置入右边大框内，然后按“OK”按钮，即完成对变量“所在地”的“数值标记(Value Labels)”工作。

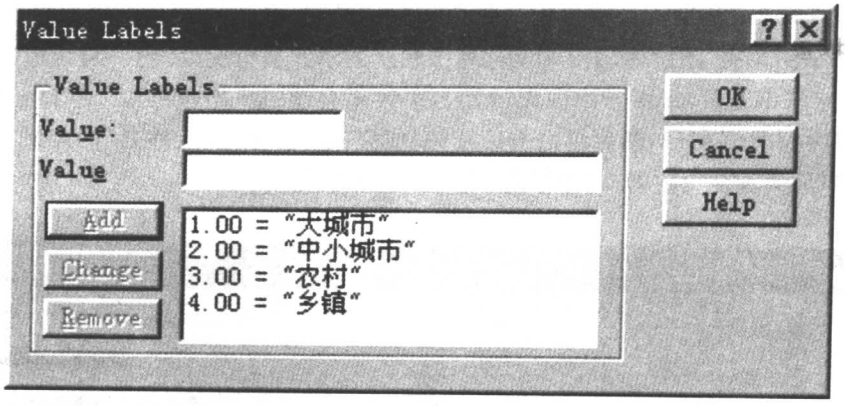

图 2.3　变量"所在地"的标记方式

对于变量"胆固醇",由于其由各种不同的数值组成,可按同样的方法将其设定为 Numeric(数字型变量),Width(宽度)为 SPSS 默认的 8 个字符,Decimal(小数点)后为 2 位。SPSS 的这种默认值,基本上可满足常用生物医学数据的需要,一般不需要进行调整。如果需要可进行调整,如宽度,SPSS 规定 Numeric(数字型变量)最大可调宽度为 99 个字符(包括小数点的宽度)。同样由于 SPSS 变量长度的规定,无法建立"血清胆固醇(mg/100 ml)"变量,为了便于记忆,可在相应的 Label 单元中,键入"血清胆固醇(mg/100 ml)"对变量"胆固醇"给予标记(图 2.1)。

完成对变量的命名及定义后,用鼠标点击工作表下方的 Data View(数据查看)命令,便进入数据操作工作表,在此工作表中可输入各自变量中的相应数据(图 2.4)。如果将鼠标指针移向相应变量,便显示出所注明内容(如"血清胆固醇(mg/100 ml)");当用鼠标点击此按钮(Values Labels),便会显示出数值所标记的内容(读者体会一下)。

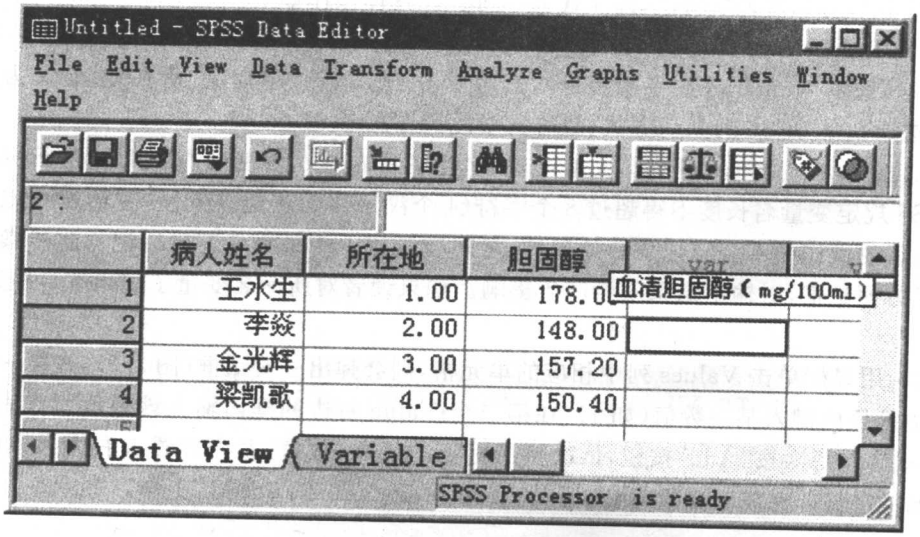

图 2.4　某医生测得病人血清胆固醇输入方式

二、数据库定义

由于数据库是个复杂的系统，难以用简单的语言对其下一个恰当的定义。但是，它主要由变量（属性，或称字段）、记录和观察值（Case，或称变量值，或称属性值）组成（图 2.5）。每一个记录中的变量值具有一定的相互关系、相互联系，彼此说明。

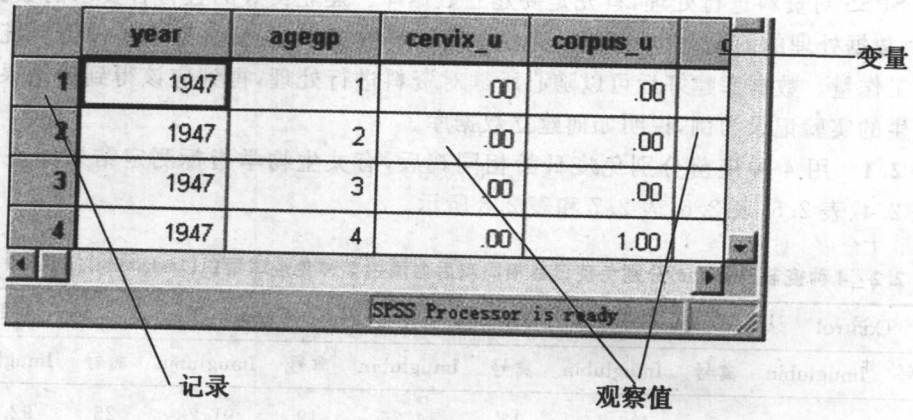

图 2.5　数据库的基本组成

变量：观察值（Case、变量值）的某些特征，如年龄、性别、职业、身高等的集合。在生物医学研究中，常用两种变量类型来描述资料：数值型变量和字符型变量。在 SPSS 中分别用 Numeric（数字型）变量和 String（字符串型）变量来表示两种变量的类型。凡是用定量方法或是用一定的单位获取或表示的资料，如身高、体重、血压、脉搏、酶活性、年龄及各种汇总所得率（或比）等，其变量类型必须设定为 Numeric（数字型）变量；凡是由一定的属性所组成的资料，如姓名、性别、家庭住址、家庭状况、等级资料等，也可以这样说，凡是要在单元中直接输入中文或字母变量，必须将该变量设定为 String（字符串型）变量，这种资料也称为归类变量资料。

三、建立数据库时应注意的问题

1. 不要破坏资料的原始性。只要能按数据库的要求输入自己所得资料，而不用顾及何者为先何者为后，可以随心所欲地整理资料，千万不可人为地破坏资料的原始性，一旦破坏很难恢复。资料的原始性是进行正确统计分析的前提，这一点应特别注意。

2. 不要将毫无关系的观察值（Case、变量值、属性值）输在同一个记录中。既然是数据库，它必然有一定要求（结构），输入资料时，应稍动动头脑，按照一定的格式输入，不可根据自己的想象输入数据，因为计算机是根据一定的格式（程序）运行的。

3. 最好以代码的形式建立数据库（输入资料）。众所周知，计算机是美国人发明的，最基本的操作系统是西（英）文命令，而中文等操作系统是在此基础上建立起来的。因此，为了保证数据库资料有较好的兼容性，而不会在相互调用时产生乱码，最好能用英文或阿拉伯数

字代替中文输入，例如，可用“1”代替“有效”，“2”代替“无效”；“y”代替“好转”，“n”代替“未见好转”。只要读者心中明白各自所代表的“内容”，做好有关的记录，就会使该数据库操作起来得心应手。

四、数据库的建立

用 SPSS 对资料进行处理，首先是要建立数据库。建立良好的且符合要求的数据库，是 SPSS 对数据处理的前提。也可以说，建立了某资料数据库，也就是完成了该资料统计处理 80%的工作量。数据库建好后可以随心所欲对资料进行处理，得到应该得到的结果。现以某研究生的实验记录为例，说明如何建立数据库。

例 2.1 用 4 种疫苗分别免疫日龄相同鸡后，有关生物学指标测定结果如表 2.2、表 2.3、表 2.4、表 2.5、表 2.6、表 2.7 和表 2.8 所示。

表 2.2 4 种疫苗(Vaccine)分别免疫日龄相同鸡后血清中某种免疫球蛋白(Imuglubin)(μg/100 ml)

Control		Vaccine A		Vaccine B		Vaccine C		Vaccine D	
禽号	Imuglubin	禽号	Imuglubin	禽号	Imuglubin	禽号	Imuglubin	禽号	Imuglubin
1	90.41	7	86.16	13	84.25	19	91.22	25	92.75
2	83.50	8	87.56	14	92.63	20	92.91	26	76.63
3	90.69	9	90.19	15	90.25	21	89.78	27	88.59
4	91.81	10	87.03	16	91.31	22	90.19	28	91.41
5	87.47	11	90.03	17	81.75	23	90.44	29	90.94
6	89.94	12	91.38	18	83.28	24	91.72	30	84.03

表 2.3 对照组鸡体重变化(g)

禽 号	第 1 周	第 3 周	第 5 周
1	29	23	25
2	27	22	20
3	30	35	38
4	22	30	32
5	15	24	27
6	21	35	36

表 2.4　疫苗 A 免疫日龄相同鸡后体重变化(g)

禽　号	第 1 周	第 3 周	第 5 周
7	23	25	13
8	26	30	35
9	20	25	20
10	20	26	21
11	16	19	22
12	20	30	21

表 2.5　疫苗 B 免疫日龄相同鸡后体重变化(g)

禽　号	第 1 周	第 3 周	第 5 周
13	21	25	27
14	28	30	31
15	25	26	28
16	15	35	30
17	18	19	23
18	24	35	30

表 2.6　疫苗 C 免疫日龄相同鸡后体重变化(g)

禽　号	第 1 周	第 3 周	第 5 周
19	26	21	30
20	14	16	20
21	18	27	24
22	22	25	30
23	18	20	23
24	24	30	28

表 2.7　疫苗 D 免疫日龄相同鸡后体重变化(g)

禽　号	第 1 周	第 3 周	第 5 周
25	27	26	28
26	15	25	26
27	21	29	24
28	26	36	30
29	23	40	35
30	26	35	40

表 2.8　4 种疫苗分别免疫日龄相同鸡后实验终点时死亡情况

禽　号	死亡情况	禽　号	死亡情况	禽　号	死亡情况	禽　号	死亡情况	禽　号	死亡情况
1	alive	7	alive	13	died	19	died	25	died
2	alive	8	died	14	alive	20	alive	26	alive
3	alive	9	alive	15	alive	21	died	27	alive
4	alive	10	died	16	alive	22	alive	28	died
5	alive	11	died	17	died	23	alive	29	alive
6	alive	12	alive	18	alive	24	died	30	alive

从上述科研记录来看，有 30 只动物(鸡)，编号(num)从 1 到 30，被分为 5 组(对照组、疫苗 A、疫苗 B、疫苗 C、疫苗 D)，分别获得每个个体的免疫球蛋白含量及第 1、3、5 周的体重变化以及在实验结束时实验动物的死亡状况。现在就用该资料对如何建立数据库作一叙述。

启动 SPSS 进入工作表后，首先点击 Variable View 进入变量查看工作表，在变量命名(Name)列中输入(定义)5 个变量名，分别为 num(表示鸡的编号)、glubin(表示免疫球蛋白)、group(表示受试动物所处的组)、dl(表示实验结束时实验动物的死亡状况)、bw1(表示受试动物第 1 周的体重)、bw3(表示受试动物第 3 周的体重)和 bw5(表示受试动物第 5 周的体重)，并将所有的变量类型(Type)定义为 Numeric(数字型变量)，这种变量类型通常为 SPSS 默认方式，只要用鼠标在任意单元格内单击一下，便会在 Type 栏的相应单元内显示 Numeric。为了有助于记住各变量所代表内容，可在 Label 列(栏)内对相应的变量进行标记，如对“glubin”标记为“Imuglubin(μg/100ml)”，对“dl”标记为“died or alive in end”……(图 2.6)。

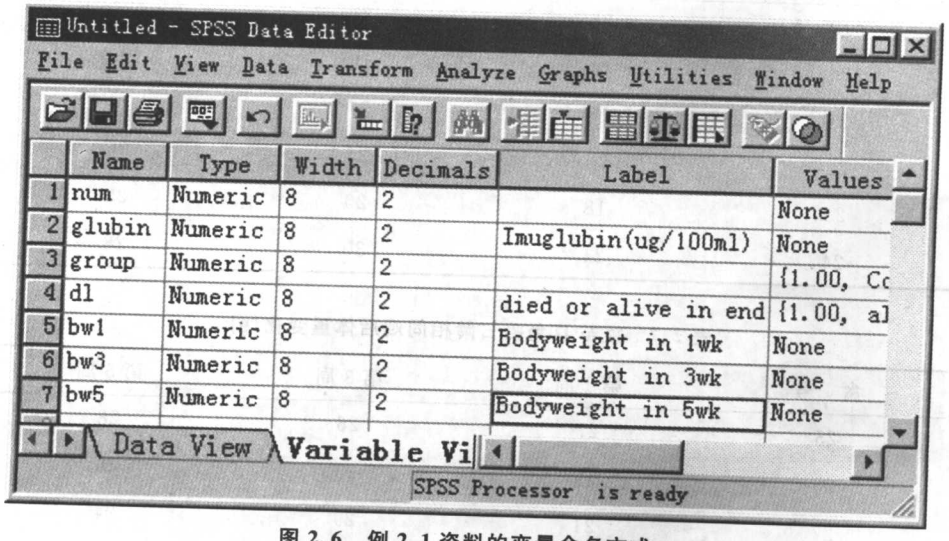

图 2.6　例 2.1 资料的变量命名方式

为了输入方便，对变量“group”中的观察值(Case)可进行标记，用 1 代表 Control(对照组)，用 2 代表 Vaccine A(接种 A 疫苗)，用 3 代表 Vaccine B(接种 B 疫苗)用 4 代表 Vaccine C(接种 C 疫苗)，用 5 代表 Vaccine D(接种 D 疫苗)；对变量“dl”中两种属性进行标记，用 1 表示 alive(存活)，用 2 表示 died(死亡)，图 2.7 和图 2.8 所示。

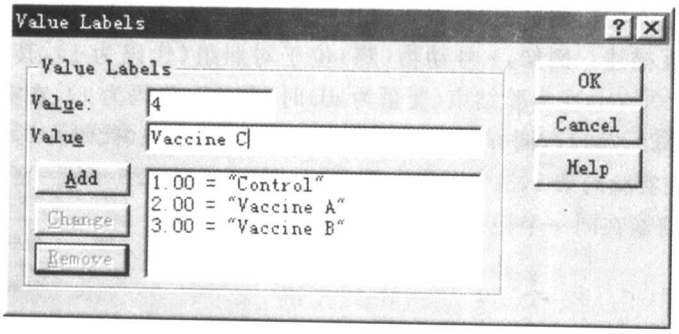

图 2.7　变量“group”中变量值的标记方式

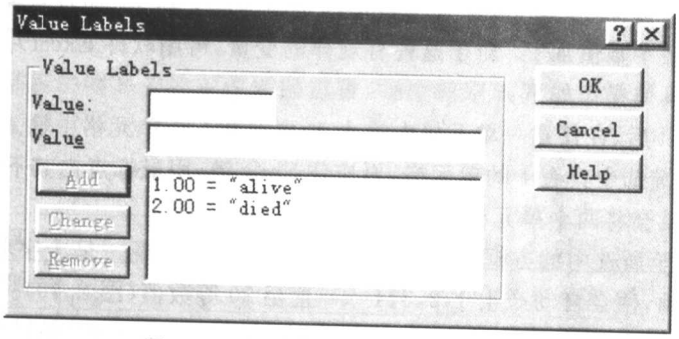

图 2.8　变量“dl”中变量值的标记方式

完成对上述变量的命名后，可点击 Data View 进入数据查看工作表，在各自变量的相应的单元格内，输入各自的应有的数值或代码(图 2.9)。

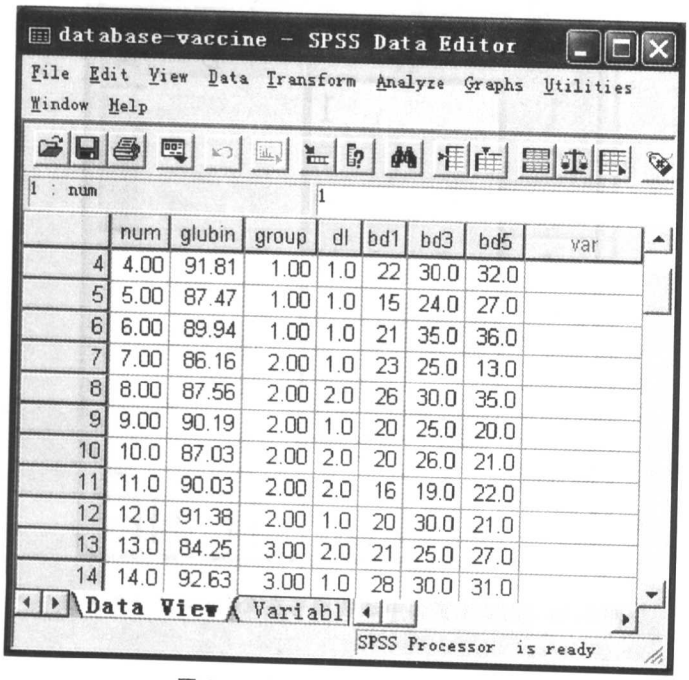

图 2.9　例 2.1 资料的输入方式