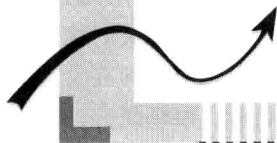


# 第 1 章



## SPSS 统计分析软件概述

### 本章要点

1. 明确 SPSS 软件是一种专业的统计分析软件，了解 SPSS 的主要应用领域。
2. 熟练掌握 SPSS 进入和退出等基本操作，了解 SPSS 的基本窗口和菜单安排。
3. 掌握 SPSS 的三种使用方式以及它们的特点和应用场合。
4. 掌握利用 SPSS 进行数据分析的基本步骤。

SPSS 是 Statistical Package for the Social Science 的英文缩写，意思是社会科学统计软件包。

SPSS 也是 Statistical Product and Service Solutions 的英文缩写，意思是统计产品与服务解决方案。

社会科学统计软件包 (SPSS) 是世界著名的统计分析软件之一。20 世纪 60 年代末，美国斯坦福大学的三位研究生研制开发了最早的统计分析软件 SPSS，并于 1975 年在芝加哥成立了专门研发和经营 SPSS 软件的 SPSS 公司。此时的 SPSS 软件主要运行在中小型计算机上，统称为 SPSSx 版，面向企事业单位用户。20 世纪 80 年代初微型计算机出现，SPSS 公司以其敏锐的市场洞察力和雄厚的技术实力，于 1984 年推出了运行在 DOS 操作系统上的 SPSS 微机版第 1 版，随后又相继推出了第 2 版、第 3 版等，统称为 SPSS/PC+ 版，并确立了个人用户市场第一的地位。90 年代，随着微机 Windows 图形操作系统的出现和盛行，SPSS 公司又研制出了以 Windows 为运行平台的 SPSS 第 5 版、第 6 版。90 年代中后期，为适应用户在 Windows 95 操作系统环境下工作的习惯，并迎合 Internet 的广泛使用，SPSS 第 7 版至第 17 版又相继诞生，统称为 SPSS for Windows 版。



SPSS 软件运行时有多个窗口，各窗口有各自的作用。但要快速入门，只需要熟悉两个基本窗口即可，它们是数据编辑器窗口和结果查看器窗口。

## 一、SPSS 数据编辑器窗口

SPSS 启动后，屏幕会显示如图 1—1 所示的窗口，这就是数据编辑器窗口。

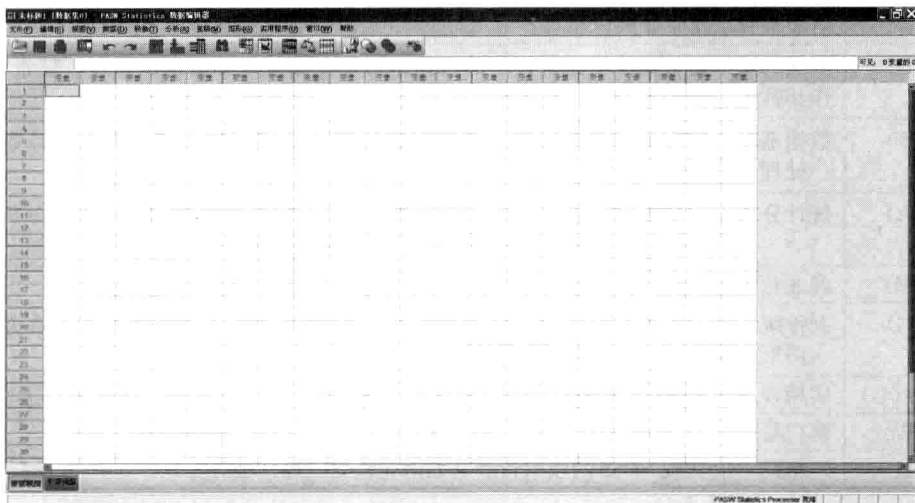


图 1—1 SPSS 数据编辑器窗口

SPSS 数据编辑器窗口是 SPSS 的主程序窗口，该窗口的主要功能是定义 SPSS 数据的结构、录入编辑和管理待分析的数据。其中，【数据视图】选项卡用于显示 SPSS 数据的内容；【变量视图】选项卡用于显示 SPSS 数据的结构。

SPSS 运行时可同时打开多个数据编辑器窗口。各数据编辑器窗口分别显示不同的数据集合（简称数据集）。按打开的先后顺序，各数据集依次自动命名为：数据集 0，数据集 1，数据集 2，等等。其中只有一个数据集为当前数据集，称为活动数据集。用户只能对某一时活动数据集中的数据进行分析。此外，数据集通常以 SPSS 数据文件的特有格式保存在磁盘上，其文件扩展名为 .sav。



sav 文件格式是 SPSS 独有的，一般无法通过 Word，Excel 等其他软件打开。

数据编辑器窗口由窗口主菜单、工具栏、数据编辑区、系统状态显示区组成。

### 1. 窗口主菜单

窗口主菜单列出了 SPSS 常用的数据编辑、加工和分析的功能。用户可以通过点击菜单完成相应的操作。菜单项对应的功能如表 1—1 所示。

表 1-1 主窗口菜单及其功能

菜单名	功能	解释
文件(F)	文件操作	对 SPSS 相关文件进行基本管理 (如新建、打开、保存、打印等)
编辑(E)	数据编辑	对数据编辑器窗口中的数据进行基本编辑 (如撤销/恢复、剪切、复制、粘贴), 并实现数据查找、软件参数设置等功能
视图(V)	窗口外观 状态管理	对 SPSS 窗口外观等进行设置 (如状态栏、表格线、变量值标签等是否显示, 字体设置等)
数据(D)	数据的操 作和管理	对数据编辑器窗口中的数据进行加工整理 (如数据的排序、转置、抽样、分类汇总、加权等)
转换(T)	数据基本 处理	对数据编辑器窗口中的数据进行基本处理 (如生成新变量、计数、分组等)
分析(A)	统计分析	对数据编辑器窗口中的数据进行统计分析和建模 (如基本统计分析、均值比较、相关分析、回归分析、非参数检验等)
直销(M)	商业营销	提供了商业营销分析的专用方法 (如 RFM 分析等)
图形(G)	制作统计 图形	对数据编辑器窗口中的数据生成各种统计图形 (如条形图、直方图、饼图、线图、散点图等)
实用程序(U)	实用程序	SPSS 其他辅助管理 (如显示变量信息、定义变量集、菜单编辑器等)
窗口(W)	窗口管理	对 SPSS 的多个窗口进行管理 (如窗口切换、最小化窗口等)
帮助(H)	帮助	实现 SPSS 的联机帮助 (如语句检索、统计辅导等)



初学 SPSS 时不必掌握菜单中的所有功能, 只需通过浏览菜单了解 SPSS 的大致功能即可。

## 2. 工具栏

同其他常用软件一样, SPSS 将一些常用功能以图形按钮的形式组织在工具栏中。图形按钮的功能都能在窗口主菜单中找到。用户可以直接点击工具栏上的某个按钮完成其相应的功能, 操作更加快捷和方便。当鼠标停留在工具栏按钮上时, 计算机会自动提示相应按钮的功能说明。

## 3. 数据编辑区

数据编辑区是显示和管理 SPSS 数据结构和数据内容的区域。在数据编辑区中有两个视图, 分别是数据视图和变量视图, 依次用来以电子表格的形式录入和编辑管理 SPSS 的数据, 以及定义和修改 SPSS 数据的结构。每条数据都有一个顺序编号显示在编辑区的最左列。数据编辑区中的表格线可以通过【视图(V)】菜单下的【网格线】选项设置成显示或不显示两种状态。

## 4. 系统状态显示区

系统状态显示区用来显示系统的当前运行状态。当系统等待用户操作时, 会出现“PASW Statistics processor 就绪”的提示信息, 该信息可以作为检查 SPSS 是否成功安装的手段。通过【视图(V)】菜单下的【状态栏(S)】选项可以将系统状态区设置成显示或不显示两种状态。



同时编辑多个 SPSS 程序, 它允许将同一批数据的不同分析步骤(程序)分别写在不同的语法窗口中, 进而可将它们以不同的文件名分别存放到磁盘上。应在工具栏的【激活:】后指定所分析的数据集。

语法窗口由窗口主菜单、工具栏、SPSS 程序编辑区组成。其中的大部分含义和功能与数据编辑器窗口和结果查看器窗口类似。除此之外, 增添了【运行(R)】菜单和相应的图形按钮, 用来执行 SPSS 程序。

SPSS 在【运行(R)】菜单中提供了四种程序执行的方式, 它们是:

- 全部执行方式 (【全部】)。即依次执行当前语法窗口中的所有 SPSS 命令程序。
- 选中执行方式 (【选择】)。即仅执行当前语法窗口中当前已被选中的 SPSS 命令程序。选择可以通过鼠标在相应语句行上拖动来完成。
- 当前行执行方式 (【逐步前进(P)】)。即仅执行当前语法窗口中当前光标所在行的 SPSS 命令程序。
- 至末尾行执行方式 (【至结束(T)】)。即仅执行当前语法窗口中当前光标所在行以后的所有 SPSS 命令程序。

以上四种程序执行方式能够帮助程序员方便地调试和运行程序。

在语法窗口中编写 SPSS 程序必须按照一定的语法规则进行, 掌握 SPSS 语法规则、熟悉 SPSS 命令语句是编写 SPSS 程序的前提。因此, 程序运行方式一般适合 SPSS 的高级用户, 或需在 SPSS 程序员的帮助下进行。



SPSS 支持程序编写。掌握 SPSS 语法规则、熟悉 SPSS 命令语句是编写 SPSS 程序的前提。

### 1.2.3 混合运行方式

混合运行方式, 是指在使用菜单的同时编辑 SPSS 程序, 是完全窗口菜单方式和程序运行方式的综合。

为实现混合运行方式, 用户应首先按照菜单运行方式选择统计分析的菜单和选项, 但并不马上点击【确定】按钮提交执行, 而是点击【粘贴(P)】按钮。于是, SPSS 将自动把用户所选择的菜单和选项转换成 SPSS 的命令程序, 并粘贴到当前语法窗口中。然后, 用户可以按照程序运行的方式, 对在语法窗口中生成的 SPSS 命令进行必要的编辑修改, 然后再一次性提交给计算机执行。

可见, 混合运行方式弥补了完全窗口菜单方式中每步分析操作都要人工干预的不足, 同时摆脱了程序运行方式中必须熟记 SPSS 命令和参数的制约, 因此是一种较为灵活且实用的操作方式。另外, 对于熟练的 SPSS 程序员, 可以借助该方式在程序中添加窗口菜单和选项中没有提供的参数。

总之, SPSS 的三种基本使用方式各有千秋, 实际中应根据应用分析的需要和掌握 SPSS 的程度进行合理的选择。



因此,排除数据中那些与目标不关联的干扰因素是数据收集中的重要环节。数据分析并不仅仅是对数据进行数学建模,收集的数据是否真正迎合数据分析的目标,其中是否包含了其他因素的影响,影响程度怎样,应如何剔除这些影响等,都是数据分析过程中必须注意的重要问题。

此外,采用恰当的抽样调查方式是保证正确收集数据的重要手段。通常,在无法获得总体数据,或者获得总体数据需要较大投入时,可利用通过抽样方式获得的来自总体的部分样本数据,对问题进行分析和研究。抽样调查方式分为概率抽样方式和非概率抽样方式两大类。当需利用样本数据对总体的特征参数进行估计和推断时,应采用概率抽样方式。当仅需对问题本身进行刻画而不涉及对总体的推断时,可采用非概率抽样方式。概率抽样(也称随机抽样)是指按一定的概率以随机原则抽取样本,总体包含的每个元素都有一定的机会被抽中进入样本。概率抽样具体包括简单随机抽样、分层抽样、整群抽样、系统抽样、多阶段抽样等。完整的抽样框是概率抽样的基础。相对于概率抽样而言,非概率抽样抽取样本时并不遵循随机原则,而是根据研究目的对数据的要求实施调查。非概率抽样主要包括方便抽样、判断抽样、自愿样本、滚雪球抽样以及配额抽样等。

### 三、加工和整理数据

在明确数据分析目标的基础上收集到的数据,往往要进行必要的加工整理后才能真正用于分析建模。数据的加工整理通常包括数据的缺失值处理、数据的分组、基本描述统计量的计算、基本统计图形的绘制、数据的变换处理等。通过数据的加工整理,能够大致掌握数据的总体分布特征,这是进一步分析和建模的基础。

### 四、选择恰当的统计分析方法进行探索分析

数据加工整理完成后一般就可进行数据分析了。分析时切忌滥用和误用统计分析方法。滥用和误用统计分析方法主要是由于对方法能解决哪类问题、方法适用的前提、方法对数据的要求不清楚等原因造成的。另外,统计软件的不断普及和应用中的不求甚解也会加重这种现象。因此,在数据分析中应避免盲目的“拿来主义”,否则,得到的分析结论可能会偏差较大甚至是错误的。



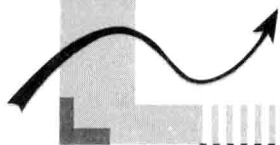
明确统计方法的含义和适用范围是非常关键的。

另外,选择几种统计分析方法对数据进行探索性的反复分析也是极为重要的。每一种统计分析方法都有自己的特点和局限性,因此一般需要选择几种方法反复进行分析,仅依据一种分析方法的结果就断然下结论是不科学的。

### 五、读懂统计分析结果

数据分析的直接结果是统计量。正确理解统计量的统计含义是一切分析结论的基础。正确理解统计量的含义不仅能够帮助人们有效避免毫无根据地随意引用分析结果的错误,同时也是证实分析结论正确性和可信性的依据。而这一切都取决于人们能否正确地把握统计分析方法的核心思想。

另外,将统计分析结果与实际问题相结合也是非常重要的。客观地说,统计方法仅仅是一种有用的数量分析工具,它绝不是万能的。统计方法能否正确地解决各学科的具体问



### 本章要点

1. 明确 SPSS 数据的基本组织方式和数据行列的含义。
2. 掌握应从哪些方面描述 SPSS 数据文件的结构特征。
3. 熟练掌握建立 SPSS 数据文件以及管理 SPSS 数据的基本操作。
4. 熟练掌握在 SPSS 中读取 Excel 工作表数据的基本操作，了解读取文本和数据库数据的基本方法。

建立 SPSS 数据文件是利用 SPSS 软件进行数据分析的首要工作。没有完整且高质量的数据，也就没有值得信赖的数据分析结论。

## 2.1 SPSS 数据文件

建立 SPSS 数据文件，应首先了解 SPSS 数据文件的特点、数据组织的基本方式和相关概念等。只有这样才能够建立一个完整且全面的数据环境，服务于以后的数据分析工作。

### 2.1.1 SPSS 数据文件的特点

SPSS 数据文件是一种有别于其他文件（如 Word 文档、文本文件）的具有特殊性的文件，也称数据集（dataset）。从应用角度理解，这种特殊性表现在两方面：第一，SPSS



**第二部分:**

1. 您当年的高考志愿是由谁确定的?(主要遵循了谁的意见或意愿?)

1. 自己            2. 父母            3. 老师            4. 其他人\_\_\_\_\_

2. 您当年的高考专业志愿主要是按哪些因素确定的?

1. 按您的兴趣爱好            2. 按市场就业前景            3. 按您的职业目标  
4. 按您的能力和学习优势    5. 按您的性格特点            6. 其他\_\_\_\_\_

3. 您是否听过职业规划专门课程或就业指导课程?

1. 没有            2. 听过一些            3. 听过很多

4. 您认为职业规划过程中最大的困难在于\_\_\_\_\_

1. 了解自己的兴趣、个性、能力            2. 确定自己的职业方向  
3. 了解本专业可能的职业通道            4. 其他\_\_\_\_\_

5. 您毕业后的意向: \_\_\_\_\_

1. 出国留学            2. 考研/博            3. 求职工作            4. 自主创业  
5. 还未确定

6. 下表是有关职业生涯问题的各种表述,请依据您的实际情况或想法作答,请在“1~5”中选择一个数字表明您同意的程度,难以回答的可选“0”:

	状态或意愿表述	非常反对—————完全赞成, 不好答					
1	我对所学专业非常感兴趣	1	2	3	4	5	0
2	我很清楚所学专业的未来职业前景	1	2	3	4	5	0
3	我有明确的职业目标	1	2	3	4	5	0
4	我对自己未来的职业很期待	1	2	3	4	5	0
5	我很了解职业生涯规划是什么	1	2	3	4	5	0
6	我认真阅读过一些职业生涯规划 and 就业指导方面的书籍(没有读过选1, 读5本以上选5)	1	2	3	4	5	0
7	我有自己明确的职业生涯规划和发展安排	1	2	3	4	5	0
8	我参加过许多次(5次以上选5)校外社会实践活动或工作实习(没有参加过选1)	1	2	3	4	5	0
9	我对现在大学生就业市场状况很了解	1	2	3	4	5	0
10	我对现在毕业生求职过程很了解	1	2	3	4	5	0
11	选择职业, 与个人兴趣和个性相符最重要	1	2	3	4	5	0
12	选择职业, 专业对口最重要	1	2	3	4	5	0
13	选择职业, 社会声望最重要	1	2	3	4	5	0
14	选择职业, 挣钱多最重要	1	2	3	4	5	0
15	选择职业, 有个人发展空间最重要	1	2	3	4	5	0
16	在校学习成绩与未来工作关系密切	1	2	3	4	5	0

收回有效问卷 919 份。具体数据在可下载的压缩包中, 文件名为“大学生职业生涯规划.sav”。

案例 2—1 的数据就是一份原始数据。在 SPSS 数据编辑器窗口中, 一行存储一份问卷数据, 是一个个案。对于案例 2—1, 919 份问卷在 SPSS 中就有 919 行数据, 即有 919 个个案。

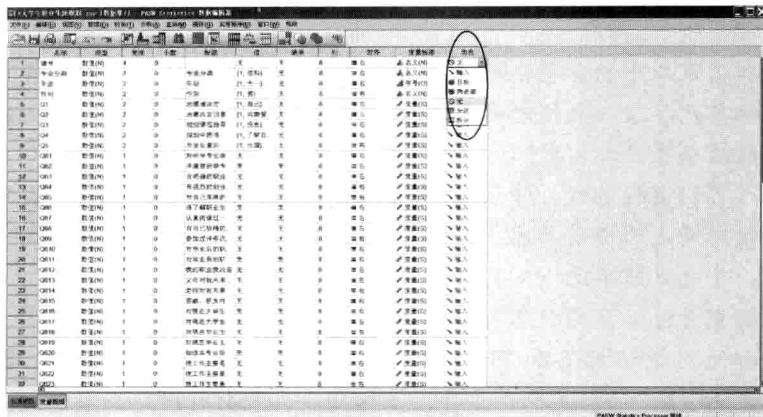


图 2—10 定义变量角色

别值分别作为训练样本集、测试样本集以及验证样本集的标志；另外，有的分类型变量可作为【拆分】角色变量，以方便后续的计算。关于拆分详见 3.8.3 节。

通常建立 SPSS 数据文件时，应首先定义 SPSS 数据的结构部分，然后再输入 SPSS 数据。但实际应用中，边录入、边分析、边修改数据结构的情况也是很常见的。

## 2.3 SPSS 数据的录入与编辑

### 2.3.1 SPSS 数据的录入

SPSS 数据的结构定义好后就可将具体的数据输入到 SPSS 中，以最终形成 SPSS 数据文件。

SPSS 数据的录入操作在数据编辑器窗口的数据视图中实现。其操作方法与 Excel 基本类似，也是以电子表格的方式进行录入，如图 2—11 所示，具体操作不再赘述，但操作时应注意以下几点：

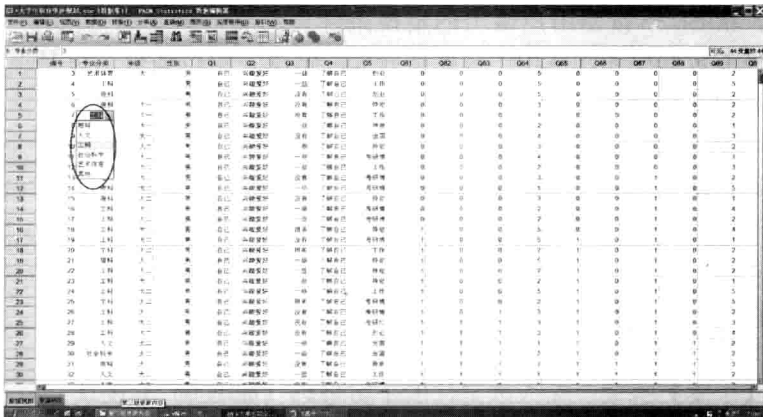


图 2—11 带变量值标签的数据录入





- 数据编辑器窗口中的黄色框框住的单元为当前数据单元，它是当前正在录入或修改数据的单元。录入数据时应首先确定当前数据单元，即将鼠标指到某个数据单元上，然后单击鼠标左键。
- 数据录入可以逐行进行，即录入完一个数据后按 Tab 键。于是当前单元的右边一个单元便自动成为当前单元。数据录入也可以逐列进行，即录入完一个数据后按回车键。于是当前单元下方的单元便自动成为当前单元。
- 录入带有变量值标签的数据可以通过下拉框完成，如图 2—10 所示。但在此之前应首先打开变量值标签的显示开关，选择的菜单是：**【视图(V)】** → **【值标签(V)】**。**【值标签(V)】** 是一个重复开关选项。开关打开时，变量值标签将显示在数据编辑器窗口中；反之，不显示变量值标签，只显示变量值。应当说明的是：在变量值标签显示开关打开的状态下，虽然屏幕显示的是变量值标签，但实际存储的数据仍是变量值。

### 2.3.2 SPSS 数据的编辑

SPSS 数据的编辑主要包括对数据的定位、增加、删除、修改、复制等操作。编辑操作也在数据编辑器窗口的数据视图中进行。

#### 一、SPSS 数据的定位

数据定位的目的是将当前数据单元定位到某个特定的单元中。SPSS 提供了两种定位方式，即人工定位和自动定位。

人工定位，指通过人工浏览数据，确定当前数据单元，适用于数据量较少的情况。用户只需用鼠标拖动数据编辑器窗口右边的滚动钮，或按键盘上的 **Page Up**，**Page Down** 键就可以完成数据的浏览和定位。

自动定位，指 SPSS 按照用户给出的定位条件自动寻找满足条件的第一个数据单元，并把它设置为当前数据单元，适用于数据量较大的情况。定位条件可以是个案的号码，也可以是某个变量的变量值。

##### 1. 按个案号码自动定位

按个案号码自动定位的基本操作如下：

- (1) 将当前单元定位在任何单元中。
- (2) 选择菜单：

**【编辑(E)】** → **【转至个案(S)】**

出现如图 2—12 所示的窗口。

- (3) 选择 **【个案】** 选项，输入欲定位的个案号码（如 200）并确认。

于是，SPSS 自动搜索该号码所对应的个案并定位。另外，该功能也可实现对某个变量列的定位。

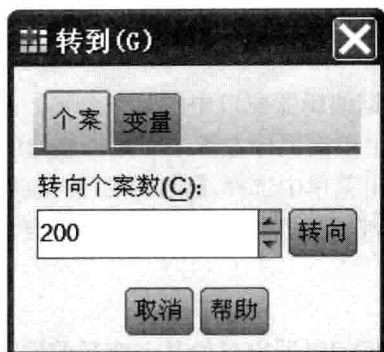


图 2—12 按个案号码自动定位窗口

## 2. 按变量值自动定位

按变量值自动定位的基本操作如下：

- (1) 将当前单元定位在某变量列的任何一条个案上（如专业分类）。
- (2) 选择菜单：

【编辑(E)】→【查找(F)】

出现如图 2—13 所示的窗口。

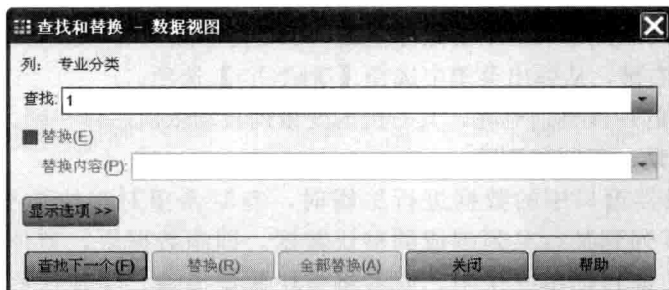


图 2—13 按变量值自动定位窗口

- (3) 输入定位变量值（如 32）并确认。

于是，SPSS 自动对第 (1) 步指定的变量进行变量值搜索，并将当前单元定位到与定位变量值相等的第一条个案（如专业分类等于 1 的个案）上。另外，该功能还可实现变量值的自动替换。

## 二、插入和删除一条个案

### 1. 插入一条个案

插入一条个案，即在数据编辑器窗口的某条个案前插入一条新个案。基本操作步骤如下：

- (1) 将当前数据单元定位在一条个案上。
- (2) 选择菜单：

【编辑(E)】→【插入个案(I)】

于是，SPSS 自动在当前数据单元所在行之前插入一空行，其中，数值型变量的变量

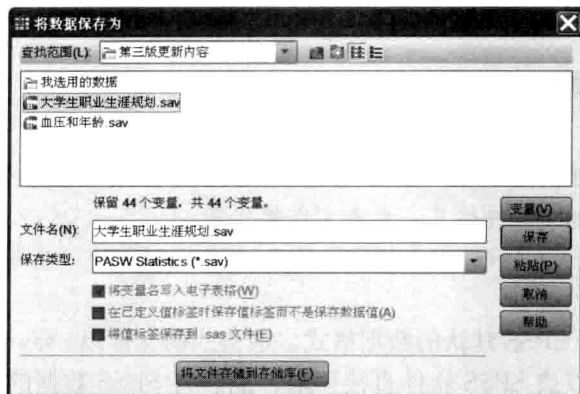


图 2—14 数据保存窗口

- 数据文件的格式通过【保存类型】后的下拉框选择。**变量(V)**按钮的作用是允许用户指定保存哪些变量, 不保存哪些变量, 窗口如图 2—15 所示。变量名前打钩的变量将被保存到磁盘中。

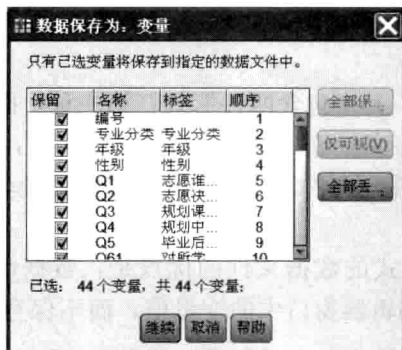


图 2—15 选择保存变量窗口

- 将数据保存为 Excel 文件格式时, 图 2—14 中的【将变量名写入电子表格(W)】选项呈可用状态。它的作用是指定是否将 SPSS 变量名写到 Excel 工作表的第一行上。

## 2.5 读取其他格式的数据文件

以上讲解了建立 SPSS 数据文件的一般方法, 即利用数据编辑器窗口定义 SPSS 数据的结构并输入数据。然而在实际应用中, 可能已经将一批待分析的数据保存在其他软件中。例如, 可能已利用数据库管理系统将一批数据存储在数据库文件中, 或已将数据存在 Excel 电子表格中。如果希望用 SPSS 对这些数据进行统计分析, 就需将这些数据转换到 SPSS 中。SPSS 能够直接将它们读入数据编辑器窗口, 用户可再将其保存为 SPSS 格式文件。因此, 读取其他格式的文件并将其转换为 SPSS 格式数据, 是另外一种建立 SPSS 数据文件的方法。



### 2.5.1 直接读入其他格式的数据文件

SPSS 能够直接打开各种类型的数据文件，常见的格式有：

- SPSS 格式文件，扩展名为 .sav。
- Excel 格式文件，扩展名为 .xls 等。
- SAS 格式文件，扩展名为 .sas7bdat。

基本操作步骤如下：

(1) 选择菜单：

【文件(F)】→【打开(O)】→【数据(A)】

(2) 选择数据文件的类型，并输入数据文件名。

例如：打开名为“大学生职业生涯规划.xls”的文件。



如果读入的是 Excel 格式文件（如图 2—16 所示），SPSS 默认将 Excel 工作表中的所有数据读到 SPSS 数据编辑器窗口中，也可以在【范围】后指定仅读取工作表中某个区域内的数据（如 A5:B10，表示仅读取以 A5 单元为左上角，B10 单元为右下角的矩形区域内的数据）。工作表上的一行数据为 SPSS 中的一条个案。如果 Excel 工作表文件第一行或指定读取区域内的第一行上为变量名，则应选择【从第一行数据读取变量名】项，即以工作表文件第一行或指定读取区域内的第一行上的文字信息作为 SPSS 的变量名；否则不选，SPSS 的变量名将自动取名为工作表中的单元名（如 A，B，C 等）。

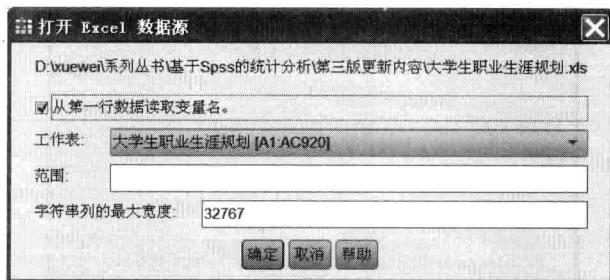


图 2—16 读入 Excel 数据文件

### 2.5.2 使用文本导向读入文本文件

前面已经讲过，文本格式的数据文件是一种最通用格式的数据文件。SPSS 提供了专门读取文本文件的功能。例如，读入名为“大学生职业生涯规划.dat”的文本文件，基本操作步骤如下：

(1) 选择菜单：

【文件(F)】→【打开文本数据(D)】

指定文本文件所在目录和文件名，出现如图 2—17 所示的窗口。应观察该窗口确认

SPSS 是否已经正确地识别和分隔了各数据项。



图 2—17 文本导入窗口 (1)

(2) 点击 **下一步 >(N)** 进入图 2—18 所示的窗口。观察该窗口回答两个问题：第一，数据项间是如何分隔的；第二，数据文件的第一行上是否有变量名，这里应回答否。随后点击 **下一步 >(N)** 进入图 2—19 所示的窗口。

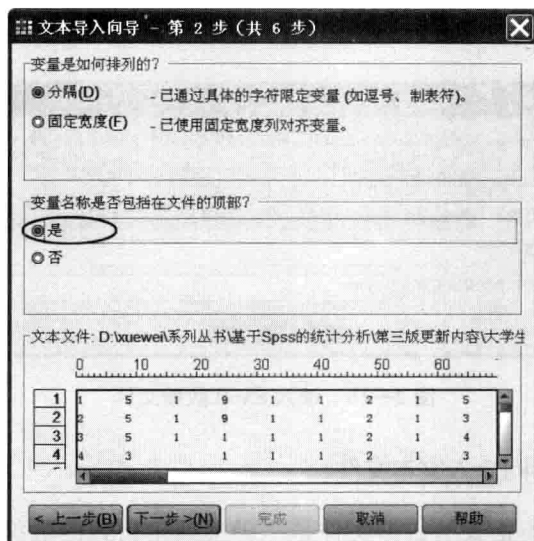


图 2—18 文本导入窗口 (2)

(3) 在图 2—19 所示的窗口中应回答的问题是：

第一，数据部分从文本文件的第几行开始。如果文本文件的第一行是变量名，这里应回答 2。



**案例 2—3**

有两份关于职工基本情况的 SPSS 数据文件，分别如表 2—2 和表 2—3 (a) 所示，表中圆括号中的内容为相应变量的变量名。数据文件在可供下载的压缩包中，文件名分别为“职工数据.sav”和“追加职工.sav”。这里，两份数据文件中的数据项不尽相同，且同一数据项的变量名也不完全一致。现需要将这两份数据合并到一起。

**表 2—2 职工基本情况数据**

职工号 (zgh)	性别 (xb)	年龄 (nl)	基本工资 (sr)	职称 (zc)	学历 (xl)	失业保险 (bx)
001	1	48	1 014.00	1	1	12.00
002	1	49	984.00	2	2	9.00
003	1	54	1 044.00	1	3	13.00
004	1	41	866.00	3	3	8.00
005	1	38	848.00	3	1	8.00
006	2	41	824.00	4	3	7.00
007	2	42	824.00	4	3	7.00
008	2	41	824.00	4	3	7.00
009	2	42	859.00	2	2	8.00
010	1	35	827.00	3	1	7.00
011	1	56	1 014.00	1	2	12.00
012	1	59	989.00	2	2	9.00
013	1	59	938.00	3	4	8.00
014	1	41	889.00	2	1	8.00
015	1	55	887.00	3	4	8.00
016	1	45	887.00	3	4	8.00

**表 2—3 (a) 其他职工的基本情况**

职工号 (zgh)	性别 (xb)	职称 (zcl)	基本工资 (income)
017	2	1	570.00
018	1	1	400.34
019	2	2	690.00
020	1	2	1 003.00
015	1	3	520.00

本例的合并实质是一个纵向合并，纵向合并后的结果如表 2—3 (b) 所示。可以看到，由于两份数据文件中均有 015 号职工的数据，SPSS 将其合并时没有剔除其中的任何一条，而是一并保留下来。由于第二份数据中没有关于年龄、学历和失业保险的数据项，因此合并之后的数据为系统缺失值。



SPSS 中算术表达式的计算是针对每个个案进行的，因此，得到的计算结果也是一系列的，每个个案都有相应的计算结果。

### 3.3.3 SPSS 条件表达式

在变量计算中通常要求对不同组（类）的个案分别按不同的方法进行计算，于是就需要通过一定的方式来指定个案。条件表达式能够帮助实现这一目标。

条件表达式是一个对条件进行判断的式子。其结果有两种取值：如果判断条件成立，则结果为真；如果判断条件不成立，则结果为假。

条件表达式包括简单条件表达式和复合条件表达式。

#### 1. 简单条件表达式

简单条件表达式是由关系运算符、常量、变量以及算术表达式等组成的式子。其中，关系运算符包括  $>$ （大于）、 $<$ （小于）、 $=$ （等于）、 $\neq$ （不等于）、 $\geq$ （大于等于）、 $\leq$ （小于等于）。

例如，对第 2 章表 2—2 中的职工基本情况数据可以写成这样的简单条件表达式： $nl > 35$ ，表示年龄大于 35 岁。在数据编辑器窗口中，对于年龄大于 35 岁的个案，该条件判断的结果为真；而对于年龄小于或等于 35 岁的个案，该条件判断的结果为假。

#### 2. 复合条件表达式

复合条件表达式又称逻辑表达式，是由逻辑运算符、圆括号和简单条件表达式等组成的式子。其中，逻辑运算符包括  $\&$  或 AND（并且）、 $|$  或 OR（或者）、 $\sim$  或 NOT（非）。NOT 的运算最优先，其次是 AND，最后是 OR。可以通过圆括号改变这种运算次序。

例如，对第 2 章表 2—2 中的职工基本情况数据可以写成这样的复合条件表达式： $(nl \leq 35) \text{ and not } (zc < 3)$ ，表示年龄小于等于 35 岁并且职称不低于 3。对于年龄小于等于 35 岁并且职称不低于 3 的个案，该条件判断的结果为真，其余个案的条件判断结果为假。

在变量计算时，如果根据实际需要给出了条件表达式，SPSS 将只对数据编辑器窗口中条件判断结果为真的那些个案进行计算。因此，构造条件表达式是很关键的。

### 3.3.4 SPSS 函数

函数是事先编好并存储在 SPSS 软件中，能够实现某些特定计算任务的一段计算机程序。这些程序段都有各自的名字，称为函数名。执行这些程序段得到的计算结果称为函数值。用户在使用这些函数时，只需通过书写相应的函数名，并给出必要的计算参数，SPSS 便会自动计算函数值。函数书写的具体形式为：

函数名（参数）



图 3—6 中，【固定值(F)】表示随机数种子为一个具体的正整数（该整数应小于等于 2 000 000），一般用于随机化结果需要重复出现的情况；【随机(N)】表示随机数种子每次自动取一个新的值，是 SPSS 默认的选项，这样随机化结果将不会重复出现。

#### 第二，精确选取。

精确抽样要求用户给出两个参数：第一个参数是希望选取的个案数；第二个参数是指定在前几个个案中选取。于是 SPSS 会自动在数据编辑器窗口的前若干个个案中随机精确地抽出相应个数的个案来。

#### 3. 选取某一区域内的样本

即选取数据编辑器窗口中的样本号范围内的所有个案，要求给出这个范围的上下限个案号码。这种选取方法通常适用于时间序列数据。

#### 4. 通过筛选器变量选取

即依据筛选器变量的取值进行样本选取。要求指定一个变量作为筛选器变量，变量值为非 0 或非系统缺失值的个案将被选中。这种方法通常用于排除包含系统缺失值的个案。

### 3.4.3 数据选取的应用举例



#### 案例 3—4

对于第 2 章案例 2—1 的大学生职业生涯规划数据，若希望仅对听过职业规划专门课程或就业指导课程的学生进行分析，应采用上述第一种数据选取方法。

SPSS 数据选取的基本操作步骤如下：

#### (1) 选择菜单：

【数据(D)】→【选择个案】

(2) 在【选择】框中指定选取方法。【全部个案(A)】表示全部选中。本例采用第一种选取方案应选择【如果条件满足(C)】项，出现如图 3—7 所示的窗口。

(3) 指定对未选中个案的处理方式。其中，【过滤掉未选定的个案(F)】表示在未被选中的个案号码上打一个“\”标记，表示暂时筛掉；也可以将未被选中的个案从数据编辑器窗口中删除；或者将筛出的个案复制到一个新的数据编辑器窗口中，应在【数据集名称(S)】后输入数据集名称。一般采取第一种处理方式。



(1) 按上述操作步骤完成数据选取后，以后的 SPSS 分析操作仅针对那些被选中的个案直到用户再次改变数据选取为止。

(2) 采用按指定条件选取和随机选取方法进行数据选取后，SPSS 将在数据编辑器窗口中自动生成一个名为 filter\_ \$ 的新变量，取值为 1 或 0。1 表示本条个案被选中，0 表示未被选中。该变量是 SPSS 产生的中间变量，如果删除它，则自动取消样本选取。



$$K=1+\frac{\ln 16}{\ln 2}=5$$

可以分成 5 组。实际中, 式 (3.1) 可以作为一种参考。

## 2. 组距的确定

组距是一个组的上限(组中的最大值)与下限(组中的最小值)之差。组距可根据全部数据的最大值和最小值及组数来确定, 即

$$\text{组距} = \frac{\text{最大值} - \text{最小值}}{\text{组数}}$$

例如, 对第 2 章表 2-2 的职工基本情况数据, 基本工资的最大值和最小值分别为 1 044 和 824, 即

$$\text{组距} = \frac{1\,044 - 824}{5} = 44$$

可以近似取为 50。

上述问题确定以后, 便可实施分组操作了。

## 3.7.3 数据分组的应用举例



### 案例 3-7

对于第 2 章案例 2-1 的大学生职业生涯规划数据, 计算专业和职业认知得分, 并以 5 为组距进行分组, 以便把握认知得分的分布特征。

SPSS 组距分组的基本操作步骤如下:

(1) 选择菜单:

【转换(T)】→【重新编码为不同变量(R)】

(2) 选择分组变量到【输入变量→输出变量(V)】框中。这里选择 X1。

(3) 在【输出变量】框中的【名称(N)】后输入存放分组结果的变量名, 并点击 **更改(H)** 按钮确认, 这里的变量名为 X1\_new。也可在【标签(L)】后输入相应的变量名标签。

(4) 点击 **旧值和新值(O)** 按钮进行分组区间定义。这里, 应根据分析要求逐个定义各分组区间, 如图 3-11 所示。

(5) 如果仅对符合一定条件的个案分组, 则点击 **如果(I)** 按钮并输入 SPSS 条件表达式。否则, 本步可略。

至此, SPSS 便自动进行组距分组, 并在数据编辑器窗口中创建存放分组结果的名为 X1\_new 的新变量。



在定义分组区间时应注意遵循“不重不漏”的原则。“不重”是指一个变量值只能分在某一个组中, 不能在其他组中重复出现; “不漏”是指所有数据都应分配在某一个组中, 不能遗漏。本例中, 为使各组的组限不重叠, 操作时进行了处理。