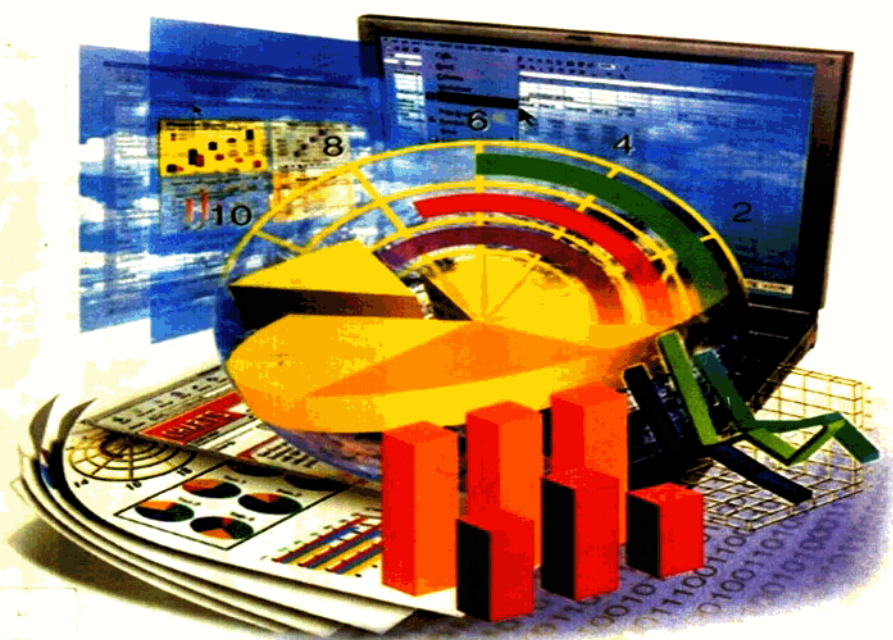


SPSS 10.0

医学统计方法与应用

上册 · 基础统计

主编 刘润幸 副主编 邝建



广东人民出版社

SPSS 10.0
医学统计方法与应用
(上册)
基础统计

刘润幸 主编 邝 建 副主编

刘润幸 邝 建
林汉生 杨 丽
王加辉 蔡 玲

广东人民出版社

图书在版编目(CIP)数据

SPSS10.0 医学统计方法与应用/刘润幸主编. —广州:
广东人民出版社, 2001.4

ISBN 7-218-03570-1

I. S... II. 刘... III. 医学统计-统计分析-软件包, SPSS10.0 IV. R195.1-39

中国版本图书馆 CIP 数据核字(2001)第 10025 号

出版发行	广东人民出版社
经 销	广东新华发行集团股份有限公司
印 刷	湛江日报印刷厂
开 本	787 毫米×1168 毫米 16 开本
印 张	32.75
字 数	650000 字
版 次	2001 年 4 月第 1 版 2001 年 4 月第 1 次印刷
印 数	0001-6000 册
书 号	ISBN7-218-03570-1/R·98
定 价	58.00 元(上下册)

如发现印装质量问题,影响阅读,请与承印厂联系调换。

售书热线:(020)83790667 83791084



刘润幸, 1991年在广州暨南大学获卫生统计学硕士学位。1996年晋升为副教授。专业特长: 卫生统计学, 流行病学与相关的计算机软件应用。

曾从事的厂研究项目:

①慢性病危机因素的研究, 如: 女性乳腺癌的危险因素, 心理应激与癌的相关, 冠心病社会心理因素等。②数学模型的应用研究, 如: 通径分析, Bootstrap法的重抽样, Logistic回归, 岭回归, Cox回归, 主成分回归, 等级回归。③SPSS统计软件的开发研究。④医学统计学与SPSS统计; 软件的计算机辅助教学研究。

前 言

SPSS 10.0 for Windows 是目前 SPSS 系列中的最高版本,它在 SPSS 9.0 版本的基础上进行了改善和增添,与之相比拥有以下的新特点:

①数据访问的改善:无需临时外部磁盘空间就可分析大型数据文件。②关系数据库界面的改善:非常容易和有效地从关系数据库中提取数据进行分析。③分布式分析:通过使用遥控式服务器,计算机执行数据的精确计算,改善分析速度。④多对话期:可以同时分析多于一个数据文件。⑤直接访问 Excel:可以读取 Excels 5 以上的版本。⑥新式数据编辑器:新颖的变量表非常直观,可以更容易定义和描述变量及其标签。⑦图形输出的改善:图形可以复制为 Windows 的亚文件(metafile)。⑧Polytomous Logit 通用模型:可以把回归技术应用在等级结果中。⑨Thematic 地图:按地理区域图形化综合数据。⑩ROC 曲线的多重检验变量:ROC 曲线过程比较多重检验变量。⑪Logistic 和 COX 回归输出的改善:产生关键结果的表格输出。⑫新式优化 scaling 过程:产生非线性主成分分析过程。

SPSS 10.0 医学统计方法与应用共分上下两册。

(1)上册介绍基础统计部分:SPSS 的基本操作,统计报表,描述性统计和探索性分析,T 检验,单因素方差分析,GLM 一单变量过程(双因素与协方差分析,线性回归),二项与 Poisson 分布,列联表分析,非参数检验,相关分析,线性回归分析,曲线拟合,生存分析(寿命表与 Kaplan-Meier 法),半数效量和统计图等。它适合于医学院与相关学科的本科生,研究生,教师,临床医生和科研人员等使用。

(2)下册介绍高级统计部分:GLM 一多变量与重复数据过程,Logistic 回归(二项与多项分类 Logistic 回归),等级回归,权重估计和两阶段最小二乘法回归,对数线性模型(广义与可选对数线性模型),聚类分析,判别分析,数据简化分析(主成分与因子分析),可靠性分析,和 Cox 回归(时间-常数与时间-依赖模型)等。它适合于预防医学系与相关学科的研究生,进一步提高统计水平的教师,科研人员和专业统计人员等使用。

本书统计部分的写作模式与 SPSS 8.0 for Windows 统计软件使用指南相似,基本分为三部分:①一般概念,简述某种统计方法的基本概念和适用条件,统计过程的主要功能。②统计过程的对话框与选择项,简述统计过程的各种对话框和选择项的功能,方便读者查阅。③例题,通过举例,详细地介绍 SPSS 工作数据表的数据排列,视窗操作,运算结果和分析。

本书的例题全部来源于参考文献和科研数据,我们对每一位参考文献的作者所付出的辛勤劳动,以及香港城市大学的萧灿培博士和香港 SGS 有限公司的郑景辉先生给予大力支持表示衷心的感谢。由于我们的知识水平有限,错误之处在所难免,恳请读者给予批评指正。

刘润幸

2000 年 10 月于暨南大学

目 录

第一章	数据库建立和结果输出	1
1.1	运行环境与操作说明	1
1.2	视窗界面	2
1.3	数据库建立	7
1.4	数据编辑	15
1.5	数据库保存、转换和读取	17
1.6	结果输出	19
1.7	数据和结果的打印	25
1.8	SPSS 系统选项设定	27
第二章	数据整理	28
2.1	一般概念	28
2.2	插入或删除变量或个体	28
2.3	个体排序	30
2.4	行列转置	31
2.5	合并文件	33
2.6	选择个体	37
2.7	加权个体	41
第三章	数据变换	42
3.1	一般概念	42
3.2	建立新变量	42
3.3	重置代码	47
3.4	计数	49
第四章	报表	53
4.1	实时分析处理立体表	53
4.2	综合表	58
4.3	行输出综合报告	63
4.4	列输出综合报告	68
第五章	统计描述	74
5.1	Frequencies 过程	74
5.2	Descriptives 过程	86
5.3	Means 过程	92

5.4 Explore 过程	102
第六章 T 检验	117
6.1 单样本 T 检验	117
6.2 独立样本 T 检验	121
6.3 配对样本 T 检验	128
第七章 单因素方差分析	132
7.1 一般概念	132
7.2 One-Way ANOVA 过程的对话框与选择项	132
7.3 例题	135
第八章 GLM—单变量过程	139
8.1 一般概念	139
8.2 GLM-Univariate 过程的对话框与选择项	139
8.3 随机区组（无重复数）设计的双因素方差分析	145
8.4 析因（有重复数，平衡数据）设计的双因素方差分析	151
8.5 析因（有重复数，不平衡数据）设计的双因素方差分析	155
8.6 完全随机设计的协方差分析	161
8.7 完全随机区组设计的协方差分析	164
8.8 线性回归分析	170
第九章 二项与 Poisson 分布	173
9.1 二项分布概率	173
9.2 Poisson 分布	175
第十章 列联表分析	177
10.1 一般概念	177
10.2 Crosstabs 对话框及选择项	177
10.3 四格表	181
10.4 $R \times C$ 表卡方检验	198
第十一章 非参数检验	210
11.1 频数分布拟合优度的卡方检验	210
11.2 二项式检验	214
11.3 游程检验	218
11.4 单样本 K-S 检验	223
11.5 两独立样本检验	228
11.6 多个独立样本检验	234
11.7 两相关样本检验	241

11.8 多个相关样本检验·····	247
第十二章 相关分析·····	250
12.1 一般概念·····	250
12.2 Bivariate 和 Partial Correlations 过程的对话框与选择项·····	250
12.3 Pearson 相关分析·····	252
12.4 Spearman 等级相关分析·····	254
12.5 偏相关分析·····	255
第十三章 线性回归分析·····	258
13.1 一般概念·····	258
13.2 Linear Regression 过程的对话框与选择项·····	258
13.3 直线回归分析·····	263
13.4 多重线性回归分析·····	269
第十四章 曲线拟合·····	274
14.1 一般概念·····	274
14.2 曲线估计·····	274
14.3 非线性回归·····	283
第十五章 生存分析·····	292
15.1 一般概念·····	292
15.2 Kaplan-Meier 和 Life table 过程的对话框与选择项·····	292
15.3 Kaplan-Meier 法: 适合小样本资料·····	297
15.4 Life table 法: 适合大样本资料·····	301
第十六章 半数效量(概率单位法)·····	307
16.1 一般概念·····	307
16.2 Probit 过程的对话框与选择项·····	307
16.3 Probit(概率单位法)例题·····	309
第十七章 统计图·····	314
17.1 一般概念·····	314
17.2 条形图(Bar Charts)·····	314
17.3 圆图(Pie Charts)·····	319
17.4 线图(Line Charts)·····	320
17.5 直方图(Histogram)·····	323
17.6 箱图(Boxplot)·····	325
17.7 误差条图(Error Bar Charts)·····	327
17.8 散点图(Scatterplots)·····	329

第一章 数据库建立和结果输出

1.1 运行环境与操作说明

1.1.1 运行环境

(1) 安装SPSS10.0 for Windows 最小硬件和软件要求

- ①Windows 9x, Windows NT 4.0, 或 Windows 2000。
- ②90MHz 以上Pentium 系列处理器。
- ③16MB (推荐64MB) 或以上内存。
- ④80MB 可用硬盘空间 (基本安装要求)。
- ⑤显示器 800 x 600 (SVGA) 或以上解像度。
- ⑥光驱。

(2) 可选组件及其要求的存储空间

- | | |
|---------------------------------|-------------------------------------|
| ①Sample Data [1MB] | ⑤Statistics Coach (IE 4.0或以上) [2MB] |
| ②Help Files [11MB] | ⑥Syntax Guide [16MB] |
| ③BASIC Scripting [2MB] | ⑦SPSS Regression Models [1.0MB] |
| ④Production mode facility [1MB] | ⑧SPSS Advanced Models [2.5MB] |

1.1.2 操作说明

(1) 鼠标 (mouse) 操作

一般鼠标有左右两键。左键相当于回车键，右键为属性/编辑键。

①单击 (single click): 如无特殊说明，指按鼠标左键一次。常用于光标的定位或选项。

②双击 (double clicks): 指快速连续按鼠标左键两次。常用于运行某种过程或进入文字编辑状态。

③拖曳 (highlight): 按住鼠标左键不放，移动光标到达某一特定位置，再放开左键，拖曳范围背景变深，表示标识了某一范围。

④右键使用: 有单击和拖曳操作，无双击功能。如在对话框中，光标箭头指向某一选项，然后单击右键，可出现关于该选项的简要英文说明。如在数据编辑器 (Data Editor) 中，单击左键选定某一对象后，单击右键，或按住右键拖曳选定范围后松开右键，出现编辑菜单。在结果阅读器 (SPSS Viewer) 中，右键没有拖曳功能，单击可选定对象并出现编辑菜单。

(2) 光标 (cursor) 形状

- ①箭头: 用于选项。

- ②沙漏斗：表示电脑在运行中。
- ③I 字形：在文书处理中的光标移动。
- ④手指形：选择名词解释。

(3) 书中的右箭头说明

右箭头→表示用鼠标在菜单或对话框中的连续向前选项过程。如：File→New→Data。
表示先用鼠标选择 File 项，再选择 New 项，最后选择 Data 项。

(4) 对话框内容说明

在使用 SPSS 过程中，会遇到各种对话框以进行功能设置，对话框一般由按钮、变量/编辑栏、选项等所组成，具体说明见图 1.1.1。

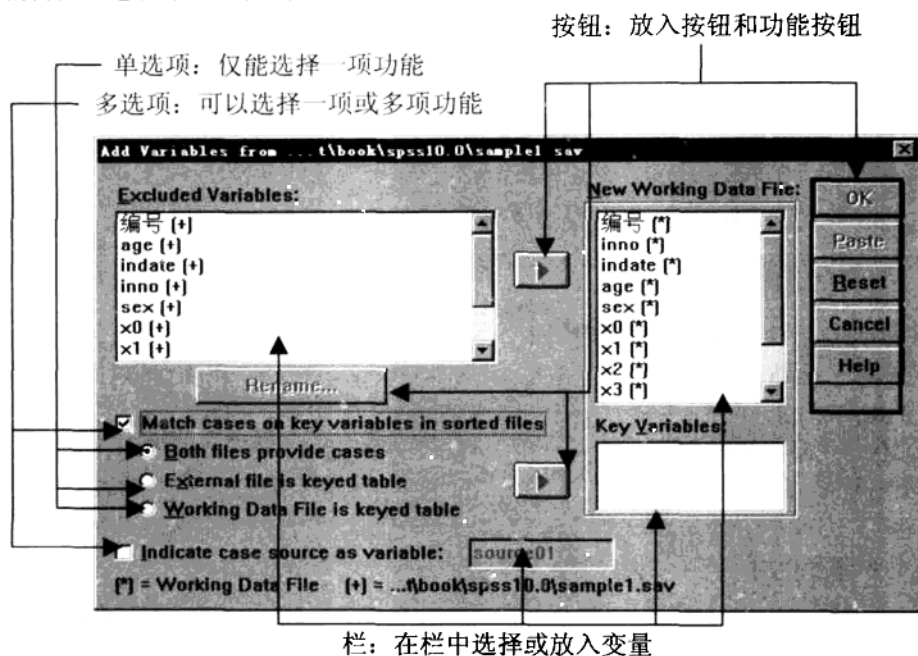


图 1.1.1 对话框的一般结构说明

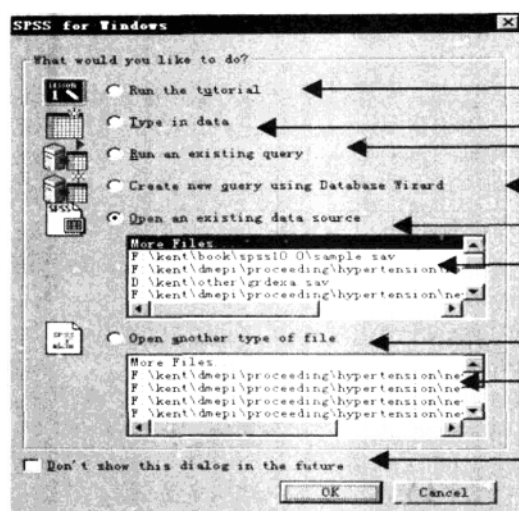
1.2 视窗界面

1.2.1 SPSS 的启动

方法有三种：

- ①在快捷工具栏单击 SPSS 10.0 图标 (icon)。
- ②开始→程序→SPSS 10.0 for Windows。
- ③在资源管理器搜索到 SPSSWIN.exe 执行文件并双击它。

SPSS 运行后，可见一个运行对话框 (图 1.2.1)，提供多个选项，单击选定后，单击 **OK** 按钮确定。

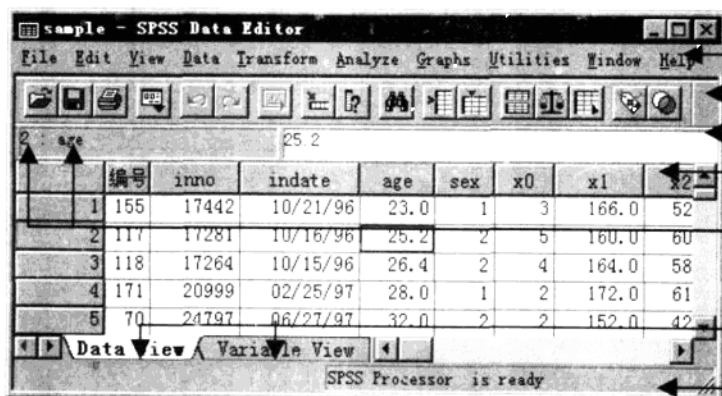


- ☐ 运行教学程序。
- ☐ 建立新数据库，输入数据。
- ☐ 运行一个已存在的查询。
- ☐ 利用数据库向导建立新查询。
- ☒ 打开一个已存在的数据库，默认。
- ☐ 最近打开过的数据库列表。单击选定
- ☐ 打开一个已存在结果文件。
- ☐ 最近打开过的其他类型文件列表。单击选定
- ☐ 关闭此对话框，以后每次启动 SPSS 不再出现此对话框。

图 1.2.1 SPSS 运行对话框

1.2.2 SPSS Data Editor: SPSS 数据编辑器

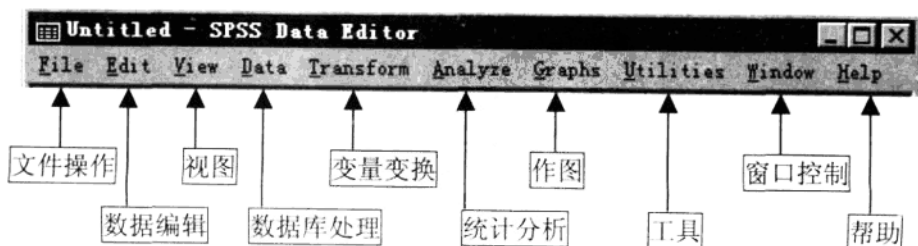
在上述运行对话框中，选定 Type in data 选项，或选定一个已存在的数据库，单击 **OK** 按钮；或直接单击 **Cancel** 按钮，即进入 SPSS 数据编辑器窗，结构如图 1.2.2。



- 菜单栏
- 快捷工具栏
- 数据输入/编辑栏
- 变量名栏
- 选定数据格的
(行:列)位置
- 数据表/变量表
- SPSS 工作状态栏

图 1.2.2 SPSS 数据编辑器

(1) SPSS 数据编辑器的菜单栏 包含 10 个下拉式 (down-list) 菜单 (submenu)。



单击上述某一菜单名，出现下拉式菜单，含有多项功能，有方向符号“▶”表示内含子菜单，如：New ▶， Open ▶ 等。有三个小点符号“...”的功能表示将打开一个对话框，如 Save As ...， Display Data Info... 等。要激活菜单或功能，单击该选项。

(2) 常用菜单主要功能

①File: 文件操作菜单。包括以下功能:

sample - SPSS Data Editor	
File	Edit View Data Transform Ans
New	新建数据库/语法/结果/手稿文件
Open	打开已有数据库/语法/结果/手稿/其他文件
Open Database	新建/编辑/运行查询
Read Text Data	读入文本格式数据
Save	存储当前应用文件
Save As...	另存当前应用文件
Display Data Info...	显示数据信息
Apply Data Dictionary...	调用外部数据库
Cache Data...	数据缓存
Print...	打印
Print Preview	打印预览
Switch Server...	转换服务器
Stop Processor	中止处理过程
Recently Used Data	最近使用过的数据库列表
Recently Used Files	最近使用过的文件列表
Exit	退出 SPSS

②Edit: 编辑菜单。包括以下功能:

sample - SPSS Data Editor	
Edit	View Data Transf
Undo	Ctrl+Z
Redo	Ctrl+Y
Cut	Ctrl+X
Copy	Ctrl+C
Paste	Ctrl+V
Clear	Del
Find...	Ctrl+F
Options...	

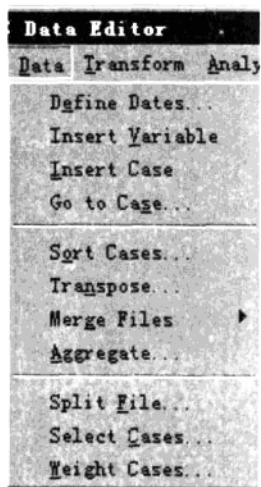
取消上一步操作
恢复上一步操作
剪切选定数据或变量
复制选定数据或变量
粘贴选定数据或变量
清除选定数据或变量
搜索数据
SPSS 参数设定

③View: 视图菜单。功能包括:

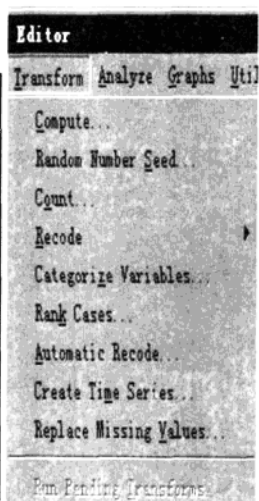
SPSS Data Editor	
View	Data Transfo
✓ Status Bar	显示状态栏
Toolbars...	显示工具栏
Fonts...	定义字体
✓ Grid Lines	显示单元格线
Value Labels	显示数值标签

④Data: 数据库处理菜单。功能包括:

⑤Transform: 变量变换菜单。功能包括:



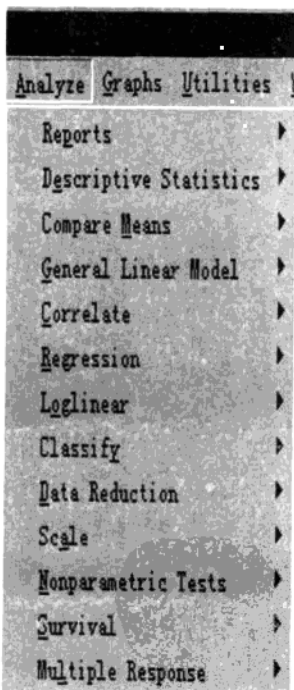
定义日期格式
插入变量
插入个体
转到某个体
个体排序
行列转置
合并文件数据
整合数据
分割文件
选择个体
加权个体



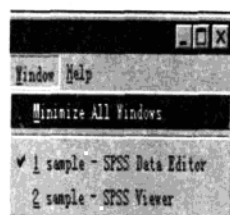
新变量赋值
产生随机数
计数
原变量重新赋值
生成新分类变量
个体排秩
自动重新赋值
生成时间序列
替代缺失值
运行未执行的变换

⑥Analyze: 统计分析菜单。功能包括:

⑦Window: 窗口菜单。功能包括:



数据概况统计
描述性统计
均数比较
一般线性模型
相关分析
回归分析
对数线性分析
分类分析
数据简化分析
尺度分析
非参数检验
生存分析
多重响应分析



最小化全部窗口
激活数据编辑器
激活结果阅读器

⑧**Graphs**: 作图菜单。功能包括:

Graphs Utilities	图形特性描述
Gallery	交互图形
Interactive ▶	
Bar...	条图
Line...	线图
Area...	区域图
Pie...	圆图
High-Low...	高-低图
Pareto...	Pareto 图
Control...	控制图
Boxplot...	箱式图
Error Bar...	误差条图
Scatter...	散点图
Histogram...	直方图
P-P...	正态 P-P 图
Q-Q...	正态 Q-Q 图
Sequence...	序列图
ROC Curve...	ROC 曲线
Time Series ▶	时间序列图

⑨**Utilities**: 工具菜单。功能包括:

Utilities Window Hel	
Variables...	变量信息
File Info	文件信息
Define Sets...	定义变量组
Use Sets...	使用变量组
✓ Auto New Case	自动增加新个体
Run Script...	运行手稿文件
Menu Editor ...	菜单编辑器

⑩ **Help**: 帮助菜单。功能包括:

Help	
Topics	帮助主题
Tutorial	运行教程
SPSS Home Page	SPSS 主页
Syntax Guide ▶	语法指南
Statistics Coach	统计指导
About...	关于 SPSS

1.3 数据库建立

1.3.1 一般概念

数据库是统计分析的基础，一个完善而简洁的数据库，可以提高研究和分析的效率。在建立数据库前，应根据研究目的，制订初步分析计划和拟定要采用的分析方法，仔细考察研究中所观察的指标（变量），确定哪些指标是目前分析所必需，哪些指标对将来进一步的研究或分析有用，哪些是主要的分析、比较指标，哪些是次要或辅助指标，哪些是分类或分组指标。既要尽量减少无关重要的指标，减少工作量，也要避免过于精简，以至在分析过程中，由于缺少某一两个指标而重新收集或输入数据。一般而言，第一次建立的数据库应较为详细，尽可能包括较多的指标或信息，尤其当考虑到以后的研究或分析还要利用到该数据库时。对于每一个个体(Case)，应确定一个或几个特异的标识（关键）变量以加以区分，如编号、身份证号、住院号、姓名等，以利日后数据库的合并。对于分类或分组指标，如是连续变量，不宜事先做划分，而应收集具体数值并输入，在实际分析时再进行分类，因为利用 SPSS 可以非常灵活地根据需要进行分组。

常用数据库为二维（行列）表结构，一般列为变量（Variable），行为个体(Case)。数据表中的一个格（Cell）放一个数据（数值或字符）。见图 1.3.1。建立一个数据库分为两个步骤：①定义变量，②输入数据。

行首为个体序号
(case number)

列首为变量名
(variable name)

数据格 (cell)

	编号	inno	indate	age	sex	x0	x1	x2
1	155	17442	10/21/96	23.0	1	3	160.0	52
2	117	17281	10/16/96	25.2	2	5	160.0	60
3	118	17264	10/15/96	26.4	2	4	164.0	58
4	171	20999	02/25/97	28.0	1	2	172.0	61
5	70	24797	06/27/97	32.0	2	2	152.0	42

图 1.3.1 数据表（库）结构

1.3.2 定义变量 (Define Variable) 过程

进入 SPSS 数据编辑器后, 单击激活左下方的 Variable View(变量表), 见图 1.3.2。定义一个变量包括十个选项。依次说明如下:

(1) **Name:** 变量名。在该选项下方的方格中, 单击定位后, 双击激活方格, 进入编辑状态, 输入变量名, 最大可输入 8 个英文字符, 4 个中文字符。系统默认值为 VAR00001。

(2) **Type:** 变量类型单选项。系统默认为数值型 (Numeric)。在该选项下方的方格中, 单击定位后, 方格右方出现按钮 。单击该按钮, 出现变量类型定义对话框, 见图 1.3.3。有下列选项供选择: ☒ Numeric, ☐ Comma, ☐ Dot, ☐ Scientific notation, ☐ Date, ☐ Dollar, ☐ Custom currency, ☐ String。

在统计数据中最常用的是数值型 Numeric 和字符型 String。系统默认值是 Numeric 型。单击选定所需变量类型后, 在对话框右侧会出现相应的选项供选择。以 Numeric 为例, 选定后在对话框右侧有两个选项, 分别为:

①**Width** : 定义数值数位。系统默认为 8 位, 双击数值栏可修改。注意: 小数点要占 1 位。

②**Decimal** : 定义小数位。系统默认为 2 位, 双击数值栏可修改。

如设定为 String, 则会出现:

③**Characters** : 定义字符串长度。系统默认为 4 位, 双击数值栏可作修改。一个中文字符占 2 位。

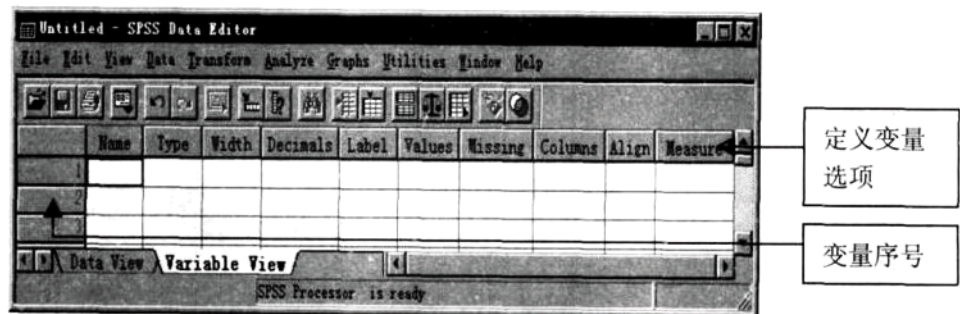


图 1.3.2 变量表视图

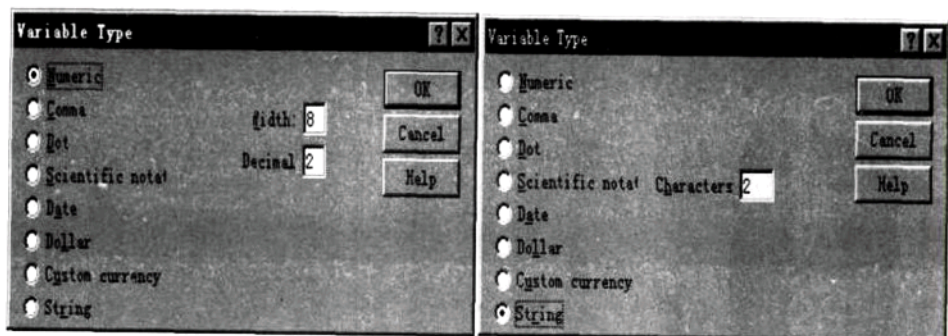


图 1.3.3 Variable Type 对话框

(3) **Width:** 数值数位。相当于 Type 对话框中的 Width 和 Characters 选项。在该选项下方的方格中, 单击定位后, 方格右方出现按钮 , 单击可增加或减少数位, 或双击后直接在方格中填入数字。

(4) **Decimals:** 小数位数。相当于 Type 对话框中的 Decimal 选项。在该选项列下方的方格中, 单击定位后, 方格右方出现按钮 , 单击可增加或减少数位, 或双击后直接在方格中填入数字。

(5) **Labels:** 变量标签。即变量说明, 所有变量均可定义标签。在该选项列下方的方格中, 单击定位后, 双击激活方格, 进入编辑状态, 输入变量标签。在输出结果中变量名可被标签所代替。

(6) **Values:** 数值标签。一般只用于分类(组)变量。在该选项列下方的方格中, 单击定位后, 方格右方出现按钮 。单击该按钮, 出现数值标签对话框, 见图 1.3.4。如分类变量中数值 1 代表高血压, 2 代表正常人, 可在对话框中的第一栏输入数值, 第二栏输入标签, 然后单击按钮 **Add** 加入显示栏。定义完成后单击按钮 **OK** 确定。在输出结果中数值可被标签所代替。

(7) **Missing:** 定义缺失值单选项。在该选项列下方的方格中, 单击定位后, 方格右方出现按钮 。单击该按钮, 出现缺失值定义对话框, 见图 1.3.5。有下列选项供选择:

☐ No missing values: 无缺失值。系统默认值。

☐ Discrete missing values: 定义离散缺失值。有三个数值栏, 每栏放入一个数值, 最多可定义三个缺失值。

☐ Range plus one optional discrete miss: 定义混合缺失值。包括连续型缺失值栏和一个离散型缺失值栏。将连续型缺失值的最小值和最大值分别放入 Low 和 High 数值栏, 将一个离散型缺失值放入 Discrete 栏。

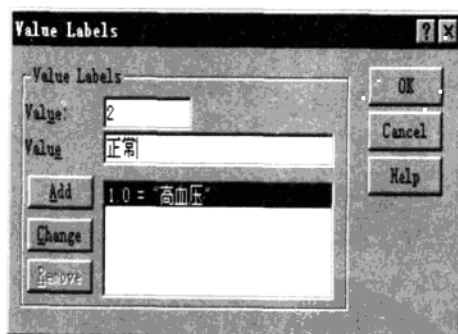


图 1.3.4 Value Labels 对话框

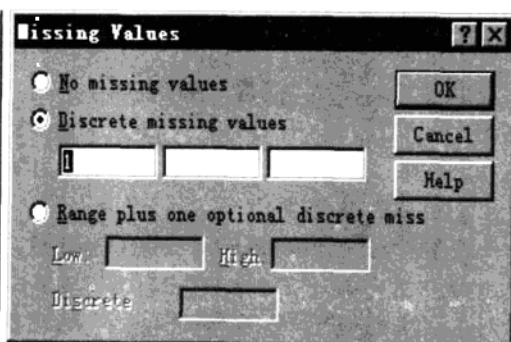
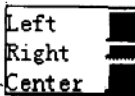
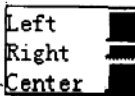


图 1.3.5 Missing Values 对话框

(8) **Columns:** 列宽。用于调整数据表中各列的显示宽度。系统默认为 8 位。在该选项列下方的方格中, 单击定位后, 方格右方出现按钮 , 点击可增加或减少数位, 或直接在方格中填入数字。列宽位数要大于该变量中 Width 选项的位数, 否则在数据表中, 数据显示不完整。

(9) **Align:** 定义字符排列方向。在该选项列  下方的方格中, 单击定位后, 方格右方出现按钮 , 点击按钮出现下拉菜单 , 分别为左对齐、右对