

统编经济管理实验实训教程书系

主 编 黄本春 李国柱
副主编 赵青霞 郇彦辉

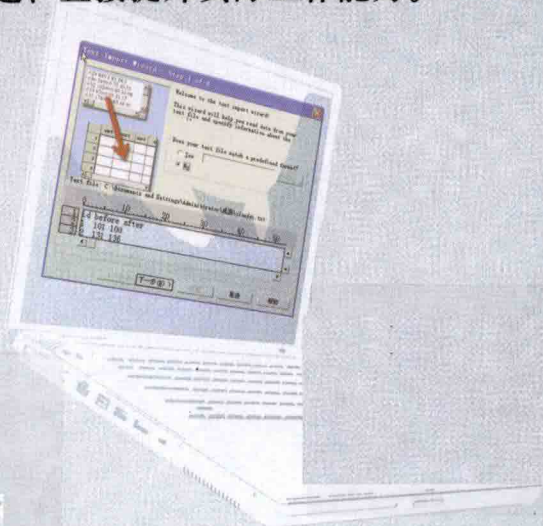
Experimental Course of
Statistics

实验教程

统计学

在高仿真的专业和工作背景下，本实验教程引导您：

1. 高度浓缩地掌握学科重要的基础理论和相关知识点；
2. 亲历各专业工作的基本流程和相关环节；
3. 运用实验软件、工具、量表有效地解决各种实际问题，快速、直接提升实际工作能力。



中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE

C8
163

统编经济管理实验实训教程书系

主 编 黄本春 李国柱
副主编 赵青霞 鄢彦辉

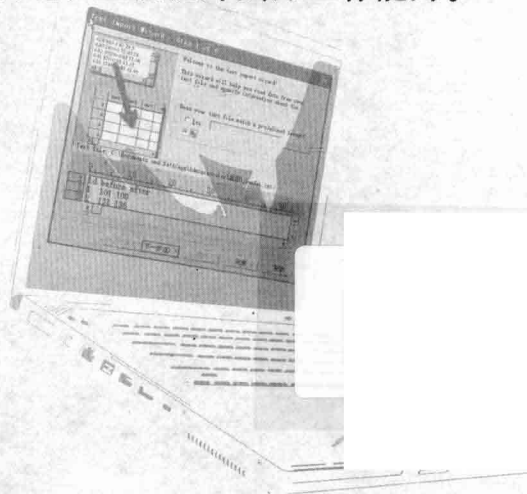
Experimental Course of
Statistics

实验教程

统计学

在高仿真的专业和工作背景下，本实验教程引导您：

1. 高度浓缩地掌握学科重要的基础理论和相关知识点；
2. 亲历各专业工作的基本流程和相关环节；
3. 运用实验软件、工具、量表有效地解决各种实际问题，快速、直接提升实际工作能力。



中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE

北 京

图书在版编目 (CIP) 数据

统计学实验教程/黄本春, 李国柱主编

北京: 中国经济出版社, 2010. 1

ISBN 978 - 7 - 5017 - 8936 - 8

I. 统… II. ①黄…②李… III. 统计学—实验—教材 IV. G649. 285. 1 - 53

中国版本图书馆 CIP 数据核字 (2010) 第 120279 号

责任编辑 师少林 张 虹

责任印制 石星岳

封面设计 巢新强

出版发行 中国经济出版社

印 刷 者 三河市佳星印装有限公司

经 销 者 各地新华书店

开 本 787mm × 1092mm 1/16

印 张 16. 75

字 数 350 千字

版 次 2010 年 1 月第 1 版

印 次 2010 年 1 月第 1 次

印 数 7000 册

书 号 ISBN 978 - 7 - 5017 - 8936 - 8/F · 6926

定 价 30. 00 元

中国经济出版社 网址 www.economyph.com 社址 北京市西城区百万庄北街 3 号 邮编 100037

本版图书如存在印装质量问题, 请与本社发行中心联系调换(联系电话: 010 - 68319116)

版权所有 盗版必究(举报电话: 010 - 68359418 010 - 68319282)

国家版权局反盗版举报中心(举报电话: 12390)

服务热线: 010 - 68344225 88386794

前 言

统计学是系统介绍有关如何测定、搜集、整理和分析客观现象总体数量特征的方法论科学。随着科学技术和社会经济的不断发展,统计学的应用领域也越来越广阔,特别是随着计算机科学的发展,基于大量数据处理的统计学在探求客观事物规律性方面越发显得重要,而统计学与计算机数据处理的结合也越来越紧密。

正是基于以上认识,我们希望编写出一本统计学知识与当前主流统计软件 SPSS 完美结合的统计实验教材。本书在编写中突出了以下特点:

1. 脉络清晰。以统计学学科包括的内容为主线,从基础的数据整理入手,分模块介绍统计知识如何借助 SPSS 软件来达到数据的快速、准确的处理。这点有别于大多数的介绍统计软件的教材以软件模块为主来展开的编写思路。

2. 理论与实践结合。目前国内的实验教材多数是就实验而实验,对相关理论要么没有介绍,要么一笔带过;而统计学教材则对相关知识如何借助计算机软件来实现介绍较少或没有介绍。本书正是针对这种现状,在每个实验前都对相关统计知识做了精练的介绍,力求做到统计理论与统计数据处理技术的有机结合,编者在编写中力求在理论与软件数据处理之间达到一个平衡。

3. 重点突出。本书在编写中围绕统计学知识展开,对 SPSS 中有包含的内容做了适当的取舍,对实践中应用较少的内容没有编入,使整个教材紧紧围绕常用的统计知识展开,努力做到重点突出。

本书是多位老师通力合作完成的。由黄本春、李国柱担任主编,赵青霞、郑彦辉担任副主编。由黄本春、李国柱等提出编写大纲,经所有编写组成员讨论后分头编写。各部分分工为:李国柱编写第一篇的实验一和第二篇;赖志花编写第一篇的实验二、四、五;李从欣编写第一篇的实验三;赵青霞编写第三篇;黄本春编写第四篇;郑彦辉编写第五篇。最后由黄本春和李国柱统稿、定稿。

在本书的编写过程中,全体编写人员进行了多次讨论,并得到了石家庄经济学院实验实践教学中心领导和老师的大力支持以及石家庄经济学院经贸学院领导的关心和指导。在编写过程中还参考和借鉴了国内同行专家、学者的有关论著和研究成果,在此一并表示感谢!

由于时间仓促,编者水平有限,书中难免有不当或疏漏之处,恳请读者和同仁批评指正,我们将不胜感激。

编者

2009 年 10 月

目录

CONTENTS

第一篇 基础数据的整理与显示 / 1

实验一 数据文件的建立	1
实验二 数据文件的整理	15
实验三 常用统计图	26
实验四 统计报表(一)	51
实验五 统计报表(二)	62

第二篇 数据分析基础 / 76

实验六 描述性统计分析	76
实验七 均值比较分析	89
实验八 相关分析	105

第三篇 非参数统计 / 115

实验九 卡方检验 (Chi-square test)	115
实验十 二项分布检验 (Binomial test)	129
实验十一 随机游程检验 (Run Test)	132
实验十二 单样本 Kolmogorov-Smirnov 检验	135
实验十三 两样本模型的非参数检验	140
实验十四 K 个样本模型的非参数检验	156

第四篇 多元统计 / 179

实验十五 判别分析	179
实验十六 聚类分析	192
实验十七 主成分分析和因子分析	204



第五篇 实验设计 / 215

实验十八 单因素实验:方差分析	215
实验十九 随机化区组设计	225
实验二十 析因分析实验	233
实验二十一 嵌套设计或分级设计	240
实验二十二 正交实验设计	246
实验二十三 拉丁方实验设计	254

第八章 实验设计 / 215

1	立象的中文设计	一盆实
12	立象的中文设计	二盆实
20	立象的中文设计	三盆实
21	(一) 立象的中文设计	四盆实
23	(二) 立象的中文设计	五盆实

第九章 实验设计 / 215

76	立象的中文设计	六盆实
80	立象的中文设计	七盆实
102	立象的中文设计	八盆实

第十章 实验设计 / 215

112	立象的中文设计	九盆实
120	立象的中文设计	十盆实
132	立象的中文设计	十一盆实
132	立象的中文设计	十二盆实
140	立象的中文设计	十三盆实
126	立象的中文设计	十四盆实

第十一章 实验设计 / 215

120	立象的中文设计	十五盆实
122	立象的中文设计	十六盆实
124	立象的中文设计	十七盆实

第一篇

基础数据的整理与显示

实验一 数据文件的建立

一、实验目的

了解数据文件的建立方法。

二、实验内容

1. 数据文件的直接录入;
2. 外部数据文件的导入;
3. 数据文件的保存。

三、实验工具

SPSS File 菜单项。

四、知识准备

在定量研究中,统计信息都是用某种数据来表示的。在对这些数据进行处理、分析时,首先要明确这些信息资料是根据何种尺度进行测定、加工的。史蒂文斯(S. S. Stevens)将尺度分为四种类型,即定类尺度、定序尺度、定距尺度和定比尺度。

定类尺度也称名义尺度,它是对统计客体类别差异所作的反映,是最粗略、计量层次最低



的测量尺度。用定类尺度进行测量时,必须符合穷尽和互斥的原则。穷尽原则是指所有的统计客体都可归到相应的类型中去;互斥原则是指每个统计客体都属于一种类型且只属于一种类型。定类尺度中使用的数字、符号、字母等具有随意性。如可用 1 表示男性,0 表示女性,也可用 0 表示男性,1 表示女性。定类变量的各类别之间是平等的,没有高低、大小、优劣之分。不能进行算术运算,只能进行逻辑运算。

定序尺度也称顺序尺度,其所使用的数值大小,是与研究对象的特定顺序相对应的。定序变量一般以文字表述,也可以用数值标识,但仅起标签作用。定序尺度仅能反映出统计客体在等级、顺序上的差别。它不能具体地测定各等级之间的间距大小。

定距尺度也称间距尺度,其所使用的数值,不仅表示测定对象所具有的量的多少,还表示它们大小的程度即间隔的大小。不过,这种尺度中的原点可以是任意设定的,并不意味着该事物的量为“无”。例如 0°C 为绝对温度 273K , 华氏 32°F 。定类尺度与定序尺度的数值不能进行加减乘除,但定距尺度的数值是可以进行加减运算的。然而,由于原点是任意设定的,所以不能进行乘除运算。例如, 5°C 和 10°C 之间的差,可以说与 15°C 与 20°C 之间的差异是相同的,都是 5°C 。但不能说 20°C 就是比 5°C 高 4 倍的温度。

定比尺度也称比例尺度,它和定距尺度的主要区别是存在固定的零点。长度、重量、时间等都是定比尺度测定的范围。定比尺度测定值的差和比都是可以比较的。例如:5 分钟与 10 分钟之间的差和 10 分钟与 15 分钟之间的差都是 5 分钟,10 分钟是 2 分钟的 5 倍。定比尺度可以进行加减乘除运算。

在统计分析中,一般要求测量的层次越高越好,因为高层次的测量尺度包含更多的数学特性,所运用的统计分析方法越多,分析时也就越方便。

五、实验背景

有 12 名志愿者服用某减肥药,服药前和服药一个疗程后各测量一次体重(kg),数据如下:

服药前:101 131 131 143 124 137 126 95 90 67 84 101

服药后:100 136 126 150 128 126 116 105 87 57 74 109

该数据已以 jianfei. xls、jianfei. mdb 和 jianfei. txt 形式保存,要求:

- (1) 将数据直接输入 SPSS;
- (2) 将 jianfei. xls、jianfei. mdb 和 jianfei. txt 导入 SPSS。

六、实验步骤

(一) 数据的直接录入

运行 SPSS, 打开 SPSS 数据编辑器, 如图 1.1 所示:

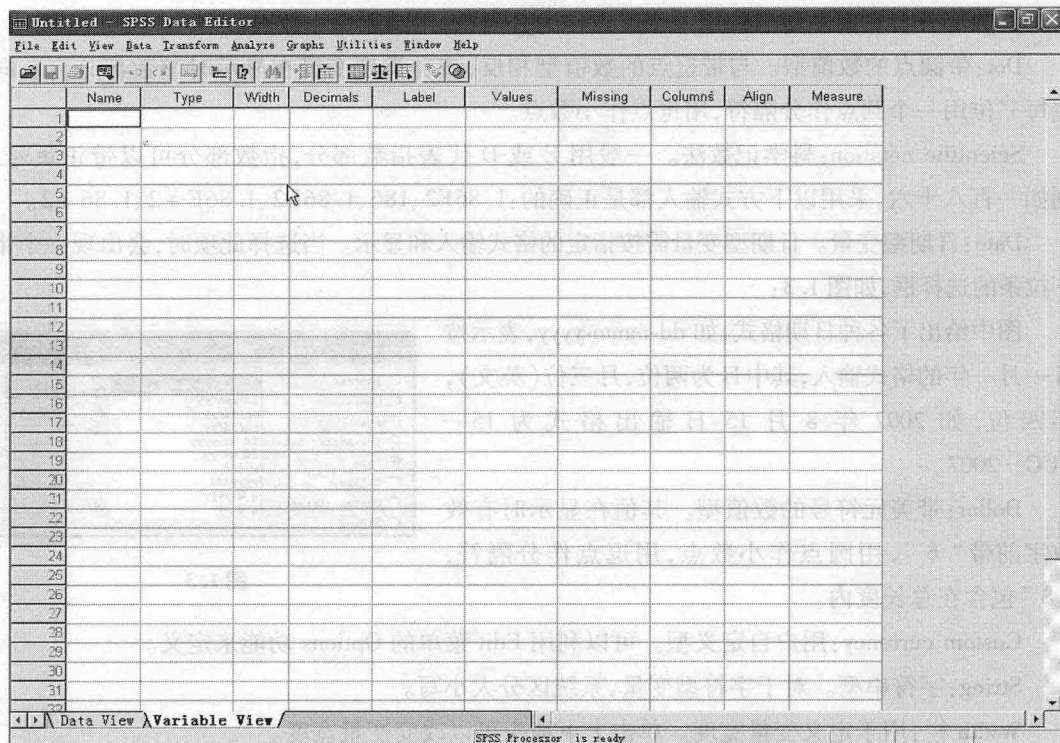


图 1.1

图 1.1 左下侧有两个标签:Data View(数据视图)和 Variable View(变量视图),点击 Variable View 标签,可以定义变量。当然也可以在 Data View 中直接输入数据,此时变量格式采用 SPSS 默认方式。

在变量视图里可以对变量的格式进行定义,其中:

Name 栏:用于定义变量名。可在此栏输入用户自定义的变量名,如不进行输入,系统默认的第一个变量名为 var00001。在定义变量名时应遵循以下原则:①变量名的长度不多于 8 个字符;②变量名不能与 SPSS 保留字相同,如变量名不能为 ALL、AND、BY、EQ、GE、GT、LE、LT、NE、NOT、OR、TO、WITH;③SPSS 不区分变量的大小写。

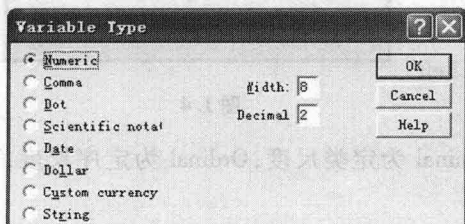


图 1.2

Type 栏:用于定义变量类型。选择该框时右侧出现 按钮,单击此按钮,弹出变量变形对话框,如图 1.2。

图 1.2 左侧列出了 8 种可供选择的变量类型,自上而下分别为:

Numeric:标准数值型。

Comma:带逗点的数值型。该种类型变量在显



示时整数部分自右向左每三位用一个逗号作为分隔符,用圆点作小数点。

Dot:带圆点的数值型。与带逗点的数值型相反,该种类型变量在显示时整数部分自右向左每三位用一个圆点作分隔符,用逗号作小数点。

Scientific notation:科学记数法。一般用 E 或 D 代表指数部分,指数部分可以带正负号。例如一百八十六,采用以下方式输入都是正确的:1.86E2、186、1.86D2、1.86E+2、1.86+2。

Date:日期型变量。日期型变量需按指定的格式输入和显示。当选择此项时,会出现一个带下拉条的选择框,如图 1.3:

图中给出了各种日期格式,如 dd-mmm-yyyy,表示按日一月一年的格式输入,其中日为两位,月三位(英文),年四位,如 2007 年 8 月 15 日输出格式为 15—AUG—2007。

Dollar:带美元符号的数值型。其值在显示时有效数字前带“\$”,用圆点作小数点,用逗号作分隔符,“\$”包含在总长度内。

Custom currency:用户自定义型。可以利用 Edit 菜单的 Options 功能来定义。

String:字符串型。对于字符型变量,系统区分大小写。

Width 栏:用于定义变量宽度。单击上下箭头可以改变变量宽度。

Decimals 栏:用于定义小数位数。单击上下箭头可以改变小数点位数。

Label 栏:用于定义变量标签,以方便他人阅读或以后自己使用方便。定义变量标签后,在统计分析结果中会在与变量名相对应的位置显示该变量标签。

Values 栏:用于定义变量值标签。该项是对变量可能取值所进行的附加说明,对分类变量一般需定义变量值标签。

Missing 栏:用于定义变量缺失值。选择该框时右侧出现按钮,单击此按钮,弹出缺失值对话框,如图 1.4 所示:

图 1.4 包含三个选项:No missing values 表示没有缺失值;Discrete missing values 用于定义离散缺失值,最多可定义三个;Range plus one optional discrete missing values 表示“缺失值范围另加一个可选的缺失值”。

Columns 栏:用于定义列宽。

Align 栏:用于定义对齐方式,共包含三种方式:Left(左对齐)、Right(右对齐)、Center(居中)。

Measure 栏:用于定义变量的测量尺度,其中,Nominal 为定类尺度,Ordinal 为定序尺度,Scale 则包含了定距尺度和定比尺度。

就本例而言,定义三个变量 id、before、after,在变量标签分别输入编号、服药前、服药后,将小

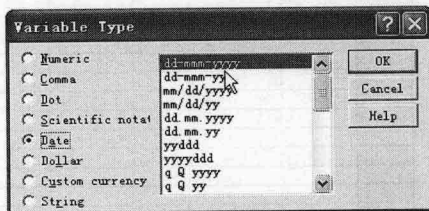


图 1.3

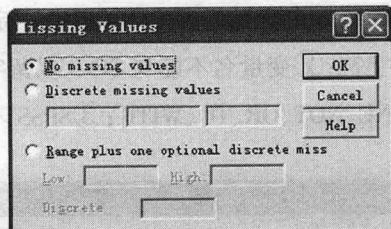


图 1.4



数位数为0,其他采用系统默认方式,用键盘直接输入数据,输入完毕后的数据格式如图 1.5:

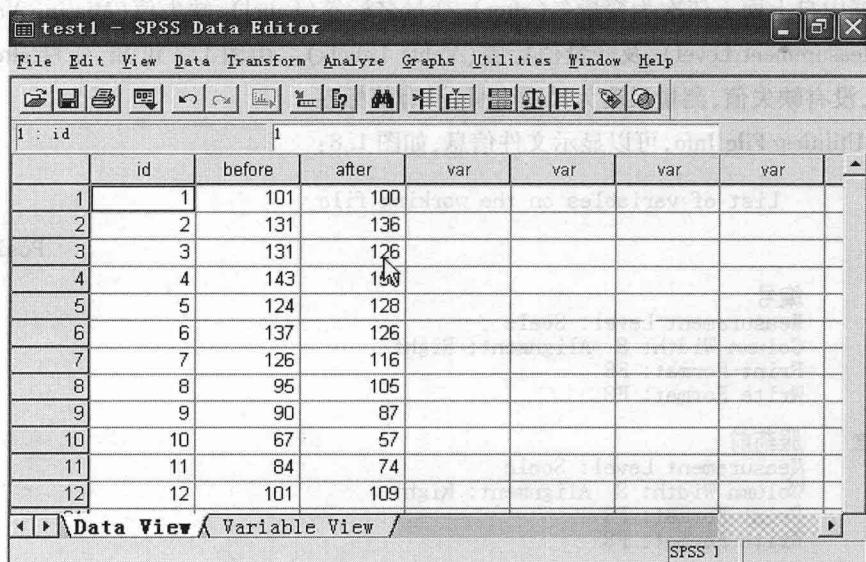


图 1.5

数据输入完毕后,可以用“查找”命令查找数据,方法是在 SPSS 主菜单中选择 Edit/Find, 打开“查找”对话框,如图 1.6:

在 Find 右侧的编辑框输入数据后,点击 Find Next 便可进行查找。Match case 表示区分大小写。

选择 Utilities/Variables,可以显示各个变量的基本信息,如图 1.7:

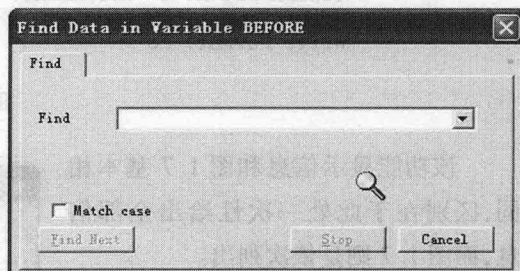


图 1.6

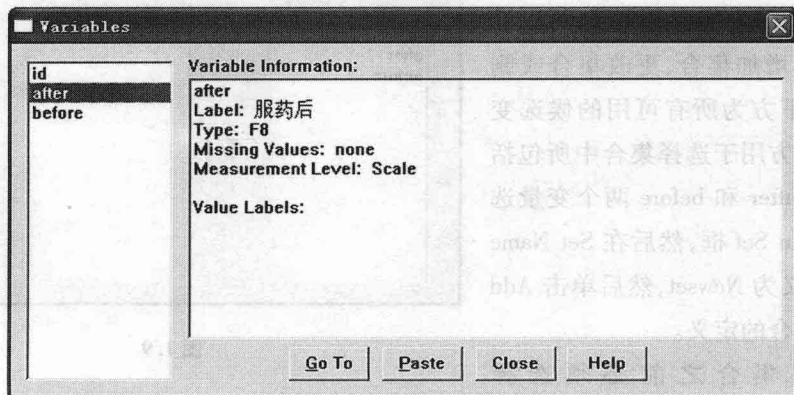


图 1.7



点击左侧变量,右侧会出现相应变量的基本信息。如左侧选择 after 后,右侧 Variable Information 框中自上而下依次为变量名(after)、变量名标签(Label)、缺失值(Missing Values)、测量尺度(Measurement Level),及变量值标签(Value Labels)。由图 1.7 可知,变量 after 的标签为服药后,没有缺失值,测量尺度为尺度变量,没有值标签。

选择 Utilities/File Info,可以显示文件信息,如图 1.8:

List of variables on the working file		
Name		Position
ID	编号 Measurement Level: Scale Column Width: 8 Alignment: Right Print Format: F8 Write Format: F8	1
BEFORE	服药前 Measurement Level: Scale Column Width: 8 Alignment: Right Print Format: F8 Write Format: F8	2
AFTER	服药后 Measurement Level: Scale Column Width: 8 Alignment: Right Print Format: F8 Write Format: F8	3

图 1.8

该功能显示信息和图 1.7 基本相同,区别在于此处一次性给出全部信息,而图 1.7 则是依次列出。

选择 Utilities/Define Sets, 打开定义数据集对话框,如图 1.9:

图 1.9 上部为集合定义区,可以设定集合名称、增加集合、更改集合或删除集合。左下方为所有可用的候选变量列表,右下方用于选择集合中所包括的变量。将 after 和 before 两个变量选入 Variables in Set 框,然后在 Set Name 框将集合定义为 Newset,然后单击 Add Set 即完成集合的定义。

在使用集合之前必须选择 Utilities/Use Sets,打开使用集合对话框,如图 1.10:

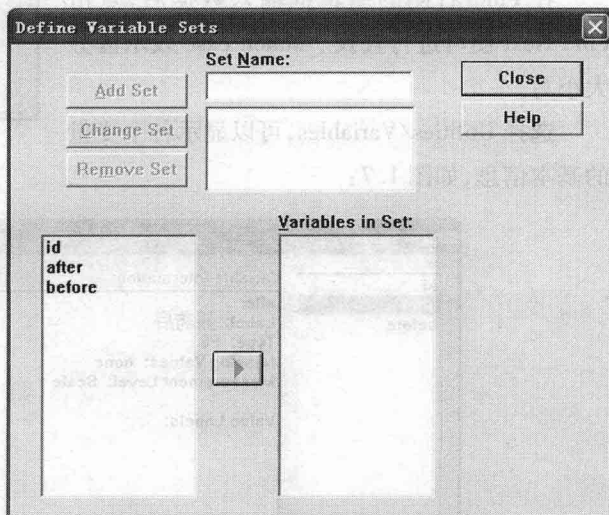


图 1.9



图 1.10 中, Sets in Use 为正在使用的集合, 可见新定义的集合并没有在其中, 单击向右箭头, 将 NEWSET 选入 Sets in Use 框中, 然后选中 ALLVARIABLES 和 NEWVARIABLES, 单击左箭头, 将二者剔除出 Sets in Use 框。新的集合便可在分析中起作用了。

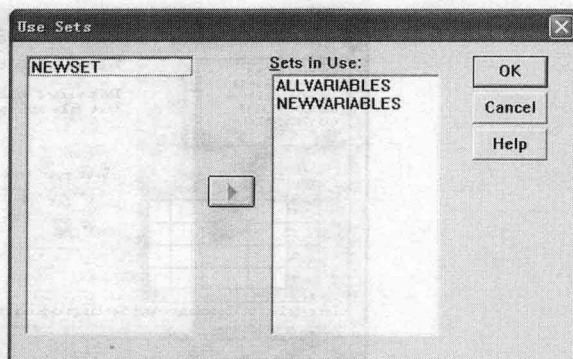


图 1.10

(二) 从外部文件导入数据

SPSS 可以读入许多非 SPSS 默认类型的数据文件, 方式主要有三种: 直接打开, 利用文本向导读入文本数据及利用数据库 ODBC 接口读取数据。

1. 电子表格数据的导入

SPSS 可以直接读入许多常用格式的数据文件, 选择 File/Open/Data, 弹出 Open File 对话框, 单击文件类型列表框, 弹出 SPSS 可直接打开的数据文件列表, 如图 1.11。

图 1.11 共列出了 16 种数据类型, 选择 Excel (*.xls), 找到 jianfei.xls, 点击“打开”, 弹出 Opening Excel Data Source 对话框, 如图 1.12。

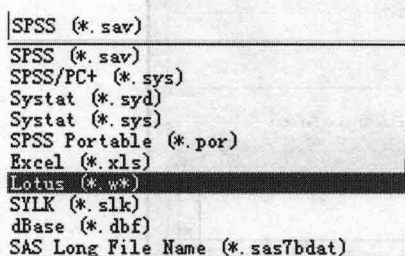


图 1.11

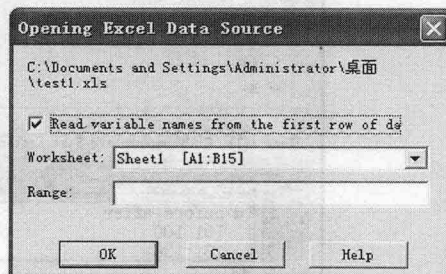


图 1.12

对话框最上面为文件所在目录, Read variable names from the first row of data 表示从数据的第一行读取变量名; Worksheet 框中用于指定哪张表; Range 框中用于指定读取数据的具体位置。指定完毕, 点击 OK 按钮, 数据便可读入 SPSS。

2. 导入文本文件

选择 File/Read Text Data, 打开文本导入向导, 如图 1.13。

图 1.13 为导入向导第一步, 主要用于询问数据文件是否和预定义格式相一致 (Does your text file match a predefined format?), 下方为按预定义格式读入的数据文件的预览, 单击下一步, 打开图 1.14。

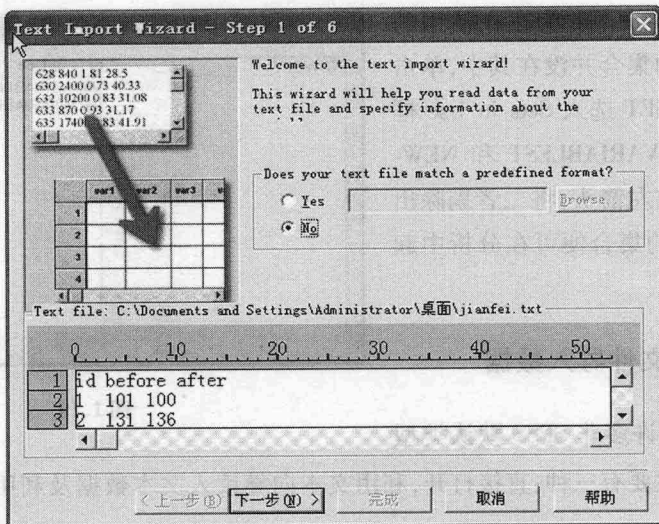


图 1.13

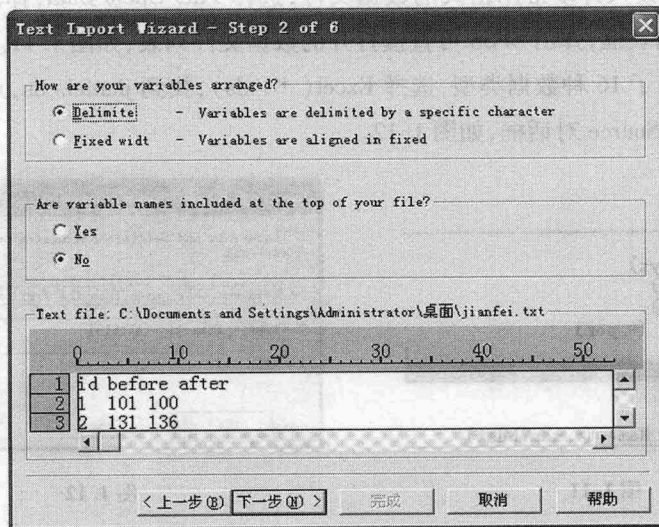


图 1.14

图 1.14 为导入向导第二步,共分两部分,第一部分为 How are your variables arranged? (你的变量是如何排列的), Delimited-Variables are delimited by a specific character 表示变量按某种字符区分, Fixed width-Variables are aligned in fixed 表示变量按固定宽度区分。第二部分为 Are variable names included at the top of your file? (变量名是否在文件顶端), 选择 Yes, 然后单击“下一步”, 打开图 1.15。

图 1.15 第一部分为 The first case of data begins on which, 表示数据的第一个观测从第几行

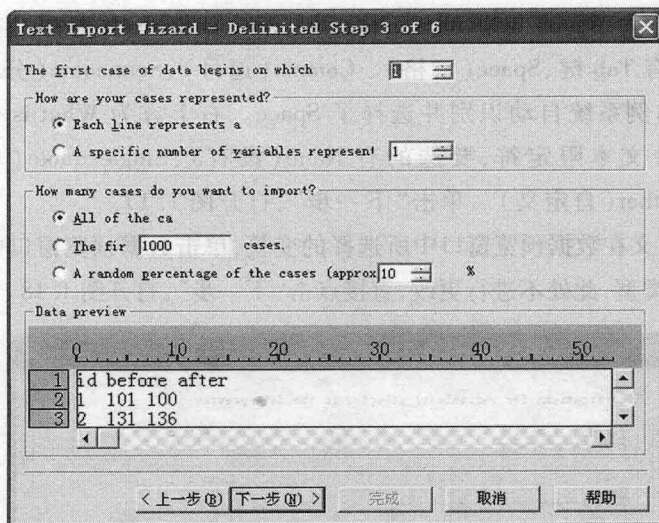


图 1.15

开始,系统默认为1,由于数据的第一行为变量名,所以此处应改为2。第二部分为 How are your cases represented? 表示记录是如何存储的,Each line represents a case 为一行代表一个记录,A specific number of variables represent 为每多少个变量代表一条记录。此处选择默认值。第三部分为 How many cases do you want to import? 用于选择需导入多少条记录,All of the cases 为导入所有记录,The first cases 为导入前多少条记录,A random percentage of the cases 为随机导入一定比例的记录,此处选择导入所有数据。选择完毕后,点击“下一步”,打开图 1.16:

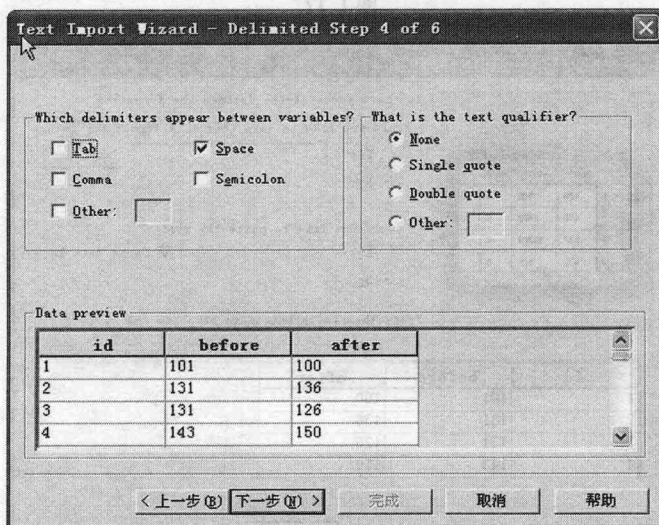


图 1.16

图 1.16 左上方为 Which delimiters appear between variables? 意思是变量间用的是哪种分隔符,可选的有 Tab 键、Space(空格)、Comma(逗号)、Semicolon(分号)、Other(自定义的其他符号),本例系统自动识别并选择了 Space。右上方为 What is the text qualifier? 意思为采用了何种文本限定符,要选的有 None(没有)、Single quote(单引号)、Double quote(双引号)、Other(自定义)。单击“下一步”,打开图 1.17。

图 1.17 用于定义在数据预览窗口中所选择的变量,单击数据预览窗口中的某一变量,可以改变其变量名和类型,此处不进行更改,直接点击“下一步”,打开图 1.18。

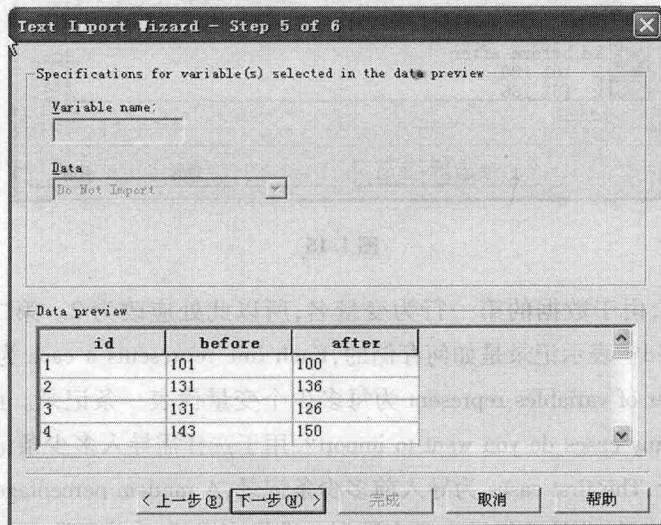


图 1.17

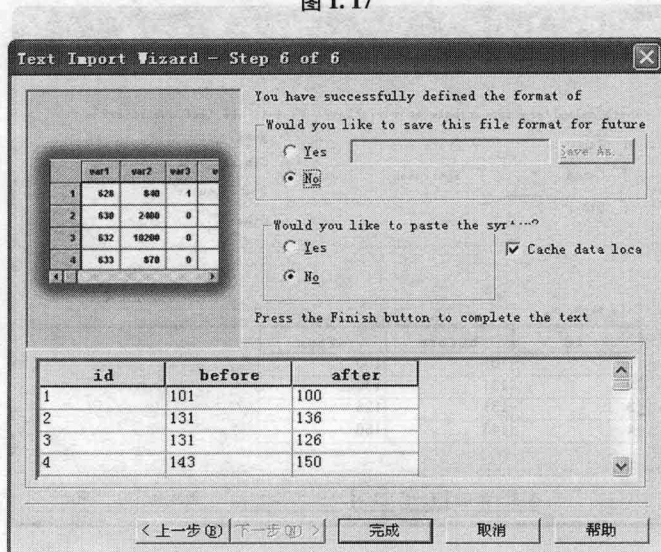


图 1.18



图 1.18 包含两个部分,上端为 Would you like to save this file format for future,意思是你愿意保存文件格式以备将来所用吗,第二部分为 Would you like to paste the syntax? 表示你是否愿意将以上操作粘贴为 SPSS 语句? 选择默认设置,单击“完成”按钮,系统即成功读入 jianfei. txt。

3. 导入数据库文件

选择 File/Open Database/New Query, 打开数据库向导对话框,如图 1.19:

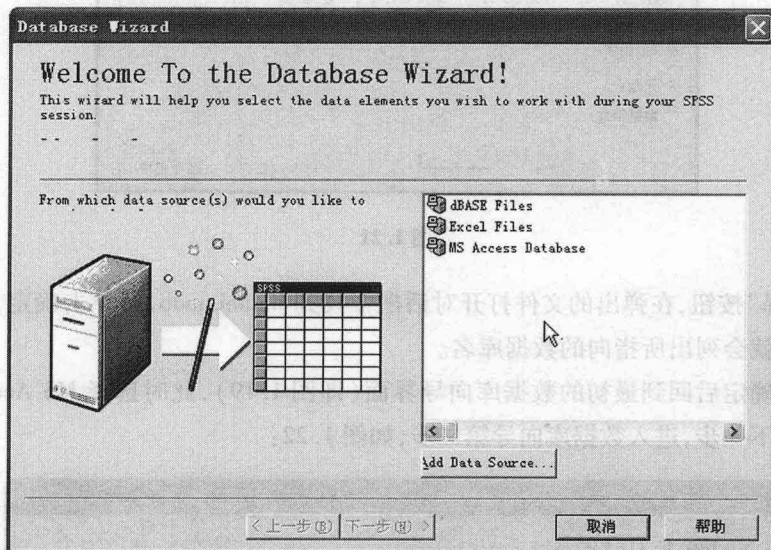


图 1.19

在图 1.19 中,单击下方的 Add Data Source,弹出 ODBC 数据管理器窗口,如图 1.20:

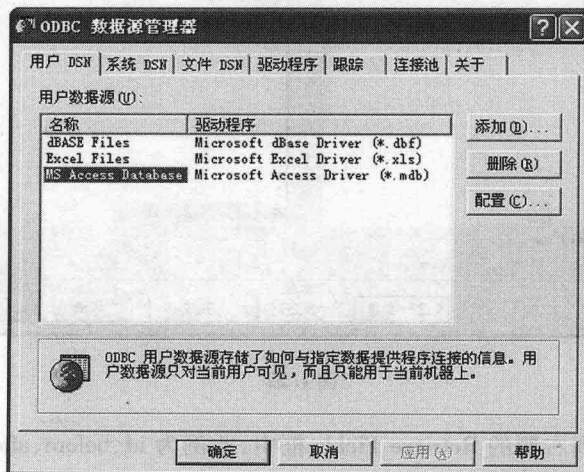


图 1.20

在用户数据源列表中选 MS Access Database,单击配置按钮,弹出 ODBC 数据源安装界