

IBM SPSS Statistics 20 简明
指南



注意：使用本信息及其支持的产品之前，请阅读注意事项第 160 页码下的一般信息。

此版本适用于 IBM® SPSS® Statistics 20 及所有后续发布和修订，除非在新版本中另有说明。

Adobe 产品屏幕截图重印已获得 Adobe Systems Incorporated 的许可。

Microsoft 产品屏幕截图重印已获得 Microsoft Corporation 的许可。

受许可保护材料 – IBM 所有

Copyright IBM Corporation 1989, 2011.

美国政府用户受限权利 – 使用、复制或披露受与 IBM Corp. 签订的 GSA ADP Schedule Contract 的限制。

前言

IBM SPSS Statistics 20 简明指南提供了一套旨在使您熟悉 IBM® SPSS® Statistics 的各种组件的教程。本指南可以和软件的所有操作系统版本一起使用，包括：Windows、Macintosh 和 Linux。您可以按顺序通览全部教程，也可以直接阅读您需要较多信息的主题。您可将本指南用作 SPSS Statistics Core 系统附带的联机教程的补充资料，或者忽略该联机教程，从此处的教程入手。

IBM SPSS Statistics 20

IBM® SPSS® Statistics 20 是一种用于分析数据的综合系统。SPSS Statistics 可以从几乎任何类型的文件中获取数据，然后使用这些数据生成分布和趋势、描述统计以及复杂统计分析的表格式报告、图表和图。

SPSS Statistics 使统计分析更易于初学者理解，且更方便有经验的用户使用。简单的菜单和对话框选择使得用户不用键入命令语法即可执行复杂的分析。数据编辑器提供了简单而有效的类似电子表格的工具，用于输入数据和浏览工作数据文件。

Internet 资源

IBM Corp. 网站 (<http://www.ibm.com/support>) 提供了常见问题解答，并提供了对数据文件和其他有用信息的访问途径。

此外，SPSS USENET 讨论组（非 IBM Corp. 主办）向任何感兴趣的人开放。该 USENET 的地址是 `comp.soft-sys.stat.spss`。

您还可以订阅传送到 USENET 组的一个电子邮件消息列表。若要订阅，请发送电子邮件消息至 `listserv@listserv.uga.edu`。电子邮件消息的正文应为：**subscribe SPSSX-L firstname lastname**。然后将电子邮件消息发送至 `listserv@listserv.uga.edu` 即可向该列表张贴消息。

附加出版物

Marija Norušis 编写的 IBM SPSS Statistics Statistical Procedures Companion 已由 Prentice Hall 出版。该书包含了对 IBM® SPSS® Statistics Base 中的过程以及 Logistic 回归和常规线性模型的概述。IBM SPSS Statistics Advanced Statistical Procedures Companion 已由 Prentice Hall 出版。它包含对 Advanced 和回归模块中的过程的概述。

IBM SPSS Statistics 选项

以下选项作为完整（非 Student Version）IBM® SPSS® Statistics 系统的附加增强功能提供：

Statistics Base 为您提供广泛的统计过程以执行基本分析和报告，其中包括计数、交叉表和描述统计、OLAP 立方和码本报告。它还提供了多种降维、分类和细分方法，例如因子分析、聚类分析、最近邻元素分析和判别函数分析。此外，SPSS Statistics Base 还提供了广泛的平均值比较算法和预测方法，例如 t 检验、方差分析、线性回归和序数回归。

Advanced Statistics 主要提供一些复杂试验和生物医学研究中常用的技术。它包括一些用于一般线性模型 (GLM)、线性混合模型、方差成分分析、对数线性分析、序数回归、保险精算寿命表、Kaplan-Meier 生存分析以及基础和扩展 Cox 回归的过程。

Bootstrap 方法可以导出稳健的标准误估计值，并能为诸如均值、中位数、比例、几率比、相关系数或回归系数等估计值导出置信区间。

Categories 执行最优尺度过程，包括对应分析。

Complex Samples 使得调查、市场、卫生和民意研究者以及使用抽样调查方法的社会学家能够将他们复杂的样本设计并入数据分析中。

Conjoint 提供度量单个产品属性如何影响消费者和市民偏好的现实的方式。使用 Conjoint，您可以轻松地在一组产品属性的环境中度量每个产品属性的折衷效应一如同消费者在进行采购决策时所做的一样。

Custom Tables 可创建各种具有演示质量的表格式报表，包括复杂的行列表和多重响应数据的显示。

Data Preparation 提供数据的快捷可视快照。使用它可以应用标识无效数据值的验证规则。您可以创建标记超出范围的值、缺失值或空值的规则。还可以保存记录个别规则的违反和每个个案的规则违反总数的变量。还提供了可以复制或修改的一组有限的预定义规则。

Decision Trees 创建基于树的分类模型。它将个案分为若干组，或根据自变量（预测变量）的值预测因变量（目标变量）的值。此过程为探索性和证实性分类分析提供验证工具。

Direct Marketing 使组织能够通过专为直销设计的方法确保其营销计划尽可能地发挥效力。

Exact Tests 可在小样本或分布非常不均匀的样本可能导致常规检验不准确的情况下计算统计检验的精确的 p 值。此选项只在 Windows 操作系统中可用。

Forecasting 通过使用多种曲线拟合模型、平滑模型和用于估计自回归函数的方法，执行综合的预测和时间序列分析。

Missing Values 描述了缺失数据的模式、估计均值和其他统计量，并为缺失观察值归因值。

Neural Networks 可以通过将产品需求预测为价格函数以及其他变量的函数或根据购买习惯和人口统计特征分类客户来制定经营决策。神经网络是非线性数据建模工具。它们可以用来建立输入与输出之间的复杂关系模型，也可用来查找数据中的模式。

Regression 提供了用于分析那些不能拟合传统线性统计模型的数据的方法。它包括一些用于 probit 分析、logistic 回归、权重估计、两阶段最小平方回归和常规非线性回归的过程。

Amos (矩™结构的™分析™) 使用结构化方程建模以确认和解释涉及态度、观念和其他驱动行为的因素的概念模型。

培训讲座

IBM Corp. 提供公开的以及现场的 IBM® SPSS® Statistics 培训讲座。所有讲座都是以实践小组为特色的。讲座将定期在美国和欧洲的主要城市举办。有关这些讲座的更多信息，请前往 <http://www.ibm.com/software/analytics/spss/training/>。

技术支持

我们提供有“技术支持”以维护客户。客户可就 IBM Corp. 产品使用或某一受支持硬件环境的安装帮助寻求技术支持。如要联系技术支持，请访问 IBM Corp. 网站，网址为 <http://www.ibm.com/support>。在请求协助时，请准备好您和您组织的 ID 以及支持协议。

IBM SPSS Statistics 20 Student Version

IBM® SPSS® Statistics 20 Student Version 是 SPSS Statistics 的受限版本，但功能依然强大。

功能

Student Version 包含 IBM® SPSS® Statistics 中所包含的许多重要的数据分析工具，包括：

- 用于输入、修改和查看数据文件的类似电子表格的数据编辑器。
- 统计过程，包括 t 检验、方差分析和交叉制表。
- 允许动态更改或添加图表元素和变量的交互式图形；一旦指定更改，更改随即显示。
- 大量分析和演示图表和表的标准高分辨率图形。

限制

Student Version 专为课堂教学而创建，仅限学生使用和教师用于教育目的。以下限制适用于 IBM® SPSS® Statistics 20 Student Version：

- 数据文件不能包含 50 个以上的变量。
- 数据文件不能包含多于 1,500 个的个案。SPSS Statistics 附加模块（如 Regression 或 Advanced Statistics）不能和 Student Version 一同使用。
- SPSS Statistics 用户无法使用命令语法。这意味着不能在像在 IBM® SPSS® Statistics 完整版中那样通过在语法或“工作”文件中保存一系列命令来重复分析。
- 用户无法使用脚本和自动化。这意味着不能像在 SPSS Statistics 完整版中那样，创建自动执行经常重复的任务的脚本。

针对学生的技术支持

如果您是使用任何学生版 IBM SPSS 软件产品的学生，请访问我们专为学生提供的在线教育解决方案 (<http://www.ibm.com/spss/rd/students/>) 页面。如果您是使用大学提供的 IBM SPSS 软件副本的学生，请联系所在大学的 IBM SPSS 产品协调员。

针对导师的技术支持

将 Student Version 用于课堂教学的导师可以联系技术支持以寻求帮助。在美国和加拿大，要联系技术支持，请拨打 (312) 651-3410，或者前往 <http://www.ibm.com/support>。

内容

1	简介	1
	样本文件	1
	打开数据文件	1
	运行分析	3
	查看结果	7
	创建图表	8
2	读取数据	11
	IBM SPSS Statistics 数据文件的基本结构	11
	读取 IBM SPSS Statistics 数据文件	11
	从电子表格中读取数据	12
	从数据库中读取数据	14
	从文本文件读取数据	20
3	使用数据编辑器	28
	输入数值数据	28
	输入字符串数据	30
	定义数据	32
	添加变量标签	32
	更改变量类型和格式	32
	为数值变量添加值标签	33
	为字符串变量添加值标签	35
	使用值标签进行数据输入	36
	处理缺失数据	37
	数值变量的缺失值	38
	字符串变量的缺失值	39
	复制并粘贴变量属性	40
	定义分类变量的变量属性	44
4	使用多数据源	51
	多数据源的基本处理	52
	使用命令语法中的多个数据集	53

在数据集之间复制和粘贴信息	54
重命名数据集	54
不显示多个数据集	54
5 检查单个变量的摘要统计量	56
度量水平	56
分类数据的摘要测量	56
分类数据的图表	58
刻度变量的摘要测量	59
刻度变量的直方图	61
6 创建和编辑图表	63
图表创建的基础知识	63
使用图表生成器库	64
定义变量和统计量	65
添加文字	67
创建图表	68
图表编辑基础知识	69
选择图表元素	70
使用“属性”窗口	70
更改条颜色	71
设置刻度标记标签中数字的格式	73
编辑文本	75
显示数据值标签	76
使用模板	77
定义图表选项	82
7 使用输出	86
使用查看器	86
使用枢轴表编辑器	87
访问输出定义	87
透视表	88
创建并显示层	90
编辑表	91

隐藏行与列	92
更改数据显示格式	92
表格外观	94
使用预先定义的格式	94
自定义表格外观样式	95
更改默认的表格格式	98
自定义初始显示设置	99
显示变量和值标签	101
在其他应用程序中使用结果	103
将结果作为 Word 表格粘贴	103
将结果作为文本进行粘贴	104
将结果导出到 Microsoft Word、PowerPoint 和 Excel 文件	106
将结果导出到 PDF	113
将结果导出到 HTML	116
8 使用语法	117
粘贴语法	117
编辑语法	118
打开并运行语法文件	120
了解错误窗格	121
使用分界点	121
9 修改数据值	124
从刻度变量创建分类变量	124
计算新变量	130
使用表达式中的函数	132
使用条件表达式	134
使用日期和时间	135
计算两个日期之间的时间长度	136
向日期添加期间	140
10 排序和选择数据	144
排序数据	144

拆分文件处理	145
为拆分文件处理排序个案.	146
打开和关闭拆分文件处理.	147
选择个案的子集	147
基于条件表达式选择个案.	148
要选择随机样本请执行下列操作	149
选择时间范围或个案范围.	150
未选定的个案的处理	151
个案选择状态	151

附录

A 样本文件	153
B 注意事项	160
索引	162

简介

本指南提供一套旨在让您能够对数据执行有用分析的教程。您可以按顺序通览全部教程，也可以直接阅读您需要较多信息的主题。

本章将向您介绍基本功能并演示一个典型会话。我们将检索以前定义的 IBM® SPSS® Statistics 数据文件，然后生成简单的统计摘要和图表。

在后面的章节中将进一步提供有关本章中所涉及的许多主题的更详细说明。在此，我们希望能够为您了解后面的教程提供一个基本框架。

样本文件

此处提供的大多数示例都使用数据文件 demo.sav。此数据文件是对数千人的虚拟调查，包括基本的人口统计和消费者信息。

如果使用的是学生版，则您的 demo.sav 版本是一个具有代表性的样本，它将原始数据文件缩减为满足 1,500 个案限制。使用该数据文件获得的结果将与此处显示的结果有所不同。

随产品一起安装的样本文件可以在安装目录的 Samples 子目录中找到。对于以下每种语言在“样本”子目录中有单独的文件夹：英语、法语、德语、意大利语、日语、韩语、波兰语、俄语、简体中文、西班牙语和繁体中文。

并非所有样本文件均提供此处的全部语言版本。如果样本文件未提供某种语言的版本，则相应语言文件夹中包含该样本文件的英语版本。

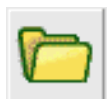
打开数据文件

要打开数据文件：

- ▶ 从菜单中选择：
文件 > 打开 > 数据...

或者，也可以使用工具栏上的“打开文件”按钮。

图片 1-1
“打开文件”工具栏按钮



将显示打开文件的对话框。

默认情况下，将显示 IBM® SPSS® Statistics 数据文件 (.sav 扩展名)。

本示例使用文件 demo.sav。

图片 1-2
“数据编辑器”中的 demo.sav 文件

	age	marital	address	income	inccat	car	carcat
1	55	M	12	72.00	\$50 - \$74	36.20	Luxury
2	56	Unmarried	29	153.00	\$75+	76.90	Luxury
3	28	Married	9	28.00	\$25 - \$49	13.70	Economy
4	24	Married	4	26.00	\$25 - \$49	12.50	Economy
5	25	Unmarried	2	23.00	Under \$25	11.30	Economy
6	45	Married	9	76.00	\$75+	37.20	Luxury
7	42	Unmarried	19	40.00	\$25 - \$49	19.80	Standard
8	35	Unmarried	15	57.00	\$50 - \$74	28.20	Standard
9	46	Unmarried	26	24.00	Under \$25	12.20	Economy
10	34	Married	0	89.00	\$75+	46.10	Luxury

数据文件将显示在“数据编辑器”中。在“数据编辑器”中，如果将鼠标光标放在变量名称（列标题）上，则会显示更具描述性的变量标签（如果已为该变量定义了标签）。

默认情况下，将显示实际的数据值。要显示标签：

- ▶ 从菜单中选择：
视图 > 值标签

或者，也可以使用工具栏上的“值标签”按钮。

图片 1-3
“值标签”按钮



当前显示了描述性值标签以便更轻松地解释响应。

图片 1-4
“数据编辑器”中显示的值标签

	age	marital	address	income	inccat	car	carcat
1	55	Married	12	72.00	\$50 - \$74	36.20	Luxury
2	56	Unmarried	29	153.00	\$75+	76.90	Luxury
3	28	Married	9	28.00	\$25 - \$49	13.70	Economy
4	24	Married	4	26.00	\$25 - \$49	12.50	Economy
5	25	Unmarried	2	23.00	Under \$25	11.30	Economy
6	45	Married	9	76.00	\$75+	37.20	Luxury
7	42	Unmarried	19	40.00	\$25 - \$49	19.80	Standard
8	35	Unmarried	15	57.00	\$50 - \$74	28.20	Standard
9	46	Unmarried	26	24.00	Under \$25	12.20	Economy
10	34	Married	0	89.00	\$75+	46.10	Luxury

运行分析

如果您有任何附加选项，“分析”菜单将包含报告和统计分析类别的列表。

我们将从创建简单的频率表（计数表）开始。本示例需要“Statistics Base”选项。

- ▶ 从菜单中选择：
分析 > 描述统计 > 频率...

此时，将显示“频率”对话框。

图片 1-5
“频率”对话框

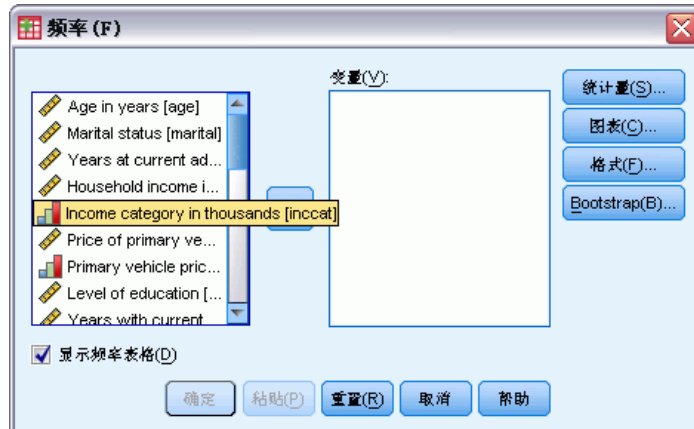


每个变量旁边的图标提供了有关数据类型和测量级别的信息。

	数值	字符串	日期	时间
尺度（连续）		n/a		
有序				
名义				

- ▶ 单击变量“Income category in thousands [inccat]”。

图片 1-6
“频率”对话框中的变量标签和名称



如果变量标签和/或名称在列表中显示为截断，则将光标置于其上时会显示完整的标签/名称。变量名称 `inccat` 显示在描述性变量标签后的方括号中。Income category in thousands 为变量标签。如果没有变量标签，则列表框中将仅显示变量名称。

您可以像窗口那样，单击并拖动外部边框或角落来调整对话框大小。例如，如果使对话框变宽，变量列表也将变得更宽。

图片 1-7
“调整大小”对话框



在对话框中，从左侧的源列表中选择要分析的变量并拖放到右侧的“变量”列表中。只有在“变量”列表中至少放置了一个变量后，用于运行分析的确定按钮才会处于启用状态。

在许多对话框中，您可以通过右键单击列表中的任何变量名称来获取附加信息。

- ▶ 右键单击Income category in thousands [inccat]并选择变量信息。
- ▶ 单击“值标签”下拉列表上的向下箭头。

图片 1-8
为收入变量定义的标签



将显示为该变量定义的所有值标签。

- ▶ 单击源变量列表中的 Gender [gender]，并将该变量拖至目标“变量”列表中。
- ▶ 单击源列表中的 Income category in thousands [inccat]，并将其拖至目标列表。

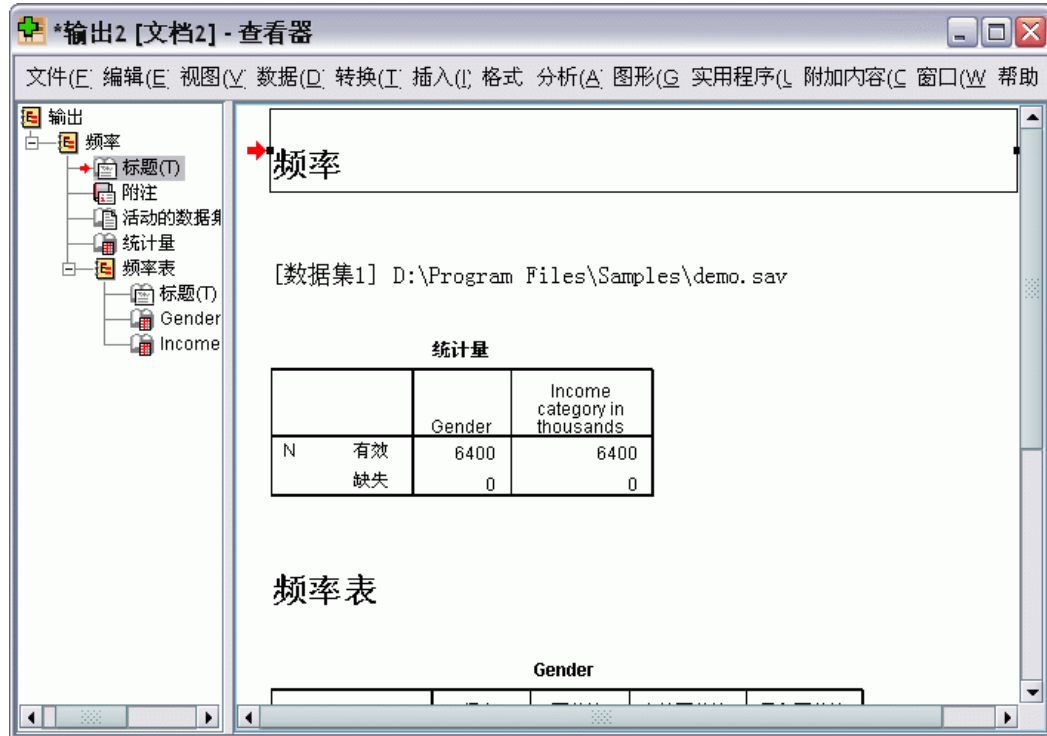
图片 1-9
为分析选择的变量



- ▶ 单击确定以运行该过程。

查看结果

图片 1-10
浏览器窗口

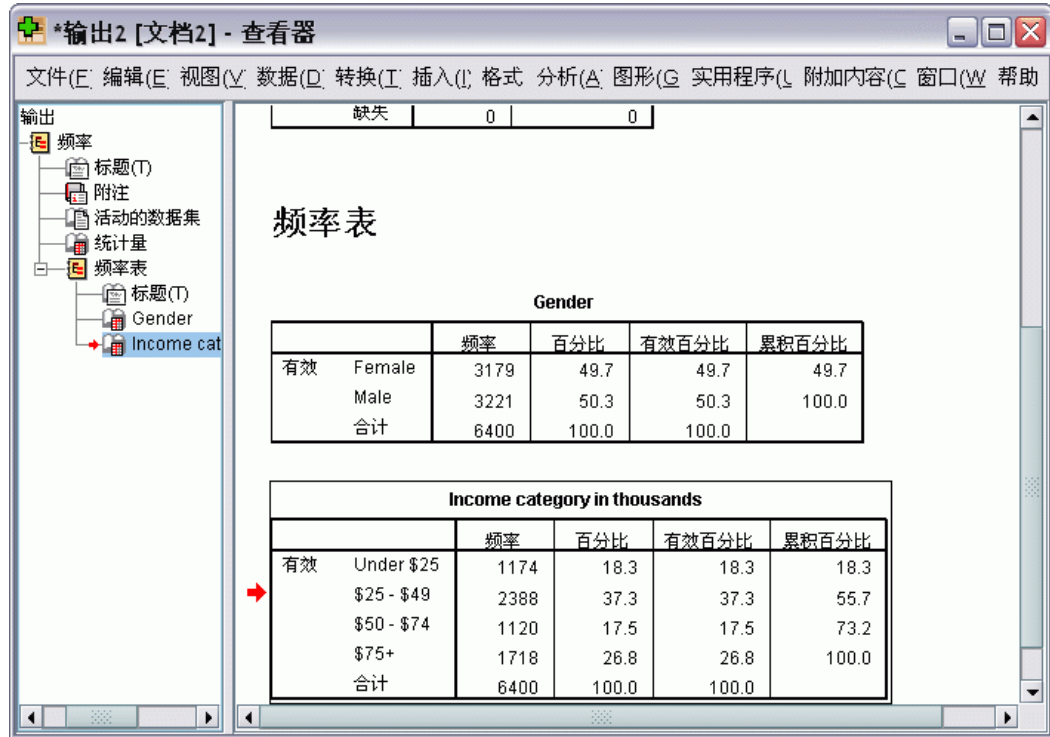


结果将显示在浏览器窗口中。

您可以通过在概要窗格中进行选择以快速转到浏览器中的任意项。

- ▶ 单击 Income category in thousands [inccat]。

图片 1-11
收入类别的频率表



此时将显示收入类别的频率表。此频率表显示了每个收入类别中的人数和百分比。

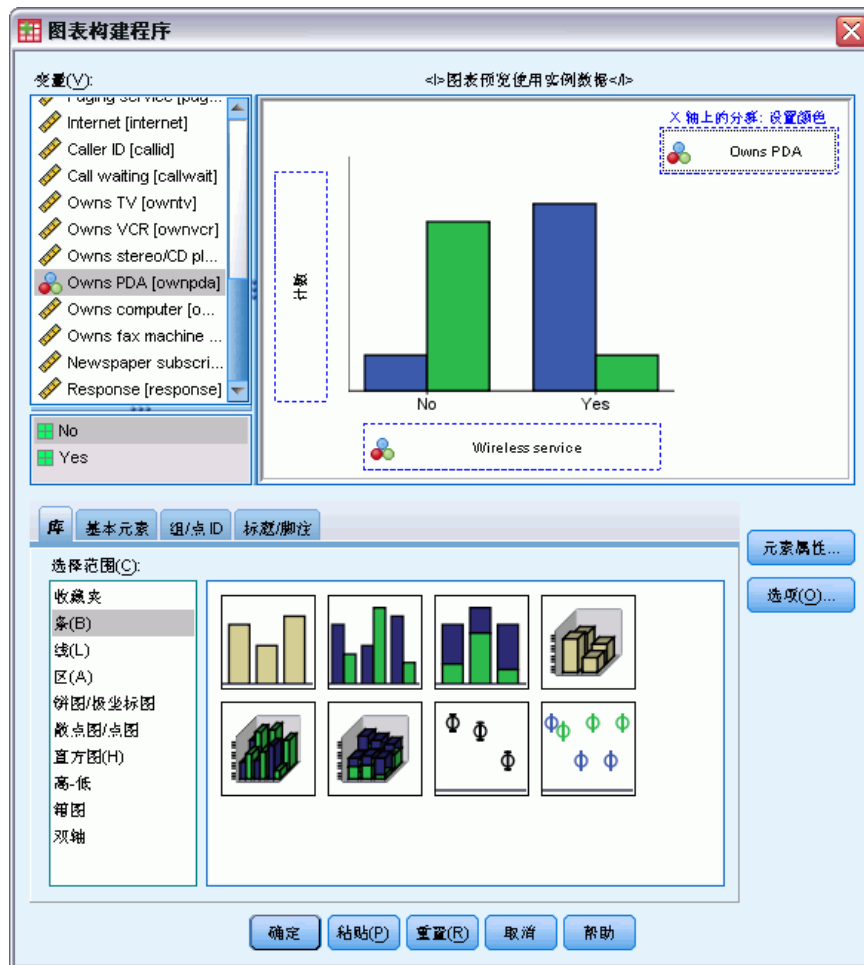
创建图表

虽然某些统计过程可以创建图表，但您也可以使用“图形”菜单来创建图表。

例如，您可以创建一个显示无线电话服务与 PDA（个人数字助理）持有情况之间关系的图表。

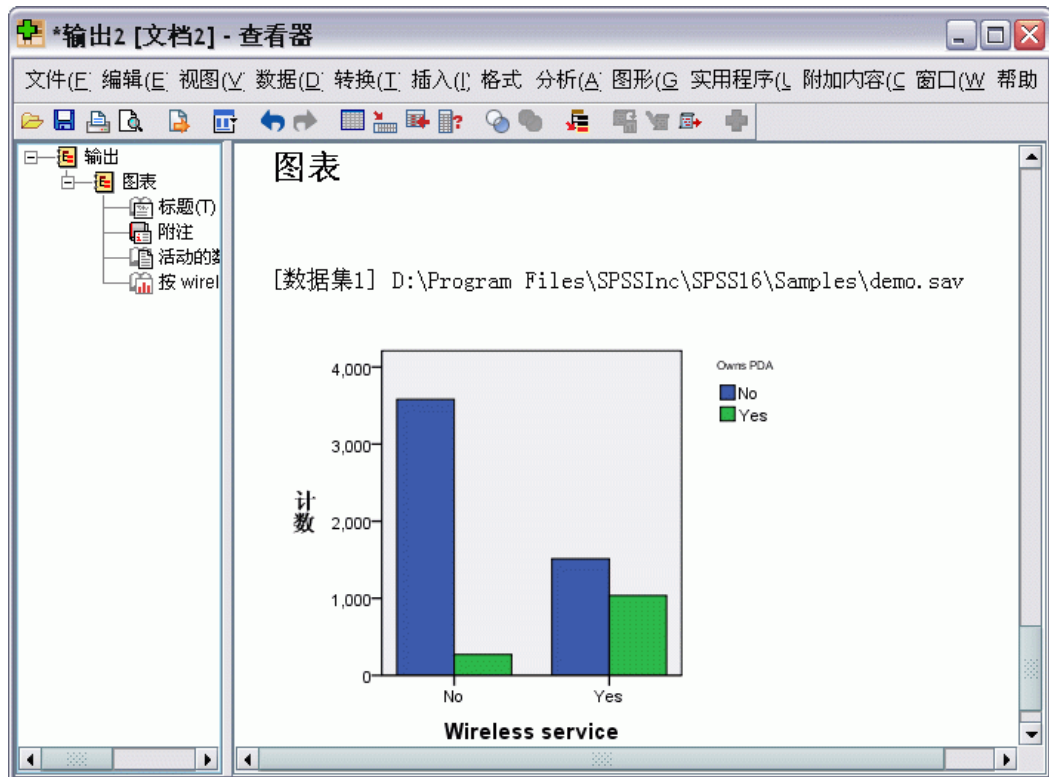
- ▶ 从菜单中选择：
图形 > 图表构建程序...
- ▶ 单击库选项卡（如果尚未将其选中）。
- ▶ 单击条形图（如果尚未将其选中）。
- ▶ 将“聚类条形图”图标拖到画布上，即“库”上方的大片区域。

图片 1-12
“图表生成器”对话框



- ▶ 向下滚动“变量”列表，右键单击 Wireless service [wireless]，然后选择名义作为其测量级别。
- ▶ 将 Wireless service [wireless] 变量拖动到 x 轴。
- ▶ 右键单击 Owns PDA [ownpda] 并选择名义作为其测量级别。
- ▶ 将 Owns PDA [ownpda] 变量拖到画布右上角的聚类拖放区域。
- ▶ 单击确定以创建图表。

图片 1-13
“浏览器”窗口中显示的条形图



条形图将显示在浏览器中。该图表显示了使用无线电话服务的人员远比不使用无线服务的人员更可能拥有 PDA。

您可以通过在浏览器窗口的内容窗格中双击来编辑图表和表，将结果复制并粘贴到其他应用程序。稍后将对这些主题进行介绍。

读取数据

可以直接输入数据，也可以从许多不同的数据源中导入数据。本章将对 IBM® SPSS® Statistics 数据文件、电子表格应用程序（如 Microsoft Excel）、数据库应用程序（如 Microsoft Access）和文本文件中所存储数据的读取过程进行详细讨论。

IBM SPSS Statistics 数据文件的基本结构

图片 2-1
数据编辑器



	age	marital	address	income	inccat	car	carcat
1	55	1	12	72.00	\$50 - \$74	36.20	Luxur
2	56	0	29	153.00	\$75+	76.90	Luxur
3	28	1	9	28.00	\$25 - \$49	13.70	Econom
4	24	1	4	26.00	\$25 - \$49	12.50	Econom
5	25	0	2	23.00	Under \$25	11.30	Econom
6	45	1	9	76.00	\$75+	37.20	Luxur
7	42	0	19	40.00	\$25 - \$49	19.80	Standar
8	35	0	15	57.00	\$50 - \$74	28.20	Standar
9	46	0	26	24.00	Under \$25	12.20	Econom
10	34	1	0	89.00	\$75+	46.10	Luxur
11	55	1	17	72.00	\$50 - \$74	35.50	Luxur

IBM® SPSS® Statistics 数据文件是按个案（行）和变量（列）组织的。在此数据文件中，个案表示各个调查对象。变量表示对调查中提出的每个问题的回答。

读取 IBM SPSS Statistics 数据文件

IBM® SPSS® Statistics 数据文件的文件扩展名为 `.sav`，包含您保存的数据。要打开 `demo.sav`（随该产品一起安装的示例文件）：

- ▶ 从菜单中选择：
文件 > 打开 > 数据...

- ▶ 浏览至 demo.sav 并将其打开。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。

数据将立即显示在数据编辑器中。

图片 2-2
打开的数据文件

	age	marital	address	income	inccat	car	carcat
1	55	1	12	72.00	\$50 - \$74	36.20	Luxur
2	56	0	29	153.00	\$75+	76.90	Luxur
3	28	1	9	28.00	\$25 - \$49	13.70	Econom
4	24	1	4	26.00	\$25 - \$49	12.50	Econom
5	25	0	2	23.00	Under \$25	11.30	Econom
6	45	1	9	76.00	\$75+	37.20	Luxur
7	42	0	19	40.00	\$25 - \$49	19.80	Standar
8	35	0	15	57.00	\$50 - \$74	28.20	Standar
9	46	0	26	24.00	Under \$25	12.20	Econom
10	34	1	0	89.00	\$75+	46.10	Luxur
11	55	1	17	72.00	\$50 - \$74	35.50	Luxur

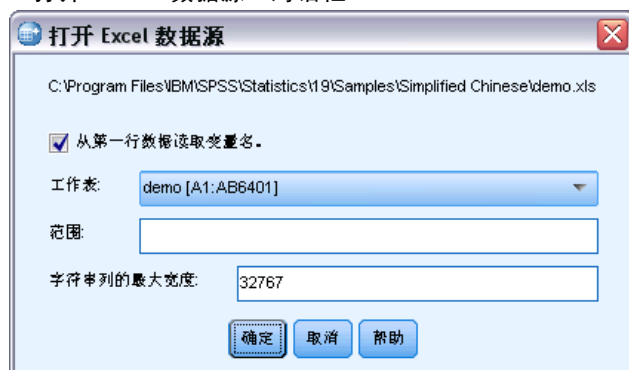
从电子表格中读取数据

您可以从应用程序如 Microsoft Excel 中读取数据，而不是向数据编辑器中直接键入所有数据。也可以将列标题作为变量名进行读取。

- ▶ 从菜单中选择：
文件 > 打开 > 数据...
- ▶ 选择 Excel (*.xls) 作为您要查看的文件类型。
- ▶ 打开 demo.xls。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。

将显示“打开 Excel 数据源”对话框，在该对话框中可以指定是否将变量名包括在电子表格和想要导入的单元格中。在 Excel 95 或更高版本中，还可以指定要导入哪些工作表。

图片 2-3
“打开 Excel 数据源”对话框



- 确保选择了从第一行数据读取变量名。选择此选项可将列标题作为变量名进行读取。

如果列标题不符合 IBM® SPSS® Statistics 变量命名规则，则这些列标题将会转换成有效的变量名，原始的列标题将另存为变量标签。如果只想导入电子表格的一部分内容，请在“范围”文本框中指定要导入的单元格范围。

- 单击确定以读取 Excel 文件。

数据将立即显示在数据编辑器中，其中的列标题被用作变量名。由于变量名中不能包含空格，所以原始的列标题中的空格已删除。例如，Excel 文件中的 Marital status 变成了变量 Maritalstatus。原始的列标题保留为变量标签。

图片 2-4
导入的 Excel 数据

	Age	Maritalstatus	Address	Income	IncomeCategory	Carpc
1	55	1	12	72.00	3.00	
2	58	0	29	153.00	4.00	
3	28	1	9	28.00	2.00	
4	24	1	4	28.00	2.00	
5	25	1	2	23.00	1.00	
6	45	0	9	76.00	4.00	
7	44	1	17	144.00	4.00	
8	46	1	20	75.00	4.00	
9	41	0	10	28.00	2.00	
10	29	0	4	19.00	1.00	

从数据库中读取数据

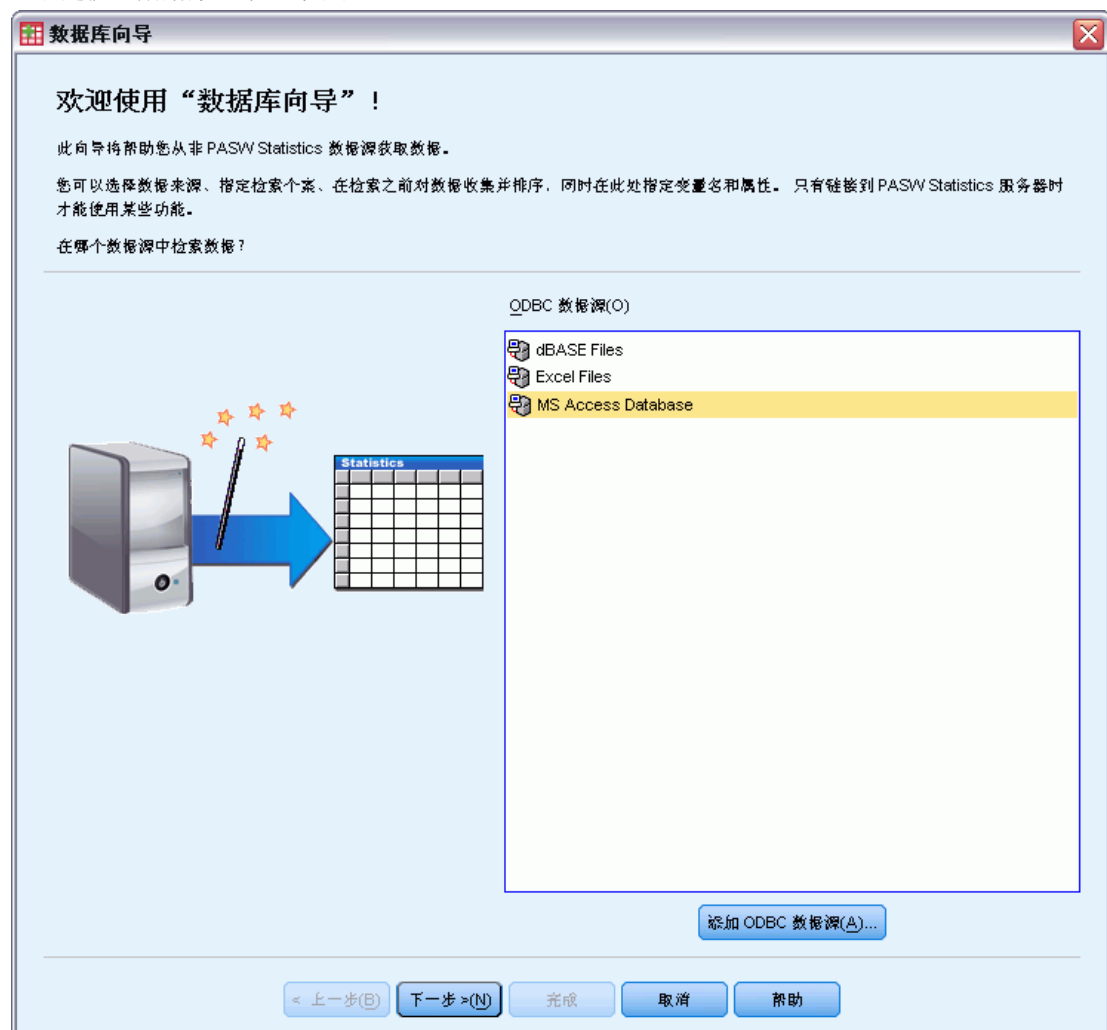
使用数据库向导可以很容易地从数据库源导入数据。在安装 ODBC（开放数据库连接）驱动程序之后，可直接读取使用 ODBC 驱动程序的任何数据库。安装 CD 中提供了适用于多种数据库格式的 ODBC 驱动程序。其他驱动程序可从第三方供应商那里获得。本示例中将对最常用的数据库应用程序之一 Microsoft Access 进行讨论。

注意：本示例特定于 Microsoft Windows 且 Access 需要 ODBC 驱动程序。其他平台上的步骤与此类似，不过可能需要 Access 的第三方 ODBC 驱动程序。

► 从菜单中选择：

文件 > 打开数据库(B) > 新建查询...

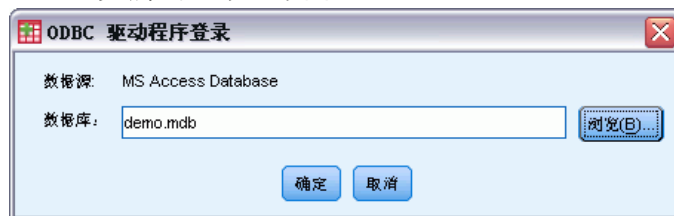
图片 2-5
“欢迎使用数据库向导”对话框



► 从数据源列表中选择 MS Access Database，然后单击下一步。

注意：根据您的安装情况，您可能在该向导左侧看到一个 OLEDB 数据源列表（仅限于 Windows 操作系统），但本示例使用该向导右侧显示的 ODBC 数据源列表。

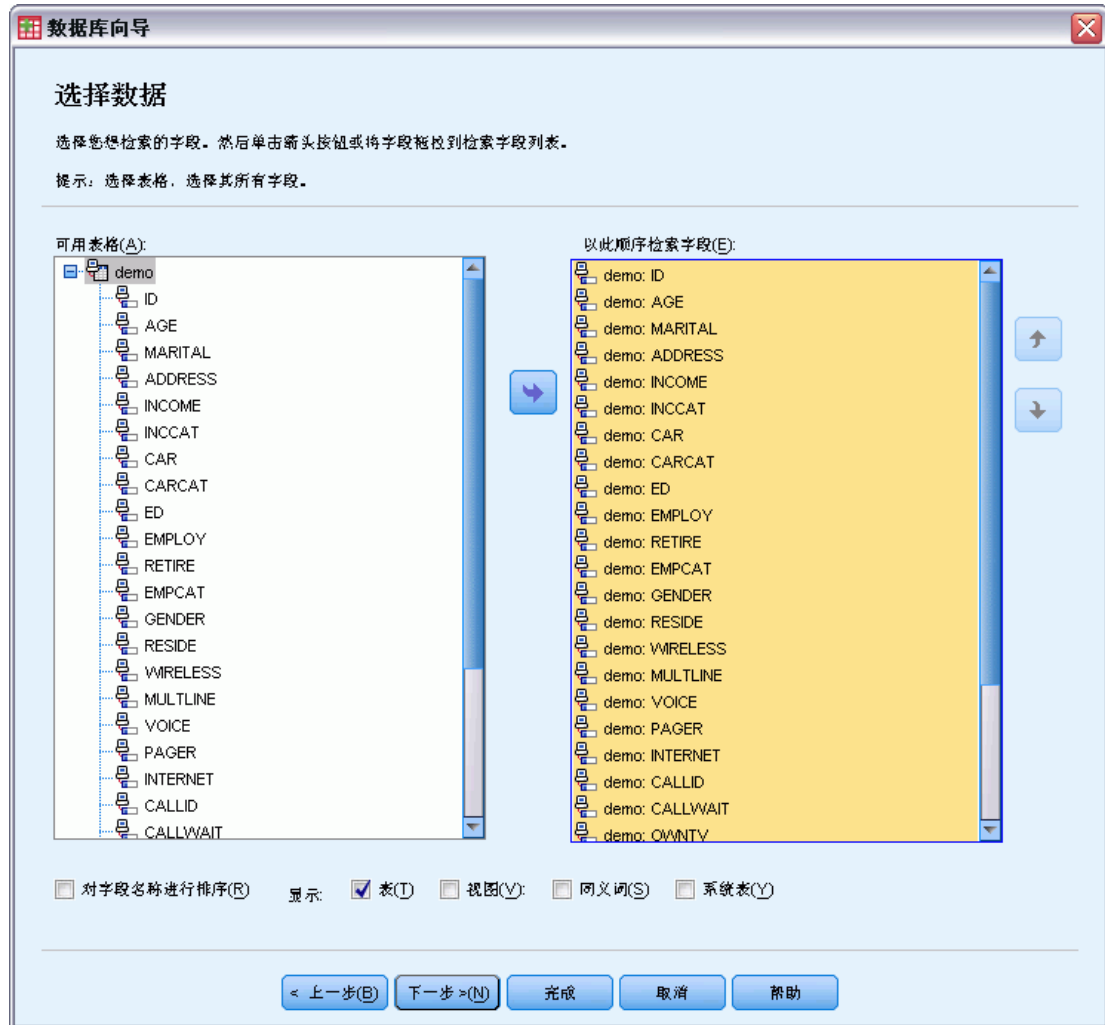
图片 2-6
“ODBC 驱动程序登录”对话框



- ▶ 单击浏览导航到要打开的 Access 数据库文件。
- ▶ 打开demo.mdb。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。
- ▶ 在登录对话框中单击确定。

在下一步骤中，可以指定要导入的表和变量。

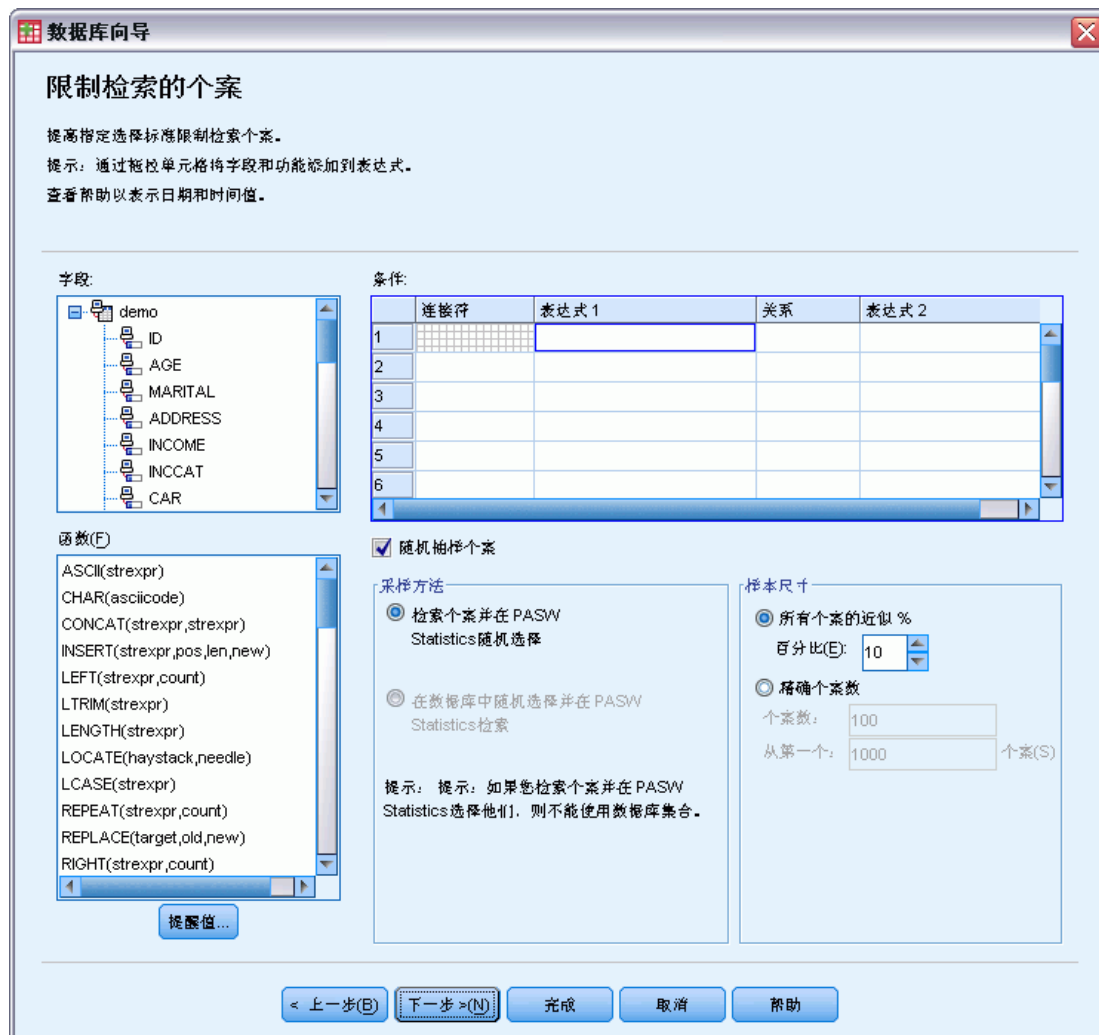
图片 2-7
“选择数据”步骤



- ▶ 将整个 demo 表拖到“以此顺序检索字段”列表中。
- ▶ 单击下一步。

在下一步骤中，选择要导入的记录（个案）。

图片 2-8
“限制检索的个案”步骤



如果不想导入所有个案，则可以导入一个个子集（例如，年龄大于 30 岁的男士），或从数据源中导入一个随机的个案示例。对于大数据源，您可能需要将个案数限制为小的、具有代表性的样本，以减少处理时间。

- ▶ 单击下一步继续。

可以使用字段名称来创建变量名。如有必要，可将这些名称转换为有效的变量名。原始字段名称将保留为变量标签。您也可以在导入数据库之前更改变量名。

图片 2-9
“定义变量”步骤



- ▶ 单击“Gender”字段中的重新编码为数值单元格。使用此选项可将字符串变量转换成整数变量，并将原始值保留为新变量的值标签。
- ▶ 单击下一步继续。

根据在数据库向导中的选择所创建的 SQL 语句将在“结果”步骤中出现。可以立即执行此语句，或将其保存到一个文件中以供将来使用。

图片 2-10
“结果”步骤



- ▶ 单击完成即可导入数据。

现在可以在“数据编辑器”中处理在 Access 数据库中选择要导入的所有数据。

图片 2-11
从 Access 数据库中导入的数据



	Age	Maritalstatus	Address	Income	IncomeCategory	Carp
1	55	1	12	72.00	3.00	
2	56	0	29	153.00	4.00	
3	28	1	9	28.00	2.00	
4	24	1	4	26.00	2.00	
5	25	1	2	23.00	1.00	
6	45	0	9	78.00	4.00	
7	44	1	17	144.00	4.00	
8	46	1	20	75.00	4.00	
9	41	0	10	26.00	2.00	
10	29	0	4	19.00	1.00	

从文本文件读取数据

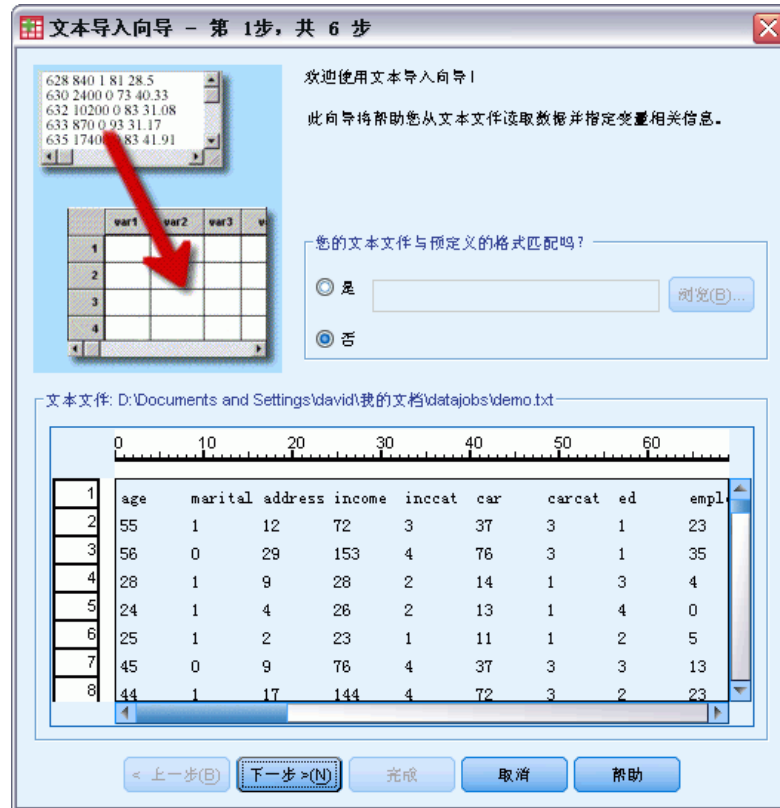
文本文件是另一种常用的数据源。许多电子表格程序和数据库都可采用多种文本文件格式的一种来保存其内容。逗号分隔文件或制表符分隔文件是指使用逗号或制表符来指示每个变量的数据行。在本示例中，数据是用制表符分隔的。

- ▶ 从菜单中选择：
文件 > 读取文本数据...
- ▶ 选择 Text (*.txt) 作为您要查看的文件类型。
- ▶ 打开demo.txt。 [有关详细信息，请参阅第 153 页码附录 A 中的样本文件。](#)

文本导入向导将指导您完成有关如何解释指定文本文件的定义过程。

图片 2-12

文本导入向导：步骤 1，共 6 个步骤

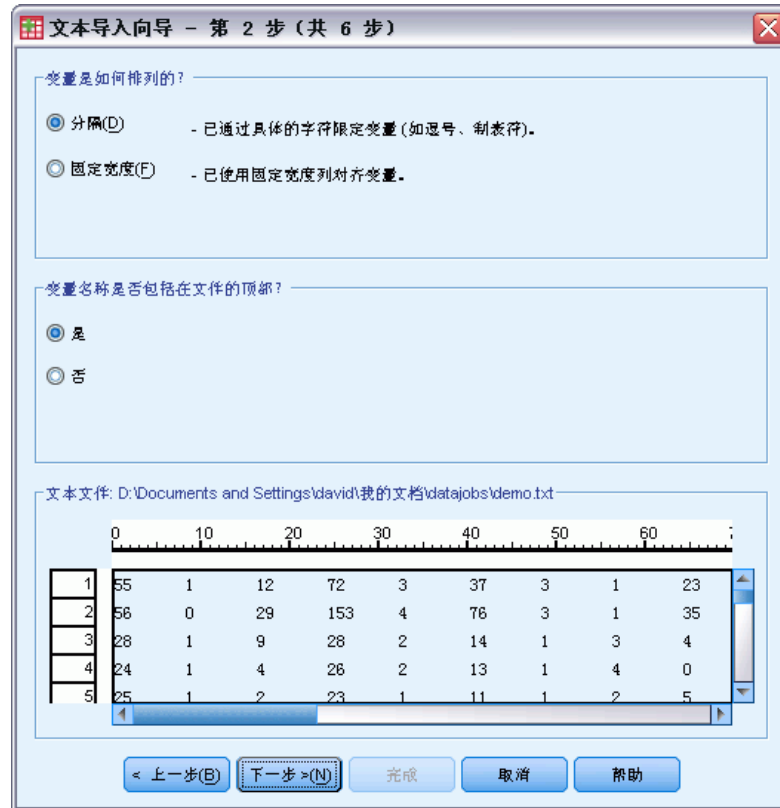


- ▶ 在第 1 步中，可以选择预定义的格式或在向导中创建新格式。选择否指示应该创建新格式。
- ▶ 单击下一步继续。

正如前面所述，此文件使用制表符分隔格式。并且，变量名是在此文件的顶行中定义的。

图片 2-13

文本导入向导：步骤 2，共 6 个步骤



- ▶ 选择分隔指示数据使用分隔格式结构。
- ▶ 选择是指示变量名应该从文件顶部读取。
- ▶ 单击下一步继续。

- ▶ 在下一个对话框的顶部键入 2，指示第一行数据是从文本文件的第二行开始。

图片 2-14

文本导入向导：步骤 3，共 6 个步骤



- ▶ 将此对话框中的其余数据保留为缺省值，然后单击下一步以继续操作。

第 4 步中的“数据预览”可提供一种快速的方式来确保您的数据被正确读取。

图片 2-15

文本导入向导：步骤 4，共 6 个步骤



- ▶ 选择制表符，然后取消选择其他选项。
- ▶ 单击下一步继续。

由于为了符合格式要求，变量名可能已被截断，所以使用此对话框可以编辑任何不合意的名称。

图片 2-16

文本导入向导：步骤 5，共 6 个步骤



还可在此处定义数据类型。例如，假定收入变量要包含一定量的美元金额，这是安全的做法。

要更改数据类型：

- ▶ 在“数据预览”下，选择要更改的变量，此个案中变量为 Income。

- ▶ 从“数据格式”下拉列表中选择美元。

图片 2-17
更改数据类型



- ▶ 单击下一步继续。

图片 2-18
文本导入向导：步骤 6，共 6 个步骤



- ▶ 保留此对话框中的缺省选择，然后单击完成以导入数据。

使用数据编辑器

数据编辑器显示活动数据文件的内容。数据编辑器中的信息由变量和个案组成。

- 在“数据视图”中，列表示变量，行表示个案（观察值）。
- 在“变量视图”中，每一行为一个变量，每一列为与该变量相关的一个属性。

变量用于表示已编译的不同类型的数据。以调查为例：对调查中每个问题的响应相当于一个变量。变量有多种不同的类型，包括数字、字符串、货币和日期。

输入数值数据

可以向数据编辑器中输入数据，这对于小型数据文件或对大型数据文件进行辅助编辑可能很有帮助。

- ▶ 单击“数据编辑器”窗口底部的变量视图选项卡。

您需要定义要使用的变量。在本示例中，只需要三个变量：age、marital status 和 income。

图片 3-1
“变量视图”中的变量名称



- ▶ 在第一列的第一行中，键入 age。
- ▶ 在第二行中，键入 marital。
- ▶ 在第三行中，键入 income。

将自动为新变量指定“数字”数据类型。

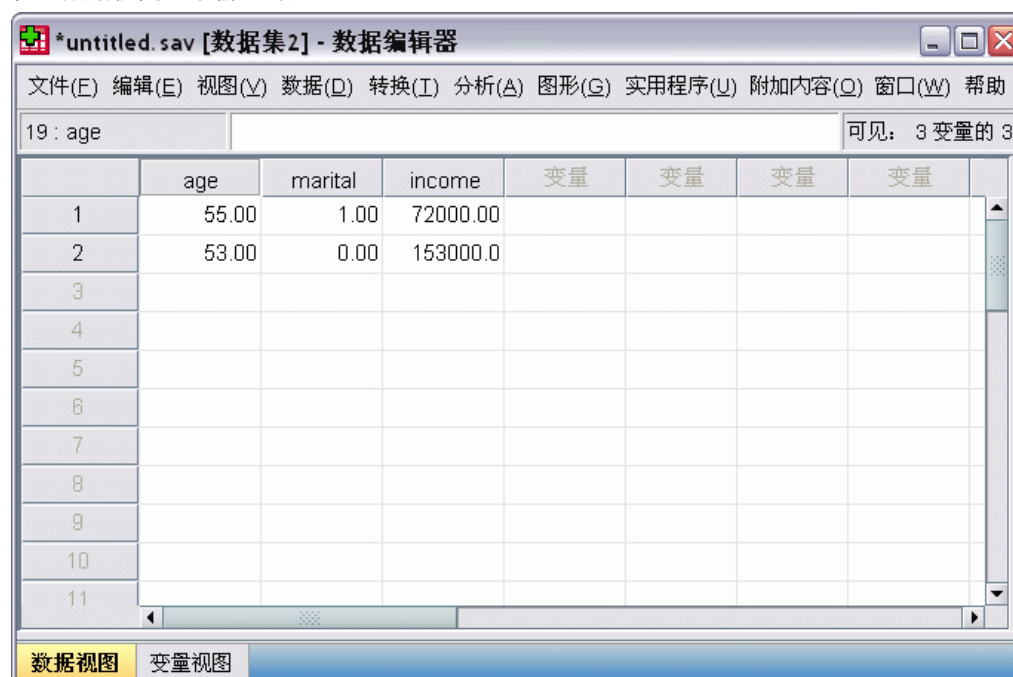
如果未输入变量名称，将自动创建唯一的名称。但是，这些名称不是描述性的，建议不要对大型数据文件使用这些名称。

- ▶ 单击数据视图选项卡继续输入数据。

在“变量视图”中输入的名称当前是“数据视图”中前三列的标题。

从第一列开始，开始在第一行中输入数据。

图片 3-2
在“数据视图”中输入的值



- ▶ 在 age 列中，键入 55。
- ▶ 在 marital 列中，键入 1。
- ▶ 在 income 列中，键入 72000。
- ▶ 将光标移动到第一列的第二行以添加下一个主体的数据。
- ▶ 在 age 列中，键入 53。
- ▶ 在 marital 列中，键入 0。
- ▶ 在 income 列中，键入 153000。

当前，age 和 marital 列显示有小数点，尽管它们的值预期为整数。要隐藏这些变量中的小数点：

- ▶ 单击“数据编辑器”窗口底部的变量视图选项卡。
- ▶ 在 age 一行的小数列中，键入 0 以隐藏小数。

- ▶ 在 marital 一行的小数列中，键入 0 以隐藏小数。

图片 3-3

age 和 marital 更新后的小数属性



The screenshot shows the SPSS Data Editor window titled '*未标题1 [数据集0] - 数据编辑器'. The menu bar includes: 文件(E), 编辑(E), 视图(V), 数据(D), 转换(I), 分析(A), 图形(G), 实用程序(U), 附加内容(O), 窗口(W), 帮助. The main table has the following columns: 名称, 类型, 宽度, 小数, 标签, 值. The data is as follows:

	名称	类型	宽度	小数	标签	值
1	age	数值(N)	8	0		无
2	marital	数值(N)	8	0		无
3	income	数值(N)	8	2		无
4						
5						
6						
7						
8						
9						
10						
11						

At the bottom, there are two tabs: 数据视图 and 变量视图, with 变量视图 selected.

输入字符串数据

也可以向数据编辑器中输入非数值数据（如文本字符串）。

- ▶ 单击“数据编辑器”窗口底部的变量视图选项卡。
- ▶ 在第一个空行的第一个单元格中，键入 sex 作为变量名称。
- ▶ 单击所输入内容后面的类型单元格。

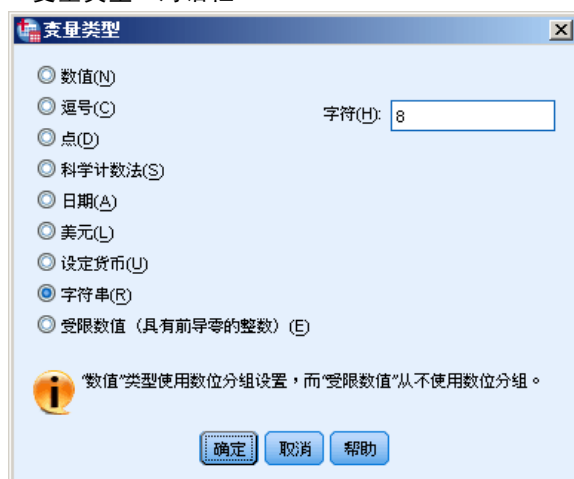
- ▶ 单击类型单元格右侧的按钮打开“变量类型”对话框。

图片 3-4
在 sex 的“类型”单元格中显示的按钮



- ▶ 选择字符串以指定变量类型。
- ▶ 单击确定保存所做的选择并返回到数据编辑器。

图片 3-5
“变量类型”对话框



定义数据

除了定义数据类型之外，还可以为变量名称和数据值定义描述性变量标签和值标签。这些描述性标签用于统计报告和图表。

添加变量标签

标签旨在提供变量的描述。这些描述通常是变量名称的较长版本。标签最多可达 255 个字节。这些标签在输出中用于标识不同的变量。

- ▶ 单击“数据编辑器”窗口底部的变量视图选项卡。
- ▶ 在 age 一行的标签列中，键入 Respondent's Age。
- ▶ 在 marital 一行的标签列中，键入 Marital Status。
- ▶ 在 income 一行的标签列中，键入 Household Income。
- ▶ 在 sex 一行的标签列中，键入 Gender。

图片 3-6
在“变量视图”中输入的变量标签

	名称	类型	宽度	小数	标签	值
1	age	数值(N)	8	0	Respondent's Age	无
2	marital	数值(N)	8	0	Marital Status	无
3	income	数值(N)	8	2	Household Income	无
4	sex	字符串	8	0	Gender	无
5						
6						
7						
8						
9						
10						
11						

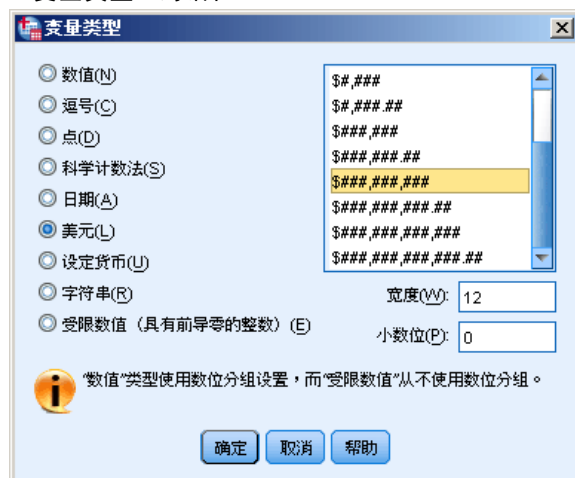
更改变量类型和格式

类型列显示每个变量的当前数据类型。最常见的数据类型为数字和字符串，但也支持大量其他格式。在当前数据文件中，income 变量定义为数字类型。

- ▶ 单击 income 一行的类型单元格，然后单击该单元格右侧的按钮打开“变量类型”对话框。

- ▶ 选择美元。

图片 3-7
“变量类型”对话框



将显示当前所选数据类型的格式选项。

- ▶ 对于本示例中的货币格式，请选择 `$###,###,###`。
- ▶ 单击确定以保存更改。

为数值变量添加值标签

值标签提供了一种将变量值映射为字符串标签的方法。在本示例中，`marital` 变量有两种可接受的值。0 值表示主体为单身，1 值表示主体已婚。

- ▶ 单击 `marital` 一行的值单元格，然后单击该单元格右侧的按钮打开“值标签”对话框。

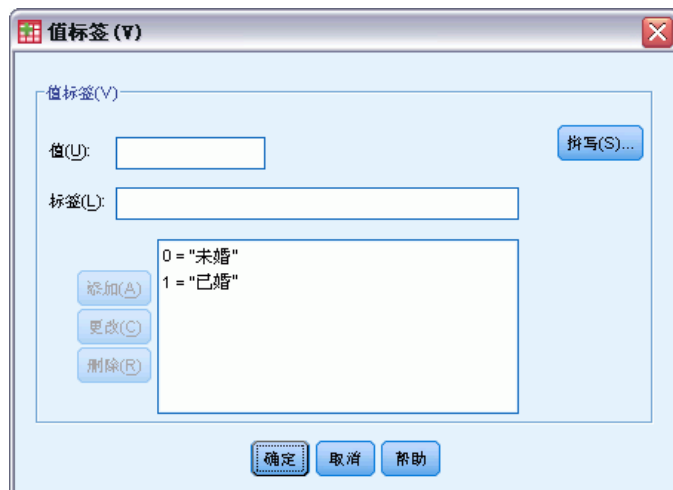
值为实际数值。

值标签是应用于指定数值的字符串标签。

- ▶ 在“值”字段中键入 0。
- ▶ 在“标签”字段中键入 `Single`。

- ▶ 单击添加将此标签添加到列表中。

图片 3-8
“值标签”对话框

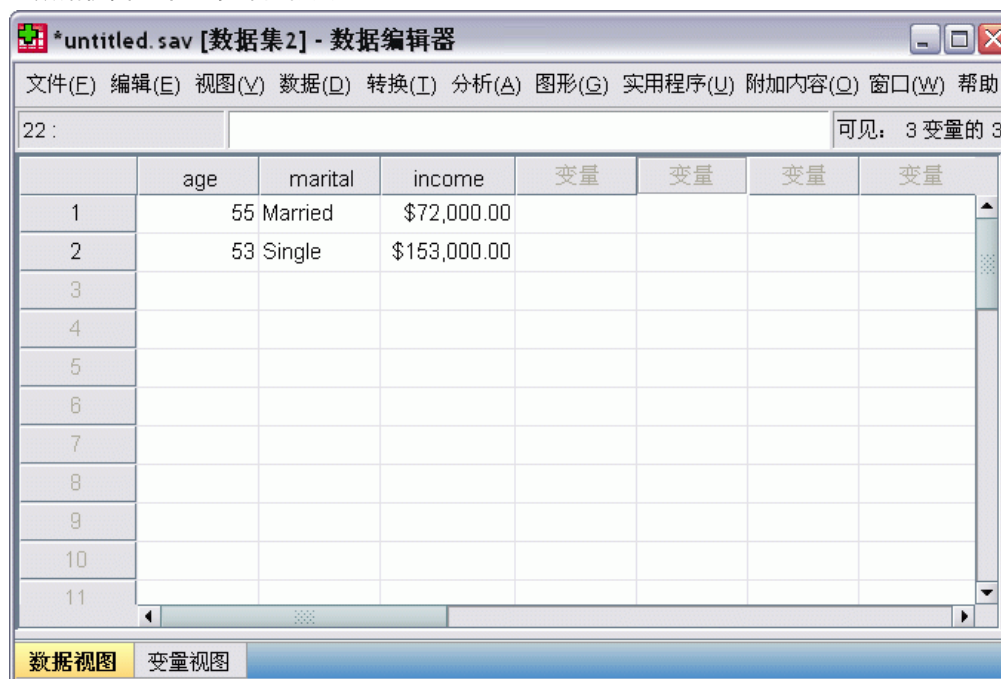


- ▶ 在“值”字段中键入 1，并在“标签”字段中键入 Married。
- ▶ 单击添加，然后单击确定保存更改并返回到数据编辑器。
这些标签还可在“数据视图”中显示，从而提高数据的可读性。
- ▶ 单击“数据编辑器”窗口底部的数据视图选项卡。
- ▶ 从菜单中选择：
视图 > 值标签

现在，在数据编辑器中输入值时会以列表的形式显示值标签。此设置在提供有效响应建议和提供更具描述性的回答方面非常有用。

如果“值标签”菜单项已经处于活动状态（旁边具有复选标记），则再次选择值标签将关闭值标签的显示。

图片 3-9
“数据视图”中显示的值标签



为字符串变量添加值标签

字符串变量可能也需要值标签。例如，您的数据可能使用单个字母（M 或 F）标识主体的性别。这时，可以使用值标签指定 M 代表 Male，F 代表 Female。

- ▶ 单击“数据编辑器”窗口底部的变量视图选项卡。
- ▶ 单击 sex 一行的值单元格，然后单击该单元格右侧的按钮打开“值标签”对话框。
- ▶ 在“值”字段中键入 F，在“标签”字段中键入 Female。

- ▶ 单击添加将此标签添加到数据文件中。

图片 3-10
“值标签”对话框



- ▶ 在“值”字段中键入 M，在“标签”字段中键入 Male。
- ▶ 单击添加，然后单击确定保存更改并返回到数据编辑器。
因为字符串区分大小写，所以应保持一致。小写字母 m 与大写字母 M 是不相同的。

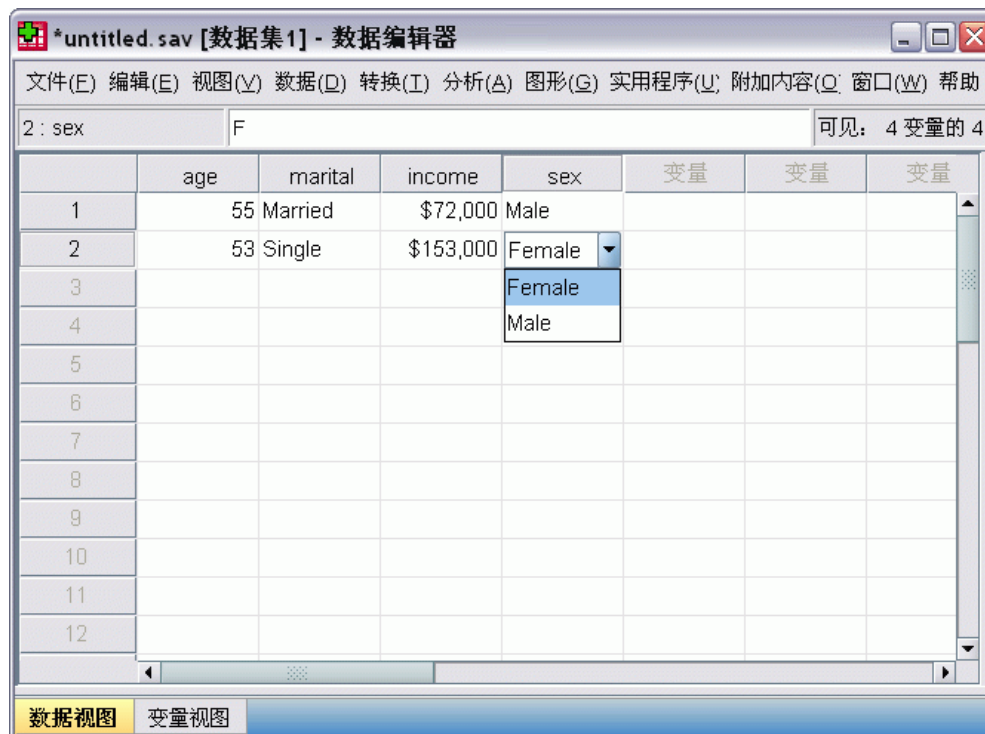
使用值标签进行数据输入

可以使用值标签进行数据输入。

- ▶ 单击“数据编辑器”窗口底部的数据视图选项卡。
- ▶ 在第一行中，选择 sex 的单元格。
- ▶ 单击该单元格右侧的按钮，然后从下拉列表中选择 Male。
- ▶ 在第二行中，选择 sex 的单元格。

- ▶ 单击该单元格右侧的按钮，然后从下拉列表中选择 Female。

图片 3-11
使用变量标签选择值



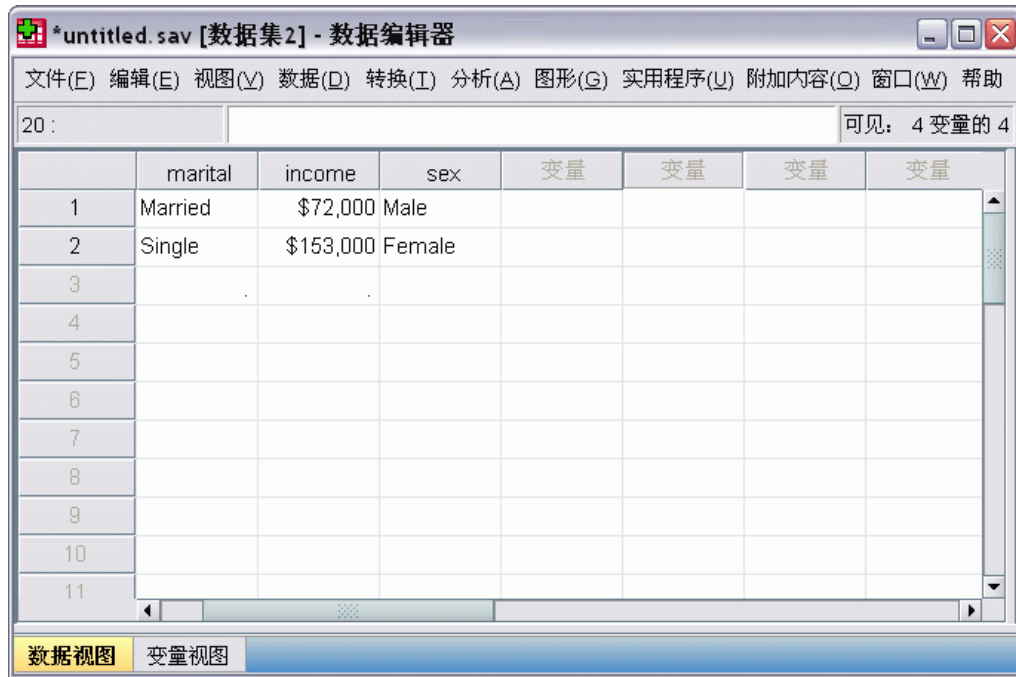
仅列出定义的值，从而确保输入的数据符合所需的格式。

处理缺失数据

缺失数据或无效数据通常很容易被忽略。调查响应者可能拒绝回答某些问题，可能不知道答案，或者可能以意外的格式回答。如果不过滤或标识这些数据，则所做分析可能无法提供准确结果。

对于数值数据，空数据字段或包含无效输入的字段将转换为系统缺失值，系统缺失值可使用单个句点来标识。

图片 3-12
显示为句点的缺失值

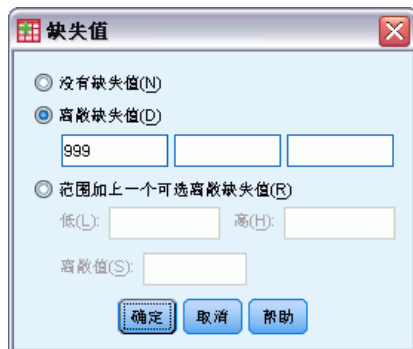


值缺失的原因可能对您的分析非常重要。例如，您可能会发现区分拒绝回答问题的响应者与由于不用而未回答问题的响应者很有帮助。

数值变量的缺失值

- ▶ 单击“数据编辑器”窗口底部的变量视图选项卡。
- ▶ 单击 age 一行的缺失单元格，然后单击该单元格右侧的按钮打开“缺失值”对话框。在此对话框中，可以指定多达三个不同的缺失值，也可以指定值范围加上一个附加的离散值。

图片 3-13
“缺失值”对话框



- ▶ 选择离散缺失值。

- ▶ 在第一个文本框中键入 999，并将另外两个文本框保留为空。
- ▶ 单击确定保存更改并返回到数据编辑器。

现在已添加了缺失数据值，接下来可对该值应用标签。

- ▶ 单击 age 一行的值单元格，然后单击该单元格右侧的按钮打开“值标签”对话框。
- ▶ 在“值”字段中键入 999。
- ▶ 在“标签”字段中键入 No Response。

图片 3-14
“值标签”对话框



- ▶ 单击添加将此标签添加到数据文件中。
- ▶ 单击确定保存更改并返回到数据编辑器。

字符串变量的缺失值

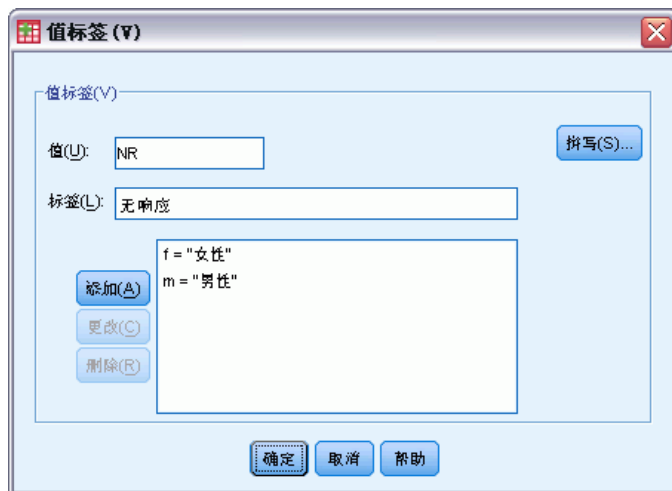
字符串变量的缺失值与数值变量的缺失值的处理方法类似。不过，与数值变量不同，字符串变量中的空字段未指定为系统缺失。相反，它们被解释为空字符串。

- ▶ 单击“数据编辑器”窗口底部的变量视图选项卡。
- ▶ 单击 sex 一行的缺失单元格，然后单击该单元格右侧的按钮打开“缺失值”对话框。
- ▶ 选择离散缺失值。
- ▶ 在第一个文本框中键入 NR。
字符串变量的缺失值区分大小写。因此，值 nr 不会被视为缺失值。
- ▶ 单击确定保存更改并返回到数据编辑器。

现在可以为缺失值添加标签。

- ▶ 单击 sex 一行的值单元格，然后单击该单元格右侧的按钮打开“值标签”对话框。
- ▶ 在“值”字段中键入 NR。
- ▶ 在“标签”字段中键入 No Response。

图片 3-15
“值标签”对话框



- ▶ 单击添加将此标签添加到您的项目中。
- ▶ 单击确定保存更改并返回到数据编辑器。

复制并粘贴变量属性

在定义了变量的变量属性之后，可以复制这些属性并将其应用到其他变量。

- ▶ 在“变量视图”中，在第一个空行的第一个单元格中键入 `agewed`。

图片 3-16

“变量视图”中的 `agewed` 变量

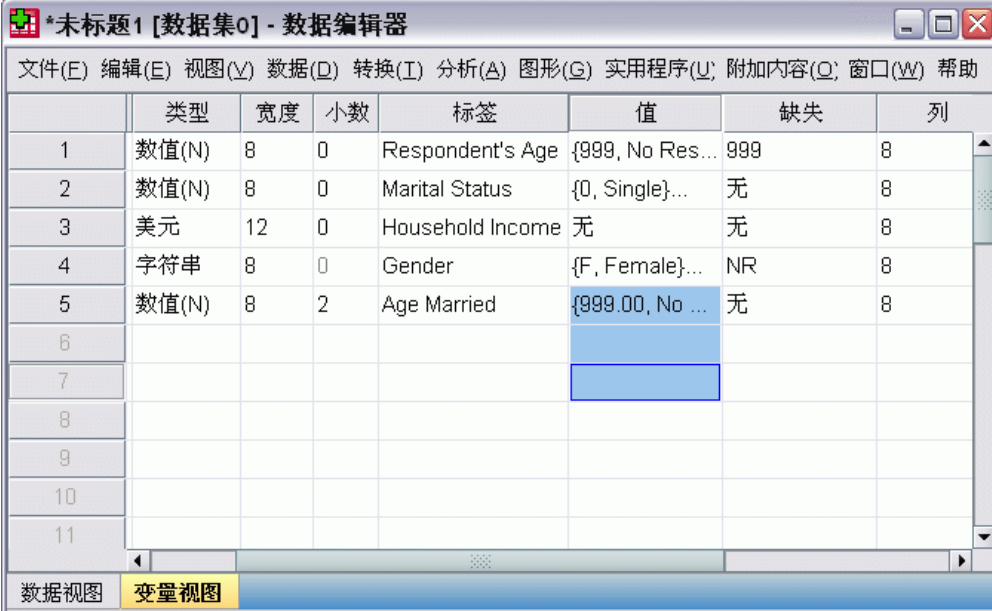
	名称	类型	宽度	小数	标签
1	age	数字	8	0	Respondent's A {999,
2	marital	数字	8	0	Marital Status {0, S
3	income	美元	12	0	Household Inco 无
4	sex	字符串	8	0	Gender {F, F
5	agewed	数字	9	2	Age Married 无
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					

- ▶ 在标签列中，键入 `Age Married`。
- ▶ 单击 `age` 一行的值单元格。
- ▶ 从菜单中选择：
编辑 > 粘贴
- ▶ 单击 `agewed` 一行的值单元格。
- ▶ 从菜单中选择：
编辑 > 粘贴

现在，在 `age` 变量中定义的值将应用到 `agewed` 变量。

要将属性应用到多个变量，只需选择多个目标单元格（单击并沿着列向下拖动）即可。

图片 3-17
选中的多个单元格



*未标题1 [数据集0] - 数据编辑器

文件(E) 编辑(E) 视图(V) 数据(D) 转换(I) 分析(A) 图形(G) 实用程序(U) 附加内容(O) 窗口(W) 帮助

	类型	宽度	小数	标签	值	缺失	列
1	数值(N)	8	0	Respondent's Age	{999, No Res...	999	8
2	数值(N)	8	0	Marital Status	{0, Single}...	无	8
3	美元	12	0	Household Income	无	无	8
4	字符串	8	0	Gender	{F, Female}...	NR	8
5	数值(N)	8	2	Age Married	{999.00, No ...	无	8
6							
7							
8							
9							
10							
11							

数据视图 变量视图

粘贴属性时，该属性将应用到所有选中的单元格。

如果将值粘贴到空行，则将自动创建新变量。

要将所有属性从一个变量复制到另一个变量：

- ▶ 单击 marital 行的行号。

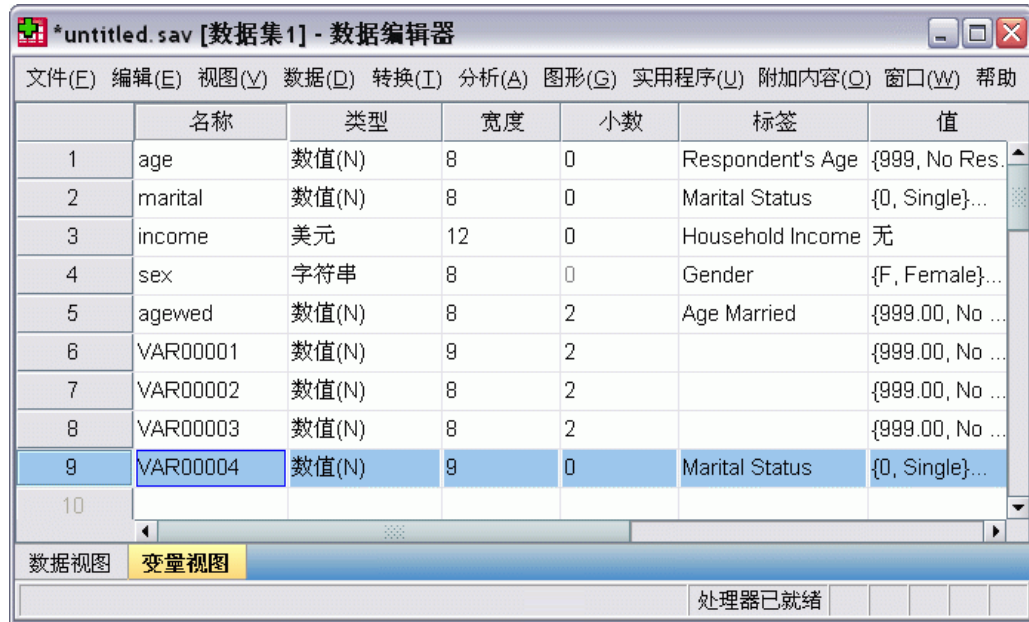
图片 3-18
选中的行



- ▶ 从菜单中选择：
编辑 > 粘贴
- ▶ 单击第一个空行的行号。
- ▶ 从菜单中选择：
编辑 > 粘贴

marital 变量的所有属性都应用到新变量。

图片 3-19
粘贴到一行的所有值



定义分类变量的变量属性

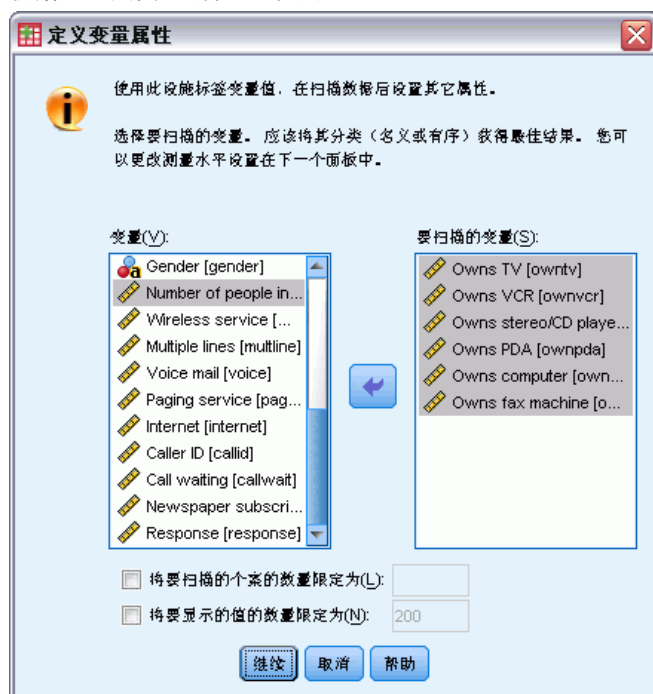
对于分类（名义、有序）数据，可以使用“定义变量属性”来定义值标签和其他变量属性。“定义变量属性”过程：

- 扫描实际数据值并为每个选定的变量列出所有的唯一数据值。
- 识别未标注的值并提供“自动标注”功能。
- 提供将被定义的值标签从另一变量复制到选定变量，或是从选定变量复制到其他变量的功能。

本示例使用数据文件demo.sav。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。该数据文件已定义了值标签，因此我们将为没有定义值标签的变量输入值。

- ▶ 在数据编辑器的“数据视图”中，单击变量 ownpc 的第一个数据单元格（可能必须向右滚动），然后输入 99。
- ▶ 从菜单中选择：
数据 > 定义变量属性...

图片 3-20
初始“定义变量属性”对话框



在初始“定义变量属性”对话框中，选择要为其定义值标签和/或其他属性的名义变量或有序变量。

- ▶ 将拥有电视 [owntv] 到拥有传真机 [ownfax] 拖动到“扫描变量”列表。

您可能会注意到，所有选定变量的测量级别图标均表明它们是刻度变量而非分类变量。本示例中选定的所有变量实际上是使用数值 0 和 1 分别代表 No 和 Yes 的分类变量；我们将使用“定义变量属性”更改的变量属性之一就是“测量级别”。

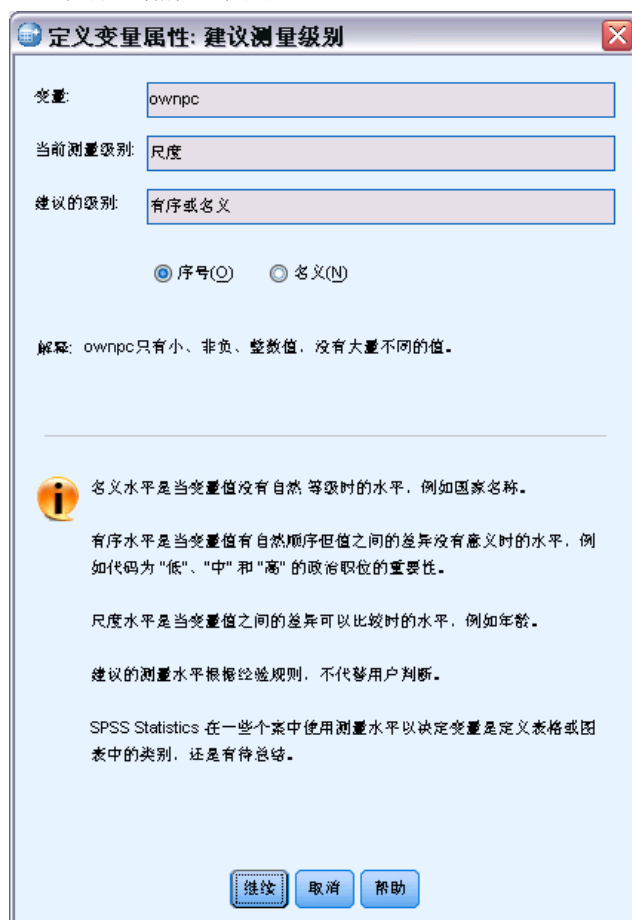
- ▶ 单击继续。

图片 3-21
“定义变量属性”主对话框



- ▶ 在“已扫描的变量列表”中，选择 ownnpc。
选定变量的当前测量级别为“刻度”。可以通过从下拉列表选择一个级别来更改测量级别，也可以让“定义变量属性”建议测量级别。
- ▶ 单击建议。
将显示“建议测量级别”对话框。

图片 3-22
“建议测量级别”对话框



由于该变量没有太多不同的值，并且所有已扫描的个案均包含整数值，因此，正确的测量级别可能为“有序”或“名义”。

- ▶ 选择有序，然后单击继续。

现在，选定变量的测量级别为“有序”。

“值标签”网格显示了选定变量的所有唯一数据值、为这些值定义的所有值标签以及每个值在扫描的个案中出现的次数（计数）。

我们在“数据视图”中输入的值 99 显示在网格中。计数仅为 1，因为我们仅更改了一个个案的值；由于我们尚未定义 99 的值标签，标签列为空。“已扫描的变量列表”的第一列中的 x 也指示选定变量至少具有一个观察值，没有定义值标签。

- ▶ 在值 99 的标签列中，输入 No answer。
- ▶ 选中值 99 的缺失列中的复选框，以将值 99 标识为用户缺失值。将指定为用户缺失值的数据值标记为进行特殊处理，并将其从大多数计算中排除。

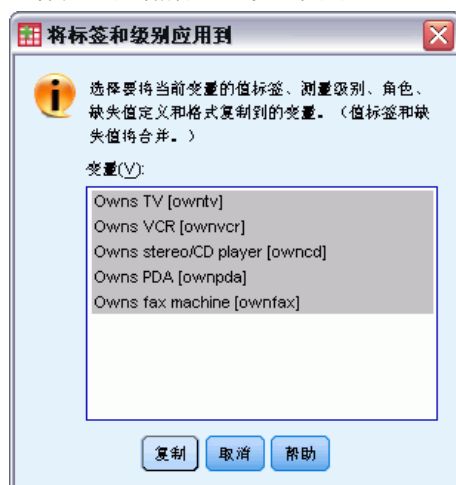
图片 3-23
为 ownpc 定义的新变量属性



在完成修改 ownpc 的变量属性工作之前，让我们将相同的测量级别、值标签和缺失值定义应用到列表中的其他变量。

- ▶ 在“复制属性”区域中，单击到其他变量。

图片 3-24
“将标签和级别应用到”对话框



- 在“将标签和级别应用到”对话框中，选择列表中的所有变量，然后单击复制。

如果当前选择了“定义变量属性”主对话框的“已扫描的变量列表”中的任何其他变量，您将看到这些变量均为有序变量，并具有值 99（定义为用户缺失值）和 No answer 值标签。

图片 3-25
为 ownfax 定义的新变量属性



- ▶ 单击**确定**保存已定义的所有变量属性。

使用多数据源

从版本 14.0 开始，可以同时打开多个数据源，从而使得以下操作更加轻松：

- 在数据源之间来回切换。
- 比较不同数据源的内容。
- 在数据源之间复制和粘贴。
- 创建个案和/或变量的多个子集以进行分析。
- 在不事先保存每个数据源的情况下，合并来自各种数据格式（例如：电子表格、数据库、文本数据）的多个数据源。

多数据源的基本处理

图片 4-1
同时打开两个数据源



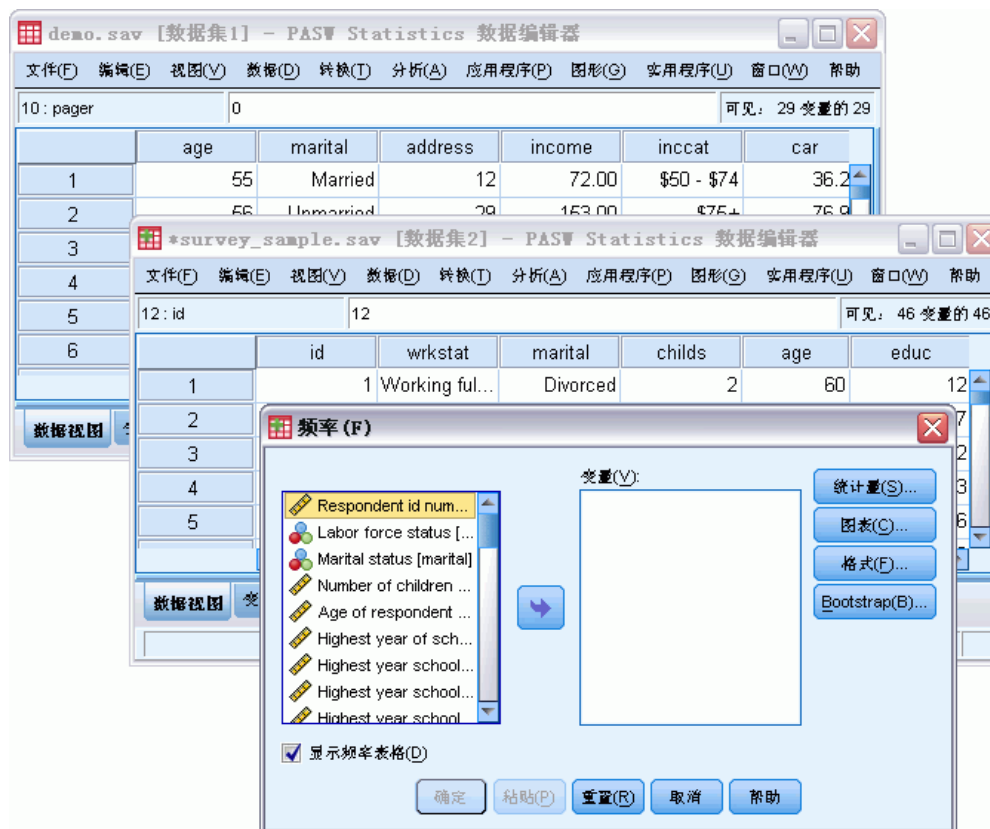
缺省情况下，所打开的每个数据源都显示在新的数据编辑器窗口中。

- 以前打开的任何数据源都保持打开状态，并可供未来使用。
- 第一次打开数据源时，它自动成为**活动数据集**。
- 只需通过单击要使用的数据源的数据编辑器窗口中任意位置，或从“窗口”菜单中选择该数据源的数据编辑器窗口，就可以更改活动数据集。

- 只有活动数据集中的变量才能用于分析。

图片 4-2

包含活动数据集中的变量的变量列表



- 访问数据的任何对话框打开（包括显示变量列表的所有对话框）时，不能更改活动数据集。
- 在会话过程中，必须至少打开一个数据编辑器窗口。关闭最后一个打开的数据编辑器窗口时，IBM® SPSS® Statistics 将自动关闭，提示首先保存更改。

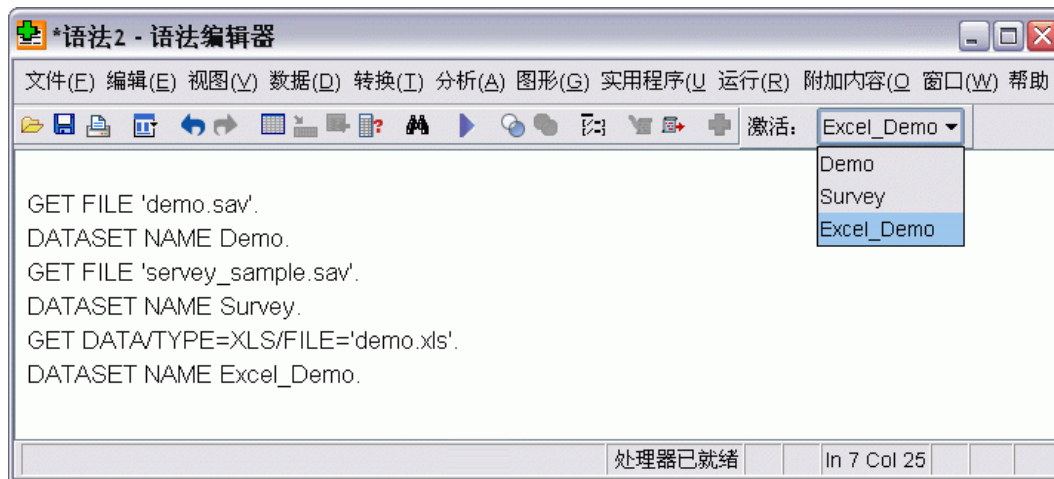
使用命令语法中的多个数据集

如果使用打开数据源的命令语法（例如 GET FILE、GET DATA），则需要使用 DATASET NAME 命令显式地命名每个数据集，以便同时打开多个数据源。

在使用命令语法时，活动数据集名称显示在语法窗口的工具栏上。所有以下操作均可更改活动数据集：

- 使用 DATASET ACTIVATE 命令。
- 单击数据集的“数据编辑器”窗口中的任意位置。
- 从语法窗口中选择数据集名。

图片 4-3
打开在语法窗口工具栏上显示的数据集



在数据集之间复制和粘贴信息

可以使用与在单个数据文件中复制和粘贴信息基本相同的方法，将数据和变量定义属性从一个数据集复制到另一数据集。

- 在“数据视图”中复制和粘贴所选数据单元将仅粘贴数据值，而不包括变量定义属性。
- 通过选择在列顶部的变量名称来复制和粘贴“数据视图”中的整个变量，会粘贴该变量的所有数据和所有数据定义属性。
- 在“变量视图”中复制和粘贴变量定义属性或整个变量将粘贴所选属性（或整个变量定义），但不粘贴任何数据值。

重命名数据集

通过菜单和对话框打开数据源时，向每个数据源都自动分配 DataSetn 的数据集名，其中 n 是序号整数值；而使用命令语法打开数据源时，不会分配数据集名，除非用 DATASET NAME 显式地指定名称。提供更具描述性的数据集名：

- ▶ 从要更改名称的数据集的“数据编辑器”窗口中的菜单，选择：
文件 > 重命名数据集...
- ▶ 输入符合变量命名规则的新数据集名。

不显示多个数据集

如果您希望一次只提供一个数据集，并取消多个数据集功能：

- ▶ 从菜单中选择：
编辑 > 选项...

- ▶ 单击常规选项卡。

选择（选中）一次只能打开一个数据集。

检查单个变量的摘要统计量

本章介绍简单的摘要测量以及变量的测量级别如何影响应使用的统计量类型。我们将使用数据文件 demo.sav。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。

度量水平

不同的摘要测量适用于不同类型的数据，这具体取决于测量级别：

分类。包含有限数量的不同值或类别（例如，性别或婚姻状况）的数据。也称为**定性数据**。分类变量可以为字符串（字母数字）数据，也可以为使用数字代码来表示类别的数值变量（例如，0 = Unmarried, 1 = Married）。存在两种基本类型的分类数据：

- **名义。**不具有固有的类别顺序的分类数据。例如，工作类别 sales 不高于也不低于工作类别 marketing 或 research。
- **有序。**具有有意义的类别顺序但各类别之间没有可测量距离的分类数据。例如，值 high、medium 和 low 具有顺序，但无法计算这些值之间的“距离”。

标度。以区间或比例刻度度量的数据，其中数据值既表示值的顺序，也表示值之间的距离。例如，72,195 美元的薪金比 52,398 美元的薪金高，这两个值之间的距离是 19,797 美元。也称为**定量或连续数据**。

分类数据的摘要测量

对于分类数据，最典型的摘要测量是每个类别中的个案的数目或百分比。**众数**是具有最大个案数目的类别。对于有序数据，如果存在大量的类别，则**中位数**（半数个案位于其上下的值）可能也是一个有用的摘要测量。

“频率”过程将生成频率表，其中显示变量的每个观察值的个案数目和百分比。

- ▶ 从菜单中选择：
分析 > 描述统计 > 频率...

注意：此功能需要 Statistics Base 选项。

- ▶ 选择 Owns PDA [ownpda] 和 Owns TV [owntv] 并将其移动到“变量”列表中。

图片 5-1
要进行分析的所选分类变量



- ▶ 单击确定以运行该过程。

图片 5-2
频率表

频率表

Owns PDA

		频率	百分比	有效百分比	累积百分比
有效	No	5093	79.6	79.6	79.6
	Yes	1307	20.4	20.4	100.0
	合计	6400	100.0	100.0	

Owns TV

		频率	百分比	有效百分比	累积百分比
有效	No	63	1.0	1.0	1.0
	Yes	6337	99.0	99.0	100.0
	合计	6400	100.0	100.0	

频率表显示在浏览器窗口中。频率表揭示仅有 20.4% 的人拥有个人数字助理，但几乎每个人都拥有电视（99.0%）。这些可能不是有趣的发现，尽管发现一小部分人没有电视可能十分有趣。

分类数据的图表

您可以用条形图或饼图来图形化显示频率表中的信息。

- ▶ 再次打开“频率”对话框。（两个变量应仍处于选中状态。）

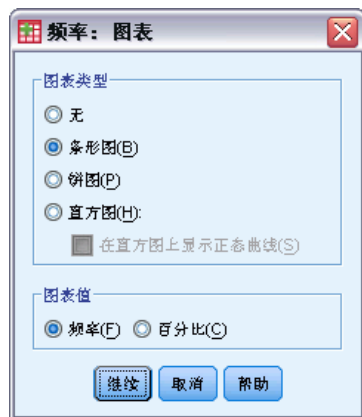
可以使用工具栏上的“对话框调用”按钮快速返回到最近用过的过程。

图片 5-3
“对话框调用”按钮



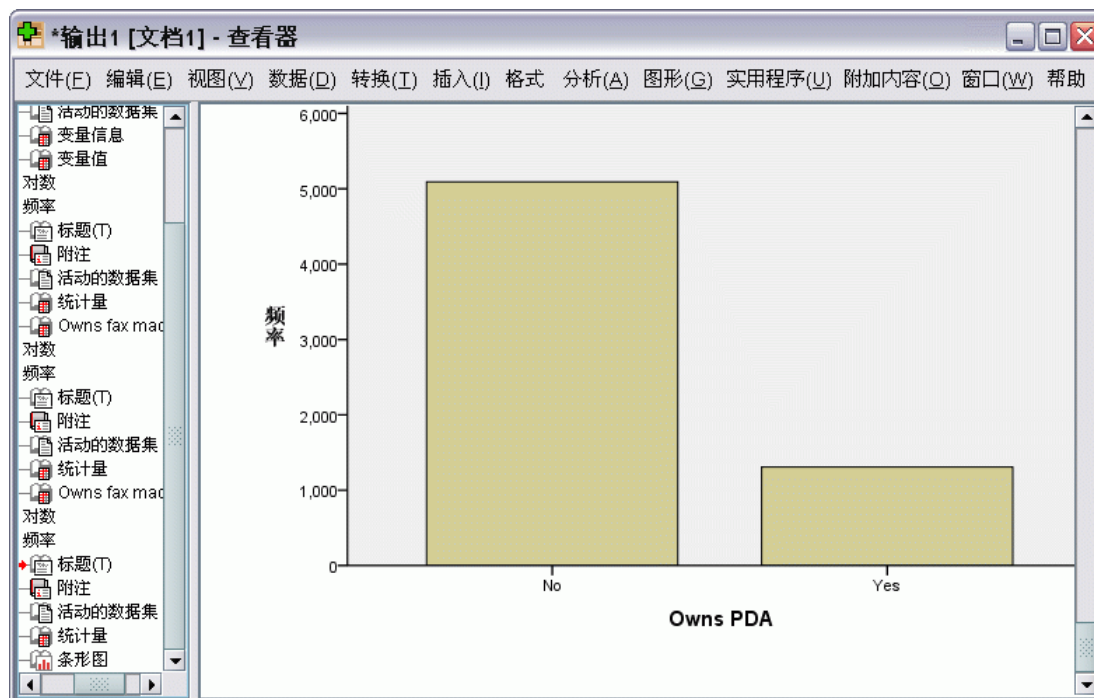
- ▶ 单击图表。
- ▶ 选择条形图，然后单击继续。

图片 5-4
“频率： 图表”对话框



- ▶ 单击主对话框中的确定以运行该过程。

图片 5-5
条形图



除了频率表之外，当前还以条形图的形式显示了相同信息，以便于轻松看出大多数人没有个人数字助理，但几乎每个人都拥有电视这一事实。

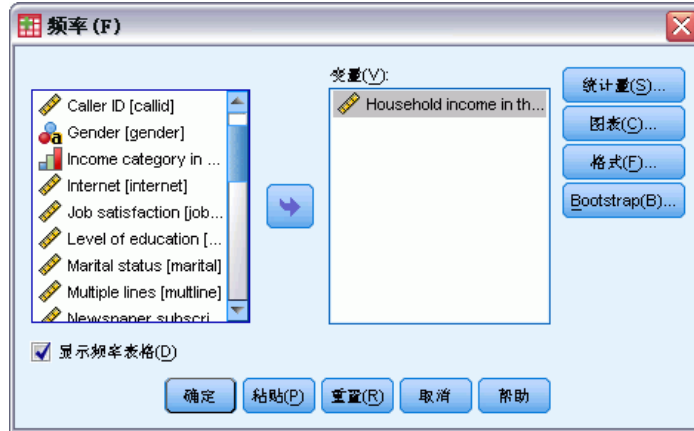
刻度变量的摘要测量

有许多可用于刻度变量的摘要测量，其中包括：

- **集中趋势的测量。**最常见的集中趋势测量是**均值**（算术平均数）和**中位数**（半数个案位于其上下的值）。
 - **离散测量。**测量数据中变异量或分散量的统计量，包括标准差、最小值和最大值。
- ▶ 再次打开“频率”对话框。
- ▶ 单击重置清除以前的所有设置。

- ▶ 选择 Household income in thousands [income] 并将其移到“变量”列表中。

图片 5-6
要进行分析的所选刻度变量



- ▶ 单击统计量。
- ▶ 选择均值、中位数、标准差、最小值和最大值。

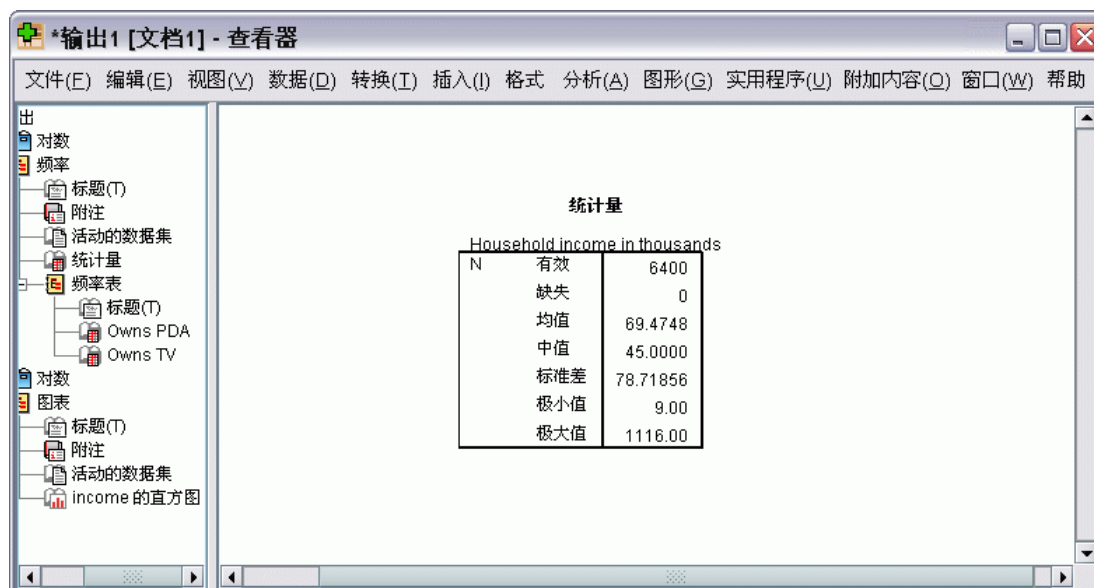
图片 5-7
“频率：统计量”对话框



- ▶ 单击继续。
- ▶ 取消选中主对话框中的显示频率表格。（频率表通常对刻度变量没有帮助，因为不同值的数目可能与数据文件中的个案数几乎一样多。
- ▶ 单击确定以运行该过程。

“频率统计量”表将显示在浏览器窗口中。

图片 5-8
“频率统计量”表

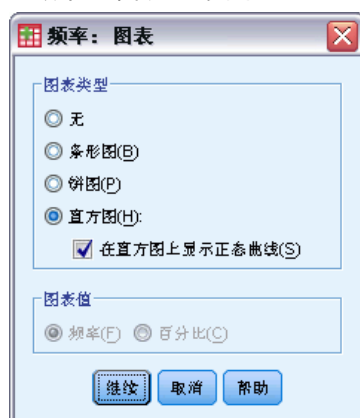


在此示例中，均值和中值之间具有较大的差异。均值几乎比中值大 25,000，这表明值为非正态分布。可以使用直方图来直观检查分布。

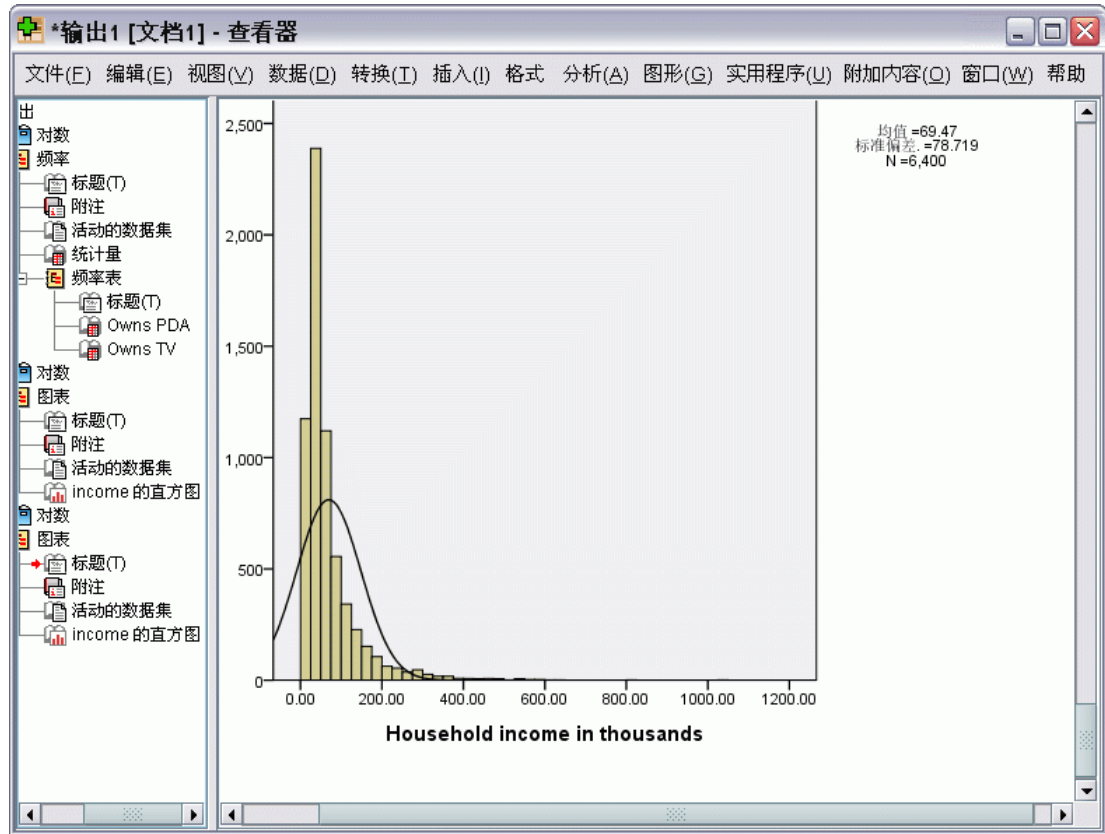
刻度变量的直方图

- ▶ 再次打开“频率”对话框。
- ▶ 单击图表。
- ▶ 选择直方图和带正态曲线。

图片 5-9
“频率：图表”对话框



- ▶ 单击继续，然后单击主对话框中的确定以运行该过程。

图片 5-10
直方图

大多数个案聚类在刻度的较低端，其中多数在 100,000 以下。但是，也有少数个案在 500,000 及以上（极少数个案甚至需要修改直方图才能可见）。这些少数个案的较高值对均值具有显著作用，但对中位数没有或仅有较小作用，因而在此示例中，中位数是对集中趋势的较好指示。

创建和编辑图表

您可以创建各种图表类型并对其进行编辑。在本章中，我们将创建和编辑条形图。您可以将以下原则应用于任何图表类型。

图表创建的基础知识

为了演示图表创建的基础知识，我们将创建平均收入条形图，用来显示不同级别的工作满意度。本示例使用数据文件demo.sav。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。

- ▶ 从菜单中选择：
图形 > 图表构建程序...

“图表生成器”对话框是一个交互式窗口，可在构建图表时用来预览其外观。

图片 6-1
“图表生成器”对话框



使用图表生成器库

- ▶ 如果尚未选择库选项卡，请单击它。

库中包括许多不同的预定义图表，这些图表都按图表类型组织在一起。“基本元素”选项卡也提供从头开始创建图表的基本元素（例如轴和图形元素），但使用库更简便。

- ▶ 如果尚未选择条形图，请单击该图表。

代表库中的可用条形图的图标随即出现在该对话框中。图形应该提供足够的信息，以便标识具体的图表类型。如果需要更多信息，也可以将光标悬停在图标上，这样可以显示该图表的工具提示说明。

- ▶ 将简单条形图的图标拖放到“画布”上，画布是库上方的一个大型区域。“图表生成器”会在画布上显示该图表的预览。请注意，用于绘制该图表的数据并非实际数据，而是示例数据。

图片 6-2
图表生成器画布上的条形图



定义变量和统计量

尽管画布上有图表，但图表并不完整，因为没有变量或统计量来控制条的高度以及指定每个条对应的变量类别。图表不能没有变量和统计量。从“变量”列表中拖出变量以添加它们，该列表位于画布左侧。

在图表生成器中，变量的测量级别至关重要。您打算在 x 轴上使用 Job satisfaction 变量。不过，该变量旁边的图标（外观像一把尺子）指示其测量级别定义为刻度。要创建正确的图表，必须使用分类测量级别。为此，可以暂时在图表生成器中更改该测量级别，而不用回到变量视图中更改该测量级别。

- ▶ 右键单击变量列表中的 Job satisfaction，并选择有序。有序是适当的测量级别，因为 Job satisfaction 中的分类可以按满意水平排序。请注意，更改了测量级别后，图标也会相应地更改。
- ▶ 现在，将 Job satisfaction 从变量列表拖到 x 轴放置区。

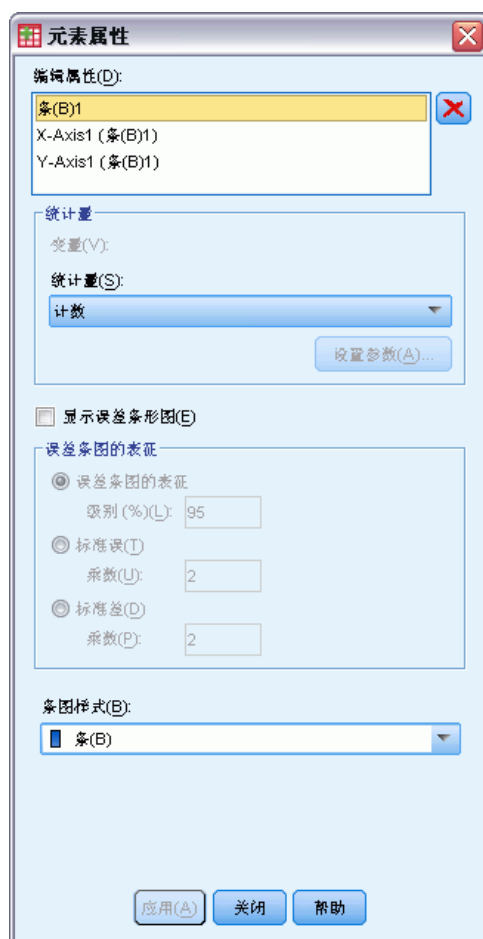
图片 6-3
Job satisfaction 在 x 轴放置区中



y 轴放置区默认为计数统计量。如果要使用另一种统计量（例如百分比或平均数），可以轻松更改它。在本示例中，将不使用这些统计量，但我们将复习这一过程，以防在其他时候需要更改这一统计量。

- ▶ 单击元素属性以显示“元素属性”窗口。

图片 6-4
“元素属性”窗口



“元素属性”窗口可用于更改各种图表元素的属性。这些元素包括图形元素（例如条形图中的条）和图表上的轴。在“编辑属性”列表中，选择一个元素，更改与该元素相关的属性。还请注意该列表右侧的红色 X。该按钮可用于删除画布上的图形元素。由于条形图 1 已选中，因此所显示的属性将应用于图形元素，在此处指条形图元素。

“统计量”下拉列表显示可用的具体统计量。对于每一种图表类型，可用的统计量通常都相同。请注意，某些统计量要求 y 轴放置区包含变量。

- ▶ 返回到“图表生成器”对话框，并将 Household income in thousands 从“变量”列表中拖到 y 轴放置区。由于 y 轴上的变量为刻度，而 x 轴变量为分类（有序是一种分类测量级别），所以 y 轴放置区默认为均值统计量。这是您希望的变量和统计量，因此，不需要更改元素属性。

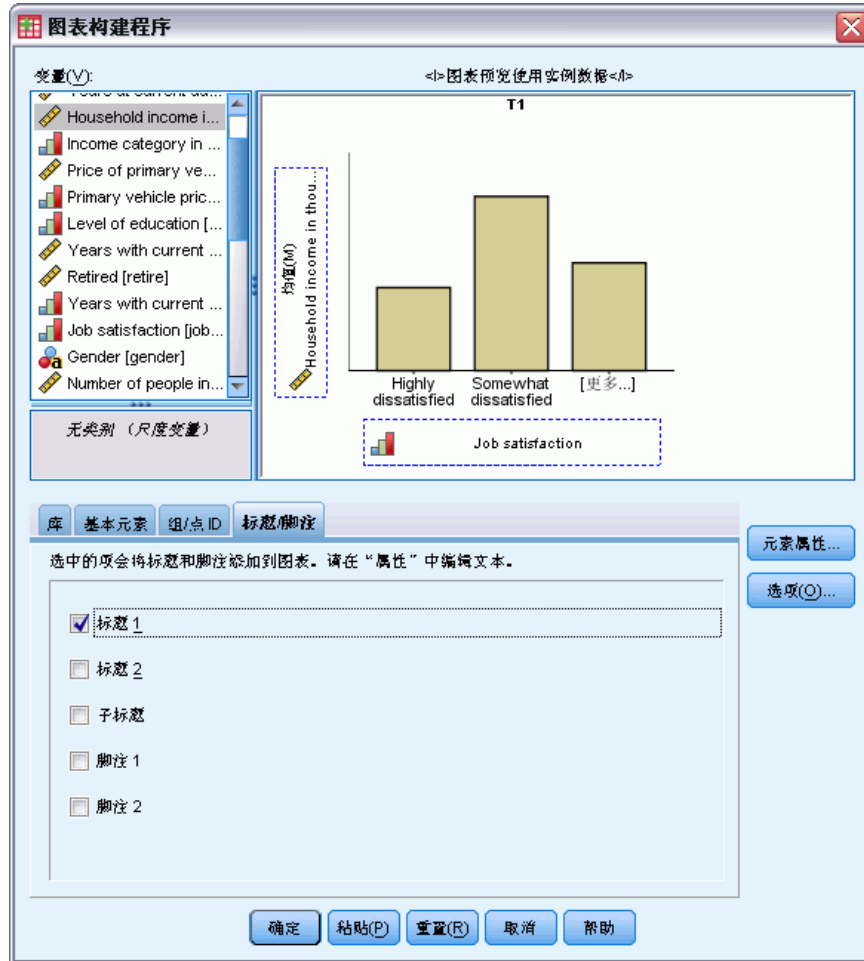
添加文字

您还可以给图表添加标题和脚注。

- ▶ 单击标题/脚注选项卡。

- ▶ 选择标题 1。

图片 6-5
标题 1 显示在画布上



该标题出现在画布上，并带有标签 T1。

- ▶ 在“元素属性”窗口中，选择“编辑属性”列表中的标题 1。
- ▶ 在“内容”文本框中，键入 Income by Job Satisfaction。这是该标题将显示的文字。
- ▶ 单击应用，保存这些文字。尽管这些文字不会显示在图表生成器中，但当您生成该图表时会出现这些文字。

创建图表

- ▶ 单击确定，创建条形图。

图片 6-6
条形图



从条形图中可以看出，被调查者对其工作越满意，其家庭收入也越高。

图表编辑基础知识

编辑图表的方法有多种。以您刚创建的条形图为例，您将进行以下操作：

- 更改颜色。
- 设置刻度标记标签中数字的格式。
- 编辑文本。
- 显示数据值标签。
- 使用图表模板。

要编辑图表，请在图表编辑器中将它打开。

- ▶ 双击该条形图，在图表编辑器中将它打开。

图片 6-7

图表编辑器中的条形图



选择图表元素

要编辑图表元素，首先要选中它。

- ▶ 单击任何一个条。条周围的矩形表明条已选中。

在简单图表中选中元素的一般规则如下：

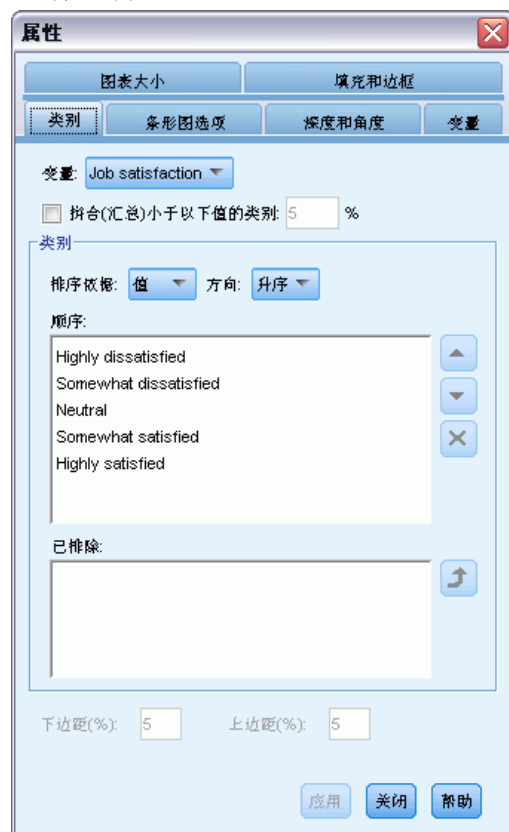
- 如果未选择任何图形元素，则单击任意图形元素可选择所有图形元素。
 - 如果所有图形元素都已选中，则单击某一图形元素可仅选择该图形元素。您可以单击另一个图形元素来选中它。要选中多个图形元素，请按住 Ctrl 键，同时单击每个要选中的元素。
- ▶ 要取消选中所有元素，请按 Esc 键。
 - ▶ 单击任何条，可再次选中所有条。

使用“属性”窗口

- ▶ 从“图表编辑器”菜单中选择：
编辑 > 属性

此操作会打开“属性”窗口，其中显示适用于所选条的选项卡。这些选项卡会因图表编辑器中选择的图表元素而异。例如，如果选择的是文本边框而不是条，则“属性”窗口中会出现不同的选项卡。在这些选项卡上，可执行大多数图表编辑操作。

图片 6-8
“属性”窗口



更改条颜色

首先，可更改条的颜色。在“填充和边框”选项卡上指定图形元素（线条和标记除外）的颜色属性。

- ▶ 单击填充和边框选项卡。
- ▶ 单击“填充”旁边指示要更改条的填充颜色的图样。图样下面的数字指定当前颜色的红色、绿色和蓝色设置。
- ▶ 单击浅蓝色，即倒数第二行左边第二个。

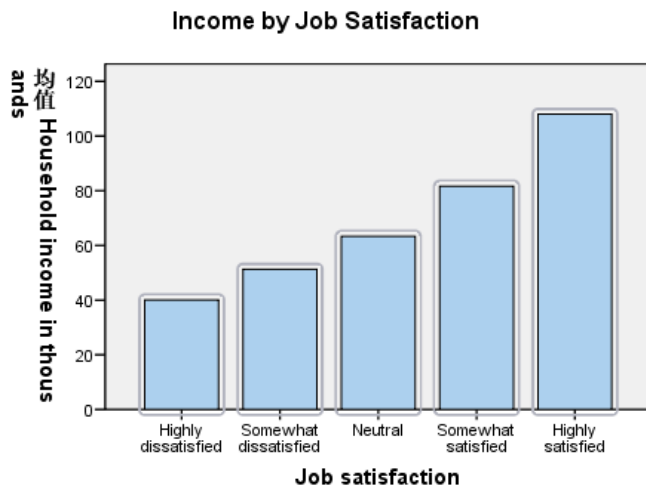
图片 6-9
“填充和边框”选项卡



- ▶ 单击应用。

图表中的条现在是浅蓝色。

图片 6-10
显示蓝色条的已编辑条形图



设置刻度标记标签中数字的格式

请注意，y 轴上的数字是以千为单位标记的。为使图表更美观，也更易于理解，我们将更改刻度标记标签中的数字格式，然后对轴标题进行相应的编辑。

- ▶ 单击 y 轴任何一个刻度标记标签，将它们选中。
- ▶ 要重新打开“属性”窗口（如果之前关闭了它），请从菜单中选择。
编辑 > 属性

注意：在此，我们假定“属性”窗口已打开。如果您已关闭了“属性”窗口，请按照前一步骤将它重新打开。您也可以使用键盘快捷方式 `Ctrl+T` 来重新打开该窗口。

- ▶ 单击**数字格式**选项卡。
- ▶ 您不希望刻度标记标签显示小数位，所以在“小数位”文本框中键入 0。
- ▶ 在“比例因子”文本框中键入 0.001。比例因子是图表编辑器用于除显示的数字的数字。由于 0.001 是分数，所以除以它将使刻度标记标签中的数字扩大 1,000 倍。因此，这些数字将不再以千为单位，它们将不再按这种比例计算。
- ▶ 选择**显示数字分组**。数字分组使用字符（由计算机的区域设置指定）来标记数字中的每个千分位。

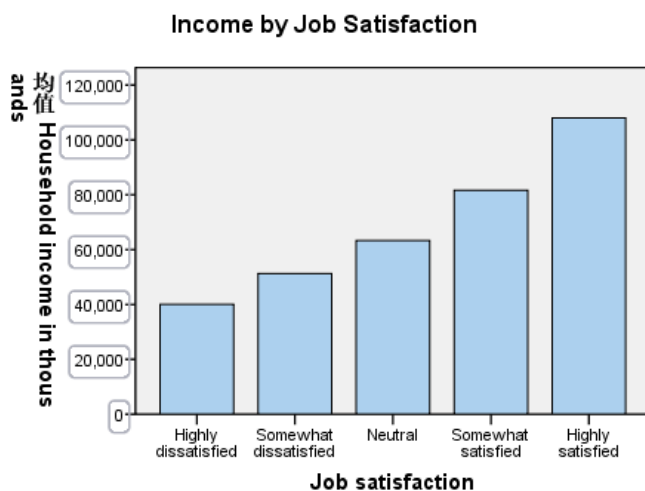
图片 6-11
“数字格式”选项卡



► 单击应用。

刻度标记标签反映新的数字格式设置。没有小数位，数字不再按比例计算，每个千分位都用字符指定。

图片 6-12
显示新数字格式的已编辑条形图



编辑文本

现在，刻度标记标签的数字格式已经更改，轴标题不再正确。接下来，将更改轴标题来反映新的数字格式。

注意：不需要打开“属性”窗口来编辑文本。可以直接在图表上编辑文本。

- ▶ 单击 y 轴标题将它选中。
- ▶ 再次单击该轴标题，启动编辑模式。在编辑模式中，图表编辑器水平放置任何旋转的文本。它也会显示不断闪烁的红色条光标（本示例中未显示）。

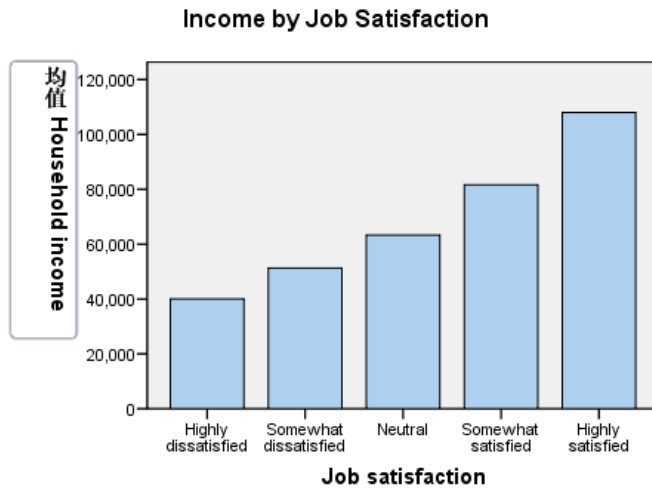
- ▶ 删除以下文字：

in thousands

- ▶ 按 Enter，退出编辑模式并更新该轴标题。现在该轴标题准确地描述了刻度标记标签的内容。

图片 6-13

显示编辑后的 y 轴标题的条形图

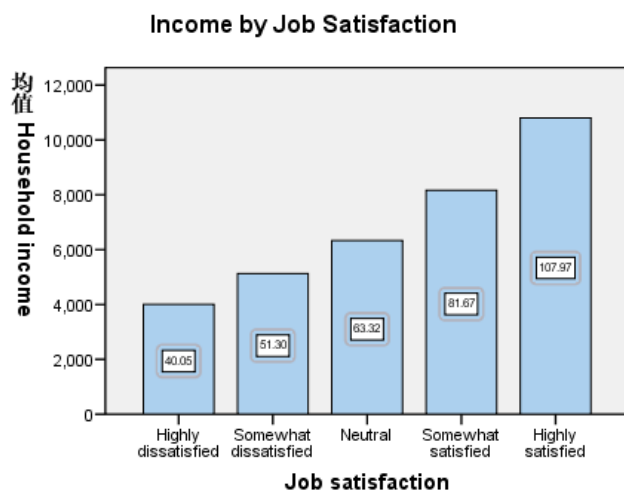


显示数据值标签

另一个常见任务是显示与图形元素（在本示例中为条）相关的确切值。这些值显示在数据标签中。

- ▶ 从“图表编辑器”菜单中选择：
元素 > 显示数据标签

图片 6-14
显示数据值标签的条形图



现在该图中的每个条都显示确切的平均家庭收入。请注意，单位是千，因此，您需要再次使用“数字格式”选项卡来更改比例因子。

使用模板

如果要对图表进行大量的常规更改，则使用图表模板可以缩短创建和编辑图表所需的时间。图表模板保存了特定图表的一些属性。在创建或编辑图表时就可以应用该模板。

我们将当前图表另存为模板，这样，创建新图表时就可以应用该模板。

- ▶ 从菜单中选择：
文件 > 保存图表模板...

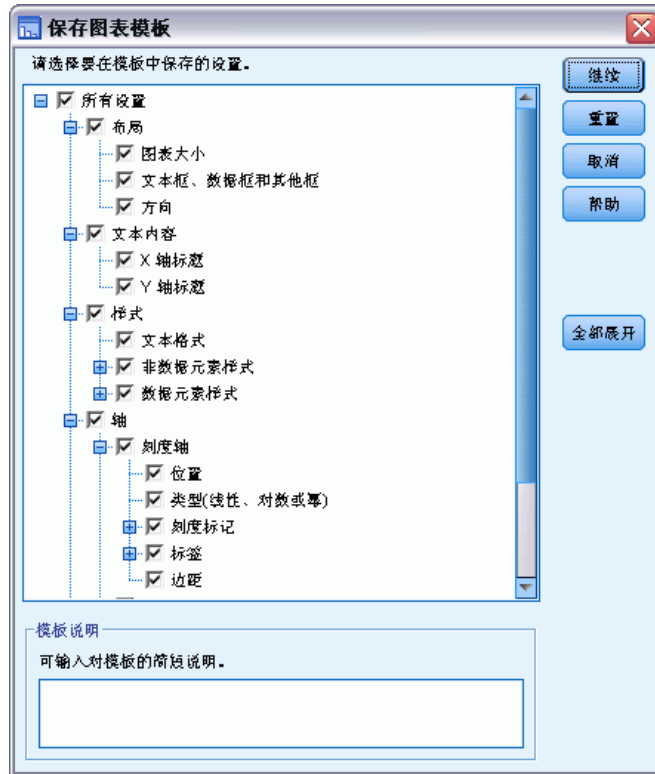
“保存图表模板”对话框可用于指定要在模板中包括哪些图表属性。

如果在树状视图中展开任何项，可以查看哪些特定的属性可以与该图表一起保存。例如，如果展开树的刻度轴部分，可以看到模板将包括的数据值标签的所有属性。您可以选择任何属性来将其包括在模板中。

- ▶ 选择所有设置，可在模板中包括所有可用的图表属性。

您还可以输入模板的说明。当您应用该模板时，会出现这一说明。

图片 6-15
“保存图表模板”对话框

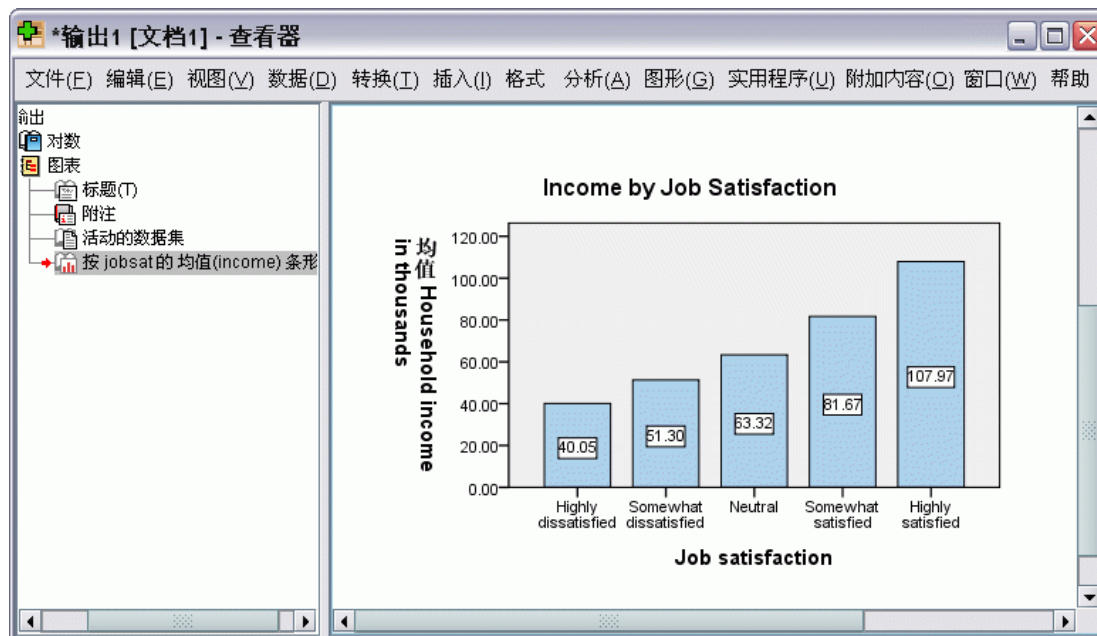


- ▶ 单击继续。
- ▶ 在“保存模板”对话框中，为模板指定位置和文件名。
- ▶ 完成操作后，单击保存。

在图表编辑器中创建图表时可以应用该模板。在下列示例中，我们在创建图表时将应用该模板。

- ▶ 关闭图表编辑器。更新后的条形图显示在查看器中。

图片 6-16
查看器中已更新的条形图



- ▶ 从“查看器”菜单中选择：
图形 > 图表构建程序...

“图表生成器”对话框会记住您在创建原始图表时输入的变量。不过，在此处将创建一个稍微不同的图表，请看如何应用模板来设置图表的格式。

- ▶ 将 x 轴的 Job satisfaction 从放置区拖回“变量”列表，从而移去它。还可以单击放置区，然后按 Delete。
- ▶ 右键单击“变量”列表中的 Level of education，并选择有序。
- ▶ 将 Level of education 从“变量”列表拖到 x 轴放置区。

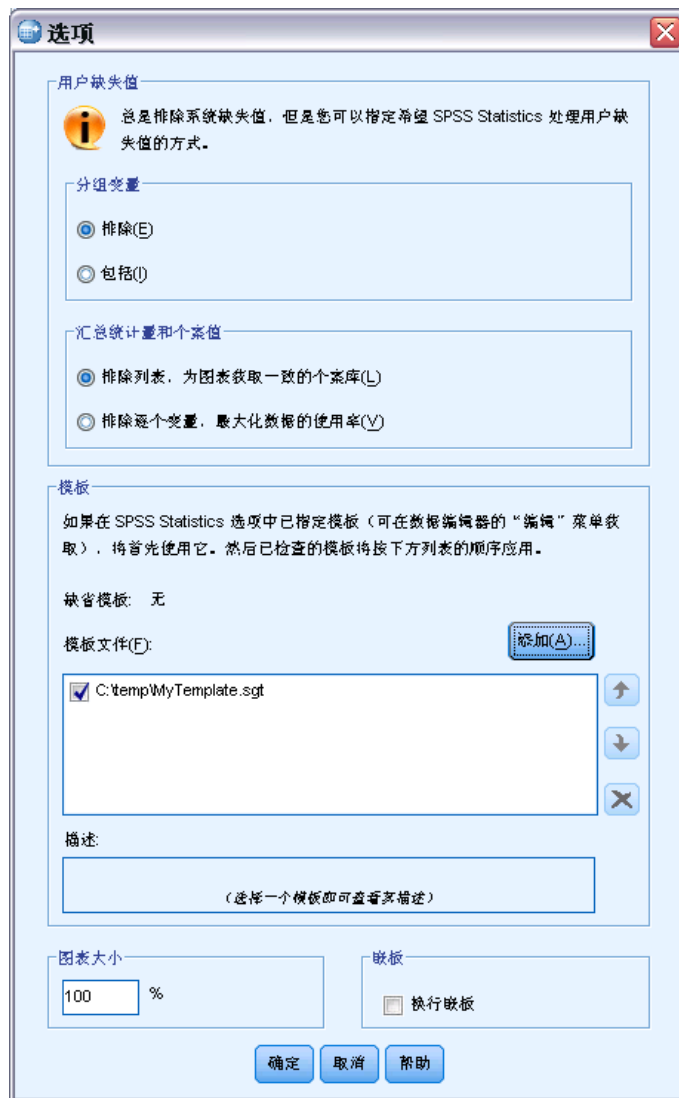
由于现在该标题不正确，我们要删除它。

- ▶ 在“标题/脚注”选项卡上，取消选择标题 1。

现在，我们将指定要应用到新图表中的模板。

- ▶ 单击选项。
- ▶ 在“选项”对话框的“模板”组中，单击添加。
- ▶ 在“查找模板文件”对话框中，找到先前使用“保存图表模板”对话框保存的模板文件。
- ▶ 选择该文件并单击打开。

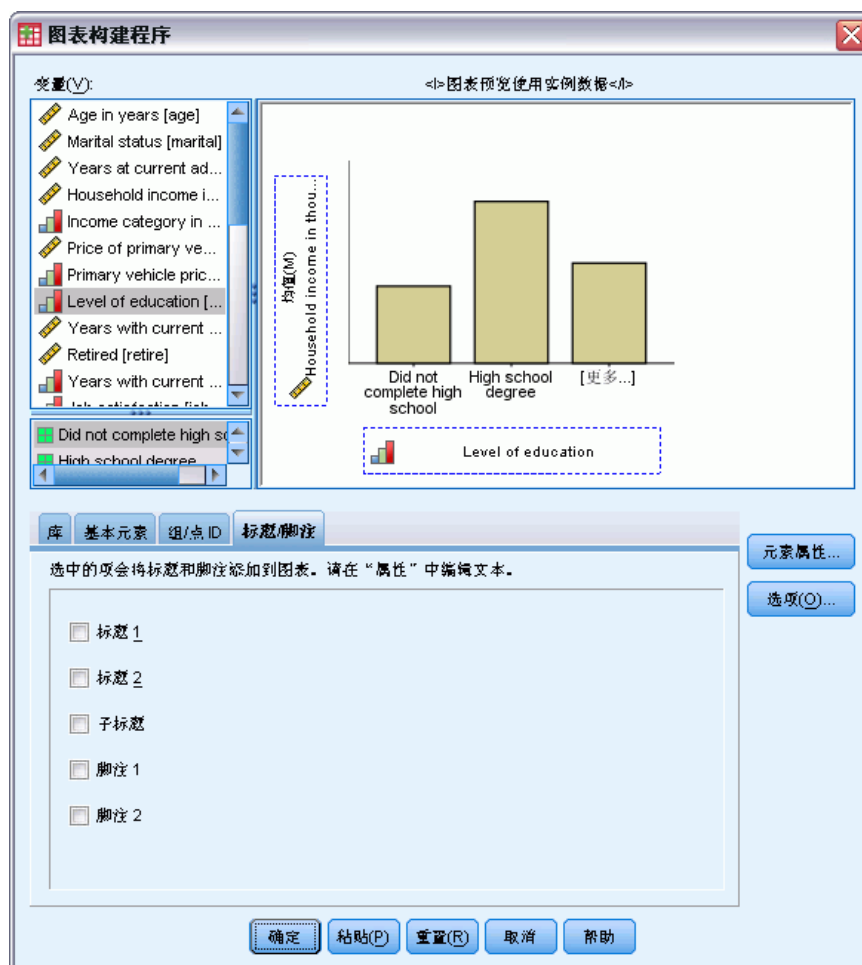
图片 6-17
带有模板的“选项”对话框



“选项”对话框随即显示您选择的模板的文件路径。

- ▶ 单击确定关闭“选项”对话框。

图片 6-18
带有完整放置区的图表生成器

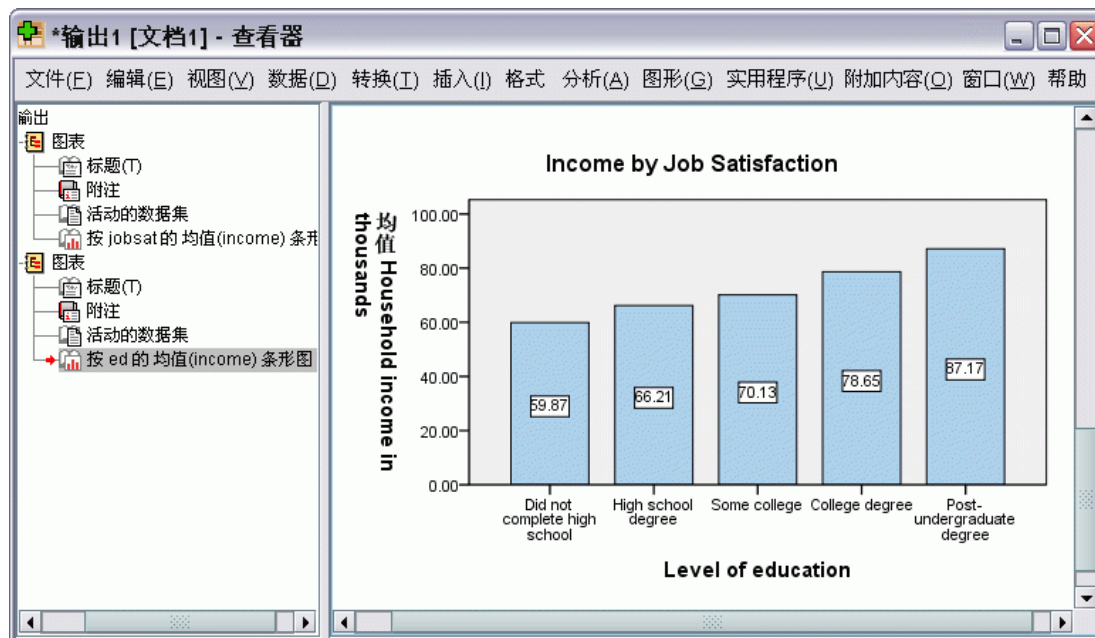


- ▶ 单击“图表生成器”对话框中的确定，创建图表并应用该模板。

新图表中的格式与先前创建并编辑的图表的格式相同。尽管 x 轴上的变量不同，但两幅图表的其他地方都相似。请注意，前一图表的标题仍保留在该模板中，尽管您在图表生成器中删除了该标题。

如果要在创建了图表后应用模板，可以在图表编辑器中执行这一操作（从“文件”菜单中，选择应用图表模板）。

图片 6-19
查看器中已更新的条形图



定义图表选项

除了使用模板来设置图表的格式外，还可以使用“选项”来控制如何创建图表的各个方面。

- ▶ 从数据编辑器或查看器菜单中选择：
编辑 > 选项...

“选项”对话框包含许多配置设置。单击图表选项卡可查看可用的选项。

图片 6-20

“选项”对话框中的“图表”选项卡



这些选项控制图表的创建方式。对于每个图表，可以指定以下设置：

- 是使用当前设置还是使用模板。
- 宽度和高度比例（宽高比）。
- 如果未使用模板，则使用这些设置进行格式设置。
- 图形元素的样式循环。

样式循环可用于指定新图表中的图形元素的样式。在本示例中，我们将查看颜色样式循环的详细信息。

- ▶ 单击颜色，打开“数据元素颜色”对话框。

对于简单的图表，图表编辑器会使用您指定的一种样式。对于分组图表，图表编辑器会将一组样式循环用于图表中的每个组（类别）。

- ▶ 选择简单图表。
- ▶ 选择浅绿色，即倒数第二行右边的第三个。

图片 6-21
“数据元素颜色”对话框



- ▶ 单击继续。
- ▶ 在“选项”对话框中，单击确定，保存颜色样式循环更改。

现在，任何新的简单图表中的图形元素都是浅绿色。

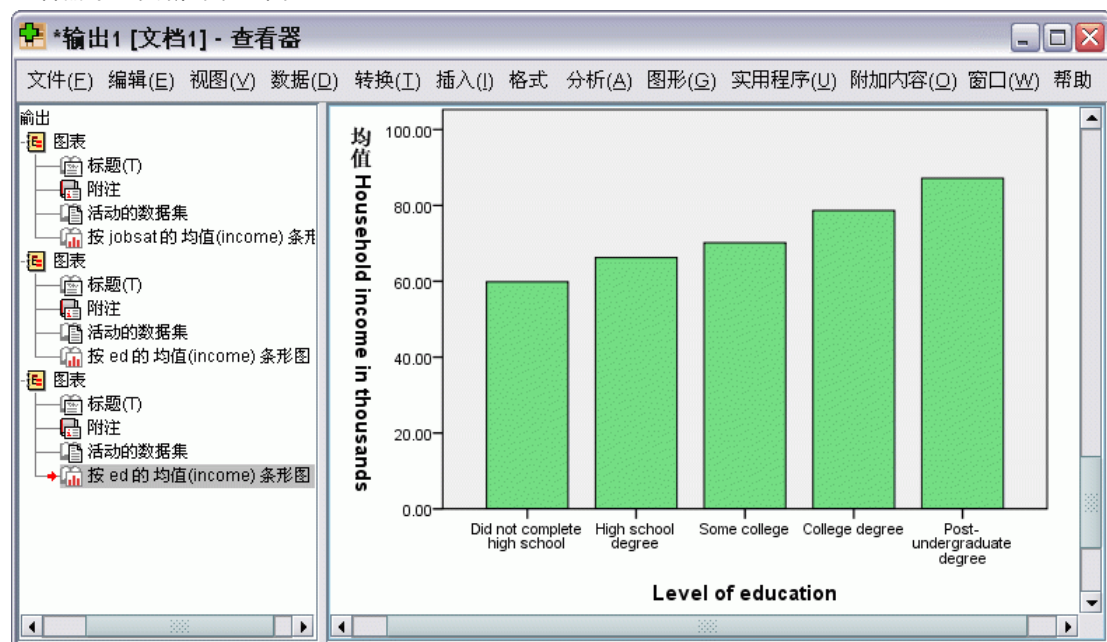
- ▶ 从数据编辑器或查看器菜单中选择：
图形 > 图表构建程序...

图表生成器显示您上次创建的图表。请记住，此图表有一个模板与之关联。我们不再需要该模板。

- ▶ 单击选项。
- ▶ 取消选择（取消选中）先前添加的模板。请注意，您也可以单击红色的 X，将该模板删除。通过取消选中而不是删除可以保留该模板，以备下次使用。
- ▶ 单击确定以创建图表。

新图表中的条是浅绿色。此外，新图表在其他方面也与前一图表不同。它没有标题、轴标签以千为单位，而且没有数据标签。之所以有这些差异，是因为没有将该模板应用于此图表。

图片 6-22
查看器中已更新的条形图

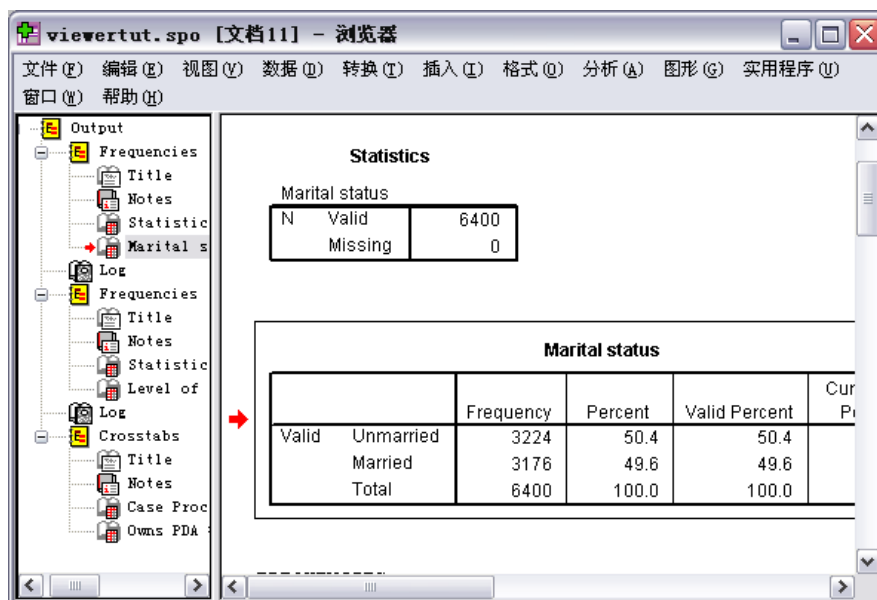


使用输出

运行统计过程所得结果将显示在查看器中。生成的输出可以为统计表、图表、图形或文本，具体取决于您在运行该过程时所作的选择。此部分使用文件 viewertut.spv 和 demo.sav。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。

使用查看器

图片 7-1
查看器



“查看器”窗口分为两个窗格。**概要窗格**包含查看器中存储的所有信息的概要。**内容窗格**包含统计表、图表和文本输出。

使用滚动条可垂直和水平地浏览窗口的所有内容。为了便于浏览，单击概要窗格中的某一项，可将它显示在内容窗格中。

如果您发现查看器中没有足够的空间来查看整个表，或概要视图太窄，则可轻松调整窗口的大小。

- ▶ 单击并拖动概要窗格的右边界可更改其宽度。

概要窗格中打开的书形图标指示当前在查看器中可看到它，尽管它当前可能不在内容窗格的可视部分。

- ▶ 要隐藏某表或图表，请在概要窗格中双击其对应的书形图标。

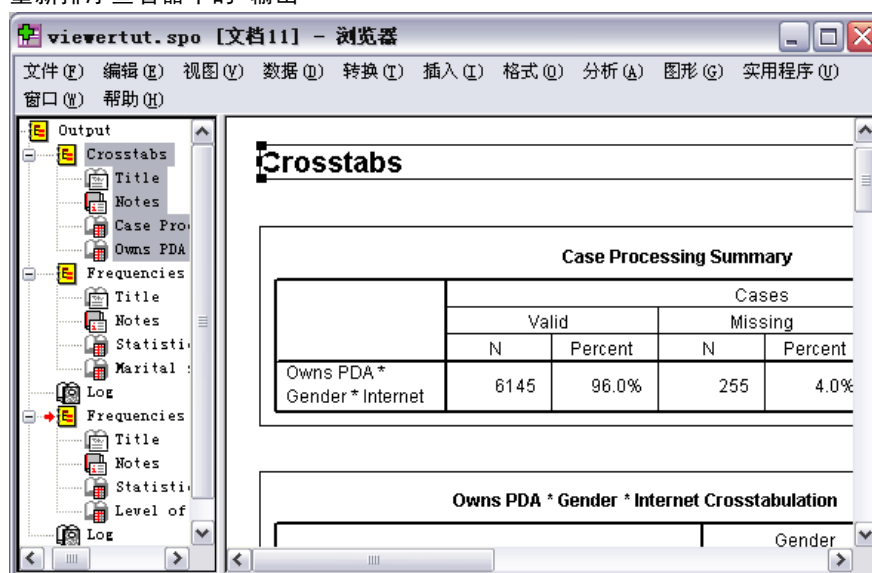
打开的书形图标随即变成合上的书形图标，这表示与之相关联的信息现已隐藏。

- ▶ 要再次显示隐藏的输出，请双击合上的书形图标。
还可以隐藏特定统计过程生成的所有输出或查看器中的全部输出。
- ▶ 单击要隐藏其结果的过程左侧的减号（-）框；或单击概要窗格中最上面一项旁边的框，隐藏所有的输出。
概要随即折叠，直观表示这些结果已隐藏。

您也可以更改输出的显示顺序。

- ▶ 在概要窗格中，单击要移动的项。
- ▶ 将选中的项拖到概要窗格中的新位置，然后释放鼠标按钮。

图片 7-2
重新排序查看器中的 输出



通过单击输出项并将其拖至内容窗格，也可移动输出项。

使用枢轴表编辑器

大多数统计过程的结果都显示在**枢轴表**中。

访问输出定义

输出中将显示许多统计术语。这些术语的定义可以直接在查看器中访问。

- ▶ 双击 Owens PDA * Gender * Internet Crosstabulation 表。
- ▶ 右键单击期望计数，然后从弹出上下文菜单中选择这是什么？。

相关定义将显示在弹出窗口中。

图片 7-3
弹出式 定义

The screenshot shows a window titled '*viewertut.spv [文档3] - 查看器'. The main content is a crosstabulation table titled 'Owns PDA * Gender * Internet Crosstabulation'. The table has columns for 'Internet' (No, Yes), 'Owns PDA' (No, Yes), and 'Gender' (Female, Male), with a 'Total' column. The table displays counts, expected counts, and percentages. A tooltip is displayed over the 'Expected Count' cell for 'Internet: No, Owns PDA: No, Gender: Female', explaining that it represents the number of cases expected if the variables are statistically independent.

				Gender		Total
				Female	Male	
Internet No	Owns PDA No	Count		1897	1962	3859
		Expected Count		1903.4	1955.6	3859.0
			0.8%		100.0%	
			323	650	650.0	
		% within Owns PDA		50.3%	49.7%	100.0%
	Total	Count		2224	2285	4509
		Expected Count		2224.0	2285.0	4509.0
		% within Owns PDA		49.3%	50.7%	100.0%
Internet Yes	Owns PDA No	Count		513	512	1025
		Expected Count		512.5	512.5	1025.0
			50.0%	50.0%	100.0%	
			305	611	611.0	
		% within Owns PDA		49.9%	50.1%	100.0%
	Yes	Count		305	306	611
		Expected Count		305.5	305.5	611.0
		% within Owns PDA		49.9%	50.1%	100.0%
	Total	Count		818	818	1636
		Expected Count		818.0	818.0	1636.0
		% within Owns PDA		50.0%	50.0%	100.0%

透视表

生成的默认表显示的信息可能不像您期望的那样美观或清晰。使用枢轴表，您可以转置行和列（“翻转”表），调整表中数据的顺序以及采用许多其他方式修改表。例如，通过转置行和列，可以将短而宽的表改成长而细的表。更改表的布局并不会影响结果，而是以不同方式或更合意的方式显示信息的一种方法。

- ▶ 如果其尚未激活，双击 Owns PDA * Gender * Internet Crosstabulation 表以激活它。
- ▶ 如果“透视托盘”窗口不可见，请从菜单中选择：
透视 > 透视托盘

透视托盘提供了在列、行和层之间移动数据的方法。

图片 7-4
旋转托盘

Owens PDA * Gender * Internet Crosstabulation

Internet				Gender		Total
				Female	Male	
No	Count	Owens PDA	No	1897	1962	3859
			Yes	327	323	650
	Total		2224	2285	4509	
Expected Count	Owens PDA	No	1903.4	1955.6	3859.0	
		Yes	320.6	329.4	650.0	
	Total					
% within Owens PDA		Owens PDA				
Total						
Yes	Count	Owens PDA	No			
			Yes			
	Total					
Expected Count	Owens PDA	No				
		Yes				
	Total					
% within Owens PDA		Owens PDA				
Total						

- ▶ 将 Statistics 元素从行维度拖动到列维度，在 Gender 项。该表会根据您的更改即刻重新配置。

图片 7-5
将行移至 列

Internet			Gender						Total		
			Female			Male			Count	Expected Count	% within Owens PDA
Count	Expected Count	% within Owens PDA	Count	Expected Count	% within Owens PDA	Count	Expected Count	% within Owens PDA			
No	Owens PDA	No	1897				3859	3859.0	100.0%		
		Yes	327				650	650.0	100.0%		
Total			2224				4509	4509.0	100.0%		
Yes	Owens PDA	No	513				1025	1025.0	100.0%		
		Yes	305				611	611.0	100.0%		
Total			818				1636	1636.0	100.0%		

透视托盘中的元素顺序反映了表中各元素的顺序。

- ▶ 先拖放 Owns PDA 元素，然后是行维度中的 Internet 元素以将这两行的顺序颠倒。

图片 7-6
交换行

		Gender						Total		
		Female			Male			Total		
		Count	Expected Count	% within Owns PDA	Count	Expected Count	% within Owns PDA	Count	Expected Count	% within Owns PDA
Owns PDA	No	1897						3859	3859.0	100.0%
	Yes	513						1025	1025.0	100.0%
	No	327						650	650.0	100.0%
	Yes	305						611	611.0	100.0%
Total	No	2224						4509	4509.0	100.0%
	Yes	818						1636	1636.0	100.0%

创建并显示层

对于具有嵌套类别的信息的大型表而言，层会比较有用。通过创建层，可以简化表的外观，使它更简单明了。

- ▶ 将 Gender 元素从列维度拖至层维度。

图片 7-7
层维度中的 Gender 透视图标

		Gender		
		Count	Expected Count	% within Owns PDA
Owns PDA	No	1897	1903.4	49.2%
	Yes			
Total	No			
	Yes			

为显示不同层，从表中的下拉列表选择类别。

图片 7-8
选择层

Gender		Gender Male		期望计数	% 内 Owns PDA
		Gender Female			
		Gender Male			
		合计			
Owns PDA	No	No	1897	1903.4	49.2%
		Yes	513	512.5	50.0%
	Yes	No	327	320.6	50.3%
		Yes	305	305.5	49.9%
合计	No	2224	2224.0	49.3%	
	Yes	818	818.0	50.0%	

编辑表

除非您花时间创建自定义的表格外观，否则枢轴表都是以标准格式创建的。表中任何文本的格式都可以更改。可更改的格式包括字体名称、字体大小、字体样式（粗体或斜体）以及颜色。

- ▶ 双击 Level of education 表。
- ▶ 如果格式设置工具栏不可见，请从菜单中选择：
视图 > 工具栏
- ▶ 单击标题文本 Level of education。
- ▶ 从工具栏上字体大小下拉列表中选择 12。
- ▶ 要更改标题文本的颜色，请单击文本颜色并选择一种新的颜色。

图片 7-9
重新设置枢轴表中 标题文字的格式



		频率	百分比	有效百分比	累积百分比
有效	未完成高中	460	54.1	54.1	54.1
	高中	235	27.6	27.6	81.8
	大专	101	11.9	11.9	93.6
	大学	49	5.8	5.8	99.4
	研究生	5	.6	.6	100.0
	合计	850	100.0	100.0	

也可以编辑表和标签的内容。例如，可以更改此表的标题。

- ▶ 双击该标题。
- ▶ 键入 Education Level 作为新的标签。

注意：如果更改表中的值，则不会重新计算合计以及其他统计数据。

隐藏行与列

表中显示的某些数据可能没有用，也可能使表变得不必要地复杂。不过您可以在不丢失任何数据的情况下隐藏整个行和列。

- ▶ 如果其尚未激活，双击 Education Level 表以激活它。
- ▶ 单击 Valid Percent 列标签以选中该列。
- ▶ 从“编辑”菜单或右键单击上下文菜单中选择：
选择 > 数据和标签单元格
- ▶ 从查看器菜单选择隐藏或从右键单击上下文菜单中选择隐藏类别。
现在，该列已隐藏，但并没有删除。

图片 7-10
表中“Valid Percent”列已隐藏

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not complete high school	1390	21.7	21.7	21.7
	High school degree	1936	30.3	30.3	52.0
	Some college	1360	21.3	21.3	73.2
	College degree	1355	21.2	21.2	94.4
	Post-undergraduate degree	359	5.6	5.6	100.0
	Total	6400	100.0	100.0	

要重新显示该列：

- ▶ 从菜单中选择：
视图 > 全部显示

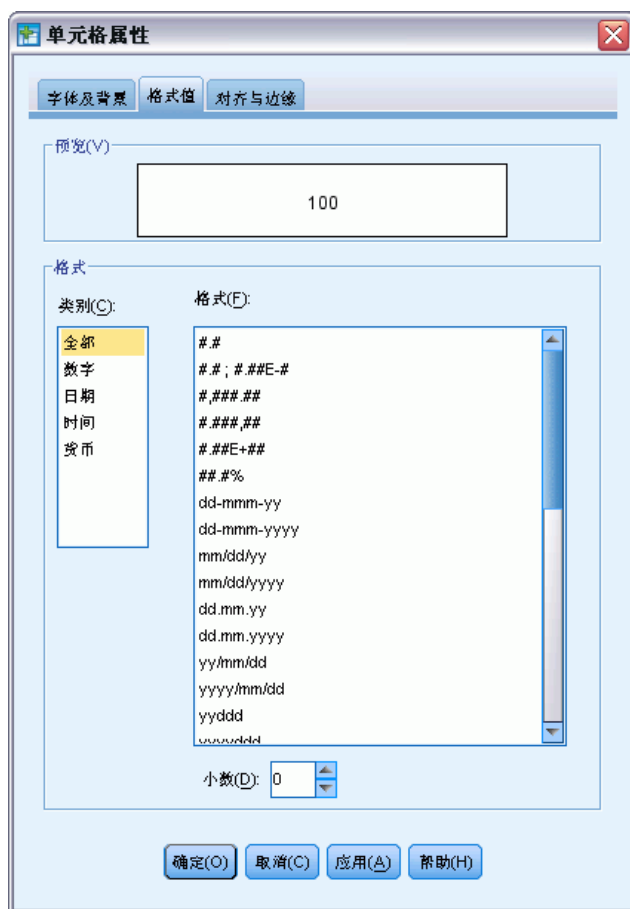
也可以隐藏列相同的方法隐藏和显示行。

更改数据显示格式

可以轻松更改枢轴表中数据的显示格式。

- ▶ 如果其尚未激活，双击 Education Level 表以激活它。
- ▶ 单击 Percent 列标签以选中该列。
- ▶ 从“编辑”菜单或右键单击上下文菜单中选择：
选择 > 数据单元格
- ▶ 从格式菜单或右键单击上下文菜单中选择单元格属性。
- ▶ 单击格式值选项卡。
- ▶ 在“小数”字段中键入 0 以隐藏该列中的所有小数点。

图片 7-11
单元格属性，格式值选项卡

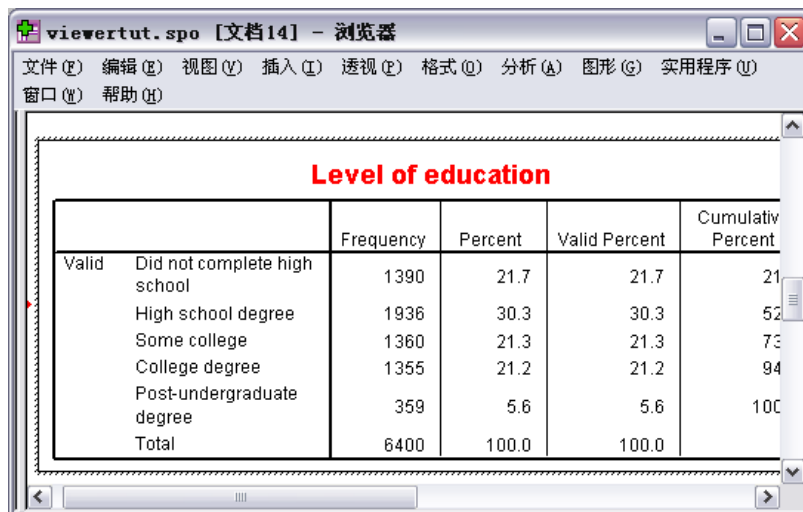


在此对话框中还可以更改数据类型和格式。

- ▶ 从“类别”列表中选择所需类型，然后在“格式”列表中选择该类型的格式。

- ▶ 单击确定或应用以应用您的更改。

图片 7-12
“Percent” 列中的小数已隐藏



		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Did not complete high school	1390	21.7	21.7	21.7
	High school degree	1936	30.3	30.3	52.0
	Some college	1360	21.3	21.3	73.3
	College degree	1355	21.2	21.2	94.5
	Post-undergraduate degree	359	5.6	5.6	100.0
	Total	6400	100.0	100.0	

现在，Percent 列中的小数已隐藏。

表格外观

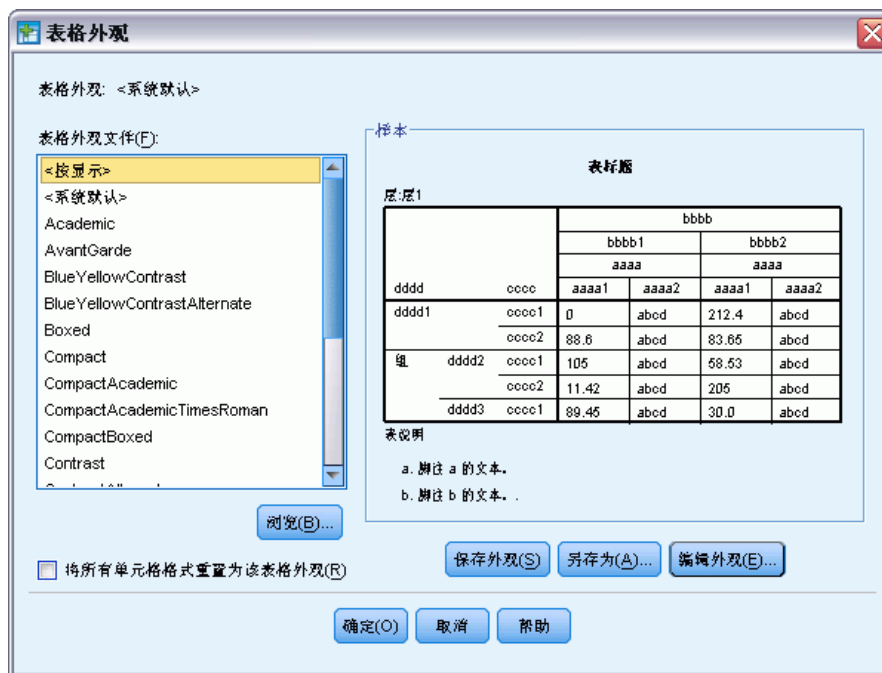
要提供清晰、简洁并有意义的结果，表的格式至关重要。如果表难以查看，则表中包含的信息也就不容易理解。

使用预先定义的格式

- ▶ 双击 Marital status 表。
- ▶ 从菜单中选择：
格式 > 表格外观...

“表格外观”对话框列出了各种预先定义的样式。从列表中选择一种样式，在右侧“示例”窗口中对其进行预览。

图片 7-13
“表格外观”对话框



可以按原样使用一种样式，也可以对现有样式进行编辑，使之更符合您的需要。

- ▶ 要使用现有样式，请选择一个样式，然后单击确定。

自定义表格外观样式

可以根据特定的需要自定义格式。对一个表而言，从背景颜色到边框样式，几乎所有方面都可以自定义。

- ▶ 双击 Marital status 表。
- ▶ 从菜单中选择：
格式 > 表格外观...
- ▶ 选择与您所需格式最接近的样式，然后单击编辑外观。

- ▶ 单击单元格格式选项卡以查看格式设置选项。

图片 7-14
“表格属性”对话框



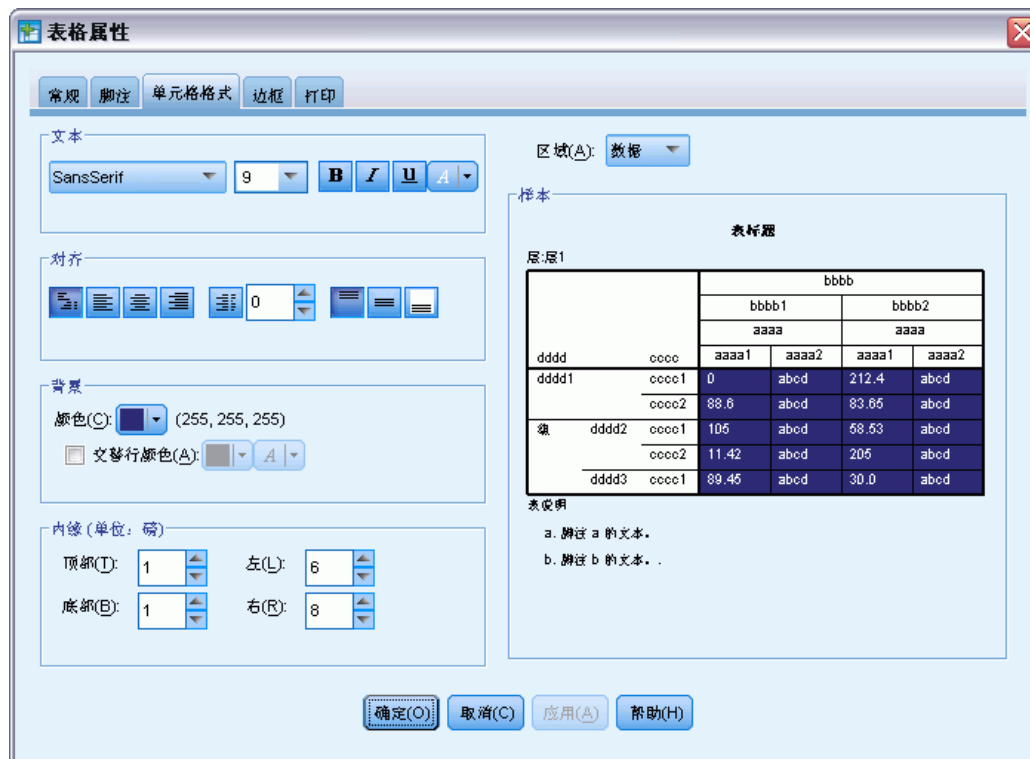
格式设置选项包括字体名称、字体大小、字体样式和颜色。其他选项包括对齐、文本和背景颜色以及边距尺寸。

右侧的“示例”窗口可用于预览格式设置更改对表的影响。表的每个区域都可以有一个不同的格式样式。例如，您可能不希望标题的样式与数据的样式相同。要选择进行编辑的表区域，则可在“区域”下拉列表中按名称选择该区域，也可以在“示例”窗口单击要更改的区域。

- ▶ 从“区域”下拉列表中选择数据。
- ▶ 从背景下拉调色板中选择新的颜色。
- ▶ 然后选择一种新的文本颜色。

此时，“示例”窗口将显示新的样式。

图片 7-15
更改表格单元格格式



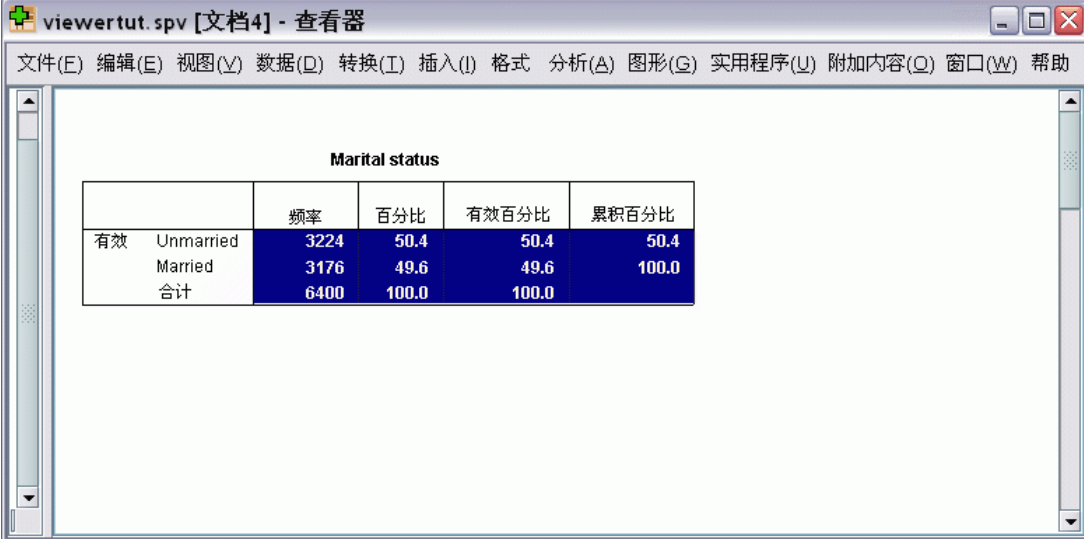
- ▶ 单击确定回到“表格外观”对话框。

您可以保存该新样式，以便将来轻松将其应用于其他表格。

- ▶ 单击另存为。
- ▶ 导航到所需目标目录，然后在“文件名”文本框中输入新样式的名称。
- ▶ 单击保存。
- ▶ 单击确定应用更改并返回到查看器。

现在，表中将包含指定的自定义格式设置。

图片 7-16
自定义表格外观



		频率	百分比	有效百分比	累积百分比
有效	Unmarried	3224	50.4	50.4	50.4
	Married	3176	49.6	49.6	100.0
	合计	6400	100.0	100.0	

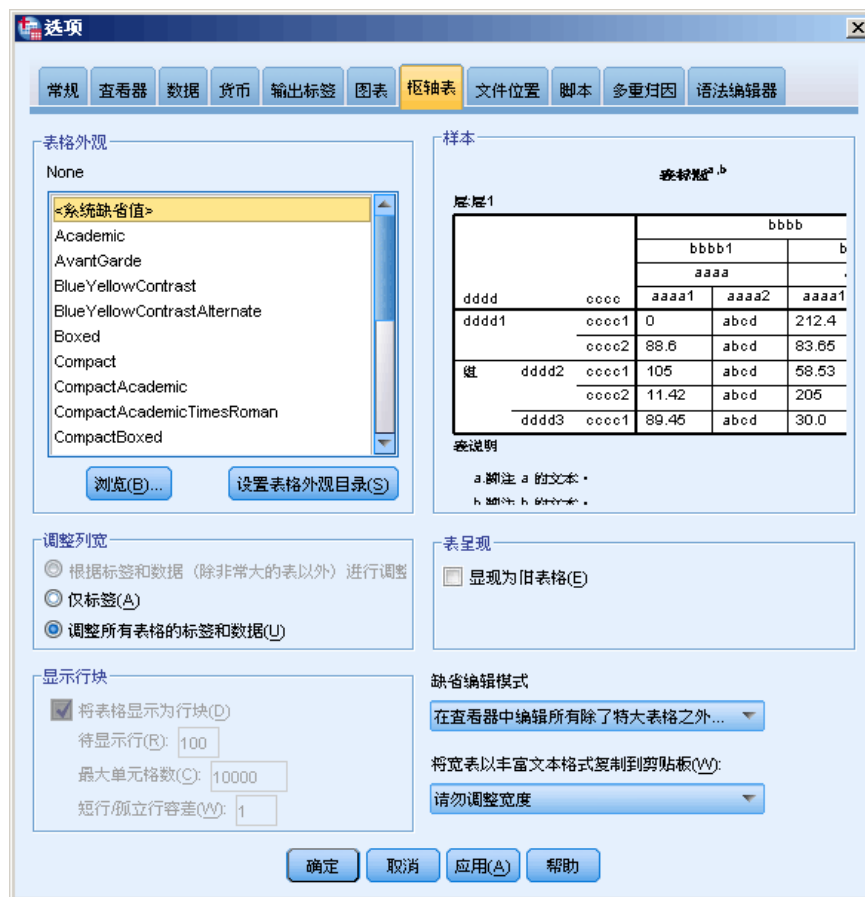
更改默认的表格格式

尽管在创建表之后可以更改其格式，但更改默认的表格外观更为高效，因为这样就不必在每次创建表后都更改其格式了。

要更改枢轴表的默认表格外观样式，请从菜单中选择：
编辑 > 选项...

- ▶ 单击“选项”对话框中的枢轴表选项卡。

图片 7-17
“选项”对话框



- ▶ 选择要用于所有新表的表格外观样式。

此时，右侧的“样本”窗口中将显示每种表格外观的预览。

- ▶ 单击确定保存设置并关闭该对话框。

在更改默认表格外观后创建的所有表都将自动遵循新的格式设置规则。

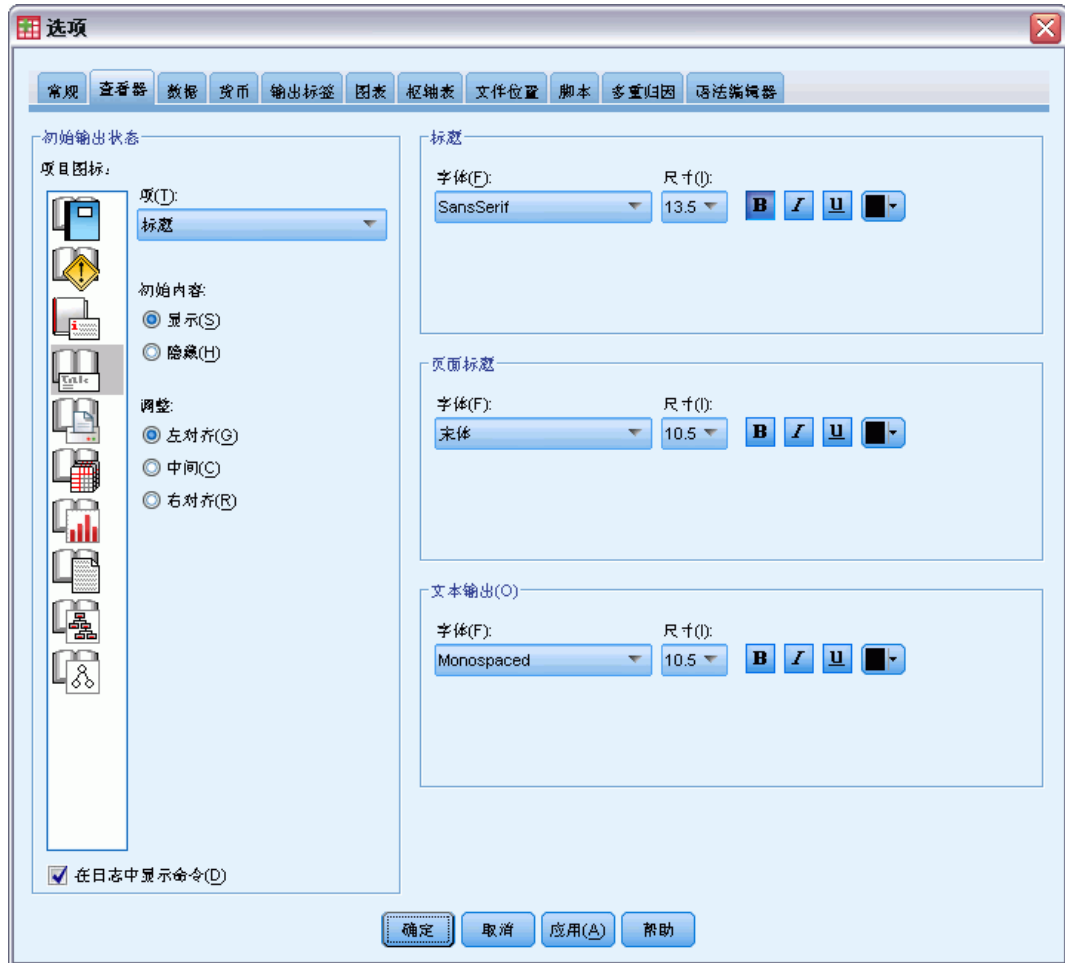
自定义初始显示设置

初始显示设置包括查看器中各对象的对齐，默认情况下对象是显示还是隐藏，以及“查看器”窗口的宽度。要更改这些设置：

- ▶ 从菜单中选择：
编辑 > 选项...

- ▶ 单击查看器选项卡。

图片 7-18
查看器 选项



逐个对对象应用设置。例如，可以自定义图表的显示方式，而无需对表的显示方式进行任何更改。只需选中要自定义的对象，然后进行所需更改即可。

- ▶ 单击标题图标以显示其设置。
- ▶ 单击居中，可将所有标题显示在查看器的（水平）中央。

也可以隐藏可能打乱输出的元素，例如日志和警告消息。双击某个图标可自动更改该对象的显示属性。

- ▶ 双击警告图标可在输出中隐藏警告消息。
- ▶ 单击确定保存所做更改并关闭该对话框。

显示变量和值标签

在大多数情况下，显示变量和值的标签比显示变量名称或实际数据值更为有效。但是，有时您可能会希望既显示名称又显示标签。

- ▶ 从菜单中选择：
编辑 > 选项...
- ▶ 单击输出标签选项卡。

图片 7-19
“输出标签”选项

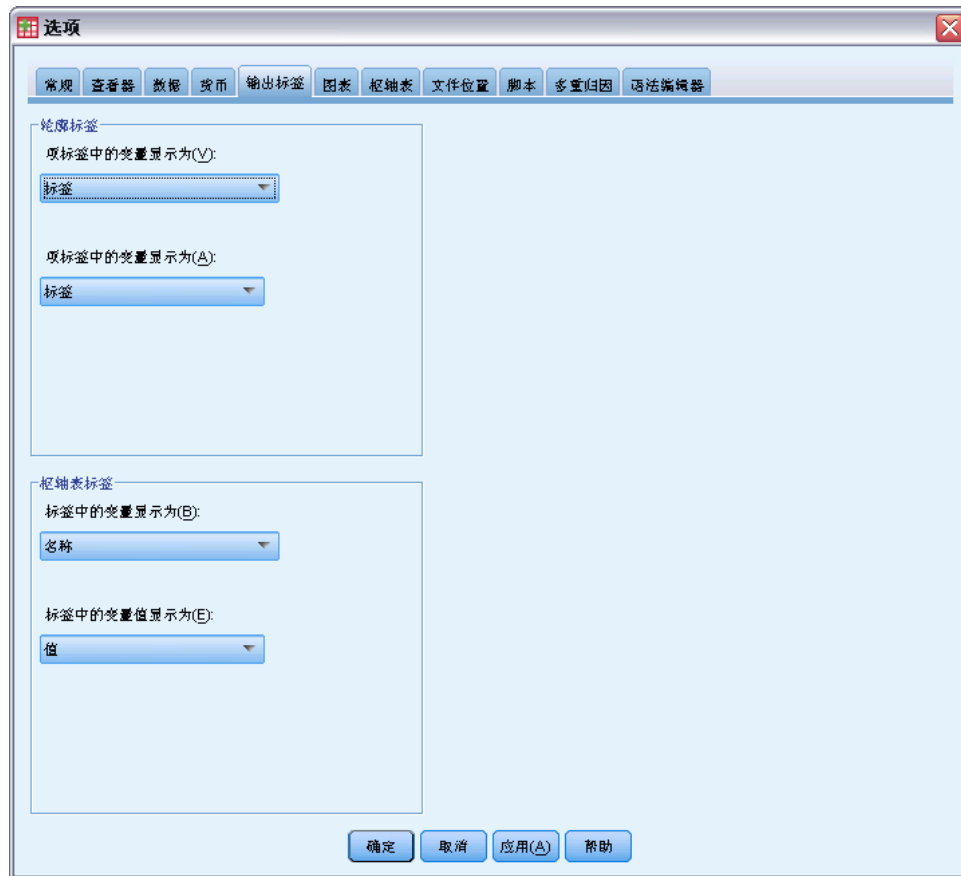


可以为概要窗格和内容窗格指定不同的设置。例如，在概要窗格中显示标签，而在内容窗格中显示变量名称和数据值。

- ▶ 在“枢轴表标签”组中，从“标签中的变量显示为”下拉列表中选择名称以显示变量名称而不是标签。

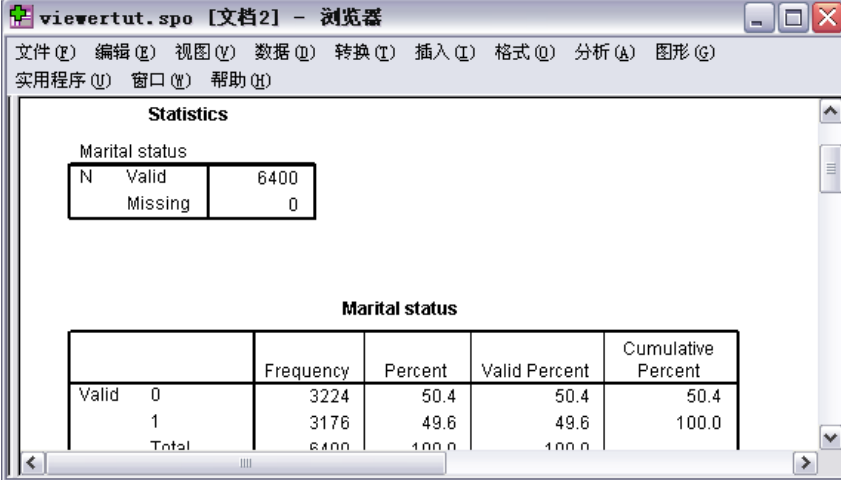
- ▶ 然后，从“标签中的变量值显示为”下拉列表中选择值以显示数据值而不是标签。

图片 7-20
枢轴表标签设置



在会话中生成的后续表将反映这些更改。

图片 7-21
显示的变量名和值



The screenshot shows the SPSS Statistics window for 'viewertut.spo'. The 'Statistics' dialog box is open for the variable 'Marital status'. The 'N Valid' is 6400 and 'Missing' is 0. Below the dialog, a summary table for 'Marital status' is displayed.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	3224	50.4	50.4	50.4
	1	3176	49.6	49.6	100.0
Total		6400	100.0	100.0	

在其他应用程序中使用结果

输出结果可用在许多应用程序中。例如您可能希望将表格或图表包含在演示文稿或报表中。

以下示例专门针对 Microsoft Word，但他们在其他文字处理应用程序中的工作方式类似。

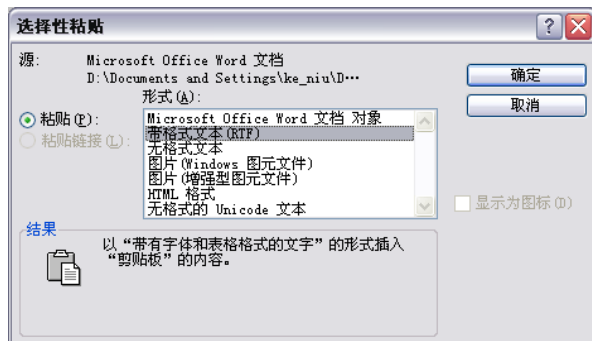
将结果作为 Word 表格粘贴

可以将枢轴表作为 Word 本地表粘贴到 Word 中。所有表属性，例如字体大小和颜色，都将保留。由于表是以 Word 表格式粘贴的，因此可以在 Word 中像对任何其他表一样进行编辑。

- ▶ 单击查看器中的 Marital status 表。
- ▶ 从菜单中选择：
编辑 > 粘贴
- ▶ 打开字处理应用程序。
- ▶ 从字处理程序的菜单中选择：
编辑 > 选择性粘贴...

- ▶ 在“选择性粘贴”对话框中，选择带格式文本(RTF)。

图片 7-22
“选择性粘贴”对话框



- ▶ 单击确定将结果粘贴到当前文档中。

图片 7-23
Word 中显示的枢轴表

Marital status					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unmarried	3224	50.4	50.4	50.4
	Married	3176	49.6	49.6	100.0
Total		6400	100.0	100.0	

现在，该表将出现在文档中。可以根据需要应用自定义的格式设置、编辑数据和调整表的大小。

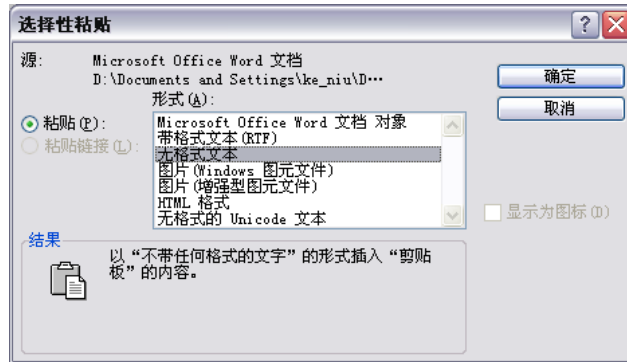
将结果作为文本进行粘贴

枢轴表可以作为纯文本格式复制到其他应用程序中。在此方法中，不会保留格式样式，但在将表数据粘贴到目标应用程序中后可以对其进行编辑。

- ▶ 单击查看器中的 Marital status 表。
- ▶ 从菜单中选择：
编辑 > 粘贴
- ▶ 打开字处理应用程序。

- ▶ 从字处理程序的菜单中选择：
编辑 > 选择性粘贴...
- ▶ 在“选择性粘贴”对话框中，选择无格式文本。

图片 7-24
“选择性粘贴”对话框



- ▶ 单击确定将结果粘贴到当前文档中。

图片 7-25
Word 中显示的枢轴表

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unmarried	3224	50.4	50.4	50.4
	Married	3176	49.6	49.6	100.0
	Total	6400	100.0	100.0	

表的每一列都使用制表符进行分隔。在文字处理应用程序中，可以通过调整制表位来更改列宽。

将结果导出到 Microsoft Word、PowerPoint 和 Excel 文件

您可以将结果导出到 Microsoft Word、PowerPoint 或 Excel 文件。可以导出查看器中的选定项或所有项。此部分使用文件 msouttut.spv 和 demo.sav。[有关详细信息，请参阅第 153 页码附录 A 中的样本文件。](#)

注意：仅在 Windows 操作系统才可以使用导出到 PowerPoint 功能，学生版未提供导出到 PowerPoint 功能。

在查看器的概要窗格中，您可以选择要导出的特定项、导出所有项或导出所有可见的项。

- ▶ 从“查看器”菜单中选择：
文件 > 导出...

可以选择只导出可视对象（概要窗格中打开的书形图标）或在概要窗格中选定的那些对象，而无需导出查看器中的全部对象。如果未在概要窗格中选择任何项，则不能选择导出选定对象。

图片 7-26
“导出输出”对话框

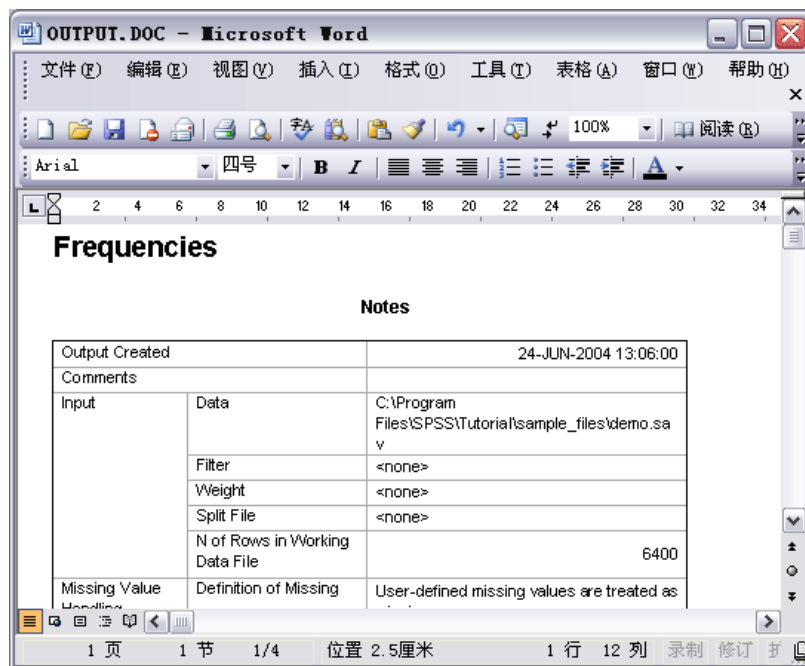


- ▶ 在导出对象组中，选择所有。
- ▶ 从“类型”下拉列表中选择 Word/RTF 文件 (*.doc)。
- ▶ 单击确定生成该 Word 文件。

在 Word 中打开生成的文件时，可以查看结果的导出方式。请注意，由于选择了导出所有对象，因此不可视对象也会出现在 Word 中。

图片 7-27

Word 中的 Output.doc



此时，枢轴表将成为 Word 表，原始枢轴表的所有格式均已保留，包括字体、颜色、边框等等。

图片 7-28
Word 中的枢轴表

Gender

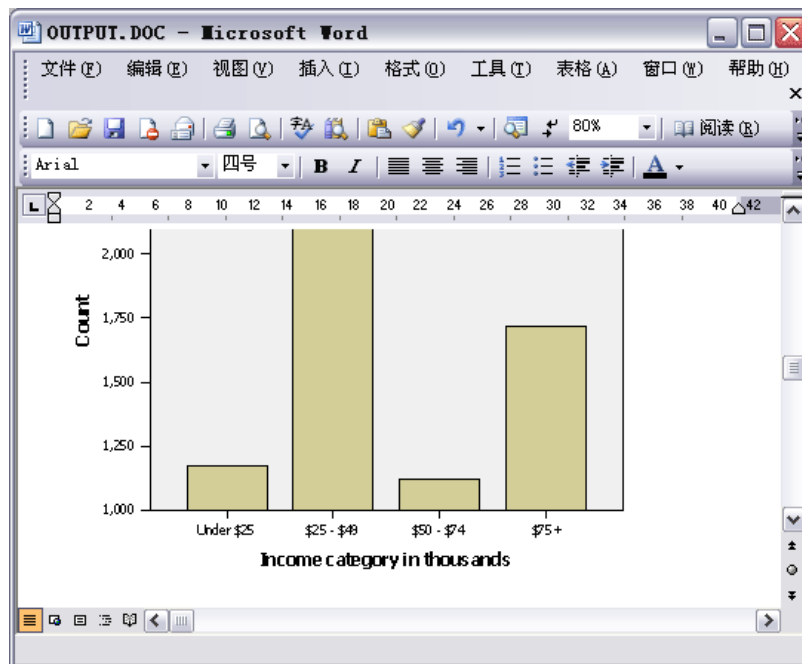
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	3179	49.7	49.7	49.7
	Male	3221	50.3	50.3	100.0
	Total	6400	100.0	100.0	

Income category in thousands

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Under \$25	1174	18.3	18.3	18.3
	\$25 - \$49	2388	37.3	37.3	55.7
	\$50 - \$74	1120	17.5	17.5	73.2
	\$75+	1918	29.9	29.9	100.0

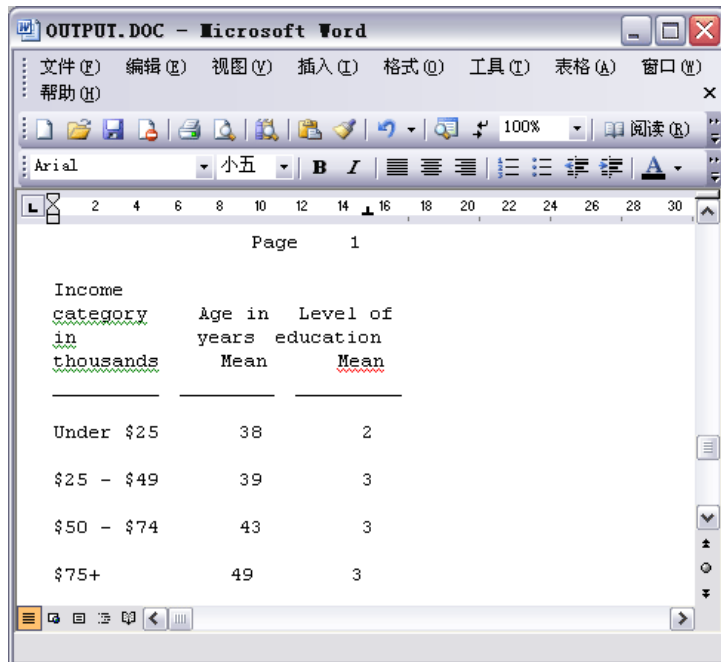
图标作为图形图像包含在 Word 文档中。

图片 7-29
Word 中的图表



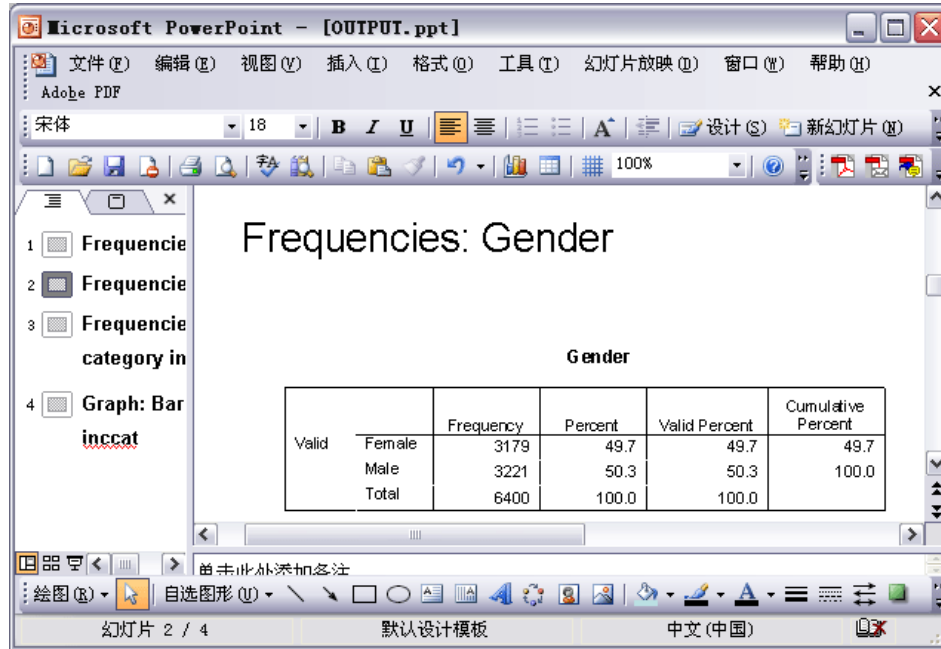
文本输出显示的字体与用于查看器文本对象的字体相同。对于正确对齐，文本输出应使用固定间距（等宽）字体。

图片 7-30
Word 中的文本输出



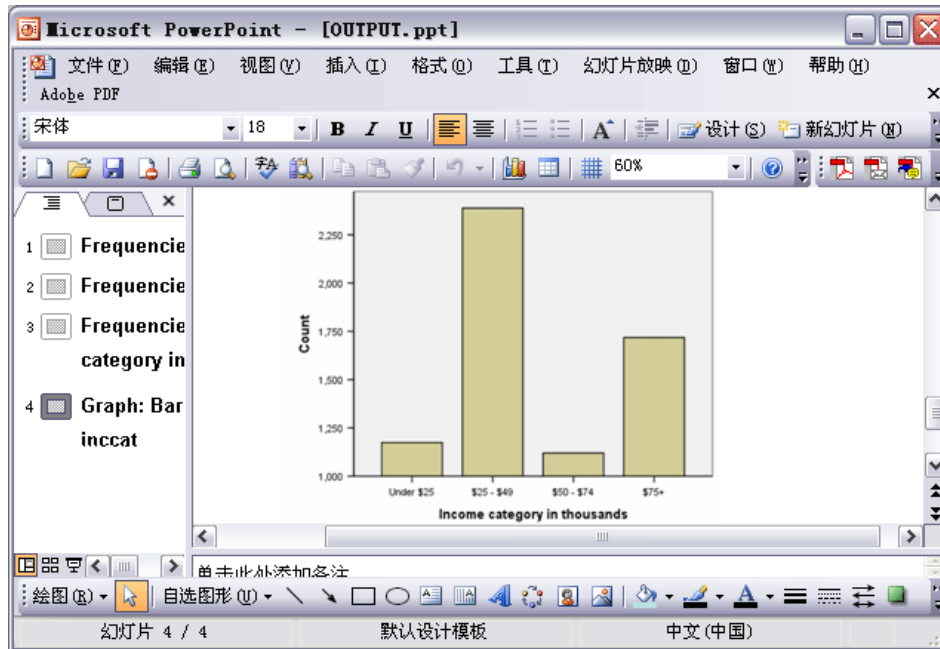
若要导出到 PowerPoint 文件，则每个导出的项均将放置于单独的幻灯片上。此时，导出到 PowerPoint 的枢轴表将成为 Word 表，原始枢轴表的所有格式都已保留，包括字体、颜色、边框等等。

图片 7-31
PowerPoint 中的枢轴表



选定导出到 PowerPoint 的图表将内嵌在 PowerPoint 文件中。

图片 7-32
PowerPoint 中的图表



若要导出到 Excel 文件，则导出的结果会有所不同。

图片 7-33
Excel 中的 Output.xls

The screenshot shows an Excel spreadsheet with a pivot table. The pivot table is structured as follows:

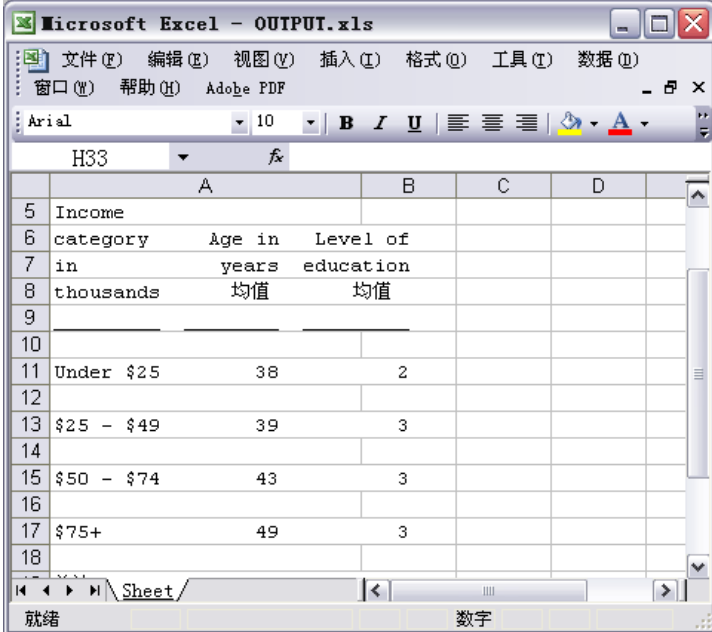
		频率	百分比	有效百分比	累积百分比
Gender					
有效	Female	3,179	49.7	49.7	49.7
	Male	3,221	50.3	50.3	100.0
	合计	6,400	100.0	100.0	
Income category in thousands					
有效	Under \$25	1,174	18.3	18.3	18.3
	\$25 - \$49	2,388	37.3	37.3	55.7
	\$50 - \$74	1,120	17.5	17.5	73.2
	\$75+	1,718	26.8	26.8	100.0
	合计	6,400	100.0	100.0	

枢轴表的行、列和单元格将成为 Excel 行、列和单元格。

文本输出中的一行即为 Excel 文件中的一行，文本输出中行的所有内容包含在单个单元格中。

图片 7-34

Excel 中的文本输出



	A	B	C	D
5	Income			
6	category	Age in	Level of	
7	in	years	education	
8	thousands	均值	均值	
9				
10				
11	Under \$25	38	2	
12				
13	\$25 - \$49	39	3	
14				
15	\$50 - \$74	43	3	
16				
17	\$75+	49	3	
18				

将结果导出到 PDF

可以将查看器中的所有项或选定项导出到 PDF（可移植文档格式）文件中。

- ▶ 从包含要导出到 PDF 的结果的“查看器”窗口的菜单中选择：
文件 > 导出...

- ▶ 从“导出输出”对话框的“导出格式文件类型”下拉列表中，选择可移植文档格式。

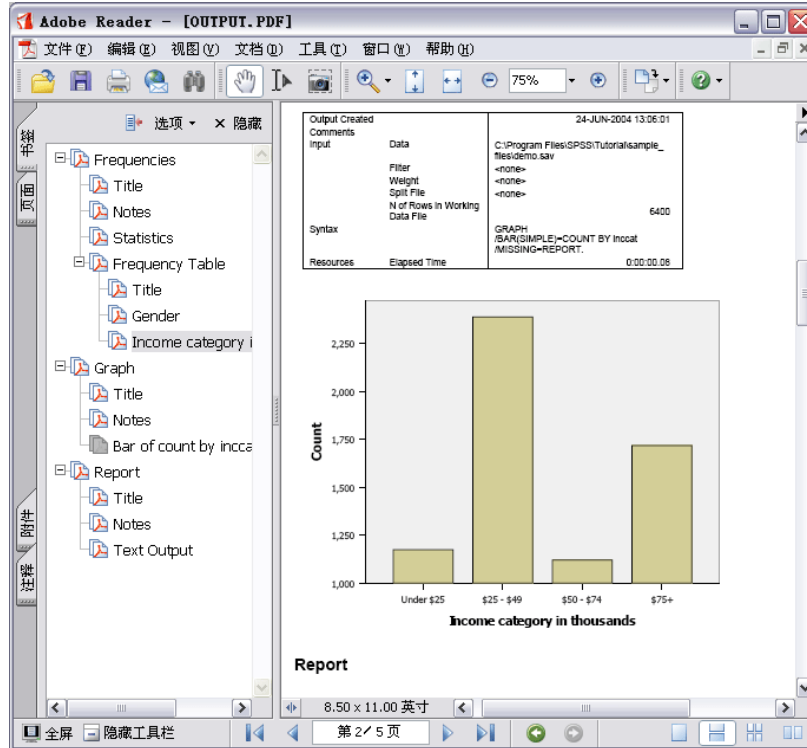
图片 7-35
“导出输出”对话框



- 查看器文档的概要窗格将转换为 PDF 文件中的书签，以便于导航。

- PDF 文档中的页面大小、方向、边距、页眉和页脚的内容和显示以及打印的图表大小都由“页面设置”选项控制（“查看器”窗口中“文件”菜单下的“页面设置”）。
- PDF 文档的分辨率（DPI）为默认打印机或当前选定的打印机的当前分辨率设置（可以使用“页面设置”进行更改）。最高分辨率为 1200 DPI。如果打印机的设置更高，则 PDF 文档的分辨率将为 1200 DPI。注意：在低分辨率的打印机上打印高分辨率文档时，输出结果的质量可能会比较差。

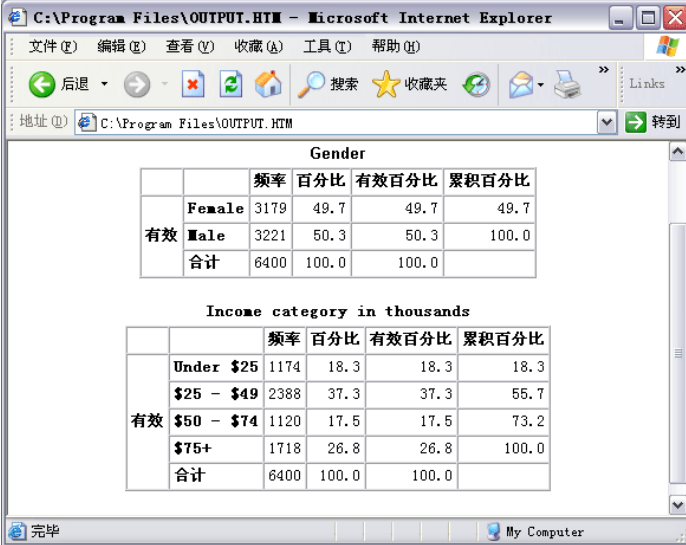
图片 7-36
带书签的 PDF 文件



将结果导出到 HTML

也可以将结果导出到 HTML（超文本标记语言）。在另存为 HTML 时，所有非图形输出均会导出到单个 HTML 文件中。

图片 7-37
Web 浏览器中的 Output.htm



The screenshot shows a Microsoft Internet Explorer window displaying two tables. The first table is titled 'Gender' and the second is titled 'Income category in thousands'. Both tables have columns for 'Frequency', 'Percentage', 'Effective Percentage', and 'Cumulative Percentage'. The 'Gender' table includes rows for 'Female', 'Male', and 'Total'. The 'Income category in thousands' table includes rows for 'Under \$25', '\$25 - \$49', '\$50 - \$74', '\$75+', and 'Total'.

	频率	百分比	有效百分比	累积百分比
	Female	3179	49.7	49.7
有效	Male	3221	50.3	100.0
	合计	6400	100.0	

	频率	百分比	有效百分比	累积百分比
	Under \$25	1174	18.3	18.3
	\$25 - \$49	2388	37.3	55.7
有效	\$50 - \$74	1120	17.5	73.2
	\$75+	1718	26.8	100.0
	合计	6400	100.0	

导出到 HTML 时，也可以导出图表，但不能导出到单个文件中。

每个图表都将以指定的格式另存为一个图形文件，而 HTML 中则放置指向这些图形文件的引用。还存在一个选项用于将所有图表（或选定的图表）导出到单独的图形文件中。

使用语法

通过使用强大的命令语言，您可以保存并自动执行许多常规任务。它还提供一些在菜单和对话框中没有的功能。大多数命令可以从菜单和对话框访问。但是，某些命令和选项只能通过命令语言使用。命令语言还允许您将作业保存在语法文件中，以便将来可以重复进行分析。

命令语法文件就是一个包含IBM® SPSS® Statistics语法命令的文本文件。您可以直接打开语法窗口并键入命令，但使用该文件通常可以更方便地使对话框为您完成部分或全部工作。

本章中的示例使用数据文件 demo.sav。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。

注意：学生版未提供命令语法。

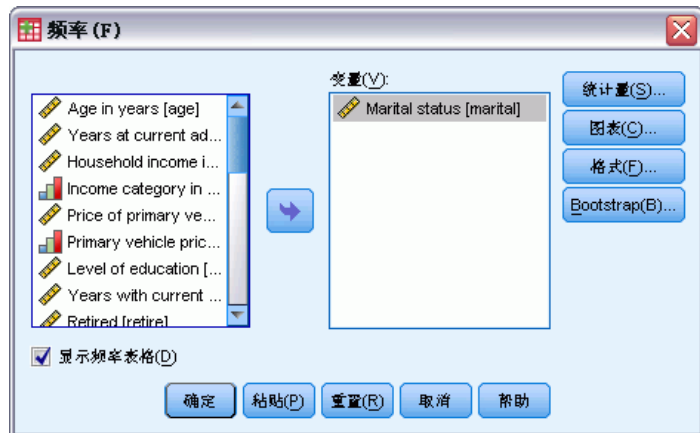
粘贴语法

创建语法最简便的方法是使用可在大多数对话框中找到的“粘贴”按钮。

- ▶ 打开数据文件 demo.sav。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。
- ▶ 从菜单中选择：
分析 > 描述统计 > 频率...

此时，将打开“频率”对话框。

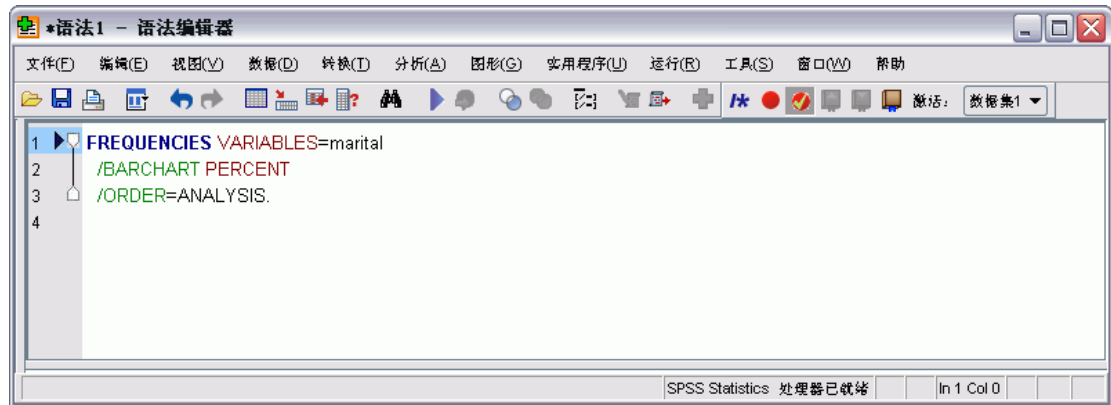
图片 8-1
“频率”对话框



- ▶ 选择婚姻状况 [marital]并将其移到“变量”列表中。
- ▶ 单击图表。
- ▶ 在“图表”对话框中，选择条形图。

- ▶ 在“图表值”组中，选择百分比。
- ▶ 单击继续。
- ▶ 单击粘贴将作为对话框选择结果创建的语法复制到语法编辑器。

图片 8-2
频率语法

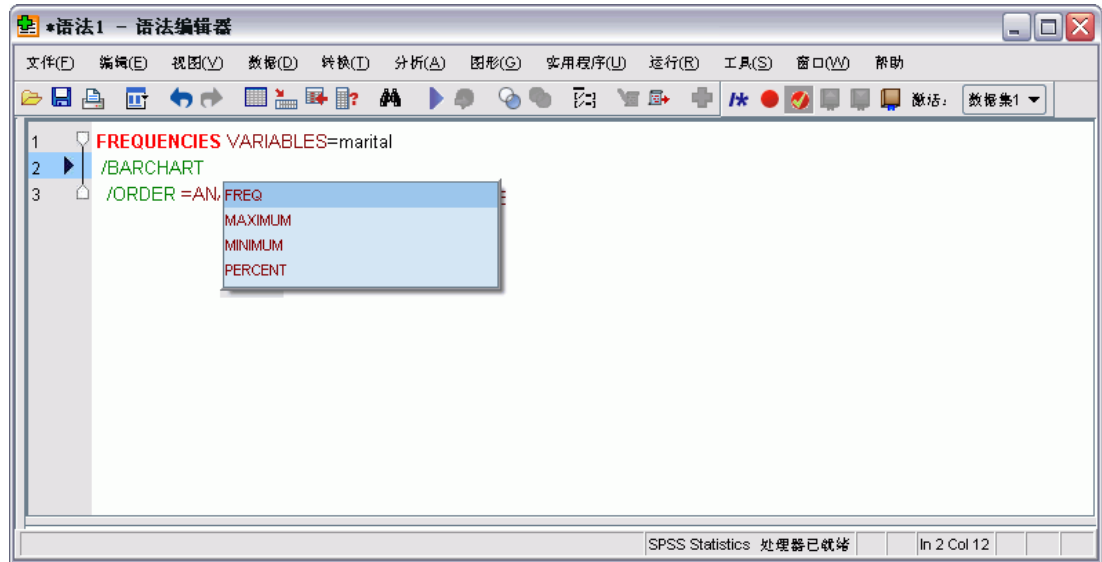


- ▶ 要运行当前显示的语法，请从菜单中选择：
运行 > 选择

编辑语法

在语法窗口中，可以编辑语法。例如，可以将子命令 `/BARCHART` 更改为显示频率而不是百分比。（子命令以斜杠表示。）如果您知道显示频率的关键字，您可以直接输入。如果您不知道关键字，您可以通过将光标置于子命令名后的任意位置并按“Ctrl+空格键”获得一个子命令的可用关键字列表。这为子命令显示自动完成控件。

图片 8-3
显示关键字的自动完成控件



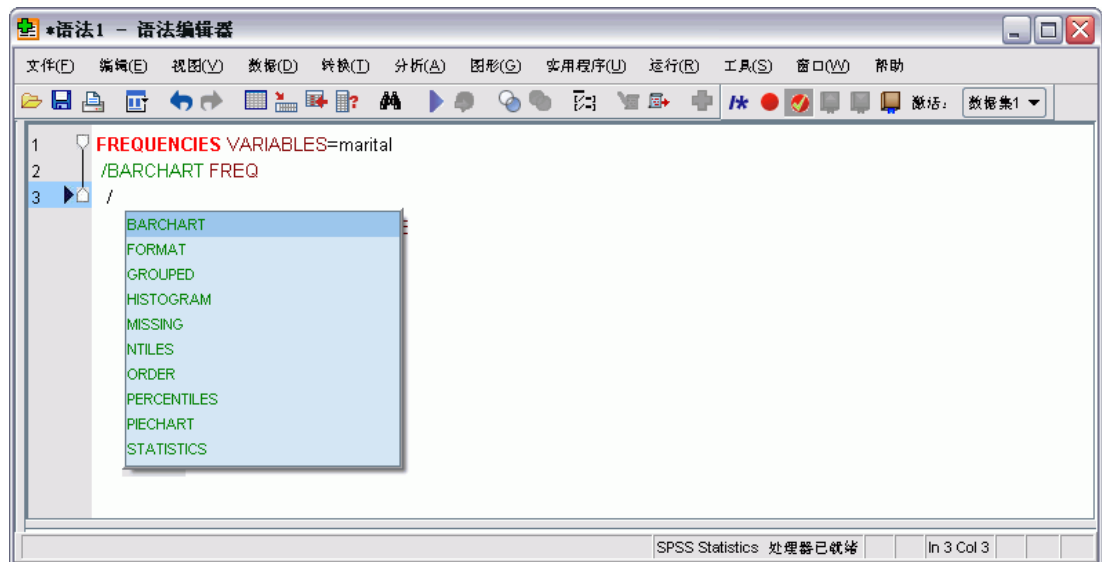
- ▶ 为频率单击标记为 **FREQ** 的项。单击自动完成控件中的项将其插入当前光标位置（原始 **PERCENT** 关键字已手动删除）。

缺省情况下，随着您的键入，自动完成控件将提示您一个可用项列表。例如，您想包括一个饼图以及条形图。使用另外的子命令指定饼图。

- ▶ 在 **FREQ** 关键字后按回车键，并键入一个正斜杠以表示子命令的开始。

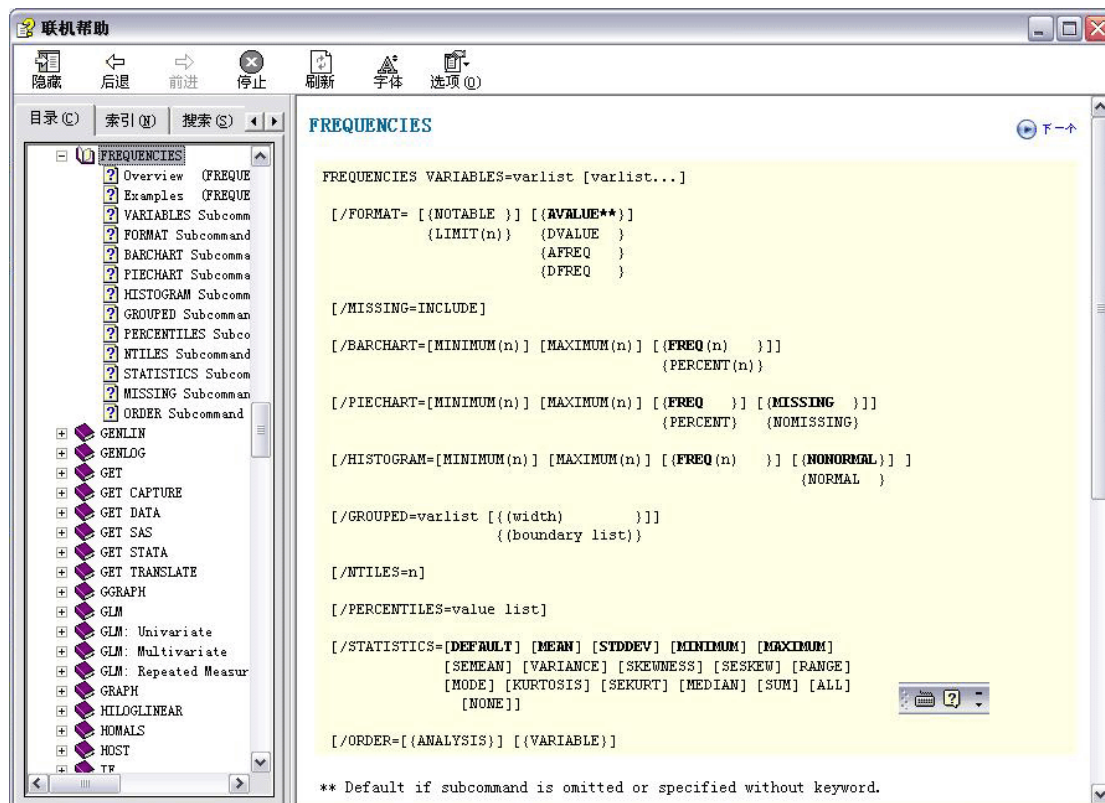
语法编辑器为当前命令提示您一个子命令列表。

图片 8-4
显示子命令的自动完成控件



要获得有关当前命令的更多详细帮助，按 F1 键。该操作会直接带您进入当前命令的命令语法参考信息。

图片 8-5
FREQUENCIES 语法帮助



您可能发现语法窗口中显示的文本显示了颜色。颜色编码允许您快速识别未识别的项，因为只有识别的项才显示颜色。例如，您将 `FORMAT` 子命令拼错为 `FRMAT`。缺省情况下，子命令显示绿色，但是文本 `FRMAT` 将显示无色，因为它未被识别。

打开并运行语法文件

- ▶ 要打开已保存的语法文件，请从菜单中选择：
文件 > 打开 > 语法...

将显示打开文件的标准对话框。

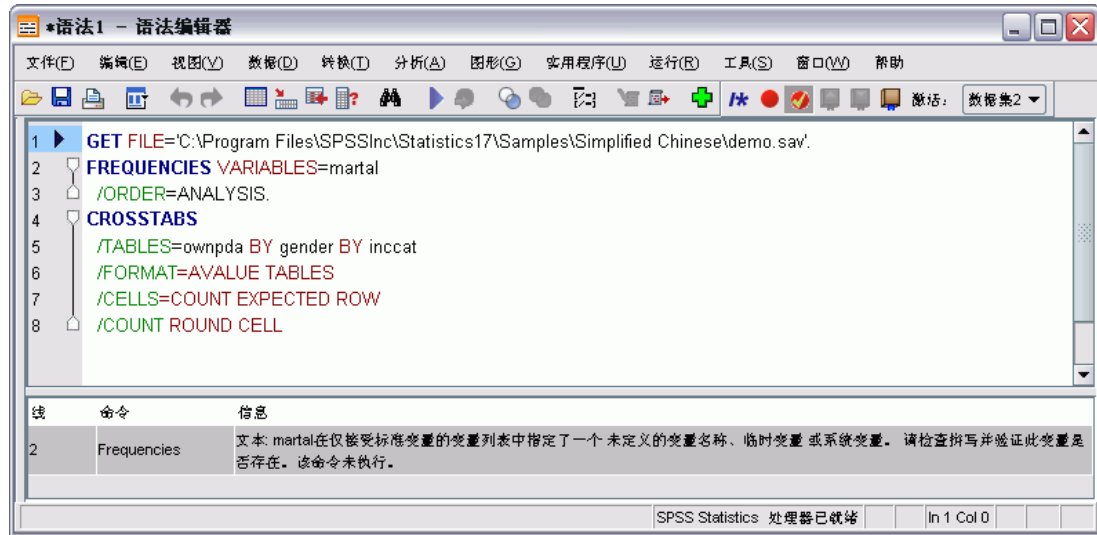
- ▶ 选择语法文件。如果未显示任何语法文件，请确保选择了语法 (*.sps) 作为您要查看的文件类型。
- ▶ 单击打开。
- ▶ 使用语法窗口中的“运行”菜单运行命令。

如果命令适用于特定的数据文件，则必须在运行命令之前打开该数据文件，或者必须包含打开该数据文件的命令。您可以从打开数据文件的对话框中粘贴此类型的命令。

了解错误窗格

错误窗格显示最当前的运行中的运行时间错误。包含每个错误的详细信息以及错误发生的命令行号。

图片 8-6
错误窗格显示在语法编辑器中



单击错误的条目将光标置于错误发生的命令第一行上。

使用分界点

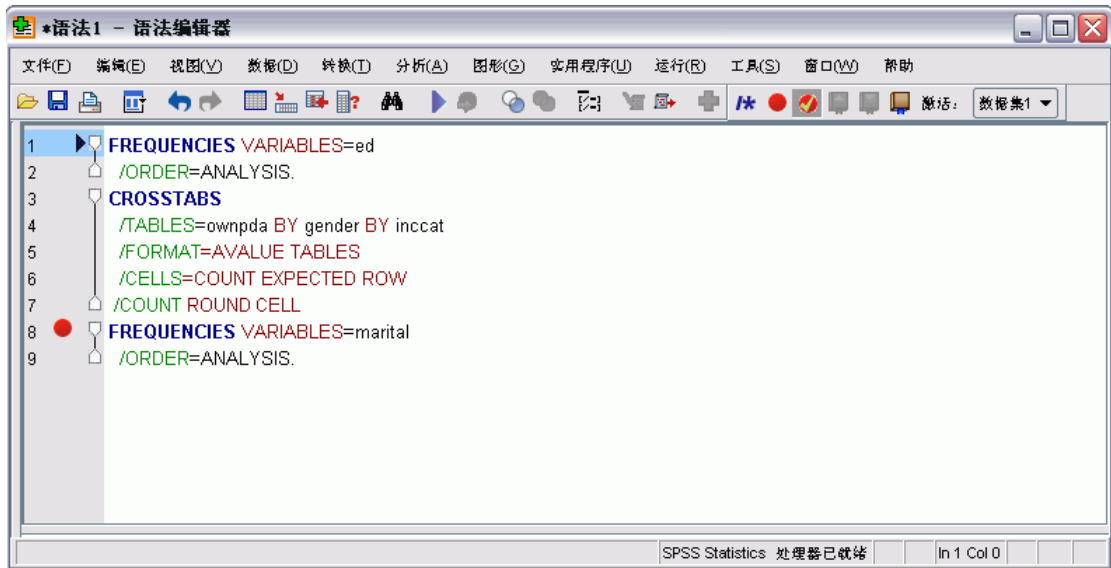
分界点允许您在语法中的指定点停止执行命令语法，并在准备好时继续执行。这允许您在语法作业中的中间某点查看输出或数据，或运行显示有关当前数据状态的信息的命令语法，如 **FREQUENCIES**。分界点只能设在命令级别，而非命令中的特定行上。

在命令上插入一个分界点：

- ▶ 单击与命令相关联的文本左侧区域的任意位置。

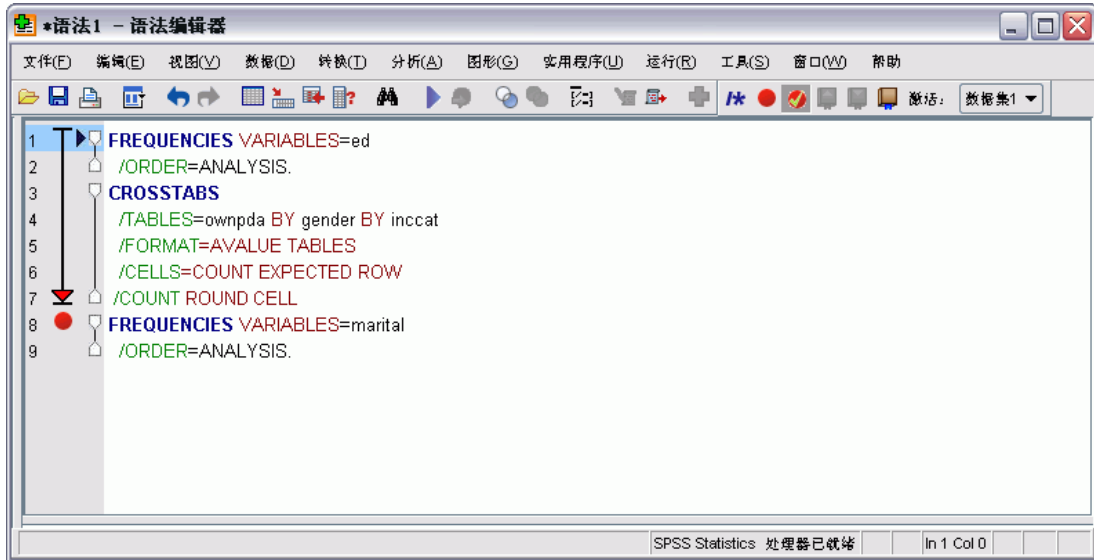
分界点在命令文本左侧区域中以及与命令名称相同的行上表示为红圈，无论您单击哪里。

图片 8-7
分界点显示在语法编辑器窗口中



当您运行包含分界点的命令语法，在每个包含分界点的命令之前执行停止。

图片 8-8
执行在分界点处停止



命令文本左侧的向下箭头显示语法运行的进度。从第一个命令运行跨到最后一个命令运行，当运行包含分界点的命令语法时尤其有用。

分界点后恢复执行：

- ▶ 从语法编辑器中的菜单选择：
运行 > 继续

修改数据值

您开始时使用的数据可能不会始终按照对您的分析或报告需求最有帮助的方式进行排列。例如，您可能想要：

- 从刻度变量创建分类变量。
- 将多个响应类别合并为单个类别。
- 创建两个现有变量之间的计算差异的新变量。
- 计算两个日期之间的时间长度。

本章使用数据文件 `demo.sav`。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。

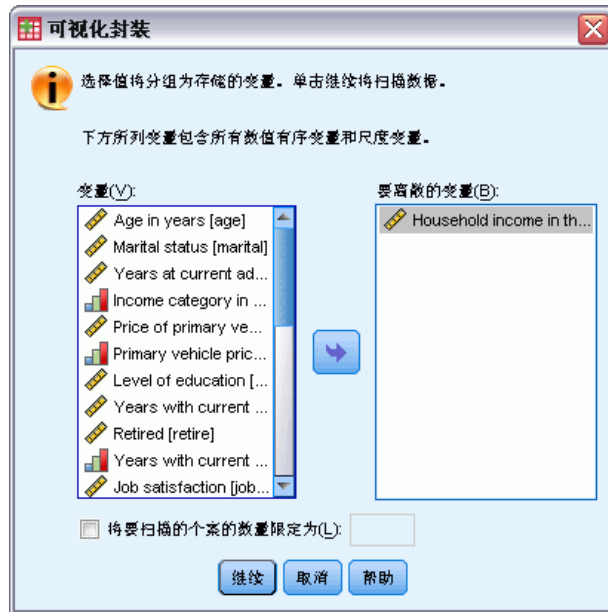
从刻度变量创建分类变量

实际上，数据文件 `demo.sav` 中的多个分类变量都是从该数据文件中的刻度变量派生的。例如，变量 `inccat` 仅是分组为四个类别的 `income`。此分类变量使用整数值 1 - 4 表示以下收入分类（以千为单位）：小于 \$25、\$25 - \$49、\$50 - \$74 以及 \$75 或更高。

要创建分类变量 `inccat`：

- ▶ 从数据编辑器窗口的菜单中选择：
转换 > 可视离散化...

图片 9-1
初始的“可视离散化”对话框

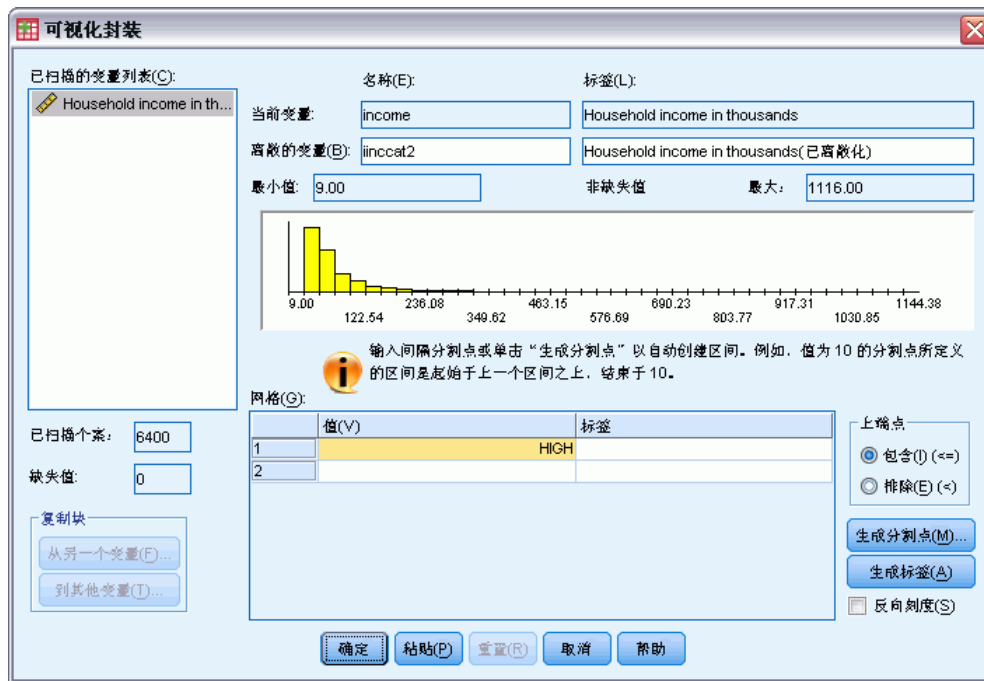


在初始的“可视离散化”对话框中，选择要为其创建新的离散化变量的刻度变量和/或有序变量。**离散化**是指取两个或多个连续值并将其分组为同一类别。

由于可视离散化依赖于数据文件中的实际值以帮助您做出良好的离散化选择，因而其需要先读取数据文件。如果您的数据文件包含大量个案，则完成此过程将需要一段时间，因此，这一初始对话框还允许您限制要读取（“扫描”）的个案数。我们的样本数据文件不需要此限制。尽管此数据文件包含 6,000 多个案，但扫描这些个案不需要太长时间。

- ▶ 将 Household income in thousands [income] 从“变量”列表中拖放到“要离散的变量”列表中，然后单击继续。

图片 9-2
“可视离散化”主对话框



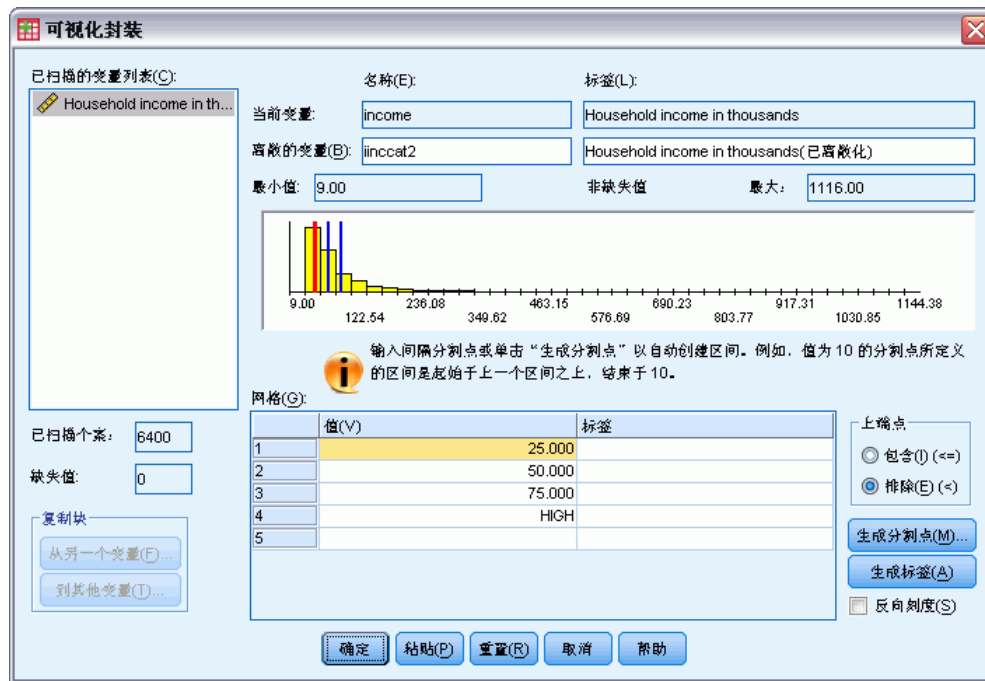
- ▶ 在“可视离散化”主对话框中，选择“已扫描的变量列表”中的 Household income in thousands [income]。
直方图显示了所选变量的分布（在此例中，分布严重偏斜）。
- ▶ 输入 inccat2 作为新的离散化变量名称，输入 Income category [in thousands] 作为变量标签。
- ▶ 单击生成分割点。

图片 9-3
“可视离散化分割点”对话框



- ▶ 选择等宽度间隔。
- ▶ 输入 25 作为第一个分割点的位置，输入 3 作为分割点数量，并输入 25 作为宽度。
离散化类别的数量比分割点数量多一个。因此在本示例中，新的离散化变量将具有四个类别，其中前三个类别中每个包含的范围为 25（千），最后一个类别包含最高分割点值 75（千）以上的所有值。
- ▶ 单击应用。

图片 9-4
包含所定义分割点的“可视离散化”主对话框

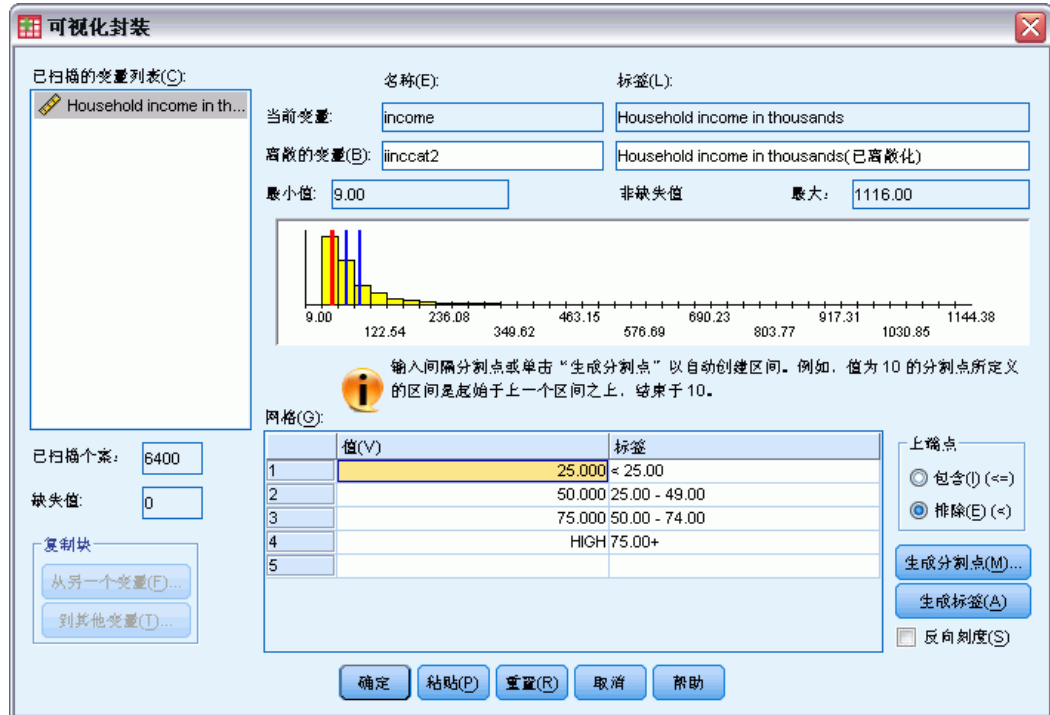


网格中当前显示的值表示所定义的分割点，这些分割点是每个类别的上端点。直方图中的垂直线还指示了各分割点的位置。

缺省情况下，这些分割点值将包含在相应的类别中。例如，第一个值 25 将包含所有小于或等于 25 的值。但在本示例中，我们希望这些类别对应于小于 25、25 - 49、50 - 74 以及 75 或更高。

- ▶ 在上端点组中，选择排除(<)。
- ▶ 然后单击生成标签。

图片 9-5
自动生成的值标签



这将自动为每个类别生成描述性值标签。由于为新的离散化变量指定的实际值只是从 1 开始的连续整数，因此这些值标签可能非常有用。

还可以在网格中手动输入或更改分割点和标签，通过在直方图中拖放分割点线来更改分割点位置以及通过将分割点线拖出直方图来删除分割点。

- ▶ 单击确定以创建新的离散化变量。

新变量将显示在数据编辑器中。由于该变量将添加到文件的末尾，因此显示在数据视图的最右侧一列，变量视图的最后一行中。

图片 9-6
数据编辑器中显示的新变量



计算新变量

使用各种数学函数，您可以基于高度复杂的方程来计算新变量。但是，在本示例中，我们将仅计算两个现有变量值之间差异的新变量。

数据文件 demo.sav 包含一个表示响应者当前年龄的变量，以及一个表示其在当前工作中年数的变量。但是，该文件中不包含表示响应者开始该工作时的年龄的变量。我们可以创建一个当前年龄与其在当前工作中年数之间的计算差异的新变量，该变量应该就是响应者开始该工作时的大约年龄。

- ▶ 从数据编辑器窗口的菜单中选择：
转换 > 计算变量...
- ▶ 对于“目标变量”，输入 jobstart。
- ▶ 选择源变量列表中的 Age in years [age] 并单击箭头按钮以将其复制到“数字表达式”文本框中。
- ▶ 单击该对话框中计算器面板上的减号 (-) 按钮（或按键盘上的减号键）。

- ▶ 选择 Years with current employer [employ] 并单击箭头按钮以将其复制到表达式。

图片 9-7
“计算变量”对话框



注意：请务必选择正确的雇用变量。该变量还存在一个经过重新编码的分类版本，它不是您所需的变量。数字表达式应为 `age - employ`，而不是 `age - empcat`。

- ▶ 单击确定以计算新变量。

新变量将显示在数据编辑器中。由于该变量将添加到文件的末尾，因此显示在数据视图的最右侧一列，变量视图的最后一行中。

图片 9-8
数据编辑器中显示的新变量

	ownfax	news	response	VAR00001	jobstart	变量	变量
1	No	Yes	No	\$50 - \$74	52.00		
2	No	Yes	Yes	\$75+	53.00		
3	No	No	No	\$25 - \$49	27.00		
4	Yes	No	No	\$25 - \$49	23.00		
5	No	No	No	Under \$25	23.00		
6	No	Yes	No	\$75+	43.00		
7	No	Yes	No	\$25 - \$49	40.00		
8	No	Yes	No	\$50 - \$74	34.00		
9	No	No	No	Under \$25	44.00		
10	Yes	No	Yes	\$75+	32.00		
11	No	No	No	\$50 - \$74	54.00		

使用表达式中的函数

还可以在表达式中使用预定义的函数。您可以使用 70 多种内置函数：

- 算术函数
- 统计函数
- 分布函数
- 逻辑函数
- 日期和时间聚合函数以及提取函数
- 缺失值函数

- 交叉个案函数
- 字符串函数

图片 9-9
显示函数分组的“计算变量”对话框



函数可以归类为逻辑上不同的各组，例如，一组用于算术运算，另一组计算统计度量。为方便起见，诸如 \$TIME（当前日期和时间）之类的许多常用系统变量也包含在相应的函数组中。“计算变量”对话框的保留区域中显示了对当前所选函数（在此例中为 SUM）或系统变量的简要描述。

将函数粘贴到表达式中

要将函数粘贴到表达式中：

- ▶ 将光标定位到您希望函数在表达式中出现的位置。
- ▶ 从“函数组”列表中选择适当的组。标记为全部的组提供了所有可用函数和系统变量的列表。
- ▶ 双击“函数和特殊变量”列表中的函数（或选择函数并单击“函数组”列表旁边的箭头）。

该数将被插入到表达式中。如果突出显示表达式的一部分，然后插入函数，则突出显示的函数部分将用作该函数中的第一个参数。

编辑表达式中的函数

只有在粘贴的函数中输入由问号表示的参数后，才会完成该函数。问号的数目指示完成该函数所需的最少参数个数。

- ▶ 在粘贴的函数中突出显示问号。
- ▶ 输入参数。如果参数为变量名称，则可以从变量列表中粘贴这些参数。

使用条件表达式

可以使用条件表达式（也称为逻辑表达式）将转换应用于所选的个案子集。条件表达式将为每个个案返回一个值：True、False 或缺失。如果条件表达式的结果为 True，则转换将应用于该个案。如果结果为 False 或缺失，则转换将不应用于该个案。

要指定条件表达式：

- ▶ 单击“计算变量”对话框中的**如果**。这将打开“If 个案”对话框。

图片 9-10
“If 个案”对话框



- ▶ 选中**如果个案满足条件则包含**。
- ▶ 输入条件表达式。

大多数条件表达式都至少包含一个关系运算符，例如：

```
age>=21
```

或

```
income*3<100
```

在第一个示例中，将仅选择 Age [age]值等于或大于 21 的个案。在第二个示例中，对于要选择的个案，Household income in thousands [income]乘以 3 必须小于 100。

还可以使用逻辑运算符来链接两个或多个条件表达式，例如：

```
age>=21 | ed>=4
```

或

```
income*3<100 & ed=5
```

在第一个示例中，将选择满足 Age [age] 条件或 Level of education [ed] 条件的个案。在第二个示例中，对于要选择的个案，必须同时满足 Household income in thousands [income] 和 Level of education [ed] 条件。

使用日期和时间

通常使用日期和时间来执行的许多任务都可以使用日期和时间向导轻松完成。使用此向导，您可以：

- 根据包含日期或时间的字符串变量创建日期/时间变量。
- 通过合并包含日期或时间的不同部分的变量来构造日期/时间变量。
- 从日期/时间变量中加减值，包括加上或减去两个日期/时间变量。
- 提取日期或时间变量的一部分，例如，从 mm/dd/yyyy 格式的日期/时间变量中提取该月中的某天。

本节中的示例使用数据文件 upgrade.sav。 [有关详细信息，请参阅第 153 页码附录 A 中的样本文件。](#)

要使用日期和时间向导：

- ▶ 从菜单中选择：
转换 > 日期和时间向导...

图片 9-11
日期和时间向导简介屏幕



日期和时间向导的介绍屏幕为您显示了一组常规任务。对当前数据不适用的任务将处于禁用状态。例如，数据文件 `upgrade.sav` 不包含任何字符串变量，因此从字符串创建日期变量的任务将处于禁用状态。

如果对 IBM® SPSS® Statistics 中的日期和时间不熟悉，则可以选择学习日期和时间的表示方式，然后单击下一步。这样将打开一个屏幕，其中提供了日期/时间变量的简要概述以及通过“帮助”按钮了解更多详细信息的链接。

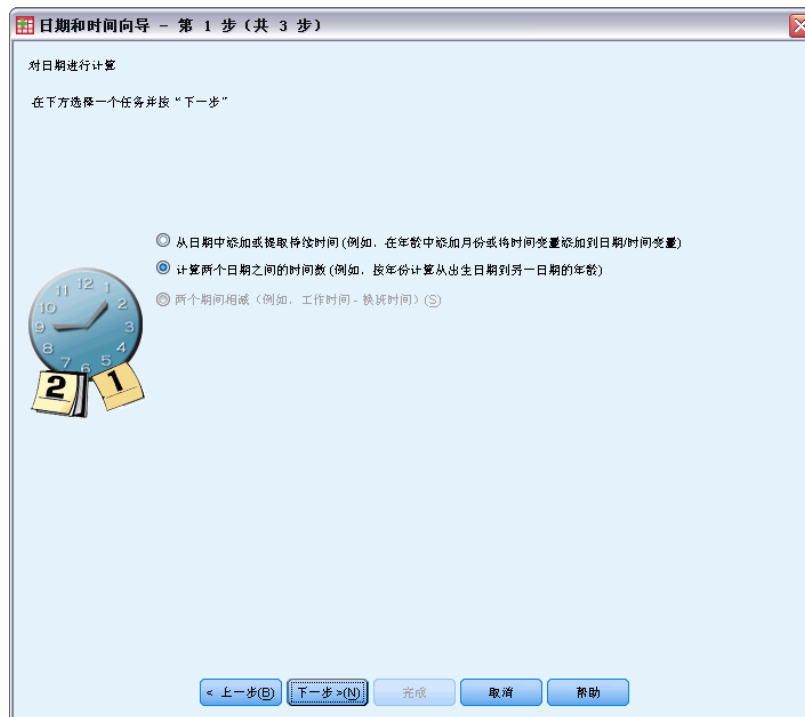
计算两个日期之间的时间长度

涉及日期的常见任务之一是计算两个日期之间的时间长度。例如，一家软件公司对通过确定每个客户上次购买升级之后的年数来分析升级许可证的购买很感兴趣。数据文件 `upgrade.sav` 包含表示每个客户上次购买升级日期的变量，但不包含表示该购买之后年数的变量。可表示上次升级日期与下个产品发布日期之间的时间长度（以年为单位）的新变量将提供对此数量的测量。

要计算两个日期之间的时间长度：

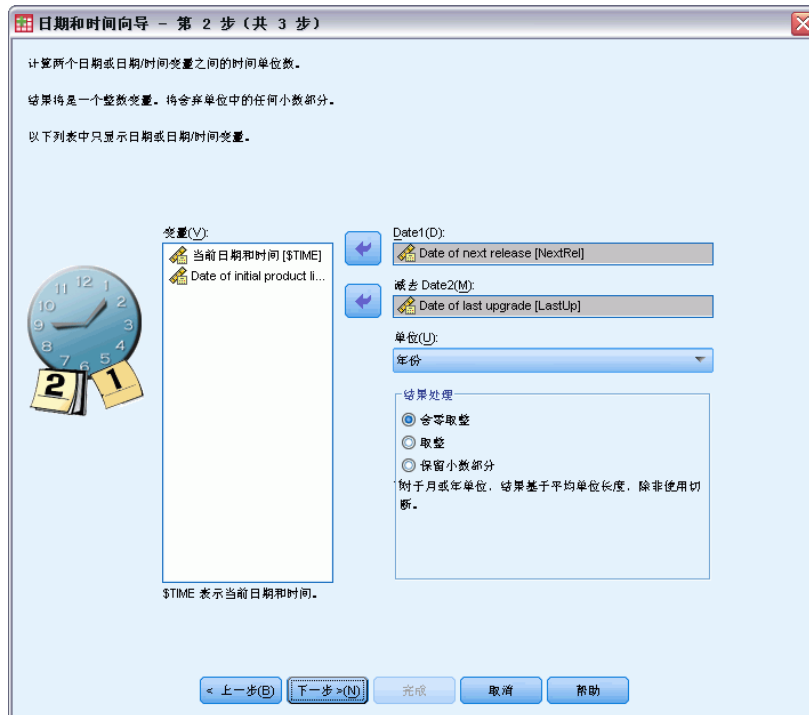
- ▶ 在日期和时间向导的介绍屏幕上选择使用日期和时间进行计算，然后单击下一步。

图片 9-12
计算两个日期之间的时间长度：第 1 步



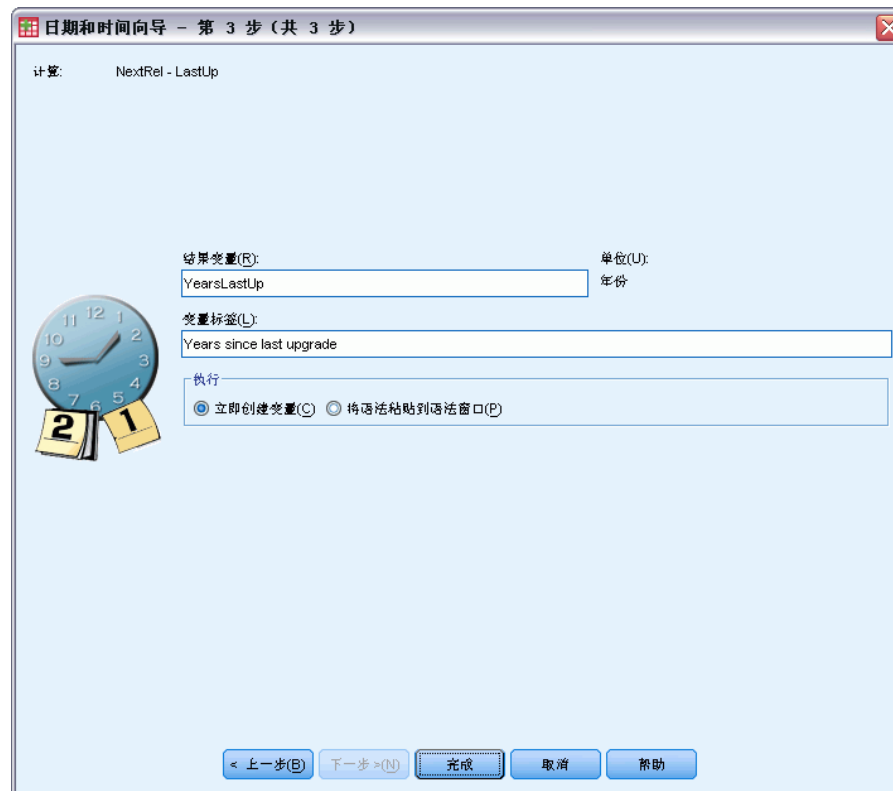
- ▶ 选择计算两个日期之间的时间单位数，然后单击下一步。

图片 9-13
计算两个日期之间的时间长度：第 2 步



- ▶ 选择 Date of next release 作为 Date1。
- ▶ 选择 Date of last upgrade 作为 Date2。
- ▶ 为“单位”选择年份，为“结果处理”选择舍零取整。（这些是缺省选择。）
- ▶ 单击下一步。

图片 9-14
计算两个日期之间的时间长度：第 3 步



- ▶ 输入 YearsLastUp 作为结果变量的名称。结果变量不能与现有变量同名。
- ▶ 输入 Years since last upgrade 作为结果变量的标签。结果变量的变量标签是可选的。
- ▶ 保留缺省选项立即创建变量，然后单击完成以创建新变量。

数据编辑器中显示的新变量 YearsLastUp 是两个日期之间的整数年数。年份的小数部分已被截去。

图片 9-15
数据编辑器中显示的新变量

	PurDate	Support	LastUp	NextRel	YearsLastUp	变量
1	12/30/1998	4	02/28/2002	06/01/2004	2	
2	06/28/2001	2	09/28/2002	06/01/2004	1	
3	08/27/1999	2	09/27/2001	06/01/2004	2	
4	02/22/2000	4	01/22/2003	06/01/2004	1	
5	01/26/2000	2	08/26/2001	06/01/2004	2	
6	07/10/1999	3	07/10/2003	06/01/2004	0	
7	01/24/2003	2	07/24/2003	06/01/2004	0	
8	06/15/1999	2	09/15/2003	06/01/2004	0	
9	01/18/2003	5	07/18/2003	06/01/2004	0	
10	12/02/2000	4	08/02/2003	06/01/2004	0	

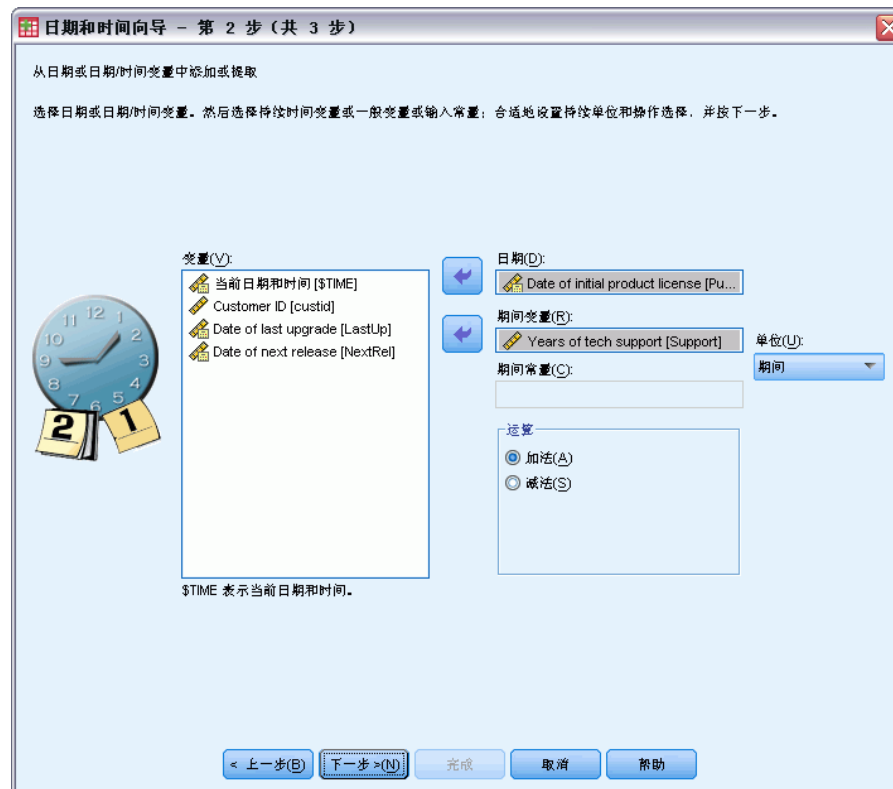
向日期添加期间

您可以从日期加上或减去期间，例如 10 天或 12 个月。我们继续使用先前部分中的软件公司示例，假设要确定每个客户的初始技术支持合同结束的日期。数据文件 upgrade.sav 包含一个表示约定的支持年数的变量，以及一个表示初始购买日期的变量。随后可以通过向购买日期添加支持年数来确定初始支持的结束日期。

要向日期添加期间：

- ▶ 在日期和时间向导的介绍屏幕上选择使用日期和时间进行计算，然后单击下一步。
- ▶ 选择日期加上或减去期间，然后单击下一步。

图片 9-16
向日期添加期间：第 2 步



► 选择 Date of initial product license 作为日期。

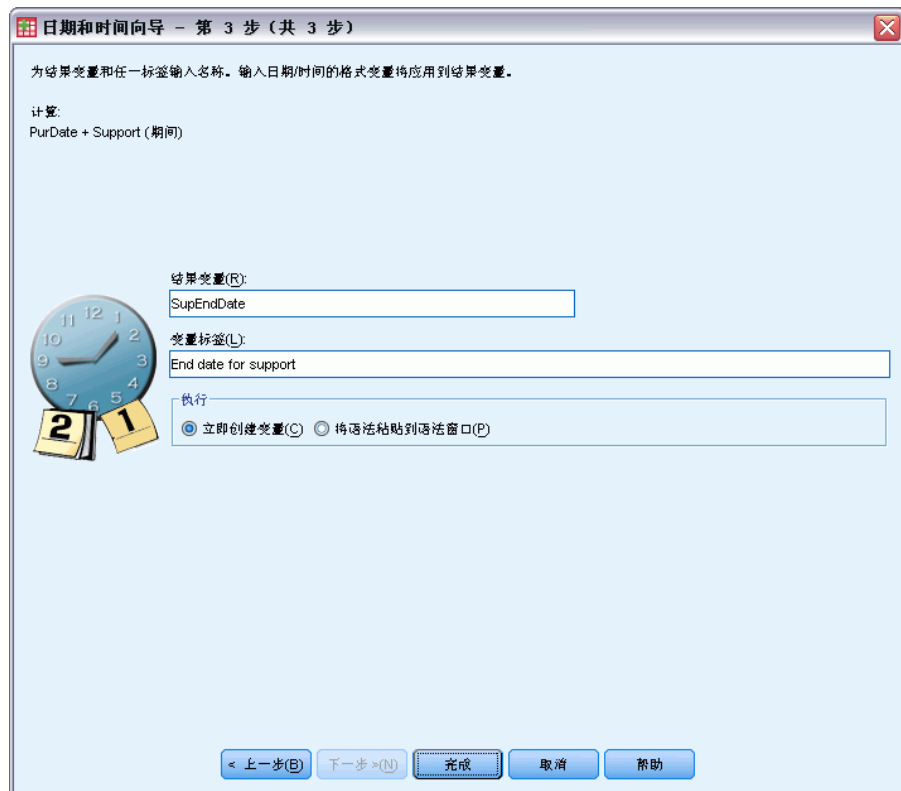
► 选择 Years of tech support 作为期间变量。

由于 Years of tech support 仅是一个数值变量，因此在添加此变量作为期间时，需要指示要使用的单位。

► 从“单位”下拉列表中选择年份。

► 单击下一步。

图片 9-17
向日期添加期间：第 3 步



- ▶ 输入 SupEndDate 作为结果变量的名称。结果变量不能与现有变量同名。
- ▶ 输入 End date for support 作为结果变量的标签。结果变量的变量标签是可选的。
- ▶ 单击完成以创建新变量。

新变量将显示在数据编辑器中。

图片 9-18
数据编辑器中显示的新变量



	Support	LastUp	NextRel	YearsLastUp	SupEndDate	变
1	4	02/28/2002	06/01/2004	2	12/30/2002	
2	2	09/28/2002	06/01/2004	1	06/28/2003	
3	2	09/27/2001	06/01/2004	2	08/27/2003	
4	4	01/22/2003	06/01/2004	1	02/22/2004	
5	2	08/26/2001	06/01/2004	2	01/26/2004	
6	3	07/10/2003	06/01/2004	0	07/10/2002	
7	2	07/24/2003	06/01/2004	0	01/24/2005	
8	2	09/15/2003	06/01/2004	0	06/15/2001	
9	5	07/18/2003	06/01/2004	0	01/18/2008	

排序和选择数据

数据文件的组织形式对于特定需要而言并非总是理想的。在准备要分析的数据时，可从大量文件转换中选择，包括执行以下操作的能力：

- **对数据排序。**您可根据一个或多个变量的值对个案排序。
- **选择个案的子集。**您可以将分析限制为个案的子集或对不同的子集执行同步分析。

本章中的示例使用数据文件 demo.sav。有关详细信息，请参阅第 153 页码附录 A 中的样本文件。

排序数据

对个案进行排序（排序数据文件的行）通常很有用，且有时对于某些类型的分析是十分必要的。

根据一个或多个排序变量的值重新排序数据文件中个案的顺序：

- ▶ 从菜单中选择：
数据 > 对个案排序...

“个案排序”对话框随即显示。

图片 10-1
“个案排序”对话框



- ▶ 将 Age in years [age] 和 Household income in thousands [income] 变量添加到“排序方式”列表中。

如果选择多个排序变量，则它们在“排序方式”列表中的显示顺序将决定个案的排序顺序。本例中，根据“排序方式”列表中的项，个案将按照 Age in years [age] 类别中的 Household income in thousands [income] 值进行排序。对于字符串变量，大写字母在排序顺序中排在其对应小写字母前面（例如，字符串值 Yes 在排序顺序中显示在 yes 之前）。

拆分文件处理

将数据文件拆分成多个不同的组以进行分析：

- ▶ 从菜单中选择：
数据 > 拆分文件...

“拆分文件”对话框随即显示。

图片 10-2
“拆分文件”对话框



- ▶ 选择比较各组或按组来组织输出。（这些步骤后面的示例显示了这两个选项之间的区别。）
- ▶ 选择 Gender [gender] 来针对这些变量将文件拆分成不同的组。

可使用数值、短字符串和长字符串变量作为分组变量。对由分组变量定义的每个子组进行单独的分析。如果选择了多个分组变量，则它们在“分组依据”列表中的显示顺序将决定个案的分组方式。

如果选择比较各组，则所有拆分文件组的结果将包含在同一个表（或多个表）中，如下面的由“频率”过程生成的摘要统计表所示。

图片 10-3
带有单个枢轴表的拆分文件输出

统计量			
Household income in thousands			
Female	N	有效	3179
		缺失	0
	均值		68.7798
	中位数		44.0000
	标准差		75.73510
Male	N	有效	3221
		缺失	0
	均值		70.1608
	中位数		45.0000
	标准差		81.56216

如果选择按组来组织输出并运行“频率”过程，则会创建两个枢轴表：一个表针对女性，另一个表针对男性。

图片 10-4
带有女性枢轴表的拆分文件输出

统计量 ^a		
Household income in thousands		
N	有效	3179
	缺失	0
均值		68.7798
中位数		44.0000
标准差		75.73510

a. Gender = Female

图片 10-5
带有男性枢轴表的拆分文件输出

统计量 ^a		
Household income in thousands		
N	有效	3221
	缺失	0
均值		70.1608
中位数		45.0000
标准差		81.56216

a. Gender = Male

为拆分文件处理排序个案

“拆分文件”过程在每次遇到某个分组变量的不同的值时将创建一个新的子组。因此，在调用拆分文件处理之前，根据分组变量的值排序个案非常重要。

缺省情况下，“拆分文件”自动根据分组变量的值对数据文件进行排序。如果文件已经以正确的顺序排序，那么选择已对文件排序可节约处理时间。

打开和关闭拆分文件处理

调用拆分文件处理之后，除非将其关闭，否则它将在余下的会话中继续保持有效。

- **分析所有个案。**此选项关闭拆分文件处理。
- **比较各组**和**按组来组织输出。**此选项打开拆分文件处理。

如果拆分文件处理有效，则消息**拆分文件范围**将显示在应用程序窗口底部的状态栏上。

选择个案的子集

可基于包含变量和复杂表达式的条件将分析限制到某个特定子组。您也可以选择个案的随机样本。用于定义子组的条件可包括：

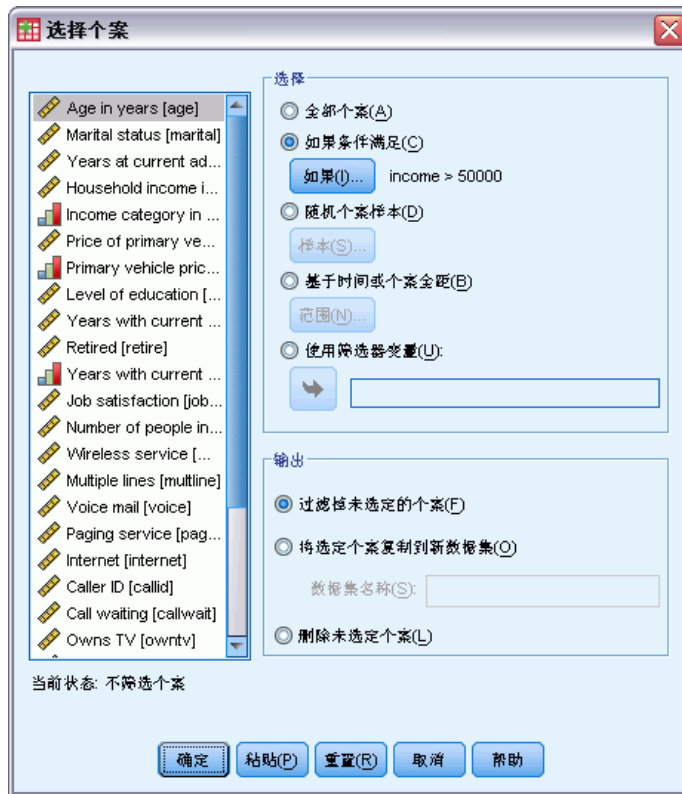
- 变量值和范围
- 日期和时间范围
- 个案（行）编号
- 算术表达式
- 逻辑表达式
- 函数

选择一个案子集进行分析：

- ▶ 从菜单中选择：
数据 > 选择个案...

这将打开“选择个案”对话框。

图片 10-6
“选择个案”对话框



基于条件表达式选择个案

基于条件表达式选择个案：

- ▶ 在“选择个案”对话框中选择如果满足条件，然后单击 If。
这将打开“选择个案：If”对话框。

图片 10-7
“选择个案：If”对话框



条件表达式可使用现有变量名称、常数、算术运算符、逻辑运算符、关系运算符和函数。可以在文本框中键入和编辑表达式，就像是处理输出窗口中的文本。还可以使用计算器键盘、变量列表和函数列表将元素粘贴到表达式中。有关详细信息，请参阅第 134 页码第 9 章中的使用条件表达式。

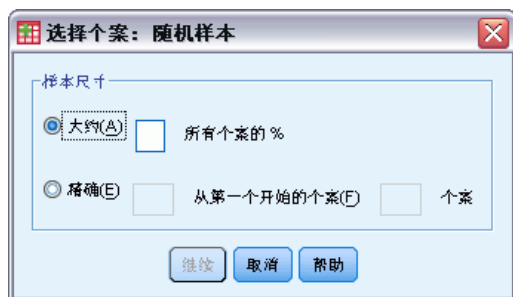
要选择随机样本请执行下列操作

获取随机样本：

- ▶ 在“选择个案”对话框中选择随机个案样本。
- ▶ 单击样本。

这将打开“选择个案：随机样本”对话框。

图片 10-8
“选择个案：随机样本”对话框



可为样本大小选择下列选项之一：

- **近似。**用户指定百分比。此选项生成近似于指定个案百分比的随机样本。
- **精确。**用户指定的个案数。您还必须指定用于生成样本的个案数。第 2 个数目应该小于或等于数据文件中个案的总数。如果该数字超过数据文件中的个案总数，则样本所包含的个案数将按请求的数目成比例减少。

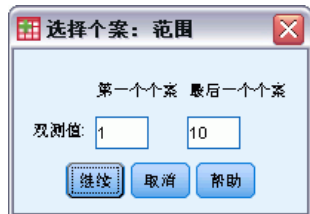
选择时间范围或个案范围

基于日期、时间或观察值（行）号选择一个个案范围：

- ▶ 在“选择个案”对话框中选择基于时间或个案范围，然后单击范围。

这将打开“选择个案：范围”对话框，可在其中选择观察值（行）号的范围。

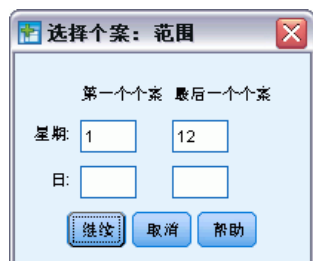
图片 10-9
“选择个案：范围”对话框



- **第一个个案。**输入范围的起始日期和/或时间值。如果未定义日期变量，则输入起始观察值号（即数据编辑器中的行号，除非“拆分文件”已打开）。如果不指定“最后一个个案”值，则从起始日期/时间到时间序列末尾这一范围内的所有个案都将被选中。
- **最后一个个案。**输入范围的结束日期和/或时间值。如果未定义日期变量，则输入末尾的观察值号（即数据编辑器中的行号，除非“拆分文件”已打开）。如果不指定“第一个个案”值，则从时间序列的开始直到结束日期/时间这一范围内的所有个案都将被选中。

对于已定义了日期变量的时间序列数据，可根据已定义的日期变量选择日期和/或时间的范围。每个个案表示不同时间的观察值，文件按时间顺序排序。

图片 10-10
“选择个案：范围”对话框（时间序列）



生成时间序列数据的日期变量：

- ▶ 从菜单中选择：
数据 > 定义日期...

未选定的个案的处理

可以选择以下可选项之一来处理未选定个案：

- **过滤掉未选定的个案。**未选定的个案不包括在分析中，但保留在数据集中。如果关闭过滤，则在会话中稍后可以使用未选定个案。如果选择随机样本，或者基于条件表达式选择个案，则此项生成名为 `filter_` 的变量，对选定个案该变量的值为 1，对未选定个案该变量的值为 0。
- **将选定的个案复制到新数据集。**选定的个案复制到新数据集，原始数据集未受影响。未选中个案不包括在新数据集中，而在初始数据集中保持其初始状态。
- **删除未选定个案。**从数据集删除未选定个案。只有退出文件而不保存任何更改，然后重新打开文件，才能恢复删除的个案。如果保存对数据文件的更改，则会永久删除个案。

注意：如果删除未选定个案并保存文件，则无法恢复个案。

个案选择状态

如果您选择了一个个案子集，但尚未抛弃未选择的个案，则数据编辑器中用穿过行号的对角线来标记未选择的个案。

图片 10-11
个案选择状态

	age	marital	address	income	inccat	car
1	55	Married	12	72.00	\$50 - \$74	36.2
2	56	Unmarried	29	153.00	\$75+	76.9
3	28	Married	9	28.00	\$25 - \$49	13.7
4	24	Married	4	26.00	\$25 - \$49	12.5
5	25	Unmarried	2	23.00	Under \$25	11.3
6	45	Married	9	76.00	\$75+	37.2
7	42	Unmarried	19	40.00	\$25 - \$49	19.8
8	35	Unmarried	15	57.00	\$50 - \$74	28.2
9	46	Unmarried	26	24.00	Under \$25	12.2

样本文件

随产品一起安装的样本文件可以在安装目录的 Samples 子目录中找到。对于以下每种语言在“样本”子目录中有单独的文件夹：英语、法语、德语、意大利语、日语、韩语、波兰语、俄语、简体中文、西班牙语和繁体中文。

并非所有样本文件均提供此处的全部语言版本。如果样本文件未提供某种语言的版本，则相应语言文件夹中包含该样本文件的英语版本。

描述

以下是对在整个文档的各种示例中使用的样本文件的简要描述。

- **accidents.sav**。该假设数据文件涉及某保险公司，该公司正在研究给定区域内汽车事故的年龄和性别风险因子。每个个案对应一个年龄类别和性别类别的交叉分类。
- **adl.sav**。该假设数据文件涉及在确定针对脑卒中患者的建议治疗类型的优点方面的举措。医师将女性脑卒中患者随机分配到两组中的一组。第一组患者接受标准的物理治疗，而第二组患者则接受附加的情绪治疗。在进行治疗的三个月时间里，将为每个患者进行一般日常生活行为的能力评分并作为原始变量。
- **advert.sav**。该假设数据文件涉及某零售商在检查广告支出与销售业绩之间的关系方面的举措。为此，他们收集了过去的销售数据以及相关的广告成本。
- **aflatoxin.sav**。该假设数据文件涉及对谷物的黄曲霉毒素的检测，该毒素的浓度会因谷物产量的不同（不同谷物之间及同种谷物之间）而有较大变化。谷物加工机从 8 个谷物产量的每一个中收到 16 个样本并以十亿分之几 (PPB) 为单位来测量黄曲霉毒素的水平。
- **anorectic.sav**。在研究厌食/暴食行为的标准症状参照时，研究人员对 55 名已知存在进食障碍的青少年进行了调查。其中每名患者每年都将进行四次检查，因此总观测数为 220。在每次观测期间，将对这些患者按 16 种症状逐项评分。但 71 号和 76 号患者的症状得分均在时间点 2 缺失，47 号患者的症状得分在时间点 3 缺失，因此有效观测数为 217。
- **bankloan.sav**。该假设数据文件涉及某银行在降低贷款拖欠率方面的举措。该文件包含 850 位过去和潜在客户的财务和人口统计信息。前 700 个个案是以前曾获得贷款的客户。剩下的 150 个个案是潜在客户，银行需要按高或低信用风险对他们进行分类。
- **bankloan_binning.sav**。该假设数据文件包含 5,000 位过去客户的财务和人口统计信息。
- **behavior.sav**。在一个经典示例中，52 名学生被要求以 10 分的标度对 15 种情况和 15 种行为的组合进行评价，该 10 分的标度介于 0 = 平均值在个人值之上，值被视为相异性。
- **behavior_ini.sav**。该数据文件包含 behavior.sav 的二维解的初始配置。

- **brakes.sav**。该假设数据文件涉及某生产高性能汽车盘式制动器的工厂的质量控制。该数据文件包含对 8 台专用机床中每一台的 16 个盘式制动器的直径测量。盘式制动器的目标直径为 322 毫米。
- **breakfast.sav**。在一项经典研究中，21 名 Wharton School MBA 学生及其配偶被要求按照喜好程度顺序对 15 种早餐食品进行评价，从 1 = 他们的喜好根据六种不同的情况加以记录，从“全部喜欢”到“只带饮料的快餐”。
- **breakfast-overall.sav**。该数据文件只包含早餐食品喜好的第一种情况，即“全部喜欢”。
- **broadband_1.sav**。该假设数据文件包含各地区订制了全国宽带服务的客户的数量。该数据文件包含 4 年期间 85 个地区每月的订户数量。
- **broadband_2.sav**。该数据文件和 broadband_1.sav 一样，但包含另外三个月的数据。
- **car_insurance_claims.sav**。在别处被提出和分析的关于汽车损坏赔偿的数据集。平均理赔金额可以当作其具有 gamma 分布来建模，通过使用逆联接函数将因变量的均值与投保人年龄、车辆类型和车龄的线性组合关联。提出理赔的数量可以作为尺度权重。
- **car_sales.sav**。该数据文件包含假设销售估计值、订价以及各种品牌和型号的车辆的物理规格。订价和物理规格可以从 edmunds.com 和制造商处获得。
- **car_sales_uprepared.sav**。这是 car_sales.sav 的修改版本，不包含字段的任何已转换版本。
- **carpet.sav**。在一个常用示例中，一家公司非常重视一种新型地毯清洁用品的市场营销，希望检验以下五种因素对消费者偏好的影响—包装设计、品牌名称、价格、优秀家用品标志和退货保证。包装设计有三个因子水平，每个因子水平因刷体位置而不同；有三个品牌名称（K2R、Glory 和 Bissell）；有三个价格水平；最后两个因素各有两个级别（有或无）。十名消费者对这些因素所定义的 22 个特征进行了排序。变量优选包含对每个特征的平均等级的排序。低排序与高偏好相对应。此变量反映了对每个特征的偏好的总体度量。
- **carpet_prefs.sav**。该数据文件所基于的示例和在 carpet.sav 中所描述的一样，但它还包含从 10 位消费者的每一位中收集到的实际排列顺序。消费者被要求按照从最喜欢到最不喜欢的顺序对 22 个产品特征进行排序。carpet_plan.sav 中定义了变量 PREF1 到 PREF22 包含相关特征的标识符。
- **catalog.sav**。该数据文件包含某编目公司出售的三种产品的假设每月销售数据。同时还包含 5 个可能的预测变量的数据。
- **catalog_seasfac.sav**。除添加了一组从“季节性分解”过程中计算出来的季节性因子和附带的日期变量外，该数据文件和 catalog.sav 是相同的。
- **cellular.sav**。该假设数据文件涉及某便携式电话公司在减少客户流失方面的举措。客户流失倾向分被应用到帐户，分数范围从 0 到 100。得到 50 分或更高分数的帐户可能会更换提供商。
- **ceramics.sav**。该假设数据文件涉及某制造商在确定新型优质合金是否比标准合金具有更高的耐热性方面的举措。每个个案代表对一种合金的单独检验；个案中会记录合金的耐热极限。
- **cereal.sav**。该假设数据文件涉及一份 880 人参与的关于早餐喜好的民意调查，该调查记录了参与者的年龄、性别、婚姻状况以及生活方式是否积极（根据他们是否每周至少做两次运动）。每个个案代表一个单独的调查对象。

- **clothing_defects.sav**。这是关于某服装厂的质量控制过程的假设数据文件。检验员要对工厂中每次大批量生产的服装进行抽样检测并清点不合格的服装的数量。
- **coffee.sav**。这是关于六种冰咖啡的认知品牌形象的数据文件。对于 23 种冰咖啡特征属性中的每种属性，人们选择了由该属性所描述的所有品牌。为保密起见，六种品牌用 AA、BB、CC、DD、EE 和 FF 来表示。
- **contacts.sav**。该假设数据文件涉及一组公司计算机销售代表的联系方式列表。根据这些销售代表所在的公司部门及其公司的秩来对每个联系方式进行分类。同时还记录了最近一次的销售量、最近一次销售距今的时间和所联系公司的规模。
- **creditpromo.sav**。该假设数据文件涉及某百货公司在评价最新信用卡促销的效果方面的举措。为此，随机选择了 500 位持卡人。其中一半收到了宣传关于在接下来的三个月内降低消费利率的广告。另一半收到了标准的季节性广告。
- **customer_dbase.sav**。该假设数据文件涉及某公司在使用数据仓库中的信息来为最有可能回应的客户提供特惠商品方面的举措。随机选择客户群的子集并为其提供特惠商品，同时记录下他们的回应。
- **customer_information.sav**。该假设数据文件包含客户邮寄信息，如姓名和地址。
- **customer_subset.sav**。来自 customer_dbase.sav 的拥有 80 个个案的子集。
- **debate.sav**。该假设数据文件涉及在某政治辩论前后对该辩论的参与者所做的调查的成对回答。每个个案对应一个单独的调查对象。
- **debate_aggregate.sav**。该假设数据文件分类汇总了 debate.sav 中的回答。每个个案对应一个辩论前后的偏好的交叉分类。
- **demo.sav**。这是关于购物客户数据库的假设数据文件，用于寄出每月的商品。将记录客户对商品是否有回应以及各种人口统计信息。
- **demo_cs_1.sav**。该假设数据文件涉及某公司在汇编调查信息数据库方面的举措的第一步。每个个案对应不同的城市，并记录地区、省、区和城市标识。
- **demo_cs_2.sav**。该假设数据文件涉及某公司在汇编调查信息数据库方面的举措的第二步。每个个案对应来自第一步中所选城市的不同的家庭单元，并记录地区、省、区、市、子区和单元标识。还包括设计前两个阶段的抽样信息。
- **demo_cs.sav**。该假设数据文件包含用复杂抽样设计收集的调查信息。每个个案对应不同的家庭单元，并记录各种人口统计和抽样信息。
- **dmdata.sav**。该假设数据文件包含直销公司的人口统计学和购买信息。dmdata2.sav 包含收到测试邮件的联系人子集的信息，dmdata3.sav 包含未收到测试邮件的其余联系人的信息。
- **dietstudy.sav**。该假设数据文件包含对“Stillman diet”的研究结果。每个个案对应一个单独的主体，并记录其在实行饮食方案前后的体重（磅）以及甘油三酸酯的水平（毫克/100 毫升）。
- **dvdplayer.sav**。这是关于开发新的 DVD 播放器的假设数据文件。营销团队用原型收集了焦点小组数据。每个个案对应一个单独的被调查用户，并记录他们的人口统计信息及其对原型问题的回答。
- **german_credit.sav**。该数据文件取自加州大学欧文分校的 Repository of Machine Learning Databases 中的“German credit”数据集。
- **grocery_1month.sav**。该假设数据文件是在数据文件 grocery_coupons.sav 的基础上加上了每周购物“累计”，所以每个个案对应一个单独的客户。所以，一些每周更改的变量消失了，而且现在记录的消费金额是为期四周的研究过程中的消费金额之和。

- **grocery_coupons.sav**。该假设数据文件包含由重视顾客购物习惯的杂货连锁店收集的调查数据。对每位顾客调查四周，每个个案对应一个单独的顾客周，并记录有关顾客购物地点和方式的信息（包括那一周里顾客在杂货上的消费金额）。
- **guttman.sav**。Bell 创建了一个表，用来阐释可能的社会群体。Guttman 引用了该表的一部分，其中包括五个变量，用于描述以下七个理论社会群体的社会交往、对群体的归属感、成员的物理亲近度以及关系正式性：观众（比如在足球比赛现场的人们）、听众（比如在剧院或听课堂讲座的人们）、公众（比如报纸或电视观众）、组织群体（与观众类似但具有紧密的关系）、初级群体（关系密切）、次级群体（自发组织）及现代社区（因在物理上亲近而导致关系松散并需要专业化服务）。
- **health_funding.sav**。该假设数据文件包含关于保健基金（每 100 人的金额）、发病率（每 10,000 人的比率）以及保健提供商拜访率（每 10,000 的比率）的数据。每个个案代表不同的城市。
- **hivassay.sav**。该假设数据文件涉及某药物实验室在开发用于检测 HIV 感染的快速化验方面的举措。化验结果为八个加深的红色阴影，如果有更深的阴影则表示感染的可能性很大。用 2,000 份血液样本来进行实验室试验，其中一半受到 HIV 感染而另一半没有受到感染。
- **hourlywagedata.sav**。该假设数据文件涉及在政府机关和医院工作的具有不同经验水平的护士的时薪。
- **insurance_claims.sav**。该假设数据文件涉及某保险公司，该公司希望构建一个模型用于标记可疑的、具有潜在欺骗性的理赔。每个个案代表一次单独的理赔。
- **insure.sav**。该假设数据文件涉及某保险公司，该公司正在研究指示客户是否会根据 10 年的人寿保险合同提出理赔的风险因子。数据文件中的每个个案代表一副根据年龄和性别进行匹配的合同，其中一份记录了一次理赔而另一份则没有。
- **judges.sav**。该假设数据文件涉及经过训练的裁判（加上一个体操爱好者）对 300 次体操表演给出的分数。每行代表一次单独的表演；裁判们观看相同的表演。
- **kinship_dat.sav**。Rosenberg 和 Kim 开始分析 15 个亲属关系项（伯母、兄弟、表兄妹、女儿、父亲、孙女、祖父、祖母、孙子、母亲、侄子或外甥、侄女或外甥女、姐妹、儿子和叔叔）。他们让四组大学生（两组女同学，两组男同学）根据相似程度将各项排序。他们让其中的两组同学（一组女同学，一组男同学）进行了两次排序，第二次排序和第一次排序采取的标准不同。这样，一共得到六组“源”。每个源对应一个 15×15 的相似性矩阵，其单元格中的值等于源中的人数减去此源中对象被划分的次数。
- **kinship_ini.sav**。该数据文件包含 kinship_dat.sav 的三维解的初始配置。
- **kinship_var.sav**。该数据文件包含自变量 gender、gener(ation) 和 degree (of separation)，这些变量可用于解释 kinship_dat.sav 的解的维数。具体而言，它们可用来将解的空间限制为这些变量的线性组合。
- **marketvalues.sav**。该数据文件涉及 1999 - 2000 年间 Algonquin, Ill. 地区新的房屋开发中的住房销售。这些销售仅仅来自公众记录。
- **nhis2000_subset.sav**。美国健康访问调查 (NHIS) 是针对美国全体公民的大型人口调查。该调查对美国的具有全国代表性的家庭样本进行了面对面的访问，并获取了每个家庭的成员的健康行为和健康状态的人口统计信息和观察数据。该数据文件包含取自 2000 年调查信息的子集。国家健康统计中心。2000 年美国健康访问调查。公用数据文件和文档。

ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/。2003 年发布。

- **ozone.sav**。这些数据包含了用来根据其余变量预测臭氧浓度的六个气象变量的 330 个观察值。在以前的研究人员中，和 发现了这些变量之间的非线性，这妨碍了标准回归方法。
- **pain_medication.sav**。该假设数据文件包含用于治疗慢性关节炎疼痛的抗炎药的临床试验结果。我们感兴趣的是该药见效的时间以及它和现有药物的比较。
- **patient_los.sav**。该假设数据文件包含被医院确诊为疑似心肌梗塞（即 MI 或“心脏病发作”）的患者的治疗记录。每个个案对应一位单独的患者，并记录与其住院期有关的一些变量。
- **patlos_sample.sav**。该假设数据文件包含在治疗心肌梗塞（即 MI 或“心脏病发作”）期间收到溶解血栓剂的患者的样本治疗记录。每个个案对应一位单独的患者，并记录与其住院期有关的一些变量。
- **poll_cs.sav**。该假设数据文件涉及民意测验专家在确定正式立法前公众对法案的支持水平方面的举措。个案对应注册的选民。每个个案记录选民居住的县、镇、区。
- **poll_cs_sample.sav**。该假设数据文件包含在 poll_cs.sav 中列出的选民的样本。该样本是根据 poll_csplan 中指定的设计来选取的，而且该数据文件记录包含概率和样本权重。请注意，由于该抽样计划使用与大小成正比（PPS）方法，因此，还有一个文件（poll_jointprob.sav）包含联合选择概率。在选取了样本之后，对应于选民人群统计信息及其对提交法案的意见的附加变量将被收集并添加到数据文件。
- **property_assess.sav**。该假设数据文件涉及某县资产评估员在利用有限的资源不断更新资产价值评估方面的举措。个案对应过去一年中县里所出售的资产。数据文件中的每个个案记录资产所在的镇、最后评估资产的评估员、该次评估距今的时间、当时的估价以及资产的出售价格。
- **property_assess_cs.sav**。该假设数据文件涉及某州资产评估员在利用有限的资源不断更新资产价值评估方面的举措。个案对应该州的资产。数据文件中的每个个案记录资产所在的县、镇和区，最后一次评估距今的时间以及当时的估价。
- **property_assess_cs_sample.sav**。该假设数据文件包含在 property_assess_cs.sav 中列出的资产的样本。该样本是根据 property_assess_csplan 中指定的设计来选取的，而且该数据文件记录包含概率和样本权重。在选取了样本之后，附加变量 Current value 将被收集并添加到数据文件。
- **recidivism.sav**。该假设数据文件涉及某政府执法机构在了解其管辖区域内的屡犯率方面的举措。每个个案对应先前的一名罪犯，并记录其人口统计信息和第一次犯罪的详细资料；如果在第一次被捕后两年内又第二次被捕，则还将记录两次被捕间隔的时间。
- **recidivism_cs_sample.sav**。该假设数据文件涉及某政府执法机构在了解其管辖区域内的屡犯率方面的举措。每个个案对应应在 2003 年 6 月期间第一次被捕释放的先前的一名罪犯，并记录其人口统计信息和第一次犯罪的详细资料，及其第二次被捕的数据（如果发生在 2006 年 6 月底之前）。根据 recidivism_cs_csplan 中指定的抽样计划从抽样部门选择罪犯；该计划使用与大小成正比（PPS）方法，因此，还有一个文件（recidivism_cs_jointprob.sav）包含联合选择概率。
- **rfm_transactions.sav**。此假设数据文件包含购买交易数据，即每笔交易的购买日期、购买商品和消费金额。

- **salesperformance.sav**。这是关于评估两个新的销售培训课程的假设数据文件。60 名员工被分成 3 组且都接受标准的培训。另外，组 2 接受技术培训；组 3 接受实践教学。在培训课程结束时，对每名员工进行测验并记录他们的分数。数据文件中的每个个案代表一名单独的受训者，并记录其被分配到的组以及测验的分数。
- **satisf.sav**。该假设数据文件涉及某零售公司在 4 个商店位置所进行的满意度调查。总共对 582 位客户进行了调查，每个个案代表一位单独客户的回答。
- **screws.sav**。该数据文件包含关于螺钉、螺栓、螺母和图钉的特征的信息。
- **shampoo_ph.sav**。这是关于某发制品厂的质量控制的假设数据文件。在规定的对六批独立输出的产品进行检测并记录它们的 pH 值。目标范围是 4.5 - 5.5。
- **ships.sav**。在别处被提出和分析的关于波浪对货船造成的损坏的数据集。在给出了船的类型、建造工期和服务期后，可以根据以泊松比率发生来为事件计数建模。在因子交叉分类构成的表格中，每个单元格的分类汇总服务月数提供遇到风险的值。
- **site.sav**。该假设数据文件涉及某公司在为扩展业务而选择新址方面的举措。该公司聘请了两名顾问分别对选址进行评估，除了提供长期报告外，他们还要以“前景颇佳”、“前景良好”或“前景不佳”来对每个选址进行总结。
- **smokers.sav**。该数据文件摘自 1998 年全国家庭药物滥用调查并且是美国家庭的概率样本。<http://dx.doi.org/10.3886/ICPSR02934> 因此，分析该数据文件的第一步应该是对数据进行加权以反映总体趋势。
- **stocks.sav** 该假设数据文件包含某一年的股票价格和成交量。
- **stroke_clean.sav**。该假设数据文件包含某医学数据库在经过“数据准备”选项中的过程清理后的状态。
- **stroke_invalid.sav**。该假设数据文件包含某医学数据库的初始状态及一些数据输入错误。
- **stroke_survival**。此假设数据文件涉及正在研究结束缺血性中风后复元计划的患者存活时间的研究人员面临着很多挑战。中风后，记录心肌梗塞、缺血性中风或出血性中风的发生及其时间。样本为左侧截短，因为只包含在中风后管理的复元计划结束后存活的患者。
- **stroke_valid.sav**。该假设数据文件包含在使用“验证数据”过程检查值后，某医学数据库的状态。它仍包含潜在异常个案。
- **survey_sample.sav**。此数据文件包含调查数据，包括人口统计学数据和各种态度测量。它基于 1998 NORC 综合社会调查的变量子集，但某些数据值已经过修改，并添加了其他虚拟变量以供演示用途。
- **telco.sav**。该假设数据文件涉及某电信公司在减少客户群中的客户流失方面的举措。每个个案对应一个单独的客户，并记录各类人口统计和服务用途信息。
- **telco_extra.sav**。该数据文件与 telco.sav 数据文件类似，但删除了“tenure”和经对数转换的客户消费变量，代替它们的是标准化的对数转换客户消费变量。
- **telco_missing.sav**。该数据文件是 telco.sav 数据文件的子集，但某些人口统计数据值已被缺失值替换。
- **testmarket.sav**。该假设数据文件涉及某快餐连锁店为其菜单添加新项目的计划。有三种可能的促销新产品的活动，所以会在多个随机选择的市场中的地点引入新的项目。在每个地点采用不同的促销方式，并记录新项目四周的每周销售情况。每个个案对应单独地点的一周。

- **testmarket_1month.sav**。该假设数据文件是在数据文件 testmarket.sav 的基础上加上了每周销售“累计”，所以每个个案对应一个单独的地点。所以，一些每周更改的变量消失了，而且现在记录的销售是为期四周的研究过程中的销售之和。
- **tree_car.sav**。该假设数据文件包含人口统计和车辆购买价格数据。
- **tree_credit.sav**。该假设数据文件包含人口统计和银行贷款历史数据。
- **tree_missing_data.sav** 该假设数据文件包含具有大量缺失值的人口统计和银行贷款历史数据。
- **tree_score_car.sav**。该假设数据文件包含人口统计和车辆购买价格数据。
- **tree_textdata.sav**。这是一个只有两个变量的样本数据文件，主要打算在指定测量级别和值标签之前显示变量的默认状态。
- **tv-survey.sav**。该假设数据文件涉及由某电视演播室进行的一项关于是否要继续制作一档成功的节目的调查。906 位调查对象被问及他们在各种情况下是否会收看该节目。每行代表一位单独的调查对象；每列代表一种单独的情况。
- **ulcer_recurrence.sav**。此文件包含某项研究的部分信息，该研究旨在比较两种用来防止溃疡复发的治疗的功效。它提供了区间数据的优秀示例并且已在别处被提出和分析。
- **ulcer_recurrence_recoded.sav**。该文件重新组织 ulcer_recurrence.sav 中的信息以允许为研究的每个区间的事件概率建模而不是简单地研究结束事件概率建模。它已在别处被提出和分析。
- **verd1985.sav**。该数据文件涉及某项调查。该调查记录了 15 个主体对 8 个变量的响应。需要处理的变量被分成 3 个集。数据集 1 包含 年龄 和 婚姻；数据集 2 包含 宠物 和 新闻；数据集 3 包含 音乐 和 居住。宠物被尺度化为多名义而年龄被尺度化为有序；所有其他变量都被尺度化为单名义。
- **virus.sav**。该假设数据文件涉及某因特网服务提供商（ISP）在确定病毒对其网络的影响方面的举措。他们从发现病毒到威胁得以遏制这段时间内跟踪其网络上受感染的电子邮件的流量的（近似）百分比。
- **wheeze_steubenville.sav**。这是关于空气污染对儿童健康影响的纵向研究的一个子集。这些数据包含儿童的气喘状况的重复二分类测量（这些儿童来自 Steubenville, Ohio, 年龄为 7 到 10 岁），以及母亲在研究的第一年中是否为吸烟者的固定记录。
- **workprog.sav**。该假设数据文件涉及一份尝试为弱势群体提供较好的工作的政府工作计划。文件后还有一个潜在计划参与者的样本，其中一些参与者是被随机选择来参加该计划的，而其他参与者则不是。每个个案代表一位单独的计划参与者。
- **worldsales.sav** 该假设数据文件包含按不同大洲和产品列出的销售收入。

注意事项

这些信息开发用于在全球提供的产品和服务。

IBM 可能在其他国家/地区中不提供在本文档中讨论的产品、服务或功能。请咨询您当地的 IBM 代表以了解有关您所在地区当前可用产品和服务的信息。任何对 IBM 产品、程序或服务的引用，并不意味着仅可使用这些 IBM 产品、程序或服务。作为替代，可以使用任何功能相当的产品、程序或服务，前提是不侵犯任何 IBM 知识产权。不过，用户应自行负责评估和验证任何非 IBM 产品、程序或服务的工作情况。

在本文档中介绍的主题可能涉及 IBM 的专利或申请中的专利。您获得该文档并不意味着授予您任何这些专利许可。您可以将书面许可查询函件发送至：

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, U. S. A.

要查询双字节字符集 (DBCS) 相关许可证信息，请联系所在国家/地区中的 IBM 知识产权部门，或者以书面形式将查询函件发送至：

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

以下段落不适用于英国或此类条款与当地法律不符的其他国家/地区： INTERNATIONAL BUSINESS MACHINES 公司“按原样”提供本出版物，不保证任何明示或暗示，包括但不限于对非侵权性、适销性或对特定用途适用性的暗示担保。有些州/省不允许针对某些交易的明示或暗示免责条款，因此本声明可能不适用于您。

本信息可能含有技术误差或排版错误。此处的信息会定期进行更改；这些更改会体现在本出版物的新版本中。IBM 可能随时对本出版物中所述的产品和/或程序进行改进和/或更改，恕不另行通知。

在本信息材料中对任何非 IBM 网站的引用仅为了方便用户，并不以任何方式表明对这些网站的认可。这些网站上的材料并非本 IBM 产品材料的一部分，您对这些网站的使用需自担风险。

IBM 可以自认为适当并且不会对您构成任何约束的任何方式使用或分发您提供的任何信息。

如果本程序的受许可方试图了解有关程序的信息以启用：(i) 在独立创建的程序和其他程序（包括本程序）之间交换信息；(ii) 相互使用交换的信息，则应联系：

IBM Software Group, Attention:Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

此类信息的提供应遵照相关条款和条件，其中包括在某些情况下支付适当费用。

在本文档中介绍的受许可保护程序，及其所有受许可保护材料由 IBM 在双方签署的“IBM 客户协议”、“IBM 国际程序许可证协议”或任何其他等同协议下提供。

非 IBM 产品的相关信息来自这些产品的供应商，及其发布的公告或其他公开来源。IBM 并未测试过这些产品，无法确认有关非 IBM 产品的性能准确性、兼容性或任何其他声明。有关非 IBM 产品功能的问题应由这些产品的供应商负责。

本信息包含日常业务运营中使用的数据和报告的示例。为了尽可能详尽地对其进行说明，示例中包含了人员的姓名、公司、品牌和产品的名称。所有这些名称均为虚构，与真实商业企业使用的名称和地址的任何雷同纯属巧合。

如果您正在查阅此信息的软拷贝，照片和彩色插图可能不会显示。

商标

IBM、IBM 徽标、ibm.com 和 SPSS 是 IBM Corporation 的商标，在全球许多司法辖区注册。有关最新的 IBM 商标列表，请访问网页 <http://www.ibm.com/legal/copytrade.shtml>。

Adobe、Adobe 徽标、PostScript 和 PostScript 徽标是 Adobe Systems Incorporated 在美国和/或其他国家或地区的注册商标或商标。

Intel、Intel 徽标、Intel Inside、Intel Inside 徽标、Intel Centrino、Intel Centrino 徽标、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 是 Intel Corporation 或其子公司在美国和其他国家或地区的商标或注册商标。

Java 和所有基于 Java 的商标和徽标是 Sun Microsystems, Inc. 在美国、其他国家或地区或两者的商标。

Linux 是 Linus Torvalds 在美国、其他国家或地区或两者的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国、其他国家或地区或两者的商标。

UNIX 是 The Open Group 在美国和其他国家或地区的注册商标。

此产品使用 Polar 工程咨询公司的 WinWrap Basic，版权 1993 年-2007 年，<http://www.winwrap.com>。

其他产品和服务名称可能是 IBM 或其他公司的商标。

Adobe 产品屏幕截图重印已获得 Adobe Systems Incorporated 的许可。

Microsoft 产品屏幕截图重印已获得 Microsoft Corporation 的许可。



索引

- Access (Microsoft), 14
- Excel (Microsoft)
 - 将结果导出到, 106
- Excel 文件
 - 读取, 12
- HTML
 - 导出结果, 116
- PowerPoint (Microsoft)
 - 将结果导出到, 106
- Word (Microsoft)
 - 将结果导出到, 106

- 个案
 - 排序, 144, 146
 - 选择, 147
- 个案的子集
 - 删除未选定的个案, 151
 - 基于日期和时间, 150
 - 如果满足条件, 148
 - 条件表达式, 148
 - 过滤未选定的个案, 151
 - 选择, 147
 - 随机样本, 149

- 值标签
 - 指定, 33, 35
 - 控制浏览器中的显示, 33, 35
 - 数值变量, 33
 - 非数值变量, 35

- 分类数据, 56
 - 摘要测量, 56
- 创建变量标签, 32
- 刻度变量
 - 摘要测量, 59
- 刻度数据, 56

- 变量, 28
 - 数据类型, 32
 - 标签, 32
- 变量属性
 - 重新使用, 40
- 变量标签
 - 创建, 32

- 名义数据, 56
- 商标, 161

- 图形
 - 创建图形, 63
 - 图表选项, 82
 - 条形, 63
 - 模板, 77
 - 编辑图形, 69

- 图表
 - 创建图表, 63
 - 图表选项, 82
 - 条形, 58, 63
 - 模板, 77
 - 直方图, 61
 - 编辑图表, 69

- 在枢轴表中转置(翻转)行与列, 88
- 在枢轴表中隐藏行与列, 92

- 复制变量属性, 40

- 多个打开的数据文件, 51
 - 取消, 54

- 字符串数据
 - 输入数据, 30

- 定性数据, 56
- 定比数据, 56
- 定距数据, 56
- 定量数据, 56

- 对个案排序, 144
- 对值重新编码, 124
- 导出结果
 - HTML, 116
 - 到 Excel, 106
 - 到 PowerPoint, 106
 - 到 Word, 106

- 层
 - 在枢轴表中创建, 90

- 拆分文件处理, 145
- 摘要测量
 - 分类数据, 56
 - 刻度变量, 59

- 数值数据, 28
- 数据库向导, 14
- 数据库文件
 - 读取, 14
- 数据文件
 - 多个打开的数据文件, 51
- 数据类型
 - 变量, 32
- 数据编辑器
 - 多个打开的数据文件, 51
 - 输入数值数据, 28
 - 输入非数值数据, 30

- 数据输入, 28, 30
- 数据集
 - 重命名, 54
- 文本导入向导, 20
- 文本数据文件
 - 读取, 20
- 日期和时间变量, 135
- 日期和时间向导, 135
- 有序数据, 56
- 条件表达式, 134
- 条形图, 58
- 枢轴表
 - 交换行和列, 88
 - 单元格数据类型, 92
 - 单元格格式, 92
 - 层, 90
 - 格式编排, 91
 - 编辑, 91
 - 访问定义, 87
 - 透视托盘, 88
 - 隐藏小数点, 92
 - 隐藏行与列, 92
- 查看器
 - 移动输出, 86
 - 隐藏和显示输出, 86
- 样本文件
 - 位置, 153
- 法律注意事项, 160
- 测量级别, 56
- 电子表格文件
 - 读取, 12
 - 读取变量名称, 12
- 直方图, 61
- 移动
 - 枢轴表中的元素, 88
 - 查看器中的项, 86
- 粘贴语法
 - 从对话框, 117
- 系统缺失值, 37
- 编辑枢轴表, 91
- 缺失值
 - 数值变量, 38
 - 系统缺失, 37
 - 非数值变量, 39
- 表达式中的函数, 132
- 计数
 - 计数表, 56
 - 计算新变量, 130
- 语法, 117
- 语法帮助工具, 120
- 语法文件
 - 打开, 120
- 语法窗口
 - 分界点, 121
 - 粘贴命令, 117
 - 编辑命令, 118
 - 自动完成, 118
 - 运行命令, 118, 120
 - 颜色编码, 120
- 输入数据
 - 数值, 28
 - 非数值, 30
- 连续数据, 56
- 选择个案, 147
- 重命名数据集, 54
- 频率表, 56