

*IBM SPSS Statistics 30* 简介指南



## 注

使用本信息及其支持的产品前，请先阅读 [第 79 页的『声明』](#) 中的信息。

## 产品信息

本版本适用于 30 版本 0、IBM® SPSS Statistics 的修改 0 以及所有后续版本和修改，除非新版本另有说明。

© Copyright International Business Machines Corporation .

# 内容

<b>第 1 章 简介</b> .....	<b>1</b>
样本文件.....	1
打开数据文件.....	1
运行分析.....	2
创建图表.....	4
<b>第 2 章 读取数据</b> .....	<b>7</b>
IBM SPSS Statistics 数据文件的基本结构.....	7
读取 IBM SPSS Statistics 数据文件.....	7
读取 Excel 数据.....	8
从数据库中读取数据.....	10
从文本文件读取数据.....	12
<b>第 3 章 使用数据编辑器</b> .....	<b>17</b>
输入数值数据.....	17
输入字符串数据.....	18
定义数据.....	19
添加变量标签.....	19
更改变量类型和格式.....	20
添加值标签.....	20
处理缺失数据.....	21
数值变量的缺失值.....	21
字符串变量的缺失值.....	22
<b>第 4 章 检查单个变量的汇总统计</b> .....	<b>23</b>
测量级别.....	23
分类数据的汇总测量.....	23
分类数据的图表.....	24
刻度变量的汇总测量.....	25
刻度变量的直方图.....	26
<b>第 5 章 创建和编辑图表</b> .....	<b>27</b>
图表创建的基础知识.....	27
使用图表构建器库.....	27
定义变量和统计.....	28
添加文字.....	29
创建图表.....	30
<b>第 6 章 使用输出</b> .....	<b>33</b>
使用查看器.....	33
使用透视表编辑器.....	34
访问输出定义.....	34
透视表.....	35
创建并显示层.....	36
编辑表.....	36
隐藏行与列.....	37
更改数据显示格式.....	37
表格外观.....	38
使用预先定义的格式.....	38
自定义表格外观样式.....	39

更改缺省的表格格式.....	41
自定义初始显示设置.....	42
显示变量和值标签.....	43
在其他应用程序中使用结果.....	44
将结果作为 Word 表格粘贴.....	44
将结果作为文本进行粘贴.....	45
将结果导出到 Microsoft Word、PowerPoint 和 Excel 文件.....	45
将结果导出到 PDF.....	51
将结果导出到 HTML.....	53
<b>第 7 章 使用语法.....</b>	<b>55</b>
粘贴语法.....	55
编辑语法.....	56
打开并运行语法文件.....	57
使用分界点.....	57
<b>第 8 章 修改数据值.....</b>	<b>59</b>
从刻度变量创建分类变量.....	59
计算新变量.....	61
使用表达式中的函数.....	61
使用条件表达式.....	62
使用日期和时间.....	63
计算两个日期之间的时间长度.....	63
向日期添加期间.....	64
<b>第 9 章 排序和选择数据.....</b>	<b>67</b>
对数据排序.....	67
拆分文件处理.....	67
为拆分文件处理排序个案.....	69
打开和关闭拆分文件处理.....	69
选择个案的子集.....	69
基于条件表达式选择个案.....	70
要选择随机样本请执行下列操作.....	71
选择时间范围或个案范围.....	71
未选定的个案的处理.....	72
个案选择状态.....	72
<b>第 10 章 样本文件.....</b>	<b>73</b>
<b>声明.....</b>	<b>79</b>
商标.....	80
<b>索引.....</b>	<b>81</b>

# 第 1 章 简介

本指南将向您介绍如何使用多种可用的功能。本教程旨在提供循序渐进的实践指南。示例中显示的所有文件均随应用程序一起安装，以便您可以遵照示例来执行相同的分析，并获得与此处所示相同的结果。

如果您需要各种统计分析技巧的详细示例，请尝试循序渐进的个案研究（可从“帮助”菜单访问）。

## 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：

1. 从菜单中选择：

文件 > 打开 > 数据...

将显示打开文件的对话框。

缺省情况下，将显示 IBM SPSS Statistics 数据文件（.sav 扩展名）。

本示例使用文件 *demo.sav*。

	age	marital	address	income	inccat	car
1	55	Marital status	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

图 1: “数据编辑器”中的 *demo.sav* 文件

数据文件将显示在“数据编辑器”中。在“数据视图”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）。

缺省情况下，将显示实际的数据值。要显示标签：

2. 从菜单中选择：

视图 > 值标签



图 2: “值标签”按钮

或者，也可以使用工具栏上的“值标签”按钮。

	age	marital	address	income	inccat	car
1	55	Married	12	72.00	\$50 - \$74	36.
2	56	Unmarried	29	153.00	\$75+	76.
3	28	Married	9	28.00	\$25 - \$49	13.
4	24	Married	4	26.00	\$25 - \$49	12.
5	25	Unmarried	2	23.00	Under \$25	11.
6	45	Married	9	76.00	\$75+	37.
7	42	Unmarried	19	40.00	\$25 - \$49	19.
8	35	Unmarried	15	57.00	\$50 - \$74	28.
9	46	Unmarried	26	24.00	Under \$25	12.
10	34	Married	0	89.00	\$75+	46.
11	55	Married	17	72.00	\$50 - \$74	35.

图 3: “数据编辑器”中显示的值标签

当前显示了描述性值标签以便更轻松地解释响应。

## 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。

1. 从菜单中选择:

分析 > 描述统计 > 频率...

此时，将显示“频率”对话框。

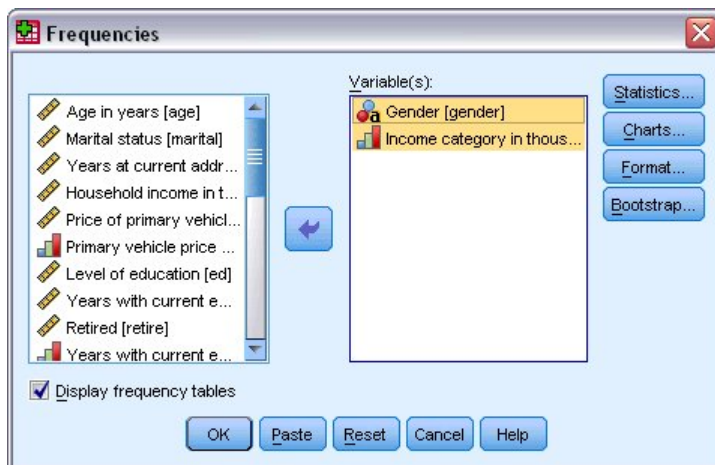













图 4: “频率”对话框

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。

	数字	String	日期	时间
标度 (连续)		不适用		
有序(O)				
名义				

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。

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并从弹出菜单中选择**变量信息**来获取附加信息。

2. 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。
3. 单击源列表中的 *Income category in thousands [inccat]*，并将其拖至目标列表。

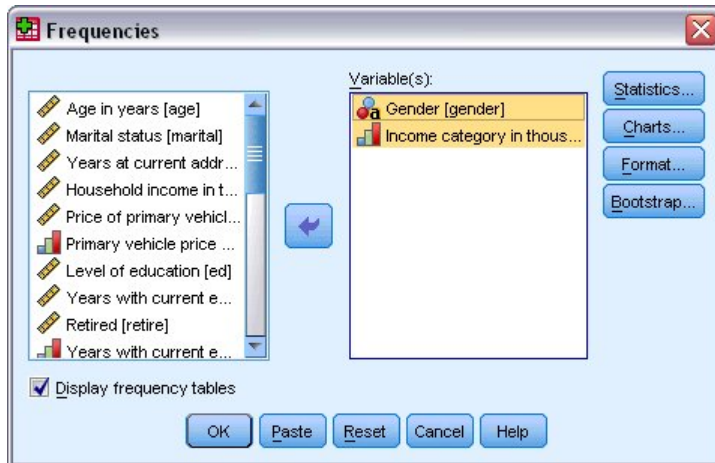


图 5: 为分析选择的变量

4. 单击**确定**以运行该过程。  
结果将显示在查看器窗口中。

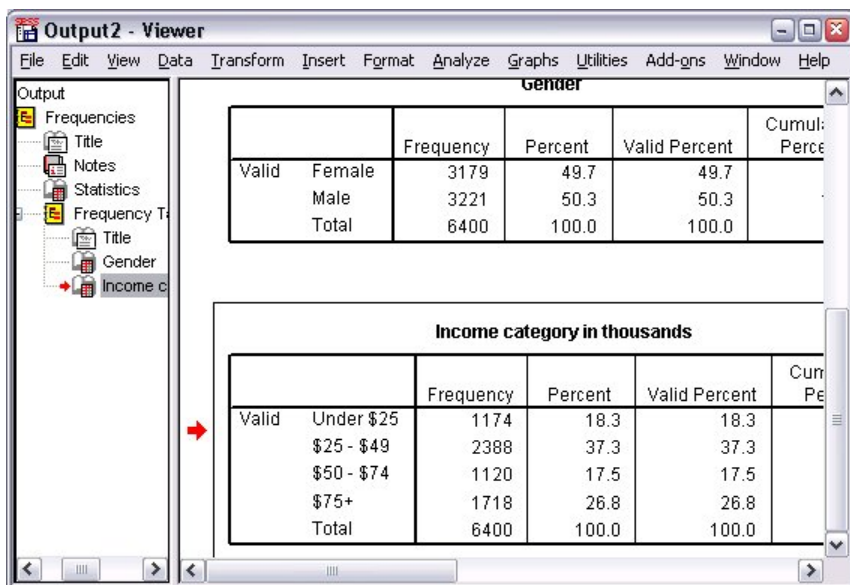


图 6: 收入类别的频率表

## 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。

1. 从菜单中选择：

图形 > 图表构建器...



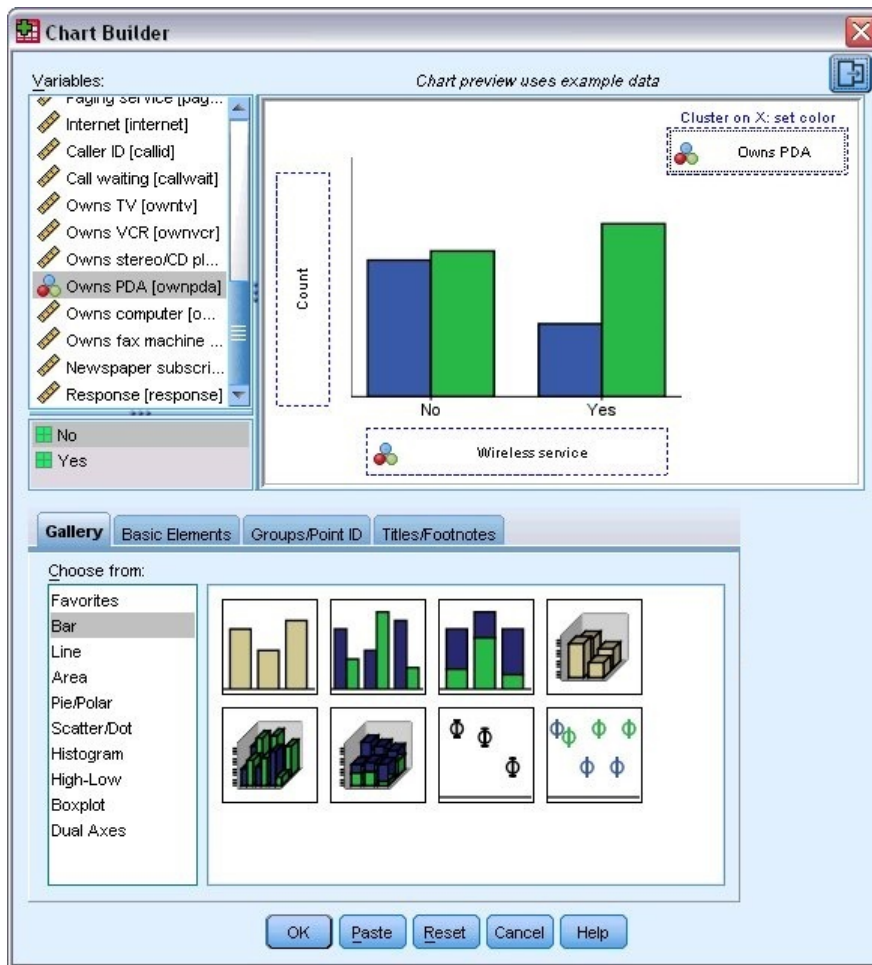


图 7: 带有完整拖放区域的“图表构建器”对话框

2. 单击库选项卡（如果尚未将其选中）。
3. 单击**条形图**（如果尚未将其选中）。
4. 将“聚类条形图”图标拖到画布上，即“库”上方的大片区域。
5. 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。
6. 将 *Wireless service [wireless]* 变量拖动到 x 轴。
7. 右键单击 *Owns PDA [ownpda]* 并选择**名义**作为其测量级别。
8. 将 *Owns PDA [ownpda]* 变量拖到画布右上角的聚类拖放区域。
9. 单击**确定**以创建图表。

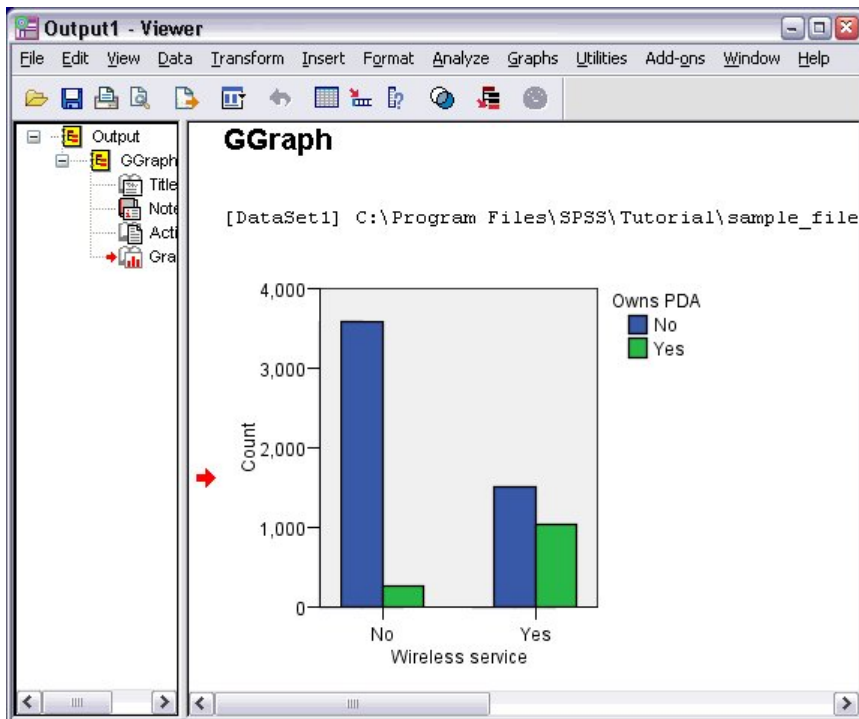


图 8: “查看器”窗口中显示的条形图

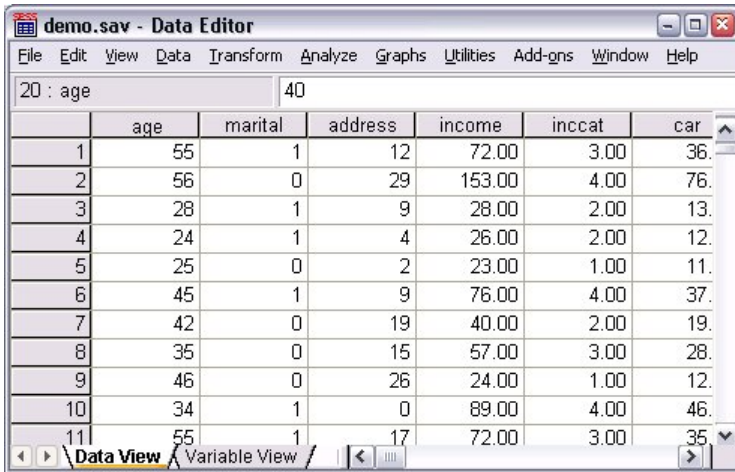
条形图将显示在查看器中。该图表显示了使用无线电话服务的人员远比不使用无线服务的人员更可能拥有 PDA。

您可以通过在查看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。

## 第 2 章 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



	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

图 9: 数据编辑器

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，个案表示各个响应者。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

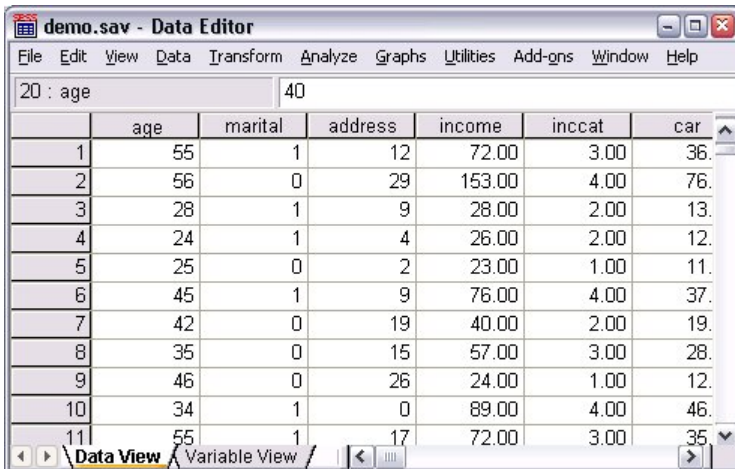
IBM SPSS Statistics 数据文件的文件扩展名为 *.sav*，包含您保存的数据。

1. 从菜单中选择：

文件 > 打开 > 数据...

2. 浏览至 *demo.sav* 并将其打开。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。

数据将立即显示在数据编辑器中。



	age	marital	address	income	inccat	car
1	55	1	12	72.00	3.00	36.
2	56	0	29	153.00	4.00	76.
3	28	1	9	28.00	2.00	13.
4	24	1	4	26.00	2.00	12.
5	25	0	2	23.00	1.00	11.
6	45	1	9	76.00	4.00	37.
7	42	0	19	40.00	2.00	19.
8	35	0	15	57.00	3.00	28.
9	46	0	26	24.00	1.00	12.
10	34	1	0	89.00	4.00	46.
11	55	1	17	72.00	3.00	35.

图 10: 打开的数据文件

## 读取 Excel 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。

1. 从菜单中选择：

文件 > 导入数据 > Excel

2. 转至 Samples\English 文件夹并选择 demo.xlsx。

“读取 Excel 文件”对话框显示数据文件的预览。将显示文件中第一个工作表的内容。如果文件具有多个工作表，那么您可以从列表中选择工作表。

您可以看到某些 *Gender* 的字符串值具有前导空格。 *MaritalStatus* 的某些值显示为句点 (.)。

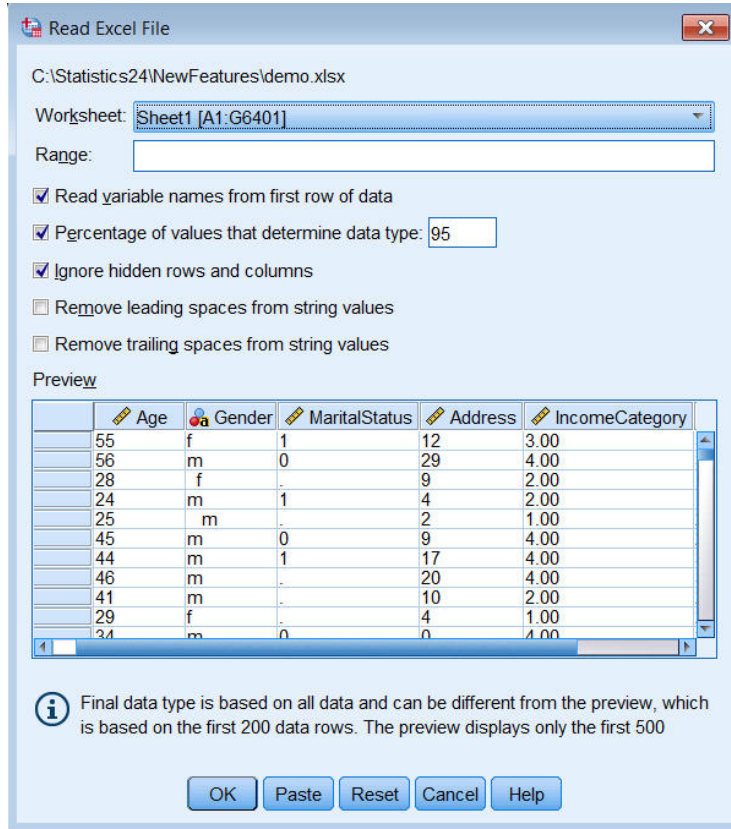
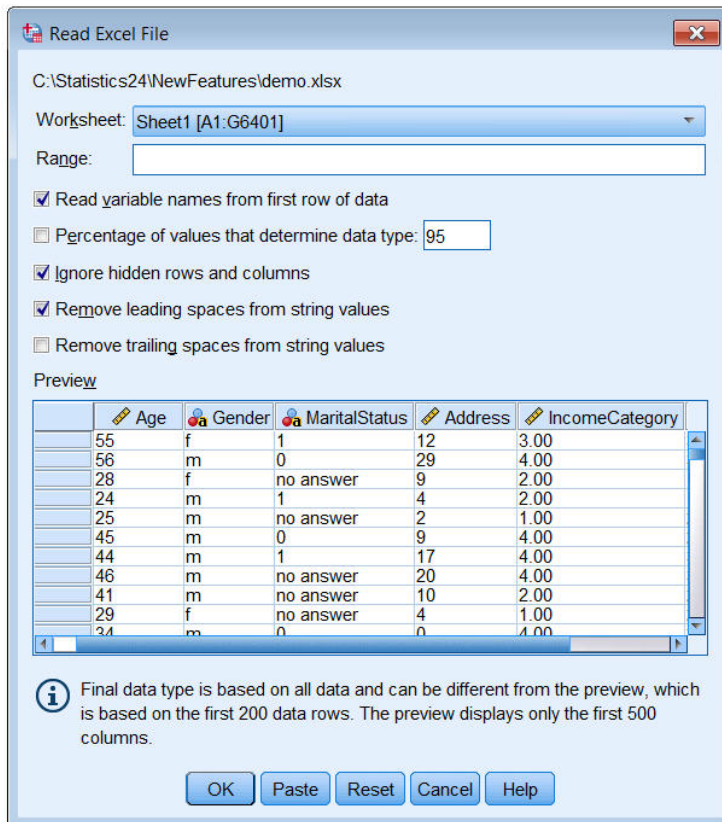


图 11: “读取 Excel 文件”对话框

3. 确保选择了**从第一行数据读取变量名**。如果列标题不符合变量名称规则，那么将转换为有效变量名称。原始列标题将另存为变量标签。
4. 选择**除去字符串值的前导空格**。
5. 取消选择**确定数据类型的值的百分比**。



现在，将在系统缺失的单元格中显示字符串值“no answer”。如果不存在值参数的百分比并且列包含混合数据类型，那么变量读取为字符串数据类型。将保留所有值，但是数字值将被视为字符串值。

6. 选择（选中）**确定数据类型的值的百分比**以将 *MaritalStatus* 读取为数字变量。
7. 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，列标题“Marital Status”将转换为变量 *MaritalStatus*。原始的列标题保留为变量标签。

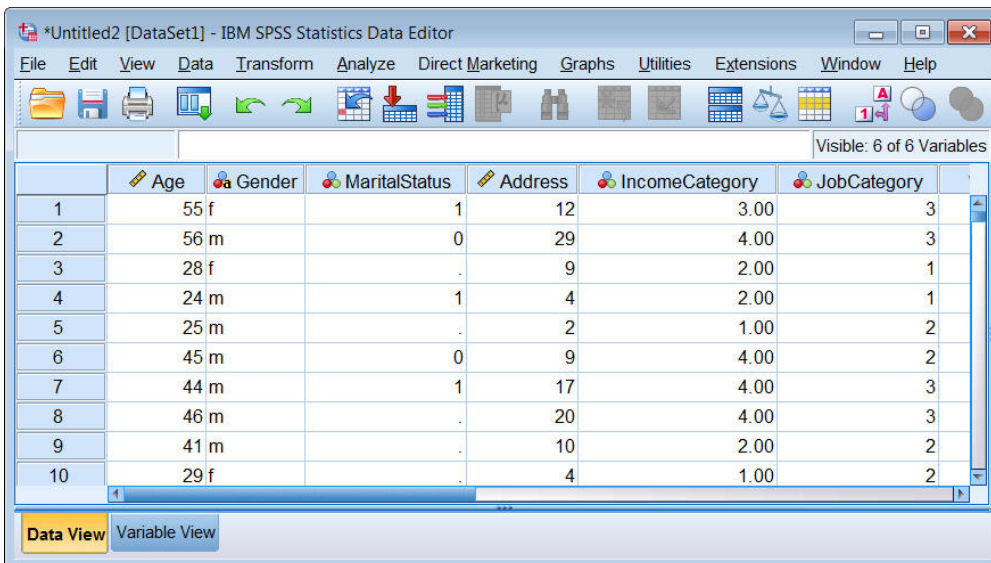


图 12: 导入的 Excel 数据

### 相关信息

第 73 页的『样本文件』

## 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：本示例特定于 Microsoft Windows 且 Access 需要 ODBC 驱动程序。Microsoft Access ODBC 驱动程序仅用于 32 位版本的 IBM SPSS Statistics。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。

1. 从菜单中选择：

文件 > 导入数据 > 数据库 > 新建查询...

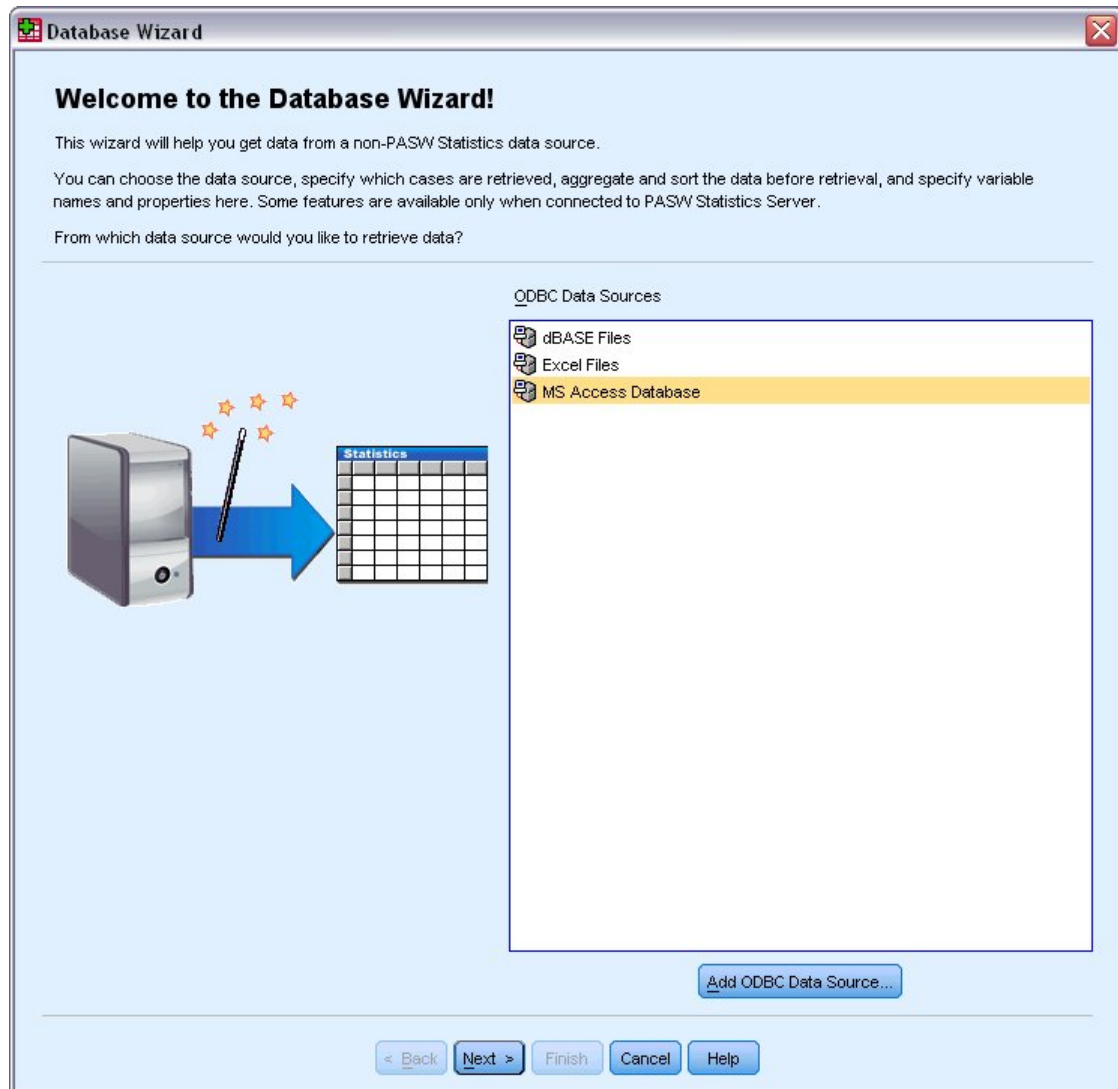


图 13: “欢迎使用数据库向导”对话框

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

注：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。

3. 单击浏览导航到要打开的 Access 数据库文件。
4. 打开 *demo.mdb*。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。
5. 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。

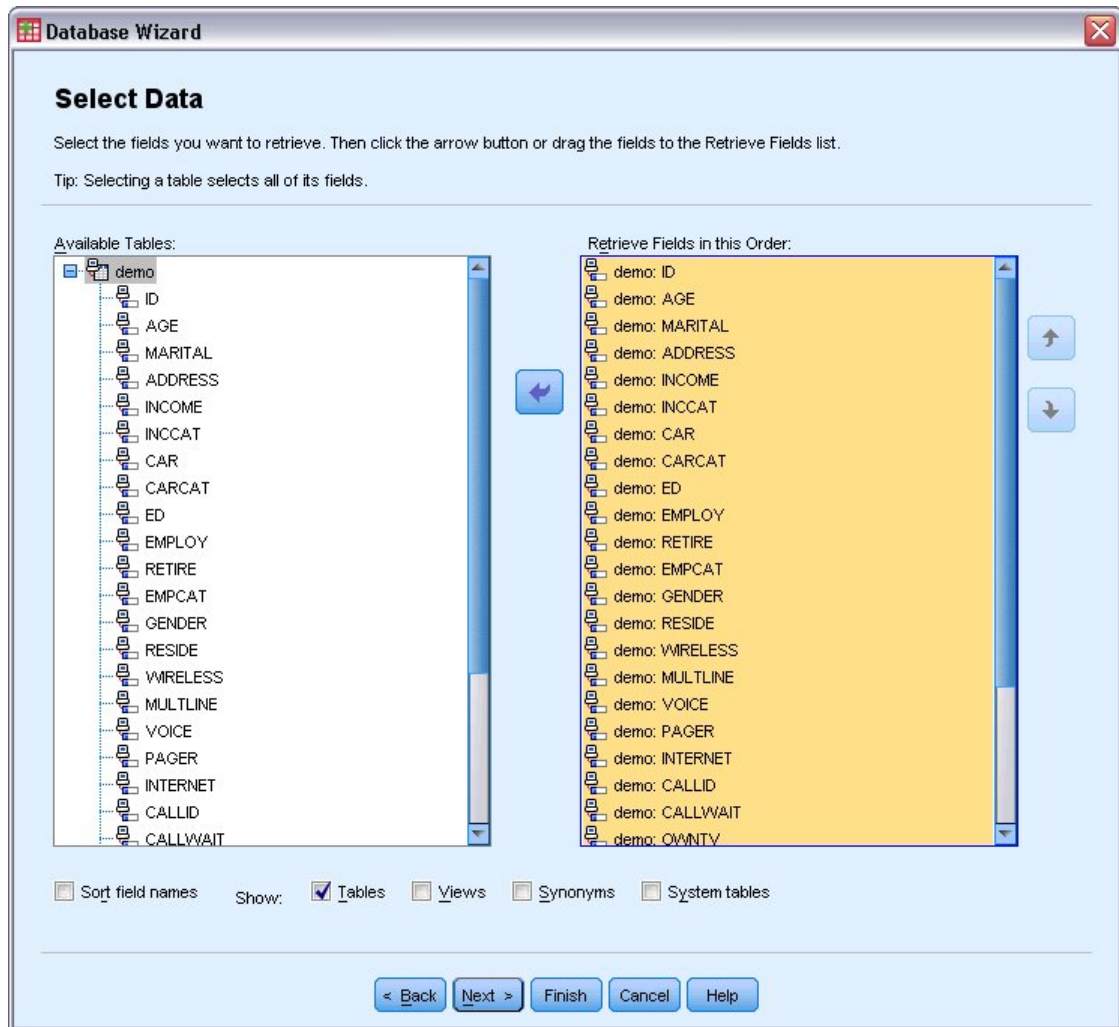


图 14: “选择数据”步骤

6. 将整个 **demo** 表拖到“以此顺序检索字段”列表中。
7. 单击下一步。

在下一步骤中，可以选择要导入的记录（个案）。

如果不想导入所有个案，则可以导入一个个案子集（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样本。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。

8. 单击下一步以继续。

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。

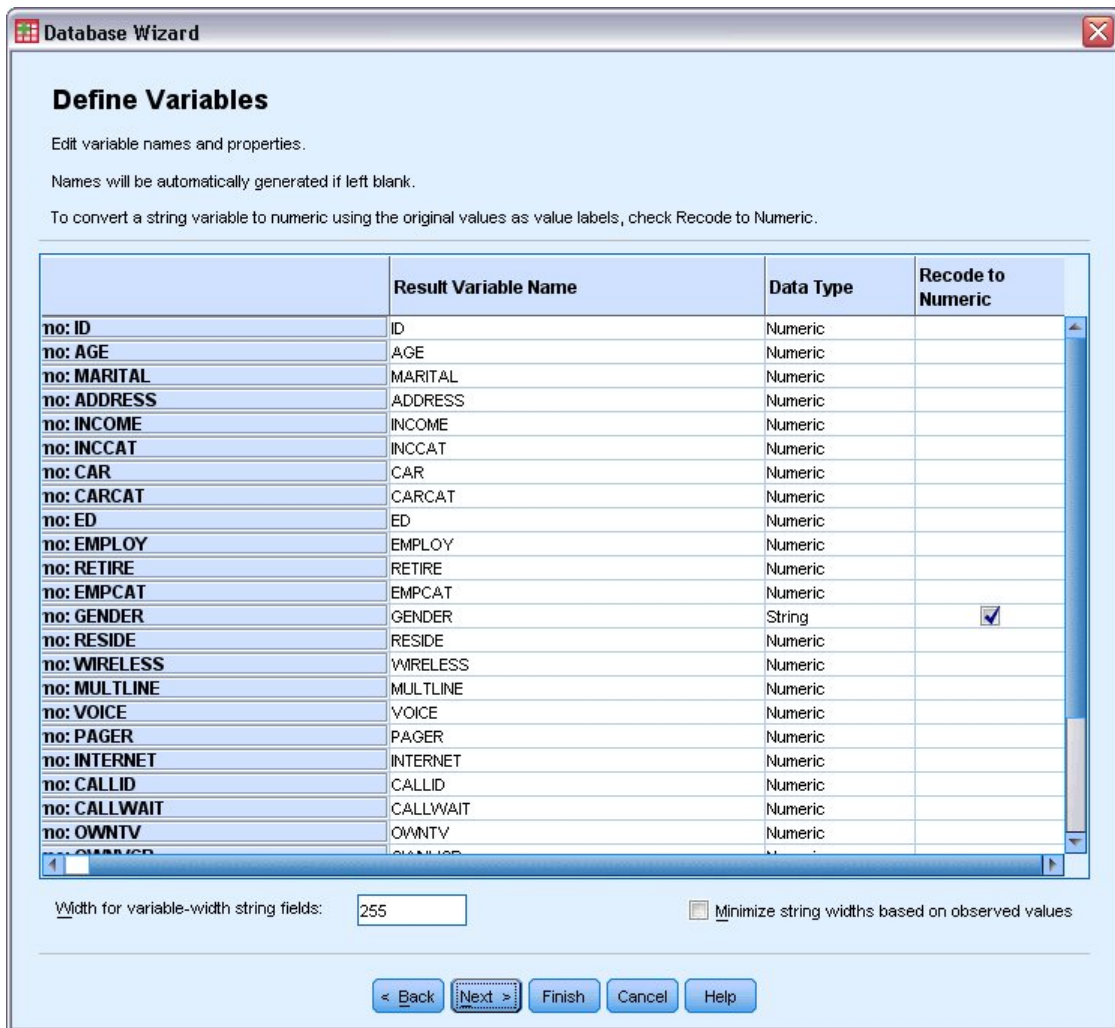


图 15: “定义变量”步骤

9. 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。
10. 单击**下一步**以继续。

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。

11. 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。

## 从文本文件读取数据

文本文件表示另一个公共数据源。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。在此示例中，数据是以制表符分隔的。

1. 从菜单中选择：  
文件 > 导入数据 > 文本数据
2. 转至 Samples\English 文件夹并选择 demo.txt。

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。



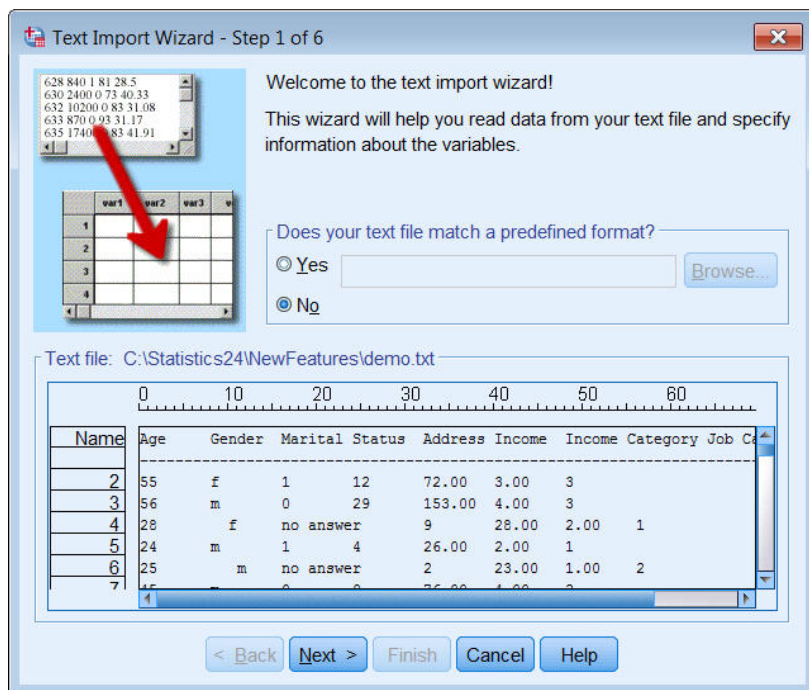


图 16: 文本导入向导: 第 1 步, 共 6 步

3. 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**无**。
4. 单击**下一步**以继续。

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。

5. 在向导的步骤 2 中, 选择**分隔**以指示文件具有分隔格式结构。
6. 选择**是**以指示文件在文件顶部包含变量名称。
7. 单击**下一步**以继续。
8. 在步骤 3 中, 输入 2 作为数据第一个个案的开始行号 (因为变量名称位于第一行)。
9. 将此步骤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文件被正确读取。

10. 选择**制表符**, 然后取消选择分隔符的其他选项。缺省情况下选择**空格**, 因为文件包含空格。对于此文件, 空格是数据值的一部分, 而非定界符。您需要取消选择**空格**以正确读取文件。
11. 选择**移除字符串值的前导空格**。字符串值开头的空格影响在表达式中对字符串值求值的方式。在此文件中, *Gender* 的某些值具有不属于值的前导空格。如果未移除这些空格, 那么值“ f”将被视为与“f”不同的值。

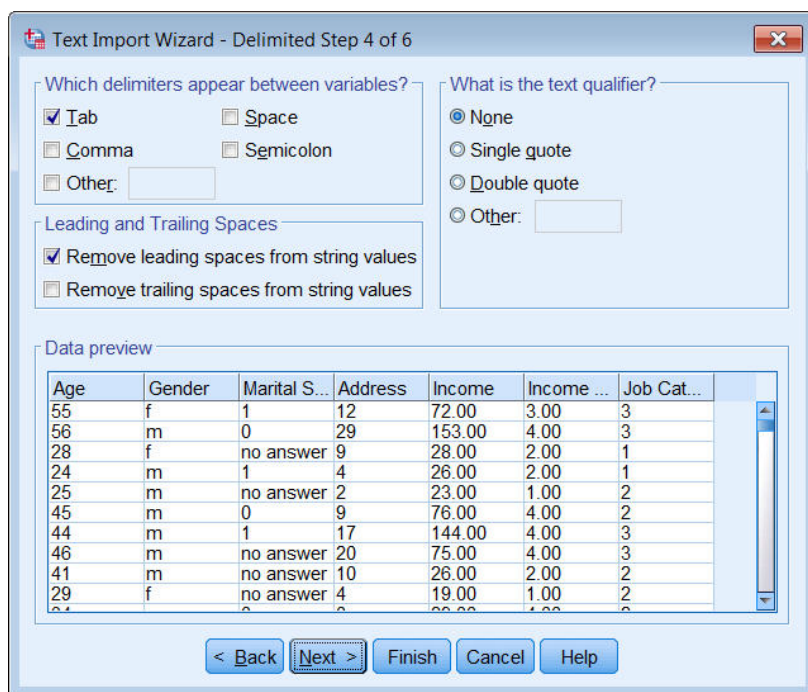


图 17: 文本导入向导: 第 4 步, 共 6 步

12. 单击下一步以继续。

由于为了符合命名规则, 变量名称可能已修改, 所以使用步骤 5 可以编辑任何不合意的名称。还可在此处定义数据类型。例如, 您可以将 *Income* 更改为美元货币格式。

要更改数据类型:

13. 在数据预览中, 选择 *Income*。

14. 从“数据格式”下拉列表中选择美元。

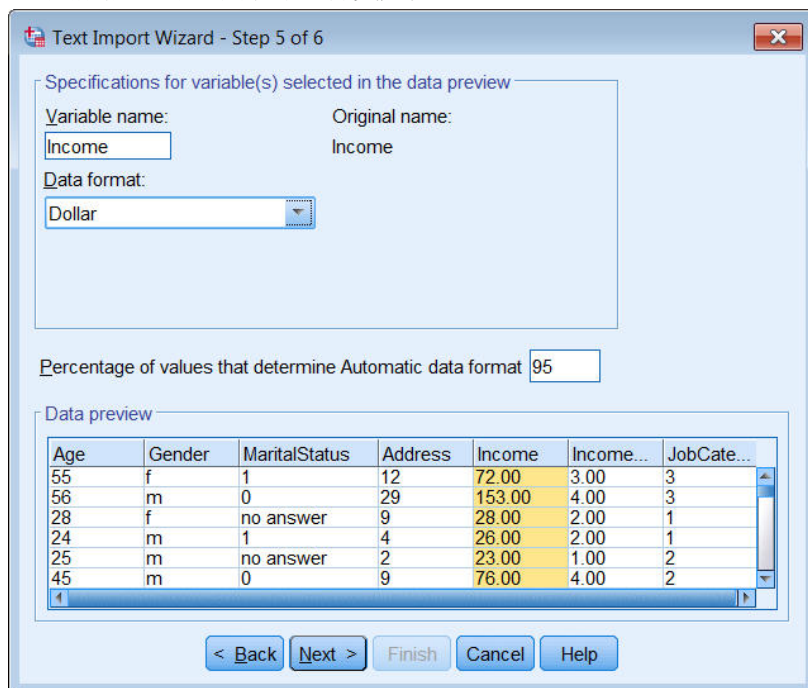


图 18: 更改数据类型

变量 *MaritalStatus* 包含字符串和数字值。不到 5% 的值是字符串。对于确定自动数据格式的百分比值使用缺省设置 95%, 变量将被视为数字值并且字符串值将设置为系统缺失。如果没有数据格式满足百

分比值，那么变量将被视为字符串变量。如果将设置更改为 100，那么将保留所有值，但是所有数字值将被视为字符串。

15. 单击**下一步**以继续。

16. 保留最后一步中的缺省选择，然后单击**完成**以导入数据。



## 第 3 章 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。

1. 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。在此情况下，只需以下三个变量：*age*、*marital status* 和 *income*。

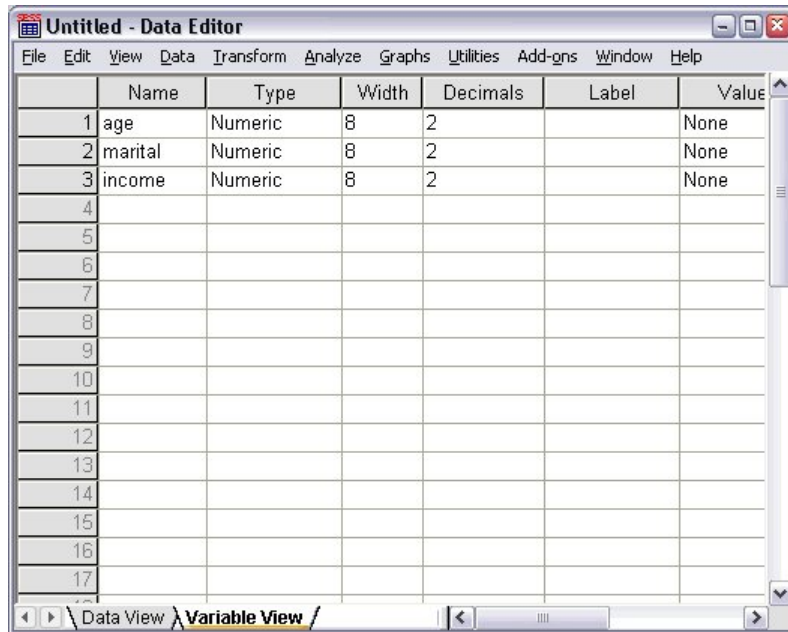


图 19: “变量视图”中的变量名称

2. 在第一列的第一行中，键入 *age*。
3. 在第二行中，键入 *marital*。
4. 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。

5. 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。

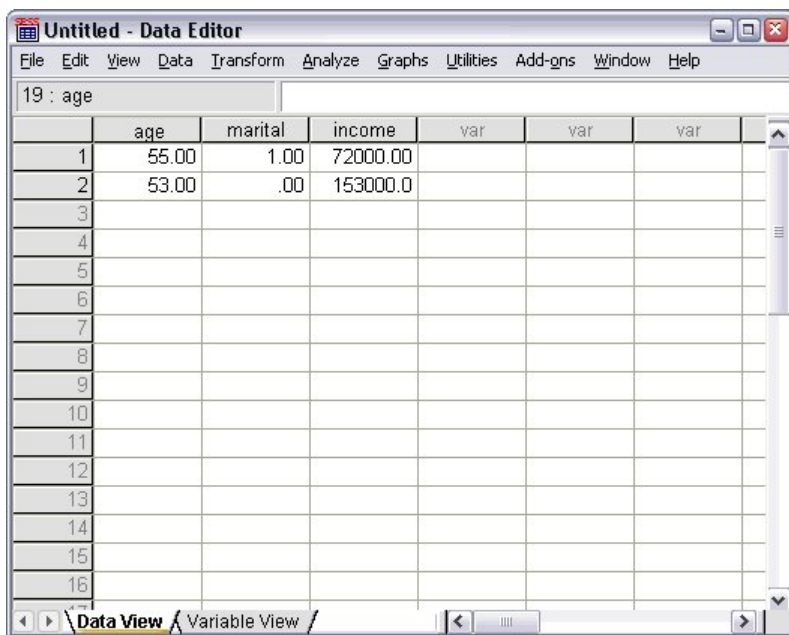


图 20: 在“数据视图”中输入的值

6. 在 *age* 列中，键入 55。
7. 在 *marital* 列中，键入 1。
8. 在 *income* 列中，键入 72000。
9. 将光标移动到第一列的第二行以添加下一个主体的数据。
10. 在 *age* 列中，输入 53。
11. 在 *marital* 列中，输入 0。
12. 在 *income* 列中，输入 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：

13. 单击“数据编辑器”窗口底部的**变量视图**选项卡。
14. 在 *age* 一行的小数列中，键入 0 以隐藏小数。
15. 在 *marital* 一行的小数列中，键入 0 以隐藏小数。

## 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）。

1. 单击“数据编辑器”窗口底部的**变量视图**选项卡。
2. 在第一个空行的第一个单元格中，键入 *sex* 作为变量名称。
3. 单击所输入内容后面的类型单元格。
4. 单击类型单元格右侧的按钮打开“变量类型”对话框。
5. 选择**字符串**以指定变量类型。
6. 单击**确定**保存所做的选择并返回到数据编辑器。

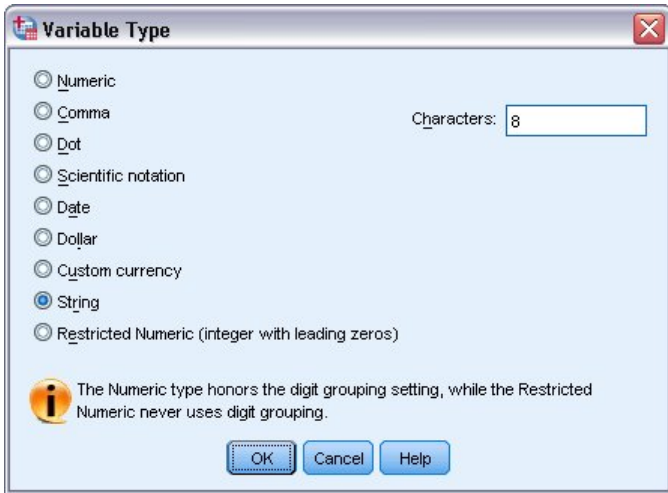


图 21: “变量类型”对话框

## 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。

1. 单击“数据编辑器”窗口底部的**变量视图**选项卡。
2. 在 *age* 一行的标签列中，键入 Respondent's Age。
3. 在 *marital* 一行的标签列中，键入 Marital Status。
4. 在 *income* 一行的标签列中，键入 Household Income。
5. 在 *sex* 一行的标签列中，键入 Gender。

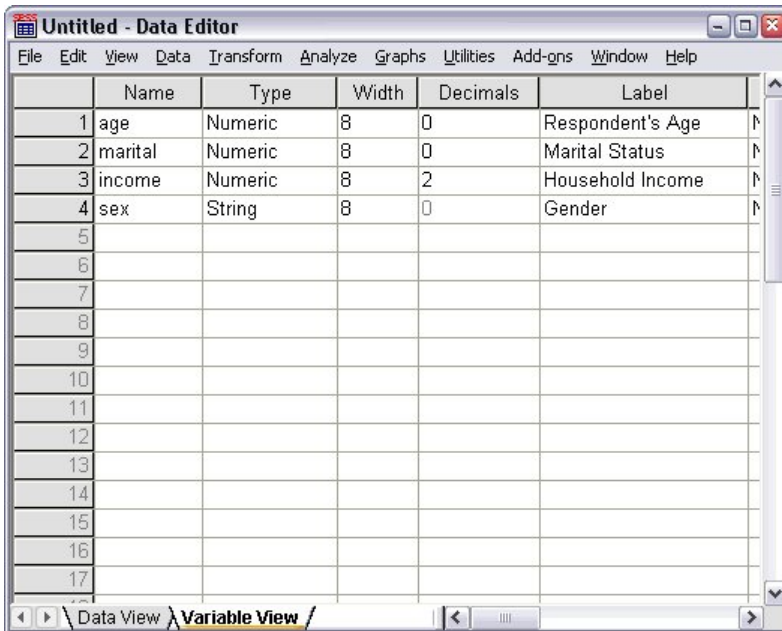


图 22: 在“变量视图”中输入的变量标签

## 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。

1. 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。
2. 选择美元。

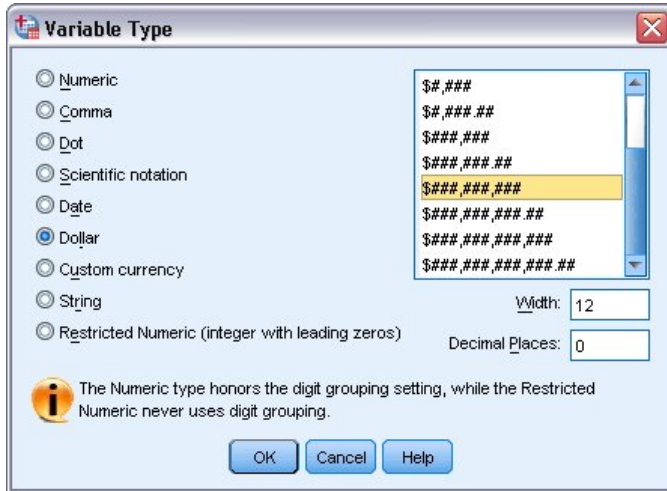


图 23: “变量类型”对话框

将显示当前所选数据类型的格式选项。

3. 对于本示例中的货币格式，请选择 **\$###,###,###**。
4. 单击**确定**以保存更改。

## 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。

1. 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。

2. 在“值”字段中键入 0。
3. 在“标签”字段中键入 Single。
4. 单击**添加**将此标签添加到列表中。

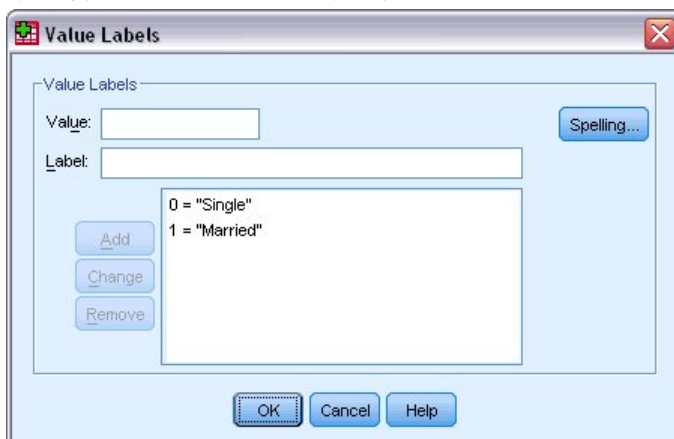


图 24: “值标签”对话框



5. 在“值”字段中键入 1，并在“标签”字段中键入 Married。
6. 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。
7. 单击“数据编辑器”窗口底部的**数据视图**选项卡。
8. 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。

## 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。

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

1. 单击“数据编辑器”窗口底部的**变量视图**选项卡。
2. 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 25: “缺失值”对话框

3. 选择**离散缺失值**。
4. 在第一个文本框中键入 999，并将另外两个文本框保留为空。
5. 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。
6. 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。
7. 在“值”字段中输入 999。
8. 在“标签”字段中键入 No Response。
9. 单击**添加**将此标签添加到数据文件中。
10. 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。

1. 单击“数据编辑器”窗口底部的**变量视图**选项卡。
2. 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。
3. 选择**离散缺失值**。
4. 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。

5. 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。

6. 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。
7. 在“值”字段中键入 NR。
8. 在“标签”字段中键入 No Response。
9. 单击**添加**将此标签添加到您的项目中。
10. 单击**确定**保存更改并返回到数据编辑器。

## 第 4 章 检查单个变量的汇总统计

该部分介绍简单的汇总测量以及变量的测量级别如何影响应使用的统计类型。我们将使用数据文件 *demo.sav*。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。

### 测量级别

不同的汇总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”。

**刻度。** 数据是按间隔或比例尺度进行测量的，其中数据值指示各值的顺序和值与值间的距离。例如，72,195 美元的薪金高于 52,398 美元的薪金，并且这两个值之间的差距为 19,797 美元。也称为**定量或连续数据**。

### 分类数据的汇总测量

对于分类数据，最典型的汇总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汇总测量。

“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。

1. 从菜单中选择：

分析 > 描述性统计 > 频率...

注：此功能需要 Statistics Base 选项。

2. 选择 *Owens PDA [ownpda]* 和 *Owens TV [owntv]* 并将其移动到“变量”列表中。

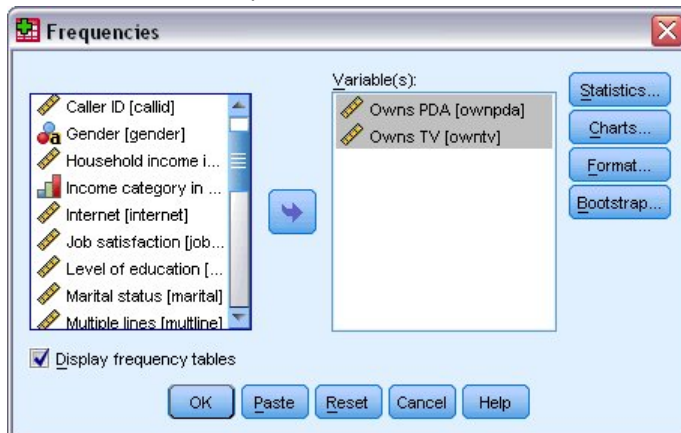


图 26: 要进行分析的所选分类变量

3. 单击**确定**以运行该过程。

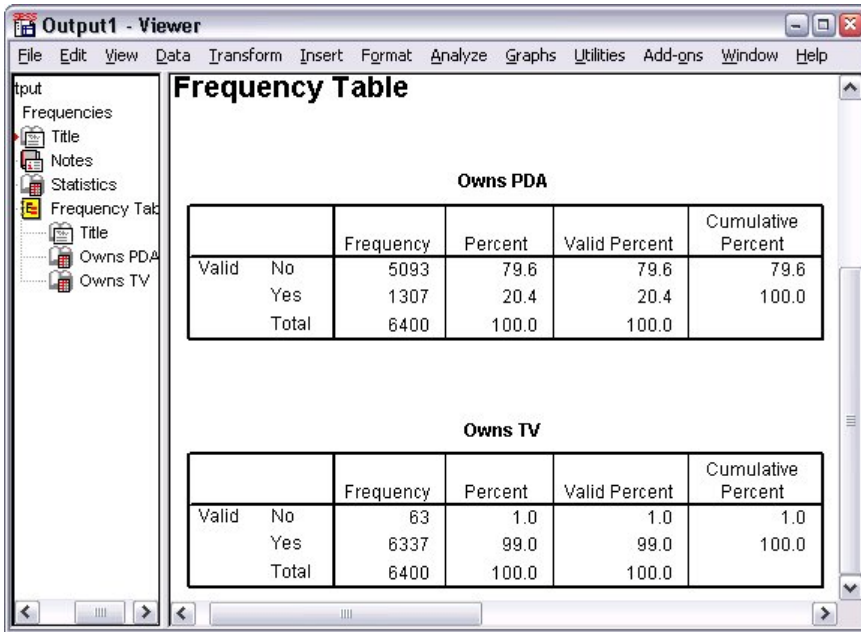


图 27: 频率表

频率表显示在查看器窗口中。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。

1. 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“对话框重新调用”按钮快速返回到最近用过的过程。



图 28: “对话框重新调用”按钮

2. 单击图表。
3. 选择条形图，然后单击继续。
4. 单击主对话框中的**确定**以运行该过程。

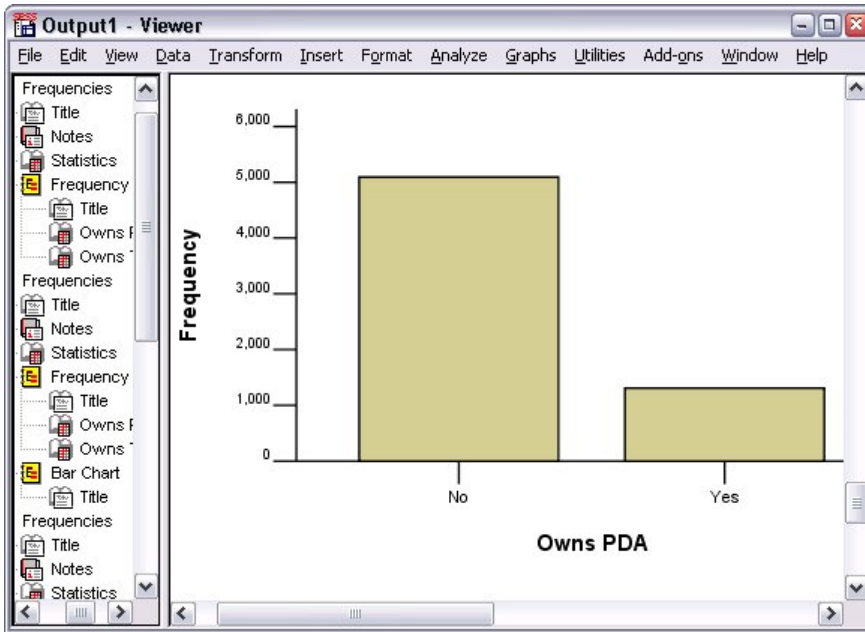


图 29: 条形图

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。

## 刻度变量的汇总测量

有许多可用于刻度变量的汇总测量，其中包括：

- **集中趋势的测量。** 最常见的集中趋势测量是**平均值**（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，包括标准差、最小值和最大值。
  1. 再次打开“频率”对话框。
  2. 单击**重置**清除以前的所有设置。
  3. 选择 *Household income in thousands [income]* 并将其移到“变量”列表中。
  4. 单击**统计**。
  5. 选择**平均值、中位数、标准差、最小值和最大值**。
  6. 单击**继续**。
  7. 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。
  8. 单击**确定**以运行该过程。

“频率统计”表将显示在查看器窗口中。

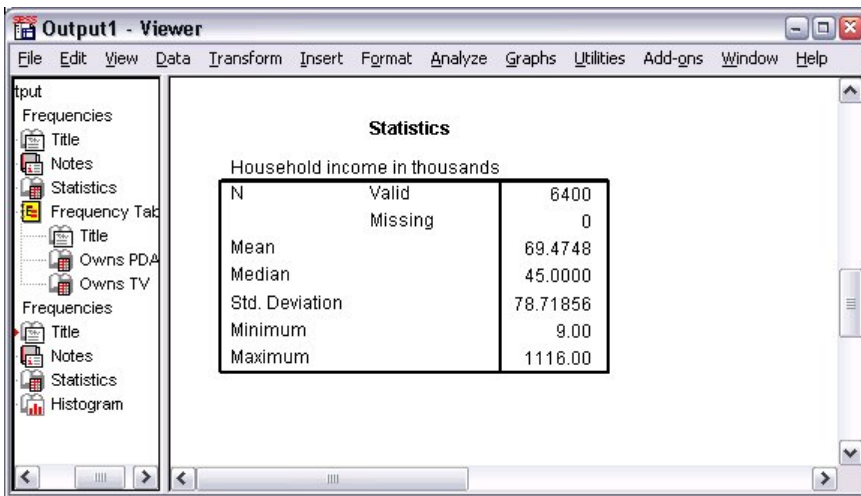


图 30: “频率统计”表

在此示例中，平均值和中值之间具有较大的差异。平均值几乎比中值大 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

1. 再次打开“频率”对话框。
2. 单击图表。
3. 选择直方图和带正态曲线。
4. 单击继续，然后单击主对话框中的**确定**以运行该过程。

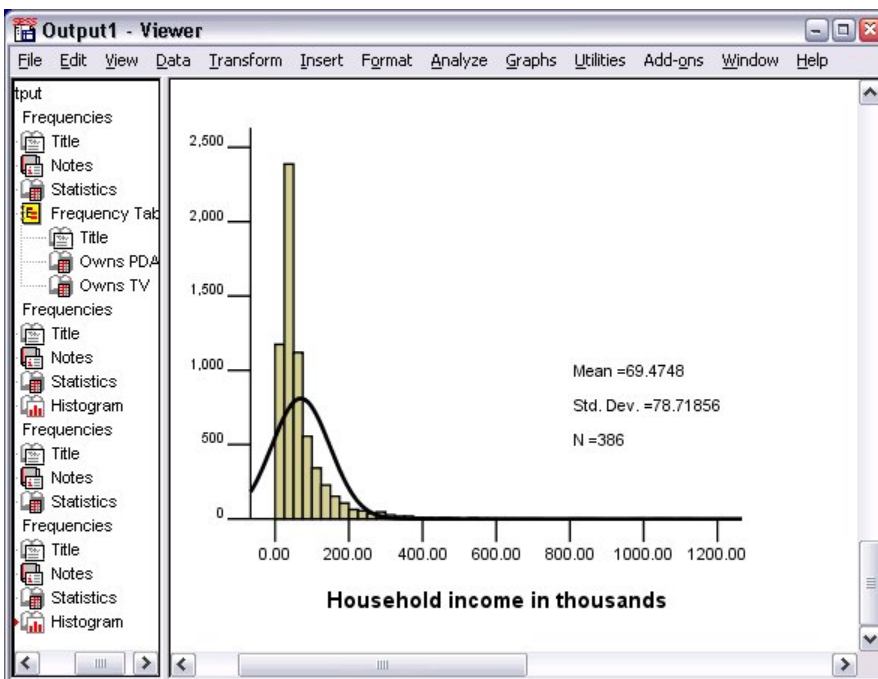


图 31: 直方图

大多数案例集中在规模的低端，大多数下降到 100,000 以下。然而，一些案例在 500,000 范围内及以上（数量太少，在不修改直方图的情况下甚至都看不到）。这些少数个案的较高值对平均值具有显著作用，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。

## 第 5 章 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*。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。

1. 从菜单中选择：

图形 > 图表构建器...

“图表构建器”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。

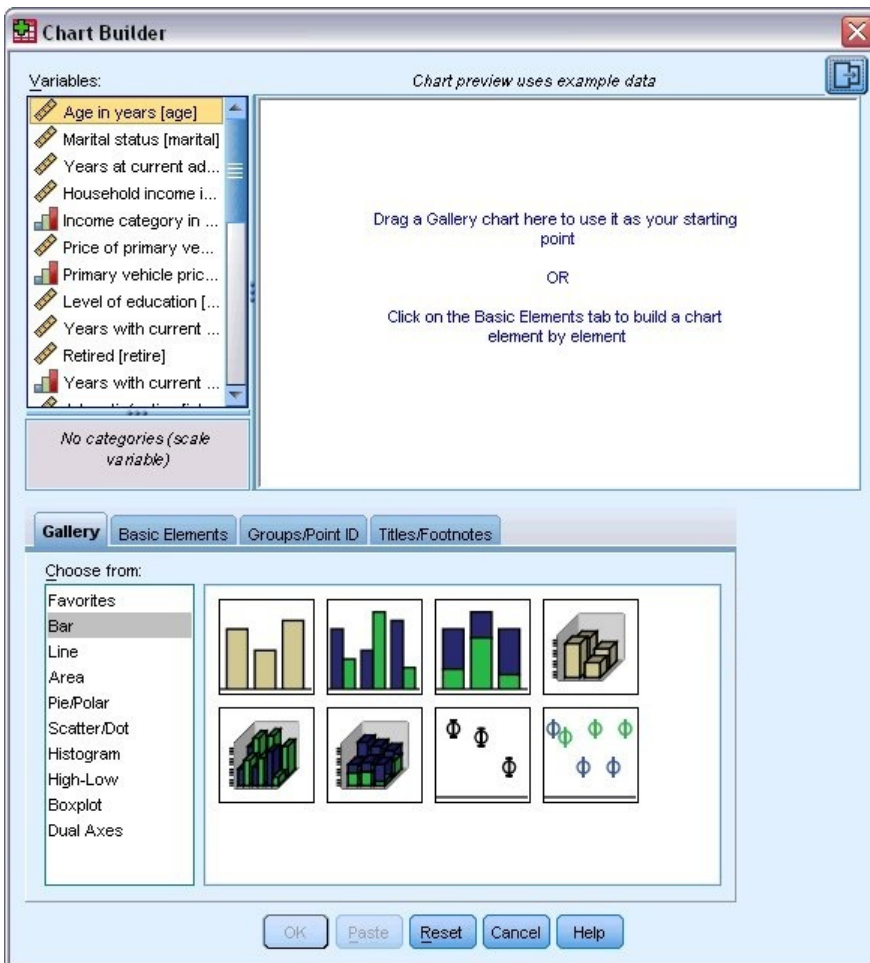


图 32: “图表构建器”对话框

### 使用图表构建器库

1. 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。

2. 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，还可以通过将光标悬停在图标上来显示图表的工具提示描述。

3. 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。“图表构建器”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。

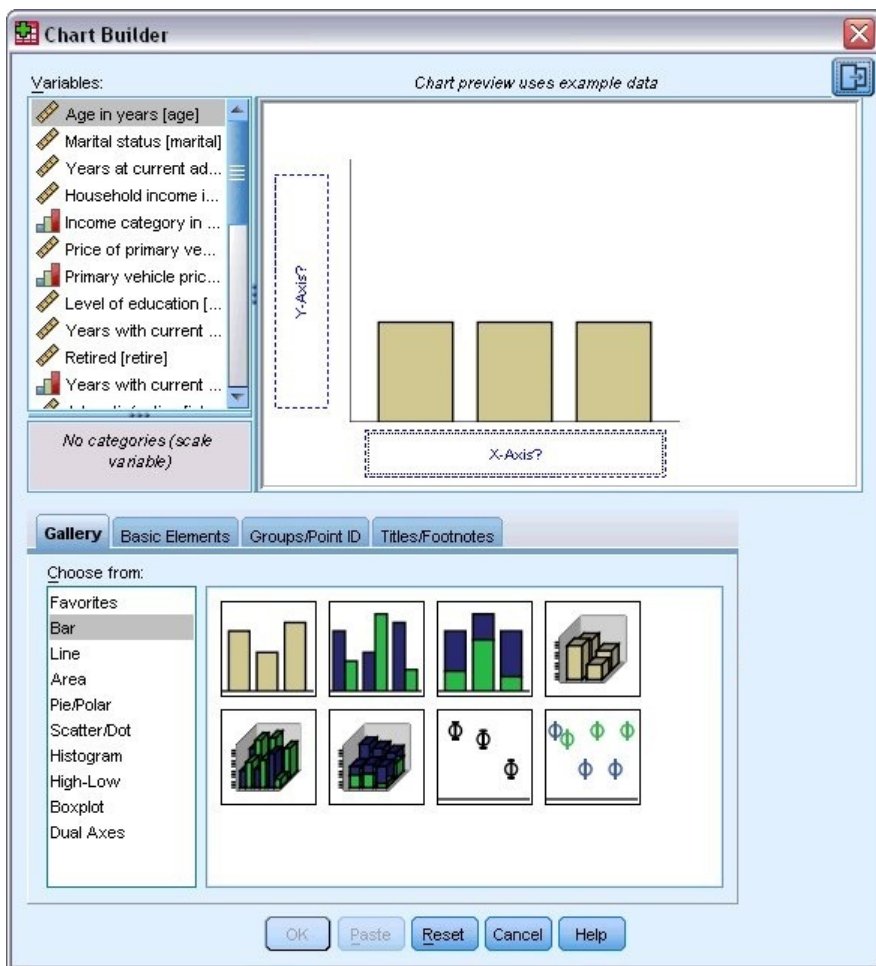


图 33: 图表构建器画布上的条形图

## 定义变量和统计

尽管画布上有图表，但图表并不完整，因为没有变量或统计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。从“变量”列表中拖出变量以添加它们，该列表位于画布左侧。

在图表构建器中，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，可以暂时在图表构建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。

1. 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。
2. 现在，将 *Job satisfaction* 从变量列表拖到 x 轴放置区。  
y 轴放置区缺省为计数统计。如果要使用另一种统计（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，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。
3. 单击图表构建器侧边栏中的**元素属性**选项卡。（如果未显示该侧边栏，单击图表构建器右上角的按钮可显示该侧边栏）。





图 34: 元素属性

“元素属性”使您能够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”下拉列表显示可用的具体统计。对于每一种图表类型，可用的统计通常都相同。请注意，某些统计要求 y 轴放置区包含变量。

4. 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 轴放置区缺省为平均值统计。这是您希望的变量和统计，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。

1. 单击**标题/脚注**选项卡。
2. 选择**标题 1**。

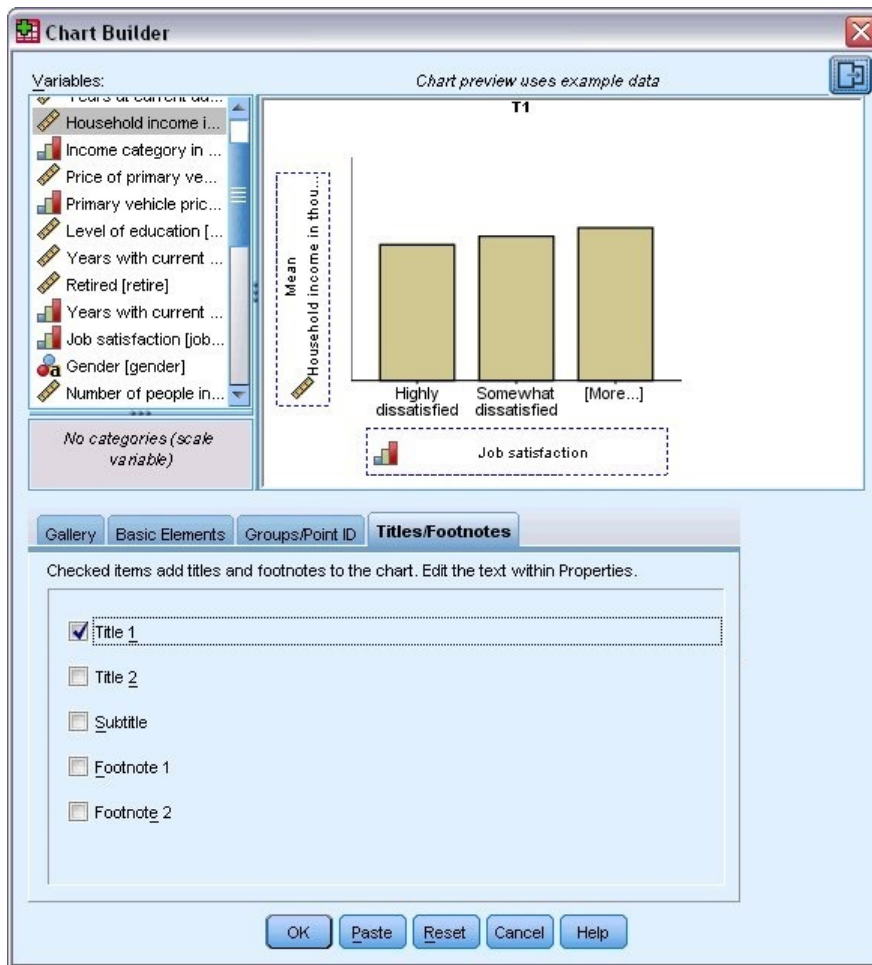


图 35: 标题 1 显示在画布上

该标题出现在画布上，并带有标签 **T1**。

3. 在**元素属性**选项卡中，选择“编辑属性”列表中的**标题 1**。
4. 在“内容”文本框中，键入 `Income by Job Satisfaction`。这是该标题将显示的文字。

## 创建图表

1. 单击**确定**，创建条形图。

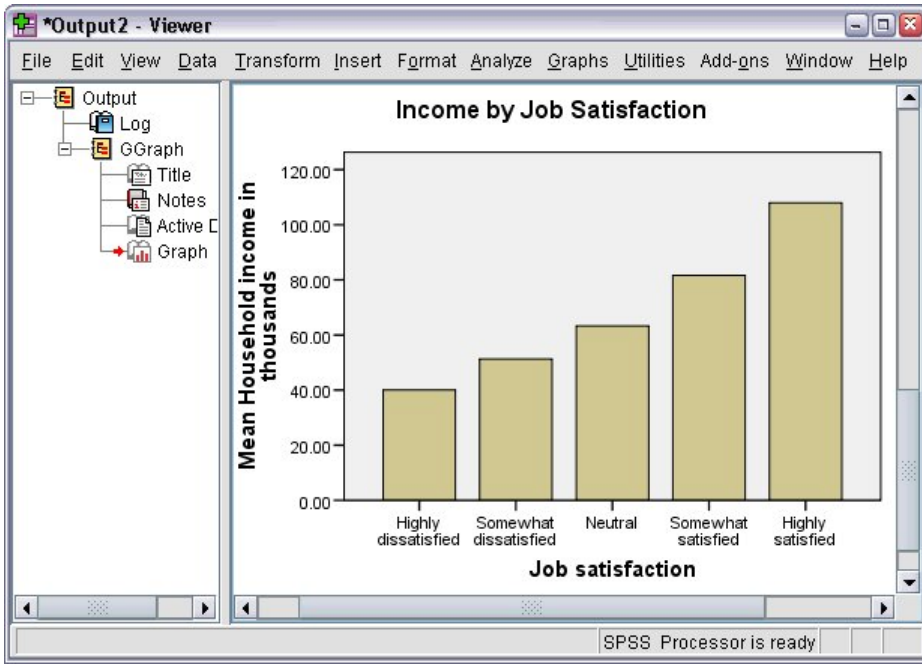


图 36: 条形图

从条形图中可以看出，响应者对其工作越满意，其家庭收入也越高。



## 第 6 章 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*。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。

### 使用查看器

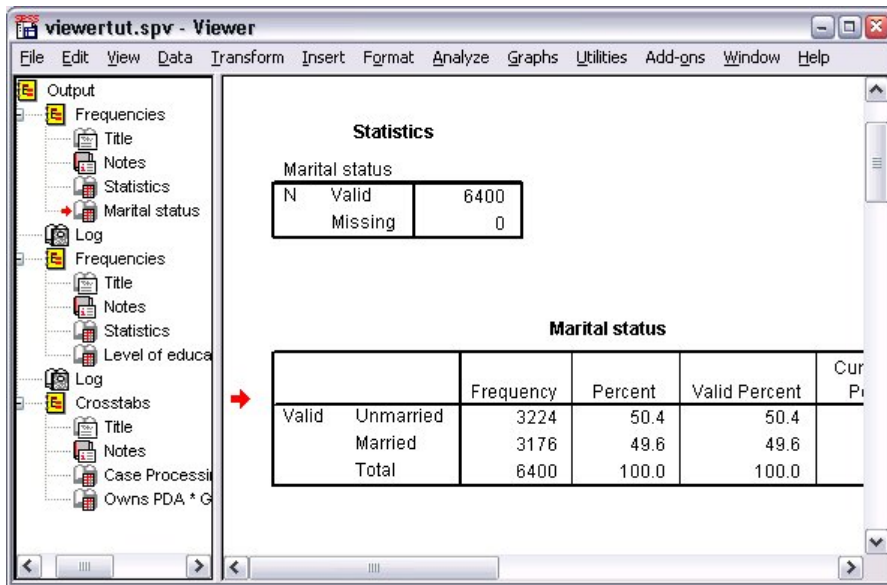


图 37: 查看器

“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，可将其显示在内容窗格中。

1. 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。

2. 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。

3. 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。

4. 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，可视化这些结果已隐藏。

您也可以更改输出的显示顺序。

5. 在概要窗格中，单击要移动的项。
6. 将选中的项拖到概要窗格中的新位置。

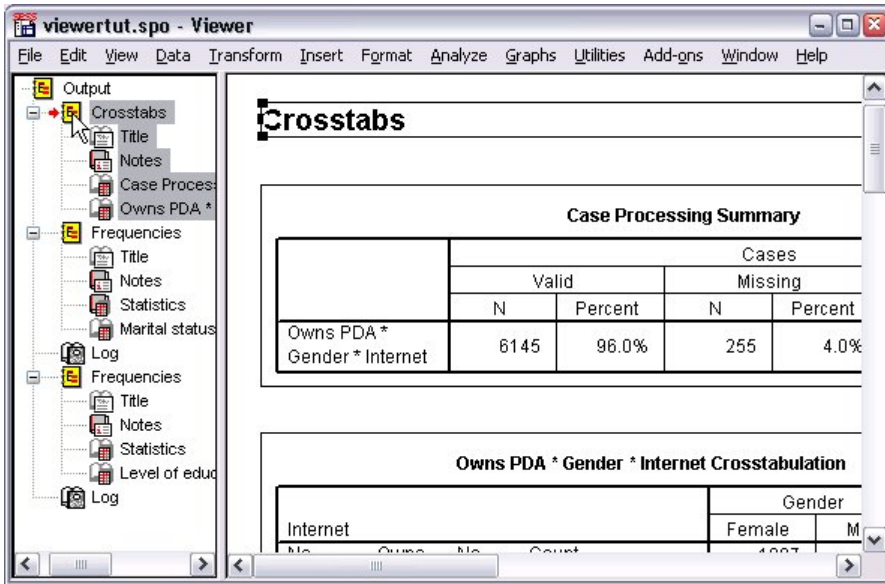


图 38: 重新排序查看器中的输出

通过单击输出项并将其拖至内容窗格，也可移动输出项。

## 使用透视表编辑器

大多数统计过程的结果都显示在透视表中。

### 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。

1. 双击 *Owns PDA \* Gender \* Internet Crosstabulation* 表。
2. 右键单击期望计数，然后从弹出菜单中选择**这是什么？**。

相关定义将显示在弹出窗口中。

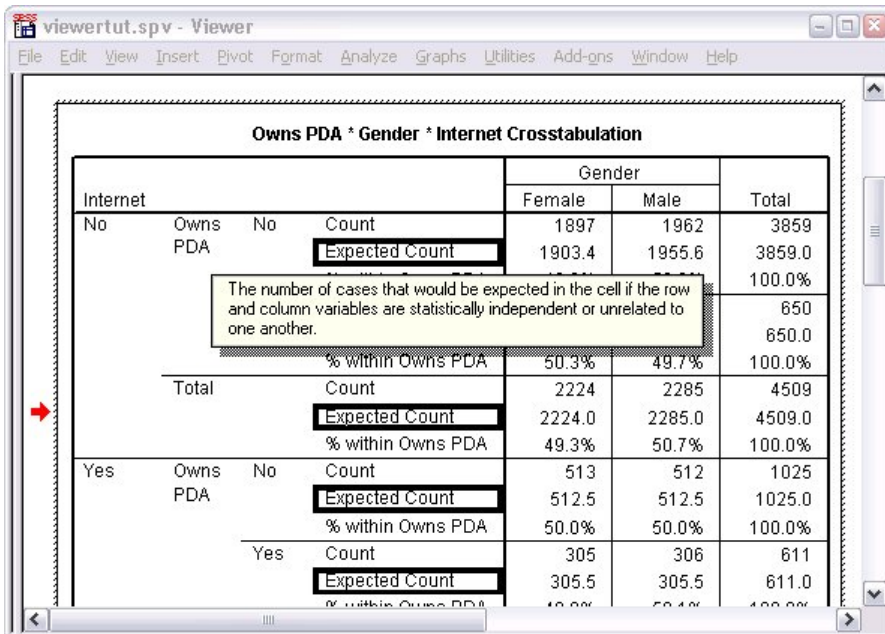


图 39: 弹出定义

## 透视表

生成的缺省表显示的信息可能不像您期望的那样美观或清晰。使用透视表，您可以变换行和列（“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，通过变换行和列，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。

1. 如果其尚未激活，双击 *Owns PDA \* Gender \* Internet Crosstabulation* 表以激活它。
2. 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。

Internet			Gender		Total	
			Female	Male		
No	Owns PDA	No	Count	1897	1962	3859
		Yes	Expected Count	1903.4	1955.6	3859.0
	Total	No	% within Owns PDA	49.2%	50.8%	100.0%
		Yes	Count	327	323	650

图 40: 透视托盘

3. 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。透视托盘中的元素顺序反映了表中各元素的顺序。
4. 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。

			Gender						Total	
			Female			Male			Count	Expected Count
Internet	Count	Expected Count	% within Owns PDA	Count	Expected Count	% within Owns PDA	Count	Expected Count		
Owns PDA	No	11	1025.0	100.0%	1025.0	100.0%	650.0	100.0%	611.0	100.0%
	Yes	22	4509.0	100.0%	1636.0	100.0%	650.0	100.0%	611.0	100.0%
Total	No	22	4509.0	100.0%	1636.0	100.0%	650.0	100.0%	611.0	100.0%
	Yes	6	1025.0	100.0%	1025.0	100.0%	650.0	100.0%	611.0	100.0%

图 41: 切换行

## 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。

1. 将 *Gender* 元素从列维度拖至层维度。

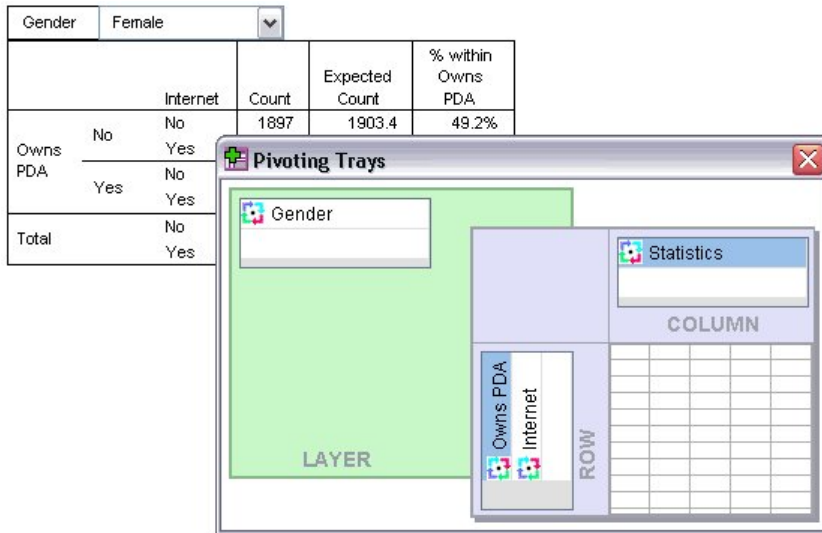


图 42: 层维度中的 *Gender* 透视图标

为显示不同层，从表中的下拉列表选择类别。

## 编辑表

除非您花时间创建自定义的表外观，否则透视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。

1. 双击 *Level of education* 表。
2. 如果格式设置工具栏不可见，请从菜单中选择：  
视图 > 工具栏
3. 单击标题文本 *Level of education*。
4. 从工具栏上字体大小下拉列表中选择 **12**。
5. 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。

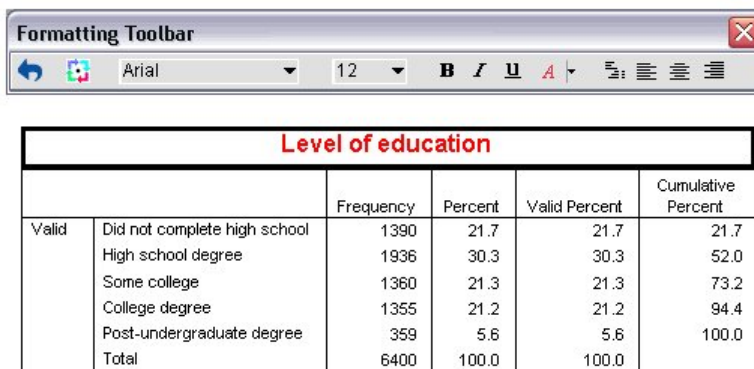


图 43: 重新设置透视表中标题文字的格式

也可以编辑表和标签的内容。例如，可以更改此表的标题。

6. 双击该标题。



7. 键入 *Education Level* 作为新的标签。

注：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无用，或者可能使表不必要地复杂化。不过您可以在不丢失任何数据的情况下隐藏整个行和列。

1. 如果其尚未激活，双击 *Education Level* 表以激活它。
2. 单击 *Valid Percent* 列标签以选中该列。
3. 从 **编辑** 菜单或右键单击弹出菜单中，选择：  
**选择 > 数据和标签单元格**
4. 从 **视图** 菜单中，选择 **隐藏**，或者从右键单击弹出菜单中选择 **隐藏类别**。

现在，该列已隐藏，但并没有删除。

**Education Level**

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

图 44: 表中 *Valid Percent* 列已隐藏

5. 要重新显示该列，请双击同一表。
6. 单击 **查看**，然后选择 **显示所有目录**。

行也可以隐藏并以与列相同的方式重新显示。

## 更改数据显示格式

可以轻松更改透视表中数据的显示格式。

1. 如果其尚未激活，双击 *Education Level* 表以激活它。
2. 单击 *Percent* 列标签以选中该列。
3. 从“编辑”菜单或右键单击弹出菜单中选择：  
**选择 > 数据单元格**
4. 从格式菜单或右键单击弹出菜单中选择 **单元格属性**。
5. 单击 **格式化值** 选项卡。
6. 在“小数”字段中键入 0 以隐藏该列中的所有小数点。

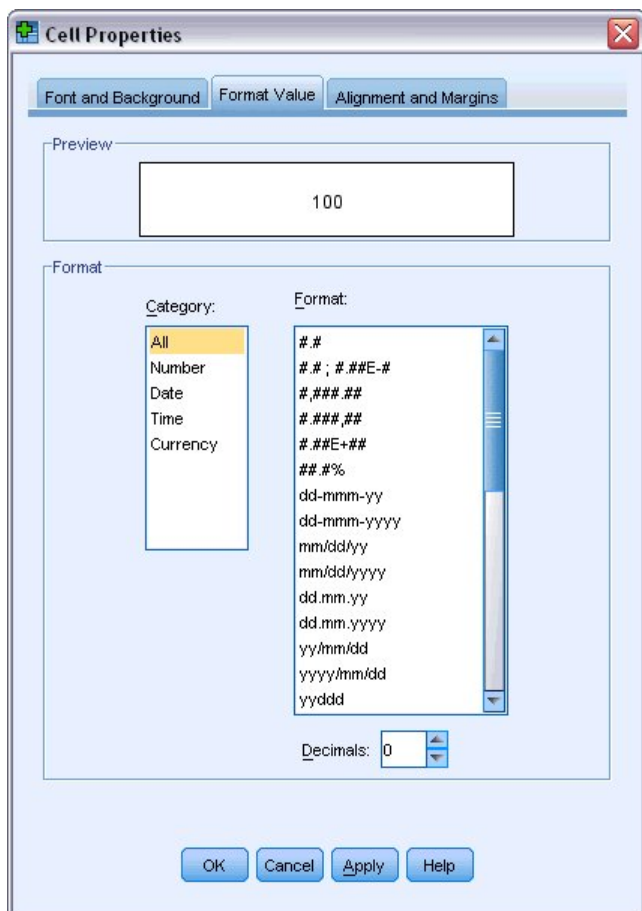


图 45: 单元格属性，格式值选项卡

在此对话框中还可以更改数据类型和格式。

7. 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。
8. 单击**确定**或**应用**以应用您的更改。

**Education Level**

		Frequency	Percent	Cumulative Percent
Valid	Did not complete high school	1390	22	21.7
	High school degree	1936	30	52.0
	Some college	1360	21	73.2
	College degree	1355	21	94.4
	Post-undergraduate degree	359	6	100.0
	Total	6400	100	

图 46: “Percent”列中的小数已隐藏

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

1. 双击 *Marital status* 表。
2. 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“样本”窗口中对其进行预览。

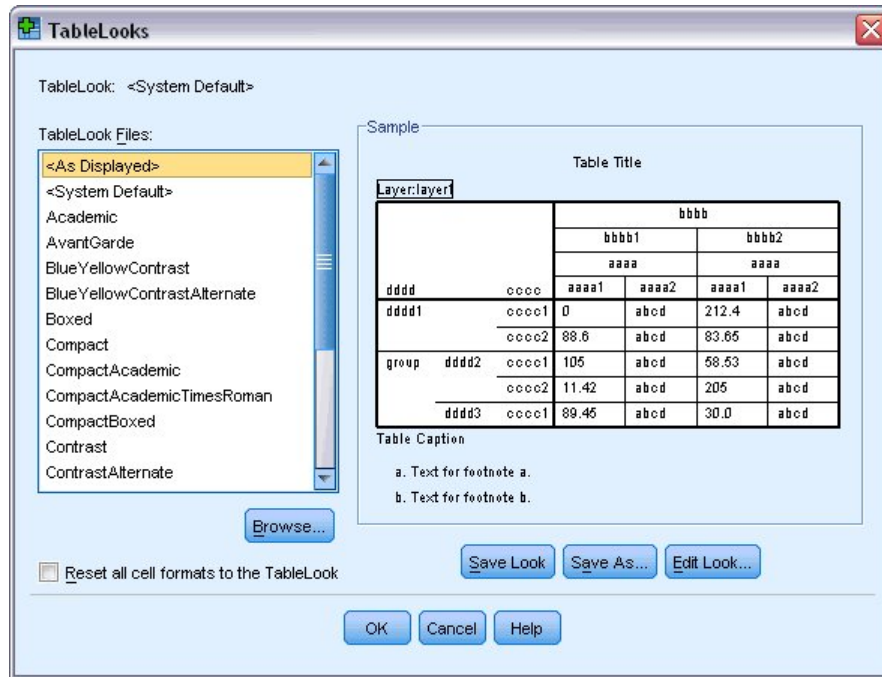


图 47: “表格外观”对话框

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。

3. 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定制格式。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。

1. 双击 *Marital status* 表。
2. 从菜单中选择：

格式 > 表格外观...

3. 选择与您所需格式最接近的样式，然后单击**编辑外观**。
4. 单击**单元格格式**选项卡以查看格式设置选项。

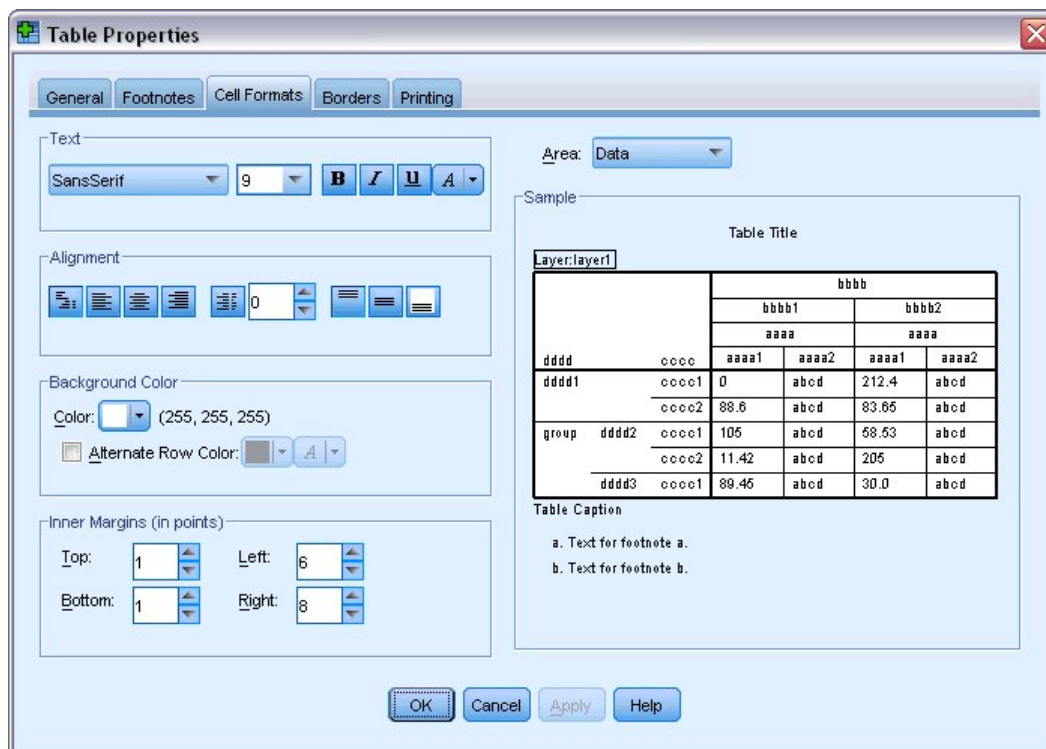


图 48: “表格属性”对话框

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“样本”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“样本”窗口单击要更改的区域。

5. 从“区域”下拉列表中选择**数据**。
6. 从背景下拉调色板中选择新的颜色。
7. 然后选择一种新的文本颜色。

此时，“样本”窗口将显示新的样式。

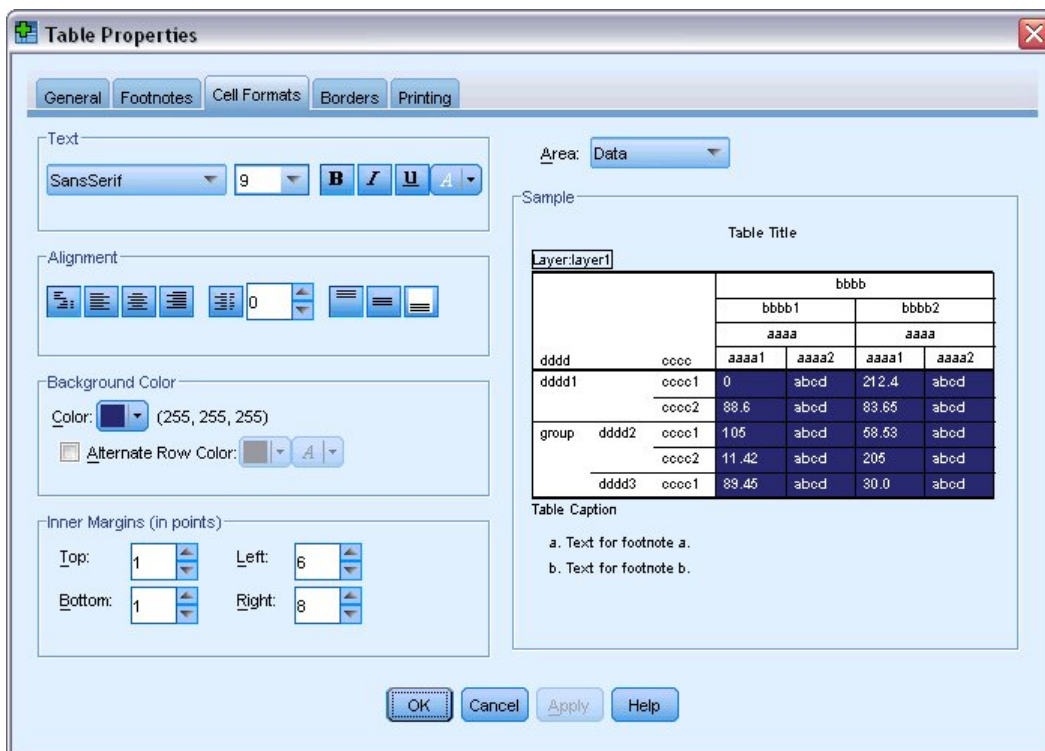


图 49: 更改表格单元格格式

8. 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。

9. 单击**另存为**。

10. 导航到目标目录，然后在“文件名”文本框中输入新样式的名称。

11. 单击**保存**。

12. 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。

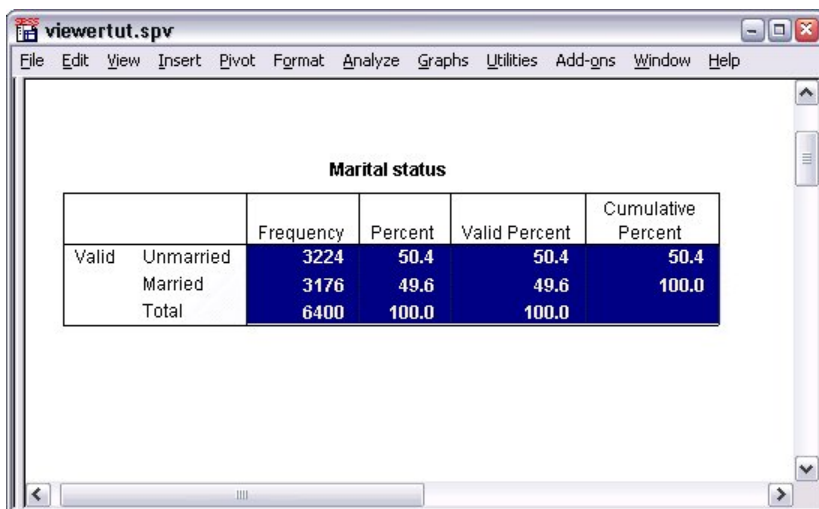


图 50: 自定义表格外观

## 更改缺省的表格格式

尽管在创建表之后可以更改其格式，但更改缺省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。

要更改透视表的缺省表格外观样式，请从菜单中选择：

编辑 > 选项...

1. 单击“选项”对话框中的透视表选项卡。

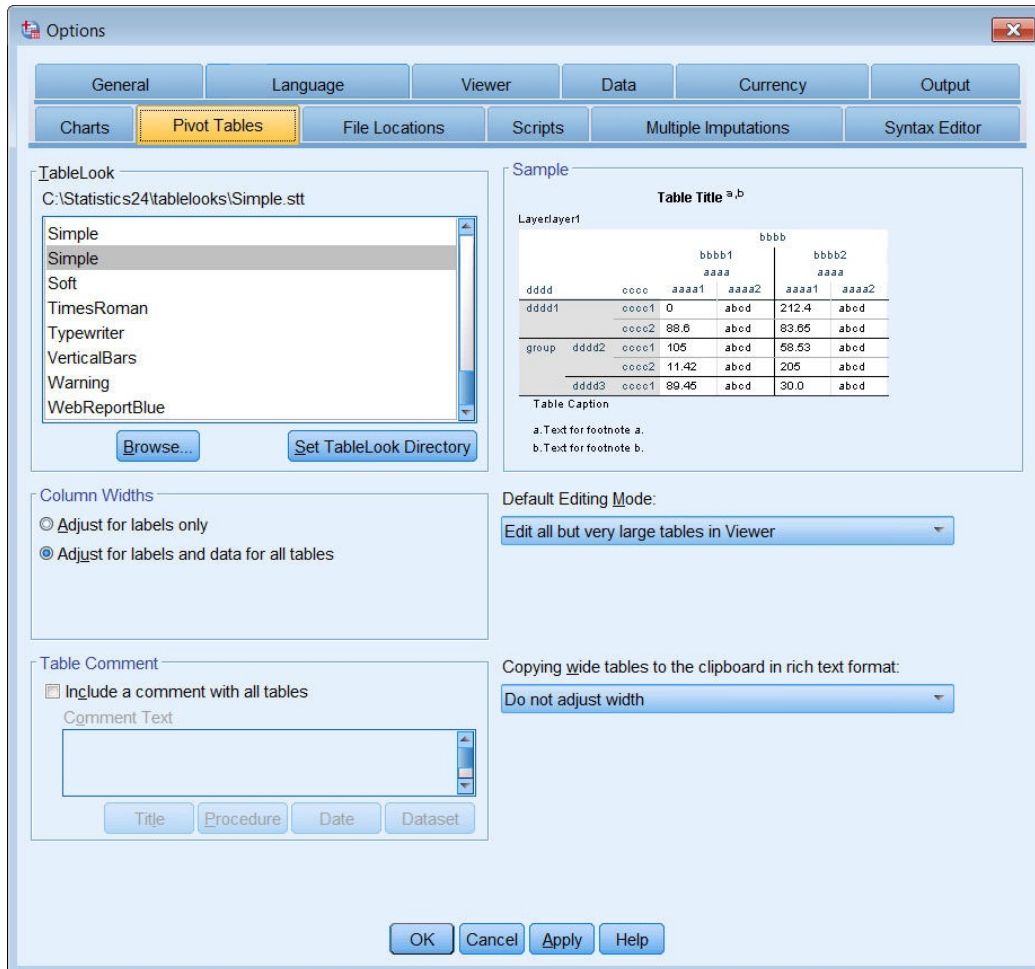


图 51: “选项”对话框

2. 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。

3. 单击确定保存设置并关闭该对话框。

在更改缺省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，缺省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：

1. 从菜单中选择：

编辑 > 选项...

2. 单击查看器选项卡。

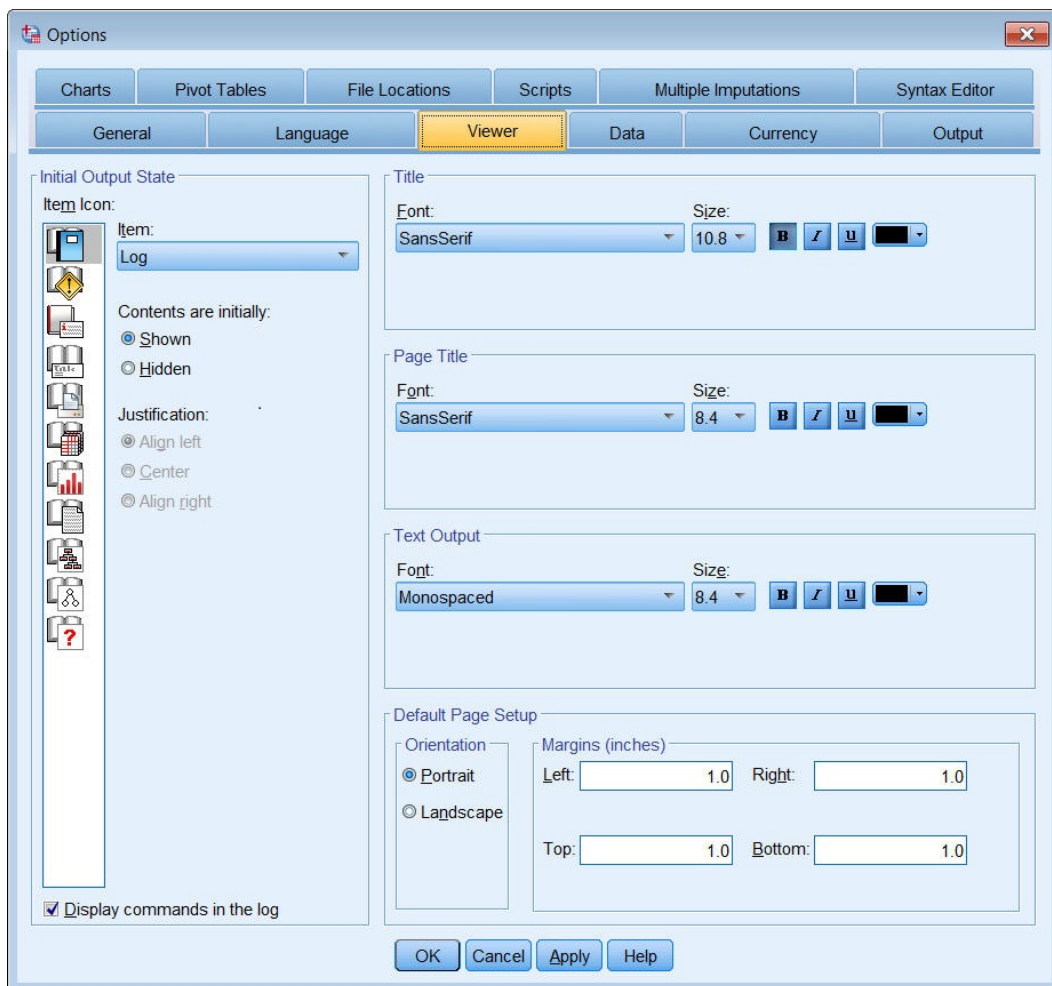


图 52: 查看器选项

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。只需选中要定制的对象，然后进行更改即可。

3. 单击**标题**图标以显示其设置。
4. 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。

5. 双击**警告**图标可在输出中隐藏警告消息。
6. 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。

1. 从菜单中选择：  
编辑 > 选项...
2. 单击**输出标签**选项卡。



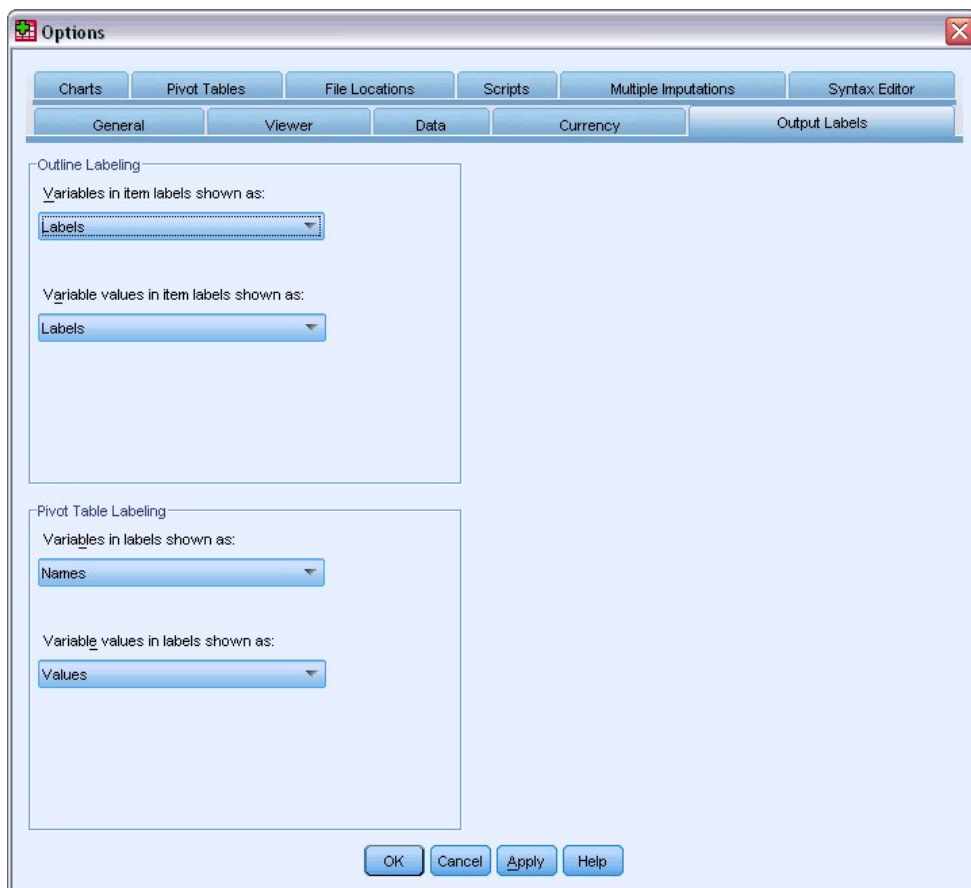


图 53: 透视表标签设置

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。

3. 在“透视表标签”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。
4. 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。

在会话中生成的后续表将反映这些更改。

marital					
		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	

图 54: 显示的变量名和值

## 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

可以将透视表作为 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。

1. 在“查看器”中单击表将其选中。
2. 从菜单中选择：

### 编辑 > 复制

3. 打开字处理应用程序。
4. 从字处理程序的菜单中选择：

### 编辑 > 选择性粘贴...

5. 在“选择性粘贴”对话框中，选择**带格式文本(RTF)**。
6. 单击**确定**将结果粘贴到当前文档中。

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

透视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。

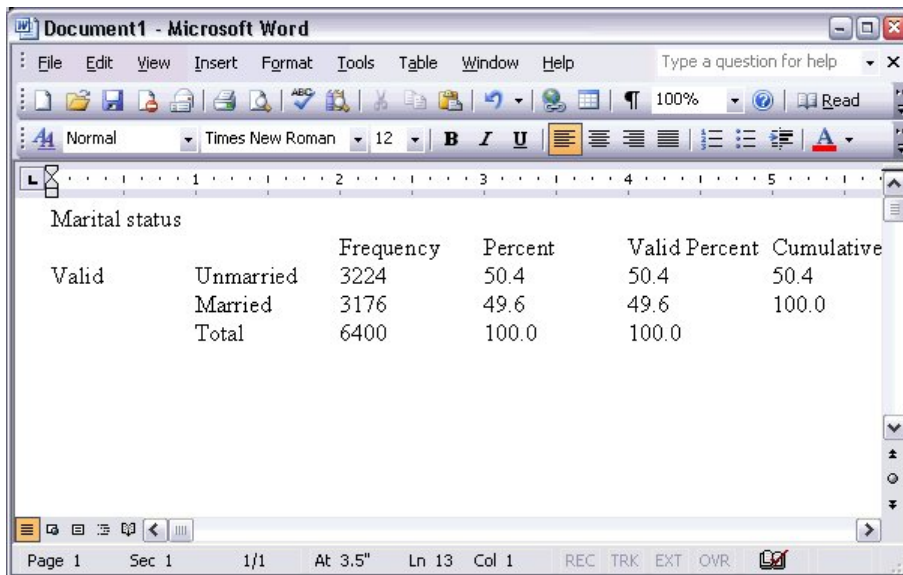
1. 在“查看器”中单击表将其选中。
2. 从菜单中选择：

### 编辑 > 复制

3. 打开字处理应用程序。
4. 从字处理程序的菜单中选择：

### 编辑 > 选择性粘贴...

5. 在“选择性粘贴”对话框中，选择**无格式文本**。
6. 单击**确定**将结果粘贴到当前文档中。



The screenshot shows a Microsoft Word window titled 'Document1 - Microsoft Word'. The ribbon is set to 'Table' and the 'Table' menu is open. A pivot table is displayed in the document area. The table has the following data:

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

图 55: Word 中显示的透视表

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*。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。

注：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，可以选择要导出的特定项或导出所有项或所有可见项。

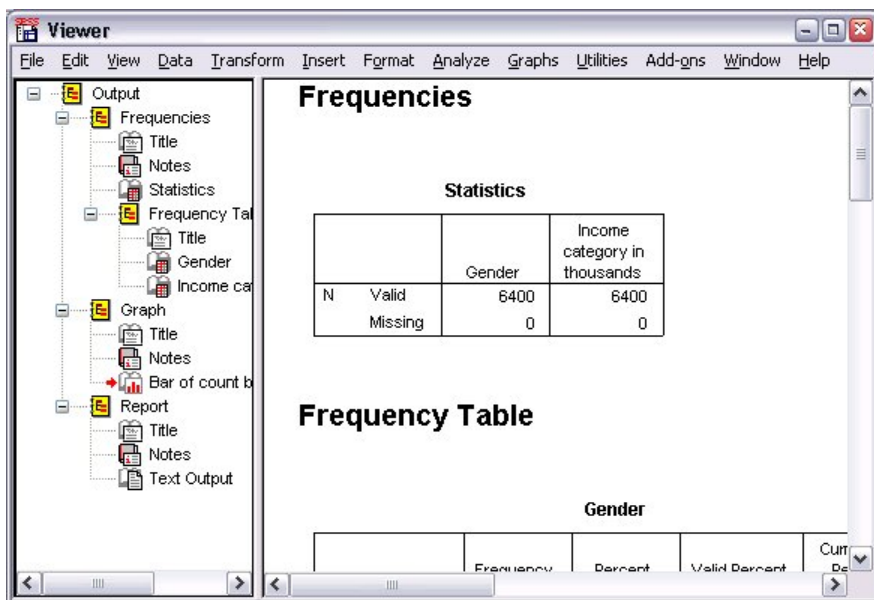


图 56: 查看器

1. 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。

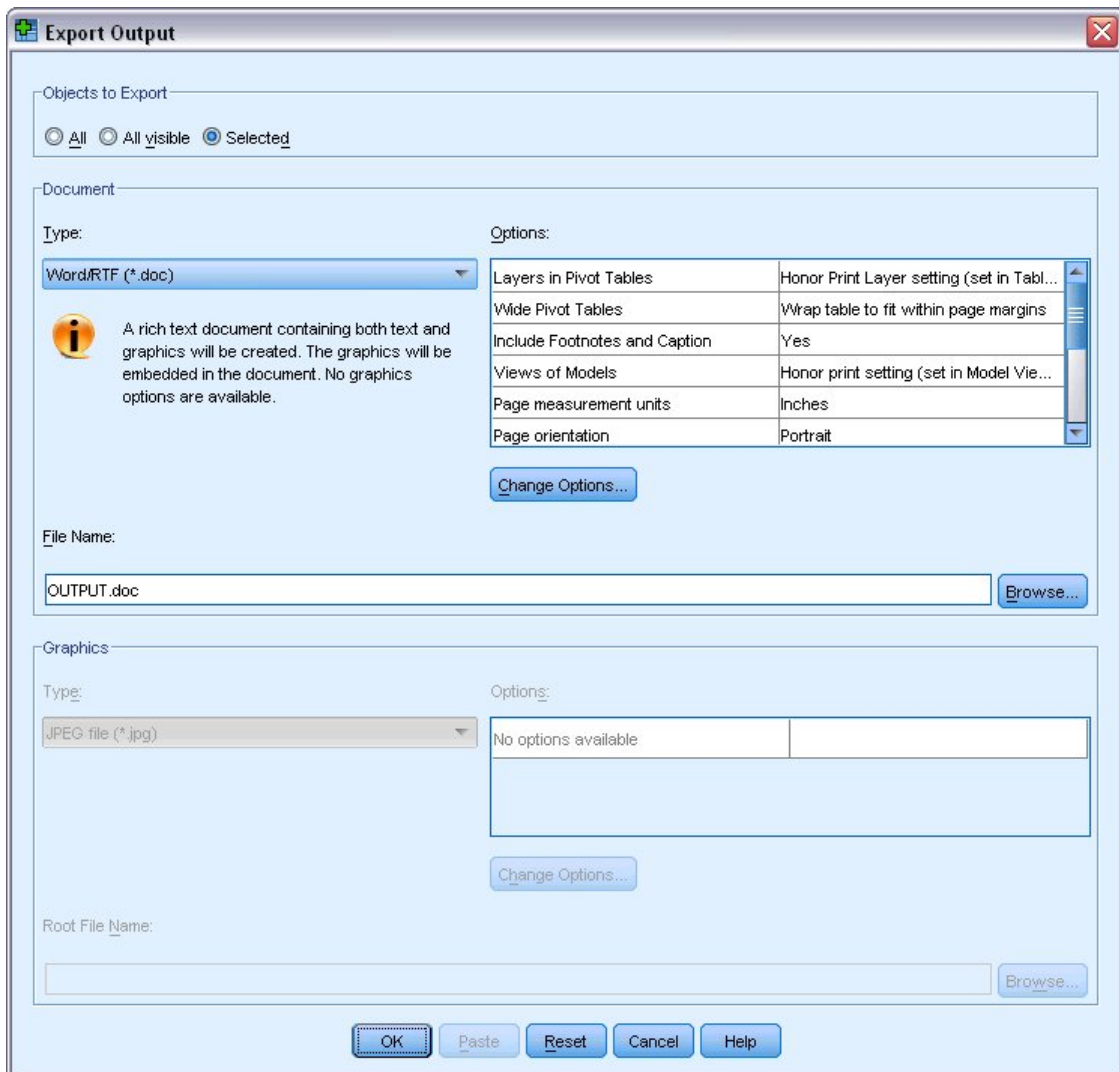


图 57: “导出输出”对话框

2. 在导出对象组中，选择**所有**。
3. 从“类型”下拉列表中选择 **Word/RTF 文件 (\*.doc)**。
4. 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。

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

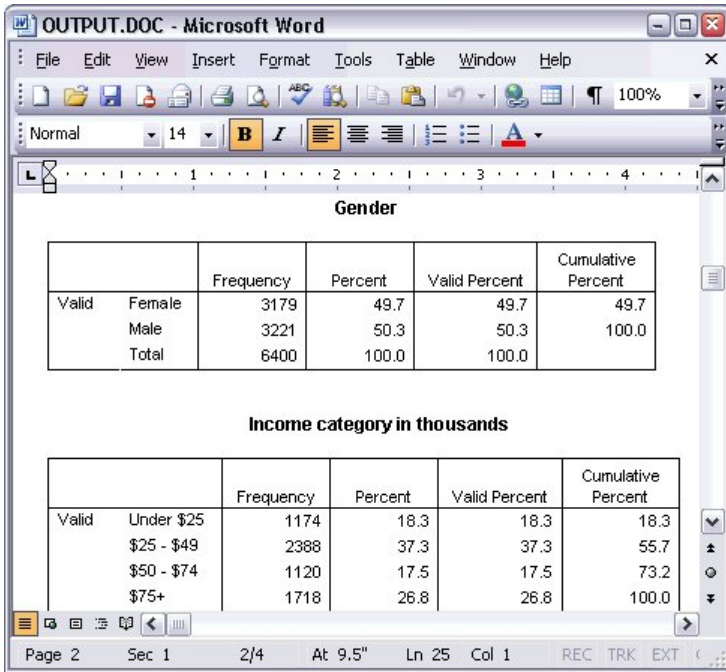


图 58: Word 中的透视表

图标作为图形图像包含在 Word 文档中。

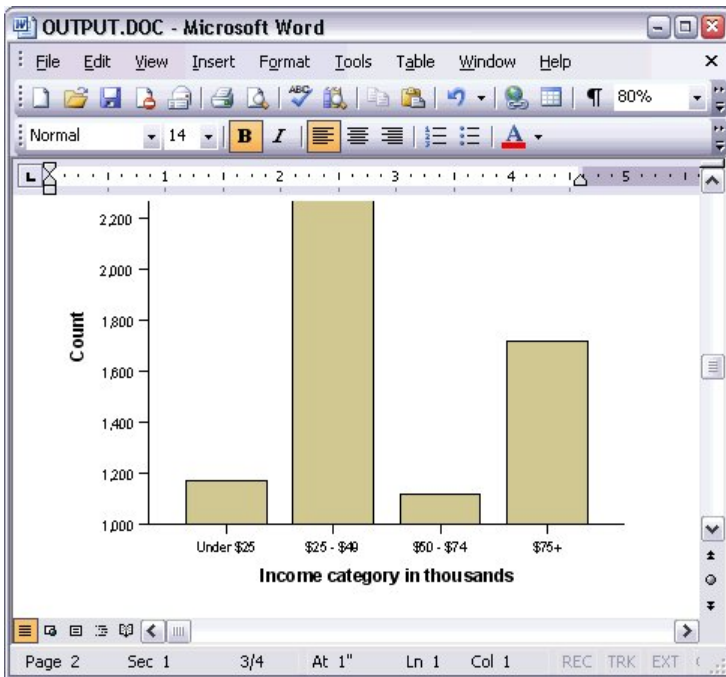


图 59: Word 中的图表

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。

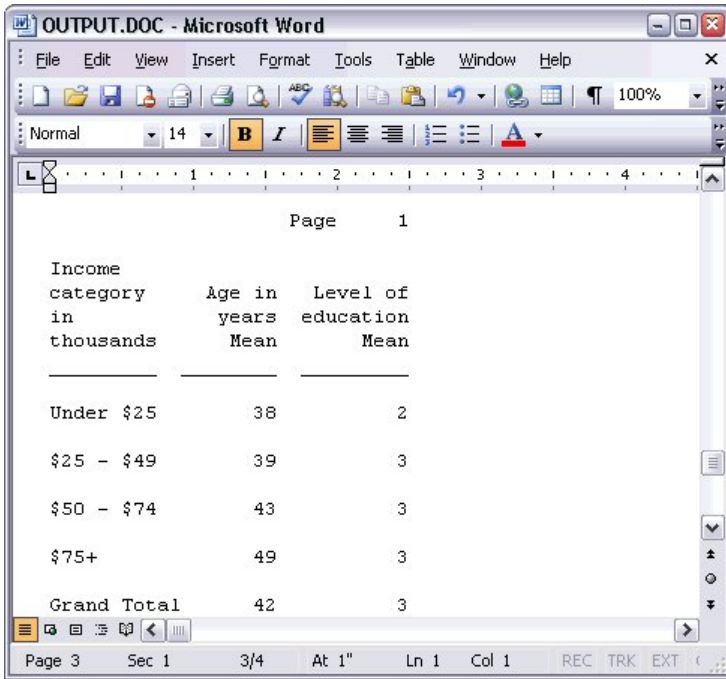


图 60: Word 中的文本输出

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 的透视表将成为 Word 表，原始透视表的所有格式都已保留，包括字体、颜色、边框等等。

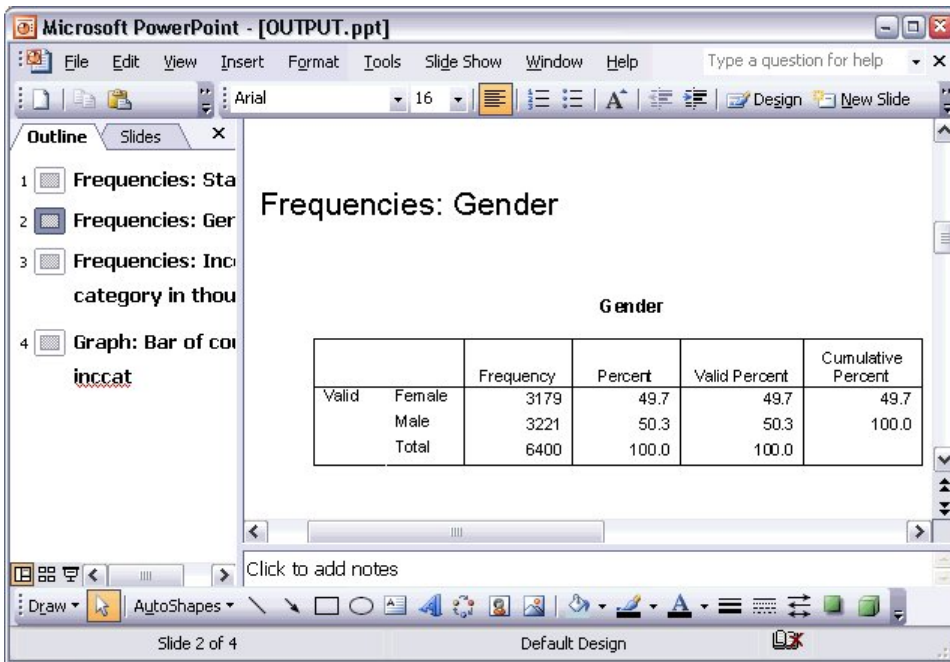


图 61: PowerPoint 中的透视表

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



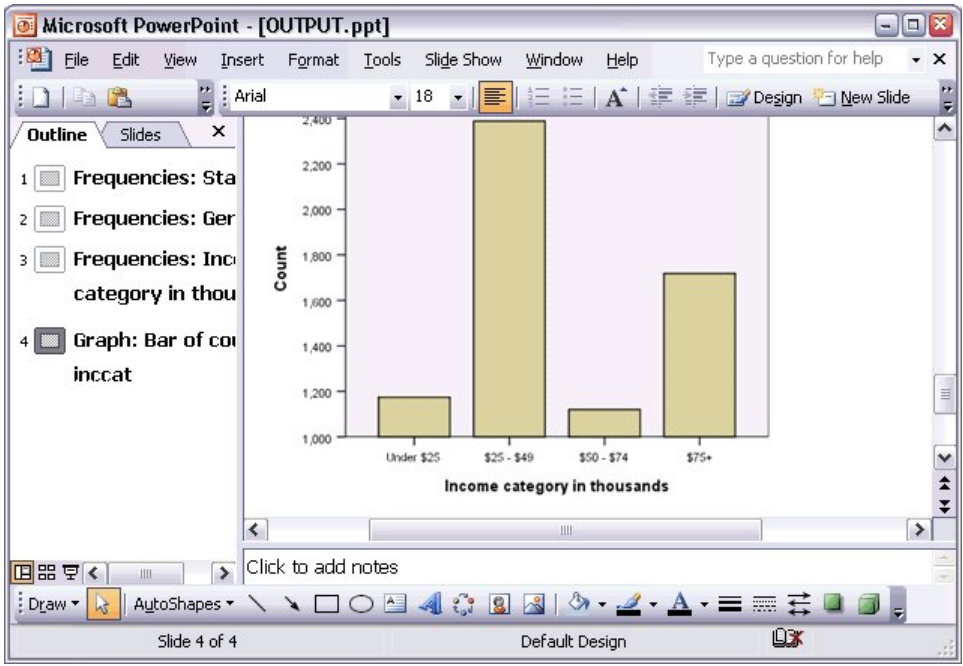


图 62: PowerPoint 中的图表

注：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。若要导出到 Excel 文件，则导出的结果会有所不同。

		Frequency	Percent	Valid Percent	Cumulative Percent
<b>Gender</b>					
Valid	Female	3,179	49.7	49.7	49.7
	Male	3,221	50.3	50.3	100.0
	Total	6,400	100.0	100.0	
<b>Income category in thousands</b>					
Valid	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
	Total	6,400	100.0	100.0	

图 63: Excel 中的 Output.xls

透视表的行、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。



	A	B	C	D	E
68	Income				
69	category	Age in	Level of		
70	in	years	education		
71	thousands	Mean	Mean		
72					
73					
74	Under \$25	38	2		
75					
76	\$25 - \$49	39	3		
77					
78	\$50 - \$74	43	3		
79					
80	\$75+	49	3		
81					
82	Grand Total	42	3		

图 64: Excel 中的文本输出

## 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。

1. 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...

2. 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**。

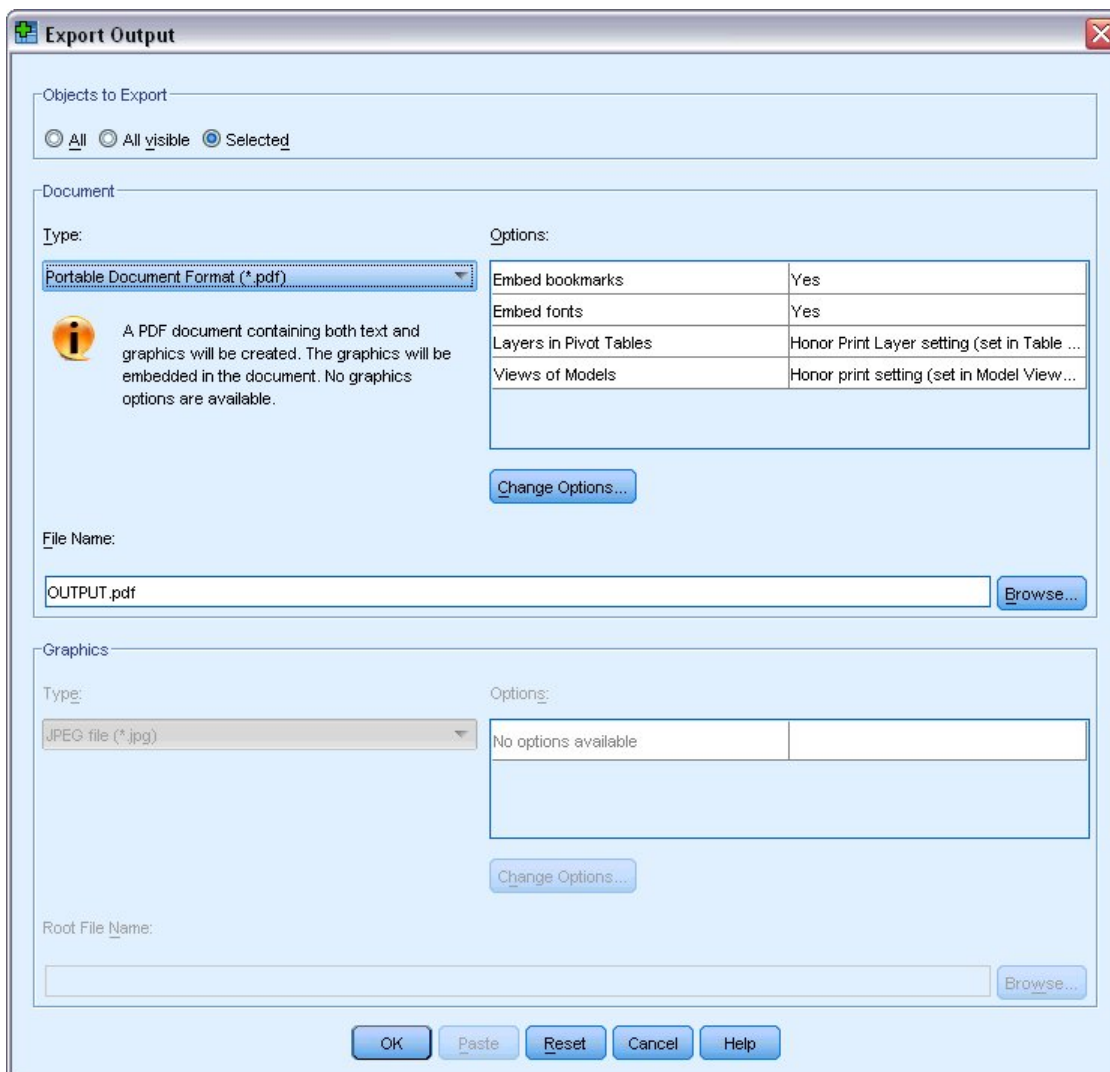


图 65: “导出输出”对话框

- 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) 是缺省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，打印的效果可能会比较差。

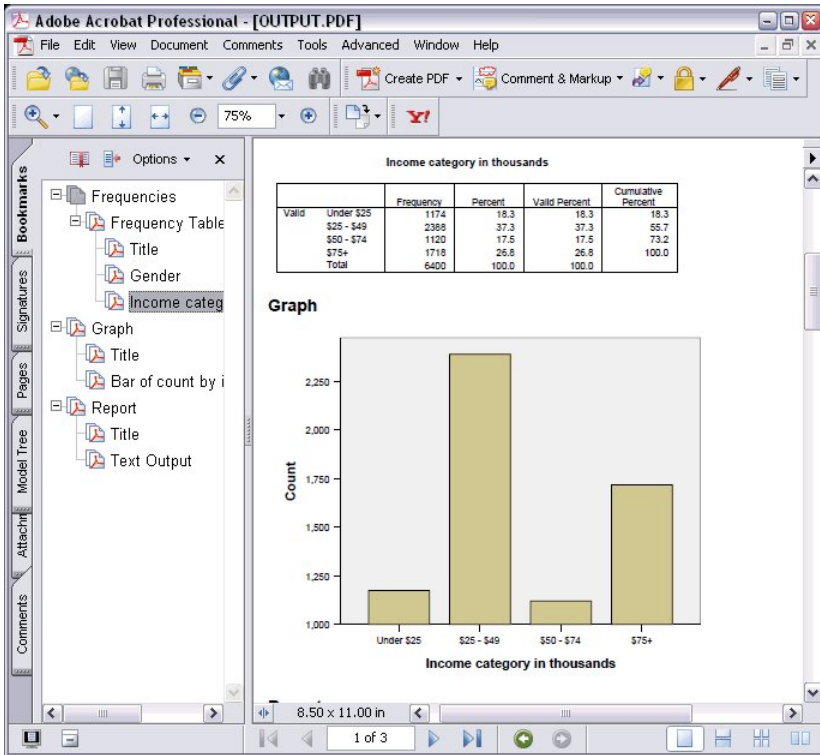


图 66: 带书签的 PDF 文件

## 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。

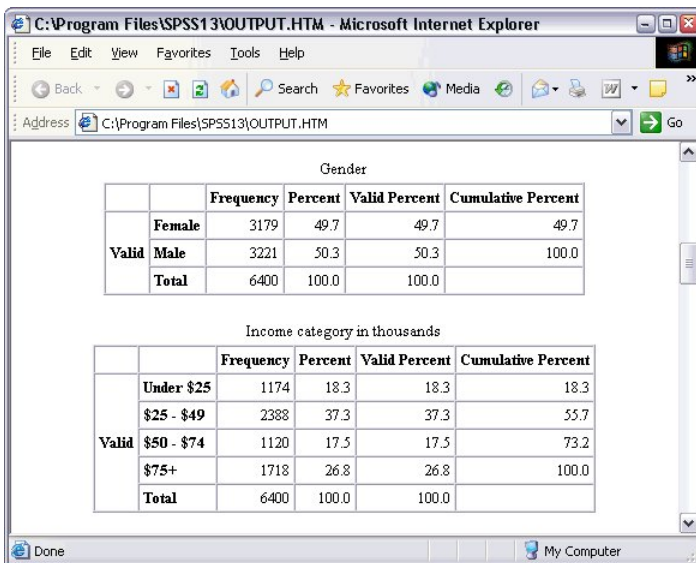


图 67: Web 浏览器中的 Output.htm

导出到 HTML 时，也可以导出图表，但不能导出到单个文件中。

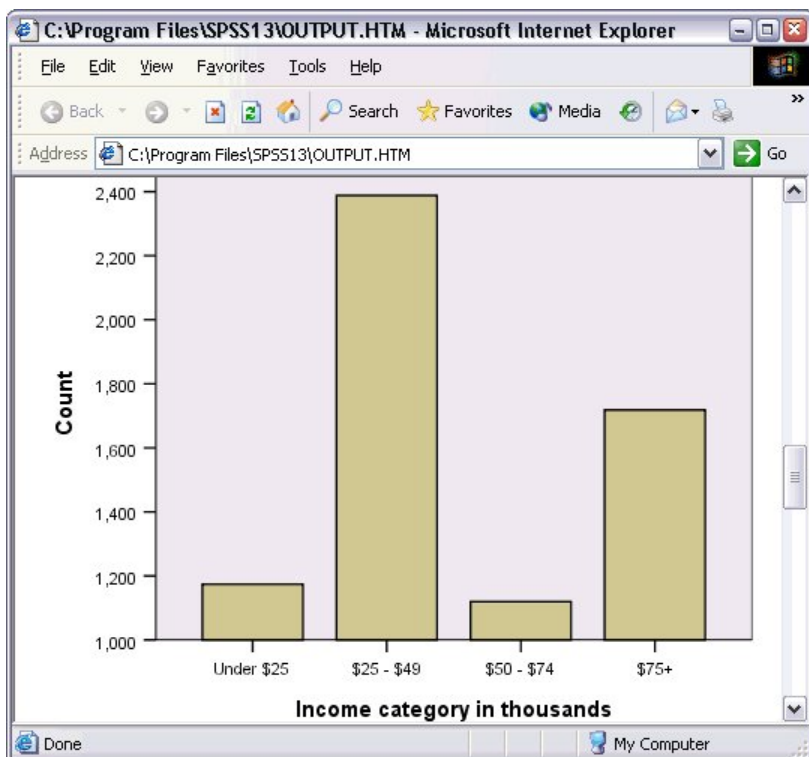


图 68: HTML 格式的图表

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。

## 第 7 章 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*。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。

注：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。

1. 打开数据文件 *demo.sav*。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。
2. 从菜单中选择：  
    **分析 > 描述统计 > 频率...**
3. 选择婚姻状况 [*marital*] 并将其移到“变量”列表中。
4. 单击**图表**。
5. 在“图表”对话框中，选择**条形图**。
6. 在“图表值”组中，选择**百分比**。
7. 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。

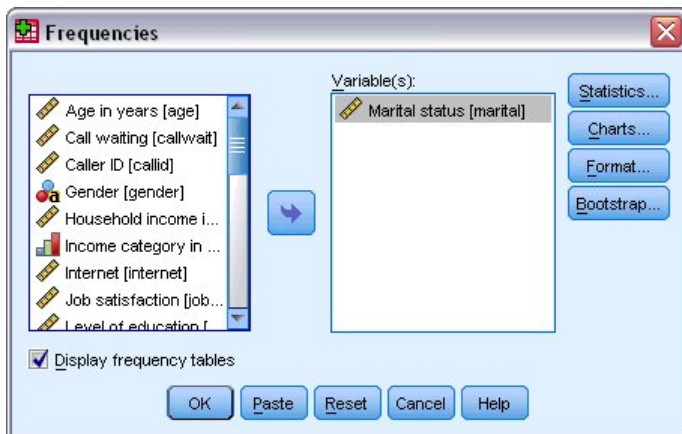


图 69: “频率”对话框

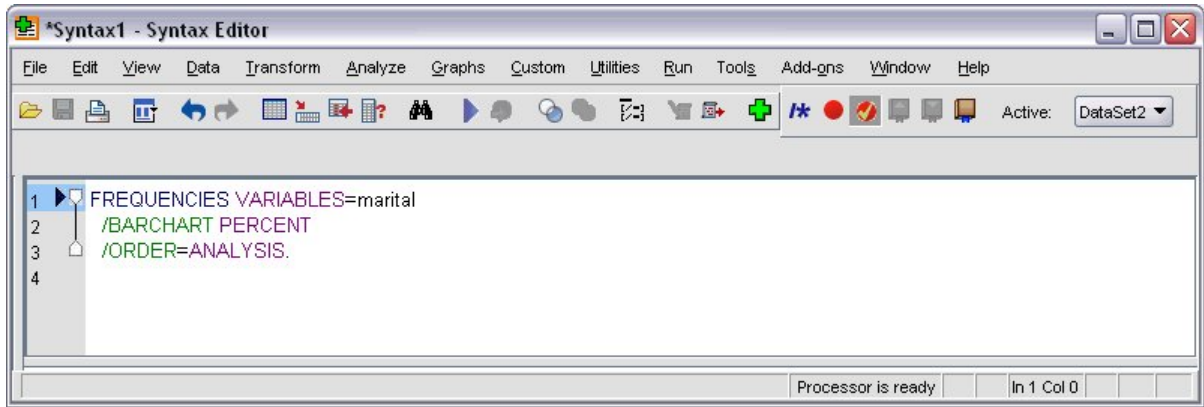


图 70: 频率语法

8. 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。

从 `BARCHART` 子命令删除关键字 `PERCENT`。

按住 `Ctrl` 的同时按 `Spacebar`。

1. 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。

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。

2. 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。

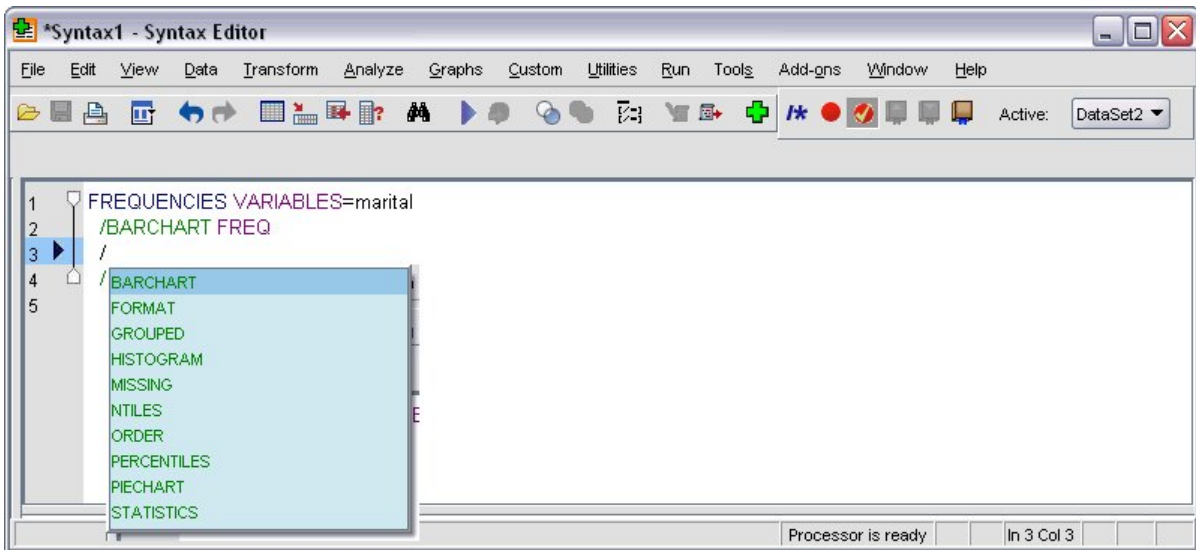


图 71: 显示子命令的自动完成控件

要获得有关当前命令的更多详细帮助，按 `F1` 键。这将使您直接转至当前命令的命令语法参考信息。

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

1. 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。

2. 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。

3. 单击**打开**。

4. 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。

## 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：

1. 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。

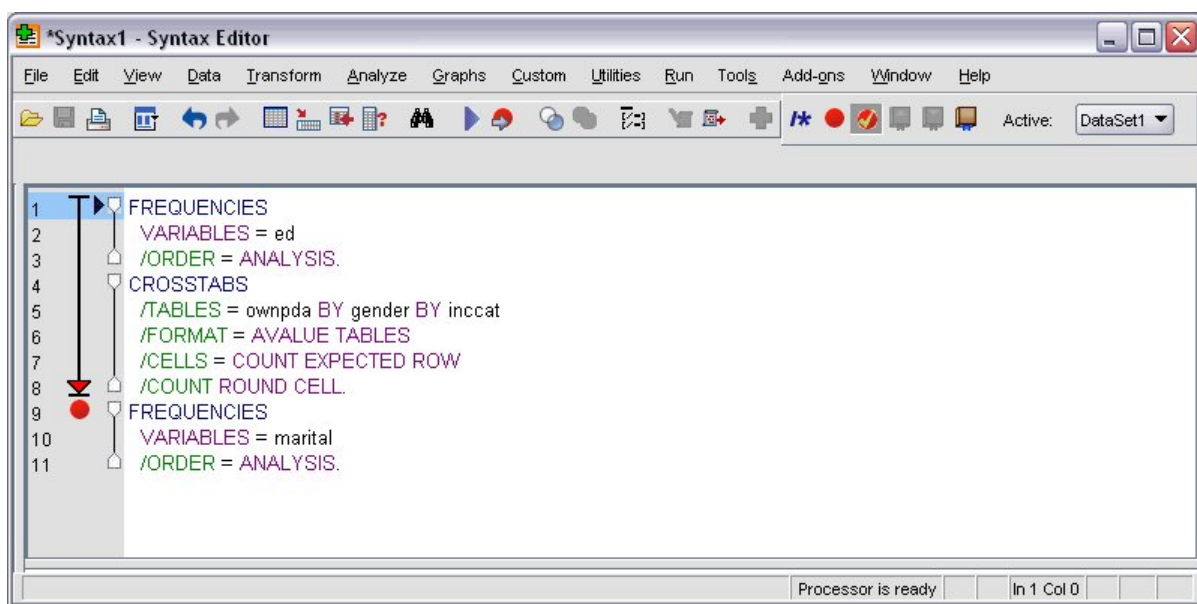


图 72: 执行在分界点处停止

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。

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：

2. 从语法编辑器中的菜单选择：

运行 > 继续





## 第 8 章 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*。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。

### 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 1–4 表示以下收入分类（以千为单位）：小于 \$25、\$25–\$49、\$50–\$74、\$75 或以上。

要创建分类变量 *inccat*：

1. 从数据编辑器窗口的菜单中选择：

**转换 > 可视分箱...**

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

由于可视分箱化依赖于数据文件中的实际值以帮助您做出良好的分箱化选择，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。

2. 将 *Household income in thousands [income]* 从“变量”列表中拖放到“要分箱的变量”列表中，然后单击**继续**。

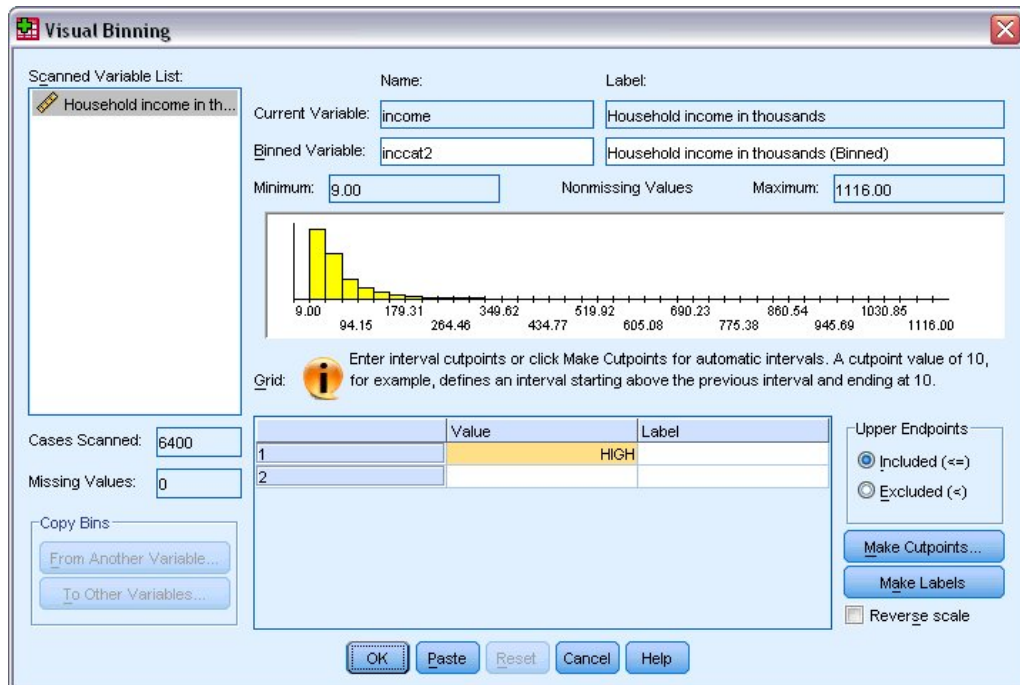


图 73: “可视分箱化”主对话框

- 在“可视分箱化”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* 作为新的分箱化变量名称，输入 *Income category [in thousands]* 作为变量标签。
  - 单击**生成分割点**。
  - 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。
- 分箱化类别的数量比分割点数量多一个。因此在本示例中，新的分箱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**。
- 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 及以上。
- 在“上端点”组中，选择**已排除 (<)**。
  - 然后单击**生成标签**。

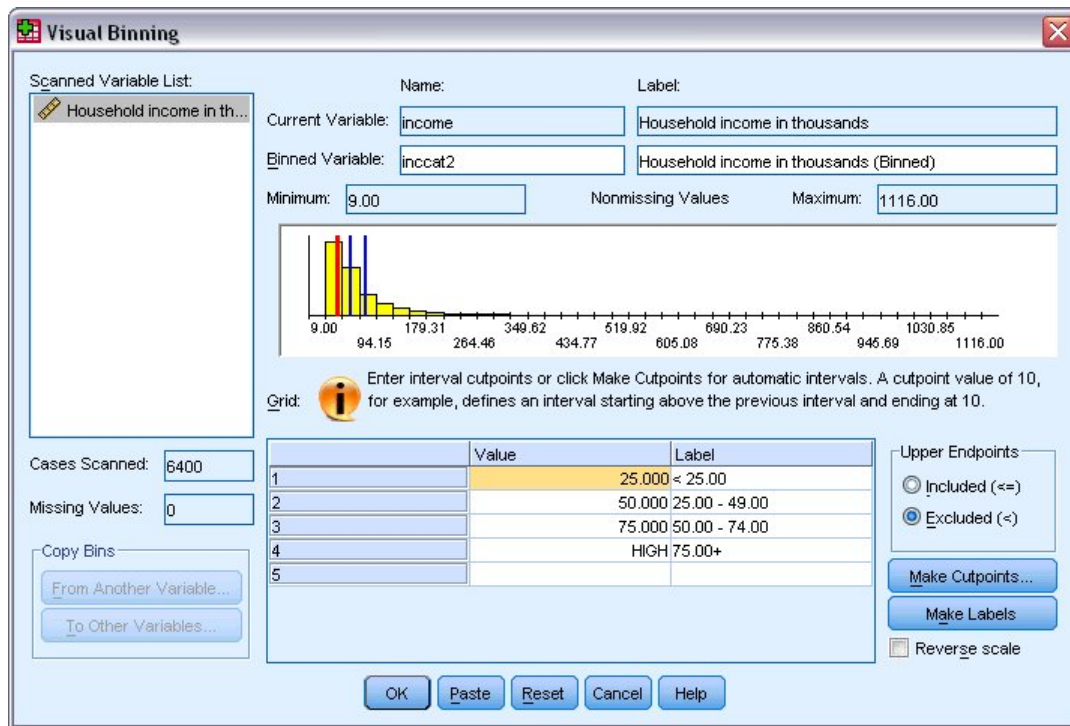


图 74: 自动生成的值标签

这将自动为每个类别生成描述性值标签。由于为新的分箱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**以创建新的分箱化变量。

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。

## 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。

1. 从数据编辑器窗口的菜单中选择：

转换 > 计算变量...

2. 对于“目标变量”，输入 `jobstart`。

3. 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。

4. 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）。

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

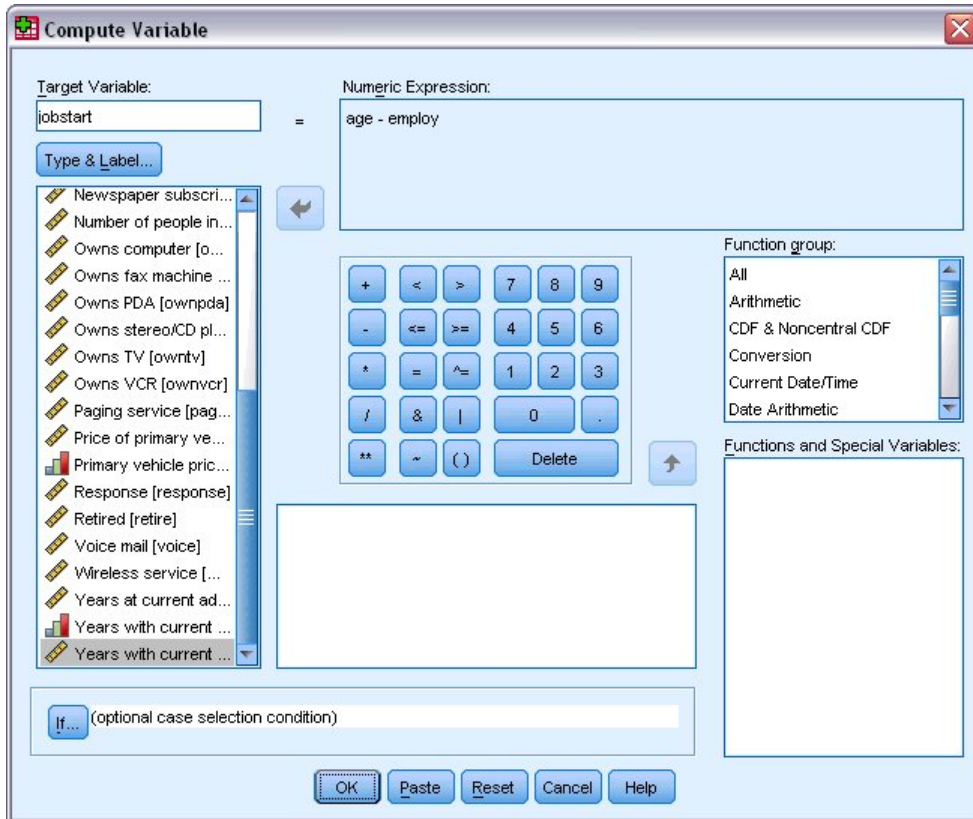


图 75: “计算变量”对话框

注：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。数字表达式应该为 `age-employ`，而不是 `age-empcat`。

6. 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。

## 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

- 分发函数
- 逻辑函数
- 日期和时间汇总函数以及提取函数
- 缺失值函数
- 交叉个案函数
- 字符串函数

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，另一组计算统计度规。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。

将函数粘贴到表达式中

要将函数粘贴到表达式中：

1. 将光标定位到您希望函数在表达式中出现的位置。
2. 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。
3. 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。

1. 在粘贴的函数中突出显示问号。
2. 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。针对每个个案，条件表达式会返回值 `true`、`false` 或缺失值。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：

1. 单击“计算变量”对话框中的**如果**。这将打开“If 个案”对话框。
2. 选中**如果个案满足条件则包含**。
3. 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*。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。

要使用日期和时间向导：

1. 从菜单中选择：

转换 > 日期和时间向导...



图 76: 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：

1. 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。
2. 选择**计算两个日期之间的时间单位数**，然后单击下一步。



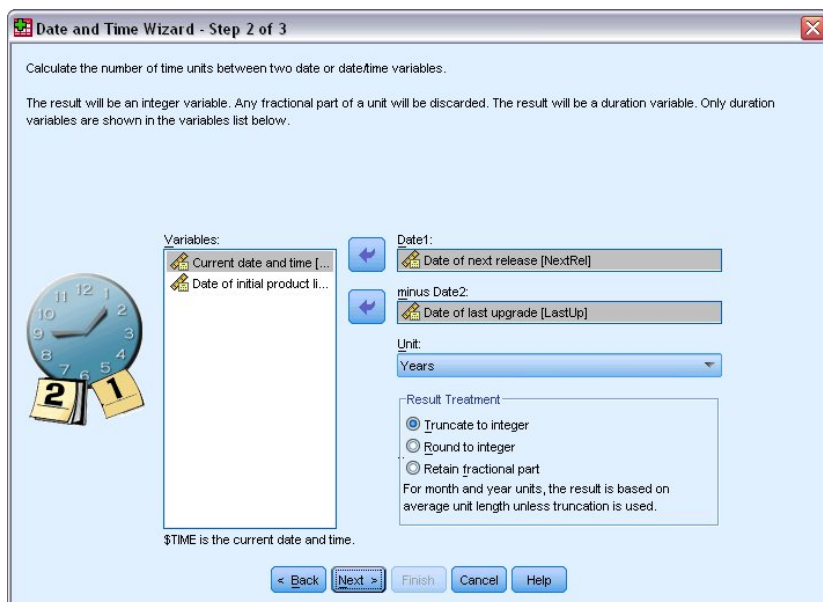


图 77: 计算两个日期之间的时间长度：步骤 2

3. 在步骤 2 中，选择 *Date of next release* 作为 Date1。
4. 选择 *Date of last upgrade* 作为 Date2。
5. 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。）
6. 单击下一步。
7. 在步骤 3 中，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。
8. 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。
9. 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。

## 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：

1. 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。
2. 在步骤 1 中，选择从日期中添加或减去期间，然后单击下一步。



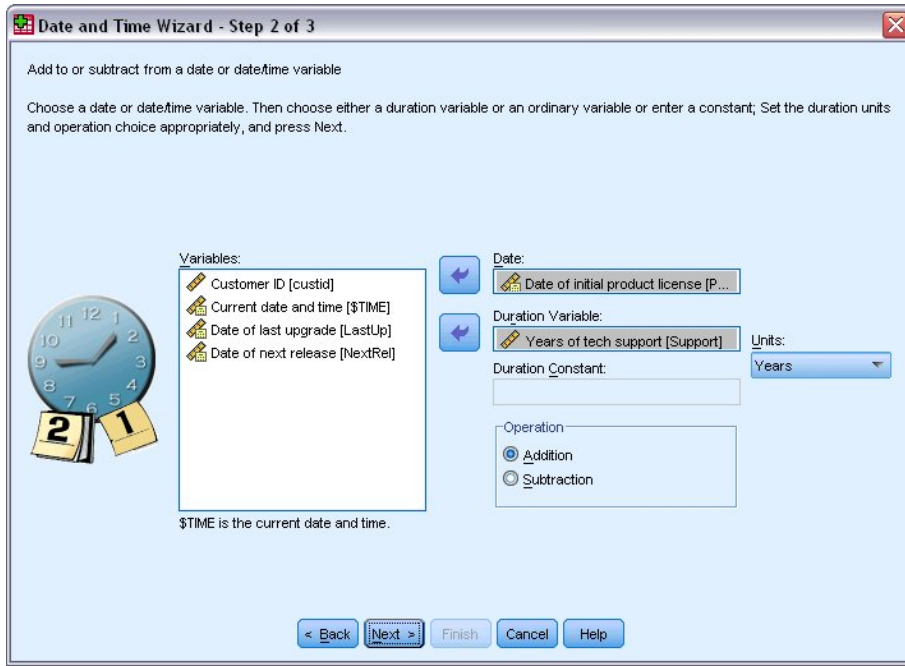


图 78: 向日期添加期间: 步骤 2

3. 选择 *Date of initial product license* 作为日期。

4. 在步骤 2 中, 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。

5. 从“单位”下拉列表中选择年份。

6. 单击下一步。

7. 在步骤 3 中, 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。

8. 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。

9. 单击完成以创建新变量。

新变量将显示在数据编辑器中。



## 第 9 章 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*。请参阅第 73 页的『第 10 章 样本文件』主题以获取更多信息。

### 对数据排序

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：

1. 从菜单中选择：

**数据 > 排序个案...**

“个案排序”对话框随即显示。

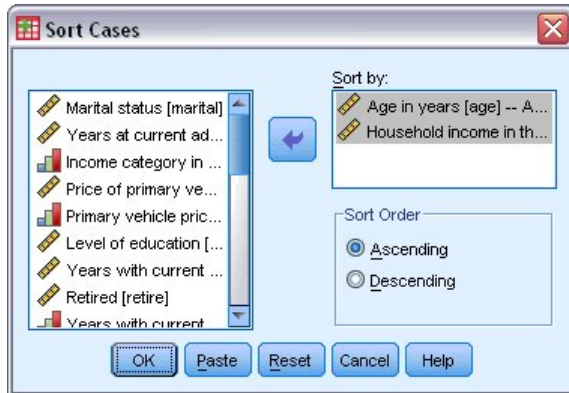


图 79: “个案排序”对话框

2. 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：

1. 从菜单中选择：

**数据 > 拆分文件...**

“拆分文件”对话框随即显示。

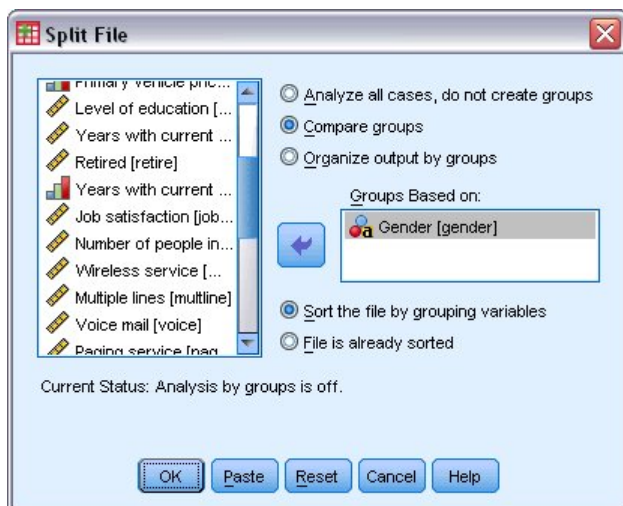


图 80: “拆分文件”对话框

2. 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。）
3. 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”过程生成的汇总统计表所示。

**Statistics**

Household income in thousands

Female	N	Valid	3179
		Missing	0
	Mean		68.7798
	Median		44.0000
	Std. Deviation		75.73510
Male	N	Valid	3221
		Missing	0
	Mean		70.1608
	Median		45.0000
	Std. Deviation		81.56216

图 81: 带有单个透视表的拆分文件输出

如果选择**按组来组织输出**并运行“频率”过程，则会创建两个透视表：一个女性透视表和一个男性透视表。

**Statistics<sup>a</sup>**

Household income in thousands

N	Valid	3179
	Missing	0
Mean		68.7798
Median		44.0000
Std. Deviation		75.73510

a. Gender = Female

图 82: 带有女性透视表的拆分文件输出

**Statistics<sup>a</sup>**

Household income in thousands

N	Valid	3221
	Missing	0
Mean		70.1608
Median		45.0000
Std. Deviation		81.56216

a. Gender = Male

图 83: 带有男性透视表的拆分文件输出

## 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

选择一个个案子集进行分析：

1. 从菜单中选择：

**数据 > 选择个案...**

这将打开“选择个案”对话框。

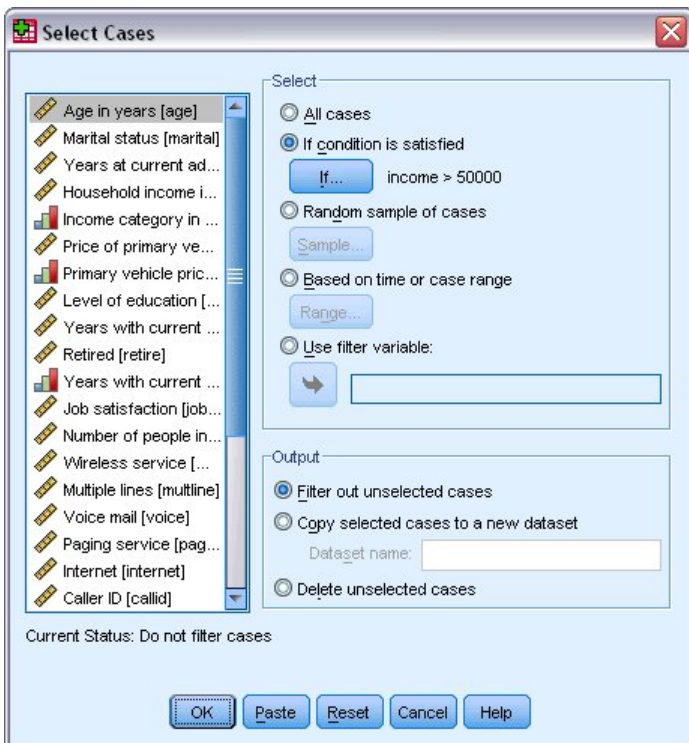


图 84: “选择个案”对话框

## 基于条件表达式选择个案

基于条件表达式选择个案:

1. 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。

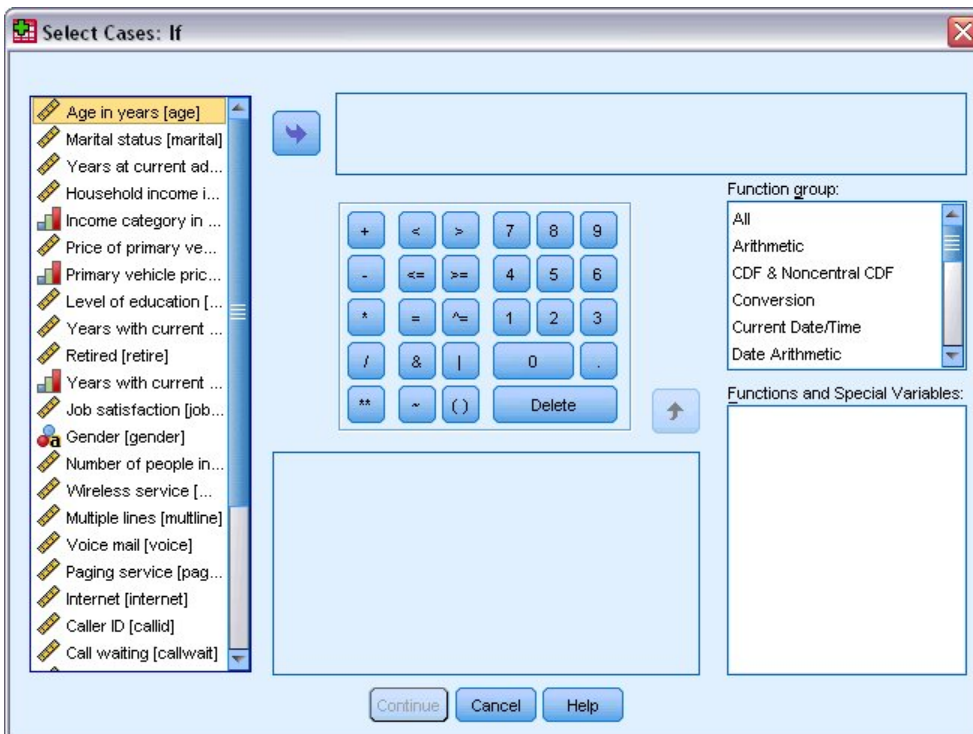


图 85: “选择个案 If”对话框

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。请参阅第 62 页的『使用条件表达式』主题，了解更多信息。

## 要选择随机样本请执行下列操作

获取随机样本：

1. 在“选择个案”对话框中选择**随机个案样本**。
2. 单击**样本**。

这将打开“选择个案：随机样本”对话框。

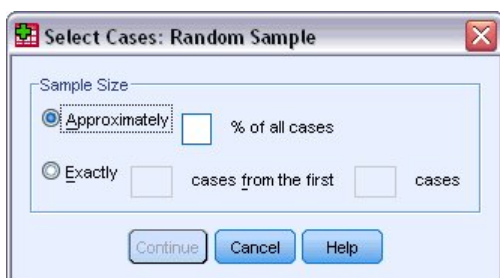


图 86: “选择个案：随机样本”对话框

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，那么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：

1. 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。

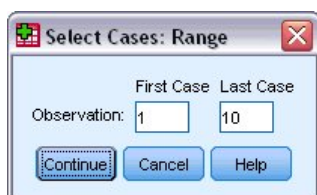


图 87: “选择个案：范围”对话框

- **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 88: “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:

2. 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：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。

1
2
3
4
5
6
7
8
9
10
11

图 89: 个案选择状态

## 第 10 章 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字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.** 在致力于标准化厌食/暴食行为症状学的同时，研究人员<sup>1</sup>对 55 名有饮食障碍的青少年进行了研究。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.** 在经典示例中<sup>2</sup>，要求 52 名学生对 15 种情况和 15 种行为的组合进行评分，10 分制，范围从 0=“非常合适”到 9=“非常不合适”。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.** 在经典研究中<sup>3</sup>，要求 21 名沃顿商学院 MBA 学生和他们的配偶按照偏好顺序对 15 种早餐项目进行排名，从 1=“最喜欢”到 15=“最不喜欢”。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.** 在其他位置显示和分析的数据集<sup>4</sup>关于汽车损坏索赔的问题 平均理赔金额可以当作其具有伽玛分布来建模，通过使用逆关联函数将因变量的平均值与投保者年龄、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标度权重。

<sup>1</sup> Van der Ham, T., J. J. Meulman, D. C. Van Strien, and H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363-368.

<sup>2</sup> Price, R. H., and D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579-586.

<sup>3</sup> Green, P. E., and V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

- **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.** 在热门示例中<sup>5</sup>，一家对营销新型地毯清洁剂感兴趣的公司想要考察五个因素对消费者偏好的影响：包装设计、品牌名称、价格、良好的内务管理印章和退款保证。包装设计有三个因子级别，每个因子级别因刷体位置而不同；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概要文件的平均等级的排序。低等级与高偏好相对应。此变量反映了对每个概要文件的偏好的总体度量。
- **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序顺序。消费者被要求按照从最喜欢到最不喜欢的顺序对 22 个产品概要文件进行排序。*carpet\_plan.sav* 中定义了变量 *PREF1* 到 *PREF22* 包含相关特征的标识。
- **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。同时还包括 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素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响应者。
- **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.** 此数据文件涉及六种冰咖啡品牌的感知图像<sup>6</sup>。对于 23 个冰咖啡图像属性中的每一个属性，人们选择了通过该属性描述的所有品牌。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等级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响应者。
- **debate\_aggregate.sav.** 该假设数据文件汇总了 *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。

<sup>4</sup> McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

<sup>5</sup> Green, P. E., and Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

<sup>6</sup> Kennedy, R., C. Riquier, and B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56-70.

- **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格，并记录地区、省、区、市、子区和单元格标识。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格，并记录各种人口统计和抽样信息。
- **diabetes\_costs.sav.** 这是假设数据文件，其中包含保险公司维护的有关患有糖尿病的投保人的信息。每个个案对应一个不同投保人。
- **dietstudy.sav.** 此假设数据文件包含对“斯蒂尔曼饮食”的研究结果<sup>7</sup>。每个案例对应一个单独的受试者，并记录他或她的饮食前后体重（磅）和甘油三酯水平（mg/100 ml）。
- **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同一部分联系人的信息，而 *dmdata3.sav* 包含其余未收到测试邮件的联系人信息。
- **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.** 此数据文件来自机器学习数据库存储库中的“德语信用”数据集<sup>8</sup> 在加州大学欧文分校
- **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.** 贝尔<sup>9</sup> 展示了一个表来说明可能的社会群体。Guttman<sup>10</sup> 使用了该表的一部分，其中描述诸如社交互动、群体归属感、成员的身体接近度和关系形式等事物的五个变量与七个理论社会群体交叉，包括人群（例如，足球比赛中的人群）、观众（例如，参与剧院或课堂演讲的人群）、公众（例如，报纸或电视观众）、聚众（如互动更加激烈的人群）、主要群体（亲密的）、次要群体（自愿）和现代社区（由于近距离接触和需要专业服务而产生的松散联盟）。
- **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来标记具有潜在欺骗性的可疑理赔。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.** 罗森伯格和金<sup>11</sup> 着手分析 15 个亲属关系术语（阿姨、兄弟、表亲、女儿、父亲、孙女、祖父、祖母、孙子、母亲、侄子、侄女、姐妹、儿子、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

<sup>7</sup> Rickman, R., N. Mitchell, J. Dingman, and J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228:, 54-58.

<sup>8</sup> Blake, C. L., and C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

<sup>9</sup> Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

<sup>10</sup> Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469-506.

<sup>11</sup> Rosenberg, S., and M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489-502.

序，第二次排序和第一次排序采取的标准不同。这样，一共得到了六组“源”。每个源对应一个 15 x 15 的近似值矩阵，该矩阵单元格中的值等于源中的人数减去该源中划分对象的次数。

- **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.** 该数据文件涉及伊利诺伊州 Algonquin 的新住房开发中的房屋销售。在 1999-2000 年期间。这些销售情况属于公共记录。
- **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/](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 个实测值。以前的研究人员<sup>12,13</sup>发现这些变量之间存在非线性，这阻碍了标准回归方法。
- **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，那么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。每个个案对应第一次被捕后于 2003 年 6 月期间释放的一名先前的罪犯，并记录其人口统计信息、第一次犯罪的详细资料以及第二次被捕的数据（如果于 2006 年 6 月底之前第二次被捕）。罪犯是根据 *recidivism\_cs.csplan* 中指定的抽样计划从抽样的部门中选择的；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 (*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 接受实践教程。在培训课程结束时，对每名员工进

<sup>12</sup> Breiman, L., and J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.

<sup>13</sup> Hastie, T., and R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.



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.** 该数据文件包含有关螺钉、螺栓、螺母和大头钉特性的信息<sup>14</sup>。
- **shampoo\_ph.sav.** 这是关于某发制品厂的质量控制的假设数据文件。在规定的间隔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.** 在其他位置显示和分析的数据集<sup>15</sup> 这涉及波浪对货船造成的损坏。在给出了船的类型、建造工期和服务期后，可以根据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《全国药物滥用家庭调查》，是美国家庭的概率样本。(http://dx.doi.org/10.3886/ICPSR02934) 因此，分析此数据文件的第一步应该是对数据进行加权以反映人口趋势。
- **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。样本为左侧截断，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。
- **tcm\_kpi.sav.** 这是一个假设数据文件，其中包含公司的每周关键绩效指标的值。它还包含相同时间段内多个可控制指标的每周数据。
- **tcm\_kpi\_upd.sav.** 该数据文件和 *tcm\_kpi.sav* 一样，但包含另外四周的数据。
- **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.** 此数据文件与 *telco.sav* 数据文件相似，但“tenure”和日志转换的客户支出变量已被除去，取而代之的是标准化的日志转换的客户支出变量。
- **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缺省状态。

<sup>14</sup> Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

<sup>15</sup> McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

- **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 位响应者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响应者；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。它提供了时间间隔检查数据的良好示例，并且已在别处进行了展示和分析<sup>16</sup>。
- **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。已在别处进行了展示和分析<sup>17</sup>。
- **verd1985.sav.** 此数据文件涉及调查<sup>18</sup>。记录了 15 名受试者对 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。宠物被标度化为多名义而年龄被标度化为有序；所有其他变量都被标度化为单名义。
- **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子集<sup>19</sup>。数据包含对来自俄亥俄州斯图本维尔的 7 岁、8 岁、9 岁和 10 岁儿童喘息状态的重复二元测量，以及在研究的第一年母亲是否吸烟的固定记录。
- **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。

<sup>16</sup> Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

<sup>17</sup> Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

<sup>18</sup> Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (in Dutch)*. Leiden: Department of Data Theory, University of Leiden.

<sup>19</sup> Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer, and B. G. Ferris Jr.. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366-374.



# 声明

---

本信息是为在美国提供的产品和服务编写的。IBM 提供了本资料的其他语言版本。但是，您可能需要拥有该语言的产品副本或产品版本才能访问这些资料。

IBM 可能在其他国家或地区不提供本文档中所讨论的产品、服务或功能特性。有关您当前所在区域的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或暗示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务，则由用户自行负责。

IBM 可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并未授予用户使用这些专利的任何许可。您可以通过书面方式将许可查询寄往：

*IBM Director of Licensing*

*IBM Corporation*

*North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
US*

有关双字节 (DBCS) 信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或以书面形式将查询寄往：

*Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.*

*19-21, Nihonbashi-Hakozakicho, Chuo-ku  
Tokyo 103-8510, Japan*

International Business Machines Corporation“按现状”提供本出版物，不附有任何种类的（无论是明示的还是默示的）保证，包括但不限于默示的有关非侵权、适销和适用于某种特定用途的保证。某些管辖区域在某些交易中不允许免除明示或默示的保证。因此本条款可能不适用于您。

本信息可能包含技术方面不够准确的地方或印刷错误。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 可以随时对本出版物中描述的产品进行改进和/或更改，而不另行通知。

本信息中对非 IBM Web 站点的任何引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是本 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的被许可方如果要了解有关程序的信息以达到如下目的：(i) 允许在独立创建的程序和其他程序（包括本程序）之间进行信息交换，以及 (ii) 允许对已经交换的信息进行相互使用，请与下列地址联系：

*IBM Director of Licensing*

*IBM Corporation*

*North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
US*

只要遵守适当的条款和条件，包括某些情形下的一定数量的付费，都可获得这方面的信息。

本文档中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际程序许可协议或任何同等协议中的条款提供。

所引用的性能数据和客户示例只用于阐述说明。根据具体配置和操作条件，实际性能结果可能有所不同。

涉及非 IBM 产品的信息是从这些产品的供应商、已出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

关于 IBM 未来方向或意向的声明都可随时变更或收回，而不另行通知，它们仅仅表示了目标和意愿而已。

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。所有这些名称都是虚构的，若实际人员或企业与此相似，纯属巧合。

版权许可：

本信息包括源语言形式的样本应用程序，这些样本说明不同操作平台上的编程方法。如果是为按照在编写样本程序的操作平台上的应用程序编程接口 (API) 进行应用程序的开发、使用、经销或分发，您可以任何形式对这些样本程序进行复制、修改、分发，而无须向 IBM 付费。这些示例并未在所有条件下作全面测试。因此，IBM 不能担保或暗示这些程序的可靠性、可维护性或功能。本样本程序仍然是“按现状”提供的，不附有任何种类的保证。对于因使用样本程序所引起的任何损害，IBM 概不负责。

凡这些实例程序的每份拷贝或其任何部分或任何衍生产品，都必须包括如下版权声明：

© Copyright IBM Corp. 2021. 此部分代码是根据 IBM Corp. 的样本程序衍生出来的。

© Copyright IBM Corp. 1989 - 2021. All rights reserved.

## 商标

---

IBM、IBM 徽标和 [ibm.com](http://ibm.com) 是 International Business Machines Corp. 在全球许多司法辖区的商标或注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。最新的 IBM 商标列表可以在 Web 上的“Copyright and trademark information”中获取，地址为：[www.ibm.com/legal/copytrade.shtml](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。

Linux 是 Linus Torvalds 在美国和/或其他国家或地区的注册商标。

Microsoft、Windows、Windows NT 和 Windows 徽标是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

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

Java 和所有基于 Java 的商标和徽标是 Oracle 和/或其子公司的商标或注册商标。

# 索引

## Special Characters

编辑透视表 [36](#)

变量

    标签 [19](#)

    数据类型 [20](#)

变量标签

    创建 [19](#)

标度变量

    汇总测量 [25](#)

表达式中的函数 [61](#)

测量级别 [23](#)

层

    在透视表中创建 [36](#)

查看器

    移动输出 [33](#)

    隐藏和显示输出 [33](#)

拆分文件处理 [67](#)

创建变量标签 [19](#)

导出结果

    到 Excel [45](#)

    到 PowerPoint [45](#)

    到 Word [45](#)

    HTML [53](#)

电子表格文件

    读取 [8](#)

    读取变量名称 [8](#)

定比数据 [23](#)

定距数据 [23](#)

定量数据 [23](#)

定性数据 [23](#)

对个案排序 [67](#)

对值重新编码 [59](#)

分类数据

    汇总测量 [23](#)

个案

    排序 [67, 69](#)

    选择 [69](#)

个案的子集

    过滤未选定的个案 [72](#)

    基于日期和时间 [71](#)

    如果满足条件 [70](#)

    删除未选定的个案 [72](#)

    随机样本 [71](#)

    条件表达式 [70](#)

    选择 [69](#)

汇总测量

    标度变量 [25](#)

    分类数据 [23](#)

计数

    计数表 [23](#)

计算新变量 [61](#)

刻度数据 [23](#)

连续数据 [23](#)

名义数据 [23](#)

频率表 [23](#)

缺失值

缺失值 (继续)

    非数值变量 [22](#)

    数值变量 [21](#)

    系统缺失 [21](#)

日期和时间变量 [63](#)

日期和时间向导 [63](#)

输入数据

    非数字 [18](#)

    数字 [17](#)

数据编辑器

    输入非数字数据 [18](#)

    输入数字数据 [17](#)

数据库文件

    读取 [10](#)

数据库向导 [10](#)

数据类型

    变量 [20](#)

数据条目 [17, 18](#)

数值数据 [17](#)

条件表达式 [62](#)

条形图 [24](#)

透视表

    编辑 [36](#)

    变换行和列 [35](#)

    层 [36](#)

    单元格格式 [37](#)

    单元格数据类型 [37](#)

    访问定义 [34](#)

    格式化 [36](#)

    透视托盘 [35](#)

    隐藏小数点 [37](#)

    隐藏行与列 [37](#)

图表

    创建图表 [27](#)

    条形 [27](#)

    条形图 [24](#)

    直方图 [26](#)

图形

    创建图形 [27](#)

    条形 [27](#)

文本导入向导 [12](#)

文本数据文件

    读取 [12](#)

系统缺失值 [21](#)

选择个案 [69](#)

样本文件

    位置 [73](#)

移动

    查看器中的项 [33](#)

    透视表中的元素 [35](#)

有序数据 [23](#)

语法 [55](#)

语法帮助工具 [56](#)

语法窗口

    编辑命令 [56](#)

    分界点 [57](#)

    颜色编码 [56](#)

语法窗口 (继续)

运行命令 [55, 57](#)

粘贴命令 [55](#)

自动完成 [56](#)

语法文件

打开 [57](#)

在透视表中变换 (翻转) 行与列 [35](#)

在透视表中隐藏行与列 [37](#)

粘贴语法

从对话框 [55](#)

直方图 [26](#)

值标签

控制查看器中的显示 [20](#)

数值变量 [20](#)

指定 [20](#)

字符串数据

输入数据 [18](#)

## A

Access (Microsoft) [10](#)

## E

Excel (Microsoft)

将结果导出到 [45](#)

Excel 文件

读取 [8](#)

## H

HTML

导出结果 [53](#)

## P

PowerPoint (Microsoft)

将结果导出到 [45](#)

## W

Word (Microsoft)

将结果导出到 [45](#)



