

# **IBM SPSS Custom Tables**

## **22**



**注释**

请先阅读 第 93 页的『声明』 中的信息，然后再使用本信息及其支持的产品。

**产品信息**

此版本适用于 IBM SPSS Statistics22.0.0 以及所有后续发行版和修订版，直至在新版本中另有说明为止。

# 目录

<b>第 1 章 表构建器界面</b> . . . . .	<b>1</b>	在已计算类别中隐藏类别 . . . . .	38
表构建器界面 . . . . .	1	在已计算类别中引用小计 . . . . .	39
构建表 . . . . .	1	使用已计算类别显示非穷举小计 . . . . .	40
构建表 . . . . .	3	<b>第 6 章 具有共享类别的变量的表</b> . . . . .	<b>43</b>
堆积变量 . . . . .	3	计数表 . . . . .	43
嵌套变量 . . . . .	4	百分比表 . . . . .	44
图层 . . . . .	4	总计和类别控制 . . . . .	45
显示和隐藏变量名/或变量标签 . . . . .	5	在表中对共享类别进行嵌套 . . . . .	45
汇总统计 . . . . .	5	<b>第 7 章 汇总统计</b> . . . . .	<b>47</b>
分类和总计 . . . . .	9	汇总统计源变量 . . . . .	47
已计算类别 . . . . .	12	分类变量的汇总统计源 . . . . .	48
具有共享类别的变量表 (共周界表) . . . . .	13	刻度变量的汇总统计源 . . . . .	48
定制表构建器 . . . . .	13	堆积变量 . . . . .	49
定制表: “选项”选项卡 . . . . .	13	分类变量的定制总计汇总统计 . . . . .	49
定制表: “标题”选项卡 . . . . .	14	显示类别值 . . . . .	50
定制表: “检验统计”选项卡 . . . . .	14	<b>第 8 章 刻度变量摘要</b> . . . . .	<b>53</b>
<b>第 2 章 分类变量的简单表</b> . . . . .	<b>17</b>	刻度变量摘要 . . . . .	53
分类变量的简单表 . . . . .	17	堆积刻度变量 . . . . .	53
单个分类变量 . . . . .	17	多个汇总统计 . . . . .	54
百分比 . . . . .	18	计数、有效 N 和缺失值 . . . . .	54
总计 . . . . .	19	不同变量的摘要 . . . . .	55
交叉列表 . . . . .	19	类别中的分组摘要 . . . . .	56
交叉制表中的百分比 . . . . .	19	多个分组变量 . . . . .	57
控制显示格式 . . . . .	20	将分类变量嵌套在刻度变量之内 . . . . .	57
边际总计 . . . . .	20	<b>第 9 章 检验统计</b> . . . . .	<b>59</b>
对类别排序和排除类别 . . . . .	21	检验统计 . . . . .	59
<b>第 3 章 对分类变量使用堆积、嵌套和层</b> <b>23</b>	<b>23</b>	独立性检验 (卡方检验) . . . . .	59
堆积分类变量 . . . . .	23	嵌套和堆积对独立性检验的影响 . . . . .	60
对交叉制表使用堆积 . . . . .	23	比较列平均值 . . . . .	61
嵌套分类变量 . . . . .	24	嵌套和堆积对列平均值检验的影响 . . . . .	62
不显示变量标签 . . . . .	25	比较列的比例 . . . . .	64
嵌套交叉制表 . . . . .	26	嵌套和堆积对列比率检验的影响 . . . . .	66
图层(L) . . . . .	27	有关权重和多响应集的说明 . . . . .	68
两个堆积分类层变量 . . . . .	28	<b>第 10 章 多响应集</b> . . . . .	<b>69</b>
两个嵌套的分类层变量 . . . . .	28	计数、响应、百分比和总计 . . . . .	69
<b>第 4 章 分类变量的总计和小计</b> . . . . .	<b>31</b>	将多响应集用于其他变量 . . . . .	71
单个变量的简单总计 . . . . .	31	统计源变量和可用的汇总统计 . . . . .	72
所见类别即总计类别 . . . . .	31	多类别集和重复响应 . . . . .	72
总计的显示位置 . . . . .	32	对多响应集进行显著性检验 . . . . .	73
嵌套表的总计 . . . . .	32	对多响应集进行独立性检验 . . . . .	74
层变量总计 . . . . .	34	使用多响应集比较列平均值 . . . . .	74
小计 . . . . .	34	<b>第 11 章 缺失值</b> . . . . .	<b>77</b>
所见类别即小计类别 . . . . .	35	没有缺失值的表 . . . . .	77
隐藏小计类别 . . . . .	35	在表中包含缺失值 . . . . .	78
层变量小计 . . . . .	36		
<b>第 5 章 分类变量的已计算类别</b> . . . . .	<b>37</b>		
简单已计算类别 . . . . .	37		

<b>第 12 章 设置表格式和定制表</b>	<b>81</b>
设置表格式和定制表	81
汇总统计显示格式	81
汇总统计的显示标签	82
列宽	83
为空单元格显示值	83
为缺失统计显示值	84

<b>第 13 章 样本文件</b>	<b>85</b>
<b>声明</b>	<b>93</b>
商标	94
<b>索引</b>	<b>97</b>

# 第 1 章 表构建器界面

## 表构建器界面

定制表使用简单的拖放表构建器界面，使您可以在选择变量和选项时预览表。它还具有典型对话框所不具备的灵活性，能够更改窗口大小以及窗口中窗格的大小。

## 构建表



图 1. “定制表”对话框，“表格”选项卡

在表构建器的“表格”选项卡中选择将在表中显示的变量和汇总测量。

**变量列表。** 数据文件中的变量显示在窗口左上方的窗格中。定制表会对变量的两个不同测量级别加以区别，并根据测量级别的不同分别处理：

**分类。** 包含有限数量的不同值或类别（例如，性别或宗教）的数据。分类变量可以是字符串（字母数值）变量或使用数值代码表示类别的数值变量（例如，0 = *male*，1 = *female*）。这种数据也称为定性数据。分类变量既可以是**名义变量**，也可以是**顺序变量**

- **名义 (Nominal)**. 当变量值表示不具有内在等级的类别时，该变量可以作为名义变量；例如，雇员任职的公司部门。名义变量的示例包括地区、邮政编码和宗教信仰。
- **有序 (Ordinal)**. 当变量值表示带有某种内在等级的类别时，该变量可以作为有序变量；例如，从十分不满意到十分满意的服务满意度水平。有序变量的示例包括表示满意度或可信度的态度分数和优先选择评分。

**刻度。** 以区间或比率刻度度量的数据，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。

分类变量定义表中的类别（行、列和层），缺省汇总统计为计数（每个类别中的个案数）。例如，分类性别变量的缺省表将只显示男性人数和女性人数。

刻度变量通常在分类变量的类别中进行摘要，缺省汇总统计为平均值。例如，性别类别的缺省收入表将显示男性的平均收入和女性的平均收入。

也可以对刻度变量自身进行摘要，无需使用分类变量定义组。这主要用于多个刻度变量的**堆积摘要**。请参阅第 3 页的『堆积变量』主题以获取更多信息。

### 多响应集

此外，定制表还支持一种称为**多响应集**的特殊“变量”。多响应集不是通常意义上真正的变量。它们不在数据编辑器中显示，也不能由其他过程识别。多响应集使用多个变量记录对问题的答复，其中响应者可以给出多个答案。多响应集以与分类变量相同的方式对待，可以对分类变量执行的大多数操作也可以对多响应集执行。请参阅第 69 页的第 10 章，『多响应集』主题以获取更多信息。

变量列表中每个变量旁的图标标识变量类型。

表 1. 测量级别图标













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

表 2. “多响应集”图标

“多响应集”类型	图标
多响应集，多类别	
多响应集，多二分	

可以在表构建器中更改变量的测量级别，方法是右键单击变量列表中相应的变量，然后从弹出菜单中选择**分类**或**刻度**。可以在数据编辑器的“变量视图”中永久更改变量的测量级别。定制表将定义为**标定**或**有序**的变量视为分类变量。

**类别。** 在变量列表中选择分类变量后，将在“类别”列表中显示为该变量定义的类别。如果在表中使用该变量，还将在画布窗格中显示这些类别。如果变量没有已定义的类别，那么“类别”列表和画布窗格将显示两个占位符类别：**类别 1** 和**类别 2**。

显示在表构建器中的已定义类别基于**值标签**，即指定给不同数据值的描述性标签（例如，数值 0 和 1 的值标签分别为 *male* 和 *female*）。可以在数据编辑器的“变量视图”中定义值标签，或者也可以使用“数据编辑器”窗口中“数据”菜单上的“定义变量属性”来定义值标签。

**画布窗格。** 通过将变量拖放到画布窗格的行和列上可以构建表。画布窗格显示要创建的表的预览。画布窗格不显示单元格中的实际数据值，但会提供十分准确的最终表的布局视图。对于分类变量，如果数据文件包含尚未定义值标签的唯一值，则实际表可能会比预览包含更多的类别。

- **普通视图**显示将包含在表中的所有行和列，包括汇总统计的行和/或列以及分类变量的类别。
- **压缩视图**只显示将包含在表中的变量，而不显示将包含在表中的行和列的预览。

构建表的基本规则和限制

- 对于分类变量，汇总统计基于统计源维度中最里面的变量。
- 分类变量的缺省统计源维度（行或列）基于这些变量拖放到画布窗格中的顺序。例如，如果首先将变量拖动到行托盘，则行维度是缺省的统计源维度。
- 刻度变量只能在行或列维度中最里面变量的类别中进行摘要。（可以将刻度变量放在表的任何级别，但将在最里面的级别对该变量进行摘要。）
- 刻度变量不能在其他刻度变量中进行摘要。可以对多个刻度变量的摘要进行堆积，也可以对分类变量的类别中的刻度变量进行摘要。不能将一个刻度变量嵌套在另一个刻度变量中，也不能将行维度中的一个刻度变量放在列维度中的另一个刻度变量中。
- 如果活动数据集中的任何变量包含 12,000 个以上的已定义值标签，则不能使用表构建器创建表。如果不需要在表中包含超过此限制的变量，可以定义和应用排除这些变量的变量集。如果需要在表中包含任何具有 12,000 个以上的已定义值标签的变量，则可以使用 CTABLES 命令语法生成这些表。

## 构建表

1. 从菜单中选择:

分析 > 表 > 定制表...

2. 将一个或多个变量拖放到画布窗格的行和/或列区域。
3. 单击**确定**以创建表。
4. 在画布窗格中选择（单击）变量。
5. 将该变量拖到画布窗格以外的任意位置，或按 **Delete** 键。

要更改变量的测量级别，请执行以下操作:

6. 在变量列表中右键单击变量（可以只在变量列表中执行，而不需要在画布上执行）。
7. 从弹出菜单中选择**分类或刻度**。

## 堆积变量

可以将堆积看作是将单个的表粘贴到一起并一同显示。例如，可以在同一个表的不同部分显示有关 *Gender* 和 *Age category* 的信息。

对变量进行堆积

1. 在变量列表中，选择要进行堆积的所有变量，然后将这些变量一起拖放到画布窗格的行或列中。

或

2. 单独拖放各变量，方法是将各变量拖到行中现有变量的上方或者下方，或拖到列中现有变量的左侧或者右侧。

表 3. 堆积分类变量.

变量(A)	类别	汇总统计
变量 1	类别 1	123
	类别 2	456
变量 2	类别 1	123
	类别 2	456
	类别 3	789

请参阅 第 23 页的『堆积分类变量』主题以获取更多信息。

## 嵌套变量

与交叉制表相似，嵌套可以显示两个分类变量之间的关系，区别在于使用嵌套时一个变量嵌套在同一维度中的另一个变量内部。例如，可以在行维度中的 *Age category* 内部嵌套 *Gender*，从而显示每个年龄组中的男性和女性人数。

也可以在分类变量中嵌套刻度变量。例如，可以在 *Gender* 中嵌套 *Income*，以显示男性和女性收入的独立平均值（中位数或其他汇总测量）。

对变量进行嵌套

1. 将一个分类变量拖放到画布窗格的行或列区域中。
2. 将一个分类变量或刻度变量拖放到分类行变量的左侧或者右侧，或拖放到分类列变量的上方或下方。

表 4. 嵌套分类变量.

变量 1	变量 2	汇总统计
类别 1	类别 1	12
	类别 2	34
	类别 3	56
类别 2	类别 1	12
	类别 2	34
	类别 3	56

请参阅 第 24 页的『嵌套分类变量』主题以获取更多信息。

注意：定制表不会显示分层拆分文件处理。要取得与分层拆分文件相同的结果，可将拆分文件变量放在表的最外嵌套层中。

## 图层

可以使用层为表添加深度维度，创建三维的“多维数据集”。层类似于嵌套或堆积；它与这两者的主要区别在于：每次只有一个层类别是可见的。例如，使用 *Age category* 作为行变量，*Gender* 作为层变量生成一个表，在该表中男性和女性的信息将显示在表的不同层中。

创建层

1. 在表构建器的“表格”选项卡上单击层以显示“层”列表。
2. 将要定义层的刻度变量或分类变量拖放到“层”列表中。



“层”列表不能既包含刻度变量又包含分类变量。所有变量的类型都必须相同。多响应集在“层”列表中将被视为分类变量。层中的刻度变量始终是堆积的。

如果有多个分类层变量，则可以对层进行堆积或嵌套。

- **将每个类别显示为一层**相当于堆积。将为每个层变量的每个类别分别显示一个层。总层数就是每个层变量的类别数的合计。例如，如果有 3 个层变量，每个层变量有 3 个类别，则表将有 9 个层。
- **将每个类别组合显示为一层**相当于对层进行嵌套或交叉制表。总层数为每个层变量的类别数的积。例如，如果有 3 个层变量，每个层变量有 3 个类别，则表将有 27 个层。

## 显示和隐藏变量名/或变量标签

以下选项可用于显示变量名和变量标签：

- 只显示变量标签。对未定义变量标签的任何变量，将显示变量名。这是缺省设置。
- 只显示变量名。
- 同时显示变量标签和变量名。
- 不显示变量名或变量标签。虽然包含变量标签或变量名的列/行仍显示在画布窗格的表预览中，但该列/行不会显示在实际的表中。

要显示或隐藏变量标签或变量名，请执行以下操作：

1. 在画布窗格的表预览中右键单击变量。
2. 从弹出菜单中选择**显示变量标签**或**显示变量名称**可显示或关闭显示变量标签或变量名称。将在所选变量标签或变量名的旁边出现一个选中标记，表示将显示该变量标签或变量名。

## 汇总统计

使用“汇总统计”对话框可以执行以下操作：

- 在表中添加和移去汇总统计。
- 更改统计的标签。
- 更改统计的顺序。
- 更改统计的格式，包括小数位数。

该对话框所提供的汇总统计（和其他选项）取决于汇总统计源变量的测量级别，如对话框顶部所示。汇总统计源（汇总统计所基于的变量）由以下因素确定：

- **测量级别**。如果表（或堆积表中的表部分）包含刻度变量，则汇总统计基于该刻度变量。
- **变量选择顺序**。分类变量的缺省统计源维度（行或列）基于将这些变量拖放到画布窗格上的顺序。例如，如果首先将变量拖动到行区域，则行维度是缺省的统计源维度。
- **嵌套**。对于分类变量，汇总统计基于统计源维度中最里面的变量。

一个堆积表可以有多个汇总统计源变量（刻度变量和分类变量均可），但每个表部分只能有一个汇总统计源。

更改汇总统计源维度

1. 在“表格”选项卡的“汇总统计”组中，从**源**下拉列表中选择相应的维度（行、列或层）。

控制在表中显示的汇总统计

1. 在“表格”选项卡的画布窗格中，选择（单击）相应的汇总统计源变量。
2. 在“表格”选项卡的“定义”组中，单击**汇总统计**。

或

3. 右键单击画布窗格上的汇总统计源变量，然后从弹出菜单中选择**汇总统计**。
4. 选择要包含在表中的汇总统计。可以使用箭头将所选择的统计从“统计”列表移动到“显示”列表，或者可以将所选择的统计从“统计”列表拖放到“显示”列表。
5. 单击向上或向下箭头更改当前选定汇总统计的显示位置。
6. 从所选汇总统计的“格式”下拉列表中选择一种显示格式。
7. 在“小数”单元格中输入要为所选汇总统计显示的小数位数。
8. 单击**应用到选定数据**可包含画布窗格上当前选定变量的选定汇总统计。
9. 单击**应用到全部**可包含画布窗格中类型相同的所有堆积变量的选定汇总统计。

注：**应用到全部**与**应用到选定数据**的唯一区别在于：画布窗格上已存在类型相同的堆积变量。这两个选项都会自动为表中添加的类型相同的任何其他堆积变量包含选定汇总统计。

## 分类变量的汇总统计

适用于分类变量的基本统计有计数和百分比。也可以为总计和小计指定定制汇总统计。这些定制汇总统计包括集中趋势度量（例如平均值和中位数）及可能适于一些序数分类变量的离差测量（例如标准差）。请参阅第 8 页的『分类变量的定制总计汇总统计』主题以获取更多信息。

**计数。** 每个表单元格中的个案数或多响应集的响应数。

**未加权的计数。** 每个表单元格中的未加权的个案数。仅当加权生效时，此统计才与计数有区别。

**列百分比。** 每一列中的百分比。子表的每一列中的百分比（简单百分比）的总和为 100%。通常仅在具有分类行变量时，列百分比才有用。

**行百分比。** 每一行中的百分比。子表的每一行中的百分比（简单百分比）的总和为 100%。通常仅在具有分类列变量时，行百分比才有用。

**分层行和分层列百分比。** 嵌套表中所有子表的行或列百分比（简单百分比）的总和为 100%。如果表包含层，则每个层中所有嵌套子表的行或列百分比的总和为 100%。

**层百分比。** 每个层中的百分比。对于简单百分比，当前可见层中的单元格百分比的总和为 100%。如果没有任何层变量，则此百分比等于表百分比。

**表百分比。** 每个单元格中的百分比基于整个表。所有单元格百分比都基于相同的个案总数且总和为 100%（简单百分比）。

**子表百分比。** 每个单元格中的百分比都基于子表。所有单元格百分比都基于相同的个案总数且总和为 100%。在嵌套表中，位于最里面的嵌套级别之前的变量定义子表。例如，在一个嵌套表中，*Marital status* 嵌套在 *Gender* 中，而后者又嵌套在 *Age category* 中，则由 *Gender* 定义子表。

多响应集可以具有基于个案数、响应数或计数的百分比。请参阅第 7 页的『多响应集的汇总统计』主题以获取更多信息。

## 堆积表

在进行百分比计算时，堆积变量所定义每个表部分都将被视为一个单独的表。每个堆积表部分中的分层行、分层列和表百分比的总和为 100%（简单百分比）。用于不同百分比计算的百分比基数基于每个堆积表部分中的个案数。

## 百分比基数

可以用三种不同的方法计算百分比，这些方法由计算基数中的缺失值处理确定：

**简单百分比。** 百分比基于表中所使用的个案数且总和始终为 100%。如果从表中排除了某一类别，也将从表中排除该类别中的个案。在任何情况下都将从基数中排除具有系统缺失值的个案。如果从表中排除用户缺失类别（缺省设置），也会排除具有用户缺失值的个案；相反，如果在表中包含用户缺失类别，也会包含具有用户缺失值的个案。任何其名称中不包含有效  $N$  或总计  $N$  的百分比都是简单百分比。

**总计  $N$  百分比。** 将向简单百分比基数中添加具有系统缺失值和用户缺失值的个案。百分比的总和可能小于 100%。

**有效  $N$  百分比。** 即使在表中包含用户缺失类别，也会从简单百分比基数中移去具有用户缺失值的个案。

注：任何情况下都将从基数中排除手动排除类别（用户缺失类别除外）中的个案。

## 多响应集的汇总统计

以下其他汇总统计适用于多响应集。

**列/行/层响应 %。** 基于响应数的百分比。

**列/行/层响应 %（基准：计数）。** 响应数是分子，总数是分母。

**列/行/层计数 %（基准：响应）。** 计数是分子，总响应数是分母。

**列/行/层响应 %。** 跨子表的百分比。基于响应数的百分比。

**分层列/行响应 %（基准：计数）。** 跨子表的百分比。响应数是分子，总数是分母。

**分层列/行响应 %（基准：响应）。** 跨子表的百分比。计数是分子，总响应数是分母。

**响应。** 响应计数。

**子表/表响应 %。** 基于响应数的百分比。

**子表/表响应 %（基准：计数）。** 响应数是分子，总数是分母。

**子表/表计数 %（基准：响应）。** 计数是分子，总响应数是分母。

## 刻度变量和分类定制总计的汇总统计

除了适用于分类变量的计数和百分比之外，还为刻度变量提供了以下汇总统计，并且这些汇总统计还作为分类变量的定制总计和小计摘要。这些汇总统计不适用于多响应集或字符串（字母数值）变量。

**平均值。** 算术平均值；总和除以个案数。

**中位数。** 一个值，大于该值和小于该值的个案数各占一半，第 50 个百分位。

**众数。** 出现频率最高的值。如果存在出现频率相等的值，则显示最小值。

**最小值。** 最小（最低）值。

**最大值。** 最大（最高）值。

**缺失。** 缺失值（用户和系统缺失值）计数。

**百分位数。** 可以包含第 5 个、第 25 个、第 75 个、第 95 个和/或第 99 个百分位。

**范围。** 最大值和最小值之差。

**平均值的标准误差。** 取自同一分布的样本与样本之间的平均值之差的测量。此方法可用于粗略地将观察到的平均值与假设值进行比较（即，如果差分与标准误差之比小于 -2 或大于 +2，那么可以断定两个值不同）。

**标准差。** 对围绕平均值的离差的测量。在正态分布中，68% 的个案在平均值的一个标准差范围内，95% 的个案在平均值的两个标准差范围内。例如，在正态分布（方差的平方根）中，如果平均年龄为 45，标准差为 10，则 95% 的个案将处于 25 到 65 之间。

**和。** 值的总和。

**合计百分比。** 基于总和的百分比。适用于行和列（在子表中）、所有行和列（跨子表）、层、子表和整个表。

**总计 N。** 无缺失值、用户缺失值和系统缺失值的计数。不包含手动排除的类别（用户缺失类别除外）中的个案。

**有效 N。** 无缺失值的计数。不包含手动排除的类别（用户缺失类别除外）中的个案。

**方差。** 对围绕平均值的离差的测量，值等于与平均值的差的平方和除以个案数减一。度量方差的单位是变量本身的单位的平方（标准差的平方）。

堆积表

堆积变量所定义每个表部分均被视为一个单独表，并将相应地为其计算汇总统计。

## 分类变量的定制总计汇总统计

对于包含总计或小计的分类变量表，可以使用不同于为每个类别所显示的摘要的汇总统计。例如，可以为序数分类行变量显示计数和列百分比，并为“总计”统计显示中位数。

要创建包含定制总计汇总统计的分类变量表，请执行以下操作：

1. 从菜单中选择：

分析 > 表 > 定制表...

将打开表构建器。

2. 将一个分类变量拖放到画布的“行”或“列”区域。
3. 右键单击画布上的该变量，然后从弹出菜单中选择**类别和总计**。
4. 单击（选中）**总计**复选框，然后单击**应用**。
5. 再次右键单击画布上的该变量，然后从弹出菜单中选择**汇总统计**。
6. 单击（选中）**定制关于总计和小计的汇总统计**，然后选择所需的定制汇总统计。

缺省情况下，所有汇总统计（包括定制摘要）都将显示在与包含分类变量的维度相反的维度中。例如，如果包含分类行变量，则汇总统计在表中定义列，如下表所示：

表 5. 有序变量类别位于行中，汇总统计计数和平均值位于列中。

变量(A)	类别	计数	平均值(M)
变量 1	1 同意	196	
	2 中立	936	
	3 不同意	744	

表 5. 有序变量类别位于行中，汇总统计计数和平均值位于列中 (续).

变量(A)	类别	计数	平均值(M)
	总计(T)	1876	2.29

要在与分类变量相同的维度中显示汇总统计，请执行以下操作：

7. 在表构建器的“表格”选项卡上，从“汇总统计”组的“位置”下拉列表中选择相应的维度。

例如，如果要在行中显示分类变量，请从下拉列表中选择行。

## 汇总统计显示格式

可用的显示格式选项有：

**nnnn**。简单数值。

**nnnn%**。在值末尾加上百分比符号。

**自动**。已定义变量显示格式，包括小数位数。

**N=nnnn**。在值前面显示  $N=$ 。在未显示汇总统计的表中，此格式可用于计数、有效  $N$  和总计  $N$ 。

**(nnnn)**。所有值都用括号括起。

**(nnnn) (负值)**。只有负值用括号括起。

**(nnnn%)**。所有值都用括号括起，并在值末尾加上一个百分比符号。

**n,nnn.n**。逗号格式。无论语言环境设置如何，均使用逗号作为分组分隔符，使用句点作为小数指示符。

**n.nnn,n**。句点格式。无论语言环境设置如何，均使用句点作为分组分隔符，使用逗号作为小数指示符。

**\$n,nnn.n**。美元格式。在值前面显示美元符号；无论语言环境设置如何，均使用逗号作为分组分隔符，使用句点作为小数指示符。

**CCA、CCB、CCC、CCD、CCE**。定制货币格式。在列表中显示每个定制货币的当前定义格式。在“选项”对话框（“编辑”菜单，“选项”）的“货币”选项卡中定义这些格式。

一般规则和限制

- 除“自动”之外，小数位数均由“小数”列设置确定。
- 除逗号、美元符号和点格式之外，所使用的小数指示符均为为当前语言环境设置（在 Windows 控制面板的“区域选项”中）定义的小数指示符。
- 虽然逗号/美元符号和句点将分别显示为作为分组分隔符的逗号或句点，但在创建时未提供根据当前语言环境设置（在 Windows 控制面板的“区域选项”中定义）显示分组分隔符的显示格式。

## 分类和总计

使用“分类和总计”对话框，可以执行以下操作：

- 重新排列和排除类别。
- 插入小计和总计。
- 插入已计算类别。

- 包含或排除空类别。
- 包含或排除定义为包含缺失值的类别。
- 包含或排除尚未定义值标签的类别。
- 此对话框仅适用于分类变量和多响应集。它不适用于刻度变量。
- 对于具有不同类别的多个选定变量，不能插入小计、插入已计算类别、排除类别或手动重新排列类别。只有在画布预览中选择多个变量并在此对话框中同时访问所有选定变量时，才能执行这些操作。但仍可以为每个变量单独执行这些操作。
- 对于未定义值标签的变量，只能排列类别和插入总计。

访问“分类和总计”对话框

1. 将某一分类变量或多响应集拖放到画布窗格中。
2. 右键单击画布窗格上的该变量，然后从弹出菜单中选择**分类和总计**。

或

3. 在画布窗格中选择（单击）该变量，然后在“表格”选项卡的“定义”组中单击**分类和总计**。

还可以在画布窗格的同一维度中选择多个分类变量，方法是：

4. 按住 **Ctrl** 并单击画布窗格中的各个变量。

或

5. 在画布窗格的表预览外部单击，然后单击并拖动鼠标以选择包含要选择变量的区域。

或

6. 右键单击维度中的任意变量，然后选择**选择全部 [维度] 变量**以选择该维度中的所有变量。

重新排列类别

要手动重新排列类别，请执行以下操作：

1. 选择（单击）列表中的一个类别。
2. 单击向上或向下箭头以在列表中上移或下移该类别。

或

3. 单击该类别的“值”列，然后将该列拖放到其他位置。

排除类别

1. 选择（单击）列表中的一个类别。
2. 单击“排除”列表旁边的箭头。

或

3. 单击该类别的“值”列，然后将该列拖到列表外的任意位置。

如果排除任何类别，则也将排除任何未定义值标签的类别。

对类别排序

可以按数据值、值标签、单元格计数或汇总统计对类别进行升序或降序排序。

1. 在“对类别排序”组中单击“依据”下拉列表，然后选择要使用的排序条件：值、标签、计数或汇总统计（例如，平均值、中位数或众数）。可用于排序的汇总统计取决于选择在表中显示的汇总统计。

2. 单击“顺序”下拉列表以选择排序顺序（升序或降序）。

如果已排除任何类别，则“对类别排序”不可用。

#### 小计

1. 在列表中选择（单击）要包含在小计中的类别范围中的最后一个类别。
2. 单击**添加小计...**。
3. 在“定义小计”对话框中，可选择性地修改小计标签文本。
4. 要只显示小计而不显示定义小计的类别，请选择**隐藏表格中的小计类别**。
5. 单击**继续添加小计**。

#### 总计

1. 单击**总计**复选框。也可以修改总计标签文本。

如果选定变量嵌套在另一个变量中，则将为每个小计插入总计。

#### 显示总计和小计的位置

总计和小计可显示在每个总计内的类别的上方或下方。

- 如果在“显示总计和小计”组中选择了下，则总计将显示在每个子表的上方，并且从选定类别（包括该类别）以上到上一个小计之间的所有类别都将包含在每个小计内。
- 如果在“显示总计和小计”组中选择了上，则总计将显示在每个小计的下方，并且从选定类别（包括该类别）以下到下一个小计的所有类别都将包含在每个小计内。

**重要提示：**必须先选择小计的显示位置，然后再定义小计。更改显示位置会影响所有小计（而不仅仅是当前选择的小计），并且还会更改小计中包含的类别。

#### 已计算类别

您可以显示从汇总统计、总计、小计和/或常量中计算得到的类别。请参阅第 12 页的『已计算类别』主题以获取更多信息。

#### 定制总计和小计汇总统计

使用“汇总统计”对话框，可以在表的总计和小计区域中显示除“总计”之外的统计。请参阅第 6 页的『分类变量的汇总统计』主题以获取更多信息。

**注意：**如果您所选的多个定制总计统计同时还位于表主体中，并且您隐藏了统计标签，那么将按与表主体相同的顺序重新排列总计；由于未显示标签，您可能不知道每个总计统计实际表示的含义。总之，选择多个统计并隐藏统计标签并不是一个好方法。

#### 总计、小计和已排除类别

已排除类别中的个案不包含在总计中。

#### 缺失值、空类别和不含值标签的值

**缺失值。** 此选项控制用户缺失值或定义为包含缺失值的值（例如，代码 99 表示怀孕“不适用”于男性）的显示。缺省情况下，将排除用户缺失值。选择（选中）此选项可在表中包含用户缺失类别。虽然变量可包含多个缺失值类别，但画布上的表预览将只显示一个一般的缺失值类别。所有已定义的用户缺失类别都将包含在表中。始终排除**系统缺失值**（数据编辑器中数值变量的空单元格）。

**空类别。** 空类别包含已定义值标签，但不包含特定表或子表的个案。缺省情况下，空类别包含在表中。取消选中此选项可从表中排除缺失类别。

**扫描数据时找到的其他值。** 缺省情况下，将数据文件中尚未定义值标签的类别值自动包含在表中。取消选中此选项可从表中排除未定义值标签的值。如果排除已定义值标签的任何类别，也将排除未定义值标签的类别。

## 已计算类别

除了显示汇总统计的汇总结果外，表格还可显示从这些汇总结果、常数值、小计和总计或上述组合后计算得到的一个或多个类别。这些结果称为已计算或后期计算的类别。已计算类别与单个变量中的类别相似，二者有下列相似和不同之处：

- 已计算类别与其他类别的位置相似。
- 已计算类别在与其他类别相同的统计上操作。
- 已计算类别不会影响小计、总计或显著性检验。
- 缺省情况下，已计算类别的值使用与其他类别相同的汇总统计格式。您可以在定义已计算类别时覆盖此格式。

由于已计算类别可以用于计算汇总结果的总计，因此它们与小计类似。不过，与小计相比，已计算类别具有下列优点：

- 可以从其他小计的结果计算已计算类别。
- 已计算类别可以相互重叠，在相同（或部分相同）类别上操作。
- 已计算类别不必包含此计算类别以上或以下所有其他类别的值。即已计算类别为非穷举。
- 已计算类别可以包括不相邻类别的值。

与总计和小计不同，已计算类别是从汇总数据而不是原始数据计算得出的。因此，已计算类别的值可能与总计及小计结果不匹配。此外，由于您可以在定义已计算类别时选择隐藏源类别，因此可能很难在结果表格中解释小计。如果使用已计算类别，建议为小计指定自定义标签。

### 定义已计算类别

从“分类和总计”对话框添加已计算类别。有关访问该对话框的信息，请参阅第 9 页的『分类和总计』主题。

1. 在“分类和总计”对话框中，单击**添加类别...**
2. 在**已计算类别的标签**中，为已计算类别指定标签。您可以从“类别”列表拖动类别以包含这些类别的标签。
3. 通过选择类别和/或总计和小计并使用运算符定义已计算类别来构建表达式。您还可以键入常数值（例如，500）以包括在表达式中。
4. 要只显示已计算类别而不显示定义已计算类别的类别，请选择**隐藏表格中表达式使用的类别**。
5. 单击**显示格式**选项卡更改已计算类别的显示格式和小数位数。请参阅『已计算类别的显示格式』主题以获取更多信息。
6. 单击**继续添加已计算类别**。

### 已计算类别的显示格式

缺省情况下，已计算类别使用与变量中其他类别相同的显示格式和小数位数。您可以在“已计算类别”对话框中的“显示格式”选项卡上覆盖这些设置。“显示格式”选项卡列出除这些统计的显示格式和小数位数之外的已计算类别操作的当前汇总统计。

对于每个汇总统计，您可以：



1. 从汇总统计的“格式”下拉列表中选择一种显示格式。有关显示格式的完整列表，请参阅第 9 页的『汇总统计显示格式』主题。
2. 在“小数”单元格中输入要为所选汇总统计显示的小数位数。

## 具有共享类别的变量表（共周界表）

调查常常包含许多问题，每个问题又都包含一组相同的、响应者可能作出的回答。可以使用堆积在同一个表中显示这些相关的变量，并在表的列中显示共用的响应类别。

为具有共享类别的多个变量创建表

表 6. 列中包含共享响应类别的堆积变量

变量(A)	类别 1	类别 2	类别 3
变量 1	12	34	56
变量 2	56	12	34
变量 3	34	56	12

请参阅第 43 页的第 6 章，『具有共享类别的变量的表』主题以获取更多信息。

## 定制表构建器

与常用对话框不同，可以用与更改任何常用窗口大小的相同方法更改表构建器的大小：

1. 单击并拖动表构建器的顶部、底部、任一边或任一角可增加或减少其大小。

此外，还可以在“表格”选项卡上更改“变量”列表、“类别”列表和画布窗格的大小。

2. 单击并拖动“变量”列表和“类别”列表之间的水平条可延长或缩短列表。下移该水平条可以延长“变量”列表并缩短“类别”列表。上移该水平条则正好相反。
3. 在画布窗格中单击并拖动“变量”列表和“类别”列表之间的垂直条可使列表变宽或变窄。画布会根据剩余空间自动调整大小。

## 定制表：“选项”选项卡

使用“选项”选项卡，可以执行以下操作：

- 指定将显示在空单元格以及不能计算统计的单元格中的内容。
- 控制如何在刻度变量统计计算中处理缺失值。
- 设置最小和/或最大数据列宽度。
- 控制如何处理多类别集中的重复响应。

**数据单元格外观。** 控制将显示在空单元格以及不能计算统计的单元格中的内容。

- **空单元格。** 对于未包含个案的表单元格（单元格计数为 0），可以选择以下任一显示选项（共三个选项）：零、空白或指定的文本值。文本值最长可以有 255 个字符。
- **无法计算的统计数据。** 如果不能计算统计（例如，不包含个案的类别的平均值），则显示文本。文本值最长可以有 255 个字符。缺省值为一个句点 (.)。

**数据列宽度。** 控制数据列的最小和最大列宽。此设置不影响行标签的列宽。

- **表格外观设置。** 根据当前缺省表格外观。可以创建自己的定制缺省表格外观以在创建新表时使用，并且可以用表格外观控制行标签列和数据列宽度。

- **定制。** 覆盖数据列宽的缺省表格外观设置。指定表的最小和最大数据列宽，度量单位为：点、英寸或厘米。

**刻度变量缺失值。** 对于包含两个或更多刻度变量的表，控制刻度变量统计缺失数据的处理方式。

- **最大限度的使用可用数据 (按变量逐个删除)。** 每个刻度变量的具有有效值的所有个案都将包含在该刻度变量的汇总统计中。
- **在所有刻度变量中使用相同的大小写格式 (按列表状态删除)。** 表中任何刻度变量的具有缺失值的个案都将从表的所有刻度变量的汇总统计中排除。

**计算多个类别集中的重复响应。** 重复响应是多类别集中两个或多个变量的同一响应。缺省情况下，不对重复响应进行计数，但这可能会成为一个要包含在计数中的完全有效条件（例如，某一多类别集表示响应者所购买的最后三辆车的制造商）。

**隐藏小计数。** 您可以选择隐藏小于指定整数的计数。隐藏值将显示为 **<N**，其中 **N** 是指定的整数。指定的整数必须大于等于 2。

---

## 定制表：“标题”选项卡

“标题”选项卡控制标题、文字说明和角标签的显示。

**标题。** 在表上方显示的文本。

**文字说明。** 在表下方和任何脚注上方显示的文本。

**脚注。** 显示在透视表的左上角的文本。仅当表包含行变量以及透视表行维度标签属性设置为**嵌套**时，才显示角文本。这不是缺省表格外观设置。

可以在表标题、文字说明或角标签中包含以下自动生成的值：

**日期。** 基于当前 Windows “区域选项”设置的格式显示的当前年、月和日。

**时间。** 基于当前 Windows “区域选项”设置的格式显示的当前小时、分钟和秒。

**表表达式。** 在表中使用的变量以及在表中使用这些变量的方式。如果变量具有已定义的变量标签，则将显示该标签。在生成的表中，以下符号表示在表中使用变量的方式：

- **+** 表示堆积变量。
- **>** 表示嵌套。
- **BY** 表示交叉制表或层。

---

## 定制表：“检验统计”选项卡

使用“检验统计”选项卡可以为定制表请求各种显著性检验，其中包括：

- 独立性卡方检验。
- 列平均值等同性检验。
- 列比例等同性检验。
- 多响应集和小计显著性检验。（有关多响应集的显著性检验的信息，请参见第 73 页的『对多响应集进行显著性检验』。）

这些检验不适用于类别标签移出其缺省表维度的表。

**比较列的平均值 (t-检验)。** 此选项为表生成列平均值等同性成对检验，该表的列中至少有一个分类变量且行中至少有一个刻度变量。该表必须包含平均值作为汇总统计。可以使用 Bonferroni 方法选择是否调整检验的  $p$  值。此外，还可以指定检验的  $\alpha$  水平， $\alpha$  水平应该是一个大于 0 且小于 1 的值。最后，虽然平均值检验的方差始终只基于多重响应检验的比较类别；但对于序数分类变量，可只根据比较的类别或所有类别估计该变量。

**比较列的比例 (z-检验)。** 此选项为表生成列比例等同性成对检验，该表的行和列中至少同时有一个类别变量。该表必须包含计数或简单的列百分比。可以使用 Bonferroni 方法选择是否调整检验的  $p$  值。还可以指定检验的  $\alpha$  水平， $\alpha$  水平应该是一个大于 0 且小于 1 的值。

**识别显著性差异。** 如果您选择了**比较列的平均值**或**比较列的比例**，则可以选择如何指示显著性差异。

- **在单独的表中。** 显著性检验结果显示在单独的表中。如果两个值显著不同，则较大值对应的单元格将显示键以标识较小值所在列。示例如下：
- **在主表中使用 APA 样式下标。** 主表本身通过 APA 样式格式使用下标字母来标识显著性差异。如果两个值显著不同，这些值将显示不同的下标字母。这些下标并非脚注。当此选项有效时，在当前表格外观中定义脚注样式将被覆盖，并且脚注将显示为上标数字。示例如下：

有关如何创建和解释这些表的完整示例，请参阅第 64 页的『比较列的比例』。

**注意：** APA 样式表包含一个文字说明，该说明用于阐述如何解释表中的下标。如果您在“标题”选项卡上指定了一个文字说明，那么“标题”选项卡上将显示该指定的文字说明，而不显示用于解释下标的文字说明。

**独立性检验 (卡方验证)。** 此选项为表生成独立性卡方检验，该表的行和列中至少同时有一个分类变量。还可以指定检验的  $\alpha$  水平， $\alpha$  水平应该是一个大于 0 且小于 1 的值。

**使用小计代替小计类别。** 如果选择此选项，则每个小计将替换其类别以用于显著性检验。否则，只有隐藏其小计类别的小计才替换其类别以用于检验。

**检验中包含多重响应变量。** 如果选择此选项，则使用多响应集类别执行检验。否则，在执行检验时，将忽略多响应集。



---

## 第 2 章 分类变量的简单表

---












### 分类变量的简单表

您要创建的大多数表都可能包含至少一个**分类变量**。分类变量是包含有限数量的不同值或类别（例如，性别或宗教）的变量。分类变量可以是**名义变量**，也可以是**有序变量**。

- **名义 (Nominal)**. 当变量值表示不具有内在等级的类别时，该变量可以作为名义变量；例如，雇员任职的公司部门。名义变量的示例包括地区、邮政编码和宗教信仰。
- **有序 (Ordinal)**. 当变量值表示带有某种内在等级的类别时，该变量可以作为有序变量；例如，从十分不满意到十分满意的服务满意度水平。有序变量的示例包括表示满意度或可信度的态度分数和优先选择评分。

变量列表中每个变量旁的图标标识变量类型。

表 7. 测量级别图标

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

定制表已经优化，能够与定义了**值标签**的分类变量一起使用。请参阅第 1 页的『构建表』主题以获取更多信息。

#### 样本数据文件

本章使用数据文件 *survey\_sample.sav*。请参阅数据文件主题以获取更多信息。

此处提供的所有示例都在对话框中按字母顺序显示变量标签。变量列表显示属性是在“选项”对话框（“编辑”菜单，“选项”）的“常规”选项卡中设置的。

---

### 单个分类变量

尽管单个分类变量的表可能是您所创建的最简单的表之一，但是它通常会提供您所需要的全部信息。

1. 从菜单中选择：

分析 > 表 > 定制表...

2. 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。

这时在画布窗格上会显示表的预览。该预览不显示真实的数据值；它只在要显示数据的位置显示占位符。

3. 单击**确定**以创建表。

“查看器”窗口中将显示该表。

		计数
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

图 2. 行中的单个分类变量

在这个简单的表中，列标题计数不是真正必需的，所以可以创建不包含该列标题的表。

- 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
- 在“汇总统计”组中，对于“位置”，选择（单击）**隐藏**。
- 单击**确定**以创建表。

Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

图 3. 没有汇总统计列标签的单个分类变量

## 百分比

除了计数以外，还可以显示百分比。对于单个分类变量的简单表，如果在行中显示变量，您可能希望查看列百分比。相反，对于在列中显示的变量，您可能希望查看行百分比。

- 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
- 在“汇总统计”组中对“位置”取消选中**隐藏**。由于表将有两个列，因此您会希望显示列标签以便了解每列所代表的内容。
- 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**汇总统计**。
- 在“汇总统计”对话框中，在“统计”列表中选择**列 N %**，然后单击箭头将其添加到“显示”列表中。
- 在“显示”列表的“标签”单元格中，删除缺省标签并键入百分比。
- 单击**应用到选定数据**，然后在表构建器中单击**确定**创建表。

		计数	百分比
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%

图 4. 计数和列百分比

## 总计

总计不自动包含在定制表中，但是很容易便可以将总计添加到表中。

1. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择类别和总计。
3. 在“分类和总计”对话框中选择（单击）总计。
4. 单击应用，然后在表构建器中单击确定创建表。

		计数	百分比
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	合计	2828	100.0%

图 5. 计数、列百分比和总计

请参阅第 31 页的第 4 章，『分类变量的总计和小计』主题以获取更多信息。

## 交叉列表

交叉制表是用于检查两个分类变量之间关系的基本方法。例如，使用 *Age category* 作为行变量，*Gender* 作为列变量，可以创建一个显示每个年龄组中男性和女性人数的二维交叉表格。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击重置删除表构建器中所有以前的选择。
3. 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
4. 将 *Gender* 从变量列表拖放到画布窗格上的“列”区域中。（可能需要向下滚动浏览变量列表才能找到此变量。）
5. 单击确定以创建表。

		Gender	
		Male	Female
		计数	计数
Age category	Less than 25	108	134
	25 to 34	276	351
	35 to 44	309	370
	45 to 54	221	260
	55 to 64	136	184
	65 or older	178	301

图 6. *Age category* 和 *Gender* 的交叉制表

## 交叉制表中的百分比

在二维交叉制表中，行和列百分比可能会提供有用的信息。

1. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 在画布窗格上右键单击 *Gender*。

您可能会注意到，弹出菜单中禁用了**汇总统计**。这是因为只能为统计源维度中最内层的变量选择汇总统计。分类变量的缺省统计源维度（行或列）基于将这些变量拖放到画布窗格上的顺序。在本例中，我们先将 *Age category* 拖动到行维度中；由于行维度中没有任何其他变量，因此 *Age category* 便是统计源变量。可以更改统计源维度，但在本例中不需要这么做。请参阅第 5 页的『汇总统计』主题以获取更多信息。

3. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**汇总统计**。
4. 在“汇总统计”对话框中，在“统计”列表中选择**列 N %**，然后单击箭头将其添加到“显示”列表中。
5. 在“统计”列表中选择**行 N %**，然后单击箭头将其添加到“显示”列表中。
6. 单击**应用到选定数据**，然后在表构建器中单击**确定**创建表。

		Gender					
		Male			Female		
		计数	列 N %	行 N %	计数	列 N %	行 N %
Age category	Less than 25	108	8.8%	44.6%	134	8.4%	55.4%
	25 to 34	276	22.5%	44.0%	351	21.9%	56.0%
	35 to 44	309	25.2%	45.5%	370	23.1%	54.5%
	45 to 54	221	18.0%	45.9%	260	16.3%	54.1%
	55 to 64	136	11.1%	42.5%	184	11.5%	57.5%
	65 or older	178	14.5%	37.2%	301	18.8%	62.8%

图 7. 具有行和列百分比的交叉制表

## 控制显示格式

可以控制显示格式，例如汇总统计中显示的小数位数。例如，缺省情况下以一位小数和一个百分号的格式显示百分比。但是如果您希望单元格值显示两位小数并且不包含百分号，该怎么做？

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**汇总统计**。
3. 对于所选择的两个百分比汇总统计（**列 N %** 和**行 N %**），从“格式”下拉列表中选择 **nnnn.n**，然后在两个统计的“小数”单元格中都键入 2。
4. 单击**确定**以创建表。

		Gender					
		Male			Female		
		计数	列 N %	行 N %	计数	列 N %	行 N %
Age category	Less than 25	108	8.79	44.63	134	8.38	55.37
	25 to 34	276	22.48	44.02	351	21.94	55.98
	35 to 44	309	25.16	45.51	370	23.13	54.49
	45 to 54	221	18.00	45.95	260	16.25	54.05
	55 to 64	136	11.07	42.50	184	11.50	57.50
	65 or older	178	14.50	37.16	301	18.81	62.84

图 8. 行和列百分比的格式化单元格显示

## 边际总计

在交叉制表中显示**边际总计**（每个行和列的总计）是相当常见的操作。由于缺省情况下“定制表”中不包含这些项，因此需要将其显式地添加到表中。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。



- 单击**重置**删除表构建器中所有以前的选择。
- 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
- 将 *Gender* 从变量列表拖放到画布窗格上的“列”区域中。（可能需要向下滚动浏览变量列表才能找到此变量。）
- 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**类别和总计**。
- 在“分类和总计”对话框中选择（单击）**总计**，然后单击**应用**。
- 右键单击画布窗格上的 *Gender*，然后从弹出菜单中选择**类别和总计**。
- 在“分类和总计”对话框中选择（单击）**总计**，然后单击**应用**。
- 在“汇总统计”组中，对于“位置”，选择（单击）**隐藏**。（由于只显示计数，因此不需要标识表的数据单元格中显示的“统计”。）
- 单击**确定**以创建表。

		Gender		
		Male	Female	合计
Age category	Less than 25	108	134	242
	25 to 34	276	351	627
	35 to 44	309	370	679
	45 to 54	221	260	481
	55 to 64	136	184	320
	65 or older	178	301	479
	合计	1228	1600	2828

图 9. 具有边际总计的交叉制表

## 对类别排序和排除类别

缺省情况下，类别按类别值标签所代表的数值值的升序显示。例如，尽管对于年龄组显示 *Less than 25*、*25 to 34*、*35 to 44* 等等的值标签，但是实际的基础数据值是 1、2、3 等等，并且控制类别的缺省显示顺序的正是这些基础数据值。

可以很容易地更改类别的顺序以及排除不希望在表中显示的类别。

对类别排序

可以手动重新排列类别或按下列项目的升序或降序对类别排序：

- 数据值。
  - 值标签。
  - 单元格计数。
  - 汇总统计。可用于排序的汇总统计取决于选择在表中显示的汇总统计。
1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
  2. 如果画布窗格上的“行”区域中尚未显示 *Age category*，则请将其拖放到此处。
  3. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**类别和总计**。

数据值和相关联的值标签按当前顺序显示。在本例中，当前顺序仍然是数据值的升序。

4. 在“对类别排序”组中，从“顺序”下拉列表中选择**降序**。

这时排列顺序逆转。

5. 从“依据”下拉列表中选择**标签**。

现在类别按值标签的字母顺序降序排列。

请注意标注为 *Less than 25* 的类别位于列表的顶端。在字母顺序中，字母排在数字之后。由于它是唯一一个以字母开头的标签并且列表以降序（反向）排列，因此该类别排在列表的顶端。

如果您希望在列表的不同位置显示特定类别，则可以很容易地将其移动到该处。

6. 单击“标签”列表中标注为 *Less than 25* 的类别。
7. 单击列表右侧的向下箭头。类别在列表中向下移动一行。
8. 继续单击向下箭头，直到该类别位于列表的底部。

### 排除类别

如果您不希望在表中显示某些类别，则可以将其排除。

1. 单击“标签”列表中标注为 *Less than 25* 的类别。
2. 单击“排除”列表左侧的箭头键。
3. 单击“标签”列表中标注为 *65 or older* 的类别。
4. 再次单击“排除”列表左侧的箭头键。

这两个类别即从“显示”列表移到“排除”列表中。如果您改变了想法，很容易便可将其移回“显示”列表。

5. 单击**应用**，然后在表构建器中单击**确定**创建表。

		性别		
		男	女	总计
年龄 类别	55 至 64	136	184	320
	45 至 54	221	260	481
	35 至 44	309	370	679
	25 至 34	276	351	627
	总计	942	1165	2107

图 10. 按值标签降序排列并排除某些类别的表

可以看到，总计要比排除两个类别之前的总计小。这是因为总计基于表中包含的类别。已排除的所有类别都被排除在总计计算之外。请参阅第 31 页的第 4 章，『分类变量的总计和小计』主题以获取更多信息。

---

## 第 3 章 对分类变量使用堆积、嵌套和层

堆积、嵌套和层都是在同一个表中显示多个变量的方法。本章 重点介绍这些方法对分类变量的应用。

样本数据文件

本章 使用数据文件 *survey\_sample.sav*。请参阅数据文件主题以获取更多信息。

此处提供的所有示例都在对话框中按字母顺序显示变量标签。变量列表显示属性是在“选项”对话框（“编辑”菜单，“选项”）的“常规”选项卡中设置的。

---

### 堆积分类变量

可以将堆积看作是将单个的表粘贴到一起并一同显示。例如，可以在同一个表的不同部分显示有关 *Gender* 和 *Age category* 的信息。

1. 从菜单中选择:

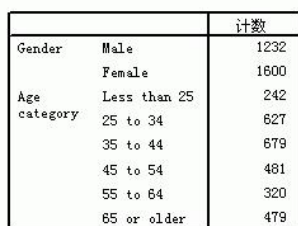
分析 > 表 > 定制表...

2. 在表构建器中，将 *Gender* 从变量列表拖放到画布窗格上的“行”区域中。

3. 将 *Age category* 从变量列表拖放到 *Gender* 下面的“行”区域中。

现在这两个变量堆积在行维度中。

4. 单击**确定**以创建表。



		计数
Gender	Male	1232
	Female	1600
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	45 to 54	481
	55 to 64	320
	65 or older	479

图 11. 在行中堆积分类变量的表

也可以用类似的方式在列中堆积变量。

### 对交叉制表使用堆积

堆积表可以在其他维度中包含其他变量。例如，可以对两个堆积在行中的变量以及显示在列维度中的第三个变量进行交叉制表。

1. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。

2. 如果 *Age category* 和 *Gender* 尚未堆积在行中，请按照上述操作堆积这两个变量。

3. 将 *Get news from internet* 从变量列表拖放到画布窗格上的“列”区域中。

4. 单击**确定**以创建表。

		Get news from internet	
		No	Yes
		计数	计数
Gender	Male	873	359
	Female	1092	508
Age category	Less than 25	146	96
	25 to 34	368	259
	35 to 44	435	244
	45 to 54	346	135
	55 to 64	252	68
	65 or older	416	63

图 12. 两个堆积行变量与一个列变量进行交叉制表

注：有多个具有以 *Get news from ...* 开头的标签的变量，因此在变量列表中可能很难对它们进行区分（因为标签可能很长，以致不能在变量列表中完整显示）。有两种方法可查看完整的变量标签：

- 将鼠标指针定位在列表中的变量上，以便在弹出的工具提示中显示完整的标签。
- 单击并拖动用于将变量列表和类别列表与画布窗格分开的垂直条，以增加这两个列表的宽度。

## 嵌套分类变量

与交叉制表相似，嵌套可以显示两个分类变量之间的关系，区别在于使用嵌套时一个变量嵌套在同一维度中的另一个变量内部。例如，可以在行维度中的 *Age category* 内部嵌套 *Gender*，从而显示每个年龄组中的男性和女性人数。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击**重置**删除表构建器中所有以前的选择。
3. 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
4. 将 *Gender* 从变量列表拖放到“行”区域中 *Age category* 的右侧。

现在画布窗格上的预览显示嵌套表将只包含计数列，并且每个单元格包含各年龄类别中男性或女性的人数。

您可能会注意到变量标签 *Gender* 被重复显示，为各个年龄类别分别显示了一次。通过将类别最少的变量放置在最外面的嵌套层可以最大限度地减少这种重复的次数。

5. 单击画布窗格上的变量标签 *Gender*。
6. 将该变量拖放到“行”区域中尽可能靠左的位置。

现在只重复两次 *Age category*，而不是重复六次 *Gender*。这个表更加整洁，但生成的结果与之前的结果在本质上是相同的。

7. 单击**确定**以创建表。

				计数
Gender	Male	Age category	Less than 25	108
			25 to 34	276
			35 to 44	309
			45 to 54	221
			55 to 64	136
			65 or older	178
	Female	Age category	Less than 25	134
			25 to 34	351
			35 to 44	370
			45 to 54	260
			55 to 64	184
			65 or older	301

图 13. 嵌套在 Gender 之内的 Age category 的表

注意: 定制表不会显示分层拆分文件处理。要取得与分层拆分文件相同的结果, 可将拆分文件变量放在表的最外嵌套层中。

## 不显示变量标签

解决嵌套表中变量标签冗余问题的另一种方法很简单: 取消变量名或标签的显示。由于在没有变量标签的情况下, Gender 和 Age category 的值标签也可能具有充分的描述性, 因此我们可以删除这两个变量的标签。

1. 打开表构建器 (“分析”菜单, “表格”, “定制表”)。
2. 右键单击画布窗格上的 Age category, 然后在弹出菜单中取消选中显示变量标签。
3. 对 Gender 执行相同的操作。

表预览中仍然显示变量标签, 但是表中将不会包含这些标签。

4. 单击确定以创建表。

		计数
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

图 14. 不带变量标签的嵌套表

如果您希望在表中的某个位置包含变量标签, 而不在表的主体中多次显示这些标签, 可以将其包含在表标题或角标签中。

5. 打开表构建器 (“分析”菜单, “表格”, “定制表”)。
6. 单击标题选项卡。
7. 单击“标题”文本框中的任意位置。
8. 单击表表达式。“标题”文本框中将显示文本 &[表表达式]。这会生成一个包含表中所用变量的变量标签的表格标题。
9. 单击确定以创建表。

		计数
Male	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178
Female	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

图 15. 表格标题中的变量标签

标题中的大于号 (>) 指示 *Age category* 嵌套在 *Gender* 之内。

## 嵌套交叉制表

嵌套表可以在其他维度中包含其他变量。例如，可以在行中将 *Age category* 嵌套在 *Gender* 之内，然后将嵌套行与列维度中的第三个变量进行交叉制表。

1. 打开表构建器 (“分析”菜单, “表格”, “定制表”)。
2. 如果 *Age category* 尚未在行中嵌套在 *Gender* 之内, 则请按照上述操作嵌套这两个变量。
3. 将 *Get news from internet* 从变量列表拖放到画布窗格上的“列”区域中。

您可能会注意到表格太大, 无法在画布窗格上完整地显示。可以在画布窗格上向上/向下或向右/向左滚动以便查看表预览的更多部分, 或者可以执行以下操作:

- 在表构建器中单击**压缩**以查看紧凑视图。紧凑视图只显示变量标签, 表中不包含任何有关类别或汇总统计的信息。
  - 单击并拖动表构建器的任何侧边或角, 以放大表构建器。
4. 单击**确定**以创建表。

				Get news from internet	
				No	Yes
				计数	计数
Gender	Male	Age category	Less than 25	59	49
			25 to 34	159	117
			35 to 44	217	92
			45 to 54	169	52
			55 to 64	112	24
			65 or older	155	23
	Female	Age category	Less than 25	87	47
			25 to 34	209	142
			35 to 44	218	152
			45 to 54	177	83
			55 to 64	140	44
			65 or older	261	40

图 16. 嵌套交叉制表

## 对调行与列

假设您花了很长时间设置一个复杂的表，并认为它接近完美，唯一需要做的改动是交换方向，即将行变量放到列中，将列变量放到行中。这时，您该怎样做？例如，您已创建了一个嵌套交叉制表，*Age category* 和 *Gender* 嵌套在行中，但是现在您希望将这两个人口统计变量嵌套在列中。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上的任意位置，然后从弹出菜单中选择**对调行与列变量**。

现在已经切换了行和列变量。

在创建表之前让我们做一些修改，以使显示更加清晰。

3. 选择**隐藏**以取消汇总统计列标签的显示。
4. 右键单击画布窗格上的 *Gender*，然后取消选中**显示变量标签**。
5. 现在单击**确定**以创建表。

		Male						Female					
		Age category						Age category					
		Less than 25	25 to 34	35 to 44	45 to 54	55 to 64	65 or older	Less than	25 to 34	35 to 44	45 to 54	55 to 64	65 or older
Get news from internet	No	59	159	217	169	112	155	87	209	218	177	140	261
	Yes	49	117	92	52	24	23	47	142	152	83	44	40

图 17. 将人口统计变量嵌套在列中的交叉制表

## 图层(L)

可以使用层为表添加深度维度，创建三维的“多维数据集”。实际上，层与嵌套或堆积颇为相似；主要区别在于一次只能看到一层类别。例如，使用 *Age category* 作为行变量，*Gender* 作为层变量生成一个表，在该表中男性和女性的信息将显示在表的不同层中。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击**重置**删除表构建器中所有以前的选择。
3. 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
4. 单击表构建器顶部的**层**以显示“层”列表。
5. 将 *Gender* 从变量列表拖放到“层”列表中。

在这里，您可能会注意到添加层变量对画布窗格上所显示的预览不会产生影响。层变量不影响画布窗格上的预览，除非层变量是统计源变量并且您更改了汇总统计。

6. 单击**确定**以创建表。

Gender Male		计数
Age category	Less than 25	108
	25 to 34	276
	35 to 44	309
	45 to 54	221
	55 to 64	136
	65 or older	178

图 18. 简单的分层表

乍一看，该表与单个分类变量的简单表并没有任何差别。唯一的差别在于表顶部显示了标签 *Gender Male*。

7. 在“查看器”窗口中双击表将其激活。
8. 现在，您可以看到标签 *Gender Male* 实际上是下拉列表中的一个选项。
9. 单击下拉列表上的向下箭头显示整个层列表。

在该表中，列表中只有一个其他选项。

10. 从下拉列表中选择 *Gender Female*。

Gender Female		计数
Age category	Less than 25	134
	25 to 34	351
	35 to 44	370
	45 to 54	260
	55 to 64	184
	65 or older	301

图 19. 显示不同层的简单分层表

## 两个堆积分类层变量

如果层中有多个分类变量，则可以堆积或嵌套层变量。缺省情况下，对层变量进行堆积。（注意：如果存在任何刻度层变量，那么只能堆积层变量。）

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 如果行中还没有 *Age category* 并且层中也没有 *Gender*，请按照上述操作创建分层表。
3. 将 *Highest degree* 从变量列表拖放到“层”列表中 *Gender* 的下方。

现在，“层”列表下方“层输出”组中的两个单选按钮被激活。缺省选择是**将每个类别显示为一层**。这等同于堆积。

4. 单击**确定**以创建表。
5. 在“查看器”窗口中双击表将其激活。
6. 单击下拉列表上的向下箭头显示整个层列表。

表中共有七层：两层对应两个 *Gender* 类别，另五层对应五个 *Highest degree* 类别。对于堆积层，层的总数是层变量的类别（包括为层变量请求的所有总计和小计类别）的数量之和。

## 两个嵌套的分类层变量

嵌套分类层变量将为层变量类别的每个组合都创建一个单独的层。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。



2. 如果您尚未创建堆积层表，请按照上述操作进行创建。
3. 在“层输出”组中，选择将每个类别组合显示为一层。这等同于嵌套。
4. 单击**确定**以创建表。
5. 在“查看器”窗口中双击表将其激活。
6. 单击下拉列表上的向下箭头显示整个层列表。

该表中有 10 个层（您需要滚动列表才能查看所有层），*Gender* 和 *Highest degree* 的每个组合使用一个层。对于嵌套的层，层的总数为各个层变量的类别数之积（本例中为  $5 \times 2 = 10$ ）。

## 打印分层表

缺省情况下，只打印当前可见的层。若要打印表的所有层，请执行以下操作：

1. 在“查看器”窗口中双击表将其激活。
2. 从“查看器”窗口菜单选择：

格式 > 表属性...

3. 单击**打印**选项卡。
4. 选择**显示所有层**。

还可以将该设置保存为表格外观的一部分，包括缺省的表格外观。



---

## 第 4 章 分类变量的总计和小计

在定制表中可以同时包含总计和小计。总计和小计可以应用到任何维度（行、列或层）中任何嵌套级别的分类变量。

样本数据文件

本章 使用数据文件 *survey\_sample.sav*。请参阅数据文件主题以获取更多信息。

此处提供的所有示例都在对话框中按字母顺序显示变量标签。变量列表显示属性是在“选项”对话框（“编辑”菜单，“选项”）的“常规”选项卡中设置的。

---

### 单个变量的简单总计

1. 从菜单中选择:

分析 > 表 > 定制表...

2. 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
3. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**汇总统计**。
4. 在“汇总统计”对话框中，在“统计”列表中选择**列 N %**，然后单击箭头将其添加到“显示”列表中。
5. 在“显示”列表的“标签”单元格中，删除缺省标签并键入百分比。
6. 单击**应用到选定数据**。
7. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**类别和总计**。
8. 在“分类和总计”对话框中选择（单击）**总计**。
9. 单击**应用**，然后在表构建器中单击**确定**创建表。

		计数	百分比
Age category	Less than 25	242	8.6%
	25 to 34	627	22.2%
	35 to 44	679	24.0%
	45 to 54	481	17.0%
	55 to 64	320	11.3%
	65 or older	479	16.9%
	合计	2828	100.0%

图 20. 单个分类变量的简单总计

---

### 所见类别即总计类别

总计基于表中所显示的类别。如果选择从表中排除一些类别，则这些类别中的个案也不会包含在总计中。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**类别和总计**。
3. 单击“标签”列表中标注为 *Less than 25* 的类别。
4. 单击“排除”列表左侧的箭头键。

- 单击“标签”列表中标注为 *65 or older* 的类别。
- 再次单击“排除”列表左侧的箭头键。

这两个类别即从“显示”列表移到“排除”列表中。

- 单击**应用**，然后在表构建器中单击**确定**创建表。

		计数	百分比
Age category	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%
	合计	2107	100.0%

图 21. 具有已排除类别的表中的总计

与包含所有类别时的总计数 2,828 相比，此表中的总计数仅为 2,107。该总计中只包含了表中所使用的类别。（百分比总计仍为 100%，因为所有百分比都基于表中所使用的个案总数，而不是数据文件中的个案总数。）

## 总计的显示位置

缺省情况下，总计显示在对其计算总计的类别的下方。可以将总计的显示位置更改为显示在对其计算总计的类别的上方。

- 打开表构建器（“分析”菜单，“表格”，“定制表”）。
- 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**类别和总计**。
- 在“显示总计和小计”组中，选择**上述类别中它们适用的类别**。
- 单击**应用**，然后在表构建器中单击**确定**创建表。

		计数	百分比
Age category	合计	2107	100.0%
	25 to 34	627	29.8%
	35 to 44	679	32.2%
	45 to 54	481	22.8%
	55 to 64	320	15.2%

图 22. 总计显示在计算总计的类别上方

## 嵌套表的总计

由于可以将总计应用于任何嵌套级别的分类变量，因此可以创建包含位于多个嵌套级别的组总计的表。

### 组总计

嵌套在其他分类变量中的分类变量的总计表示组总计。

- 打开表构建器（“分析”菜单，“表格”，“定制表”）。
- 在画布窗格上将 *Gender* 拖放到 *Age category* 的左侧。
- 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**类别和总计**。

在创建表之前，将总计移回总计类别的下方。

4. 在“显示总计和小计”组中，选择以下类别中它们适用的类别。
5. 单击应用保存设置并返回到表构建器。
6. 单击确定以创建表。

				计数	百分比
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			合计	942	100.0%
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
			合计	1165	100.0%

图 23. Gender 类别中的 Age category 总计

此时该表将显示两个组总计：一组对应男性，另一组对应女性。

总计

应用于嵌套变量的总计始终为组总计，而不是总计。如果需要整个表的总计，可以将总计应用于位于最外面嵌套级别的变量。

1. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上的 Gender，然后从弹出菜单中选择类别和总计。
3. 在“分类和总计”对话框中选择（单击）总计。
4. 单击应用，然后在表构建器中单击确定创建表。

				计数	百分比
Gender	Male	Age category	25 to 34	276	29.3%
			35 to 44	309	32.8%
			45 to 54	221	23.5%
			55 to 64	136	14.4%
			合计	942	100.0%
	Female	Age category	25 to 34	351	30.1%
			35 to 44	370	31.8%
			45 to 54	260	22.3%
			55 to 64	184	15.8%
			合计	1165	100.0%
	合计	Age category	25 to 34	627	29.8%
			35 to 44	679	32.2%
			45 to 54	481	22.8%
			55 to 64	320	15.2%
			合计	2107	100.0%

图 24. 嵌套表的总计

您会看到总计仅为 2,107，而不是 2,828。由于表中仍排除了两个年龄类别，因此这两个类别中的个案也将所有总计中排除。

---

## 层变量总计

层变量的总计在表中显示为单独层。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 在表构建器中单击层以显示“层”列表。
3. 将 *Gender* 从画布窗格的“行”区域拖放“层”列表中。

注：由于已为 *Gender* 指定了总计，因此此时无需执行该操作。在维度之间移动变量不会影响该变量的任何设置。

4. 单击**确定**以创建表。
5. 在查看器中双击表将其激活。
6. 单击“层”下拉列表中的向下箭头以在表中显示所有层列表。

表中共有三层：*Gender Male*、*Gender Female* 和 *Gender Total*。

层总计的显示位置

对于层变量总计，总计的显示位置（上或下）确定总计的层位置。例如，如果为层变量总计指定了上述类别中它们适用的类别，则总计层是第一个显示的层。

---

## 小计

可以为变量的类别子集包含小计。例如，可以为以下年龄类别包含小计：表示抽样调查中 45 岁以下和以上的所有响应者。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击**重置**清除表构建器中所有以前的设置。
3. 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
4. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**类别和总计**。
5. 在“值”列表中选择 **3.00**。
6. 单击**添加小计**显示“定义小计”对话框。
7. 在“标签”文本字段中，键入小计 < 45。
8. 然后单击**继续**。

此操作将会插入一行，其中包含前三个年龄组的小计。

9. 在“值”列表中选择 **6.00**。
10. 单击**添加小计**显示“定义小计”对话框。
11. 在“标签”文本字段中，键入小计 45+。
12. 然后单击**继续**。

**重要提示：**必须先选择总计和小计的显示位置（在应用的类别上方或在应用的类别下方），然后再定义小计。更改显示位置会影响所有小计（而不仅仅是当前选择的小计），并且还会更改小计中包含的类别。

13. 单击**应用**，然后在表构建器中单击**确定**创建表。

		计数
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Subtotal <45	1548
	45 to 54	481
	55 to 64	320
	65 or older	479
	Subtotal 45+	1280

图 25. Age category 的小计

## 所见类别即小计类别

类似总计，小计也基于表中包含的类别。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择类别和总计。

注：对第一个小计显示的值（不是标签）为 **1.00...3.00**，这表明小计包含列表中 1 和 3 之间的所有值。

3. 在“值”列表中选择 **1.00**（或单击标签 *Less than 25*）。
4. 单击“排除”列表左侧的箭头键。

此时已排除第一个年龄类别，因此对第一个小计所显示的值将更改为 **2.00...3.00**，这表明将不包含在小计中的已排除类别基于表中所包含的类别。如果排除某一类别，也会自动从任何小计中排除该类别，因此不能执行实现类似于以下的功能：只显示小计而不显示包含小计所基于的类别。

## 隐藏小计类别

可以不显示定义小计的类别并只显示小计，从而有效地“拼并”类别而不会影响基本数据。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击重置清除表构建器中所有以前的设置。
3. 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
4. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择类别和总计。
5. 在“值”列表中选择 **3.00**。
6. 单击添加小计显示“定义小计”对话框。
7. 在“标签”文本字段中，键入小于 45。
8. 选择（复选）隐藏表格中的小计类别。
9. 然后单击继续。

此操作将会插入一行，其中包含前三个年龄组的小计。

10. 在“值”列表中选择 **6.00**。
11. 单击添加小计显示“定义小计”对话框。
12. 在“标签”文本字段中，键入 45 岁或以上。
13. 选择（复选）隐藏小计类别。
14. 然后单击继续。
15. 要包含小计的总计，请选择（选中）“显示”组中的总计。

16. 单击**应用**。

画布将显示小计但将排除定义小计类别。

17. 单击**确定**以生成表。

		计数
Age category	Less than 45	1548
	45 or older	1280
	合计	2828

图 26. 只显示小计和总计的表

## 层变量小计

类似于总计，层变量的小计在表中也显示为单独层。这些小计通常被视为类别。每个类别都是表中的单独层，并且层类别的显示顺序由“分类和总计”对话框（包括小计类别的显示位置）中的类别顺序确定。



---

## 第 5 章 分类变量的已计算类别

在定制表中可以包含已计算类别。它们是新的类别，并可由任何维度（行、列或层）中任何嵌套级别的相同变量中的类别计算得出。例如，您可以包括一个显示两个类别之间差别的已计算类别。

样本数据文件

本章 使用数据文件 *survey\_sample.sav*。请参阅数据文件主题以获取更多信息。

---

### 简单已计算类别

1. 从菜单中选择:

分析 > 表 > 定制表...

2. 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
3. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择**类别和总计**。
4. 在“值”列表中选择 **3.00**。
5. 单击**添加类别**显示“定义已计算类别”对话框。
6. 在“已计算类别的标签”文本字段中，输入小于 45。
7. 选择“类别”列表中的**小于 25 (1.00)**并单击箭头按钮以将其复制到“已计算类别的表达式”文本框中。[1] 将显示在表达式中。
8. 单击该对话框中的加号 (+) 运算符按钮（或按键盘上的 + 键）。
9. 选择“类别”列表中的**25 到 34 (2.00)**并单击箭头按钮以将其复制到“已计算类别的表达式”文本框中。
10. 单击该对话框中的加号 (+) 运算符按钮（或按键盘上的 + 键）。
11. 在“类别”列表中选择 **35 至 44 (3.00)**，然后单击方向按钮将其复制到“已计算类别的表达式”文本框中。
12. 然后单击**继续**。

此操作将会插入一行，其中包含前三个年龄组的小计。

13. 在“值”列表中选择 **5.00**。
14. 单击**添加小计**显示“定义小计”对话框。
15. 在“标签文本”字段中输入小于 65。
16. 然后单击**继续**。

此操作将会插入一行，其中包含前五个类别的小计。

17. 单击**应用**，然后在表构建器中单击**确定**创建表。

		计数
Age category	Less than 25	242
	25 to 34	627
	35 to 44	679
	Less than 45	1548
	45 to 54	481
	55 to 64	320
	Less than 65	2349
	65 or older	479

图 27. 带小计的已计算类别

该表包含一个已计算类别（小于 45）和一个小计（小于 65）。小计包含的类别也包括在已计算类别中。您不能单独使用小计来创建相同表，因为小计不能共用相同的类别。

## 在已计算类别中隐藏类别

与小计相同，您可以不显示在已计算类别的表达式中使用的类别，而只显示已计算类别本身。以下示例建立在先前示例的基础之上。

1. 从菜单中选择:

分析 > 表 > 定制表...

2. 右键单击画布窗格上的 *Age category*，然后从弹出菜单中选择类别和总计。
3. 在“值”列表中选择小于 45 已计算类别。
4. 单击编辑显示“定义已计算类别”对话框。
5. 选择隐藏表格中表达式使用的类别。
6. 然后单击继续。
7. 在“值”列表中选择小于 65 小计。
8. 单击编辑显示“定义小计”对话框。
9. 选择隐藏表格中的小计类别。
10. 然后单击继续。
11. 单击应用，然后在表构建器中单击确定创建表。

		计数
Age category	Less than 45	1548
	Less than 65	2349
	65 or older	479

图 28. 带小计和隐藏类别的已计算类别

与前一示例相同，该表包含一个已计算类别和一个小计。但在本例中已隐藏了各自的类别，因此只显示这些总计。

---

## 在已计算类别中引用小计

可以在已计算类别的表达式中包括小计。

1. 从菜单中选择:

分析 > 表 > 定制表...

2. 单击**重置**清除表构建器中所有以前的设置。
3. 在表构建器中，将 *Labor force status* 从变量列表拖放到画布窗格的“行”区域中。
4. 将 *Marital status* 从变量列表拖放到“列”区域中。
5. 右键单击画布窗格上的 *Labor force status*，然后从弹出菜单中选择**类别和总计**。
6. 在“值”列表中选择 **2**。
7. 单击**添加小计**显示“定义小计”对话框。
8. 在“标签”文本字段中，键入工作。
9. 选择**隐藏表格中的小计类别**。
10. 然后单击**继续**。

此操作将会插入一行，其中包含前两个工作状态类别的小计。

11. 在“值”列表中选择 **8**。
12. 单击**添加小计**显示“定义小计”对话框。
13. 在“标签”文本字段中，键入未工作。
14. 选择**隐藏小计类别**。
15. 然后单击**继续**。

此操作将会插入一行，其中包含其他工作状态类别的小计。

16. 在“值”列表中选择未工作小计。
17. 单击**添加类别**显示“定义已计算类别”对话框。
18. 在“已计算类别的标签”文本字段中，输入工作/未工作。
19. 在“总计和小计”列表中，选择**工作（工作 #1）**并单击箭头按钮以将其复制到“已计算类别的表达式”文本框中。
20. 单击该对话框中的除号 (*/*) 运算符按钮（或按键盘上的 */* 键）。
21. 在“总计和小计”列表中，选择**未工作（未工作 #2）**并单击箭头按钮以将其复制到“已计算类别的表达式”文本框中。

缺省情况下，已计算类别使用与变量的统计（在本例中为计数）相同的格式。由于我们需要显示在已计算类别的表达式中相除后产生的小数位，而计数的缺省格式不含小数位，因此必须更改格式。

22. 单击“显示格式”选项卡。
23. 将计数的小数设置更改为 **2**。
24. 然后单击**继续**。
25. 单击**应用**，然后在表构建器中单击**确定**创建表。

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		计数	计数	计数	计数	计数
Labor force status	Working	916	64	330	67	494
	Not Working	429	219	116	26	169
	Not Working / Working	2.14	.29	2.84	2.58	2.92

图 29. 显示小计比率的已计算类别

该表包含两个小计和一个已计算类别。已计算类别显示小计的比率，以方便您比较每个小计所代表的组。与其他组相比，工作与未工作丧偶响应者的比率相当低。此外，已婚响应者的比率略微较低，这可能是由于配偶不再工作而在家带小孩。

## 使用已计算类别显示非穷举小计

小计是穷举型的。即，表中的所有小计包含了在其表中位置上方或下方的所有值。另一方面，已计算类别不是穷举型的，它允许您对表中的类别组合进行合计。

1. 从菜单中选择:

分析 > 表 > 定制表...

2. 单击**重置**清除表构建器中所有以前的设置。
3. 在表构建器中，将**考虑自身为自由或保守**从变量列表拖放到画布窗格的“行”区域中。
4. 右键单击画布窗格上的**考虑自身为自由或保守**，然后从弹出的上下文菜单中选择**分类和总计**。
5. 在“值”列表中选择 **3**。
6. 单击**添加类别**显示“定义已计算类别”对话框。
7. 在“已计算类别的标签”文本字段中输入自由小计。请注意，在文本前面有四个空格。这些空格用来表示结果表中的缩进。
8. 选择“类别”列表中的**极为自由 (1)**并单击箭头按钮以将其复制到“已计算类别的表达式”文本框中。
9. 单击该对话框中的加号 (+) 运算符按钮（或按键盘上的 + 键）。
10. 选择“类别”列表中的**自由 (2)**并单击箭头按钮以将其复制到“已计算类别的表达式”文本框中。
11. 单击该对话框中的加号 (+) 运算符按钮（或按键盘上的 + 键）。
12. 选择“类别”列表中的**较自由 (3)**并单击箭头按钮以将其复制到“已计算类别的表达式”文本框中。
13. 单击**继续**。

此操作将会插入一行，其中包含自由类别的小计。

14. 在“值”列表中选择 **7**。
15. 单击**添加类别**显示“定义已计算类别”对话框。
16. 在“已计算类别的标签”文本字段中输入保守小计。请注意，在文本前面有四个空格。这些空格用来表示结果表中的缩进。
17. 选择“类别”列表中的**较保守 (5)**并单击箭头按钮以将其复制到“已计算类别的表达式”文本框中。
18. 单击该对话框中的加号 (+) 运算符按钮（或按键盘上的 + 键）。
19. 选择“类别”列表中的**保守 (6)**并单击箭头按钮以将其复制到“已计算类别的表达式”文本框中。
20. 单击该对话框中的加号 (+) 运算符按钮（或按键盘上的 + 键）。
21. 选择“类别”列表中的**极为保守 (7)**并单击箭头按钮以将其复制到“已计算类别的表达式”文本框中。
22. 单击**继续**。

此操作将会插入一行，其中包含保守类别的小计。

23. 单击**应用**，然后在表构建器中单击**确定**创建表。

		计数
Think of self as liberal or conservative	Extremely liberal	64
	Liberal	357
	Slightly liberal	351
	Liberal Subtotal	772
	Moderate	986
	Slightly conservative	432
	Conservative	415
	Extremely conservative	86
	Conservative Subtotal	933

图 30. 显示非穷举小计的已计算类别

该表包含两个已计算类别，在已计算类别中未包括表中显示的全部类别。适度类别未包含在任一已计算类别中。您不能使用小计创建相同表，因为小计是穷举型的。



---

## 第 6 章 具有共享类别的变量的表

调查常常包含许多问题，每个问题又都包含一组相同的、响应者可能作出的回答。例如，我们的抽样调查包含一些变量，这些变量涉及响应者对各种公共及私人机构和服务的信心，并且具有一组相同的响应类别：1 = *A great deal*, 2 = *Only some* 以及 3 = *Hardly any*。可以使用堆积在同一个表中显示这些相关的变量，并在表的列中显示共用的响应类别。在您使用已计算类别时，这些功能同样可用，但条件是任何已计算类别的标签和表达式在所有变量中均相同。

	大量	少许	极少
银行与金融机构的置信度	490	1068	306
教育的置信度	511	1055	315
大公司的置信度	500	1078	243
医疗的置信度	844	864	167
媒体的置信度	176	878	808
电视的置信度	196	936	744

图 31. 具有共享类别的变量表

注：在以前版本的定制表中，该表称为“频率表”。

样本数据文件

本章 使用数据文件 *survey\_sample.sav*。请参阅数据文件主题以获取更多信息。

此处提供的所有示例都在对话框中按字母顺序显示变量标签。变量列表显示属性是在“选项”对话框（“编辑”菜单，“选项”）的“常规”选项卡中设置的。

---

### 计数表

1. 从菜单中选择：

分析 > 表 > 定制表...

2. 在表构建器中的变量列表中，单击 *Confidence in banks...*，然后按住 Shift 并单击 *Confidence in television* 以选择所有“confidence”变量。（注意：假设变量标签在变量列表中按字母顺序而非文件顺序显示。）
3. 将这六个 confidence 变量拖放到画布窗格的“行”区域中。

这会使变量堆积在行维度中。缺省情况下，每个变量的类别标签也显示在行中，这导致表很长且很窄（6 个变量 x 3 个类别 = 18 行）。但是由于所有六个变量共享相同的已定义类别标签（值标签），因此可以将类别标签放置在列维度中。

4. 从“类别位置”下拉列表中选择列中的行标签。

现在表只有六行，每行表示堆积变量中的一个，而定义的类别变成了表中的列。

5. 由于汇总统计标签计数不是必需的，因此在创建表之前，为“汇总统计”组中的“位置”选择（单击）隐藏。
6. 单击确定以创建表。

	大量	少许	极少
银行与金融机构的置信度	490	1068	306
教育的置信度	511	1055	315
大公司的置信度	500	1078	243
医疗的置信度	844	864	167
媒体的置信度	176	878	808
电视的置信度	196	936	744

图 32. 列中含有共享类别标签的堆积行变量表

可以创建在列中堆积变量并在行中显示类别的表，而不是在行中显示变量并在列中显示类别。如果类别多于变量，这会是一个更好的选择，而在我们的示例中变量多于类别。

## 百分比表

对于在行中堆积变量并在列中显示类别的表，要显示的最有意义的（或者至少是最易于理解的）百分比是行百分比。（对于在列中堆积变量并在行中显示类别的表，所需要的可能是列百分比。）

1. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上表预览中的任一置信度变量，然后从弹出菜单中选择**汇总统计**。
3. 在“统计”列表中选择**行 N %**，然后单击箭头按钮将其移动到“显示”列表中。
4. 在“显示”列表中，单击计数行中的任意一个单元格，然后单击箭头按钮将其移动回“统计”列表，以将其从“显示”列表中删除。
5. 单击**应用到全部**以将汇总统计更改应用到表中所有的堆积变量。

注：如果您的表预览与该图所示的不同，则您可能单击了**应用到选定数据**而不是**应用到全部**，选取前者只会将新的汇总统计应用到所选择的变量。在本示例中，该操作会为每个类别生成两列：一列显示所有其他变量的计数占位符，另一列显示选定变量的行百分比占位符。这恰好是要生成的表，却不是我们在本示例中所需要的表。

6. 单击**确定**以创建表。

	A great deal	Only some	Hardly any
Confidence in banks & financial institutions	26.3%	57.3%	16.4%
Confidence in education	27.2%	56.1%	16.7%
Confidence in major companies	27.5%	59.2%	13.3%
Confidence in medicine	45.0%	46.1%	8.9%
Confidence in press	9.5%	47.2%	43.4%
Confidence in television	10.4%	49.9%	39.7%

图 33. 在行中堆积变量的行百分比、在列中显示类别的表

注：可以在具有共享类别的变量表中包含任意数目的汇总统计。为简单起见，在我们的示例中每次只显示一个。



## 总计和类别控制

只有当表中的所有变量都具有相同的类别并以相同的顺序显示时，才可以创建与变量维度相对应的类别的表。这包括总计、小计和所做的所有其他类别调整。这意味着在“分类和总计”对话框中所做的所有修改都必须应用于表中共享类别的所有变量。

1. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上表预览中的第一个置信度变量，然后从弹出菜单中选择**分类和总计**。
3. 在“分类和总计”对话框中选择（复选）**总计**，然后单击**应用**。

您注意到的第一件事情可能是类别标签已经从列移动回行。您还会注意到现在禁用了“类别位置”控件。这是因为变量不再共享完全相同的一组“类别”。其中一个变量现在有了一个总计类别。

4. 右键单击画布窗格上的任意一个置信度变量，然后从弹出菜单中选择**选择所有行变量**；或者在画布窗格上按住 **Ctrl** 单击每个堆积变量，直到选中所有变量（您可能需要向下滚动窗格或展开表构建器窗口）。
5. 在“定义”组中，单击**分类和总计**。
6. 在“分类和总计”对话框中，如果尚未选择（复选）**总计**，则现在选择该选项，然后单击**应用**。
7. 现在所有变量都具有该附加的总计类别，因此“类别位置”下拉列表会再次启用，请选择列中的行标签。
8. 单击**确定**以创建表。

	A great	Only some	Hardly	合计
Confidence in banks & financial institutions	26.3%	57.3%	16.4%	100.0%
Confidence in education	27.2%	56.1%	16.7%	100.0%
Confidence in major companies	27.5%	59.2%	13.3%	100.0%
Confidence in medicine	45.0%	46.1%	8.9%	100.0%
Confidence in press	9.5%	47.2%	43.4%	100.0%
Confidence in television	10.4%	49.9%	39.7%	100.0%

图 34. 在行中堆积变量的行百分比、在列中显示类别和总计的表

## 在表中对共享类别进行嵌套

在嵌套表中，如果要相对应的维度显示类别标签，则具有共享类别的堆积变量必须位于维度最里面的嵌套层中。

1. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 将 *Gender* 从变量列表拖放到“行”区域的左侧。

具有共享类别的堆积变量，在表预览中现在嵌入“gender”类别中。

3. 现在，在表预览中将 *Gender* 拖放到其中一个堆积 *confidence* 变量的右侧。

类别标签再次回到行维度中，“类别位置”控件又一次被禁用。现在，有一个堆积变量又嵌套了 *Gender*，而其他堆积变量不包含嵌套变量。可以将 *Gender* 作为嵌套变量添加到每个堆积变量中，但在这之后将行标签移动到列会导致在列中显示 *Gender* 的类别标签，而不是具有共享类别的堆积变量的类别标签。这是因为 *Gender* 现在成为最内层的嵌套变量，并且更改类别位置总是应用到最内层的嵌套变量。



---

## 第 7 章 汇总统计

汇总统计包括从分类变量的简单计数到离差测量的所有统计，例如刻度变量平均值的标准误差。汇总统计不包括“定制表”对话框中“检验统计”选项卡上提供的显著性检验。请参阅第 59 页的『检验统计』主题以获取更多信息。

分类变量和多响应集的汇总统计包括计数和多种百分比计算，例如：

- 行百分比
- 列百分比
- 子表百分比
- 表百分比
- 有效 N 百分比

除了可用于分类变量的汇总统计之外，刻度变量的汇总统计和分类变量的定制总计摘要还包括：

- 平均值
- 中位数
- 百分位数
- 合计
- 标准差
- 范围(A)
- 最小值和最大值

另有一些汇总统计可用于多响应集。另外，还提供了汇总统计的完整列表。请参阅第 5 页的『汇总统计』主题以获取更多信息。

样本数据文件

本章 使用数据文件 *survey\_sample.sav*。请参阅数据文件主题以获取更多信息。

此处提供的所有示例都在对话框中按字母顺序显示变量标签。变量列表显示属性是在“选项”对话框（“编辑”菜单，“选项”）的“常规”选项卡中设置的。

---

### 汇总统计源变量

可用的汇总统计取决于汇总统计源变量的测量级别。汇总统计源（汇总统计所基于的变量）由以下因素确定：

- **测量级别。**如果表（或堆积表中的表部分）包含刻度变量，则汇总统计基于该刻度变量。
- **变量选择顺序。**分类变量的缺省统计源维度（行或列）基于将这些变量拖放到画布窗格上的顺序。例如，如果首先将变量拖动到行区域，则行维度是缺省的统计源维度。
- **嵌套。**对于分类变量，汇总统计基于统计源维度中最里面的变量。

一个堆积表可以有多个汇总统计源变量（刻度变量和分类变量均可），但每个表部分只能有一个汇总统计源。

## 分类变量的汇总统计源

1. 从菜单中选择:  
  
    分析 > 表 > 定制表...
2. 在表构建器中, 将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
3. 右键单击画布窗格上的 *Age category*, 然后从弹出菜单中选择**汇总统计**。(由于该变量是表中唯一的变量, 因此便是统计源变量。)
4. 在“汇总统计”对话框中, 在“统计”列表中选择 *列 N %*, 然后单击箭头将其添加到“显示”列表中。
5. 单击**应用到选定数据**。
6. 在表构建器中, 将 *Get news from internet* 拖放到画布窗格上 *Age category* 的右侧。
7. 再次右键单击画布窗格上的 *Age category*。此时弹出菜单中的**汇总统计**项被禁用, 因为 *Age category* 已不是统计源维度中最里面的嵌套变量。
8. 右键单击画布窗格上的 *Get news from internet*。现在**汇总统计**项被启用, 因为所选变量是统计源维度中最里面的嵌套变量, 因此也是汇总统计源变量。(由于该表只有一个维度(即行), 所以行就是统计源维度。)
9. 将 *Get news from internet* 从画布窗格上的“行”区域拖放到“列”区域中。
10. 再次右键单击画布窗格上的 *Get news from internet*。此时弹出菜单中的**汇总统计**项被禁用, 因为该变量不再位于统计源维度中。

因为分类变量的缺省统计源维度是创建表时放置变量的第一个维度, 所以 *Age category* 再次成为统计源变量。在本例中, 我们所做的第一件事情是将变量放置到行维度中。这样, 行维度便成为缺省统计源维度; 由于 *Age category* 现在是在该维度中的唯一变量, 因此成为统计源变量。

## 刻度变量的汇总统计源

1. 将刻度变量 *Hours per day watching TV* 拖放到画布窗格上“行”区域中 *Age category* 的左侧。  
  
    您首先会注意到计数和 *列 N %* 摘要已被平均值取代--如果右键单击画布窗格上的 *Hours per day watching TV*, 您将会看到汇总统计源变量。对于包含刻度变量的表, 无论嵌套层或维度如何, 刻度变量总是统计源变量, 并且刻度变量的缺省汇总统计是平均值。
2. 将 *Hours per day watching TV* 从“行”区域拖放到“列”区域中 *Get news from internet* 的上方。
3. 右键单击 *Hours per day watching TV*, 然后从弹出菜单选择**汇总统计**。(即使将该变量移动到不同维度中, 它仍然是统计源变量。)
4. 在“汇总统计”对话框中, 单击“显示”列表中平均值的**格式**单元格, 然后从“格式”下拉列表中选择 **nnnn**。(可能必须向上滚动列表才能看到此选项。)
5. 在“小数”单元格中输入 2。
6. 单击**应用到选定数据**。  
  
    画布窗格上的表预览现在表明将以两位小数的格式显示平均值。
7. 单击**确定**以创建表。

		Hours per day watching TV	
		Get news from internet	
		No	Yes
		均值	均值
Age category	Less than 25	3.54	2.12
	25 to 34	3.42	2.14
	35 to 44	3.00	2.01
	45 to 54	2.83	2.06
	55 to 64	3.24	2.37
	65 or older	3.82	2.33

图 35. 在交叉制表的分类变量中汇总的刻度变量

## 堆积变量

由于堆积表可以包含多个统计源变量，并且可以为每个统计源变量指定不同的汇总统计，因此对于在堆积表中指定汇总统计，有一些需要特别注意的事项。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击**重置**清除表构建器中所有以前的设置。
3. 在变量列表中单击 *Get news from internet*，然后按住 **Shift** 并在变量列表中单击 *Get news from television* 以选择所有“news”变量。（注意：假设变量标签在变量列表中按字母顺序而非文件顺序显示。）
4. 将这五个 news 变量拖放到画布窗格的“行”区域中。

五个 news 变量堆积在行维度中。

5. 在画布窗格上单击 *Get news from internet* 以便只选择该变量。
6. 此时右键单击 *Get news from internet*，然后从弹出菜单中选择**汇总统计**。
7. 在“汇总统计”对话框中，从“统计”列表中选择 *列 N %*，然后单击箭头将其添加到“显示”列表中。（可以使用箭头将所选择的统计从“统计”列表移动到“显示”列表中，或者可以将所选择的统计从“统计”列表拖放到“显示”列表中。）
8. 然后单击**应用到选定数据**。

添加一个用于显示列百分比的列--但是画布窗格上的表预览表明将只为一个变量显示列百分比。这是因为在堆积表中有多个统计源变量，并且每个统计源变量都可以包含不同的汇总统计。但是在本例中，我们要为所有变量显示相同的汇总统计。

9. 右键单击画布窗格上的 *Get news from newspapers*，让从弹出菜单中选择**汇总统计**。
10. 在“汇总统计”对话框中，从“统计”列表中选择 *列 N %*，然后单击箭头将其添加到“显示”列表中。
11. 然后单击**应用到全部**。

现在表预览表明将为所有堆积变量显示列百分比。

## 分类变量的定制总计汇总统计

对于分类统计源变量，可以包含不同于为变量类别显示的统计的定制总计汇总统计。例如，对于有序变量，可以为每个类别显示百分比，为定制总计汇总统计显示平均值或中位数。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击**重置**清除表构建器中所有以前的设置。

3. 在变量列表中单击 *Confidence in press*，然后按住 Ctrl 并单击 *Confidence in TV* 以选择这两个变量。
4. 将这两个变量拖放到画布窗格的“行”区域中。这会使这两个变量堆积在行维度中。
5. 右键单击画布窗格上任意一个变量，然后从弹出菜单中选择**选择所有行变量**。（它们可能已经全部被选中，但是我们需要确认一下。）
6. 再次右键单击该变量，然后从弹出菜单中选择**分类和总计**。
7. 在“分类和总计”对话框中单击（复选）**总计**，然后单击**应用**。

画布窗格上的表预览现在为两个变量显示一个总计行。为了显示定制总计汇总统计，必须为表指定总计和/或小计。

8. 右键单击画布窗格上任意一个变量，然后从弹出菜单中选择**汇总统计**。
9. 在“汇总统计”对话框中，在“显示”列表中单击**计数**，然后单击箭头将其移动到“统计”列表中，以将其从“显示”列表中移去。
10. 在“统计”列表中单击**列 N %**，然后单击箭头键将其移动到“显示”列表中。
11. 单击（复选）**定制关于总计和小计的汇总统计**。
12. 在定制摘要“显示”列表中单击**计数**，然后单击箭头将其移动到定制汇总统计列表，以将其从“显示”列表中移去。
13. 在定制摘要“统计”列表中单击**平均值**，然后单击箭头将其移动到定制摘要“显示”列表中。
14. 在“显示”列表中单击平均值的**格式**单元格，然后从格式的下拉列表中选择 **nnnn**。（可能必须向上滚动列表才能看到此选项。）
15. 在“小数”单元格中输入 2。
16. 单击**应用到全部**以将这些设置应用到表中的两个变量。

为定制总计汇总统计添加了一个新列，但这可能不是您需要的结果，因为画布窗格上的预览明确表明产生的表会包含许多空白单元格。

17. 在表构建器的“汇总统计”组中，从“位置”下拉列表中选择**行**。

这会将所有汇总统计都移动到行维度，并在表的一个列中显示所有汇总统计。

18. 单击**确定**以创建表。

Confidence in press	A great deal	列 N %	9.5%
	Only some	列 N %	47.2%
	Hardly any	列 N %	43.4%
	合计	均值	2.34
Confidence in television	A great deal	列 N %	10.4%
	Only some	列 N %	49.9%
	Hardly any	列 N %	39.7%
	合计	均值	2.29

图 36. 具有定制总计汇总统计的分类变量

## 显示类别值

上一个表中只有一个小问题：如果不知道平均值基于的基础类别值，可能很难解释平均值。值为 2.34 的平均值是处于 *A great deal* 和 *Only some* 之间，还是处于 *Only some* 和 *Hardly any* 之间？

尽管我们无法直接在“定制表”中解决该问题，但是我们可以用更常规的方法解决它。

1. 从菜单中选择:

编辑 > 选项...

2. 在“选项”对话框中，单击输出标签选项卡。
3. 在“透视表标签”组中，从标签中的变量值显示为下拉列表中选择值和标签。
4. 单击确定保存此设置。
5. 打开表构建器（“分析”菜单，“表格”，“定制表”），然后单击确定再次创建表。

Confidence in press	1 A great deal	列 N %	9.5%
	2 Only some	列 N %	47.2%
	3 Hardly any	列 N %	43.4%
	合计	均值	2.34
Confidence in television	1 A great deal	列 N %	10.4%
	2 Only some	列 N %	49.9%
	3 Hardly any	列 N %	39.7%
	合计	均值	2.29

图 37. 为变量类别显示的值和标签

类别值清楚地指出值为 2.34 的平均值位于 *Only some* 和 *Hardly any* 之间。在表中显示类别值便于解释定制总计汇总统计（例如平均值）的值。

该显示设置是全局设置，会影响从所有过程得到的所有透视表输出，并在会话之间保持不变，直到您对其进行更改。若要将该设置重新更改为只显示值标签，请执行以下操作:

6. 从菜单中选择:

编辑 > 选项...

7. 在“选项”对话框中，单击输出标签选项卡。
8. 在“透视表标签”组中，从标签中的变量值显示为下拉列表中选择标签。
9. 单击确定保存此设置。





---

## 第 8 章 刻度变量摘要

---

### 刻度变量摘要

有多种可用于刻度变量的汇总统计。除了可用于分类变量的计数和百分比之外，用于刻度变量的汇总统计还包括：

- 平均值(M)
- 中位数
- 百分位数
- 合计
- 标准差
- 范围(R)
- 最小值和最大值

请参阅第 7 页的『刻度变量和分类定制总计的汇总统计』主题以获取更多信息。

样本数据文件

本章使用数据文件 *survey\_sample.sav*。请参阅数据文件主题以获取更多信息。

此处提供的所有示例都在对话框中按字母顺序显示变量标签。变量列表显示属性是在“选项”对话框（“编辑”菜单，“选项”）的“常规”选项卡中指定的。

---

### 堆积刻度变量

通过在表中堆积多个刻度变量可以在同一个表中对这些变量进行摘要。

1. 从菜单中选择：

分析 > 表 > 定制表...

2. 在表构建器中，单击变量列表中的 *Age of respondent*，按住 Ctrl 并单击 *Highest year of school completed*，然后按住 Ctrl 并单击 *Hours per day watching TV* 选择所有三个变量。
3. 将已选择的三个变量拖放到画布窗格的“行”区域。

这三个变量在行维度中堆积起来。由于这三个变量都是刻度变量，因此不会显示类别，并且缺省汇总统计是平均值。

4. 单击**确定**以创建表。

	均值
Age of respondent	46
Highest year of school completed	13
Hours per day watching TV	3

图 38. 堆积刻度变量的平均值表

## 多个汇总统计

缺省情况下显示刻度变量的平均值；但是可以为刻度变量选择其他汇总统计，还可以显示多个汇总统计。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上表预览中的任意一个刻度变量（共三个），然后从弹出菜单中选择**汇总统计**。
3. 在“汇总统计”对话框中，在“统计”列表中选择**中位数**，然后单击箭头将其添加到“显示”列表中。（可以使用箭头将所选择的统计从“统计”列表移动到“显示”列表，或者可以将所选择的统计从“统计”列表拖放到“显示”列表。）
4. 在“显示”列表中单击中位数的**格式**单元格，然后从格式的下拉列表中选择 **nnnn**。
5. 在“小数”单元格中，键入 1。
6. 对“显示”列表中的平均值做相同的更改。
7. 单击**应用到全部**对所有三个刻度变量应用这些更改。
8. 在表构建器中单击**确定**创建表。

	均值	中值
Age of respondent	45.6	42.0
Highest year of school completed	13.3	13.0
Hours per day watching TV	2.9	2.0

图 39. 堆积刻度变量表中显示的平均值和中间值

## 计数、有效 N 和缺失值

显示用于计算汇总统计（例如平均值）的个案数量通常是有用处的，您可以假定（没什么不合理）汇总统计计数会提供该信息。但是，如果存在任何缺失值，则这将不会为您提供精确的个案基数。若要获取精确的个案基数，请使用有效 *N*。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上表预览中的任意一个刻度变量（共三个），然后从弹出菜单中选择**汇总统计**。
3. 在“汇总统计”对话框中，在“统计”列表中选择**计数**，然后单击箭头将其添加到“显示”列表中。
4. 随后，在“统计”列表中选择**有效 N**，然后单击箭头将其添加到“显示”列表中。
5. 单击**应用到全部**对所有三个刻度变量应用这些更改。
6. 在表构建器中单击**确定**创建表。

	均值	中值	计数	有效的 N
Age of respondent	46	42	2832	2828
Highest year of school completed	13	13	2832	2820
Hours per day watching TV	3	2	2832	2337

图 40. 计数与有效 N

对于所有三个变量，计数均为：2,832。这并非巧合，该数值就是数据文件中个案的总数。由于刻度变量未嵌套在任何分类变量中，因此计数只代表数据文件中个案的总数。

另一方面，每个变量的有效 N 都是不同的，并且与 *Hours per day watching TV* 的计数差异很大。这是因为该变量存在大量的缺失值 -- 即，有些个案没有为该变量记录任何值或值被定义为代表缺失数据（例如，代码 99 代表怀孕不适用于男性）。

7. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
8. 右键单击画布窗格上表预览中的任意一个刻度变量（共三个），然后从弹出菜单中选择**汇总统计**。
9. 在“汇总统计”对话框中，在“显示”列表中选择**有效 N**，然后单击箭头键将其移动回“统计”列表，以将其从“显示”列表中删除。
10. 在“显示”列表中选择**计数**，然后单击箭头键将其移动回“统计”列表，以将其从“显示”列表中删除。
11. 在“统计”列表中选择**缺失**，然后单击箭头键将其添加到“显示”列表中。
12. 单击**应用到全部**对所有三个刻度变量应用这些更改。
13. 在表构建器中单击**确定**创建表。

	均值	中值	缺失
Age of respondent	45.6	42.0	4
Highest year of school completed	13.3	13.0	12
Hours per day watching TV	2.9	2.0	495

图 41. 刻度汇总统计表中显示的缺失值数量

现在，该表显示每个刻度变量的缺失值个数。从该表可以很明显地看出，*Hours per day watching TV* 包含大量的缺失值，而其他两个变量的缺失值很少。这可能是在充分相信变量的摘要值之前应考虑的一个因素。

## 不同变量的摘要

除了显示多个汇总统计之外，还可以在堆积表中为不同刻度变量显示不同的汇总统计。例如，先前的表揭示了三个变量中只有一个具有大量的缺失值；因此您可能希望只为该变量显示缺失值个数。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 在画布窗格上的表预览中单击 *Age of respondent*，然后按住 Ctrl 并单击 *Highest year of school completed* 以选择这两个变量。
3. 右键单击任意一个选定的变量（共两个），然后从弹出菜单中选择**汇总统计**。
4. 在“汇总统计”对话框中，在“显示”列表中选择**缺失**，然后单击箭头键将其移动回“统计”列表，以将其从“显示”列表中删除。
5. 单击**应用到选定数据**以将更改只应用到所选择的两个变量。

表的数据单元格中的占位符指示将只为 *Hours per day watching TV* 显示缺失值个数。

6. 单击**确定**以创建表。

	均值	中值	缺失
Age of respondent	45.6	42.0	
Highest year of school completed	13.3	13.0	
Hours per day watching TV	2.9	2.0	495

图 42. 不同变量的汇总统计表

尽管这个表提供了我们需要的信息，但是其布局使我们很难解释该表。查阅该表的人可能会认为缺失列中的空白单元格表示这些变量的缺失值个数为零。

7. 打开表构建器（“分析”菜单，“表格”，“定制表”）。

8. 在表构建器的“汇总统计”组中，从“位置”下拉列表中选择行。

9. 单击**确定**以创建表。

Age of respondent	均值	45.6
	中值	42.0
Highest year of school completed	均值	13.3
	中值	13.0
Hours per day watching TV	均值	2.9
	中值	2.0
	缺失	495

图 43. 行维度中同时显示的汇总统计和变量

现在很清楚的一点是，该表只报告一个变量的缺失值个数。

## 类别中的分组摘要

可以使用分类变量作为分组变量，以在由该分类变量的类别所定义的组内显示刻度变量摘要。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。

2. 将 *Gender* 从变量列表拖放到画布窗格的“列”区域中。

如果右键单击画布窗格上表预览中的 *Gender*，那么会看到弹出菜单中的**汇总统计**已被禁用。这是因为在包含刻度变量的表中，刻度变量总是统计源变量。

3. 单击**确定**以创建表。

		Gender	
		Male	Female
Age of respondent	均值	44.6	46.3
	中值	42.0	43.0
Highest year of school completed	均值	13.4	13.2
	中值	13.0	13.0
Hours per day watching TV	均值	2.8	2.9
	中值	2.0	2.0
	缺失	213	282

图 44. 使用分类列变量分组的刻度摘要

通过该表可以很容易地比较男性和女性的平均数（平均值和中位数），并且清晰地显示出两者之间没有很大的差异。这一信息也许不至关重要，但可能很有用处。

## 多个分组变量

通过嵌套和/或使用行和列分类分组变量，可以对组进行进一步细分。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 将 *Get news from internet* 从变量列表拖放至画布窗格的“行”区域的最左侧。对该变量进行定位时，务必使所有三个刻度变量都嵌套在其中，而不是仅嵌套其中一个。

您有时可能需要类似以上第二个示例的形式，但不符合本例中的需要。

3. 单击**确定**以创建表。

				Gender	
				Male	Female
Get news from internet	No	Age of respondent	均值	47.0	48.8
			中值	45.0	46.0
		Highest year of school completed	均值	13.4	13.1
			中值	13.0	12.0
		Hours per day watching TV	均值	3.2	3.4
			中值	2.0	3.0
	Yes	Age of respondent	均值	38.7	41.1
			中值	35.0	38.0
		Highest year of school completed	均值	13.2	13.3
			中值	13.0	13.0
		Hours per day watching TV	均值	2.1	2.1
			中值	2.0	2.0
		缺失	213	282	
		缺失	0	0	

图 45. 按分类行和列变量分组的刻度摘要

## 将分类变量嵌套在刻度变量之内

尽管上表可能提供了您所需要的信息，但提供信息的格式可能不是最易于解释的。例如，您可以比较使用互联网获取新闻的男性与不使用互联网获取新闻的男性的平均年龄，但是如果所比较的值相邻而不是彼此远离，则比较起来会较为容易。交换两个行变量的位置并将分类分组变量嵌套在三个刻度变量之内可以对表做出改进。有了刻度变量，嵌套层便不会对统计源变量产生影响。无论嵌套层如何，刻度变量总是统计源变量。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 在画布窗格上的表预览中，单击 *Age of respondent*，按住 **Ctrl** 并单击 *Highest year of school completed*，然后按住 **Ctrl** 并单击 *Hours per day watching TV* 以选择所有三个刻度变量。
3. 将三个刻度变量拖放到“行”区域的最左侧，以将分类变量 *Get news from internet* 嵌套在这三个刻度变量的每一个之内。
4. 单击**确定**以创建表。

				Gender	
				Male	Female
Age of respondent	Get news from internet	No	均值	47.0	48.8
			中值	45.0	46.0
		Yes	均值	38.7	41.1
			中值	35.0	38.0
Highest year of school completed	Get news from internet	No	均值	13.4	13.1
			中值	13.0	12.0
		Yes	均值	13.2	13.3
			中值	13.0	13.0
Hours per day watching TV	Get news from internet	No	均值	3.2	3.4
			中值	2.0	3.0
			缺失	213	282
		Yes	均值	2.1	2.1
			中值	2.0	2.0
			缺失	0	0

图 46. 嵌套在堆积刻度变量中的分类行变量

嵌套顺序的选择取决于要在表中强调的关系或比较。更改刻度变量的嵌套顺序不会改变汇总统计值，只会改变它们在表中的相对位置。

## 第 9 章 检验统计

### 检验统计

可以使用三种不同的显著性检验来研究列和行变量之间的关系。本章将讨论这些检验的输出，并将重点讨论嵌套和堆积的作用。请参阅第 23 页的第 3 章，『对分类变量使用堆积、嵌套和层』主题以获取更多信息。

样本数据文件

本章中的示例使用数据文件 *survey\_sample.sav*。请参阅数据文件主题以获取更多信息。

### 独立性检验 (卡方检验)

卡方独立性检验用于确定两个分类变量之间是否存在关系。例如，您可能希望确定 *Labor force status* 是否与 *Marital status* 相关。

1. 从菜单中选择:

分析 > 表 > 定制表...

2. 在表构建器中，将 *Labor force status* 从变量列表拖放到画布窗格的“行”区域中。
3. 将 *Marital status* 从变量列表拖放到“列”区域中。
4. 选择行作为汇总统计的位置。
5. 在“定义”组中选择 *Labor force status*，然后单击汇总统计。
6. 在“统计”列表中选择列 **N %**，然后将其添加到“显示”列表中。
7. 单击应用到选定数据。
8. 在“定制表”对话框中，单击检验统计选项卡。
9. 选择独立性检验 (卡方验证)。
10. 单击确定以创建表并获取卡方检验。

			Marital status				
			Married	Widowed	Divorced	Separated	Never married
Labor force status	Working full time	计数	778	44	295	58	392
		列 %	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	计数	138	20	35	9	102
		列 %	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	计数	23	2	9	1	11
		列 %	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	计数	13	3	10	0	32
		列 %	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	计数	168	150	53	6	17
		列 %	12.5%	53.0%	11.9%	6.5%	2.6%
	School	计数	9	1	7	2	60
		列 %	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	计数	200	55	25	13	35
		列 %	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	计数	16	8	12	4	14
		列 %	1.2%	2.8%	2.7%	4.3%	2.1%

图 47. *Labor force status* 与 *Marital status* 的关系

此表是 *Labor force status* 与 *Marital status* 的交叉制表，其中计数和列比例显示为汇总统计。计算列比例的目的是为了表明每列的百分比合计为 100%。如果这两个变量无关，则所有列中每个行中的比例都应基本相同。

比例上看起来有差异，但可以通过检查卡方检验来确认。

		Marital status
Labor force status	卡方	729.242
	df	28
	Sig.	.000*

\*. 卡方统计量在 0.05 级别处有意义。

图 48. Pearson 卡方检验

独立性检验假定 *Labor force status* 与 *Marital status* 无关，也就是说所有列的列比例相同，并且所有观察到的差异都归因于几率变异。卡方统计测量观察到的单元格计数与所有列的列比例相同的情况下所预期的计数之间的总体差异。卡方统计越大，观察到的和预期的单元格计数之间所存在的差异就越大，这明显表明列比例不相同，即独立性假设不正确；因此，*Labor force status* 与 *Marital status* 是相关的。

计算所得的卡方统计的值为 729.242。为了确定这是否足以否定独立性检验假设，将计算该统计的显著性值。该显著性值表示以下事件的概率：从自由度为 28 的卡方分布中抽取的随机变量大于 729.242。由于该值小于在“检验统计”选项卡上指定的 alpha 水平，因此可以否定独立性水平为 0.05 的假设。因此，*Labor force status* 与 *Marital status* 实际上是相关的。

## 嵌套和堆积对独立性检验的影响

独立性检验规则如下所示：对每个最里面的子表执行独立检验。要了解嵌套对该检验的影响，请考虑上一示例，但需要将 *Marital status* 嵌套在 *Gender* 级别中。

1. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 将 *Gender* 从变量列表拖放到画布窗格中 *Marital status* 上方的“列”区域。
3. 单击确定以创建表。

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	卡方	246.637	542.589
	df	28	28
	Sig.	.000*, a, b	.000*, a, b

结果基于每个最深处的子表中的非空行和列。

- \*. 卡方统计量在 0.05 级别处有意义。
- a. 该子表中超过 20% 单元格的期望单元格计数小于 5。卡方结果可能无效。
- b. 该子表中的最小期望单元计数小于 1。卡方结果可能无效。

图 49. Pearson 卡方检验

在将 *Marital status* 嵌套在 *Gender* 级别后，将执行两个检验，每个检验分别对应 *Gender* 的一个级别。每组检验的显著性值表明可以否定有关男性和女性的 *Marital status* 与 *Labor force status* 之间的独立性假设。但是，表注释指出每个表中超过 20% 的单元格的预期计数小于 5，并且最小预期单元格计数小于 1。这些注释表明这些表可能不符合卡方检验假设，因此检验的结果值得怀疑。



注：可以通过单元格边界去掉视图中的脚注。通过在“单元格属性”对话框中更改这些单元格的对齐方式，可以重新显示它们。

查看堆积对检验的影响方式：

- 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
- 将 *Highest degree* 从变量列表拖放到 *Labor force status* 下的“行”区域中。
- 单击**确定**以创建表。

		Gender	
		Male	Female
		Marital status	Marital status
Labor force status	卡方	246.637	542.589
	df	26	26
	Sig.	.000* <sup>a, b</sup>	.000* <sup>a, b</sup>
Highest degree	卡方	43.844	105.506
	df	16	16
	Sig.	.000*	.000*

结果基于每个最深处的子表中的非空行和列。

\*. 卡方统计量在 0.05 级别处有意义。

- 该子表中超过 20% 单元格的期望单元格计数小于 5。卡方结果可能无效。
- 该子表中的最小期望单元计数小于 1。卡方结果可能无效。

图 50. Pearson 卡方检验

在将 *Highest degree* 与 *Labor force status* 进行堆积后，将执行四个检验，其中包括 *Marital status* 和 *Labor force status* 的独立性检验以及每个 *Gender* 级别的 *Marital status* 和 *Highest degree* 的检验。*Marital status* 和 *Labor force status* 的检验结果与以前相同。*Marital status* 和 *Highest degree* 的检验结果表明这两个变量不是独立的。

## 比较列平均值

列平均值检验用于确定列中的分类变量与行中的连续变量是否存在关系。此外，可以使用该检验结果根据连续变量的平均值确定分类变量的类别的相对排序。例如，您可能需要确定 *Hours per day watching TV* 与 *Get news from newspapers* 是否相关。

- 从菜单中选择：
  - 分析 > 表 > 定制表...
- 单击**重置**恢复所有选项卡的缺省设置。
- 在表构建器中，将 *Hours per day watching TV* 从变量列表拖放到画布窗格的“行”区域中。
- 将 *Get news from newspapers* 从变量列表拖放到“列”区域中。
- 选择 *Hours per day watching TV* 并在“定义”组中单击**汇总统计**。
- 选择 **nnnn** 作为统计格式。
- 选择 **2** 作为要显示的小数位数。此时您会看到统计格式将显示为 **nnnn.nn**。
- 单击**应用到选定数据**。
- 在“定制表”对话框中，单击**检验统计**选项卡。
- 选择**比较列的平均值 (t-检验)**。

11. 单击**确定**以创建表并获取列平均值检验。

	Get news from newspapers	
	No	Yes
	均值	均值
Hours per day watching TV	2.92	2.74

图 51. “Get news from newspapers”与“Hours per day watching TV”的关系

此表显示通过和不通过报纸了解新闻的人的 *Hours per day watching TV* 平均值。这些平均值的检验差异表明：不通过报纸了解新闻的人每天看电视的时间比通过报纸了解新闻的人每天看电视的时间多出 0.18 小时。要了解此差异是否归因于几率变异，请检查列平均值检验。

	Get news from newspapers	
	No	Yes
	(A)	(B)
Hours per day watching TV		

图 52. 列平均值比较

列平均值检验表向列变量的每个类别分别指定一个字母关键字。对于 *Get news from newspapers*，类别 *No* 指定给字母 *A*，为类别 *Yes* 指定给字母 *B*。对于每对列，使用 *t* 检验比较列的平均值。由于该表只有两列，因此就只执行一个检验。对于每一显著性对，具有较小平均值的类别关键字将放在具有较大平均值的类别下方。由于表单元格中未出现任何关键字，因此这表明列平均值在统计上是不同的。

以 APA 样式符号显示显著性结果

如果您不想在单独的表中查看显著性结果，可以选择在主表中显示它们。显著性结果使用带有下标字母的 APA 样式符号进行标识。您需要完成比较列平均值的前面步骤，但在“检验统计”选项卡中进行以下更改：

1. 在“标识显著性差异”区域中，选择在主表中使用 **APA 样式下标**。
2. 单击**确定**以创建表并使用 APA 样式符号获取列平均值检验。

	Get news from newspapers	
	No	Yes
	Mean	Mean
Hours per day watching TV	2.92	2.74

图 53. 使用 APA 样式符号的列平均值比较

列平均值检验表向列变量类别指定一个下标字母。对于每对列，使用 *t* 检验比较列的平均值。如果某对值明显不同，则这些值将被指定不同的下标字母。由于该表只有两列，因此就只执行一个检验。由于本例中的列平均值共用相同下标字母，因此列平均值并非在统计上不同。

## 嵌套和堆积对列平均值检验的影响

列平均值检验规则如下所示：对每个最里面的子表执行一组独立的成对检验。要了解嵌套对该检验有何影响，请考虑上一示例，但需要将 *Hours per day watching TV* 嵌套在 *Labor force status* 级别中。

1. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 将 *Labor force status* 从变量列表拖放到画布窗格的“行”区域中。
3. 单击**确定**以创建表。

			Get news from newspapers	
			No	Yes
			(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B	
	Working part-time	Hours per day watching TV		
	Temporarily not working	Hours per day watching TV		
	Unemployed, laid off	Hours per day watching TV		
	Retired	Hours per day watching TV		
	School	Hours per day watching TV		
	Keeping house	Hours per day watching TV		
	Other	Hours per day watching TV		

图 54. 列平均值比较

将 *Hours per day watching TV* 嵌套在 *Labor force status* 级别中之后，将执行七组列平均值检验：每个 *Labor force status* 级别对应一组。为 *Get news from newspapers* 的类别指定相同的字母关键字。对于 *working full time* 响应者，会在 A 列中显示关键字 B。这表明全职员工 *Hours per day watching TV* 的平均值低于通过报纸了解新闻的员工。列中不会显示其他关键字，因此可以断定列平均值中没有任何其他在统计上的显著差异。

**Bonferroni 调节。** 在执行多个检验时，将对列平均值检验应用 Bonferroni 调节，以确保“检验统计”选项卡上所指定的 alpha 水平（或误确认率）可以应用于每组检验。在此表中，虽然执行七组检验，但在每组检验中只比较一对列，因此未应用 Bonferroni 调节。

查看堆积对检验的影响方式：

4. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
5. 将 *Get news from internet* 从变量列表中拖放到 *Get news from newspapers* 左侧的“列”区域中。
6. 单击**确定**以创建表。

			Get news from internet		Get news from newspapers	
			No	Yes	No	Yes
			(A)	(B)	(A)	(B)
Labor force status	Working full time	Hours per day watching TV	B		B	
	Working part-time	Hours per day watching TV	B			
	Temporarily not working	Hours per day watching TV				
	Unemployed, laid off	Hours per day watching TV	B			
	Retired	Hours per day watching TV	B			
	School	Hours per day watching TV	B			
	Keeping house	Hours per day watching TV	B			
	Other	Hours per day watching TV	B			

图 55. 列平均值比较

在将 *Get news from internet* 与 *Get news from newspapers* 进行堆积后，将执行 14 组列平均值检验，每组检验分别对应于 *Get news from internet* 和 *Get news from newspapers* 的每个 *Labor force status* 级别。此外，因为在每组检验中只比较一对列，所以不应用 Bonferroni 调节。对 *Get news from newspapers* 的检验与前面相同。对于 *Get news from internet*，分别向类别 *No* 和 *Yes* 指定了字母 A 和 B。除暂时无工作的响应者之外，对每组列平均值检验均在 A 列中显示关键字 B。这意味着通过 Internet 了解新闻的人 *Hours per day watching TV* 的平均值低于不通过报纸了解新闻的人。对 *Temporarily not working* 组未显示任何关键字，因此这些响应者的列平均值在统计上没有不同。

## 比较列的比例

列比例检验用于根据“行”分类变量的类别比例确定“列”分类变量的类别的相关排序。例如，在使用卡方检验确定 *Labor force status* 与 *Marital status* 相关后，您可能希望确定了解哪些行和列与此关系相关。

1. 从菜单中选择:

分析 > 表 > 定制表...

2. 单击重置恢复所有选项卡的缺省设置。
3. 在表构建器中，将 *Labor force status* 从变量列表拖放到画布窗格的“行”区域中。
4. 将 *Marital status* 从变量列表拖放到“列”区域中。
5. 在“定义”组中选择 *Labor force status*，然后单击汇总统计。
6. 在“统计”列表中选择列 **N %**，然后将其添加到“显示”列表中。
7. 从“显示”列表中取消选择计数。
8. 单击应用到选定数据。
9. 在“定制表”对话框中，单击检验统计选项卡。
10. 选择比较列的比例 (**z-检验**)。
11. 单击确定以创建表并获取列比例检验。

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		列 N %	列 N %	列 N %	列 N %	列 N %
Labor force status	Working full time	57.8%	15.5%	66.1%	62.4%	59.1%
	Working part-time	10.3%	7.1%	7.8%	9.7%	15.4%
	Temporarily not working	1.7%	.7%	2.0%	1.1%	1.7%
	Unemployed, laid off	1.0%	1.1%	2.2%	.0%	4.8%
	Retired	12.5%	53.0%	11.9%	6.5%	2.6%
	School	.7%	.4%	1.6%	2.2%	9.0%
	Keeping house	14.9%	19.4%	5.6%	14.0%	5.3%
	Other	1.2%	2.8%	2.7%	4.3%	2.1%

图 56. Labor force status 与 Marital status 的关系

此表是 Labor force status 与 Marital status 的交叉制表，其中列比例显示为汇总统计。

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		(A)	(B)	(C)	(D)	(E)
Labor force status	Working full time	B		AB	B	B
	Working part-time					ABC
	Temporarily not working					
	Unemployed, laid off					AB
	Retired	E	ACDE	E		
	School					ABC
	Keeping house	CE	CE		CE	
	Other					

图 57. 列比例比较

列比例检验表向列变量的每个类别分别指定了一个字母关键字。对于 Marital status，向类别 Married 指定了字母 A，向 Widowed 指定了字母 B，依此类推，向最后一个类别 Never married 指定了字母 E。对于每一对列，将使用  $z$  检验比较列的比例。将执行七组列比例检验，其中每组对应于 Labor force status 的每个级别。由于 Marital status 有五个级别，因此将在每组检验中对  $(5*4)/2 = 10$  对列进行比较，并使用 Bonferroni 调节来调整显著性值。对于每个显著性对，将具有较小比例的类别的关键字放在具有较大比例的类别下方。

对于与 Working full time 相关联的检验组，将在其他每一列中显示关键字 B。此外，还将在 C 列中显示关键字 A。其他列中没有出现其他的关键字。因此，可以断定全职的离婚人数比例大于全职的已婚人数比例，而后者又大于全职的丧偶人数比例。全职的单身或未婚人数比例与全职的离婚或已婚人数比例无法进行区分，但这两个比例大于全职的丧偶人数比例。

对于与 Working part time 或 School 相关联的检验，将在 E 列中显示关键字 A、B 和 C。其他列中没有出现其他的关键字。因此，求学或兼职的未婚人数比例大于求学或兼职的已婚、丧偶或离婚人数。

对于与 Temporarily not working 或 Other 劳动力状况相关联的检验，在任何列中均未显示任何关键字。因此，暂时未工作或处于其他未分类雇佣状况的已婚、丧偶、离异、单身或未婚人数比例没有显著差异。

与 Retired 相关联的检验表明退休的丧偶人数比例大于退休的所有其他婚姻类别人数比例。此外，退休的已婚或离异人数比例大于退休的未婚人数比例。

料理家务的已婚、丧偶、单身人数比例大于料理家务的离异或未婚人数比例。

未婚的 *Unemployed, laid off* 人数比例大于已婚或丧偶的未工作人数比例。此外，您还会看到 *Separated* 列标记为“.”，这表明由于观察到的 *Unemployed, laid off* 行中的单身人数比例为 0 或 1，因此未使用该列对无工作的响应者进行比较。

以 APA 样式符号显示显著性结果

如果您不想在单独的表中查看显著性结果，可以选择在主表中显示它们。显著性结果使用带有下标字母的 APA 样式符号进行标识。您需要完成比较列比例的前面步骤，但在“检验统计”选项卡中进行以下更改：

1. 在“标识显著性差异”区域中，选择在主表中使用 **APA** 样式下标。
2. 单击**确定**以创建表并使用 APA 样式符号获取列平均值检验。

		Marital status				
		Married	Widowed	Divorced	Separated	Never married
		Column %	Column %	Column %	Column %	Column %
Labor force status	Working full time	57.8% <sub>a</sub>	15.5% <sub>b</sub>	66.1% <sub>c</sub>	62.4% <sub>a,c</sub>	59.1% <sub>a,c</sub>
	Working part-time	10.3% <sub>a</sub>	7.1% <sub>a</sub>	7.8% <sub>a</sub>	9.7% <sub>a,b</sub>	15.4% <sub>b</sub>
	Temporarily not working	1.7% <sub>a</sub>	.7% <sub>a</sub>	2.0% <sub>a</sub>	1.1% <sub>a</sub>	1.7% <sub>a</sub>
	Unemployed, laid off	1.0% <sub>a</sub>	1.1% <sub>a</sub>	2.2% <sub>a,b</sub>	.0% <sup>1</sup>	4.8% <sub>b</sub>
	Retired	12.5% <sub>a</sub>	53.0% <sub>b</sub>	11.9% <sub>a</sub>	6.5% <sub>a,c</sub>	2.6% <sub>c</sub>
	School	.7% <sub>a</sub>	.4% <sub>a</sub>	1.6% <sub>a</sub>	2.2% <sub>a,b</sub>	9.0% <sub>b</sub>
	Keeping house	14.9% <sub>a</sub>	19.4% <sub>a</sub>	5.6% <sub>b</sub>	14.0% <sub>a</sub>	5.3% <sub>b</sub>
	Other	1.2% <sub>a</sub>	2.8% <sub>a</sub>	2.7% <sub>a</sub>	4.3% <sub>a</sub>	2.1% <sub>a</sub>

图 58. 使用 APA 样式符号的列比例比较

列比例检验表向列变量类别指定一个下标字母。对于每对列，使用 *z* 检验比较列的比例。如果某对值明显不同，则这些值将被指定不同的下标字母。

对于与 *Working full time* 相关联的检验组，“丧偶”类别的下标字母不在其他列中使用，而“单身”和“未婚”类别具有相同的两个下标。考虑这些下标字母和表中显示的实际比例，您可以像前面使用单独表的示例那样进行相同的分析。因此，可以断定全职的离婚人数比例大于全职的已婚人数比例，而后者又大于全职的丧偶人数比例。全职的单身或未婚人数比例与全职的离婚或已婚人数比例无法进行区分，但这两个比例大于全职的丧偶人数比例。前面示例的其余分析部分均适用。

## 嵌套和堆积对列比率检验的影响

列比例检验规则如下所示：对每个最里面的子表执行一组独立的成对检验。要了解嵌套对该检验有何影响，请考虑上一示例，但需要将 *Labor force status* 嵌套在 *Gender* 的级别中。

1. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 将 *Gender* 从变量列表拖放到画布窗格的“行”区域中。
3. 单击**确定**以创建表。

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A C
			Keeping house					
	Other				A			
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A B C
Keeping house			C E	C E		C		
Other								

图 59. 列比例比较

在将 *Labor force status* 嵌套在 *Gender* 的级别后，将执行 14 组列比例检验，每一组检验分别对应于每个 *Gender* 级别的每个 *Labor force status* 级别。为 *Marital status* 类别指定了同样的字母关键字。

对于表结果，有以下几点需要注意：

- 检验数量越多，具有零列比例的列也就越多。在单身响应者和丧偶男性中，最常出现这种情况。
- 前面在 *keeping house* 响应者之间看到的列差异似乎完全归因于女性。

查看堆积对检验的影响方式：

4. 再次打开表构建器（“分析”菜单，“表格”，“定制表”）。
5. 将 *Highest degree* 从变量列表拖放到 *Gender* 下方的“行”区域中。
6. 单击**确定**以创建表。

				Marital status				
				Married	Widowed	Divorced	Separated	Never married
				(A)	(B)	(C)	(D)	(E)
Gender	Male	Labor force status	Working full time	B		B	B	B
			Working part-time					A
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A C
			Keeping house					
	Other				A			
	Female	Labor force status	Working full time	B		A B	B	B
			Working part-time	B				B
			Temporarily not working					
			Unemployed, laid off					A
			Retired	E	A C D E	E		
			School					A B C
Keeping house			C E	C E		C		
Other								
Highest degree	LT High school		A C E					
	High school							
	Junior college	B		B		B		
	Bachelor	B				B		
	Graduate	B						

图 60. 列比例比较

在将 *Highest degree* 与 *Gender* 进行堆积之后，将执行 19 组列平均值检验，其中包括前面讨论的 14 组检验以及与每个 *Highest degree* 级别相对应的检验。为 *Marital status* 类别指定了同样的字母关键字。

对于表结果，有以下几点需要注意：

- 前面运行的 14 组检验的检验结果相同。
- 低于高中学历的丧偶响应者往往比已婚、离异或未婚响应者更常见。
- 受到高等教育的已婚、离异和未婚人员往往比丧偶人员更常见。

## 有关权重和多响应集的说明

即使其中某一变量是多重响应变量，个案权重也始终基于计数而不是响应数。



---

## 第 10 章 多响应集

定制表和图表构建器支持一种称为**多响应集**的特殊“变量”。多响应集不是通常意义上真正的“变量”。多响应集不显示在数据编辑器中，也不能由其他过程识别。多响应集使用多个变量记录对问题的答复，其中响应者可以给出多个答案。多响应集以与分类变量相同的方式对待，可以对分类变量执行的大多数操作也可以对多响应集执行。

多响应集根据数据文件中的多个变量进行构造。多响应集是数据文件中的特殊结构。可以在 IBM® SPSS® Statistics 数据文件中定义和保存多响应集，但不能从其他文件格式导入或向其导出多响应集。可以使用“复制数据属性”从其他 IBM SPSS Statistics 数据文件复制多响应集，从“数据编辑器”窗口的“数据”菜单可以访问该功能。

样本数据文件

本章中的示例使用数据文件 *survey\_sample.sav*。请参阅数据文件主题以获取更多信息。

此处提供的所有示例都在对话框中按字母顺序显示变量标签。变量列表显示属性是在“选项”对话框（“编辑”菜单，“选项”）的“常规”选项卡中指定的。

---

### 计数、响应、百分比和总计

所有可用于分类变量的汇总统计也都可用于多响应集。另有一些统计也可用于多响应集。

1. 从菜单中选择：

分析 > 表 > 定制表...

2. 将 *News sources*（多响应集 *\$mltnews* 的描述性标签）从变量列表拖放到画布窗格的“行”区域中。

变量列表中，“变量”旁边的图标将其标识为多二分集。



图 61. 多二分集图标

事实上，对于多二分集，每个“类别”都是单独的变量，并且类别标签就是变量标签（对于没有定义变量标签的变量，则为变量名）。在本例中，将要显示的计数表示对多二分集中每个变量都响应为 *Yes* 的个案的数量。

3. 右键单击画布窗格上表预览中的 *News sources*，然后从弹出菜单中选择**分类和总计**。

4. 在“分类和总计”对话框中选择（单击）**总计**，然后单击**应用**。

5. 再次右键单击 *News sources*，然后从弹出菜单中选择**汇总统计**。

6. 在“汇总统计”对话框中，在“统计”列表中选择**列 N %**，然后单击箭头将其添加到“显示”列表中。

7. 单击**应用到选定数据**，然后单击**确定**创建表。

		计数	列 N %
News sources	Get news from internet	867	41.7%
	Get news from radio	551	26.5%
	Get news from television	1077	51.8%
	Get news from news magazines	294	14.1%
	Get news from newspapers	805	38.7%
	合计	2081	100.0%

图 62. 多二分计数和列百分比

### 不全加的总计

如果查看表中的数字，您会发现“总计”与意欲加总的值差别很大。具体而言，总计似乎比它应达到的值低很多。这是因为表中每个“类别”的计数是变量值为 1（响应为 *Yes*）的个案的数量，而多二分集中所有五个变量的 *Yes* 响应的总数可能会轻易地超过数据文件中的个案总数。

不过，总“计数”是对多二分集中至少一个变量做出 *Yes* 响应的个案的总数，该计数从不会超过数据文件中个案的总数。在本例中，2,081 的总计数比数据文件中的个案总数少了近 800。如果这些变量都没有缺失值，则意味着近 800 个调查响应者表示他们不从任何这些新闻来源获取新闻。总数是列百分比的基数；因此本例中的列百分比合计大于显示为 100% 的总列百分比。

### 全加的总计

虽然“计数”通常是一个相当明确的术语，但是以上的示例显示了它在多响应集总计上下文中可能引发很大的混淆，对于多响应集，*响应* 通常是您真正需要的汇总统计。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上表预览中的 *News sources*，然后从弹出菜单中选择**汇总统计**。
3. 在“汇总统计”对话框中，在“统计”列表中选择**响应**，然后单击箭头将其添加到“显示”列表中。
4. 在“统计”列表中选择**列响应 %**，然后单击箭头将其添加到“显示”列表中。
5. 单击**应用到选定数据**，然后单击**确定**创建表。

		计数	列 N %	响应	列响应 %
News sources	Get news from internet	867	41.7%	867	24.1%
	Get news from radio	551	26.5%	551	15.3%
	Get news from television	1077	51.8%	1077	30.0%
	Get news from news magazines	294	14.1%	294	8.2%
	Get news from newspapers	805	38.7%	805	22.4%
	合计	2081	100.0%	3594	100.0%

图 63. 多二分响应和列响应百分比

对于多二分集中的每个“类别”，*响应*与*计数*是相同的，并且对于多二分集总是如此。但总计与它们之间的差别很大。响应的总数是 3,594，比总计数多出 1,500 多，比数据文件中个案的总数多出 700 多。

对于百分比，列 N % 和列响应 % 的总计都是 100%，但是多二分集中每个类别的百分比都比列响应的百分比低很多。这是因为列响应百分比的百分比基数是响应的总数（在本例中为 3,594），这导致计算出的百分比比用值为 2,081 的列百分比基数计算出的百分比低很多。

### 大于 100% 的百分比总计

在我们的示例中，虽然列 N % 列中各个值的合计明显大于 100%，但是列百分比和列响应百分比都会产生 100% 的总百分比。这样，如果您要显示基于总数而不是总响应数的百分比，但又希望“总计”百分比能准确地反映单个类别百分比的和，该怎么做？

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上表预览中的 *News sources*，然后从弹出菜单中选择**汇总统计**。
3. 在“汇总统计”对话框的“统计”列表中选择**列响应 % (基准: 计数)**，然后单击箭头将其添加到“显示”列表中。
4. 单击**应用到选定数据**，然后单击**确定**创建表。

		计数	列 N %	响应	列响应 %	列响应 % (基数: 计数)
News sources	Get news from internet	867	41.7%	867	24.1%	41.7%
	Get news from radio	551	26.5%	551	15.3%	26.5%
	Get news from television	1077	51.8%	1077	30.0%	51.8%
	Get news from news magazines	294	14.1%	294	8.2%	14.1%
	Get news from newspapers	805	38.7%	805	22.4%	38.7%
	合计	2081	100.0%	3594	100.0%	172.7%

图 64. 以计数作为百分比基数的列响应百分比

## 将多响应集用于其他变量

通常，可以像使用分类变量那样使用多响应集。例如，可以对包含分类变量的多响应集进行交叉制表，或者在分类变量内嵌套多响应集。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 将 *Gender* 从变量列表拖放到预览窗格上“行”区域的左侧，使多响应集 *News sources* 嵌套在 *Gender* 类别内。
3. 右键单击画布窗格上表预览中的 *Gender*，然后在弹出菜单中取消选中**显示变量标签**。
4. 对 *News sources* 执行相同的操作。

这将从表中移去具有变量标签的列（因为在这种情况下实际上不需要这些列）。

5. 单击**确定**以创建表。

		计数	列 N %	响应	列响应 %	列响应 % (基数 : 计数)
Male	Get news from internet	359	40.1%	359	23.3%	40.1%
	Get news from radio	233	26.0%	233	15.1%	26.0%
	Get news from television	451	50.3%	451	29.3%	50.3%
	Get news from news magazines	121	13.5%	121	7.9%	13.5%
	Get news from newspapers	375	41.9%	375	24.4%	41.9%
	合计	896	100.0%	1539	100.0%	171.8%
Female	Get news from internet	508	42.9%	508	24.7%	42.9%
	Get news from radio	318	26.8%	318	15.5%	26.8%
	Get news from television	626	52.8%	626	30.5%	52.8%
	Get news from news magazines	173	14.6%	173	8.4%	14.6%
	Get news from newspapers	430	36.3%	430	20.9%	36.3%
	合计	1185	100.0%	2055	100.0%	173.4%

图 65. 嵌套在分类变量内的多响应集

## 统计源变量和可用的汇总统计

就统计源变量而言，如果表中缺少刻度变量，那么将以相同的方式处理分类变量和多响应集：统计源维度中嵌套在最里面的变量是统计源变量。有些汇总统计只能指定给多响应集，这意味着如果您需要任何特殊的多重响应汇总统计，则多响应集必须是统计源维度中嵌套在最内层的变量。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 在画布窗格上的表预览中，将 *News sources* 拖放到 *Gender* 的左侧，以改变嵌套顺序。

因为现在分类变量 *Gender* 是嵌套在最内层的变量，从而也是统计源变量，所以将从表预览中移去所有特殊的多重响应汇总统计（响应数、列响应百分比）。

幸运的是表构建器可以“记住”这些设置。如果将 *News sources* 移动回其先前的位置，嵌套在 *Gender* 内，则所有与响应相关的汇总统计都会被恢复到表预览中。

## 多类别集和重复响应

多类别集提供了一种不可用于多二分集的功能：对重复响应进行计数的功能。许多情况下，多类别集中的重复响应可能表示存在编码错误。例如，对于诸如“您认为哪三个国家/地区制造的汽车最好？”的调查问题，回答瑞典、德国和瑞典可能不是有效的。

但是在其他情况下，重复响应可能会完全有效。例如，如果问题是“您最近拥有的三辆汽车是哪些国家/地区制造的？”，则回答瑞典、德国和瑞典完全合理。

定制表提供了在多类别集中选择重复响应的功能。缺省情况下不对重复响应进行计数，但是可以请求包含重复响应。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击重置清除以前的所有设置。
3. 将 *Car maker, most recent cars* 从变量列表拖放到画布窗格的“行”区域中。

在变量列表中，“变量”旁边的图标将其标识为多类别集。



图 66. 多类别集图标

对于多类别集，所显示的类别代表集合中所有变量的已定义值标签的公共集（而对于多二分集，“类别”实际上是集中每个变量的变量标签）。

4. 右键单击画布窗格上表预览中的 *Car maker, most recent cars*，然后从弹出菜单中选择**分类和总计**。
5. 在“分类和总计”对话框中选择（单击）**总计**，然后单击**应用**。
6. 再次右键单击 *Car maker, most recent cars*，然后从弹出菜单中选择**汇总统计**。
7. 在“汇总统计”对话框中，在“统计”列表中选择**响应**，然后单击箭头将其添加到“显示”列表中。
8. 单击**应用到选定数据**，然后单击**确定**创建表。

		计数	响应
Car maker, most recent cars	American	1938	1938
	Japanese	1327	1327
	Korean	695	695
	German	693	693
	Swedish	360	360
	Other	343	343
	合计	2832	5356

图 67. 多类别集：不含重复的计数和响应

缺省情况下不对重复响应进行计数；因此在该表中，计数和响应列中每个类别的值都相等。只有总计不同。

9. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
10. 单击**选项**选项卡。
11. 单击（复选）**计算多类别集中的重复响应**。
12. 单击**确定**以创建表。

		计数	响应
Car maker, most recent cars	American	1938	2797
	Japanese	1327	1717
	Korean	695	760
	German	693	754
	Swedish	360	363
	Other	343	359
	合计	2832	6770

图 68. 包含重复响应的多类别集

在该表中，计数和响应列中的值之间存在着颇为明显的差异，特别是对于 *American* 汽车；这意味着许多响应者拥有多辆 *American* 汽车。

## 对多响应集进行显著性检验

可以用与分类变量在本质上相同的方法对多响应集进行显著性检验。

- 对于独立性检验（卡方检验）或比较列的比例（z-检验），将基于计数执行检验，而“计数”必须是表中所显示的汇总统计之一。
- 对于多类别集，如果“选项”选项卡上选择了**计算多类别集中的重复响应**，则不会执行比较列的比例或列平均值的检验（t-检验）。请参阅第 13 页的『定制表：“选项”选项卡』主题以获取更多信息。

## 对多响应集进行独立性检验

本示例对一个分类变量和一个多响应集创建交叉制表，并对该交叉制表执行独立性的卡方检验。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击**重置**清除以前的所有设置。
3. 将 *News sources*（多二分集 *\$mltnews* 的描述性标签）从变量列表拖放到画布窗格的“列”区域中。
4. 将 *Gender* 从变量列表拖放到画布窗格的“行”区域中。
5. 单击**检验统计**选项卡。
6. 选择（复选）**独立性检验(卡方验证)**。
7. 如果尚未选择**检验中包含多重响应变量**，请选择该选项。
8. 单击**确定**以运行该过程。

		News sources
Gender	卡方	10.266
	df	5
	Sig.	.068

图 69. 卡方检验结果

卡方检验的显著性水平为 0.068，这表示男性和女性在新闻来源的选择上可能不存在显著的差异（假设您使用 0.05 或更低的显著性值作为确定统计显著性的准则）。

## 使用多响应集比较列平均值

本示例计算由多响应集定义类别内某刻度变量的平均值，并将每个类别平均值与其他类别平均值逐一比较以确定显著性差异。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击**重置**清除以前的所有设置。
3. 将 *News sources*（多二分集 *\$mltnews* 的描述性标签）从变量列表拖放到画布窗格的“列”区域中。
4. 将 *Age of respondent* 拖放到画布窗格的“行”区域中。
5. 单击**检验统计**选项卡。
6. 选择（复选）**比较列的平均值 (t-检验)**。
7. 如果尚未选择**检验中包含多重响应变量**，请选择该选项。
8. 单击**确定**以运行该过程。

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	均值	均值	均值	均值	均值
Age of respondent	52	40	48	40	40

列均值的比较

	News sources				
	Get news from newspapers	Get news from news magazines	Get news from television	Get news from radio	Get news from internet
	(A)	(B)	(C)	(D)	(E)
Age of respondent	B C D E		B D E		

结果基于双侧检验，假定方差相等且显著性水平为 0.05。对于每个显著对，较小类别的关键字显示在具有较大均值的类别下。

图 70. 显著性检验结果

- 多响应集的每个类别都用字母 (A、B、C、D、E) 来标识，并且对于另一个类别的平均值既低于该类别的平均值又与其差别颇大的每个类别，显示代表具有较低平均值的类别的字母。
- *Get news from newspapers* (A) 的平均年龄最高，并且所有其他类别的平均值都与其有较大的差异。
- *Get news from television* (C) 的平均年龄次之，并且所有剩余类别的平均值 (B、D 和 E) 都与其有较大的差异。（如先前所述，C 与 A 也有较大的差异。）
- *Get news from magazine* (B)、*Get news from radio* (D) 和 *Get news from internet* (E) 的平均年龄之间没有显著的差异。





---

## 第 11 章 缺失值

很多数据文件都包含一些缺失数据。导致缺失数据的原因有多种。例如，调查的响应者可能不会对每一个问题都作出回答；一些变量可能不适用于某些个案；编码错误可能会导致丢弃某些值。

在 IBM SPSS Statistics 中有两种缺失值：

- **用户缺失值。** 定义为包含缺失数据的值。可以为这些值指定值标签，以标识数据缺失的原因（例如，为男性的怀孕项使用代码 99，不适用作为值标签）。
- **系统缺失值。** 如果数值变量没有值，则为其指定系统缺失值。在数据编辑器的“数据视图”中用句点表示系统缺失值。

存在很多可以帮助补偿缺失数据的影响甚至分析缺失数据模式的功能。不过，本章的作用较为简单：用于描述“定制表”如何处理缺失数据以及缺失数据如何影响汇总统计的计算。

样本数据文件

本章使用数据文件 *missing\_values.sav*。请参阅数据文件主题以获取更多信息。这是一个非常简单的全人工数据文件；该文件只包含一个变量和十个个案，用于说明缺失值的基本概念。

---

### 没有缺失值的表

缺省情况下，定制表中不显示用户缺失类别（并且从不显示系统缺失值）。

1. 从菜单中选择：

分析 > 表 > 定制表...

2. 在表构建器中，将 *Variable with missing values*（文件中的唯一变量）从变量列表拖放到画布窗格的“行”区域。
3. 右键单击画布窗格上的该变量，然后从弹出菜单中选择**分类和总计**。
4. 在“分类和总计”对话框中单击（选中）**总计**，然后单击**应用**。
5. 再次右键单击画布窗格上表预览中的 *Variable with missing values*，然后从弹出菜单中选择**汇总统计**。
6. 在“汇总统计”对话框中，在“统计”列表中选择**列 N %**，然后单击箭头将其添加到“显示”列表中。
7. 单击**应用到选定数据**。

您可能会注意到，画布窗格上表预览中所显示的类别与“类别”列表（位于表构建器左侧的变量列表下方）中所显示的类别之间存在细微的差异。“类别”列表包含一个标注为缺失值的类别，而表预览中不包含该类别，原因是缺省情况下会排除缺失值类别。在标签中“值”是复数，这表示变量包含两个或更多用户缺失类别。

8. 单击**确定**以创建表。

		计数	列 N %
Variable with missing values	Low	2	28.6%
	Medium	3	42.9%
	High	2	28.6%
	合计	7	100.0%

图 71. 没有缺失值的表

该表中的每项数据都很完整。类别值相加得到总计，百分比准确地反映了使用总数作为百分比基数所得到的值（例如， $3/7 = 0.429$ ，即 42.9%）。不过，该总数并不是数据文件中个案的总数，而是具有非缺失值的个案（即该变量既不包含用户缺失值也不包含系统缺失值的个案）总数。

## 在表中包含缺失值

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 右键单击画布窗格上表预览中的 *Variable with missing values*，然后从弹出菜单中选择分类和总计。
3. 在“分类和总计”对话框中单击（复选）缺失值，然后单击应用。

现在，表预览包含一个缺失值类别。尽管表预览中仅为缺失值显示一个类别，但表中会显示所有用户缺失类别。

4. 再次右键单击画布窗格上表预览中的 *Variable with missing values*，然后从弹出菜单中选择汇总统计。
5. 在“汇总统计”对话框中，单击（复选）定制关于总计和小计的汇总统计。
6. 在定制汇总统计列表中选择有效 N，然后单击箭头将其添加到“显示”列表中。
7. 对总计 N 执行相同的操作。
8. 单击应用到选定数据，然后在表构建器中单击确定创建表。

		计数	列 N %	有效的 N	合计 N
Variable with missing values	Low	2	22.2%		
	Medium	3	33.3%		
	High	2	22.2%		
	Don't know	1	11.1%		
	Not applicable	1	11.1%		
	合计	9	100.0%	7	10

图 72. 具有缺失值的表

现在，表中显示了两个已定义的用户缺失类别：*Don't know* 和 *Not applicable*，并且总数是 9 而不是 7，反映出添加了两个具有用户缺失值的个案（每个用户缺失类别中一个）。因为列百分比基于非缺失值和用户缺失值的数量，所以列百分比现在也发生了变化。只有系统缺失值不包含在百分比计算中。

有效 N 显示非缺失个案的总数 (7)，而总计 N 显示个案总数（包括用户缺失和系统缺失个案）。个案的总数是 10，比计数列中作为总计显示的非缺失加用户缺失值的计数多一个。这是因为有一个包含系统缺失值的个案。

9. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
10. 右键单击画布窗格上表预览中的 *Variable with missing values*，然后从弹出菜单中选择汇总统计。
11. 在上方的“统计”列表（不是“定制关于总计和小计的摘要”）中选择列有效 N %，然后单击箭头将其添加到“显示”列表。
12. 对列总计 N % 执行相同的操作。
13. 还可以将这两项都添加到“定制关于总计和小计的汇总统计”列表中。

14. 单击应用到选定数据，然后单击确定创建表。

		计数	列 N %	有效列 N %	列总数 N %	有效的 N	合计 N
Variable with missing values	Low	2	22.2%	28.6%	20.0%		
	Medium	3	33.3%	42.9%	30.0%		
	High	2	22.2%	28.6%	20.0%		
	Don't know	1	11.1%	.0%	10.0%		
	Not applicable	1	11.1%	.0%	10.0%		
	合计	9	100.0%	100.0%	100.0%	7	10

图 73. 具有缺失值以及有效和总计百分比的表

- 列 N % 是每个类别中基于非缺失值和用户缺失值的百分比（因为表中已显式地包含了用户缺失值）。
- 列有效 N % 是每个类别中仅基于有效的非缺失个案的百分比。这些值与不包含用户缺失值的原始表中的列百分比相同。
- 列总计 N % 是每个类别中基于所有个案（包括用户缺失个案和系统缺失个案）的百分比。如果将此类别中的单个类别百分比相加，您将看到相加的结果仅仅等于 90%，这是因为在总共 10 个个案中有一个个案 (10%) 包含系统缺失值。尽管百分比计算的基数中包含了该个案，但是表中没有为具有系统缺失值的个案提供类别。



---

## 第 12 章 设置表格式和定制表

---

### 设置表格式和定制表

使用定制表可以在构建表的过程中控制一些表格式属性，包括：

- 汇总统计的显示格式和标签
- 最小和最大数据列宽
- 空单元格中显示的文本或值

这些设置会一直保留在表构建器界面中（直到您更改这些设置、重置表构建器设置或打开不同的数据文件），这使您可以创建具有相同格式属性的多个表，而无需在创建表之后手动编辑这些表。通过使用表构建器界面中的“粘贴”按钮将命令语句粘贴到语法窗口，然后将该窗口中的内容保存为文件，还可以保存这些格式设置以及所有其他表参数。

还可以在创建表之后，使用透视表的“查看器”中提供的所有格式功能更改表的许多格式属性。但是，本章重点介绍在创建表之前对表格式属性的控制。有关透视表的更多信息，请使用“帮助”系统中的“索引”选项卡，并键入透视表作为关键字。

样本数据文件

本章中的示例使用数据文件 *survey\_sample.sav*。请参阅数据文件主题以获取更多信息。

此处提供的所有示例都在对话框中按字母顺序显示变量标签。变量列表显示属性是在“选项”对话框（“编辑”菜单，“选项”）的“常规”选项卡中设置的。

---

### 汇总统计显示格式

定制表会力图对汇总统计应用相对智能的缺省格式，但有时您可能希望覆盖这些缺省值。

1. 从菜单中选择：

分析 > 表 > 定制表...

2. 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
3. 将 *Confidence in television* 拖放到“行”区域中 *Age category* 的下面，使这两个变量堆积到行维度中。
4. 右键单击画布窗格上表预览中的 *Age category*，然后从弹出菜单中选择**选择所有行变量**。
5. 再次右键单击 *Age category*，然后从弹出菜单中选择**类别和总计**。
6. 在“分类和总计”对话框中，选择（复选）**总计**，然后单击**应用**。
7. 右键单击画布窗格上表预览中的任意一个变量，然后从弹出菜单中选择**汇总统计**。
8. 在“统计”列表中选择**列 N %**，然后单击箭头键将其添加到“显示”列表中。
9. 选择（复选）**定制关于总计和小计的汇总统计**。
10. 在定制汇总统计的“统计”列表中，选择**列 N %**，然后单击箭头将其添加到“显示”列表中。
11. 对**平均值**执行相同的操作。
12. 然后单击**应用到全部**。

表预览中的占位符值反映了每个汇总统计的缺省格式。

- 对于计数，缺省的显示格式是 **nnnn**，即没有小数位的整数值。
- 对于百分比，缺省的显示格式是 **nnnn.n%**，即包含一个小数位的数值，后面有一个百分号。
- 对于平均值，这两个变量的缺省显示格式是不同的。

对于不是某种格式的计数（包括有效 N 和总计 N）或百分比的汇总统计，缺省的显示格式是在数据编辑器中为变量定义的显示格式。如果在数据编辑器中的“变量视图”中查看变量，您将会看到 *Age category*（变量 *agecat*）被定义为包含两个小数位，而 *Confidence in television*（变量 *contv*）被定义为包含零个小数位。

如果两个平均值都显示相同数量的小数位，则可能会更好，因此缺省格式有时可能不是您所希望的格式，而上述情况就属于此类情况。

13. 右键单击画布窗格上表预览中的任意一个变量，然后从弹出菜单中选择**汇总统计**。

对于平均值，“显示”列表中的“格式”单元格表示格式为**自动**，这意味着将使用已定义的变量显示格式，并禁用“小数”单元格。若要指定小数位数，首先需要选择一种不同的格式。

14. 在定制汇总统计“显示”列表中，单击平均值的“格式”单元格，然后从格式下拉列表中选择 **nnnn**。
15. 在“小数”单元格中，输入值 1。
16. 然后单击**应用到全部**将设置应用到两个变量。

现在表预览指示两个平均值都将以一个小数位的格式显示。（现在可以继续创建表，但是由于 *Age category* 的实际数值编码只在 1 到 6 的范围内变动，因此您会发现较难解释该变量的“平均值”。）

---

## 汇总统计的显示标签

除了汇总统计的显示格式以外，还可以控制每个汇总统计的描述标签。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击**重置**清除表构建器中所有以前的设置。
3. 在表构建器中，将 *Age category* 从变量列表拖放到画布窗格上的“行”区域中。
4. 将 *How get paid last week* 从变量列表拖放到画布窗格上的“列”区域中。
5. 右键单击画布窗格上表预览中的 *Age category*，然后从弹出的上下文菜单中选择**汇总统计**。
6. 在“统计”列表中选择**列 N %**，然后单击箭头键将其添加到“显示”列表中。
7. 双击“显示”列表中“标签”单元格中的**列字**，以编辑该单元格的内容。从标签中删除**列字**，将标签更改为简单的 **%**。
8. 用相同的方法编辑计数的“标签”单元格，将标签更改为简单的 **N**。

在此处，我们来更改“列 N %”统计的格式，以移去不必要的百分号（因为列标签指示列包含的是百分比）。

9. 单击**列 N %**的“格式”单元格，然后从格式下拉列表中选择 **nnnn.n**。
10. 然后单击**应用到选定数据**。

表预览显示了已修改的显示格式和已修改的标签。

11. 单击**确定**以创建表。

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

图 74. 具有已修改的汇总统计标签的表

## 列宽

您可能已经注意到，以上示例中的表非常宽。解决该问题的一种方法是简单地交换行和列变量。另一种方法是使列变窄，因为列看起来没有必要那么宽。（实际上，我们之前缩短汇总统计标签就是为了能够使列更窄。）

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击选项选项卡。
3. 在“数据列宽度”组中，选择**定制**。
4. 对于“最大值”，键入 36。（确保“单位”的设置为磅。）
5. 单击**确定**以创建表。

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	0	.0	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	0	.0	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

图 75. 已缩短列宽的表

表现在紧凑得多了。

## 为空单元格显示值

缺省情况下，空单元格（不包含个案的单元格）中显示 0。可以让这些单元格不显示任何内容（留为空白），或者指定一个文本字符串显示在空单元格中。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 单击选项选项卡。
3. 在“数据单元格外观”组中，为“空单元格”选择**文本**，并键入无。
4. 单击**确定**以创建表。

		How get paid last week											
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate	
		N	%	N	%	N	%	N	%	N	%	N	%
Age category	Less than 25	91	14.0	无	无	12	9.7	3	2.0	7	3.1	14	7.7
	25 to 34	175	26.9	5	29.4	33	26.6	37	24.8	63	28.0	31	17.1
	35 to 44	185	28.5	5	29.4	42	33.9	45	30.2	66	29.3	61	33.7
	45 to 54	124	19.1	5	29.4	25	20.2	38	25.5	58	25.8	41	22.7
	55 to 64	52	8.0	无	无	10	8.1	23	15.4	29	12.9	19	10.5
	65 or older	23	3.5	2	11.8	2	1.6	3	2.0	2	.9	15	8.3

图 76. 在空单元格中显示“无”的表

现在，表中的四个空单元格都显示文本无，而不是值 0。

## 为缺失统计显示值

如果不能计算某统计，则缺省的显示值是一个句点 (.)，该符号用于表示系统缺失值。系统缺失值与“空”单元格不同，因此应分别控制缺失统计的显示值与不包含个案的单元格的显示值。

1. 打开表构建器（“分析”菜单，“表格”，“定制表”）。
2. 将 *Hours per day watching TV* 从变量列表拖放到画布上“列”区域的顶部，*How get paid last week* 的上面。

由于 *Hours per day watching TV* 是刻度变量，因此它会自动成为统计源变量，而汇总统计则更改为平均值。

3. 右键单击画布窗格上表预览中的 *Hours per day watching TV*，然后从弹出的上下文菜单中选择**汇总统计**。
4. 在“统计”列表中选择**有效 N**，然后单击箭头键将其添加到“显示”列表中。
5. 单击**应用到选定数据**。
6. 单击**选项选项卡**。
7. 在“无法计算的统计数据”文本字段中，键入 *NA*。
8. 单击**确定**以创建表。

		Hours per day watching TV													
		How get paid last week													
		Hourly wage		Daily wage		Weekly wage		Monthly salary		Annual salary		Other pay rate			
		均值	有效的 N	均值	有效的 N	均值	有效的 N	均值	有效的 N	均值	有效的 N	均值	有效的 N	均值	有效的 N
Age category	Less than 25	3	71	.NA	0	3	10	2	3	2	6	2	8		
	25 to 34	3	134	5	2	2	30	2	29	2	52	2	22		
	35 to 44	3	136	2	5	3	30	2	34	2	47	3	46		
	45 to 54	2	90	2	4	2	22	2	36	2	45	2	34		
	55 to 64	3	40	.NA	0	3	7	2	15	2	23	3	15		
	65 or older	3	18	2	2	1	1	.NA	0	1	2	3	11		

图 77. 为缺失统计显示“NA”的表

表中三个单元格中的平均值显示文本 *NA*。在每个个案中，相应的有效 *N* 值说明了原因：不存在任何用于计算平均值的个案。

不过，您可能会注意到细微的差异：其中一个有效 *N* 值（共三个）显示为 0，而不是在没有个案的单元格中所显示的标签无。这是因为虽然没有可用于计算平均值的有效个案，但是类别并不真是空的。如果回到仅仅包含这两个分类变量的原始表，您将看到此交叉制表类别中确实有三个个案。但是，因为所有三个个案的刻度变量 *Hours per day watching TV* 都具有缺失值，所以没有有效个案。



---

## 第 13 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。以下每种语言的“Samples”子目录中都有一个独立的文件夹：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。
- **afatoxin.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<sup>1</sup>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 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<sup>2</sup>，52 名学生被要求以 10 点的标度对 15 种情况和 15 种行为的组合进行评价，该 10 点的标度从 0 = “极得体”到 9 = “极不得体”。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。

---

1. 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.

2. 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.

- **breakfast.sav**。 在一项经典研究中<sup>3</sup>，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<sup>4</sup>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刻度权重。
- **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**。 在常用示例<sup>5</sup>中，有一家公司想要销售一种新型的地毯清洁用品。该公司希望调查以下会对消费者偏好产生影响的五个因素：包装设计、品牌名称、价格、优质家用品标志、以及退货保证。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**。 这是关于六种冰咖啡的认知品牌形象<sup>6</sup>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。

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

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

5. Green, P. E., and Y. Wind. 1973. *Multiaattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

6. 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.

- **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。
- **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。
- **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。
- **dietstudy.sav**。该假设数据文件包含对 "Stillman diet"<sup>7</sup> 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<sup>8</sup> 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）。

7. 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.

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

- **guttman.sav**。Bell<sup>9</sup>创建了一个表，用来阐释可能的社会群体。Guttman<sup>10</sup>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<sup>11</sup> 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 1999–2000 年期间 Algonquin, Ill. 地区新住宅开发中的房屋销售情况。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 个观察值。在以前的研究人员中，<sup>12</sup> 和 <sup>13</sup> 发现了这些变量之间的非线性，这妨碍了标准回归方法。

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

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

11. 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.

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

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

- **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 接受实践教程。在培训课程结束时，对每名员工进行测试并记录他们的分数。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试的分数。
- **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<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。在规定的時間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。

---

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

- **ships.sav**。在别处被提出和分析的<sup>15</sup>关于波浪对货船造成的损坏的数据集。在给定了船的类型、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，从而反映总体趋势。
- **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。
- **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。
- **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。它提供了区间数据的优秀示例并且已在别处被提出和分析<sup>16</sup>。

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

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

- **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<sup>17</sup>。
- **verd1985.sav**。该数据文件涉及某项调查<sup>18</sup>。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。宠物被刻度化为多名义而年龄被刻度化为有序；所有其他变量都被刻度化为单名义。
- **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<sup>19</sup>。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。

---

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

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

19. 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 可能在其他国家或地区不提供本文中讨论的产品、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。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。

IBM 公司可能已拥有或正在申请与本文档内容有关的各项专利。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.

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，或用书面方式将查询寄往：

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

本条款不适用英国或任何这样的条款与当地法律不一致的国家或地区：International Business Machines Corporation“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某种特定用途的保证。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 可以随时对本资料中描述的产品和/或程序进行改进和/或更改，而不另行通知。

本信息中对任何非 IBM Web 站点的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 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 Software Group  
ATTN: Licensing

200 W. Madison St.  
Chicago, IL; 60606  
U.S.A.

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。

本资料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 国际软件许可协议或任何同等协议中的条款提供。

此处所含的性能数据均在受控环境下决定。因此，在其他操作环境中获得的结果可能差异较大。有些测量可能在开发级的系统中进行，不保证这些测量结果与常用系统上的测量结果相同。此外，有些测量结果可能通过推断来估计得出。实际结果可能有所差异。此文档的用户应针对其具体环境验证适用的数据。

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 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。

有关 IBM 未来方向或意向的所有声明均可能未经通知即变更或撤销，并且仅代表目标和目的。

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。

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：

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。

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：

©（贵公司的名称）（年）。此部分代码是根据 IBM Corp. 公司的样本程序衍生出来的。

© Copyright IBM Corp. \_（输入年份）\_. All rights reserved.

---

## 商标

IBM、IBM 徽标和 ibm.com 是 International Business Machines Corp.，在全球许多管辖区域注册的商标或注册商标。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 商标列表可从 Web 站点 [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) 上的『版权和商标信息』部分获取。

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、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。

Java 和所有基于 Java 的商标和徽标是 Oracle 及/或其分支机构的商标和注册商标。



# 索引

## [ B ]

- 百分比
  - 多响应集 7
  - 缺失值 78
  - 在定制表中 6, 7
- 百分比值
  - 在定制表中 18, 19
- 变量标签
  - 不在定制表中显示 5
- 标签
  - 更改汇总统计的标签文本 82
- 标题
  - 设定表格 14
- 标准差
  - 设定表格 7
- 表
  - 设定表格 1
- 不同变量的汇总统计
  - 堆积表 55

## [ C ]

- 测量级别
  - 在定制表中更改 1
- 层变量
  - 打印分层表 29
  - 堆积层变量 28
  - 嵌套层变量 28
  - 设定表格 27, 28
- 拆分文件处理
  - 定制表 4

## [ D ]

- 打印包含层的表 29
- 定制表
  - 百分比值 6, 7
  - 标题 14
  - 拆分文件处理 4
  - 重新排列类别 9
  - 定制总计 8
  - 多类别集 13
  - 多响应集 1
  - 多响应集的百分比 7
  - 分类变量 1
  - 分类变量的值标签 1
  - 更改测量级别 1
  - 更改汇总统计维度 8
  - 共周界表 13, 43
  - 汇总统计 6, 7

## 定制表 (续)

- 汇总统计显示格式 9
- 检验统计 14
- 角标签 14
- 具有共享类别的变量表 13, 43
- 刻度变量 1
- 刻度摘要的缺失值排除 13
- 空单元格 13
- 控制显示的小数位数 5
- 列宽 13
- 排除类别 9
- 频率表 13, 43
- 平均频率表 8
- 如何构建表 3
- 文字说明 14
- 显示格式 5
- 显示和隐藏变量名以及变量标签 5
- 小计 9
- 已计算类别 9
- 总计 9
- 定制总计汇总统计 49
- 堆积变量
  - 不同变量的汇总统计 55
  - 堆积层变量 28
  - 多个汇总统计源变量 49
  - 刻度变量 53
  - 设定表格 23
- 对类别进行重新排序
  - 设定表格 9
- 对类别排序
  - 设定表格 21
- 多响应集 69
  - 百分比值 7
  - 多类别集中的重复响应 13
  - 显著性检验 68, 74

## [ F ]

- 范围
  - 设定表格 7
- 方差
  - 设定表格 7
- 分组摘要
  - 刻度变量 56

## [ G ]

- 共周界表 13, 43

## [ H ]

- 后期计算类别
  - 设定表格 12, 37
- 汇总统计 47
  - 定制总计汇总统计 49
  - 堆积表 49
  - 堆积表中不同变量的摘要 55
  - 更改标签文本 82
  - 显示格式 81
  - 源变量 47
  - 源维度 47
- 汇总统计源变量
  - 刻度变量 57

## [ J ]

- 计数
  - 与有效 N 54
- 检验统计
  - 设定表格 14, 59
- 交叉制表
  - 设定表格 19
- 角标签
  - 设定表格 14

## [ K ]

- 卡方统计
  - 设定表格 59
- 刻度变量
  - 按行和列分类变量分组的摘要 57
  - 堆积 53
  - 多个汇总统计 54
  - 分组摘要 56
  - 汇总统计 53
  - 嵌套 57
- 空单元格
  - 定制表中显示的值 13, 83
- 控制显示的小数位数 20

## [ L ]

- 列比例统计
  - 定制表 64
- 列宽
  - 定制表中的控制 13, 83
- 列平均值统计
  - 定制表 61

## [ P ]

- 排除类别
  - 设定表格 9, 21
- 拼并类别
  - 设定表格 35
- 频率表
  - 设定表格 13, 43
- 平均频率表 8, 49
- 平均值 53
  - 设定表格 7

## [ Q ]

- 嵌套变量
  - 刻度变量 57
  - 设定表格 24, 26
- 缺失值 54, 77
  - 对百分比计算的影响 78
  - 在定制表中包含 78

## [ R ]

- 日期
  - 在定制表中包含当前日期 14

## [ S ]

- 删除类别
  - 设定表格 9, 21
- 设定表格
  - 百分比值 18, 19
  - 边际总计 20
  - 层变量 27, 28
  - 打印分层表 29
  - 堆积变量 23
  - 对调行与列变量 27
  - 对类别排序 21
  - 多响应集 69
  - 分类变量的简单表 17
  - 更改汇总统计的标签 18
  - 行与列百分比 18
  - 后期计算类别 12, 37
  - 检验统计 59
  - 交叉制表 19
  - 紧凑视图 26
  - 具有已排除类别的表中的总计 21
  - 排除类别 21
  - 拼并类别 35
  - 嵌套变量 24, 26
  - 嵌套层变量 28
  - 统计源维度 19
  - 显著性检验和多重响应 68
  - 小计 31
  - 已计算类别 12, 37

- 设定表格 (续)
  - 隐藏统计标签 17
  - 隐藏小计类别 35
  - 总计 19, 31
- 省略类别
  - 设定表格 21
- 时间
  - 在定制表中包含当前时间 14

## [ T ]

- 统计信息
  - 定制总计汇总统计 49
  - 堆积表 49
  - 汇总统计 47

## [ W ]

- 文字说明
  - 设定表格 14

## [ X ]

- 系统缺失值 77
- 显示格式 20
  - 定制表中的汇总统计 9, 81
- 显示类别值 50
- 显著性检验
  - 多响应集 74
  - 设定表格 14
- 小计 34
  - 设定表格 9, 31
  - 隐藏小计类别 35
- 小数位
  - 控制定制表中显示的小数位数 5, 20, 81

## [ Y ]

- 样本文件
  - 位置 85
- 已计算类别
  - 从小计 39
  - 设定表格 12, 37
  - 显示格式 12
  - 在表达式中隐藏类别 38
- 用户缺失值 77
- 有效 N 54, 78
  - 设定表格 7

## [ Z ]

- 在定制表中隐藏统计标签 17

- 值
  - 显示类别标签和值 50
- 值和值标签 50
- 中间值
  - 设定表格 7
- 中位数 54
- 众数
  - 设定表格 7
- 子组总计 32
- 总和
  - 设定表格 7
- 总计
  - 层 34
  - 定制表的边际总计 20
  - 嵌套表 32
  - 设定表格 9, 19, 31
  - 显示位置 32
  - 已排除的类别 31
  - 组总计 32
- 总计 N 78
- 组总计 32
- 最大值
  - 设定表格 7
- 最小值
  - 设定表格 7





Printed in China