

信息与计算机科学丛书

SPSS软件应用指导

——常用医学生物统计

李湘鸣 主编



东南大学出版社

中国统计出版社 中国统计出版社

SPSS软件应用指导

中国统计出版社 中国统计出版社

中国统计出版社

中国统计出版社 中国统计出版社

信息与计算科学丛书

SPSS 软件应用指导

——常用医学生物统计

主 编 李湘鸣

副主编 唐 尧 孙 峰

东南大学出版社

·南京·

内 容 提 要

本书主要以 SPSS8.0 for Windows 基本使用方法为基础,系统地讲述医学上常用统计方法,包括以下统计内容:描述性统计、 t 检验、方差分析(F 检验)、 χ^2 (卡方)检验、正态性检验、直线相关与回归、多元相关与回归、统计绘图及使用技巧。书中实例均来自实际的医学科研中,因此本书主要供医学院校各专业的本科生、研究生、青年教师、临床医生学习及统计处理科研数据之用,也可供综合性大学、农林院校及师范院校有关专业的师生学习参考之用。

图书在版编目(CIP)数据

SPSS 软件应用指导:常用医学生物统计/李湘鸣主编. -
南京:东南大学出版社,2000.10

(信息与计算科学丛书)

ISBN 7-81050-674-9

I. S... II. 李... III. 统计分析-软件包,SPSS-应
用-医学统计:生物统计 IV. R311

中国版本图书馆 CIP 数据核字(2000)第 52350 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 大丰市科星印刷有限责任公司印刷

开本:787mm×1092mm 1/16 印张:7.5 字数:187 千字

2000 年 10 月第 1 版 2000 年 10 月第 1 次印刷

印数:1~3500 册 总定价:180.00 元 本册定价:14.00 元

(凡因印装质量问题,可直接向发行科调换。电话:025-3792327)

前 言

SPSS 是美国 SPSS 公司所研制的一套大型集成应用软件系统,具有完备的数据存取、数据管理、数据分析和数据展现功能,在数据处理和统计分析领域,被誉为国际上的标准软件,广泛应用于政府行政管理、科研、教育、生产和金融等不同领域,发挥着重要的作用。

SPSS 软件自开发至今版本不断更新升级,功能更加全面,因此,一个人不可能将它的所有统计功能全部掌握。本书着眼于用 SPSS 软件处理大多数医务工作者所经常遇到的科研资料,以使其结果更加科学、严谨。

目前所流行的 SPSS 多为英文菜单操作,因为涉及很多术语,用普通汉化软件汉化有一定的困难。为此,我约请从事多年医学统计教学工作的唐尧、孙峰两位老师合编本书。书中对医学生物统计学中常用的某些英文菜单、命令、术语给予必要的中文解释,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。

系统地学习和掌握 SPSS,需要花费大量的时间和精力,而对于大多数医务工作者而言,需要的仅仅是如何用 SPSS 来处理资料,解决实际问题。本书正是以解决实际问题作为出发点,写作时主要以实例为主,具有以下优点:1. 使略有点医学统计学知识且具一定的计算机知识者,根据自己的资料性质,进行临摹套用,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科研中,因此本书主要可供医学院校各专业的本科生、研究生、青年教师、临床医生学习及统计处理科研数据之用,也可供综合性大学、农林院校及师范院校有关专业的师生学习参考。

随着计算机不断换代,软件不断更新,SPSS 在不断升级,但万变不离其宗。本书写作主要参照 SPSS8.0 for Windows,与其他版本大同小异。

学习计算机软件除了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。

SPSS 的各项功能由各种功能模块完成,因此各章节之间自然会出现数据共享或统计模块共用的可能。为了便于自学和实际应用,本书原则上以各章节为独立单元进行写作,使初学者能根据自己的资料性质,在较短的时间内掌握某一章节的内容而不必从头学起。

SPSS 是以数据库为基本框架结构的,因此要学好 SPSS,建议读者抽时间学点数据库方面的知识。

由于“任何个人错误总是难免的”,故书中错误之处,恳请各位同行专家和读者提出批评意见。

李湘鸣

于扬州大学医学院

2000 年 8 月

目 录

第一章 概述	(1)
一、SPSS 的设计思想	(1)
二、SPSS 的特点	(1)
三、安装与启动	(2)
四、SPSS 视窗介绍	(2)
五、基本操作	(3)
第二章 描述性统计	(6)
一、正态性资料	(6)
二、非正态性资料	(15)
第三章 t 检验	(19)
一、样本均数与总体均数的比较	(19)
二、成组资料的两个样本均数比较	(20)
三、配对资料的比较	(28)
第四章 方差分析	(31)
一、完全随机设计的多个样本均数比较	(31)
二、配伍组设计的多个样本均数比较	(36)
第五章 正态性检验	(44)
一、P - P Plot 法	(44)
二、Q - Q Plot 法	(46)
第六章 χ^2 (卡方)检验	(47)
一、独立(成组)样本资料的 χ^2 检验	(47)
二、配对样本资料的 χ^2 检验	(53)
三、行(R) $\times$ 列(C)样本资料的 χ^2 检验	(57)
第七章 相关与回归分析	(61)
一、相关分析	(61)
二、回归分析	(66)

第八章 多元线性回归与相关	(70)
一、多元线性回归	(70)
二、多元线性相关	(77)
第九章 统计制图	(79)
一、线图	(79)
二、直条图	(88)
三、圆形图	(93)
四、小结	(95)
第十章 使用技巧	(97)
一、如何用 SPSS 调用 Excel 数据库?	(97)
二、如何将 SPSS 工作表中数据转化为 Word 表格式?	(104)
三、如何求多层分组间描述性统计指标?	(106)
四、如何将 SPSS 所绘图形转入 Word 系统进行编辑?	(108)
附录 本书专用术语中英文对照表	(109)

第一章 概 述

SPSS 是美国使用最为广泛的三大著名统计分析软件(SPSS, SAS 和 SYSTAT)之一,是目前国际上最为流行的一种大型统计分析系统,被誉为统计分析的标准软件。

1968 年,美国 SPSS 公司董事长 Norman H. Nie 与 Stanford 大学的两名研究生(C. Hadlai Hull 和 Dale Bent)首次开发研制出 SPSS 软件,当时的名字为 Statistical Package for the Social Science,缩写为 SPSS,意思是“社会科学统计软件包”。今天的 SPSS 软件功能与最初大不相同,具有多种统计分析功能,可满足许多行业统计的需求,但公司仍使用最初的缩写名。

一、SPSS 的设计思想

SPSS 的设计思想基于为科技工作者提供一个工具,使用这一工具可以完成包括从简单的描述性统计及绘图到复杂的多变数分析的各种运算,从而使人们从繁重的计算任务中解脱出来,有更多的时间和精力用于分析和解释计算结果,而不必为如何获得这些结果花费过多的时间和精力。

二、SPSS 的特点

1. 统计方法齐全

(1) SPSS 提供了基本的统计指标计算,如算术均数、几何均数、中位数、各种百分位数、组中值、标准差、标准误、可信区间、相关系数、回归系数等。

(2) SPSS 几乎囊括了所有的医学统计推断或分析方法,如 t 检验、方差分析、协方差分析、正态性检验、卡方检验、二项分布、直线相关与回归分析、非线性回归分析、偏相关分析、多元回归分析、Logistic 回归分析、聚类分析以及寿命表的计算等。

2. 统计制图方便快捷

(1) SPSS 几乎可提供所有的常用医学制图,如线图、散点图、直条图、直方图、叠加图、圆形图等,制图与图形编辑完全由菜单操作,并不断出现提示对话框或图标,可使读者一目了然。

(2) 所绘图形与 Word、WPS2000、画图板等具有良好的兼容性,可发挥各自软件的优点对图形进行编辑与修改。

3. 操作简便灵活

(1) 主要为菜单式操作,整个过程以对话框的形式展现,一般不需要编程,只要告诉 SPSS “做什么”,而不必告诉其“怎样做”,具有所谓的“傻瓜型(fool-proof)”特点,通常只需几步操作便可完成一项复杂的运算或分析。

(2) 结果输出以简明的英文给出提示,统计术语规范易懂,使用者具有初步英语水平和统计基础即可。在输出结果中,会自动显示出资料进行统计分析所符合的条件。如在进行两样本均数比较时,会显示出它们的方差是否齐性(F 检验结果),若不齐性时,自动输出相应的检

验方法结果;在进行 χ^2 检验时,会输出其资料的理论值等。

(3) 在帮助菜单中,具有 Tutorial(指导手册)、Statistics Coach(统计学指导)和 Ask me(咨询)等下拉式菜单,为读者自学 SPSS 和统计处理资料提供了极好的捷径。

4. 兼容性强

SPSS 数据库与 dBASE, ForBASE, Excel, Lotus, Systat 等多种数据库具有良好的兼容性。

三、安装与启动

1. 硬件要求

计算机至少应为 486 机型,内存至少 8M,为了充分发挥软件的性能,建议最好选用 P II 以上机型,内存 32M,显示器最好不少于 800 × 600 像素(pixels),显卡可支持 256 色。

2. 软件要求

SPSS8.0 要求以 Windows 操作系统为平台,最好为 Windows 9X 系统。

3. 安装方法

将有 SPSS 8.0 的软件置入光驱,进入 SPSS 子目录(文件夹),双击 Setup 命令,便开始运行 SPSS 安装程序。然后根据屏幕提示,键入公司密码或系列号(serial number),便可完成安装。

4. 运行

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 8.0 for Windows”及“SPSS 8.0 Production Facility”菜单,如图 1.1 所示。单击“SPSS 8.0 for Windows”便可启动 SPSS,产生图 1.2 对话框。常用的三个菜单为:Type in data(输入数据)、Open an existing file(打开已存在的文件)和 Run the tutorial(运行指导手册)。用鼠标选中相应命令,便可进入有关操作。亦可点击 Cancel 按钮,直接进入 SPSS 视窗(图 1.3)进行相关操作,这是常见的操作方式。



图 1.1

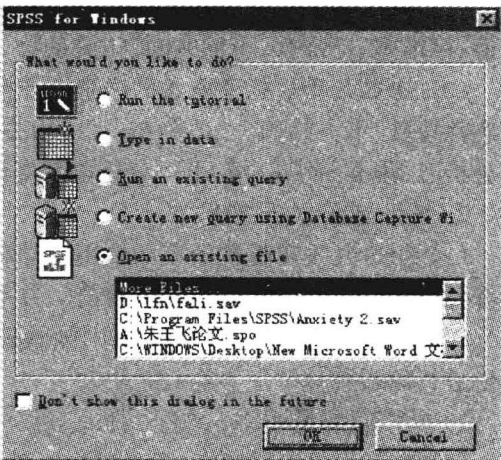


图 1.2

四、SPSS 视窗介绍

视窗最上方是标题栏。标题栏下方是功能菜单表,分别为 File(文件)、Edit(编辑)、View

(视图)、Data(数据)、Transform(转换)、Statistics(统计)、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表,在其中可选取相应的命令。在功能菜单表的下方是工具栏,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,右边用以输入变量中的数据。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。在列标栏中可选相应的变量名进行编辑。最下方是状态栏,显示当前执行的命令或提示的信息。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)。

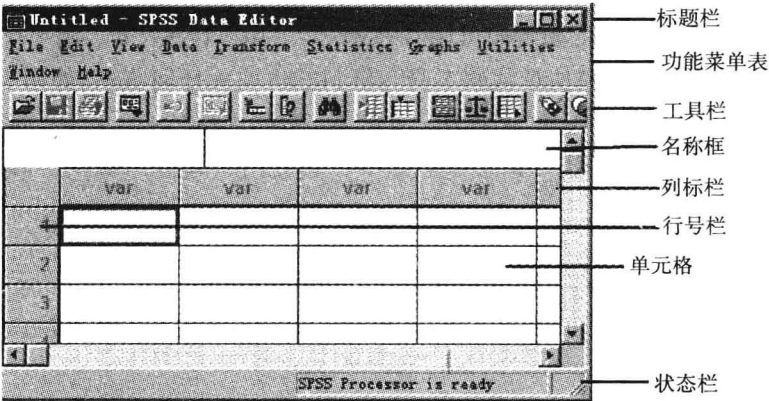


图 1.3

五、基本操作

(一) 建立文件

建立文件时,首先应定义变量,即告诉计算机当前输入数据的名称、大小、类型及性质。具体操作如下:

1. 将鼠标指针移向列标栏单击左键选中任意一 Var 列,单击鼠标右键,则显示一菜单(图 1.4),用鼠标左键选中 Define Variable(定义变量)命令,则弹出图 1.5 所示的对话框。此步操作亦可单击标题栏上 Data 菜单,在下拉式菜单中选 Define Variable 进行。

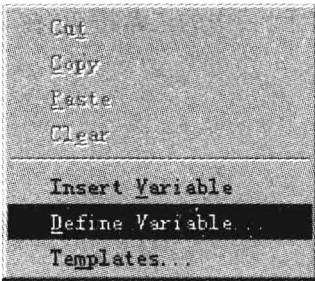


图 1.4

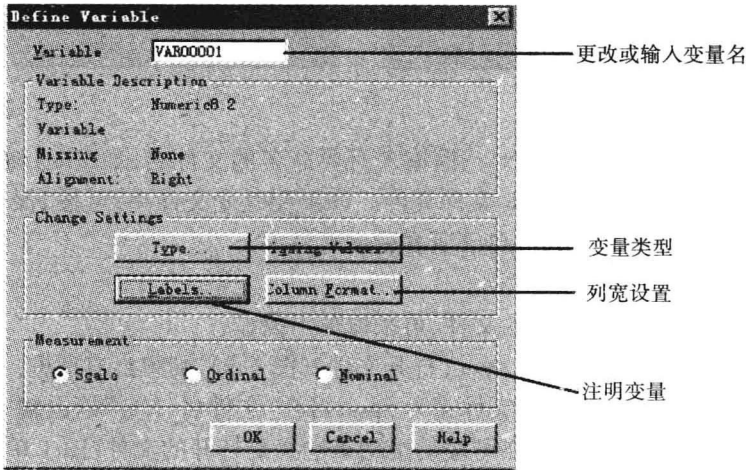


图 1.5

2. 在图 1.5 中 Variable(变量名框)内,最多可输入 8 个字符即 4 个汉字。选择变量类型(Type)按钮,可根据资料的性质选择所输入资料的类型,常用的有数字型变量(Numeric)和字符串变量(String)两种类型。宽度可输入 1~99 字符,默认为 8 个字符。

3. 列宽(Column Format)默认为 8 个字符,可变范围为 1~99。

4. 注明(Labels)按钮有以下用处:

(1) 当变量超过 8 个字符(4 个汉字)时,可在此对话框中的 Variable 栏输入,当完成 Labels 操作后,将鼠标指针移向相应变量,便显示出所注明内容。

(2) 对 Ordinal(等级的)或 Nominal(形式上的)两种性质资料进行注明。Ordinal 类资料,相当于医学上半计量(等级)资料,如我们可用 0 表示“-”,用 1 代表“+”,用 2 代表“++”,用 3 代表“+++”等。Nominal 资料所选代码本身与资料的性质无任何联系,如用 1 代表“男”,2 代表“女”;用 1 代表“阴性组”,2 代表“阳性组”,3 代表“高剂量组”,4 代表“中剂量组”等。有关注明及所代表的内容将会在输出视窗中显示出来。本节只对定义变量作一概要介绍,具体操作见以后有关章节。

(二) 保存文件

当文件建立好后,一般应将文件保存起来。具体操作如下:

1. 单击 File(文件) → Save(保存)(图 1.6 所示),则弹出对话框(图 1.7)。

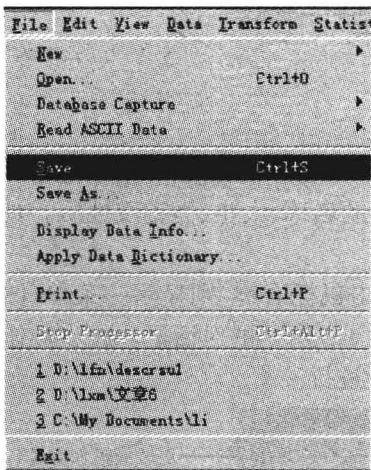


图 1.6

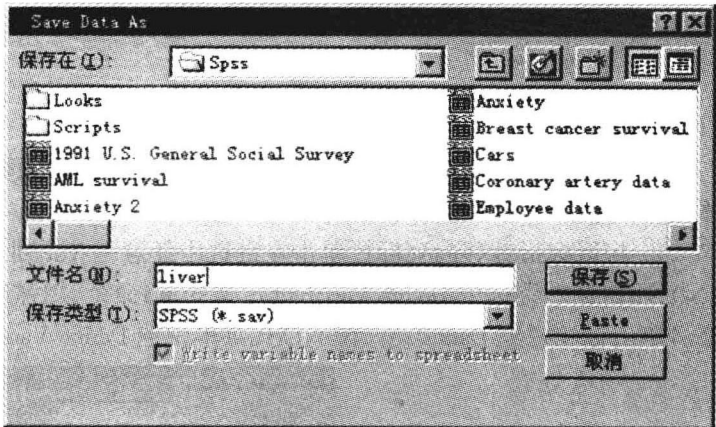


图 1.7

2. 在“文件名(N):”框输入所要命名的文件名,如 liver,然后按保存(S)按钮,便会将所要保存的文件存入 Spss 文件夹内。如果要改变文件夹,可按“保存在(I):”中的三角按钮,在下拉式选择框中,选择自己所喜爱的文件夹。

(三) 打开文件

要打开已保存的文件,可采取如下操作:

1. 单击 File → Open(打开文件)(图 1.8),则弹出对话框。

2. SPSS 默认的文件夹为 Spss,可在“文件名(N):”框内输入要打开的文件名,即可打开该文件夹中的文件。亦可按动滚动条选择要打开的文件(图 1.9 所示)。

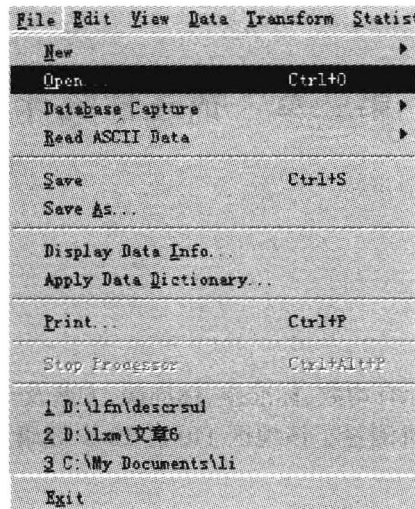


图 1.8

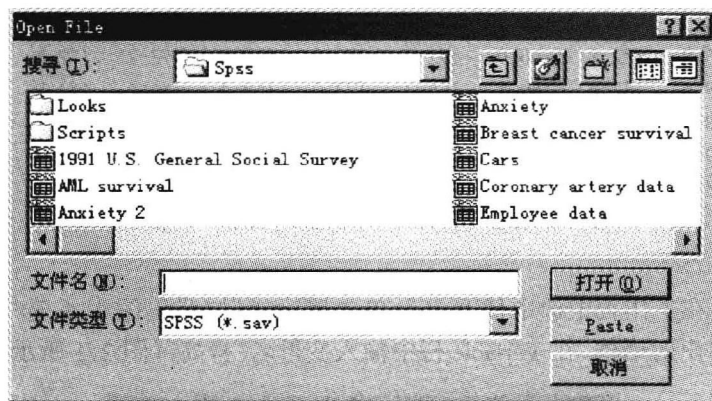


图 1.9

如果读者对上述有关文件的建立与打开仍不清楚,请学习有关 Windows 方面的操作,这里不再详述。

第二章 描述性统计

一、正态性资料

(一) 常用统计指标

描述正态性资料常用指标有均数、标准差、标准误、方差及全距等。

【例 2.1】 某地 1993 随机测量了该地区 110 名 20 岁健康男大学生的身高, 资料如下, 试对其进行统计学描述。

173.9	173.9	166.9	179.5	171.2	167.8	177.1	174.7	173.8	182.5
173.6	165.8	168.7	173.6	173.7	177.8	180.3	173.1	173.0	172.6
173.6	175.3	178.4	181.5	170.5	176.4	170.8	171.8	180.7	170.7
173.8	164.4	170.0	175.0	177.7	171.4	162.9	179.0	174.9	178.3
174.5	174.3	170.4	173.2	174.5	173.7	173.4	173.9	172.9	177.9
168.3	175.0	172.1	166.9	172.7	172.2	168.0	172.7	172.3	175.2
171.9	168.6	167.6	169.1	166.8	172.0	168.4	166.2	172.8	166.1
173.5	168.6	172.4	175.7	178.8	169.1	175.5	170.8	171.7	164.6
171.2	169.1	170.7	173.6	167.2	170.7	174.7	171.8	167.3	174.8
168.5	178.7	177.3	165.9	170.4	170.2	169.5	172.1	178.2	170.9
171.3	176.1	169.7	177.9	171.1	179.3	183.5	168.5	175.5	175.9

【操作】

1. 定义变量: 在工作表中选中任意 Var, 单击鼠标右键, 弹出图 1.4, 选中菜单中的 Define Variable 命令, 则弹出对话框, 在 Variable 栏中输入变量名“身高”(图 2.1 所示)。

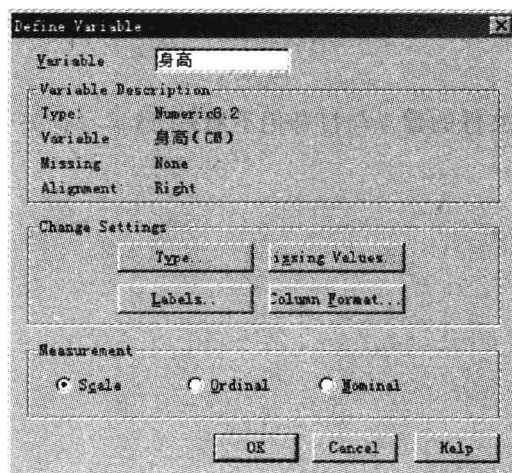


图 2.1

2. 可按 Labels 按钮,在 Variable 栏输入“身高(cm)”,此内容会在输出视窗的表格中出现。

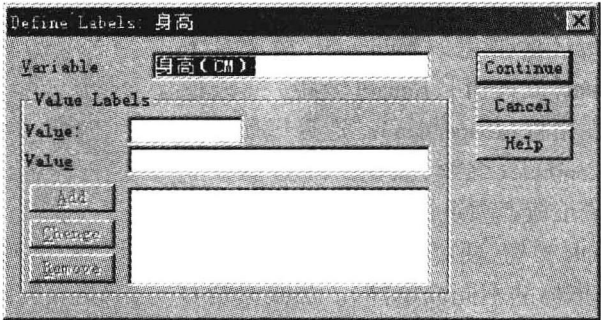


图 2.2

3. 点击图 2.1 中 OK 按钮,此时便会在工作表列标栏出现一为“身高”的变量名。

4. 数据输入:用鼠标选中该变量行号为 1(第一个记录)中的单元格,键盘输入上述数据;注意应将全部数据逐步输入变量名为“身高”的列内,不可输入其他列内。

5. 用鼠标单击 Statistics 菜单,在其下拉式菜单中选中 Summarize(总结)菜单中的 Descriptives(描述)命令(图 2.3 所示),则弹出对话框。

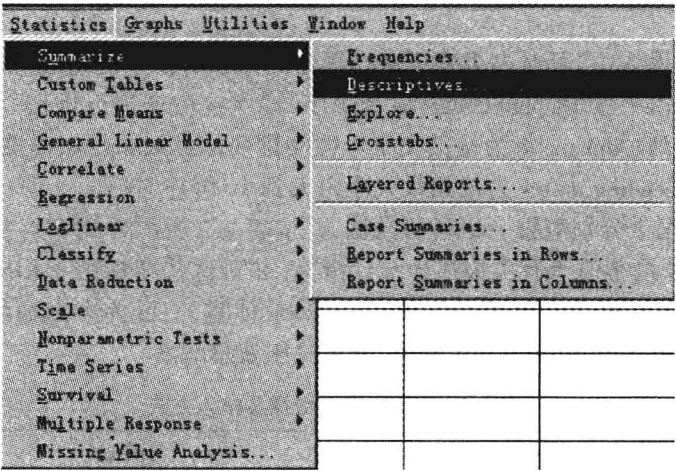


图 2.3

6. 选中变量“身高”,按三角按钮,使其进入 Variable(s)框内(图 2.4 所示)。

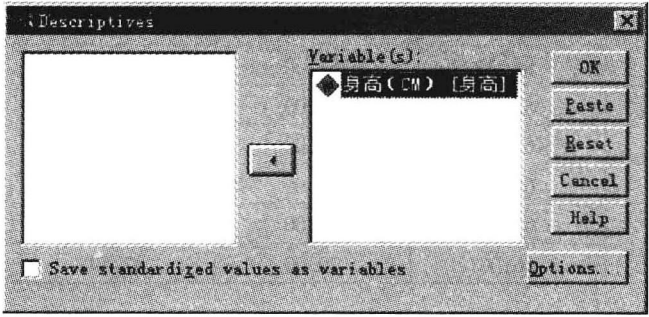


图 2.4

7. 点击 Options(选项)按钮,便弹出图 2.5 所示对话框。读者可根据需要选择所需统计指标。Mean 为均数, Sum 为总和, Std. deviat (标准差)、Minimum(最小值)、Variance(方差)、Maximum(最大值)、Range(全距)和 S. E. mean(标准误)为离散性指标(Dispersion)。选择好所需的统计指标后,再点击 Continue(继续)按钮。SPSS 默认均数、标准差、最大值和最小值四项统计指标。

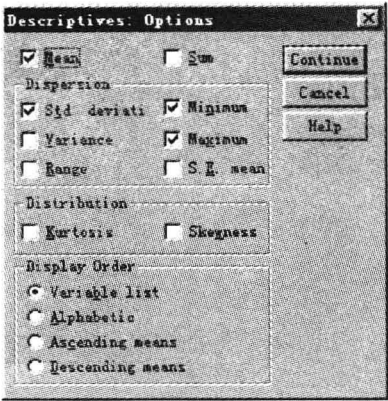


图 2.5

8. 再按图 2.1 中 OK 按钮,便可在输出视窗中输出结果(此输出结果因所选指标的不同而不同)。如果激活打印命令,便可打印出结果,如表 2.1 所示(为了便于初学者学习,本书所有输出表格均为加工过的示意表)。

表 2.1 Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
身高(cm)	110	162.90	183.50	172.695 5	4.141 7
Valid N (listwise)	110				

在 Descriptives 操作中,可对多个变量进行操作。如果对多个变量进行描述时,多个变量的输出结果在输出视窗中按 Display Order(显示顺序)栏所选命令进行排列。Variable list 为按所选择变量的顺序排列;Alphabetic 为按英文字母顺序排列;Ascending means 为按均数大小由小到大(升序)排列;Descending means 为按均数大小由大到小(降序)排列。

【例 2.2】 某医生分别测得一组科研数据,并按表 2.2 所示方式输入 SPSS,请分别求其均数、标准差、最大值、最小值等有关描述性统计指标,并对这些变量给予相应的注明,具体操作请参阅例 2.1。表中质量为大鼠甲状腺质量(g/200g 体重);Hb 为大鼠血红蛋白(g/L);RATE 为对大鼠甲状腺肿瘤的抑制率(%);URINE 为大鼠尿液总酸度。

表 2.2 例 2.2 有关数据及输入方式

记录号	质量	HB	RATE	URINE
1	65.00	68.00	85.96	33.00
2	82.00	80.00	94.62	30.50
3	73.00	75.00	89.04	33.00
4	92.00	50.00	80.47	29.90
5	81.00	82.00	86.74	28.90
6	85.00	72.00	73.82	31.40
7	63.00	76.00	83.11	37.10
8	68.00	75.00	89.10	34.60
9	67.00	68.00	77.61	30.70
10	56.00	70.00	82.59	28.40

【操作】

1. 单击 Statistics → Summarize → Descriptives, 弹出对话框。
2. 将“质量”、“Hb”、“RATE”、“URINE”等变量依次置入 Variable(s) 栏内(图 2.6 所示), 按 OK 按钮, 则输出如表 2.3 所示的结果。

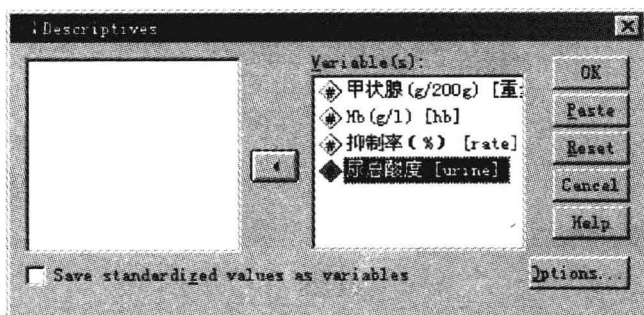


图 2.6

表 2.3 Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
甲状腺(g/200g)	10	56.00	92.00	73.200 0	11.370 5
Hb(g/L)	10	50.00	82.00	71.600 0	8.921 9
抑制率(%)	10	73.82	94.62	84.306 0	6.094 8
尿总酸度	10	28.40	37.10	31.750 0	2.696 2
Valid N (listwise)	10				

由于 SPSS 默认为 Variable list, 所以输出结果按甲状腺、Hb、抑制率、尿总酸度的顺序排列。

(二) 计算频数

在实际工作中, 我们往往想知道某计量资料在某一组段的频数, 如身高在某范围内有多少人, 不同年龄段有多少人等。现以例 2.3 说明如何计算频数。

【例 2.3】 将例 2.1 中 110 个身高数据分为 11 个组段, 组距为 2cm, 其分组依次为 162 ~, 164 ~, 166 ~, 168 ~, 170 ~, 172 ~, 174 ~, 176 ~, 178 ~, 180 ~, 182 ~ 184, 试计算各组段中有多少人。

【操作】

1. 单击 Transform 菜单, 选中 Compute(计算)命令(图 2.7 所示), 则弹出对话框。

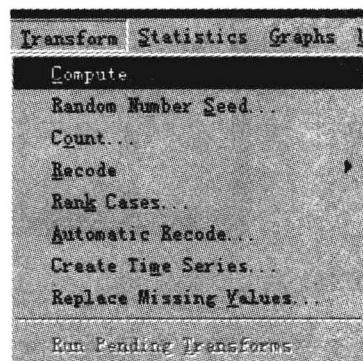


图 2.7

2. 在 Target(目标变量)框内键入任意变量名, 如 sp; 在 Numeric Expression(数式表达)框内输入任意代码, 如 1 (图 2.8 所示)。

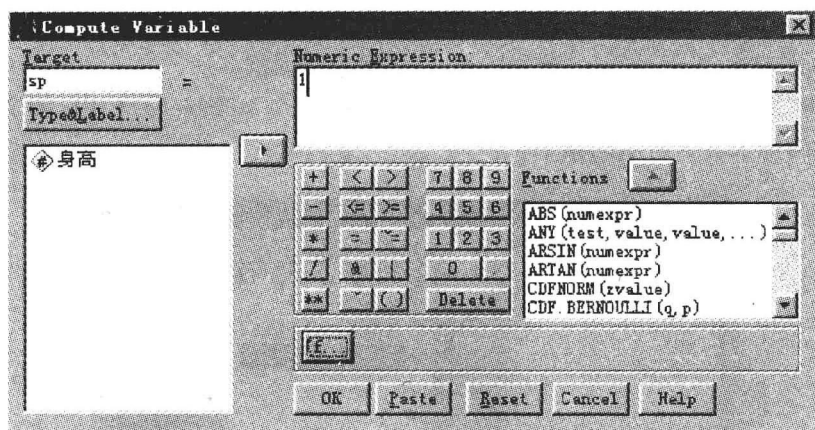


图 2.8

3. 可根据需要按 Type&Label 按钮,弹出图 2.9 所示的对话框;在 Label 栏中键入“各组段频数”,单击 Continue 按钮。

4. 按图 2.8 中 If 按钮,则弹出图 2.10 对话框,选中 Include if case satisfies condition,意思是对符合下列条件的单元(观察值)进行操作。

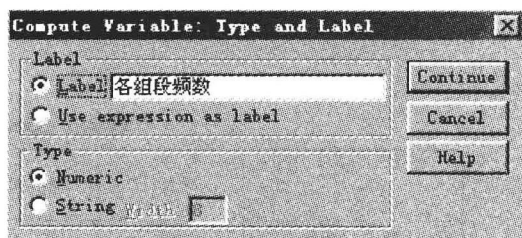


图 2.9

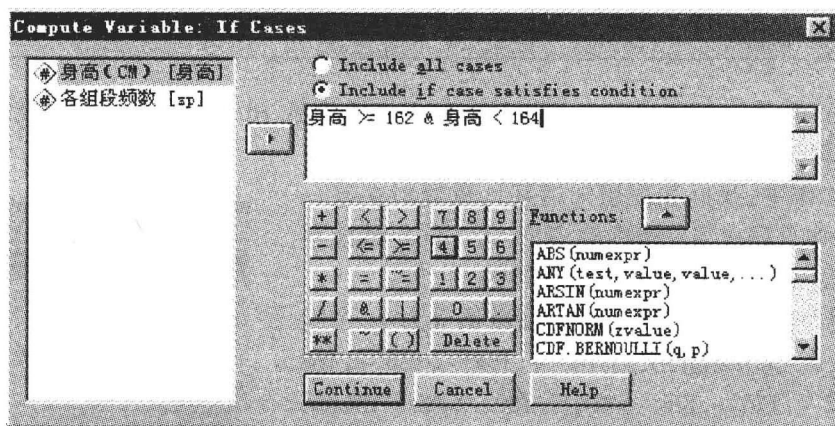


图 2.10

5. 选中“身高”变量,按左边三角按钮,置入条件表达框内,点击图中键盘键,依次输入“身高 ≥ 162 & 身高 < 164 ”,再按 Continue 按钮,然后按图 2.8 中 OK 按钮。此时便会在工作表变量名为“身高”的右边产生一个变量名为“sp”的新变量。我们会发现凡是身高在 162 ~ 164 cm 之间的记录,其相应的 sp 变量为 1(代码)。也可以这样说,凡是记录(观察值)在“62 ~ ”组的均用代码 1 表示。如果把指针移到 sp 变量时,便会显示出图 2.9 所注明的内容(图 2.11 所示)。