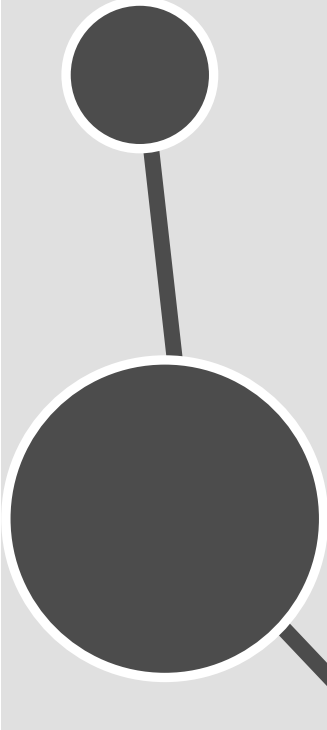




第 5 章

报表与 BootStrap

在前面章节讲解了常用的基本统计分析的方法。在本章将着重讲解应用分析后的结果输出。在 SPSS 的操作过程中，选项、格式的设置、统计量的选择，直接影响统计结果的输出。

例如，用户需要合理的选择统计量，这样可以得出用户需要的数据。而通过合理的选项设置，用户可以使运算的结果更为精确。其次，用户通过合理的格式设置，可以方便地将分析结果直接打印输出或通过其他媒介输出结果。

本章将详细介绍 SPSS 中常用的报表及 BootStrap 的使用方法。通过在本章的学习，用户可以熟练地掌握使用报表输出数据，对输出结果进行解读分析，并根据不同的数据类型、数据结构或运算要求，选择不同的报表形式。

本章学习要点：

- 了解报表的概念
- 熟悉代码本的使用方法
- 掌握 OLAP 的使用方法
- 掌握个案汇总的使用方法
- 掌握按列汇总的使用方法
- 了解 BootStrap 的计算原理
- 掌握 BootStrap 的使用方法

5.1

代码本

代码本功能是 SPSS 新增的一个选项，主要用于对数据中的指定变量分别进行描述性统计分析。它的特点是简明清晰，使用方便。

代码本对于名义类型和序号类型变量的统计包括计数和百分比。对于度量类型变量的统计包括均值、标准差和四分位数。

5.1.1 报表分析概述

报表就是指采用表格、图表等形式动态地呈现数据的统计情况。在各种报告里，除了使用各种图形直观地呈现资料以外，采用表格的形式汇报具体的数据也是必不可少的。

在 SPSS 中，可以通过【分析】|【报告】命令来实现。而在【报告】子菜单主要用来处理定量数据，包括对单个变量及多个变量进行描述性统计，这也是在对数据进行初步整理时经常需要用到的内容，然后得出不同分组的数据情况及该数据简单的描述性统计情况。

例如，统计不同年龄层次及不同收入的人群，花费在买衣服上的费用。在这个子菜单下，有以下可供使用的数据处理过程。

❑ 代码本 可以直接分析各个变量基本描

述性统计情况，例如对于名称变量，可以给出各个分类的计数。对于度量变量，则可以给出该变量的均值和标准差等集中和离散趋势。

- ❑ 在线分析处理——OLAP 过程 可以在指定的范围内对一个或多个变量进行描述性统计分析，并可以呈现不同分组数据的情况。
- ❑ 个案汇总 可以对指定数量的个案情况进行汇总分析，并在同一张表格内呈现对多个变量的组合分析情况。
- ❑ 按行/列汇总 这两个过程类似，均可以对输出的表格进行更为精细的定义。

5.1.2 代码本的应用

代码本功能允许用户选择一个或多个变量和/或多重响应集。根据实际需要，用户可以使用代码本进行以下操作。

❑ 控制显示的变量信息。

- ❑ 控制显示的统计（或排除所有摘要统计）。
- ❑ 控制变量和多重响应集显示的顺序。
- ❑ 更改源列表中任何变量的测量级别以更改显示的摘要统计。

5.1.3 代码本的操作步骤

在做任何分析之前，首先要对变量的属性有所了解，特别是有些分析过程只能针对特定属性的变量进行分析。了解变量的属性情况非常重要。本例使用代码本的方式分析查看“学生成绩”中各个变量的属性情况。

(1) 打开学生成绩.sav 文件，执行【分析】|【报告】|【代码本】命令，如图 5-1 所示。

(2) 在弹出的【代码本】对话框中，选择【学号】变量，按 Shift 键，再选择【第二次测试】变量。然后单击【代码本变量】中的【添加】按钮，如图 5-2 所示。

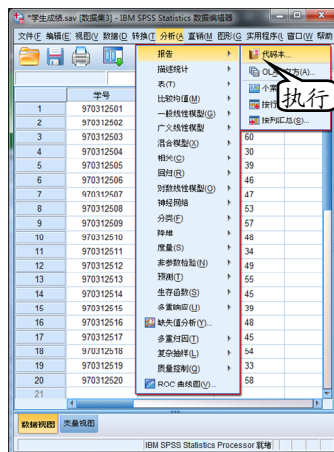


图 5-1 执行命令

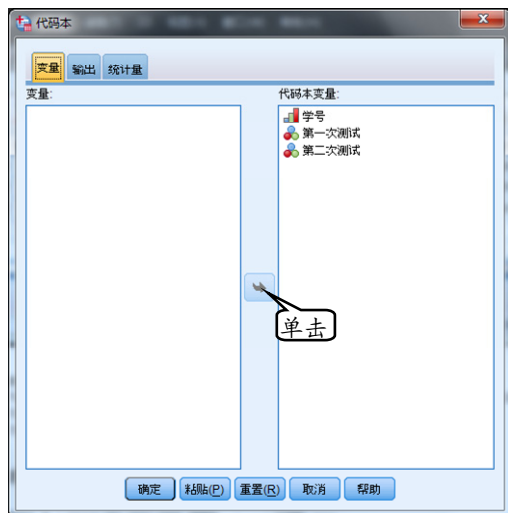


图 5-2 选择变量

(3) 在【代码本】对话框中,选择【输出】选项卡,选择【变量显示顺序】栏中的【降序】选项,如图 5-3 所示。

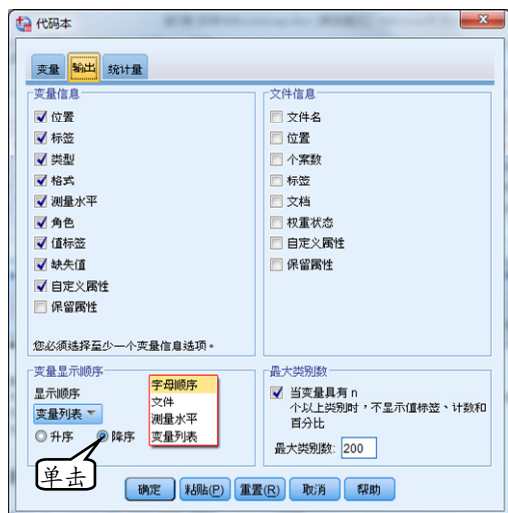


图 5-3 设置输出格式

在【代码本】对话框的【输出】选项中,包含四栏内容,各个选项组的设置和含义如下。

在【变量信息】栏中用于控制每个变量显示的信息。

- **位置** 代表变量在文件顺序中的位置的整数。
- **标签** 与变量或多重响应集相关联的描述性标签。
- **类型** 基本数据类型。可以是数值、字符串或多重响应集。

□ **格式** 变量的显示格式,如 B5、FM130.2 或 YEAR2000。在多重响应集中不可用。

□ **测量水平** 可能的值是名义、有序、刻度和未知。显示的值是 SPSS 字典中存储的测量水平,不受由更改【变量】选项卡上源变量列表中测量级别所指定的任何临时测量级别覆盖的影响。在多重响应集中不可用。

□ **角色** 某些对话框支持基于定义的角色预先选择分析变量的功能。

□ **值标签** 与特定数据值相关联的描述性标签。如果在【统计量】选项卡上选择了计数或百分比,则即使用户未在此处选择值标签,输出中仍包括定义的值标签。对于多二分集,【值标签】是集中基本变量的变量标签还是已计算值的标签,这取决于集的定义方式。

□ **缺失值** 用户定义的缺失值。如果在【统计量】选项卡上选择了计数或百分比,则即使用户未在此处选择缺失值,输出中仍包括定义的值标签。在多重响应集中不可用。

□ **自定义属性** 用户自定义的变量属性。对于任何与每个变量相关联的自定义变量属性,输出都包括名称和值。在多重响应集中不可用。

□ **保留属性** 保留系统变量属性。用户可以显示系统属性,但是用户不能改变这些属性。系统属性名称以美元符号“\$”开头。不包括名称以及对于任何与每个变量相关联的系统属性,输出都包括名称和值。在多重响应集中不可用。

在【文件信息】栏中用于显示文件的属性信息。

□ **文件名** SPSS 数据文件的名称。如果数据集从未以 SPSS Statistics 格式保存,那么就没有数据文件名。如果在“数据编辑器”窗口的标题栏中没有显示文件名,则活动数据集没有文件名。

□ **位置** SPSS 数据文件的目录(文件夹)位置。如果数据集从未以 SPSS Statistics

格式保存，那么就没有位置。

- ❑ **个案数** 活动数据集中的个案个数。这是个案的总数，包括任何由于过滤条件而从摘要统计中排除的个案。
- ❑ **标签** 由 FILE LABEL 命令定义的文件标签。
- ❑ **文档** 数据文件文档文本。
- ❑ **权重状态** 如果采用加权的方式，则显示加权变量的名称。
- ❑ **自定义属性** 用户自定义文件属性。使用 DATAFILE ATTRIBUTE 命令定义的数据文件属性。
- ❑ **保留属性** 保留系统数据文件属性。用户可以显示系统属性，但是用户不能改变这些属性。系统属性名称以美元符号“\$”开头。不包括名称以及对于任何系统数据文件属性，输出都包括名称和值。

在【变量显示顺序】栏中，用户可使用以下选项来控制变量和多重响应集的显示顺序。

- ❑ **依字母顺序排列** 依变量名称的字母顺序。
- ❑ **文件** 变量在数据集中的显示顺序（变量在数据编辑器中的显示顺序）。在升序方式中，多重响应集最后显示（在所有选定变量之后）。
- ❑ **测量水平** 按测量水平排序。这将创建四个排序组：名义、有序、刻度和未知。其中多重响应集被视为名义。当测量水平未明确设置时，数值变量的测量级别在第一次数据传输之前可能是“未知”，例如从外部源或新建变量读取的数据。
- ❑ **变量列表** 变量和多重响应集在【变量】选项卡上的选定变量列表中显示的顺序。
- ❑ **【最大类别数】** 如果输出包括每个唯一值的值标签、计数或百分比，那么如果值的数量超过指定的值，用户可以在表中不显示此信息。默认情况下，如果变量唯一值的数量超过 200，则不显示此信息。

(4) 在【代码本】对话框中，选择【统计量】选项卡，启用所示复选框，并单击【确定】按钮，

如图 5-4 所示。

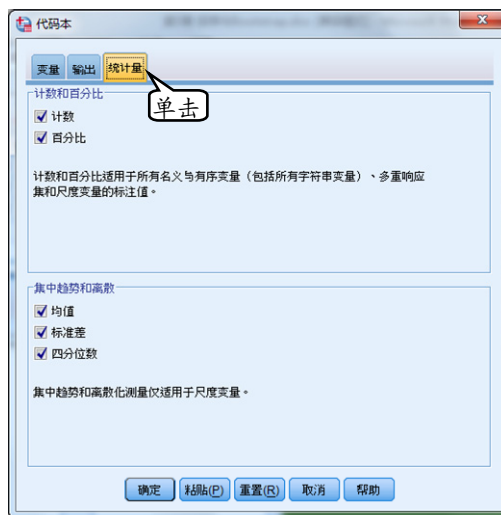


图 5-4 设置统计量

在【统计量】选项卡中，允许用户控制输出包括摘要统计量或不显示整个摘要统计量。该对话框中，还包括【计数和百分比】和【集中趋势和离散】两组选项栏。

在【计数和百分比】栏中，主要对名义和有序变量、多重响应集以及刻度变量的标签值。其各个选项的含义如下。

- ❑ **【计数】** 有变量的每个值（或值范围）个案的计数或个数。

- ❑ **【百分比】** 具有特定值的个案的百分比。

在【集中趋势和离散】栏中，主要对于刻度变量进行统计量。其中各选项的含义如下所示：

- ❑ **【均值】** 集中趋势的测量。算术平均，总和除以个案个数。

- ❑ **【标准差】** 对围绕均值的离差的测量。在正态分布中，68%的个案在均值的一倍标准差范围内，95%的个案在均值的两倍标准差范围内。例如，在正态分布中，如果平均年龄为 45，标准差为 10，则 95%的个案将处于 25~65 之间。

- ❑ **【四分位数】** 显示对应于第 25 个、第 50 个和第 75 个百分位的值。

提示

用户可以在【代码本】选项卡上【变量】列表中临时更改与变量相关联的测量级别（从而更改该变量显示的摘要统计量）。



(5) 对数据进行代码本分析, 则显示以下结果, 如图 5-5、图 5-6 和图 5-7 所示。

第二次测试			
标准属性	位置	值	计数
标签	<无>	3	
类型	数字		
格式	F22		
测量	标称		
角色	输入		
有效值	30	1	5.0%
	33	1	5.0%
	34	1	5.0%
	39	1	5.0%
	39	1	5.0%
	45	1	5.0%
	45	1	5.0%
	46	1	5.0%
	47	1	5.0%
	48	2	10.0%
	49	1	5.0%
	49	1	5.0%
	53	1	5.0%
	54	1	5.0%
	55	1	5.0%
	56	1	5.0%
	57	1	5.0%
	58	1	5.0%
	60	1	5.0%

图 5-5 【第二次测试】代码结果

第一次测试			
标准属性	位置	值	计数
标签	<无>	2	
类型	数字		
格式	F22		
测量	标称		
角色	输入		
有效值	23	1	5.0%
	26	1	5.0%
	27	2	10.0%
	33	1	5.0%
	33	1	5.0%
	36	1	5.0%
	36	1	5.0%
	37	1	5.0%
	41	2	10.0%
	42	1	5.0%
	43	1	5.0%
	44	2	10.0%
	44	1	5.0%
	45	1	5.0%
	46	1	5.0%
	60	1	5.0%
	64	1	5.0%

图 5-6 【第一次测试】代码结果

学号			
标准属性	位置	值	计数
标签	<无>	1	
类型	数字		
格式	F11		
测量	序数		
角色	输入		
有效值	970312501	1	5.0%
	970312502	1	5.0%
	970312503	1	5.0%
	970312504	1	5.0%
	970312505	1	5.0%
	970312506	1	5.0%
	970312507	1	5.0%
	970312508	1	5.0%
	970312509	1	5.0%
	970312510	1	5.0%
	970312511	1	5.0%
	970312512	1	5.0%
	970312513	1	5.0%
	970312514	1	5.0%
	970312515	1	5.0%
	970312516	1	5.0%
	970312517	1	5.0%
	970312518	1	5.0%
	970312519	1	5.0%
	970312520	1	5.0%

图 5-7 【学号】代码结果

图 5-8 各变量在【变量视图】的显示结果

在上述图中, 显示了各个变量的【标签属性】和【有效值】的结果。与图 5-8 不同的是, 虽然【标签属性】和【变量视图】下显示的一致, 但在结果输出窗口显示的【标签属性】等结果, 可以再次对数据进行编辑, 也可以根据实际需要打印输出【标签属性】结果。

5.2

在线分析处理

在线分析处理 (On-Line Analytical Processing, 简称 OLAP), 也可以用于操作储存在静态数据仓储 (Data Warehouse) 内广泛资源的技术。

OLAP 其透过快速、一致、交谈式的界面对同一数据提供各种不同的呈现方式, 供不同层面的使用者如分析师、经理及高阶主管等使用, 使其具备透析数据反应出来信息的能力。

5.2.1 了解 OLAP 分析

使用 OLAP 分析功能, 可以在指定的范围内对一个或多个变量进行描述性统计分析, 并可以呈现不同分组数据分别的情况。

OLAP 有三项要件: 一是, 动态多维度分析; 二是, 可执行复杂计算; 三是, 有时间导向处理能力。

在线分析处理过程具有快捷、灵活多样的交互方式, 能够根据用户的要求自由选择表格的报告方式和报告内容, 而且分析结果简洁明了, 便于理解。

对于多维数据资料, 在线分析处理过程可以从不同的角度给出分析报告。它还有一个特点在

于制作完成以后输出的表格是多层动态的,即在

已完成的表格上可以选择只显示某一类别的数据。

5.2.2 在线分析的应用

OLAP 过程主要用于计算一个或多个分类分组变量类别中,连续摘要变量的总和、均值和其他单变量统计量。在数据中,为每个分组变量的每个类别创建单独的层。

此外,OLAP 立方过程还允许用户进行以下操作。

- ❑ 选择不同的摘要统计量(单击统计量)。
- 在选择摘要统计量之前,必须选择一个

或多个分组变量。

- ❑ 计算变量对和由分组变量定义的组对之间的差(单击差分)。
- ❑ 创建自定义表标题(单击标题)。
- ❑ 隐藏小于指定整数的计数。隐藏值将显示为<N,其中N代表指定整数。指定的整数必须大于或等于2。

5.2.3 OLAP 立方分析的操作步骤

学生在两次测试中是否有进步,班级的平均分是否提高了?分析诸如此类的问题时,可以采用立方分析的方法制作动态的分析表格。使用 OLAP 立方制作的表格,不仅可以查看学生的整体成绩,还可以实时查看每个学生的成绩。

(1) 打开学生成绩.sav 文件,执行【分析】|【报表】|【OLAP 立方】命令,如图 5-9 所示。

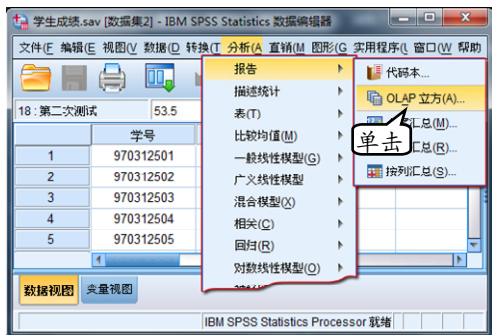


图 5-9 执行命令

(2) 在弹出的【OLAP 立方体】对话框中,选择【第一次测试】变量,单击【摘要变量】中的【添加】按钮;选择【第二次测试】变量,单击【摘要变量】中的【添加】按钮;选择【学号】变量,单击【分组变量】中的【添加】按钮。最后单击【确定】按钮,如图 5-10 所示。

用户可以在该对话框中设置【摘要变量】和【分组变量】,也可以设置立方体的【统计量】、【差分】和立方体报表的显示标题。各个选项具体的参数设置如下。

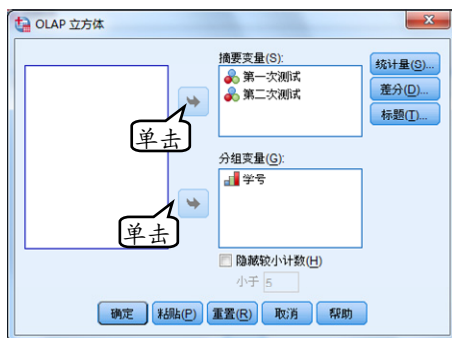


图 5-10 选择描述性分析的变量

- ❑ 【摘要变量】 用于选择一个或多个连续摘要变量。
- ❑ 【分组变量】 用于选择一个或多个分类分组变量。或者选择不同的摘要统计量(单击统计量)。在选择摘要统计量之前,必须选择一个或多个分组变量。
- ❑ 【隐藏较小的计数】 隐藏值将显示为<N,其中N代表指定整数。指定的整数必须大于或等于2。

(3) 在【OLAP 立方体】对话框中,单击【统计量】按钮,弹出【OLAP 立方: 统计量】对话框。然后,在该对话框中,选择【统计量】栏中的【最小值】选项,单击【单元格统计量】中的【添加】按钮;选择【统计量】栏中的【最大值】选项,单击【单元格统计量】中的【添加】按钮。最后,单击【继续】按钮,返回【OLAP 立方体】对话框,如图 5-11 所示。

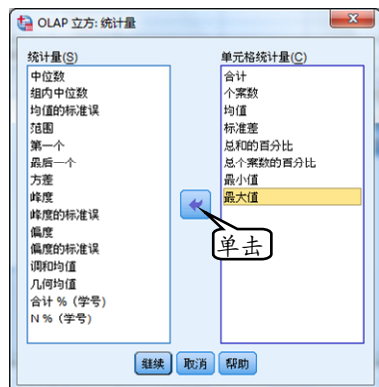


图 5-11 设置统计量

在【OLAP 立方: 统计量】对话框的【统计量】列表中, 包含了多种统计方法, 各统计方法的含义如下。

① 中位数

第 50 个百分位, 大于该值和小于该值的个案数各占一半。如果个案数为偶数, 则中位数是中间位置个案的平均数。中位数是集中趋势的测量, 但对于远离中心的值不敏感 (这与均值不同, 均值容易受到少数多个非常大或非常小的值的影响)。

② 组内中位数

针对编码到组中的数据计算的中位数。例如, 如果对于每个 30 年代的年龄数据的值都编码为 35, 40 年代的编码为 45, 依次类推, 则组内中位数是由已编码的数据计算得出的。

③ 均值的标准误

取自同一分布的样本与样本之间的均值之差的测量。它可以用来粗略地将观察到的均值与假设值进行比较 (即, 如果差与标准误的比值小于 -2 或大于 +2, 则可以断定两个值不同)。

- ❑ **范围** 即全距。数值变量最大值和最小值之间的差; 最大值减去最小值。
- ❑ **第一个** 显示在数据文件中遇到的第一个数据值。
- ❑ **最后一个** 显示在数据文件中遇到的最后一个数据值。
- ❑ **方差** 对围绕均值的离差的测量, 值等于与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。

④ 峰度

观察值聚集在中点周围的程度的测量。对于正态分布, 峰度统计量的值为 0。正峰度值表示相对于正态分布, 观察值在分布中心的聚集更多。同时, 尾部更薄, 直到分布极值。

在这一点, leptokurtic 分布的尾部比正态分布的尾部要厚。负峰度值表示相对于正态分布, 观察值聚集得少并且尾部较厚, 直到分布极值。在这一点, platykurtic 分布的尾部比正态分布的尾部要薄。

⑤ 峰度标准误

峰度与其标准误的比可用作正态性检验 (即如果比值小于 -2 或大于 +2, 可以拒绝正态性)。大的正峰度值表示分布的尾部比正态分布的尾部要长一些; 负峰度值表示比较短的尾部 (变为像框状的均匀分布尾部)。

⑥ 偏度

分布的不对称性度量。正态分布是对称的, 偏度值为 0。具有显著正偏度值的分布有很长的右尾。具有显著的负偏度的分布有很长的左尾。作为一个指导, 当偏度值超过标准误的两倍时, 则认为不具有对称性。

❑ **偏度标准误** 偏度与其标准误的比可以用作正态性检验 (即, 如果比值小于 -2 或大于 +2, 就可以拒绝正态性)。大的正偏度值表示长右尾; 极负值表示长左尾。

❑ **调和均值** 在组中的样本大小不相等的情况下用来估计平均组大小。调和均值是样本总数除以样本大小的倒数总和。

❑ **几何均值** 数据值的乘积的 n 次根, 其中 n 代表个案数目。

❑ **合计%** 每个类别中的总和的百分比。

❑ **N%** 每个类别中的个案总数的百分比。

❑ **合计** 所有带有非缺失值的个案的值的合计或总计。

❑ **个案数** 个案 (观察值或记录) 的数目。

❑ **均值** 集中趋势的测量。算术平均, 总和除以个案个数。

⑦ 标准差

对围绕均值的离差的测量。在正态分布中, 68% 的个案在均值的一倍标准差范围内, 95% 的

个案在均值的两倍标准差范围内。例如，在正态分布中，如果平均年龄为 45，标准差为 10，则 95% 的个案将处于 25~65 之间。

❑ **总和的百分比** 其他分组变量的类别内指定分组变量总和的百分比。如果只有一个分组变量，则此值与总和百分比相同。

❑ **总个案数的百分比** 其他分组变量的类别内指定分组变量的个案数的百分比。如果只有一个分组变量，则此值与个案总数百分比相同。

❑ **最小值** 数值变量的最小值。

❑ **最大值** 数值变量的最大值。

(4) 在【OLAP 立方体】对话框中，单击【差分】按钮。在弹出的【OLAP 立方：差分】对话框中，启用【算术差值】复选框，单击【继续】按钮，如图 5-12 所示。



图 5-12 设置差分

在【OLAP 立方：差分】对话框中，允许用户计算摘要变量间或由分组变量定义的组间的百分比和算术差。将计算【OLAP 立方：统计量】对话框中，选定的所有度量的差。

① 变量之间移动的差值

计算变量对之间的差值。每一对中第一个变量的摘要统计量值减去第二个变量（减去的变量）的摘要统计量值。对百分比差而言，减去的变量的摘要变量值用作分母。在可以指定变量间的差之前，用户必须在主对话框中至少选择两个

摘要变量。

② 个案组之间的差值

计算由分组变量定义的组对间的差。每一对中第一个类别的摘要统计量值减去第二个类别（减去的类别）的摘要统计量值。百分比差将减去的类别的摘要统计值作为分母。必须在主对话框中选择一个或多个分组变量，之后才能指定组之间的差分。

(5) 返回【OLAP 立方体】对话框，并单击【标题】按钮，弹出【OLAP：标题】对话框。然后，在【标题】文本框中，输入 OLAP 立方体分析结果文本，单击【继续】按钮，如图 5-13 所示。返回【OLAP 立方】对话框，单击【确定】按钮。

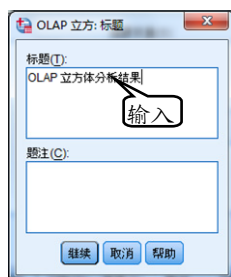


图 5-13 设置标题

提示

用户可以更改输出标题或添加在输出表下面显示的题注。在文本框中，用户也可以在想插入换行符的地方输入“\n”以控制标题和题注的换行。

(6) 用户通过对数据进行立方体分析后，则得出的结果如图 5-14 和图 5-15 所示。

案例处理摘要						
	案例				总计	
	已包含		已排除		N	百分比
	N	百分比	N	百分比	N	百分比
第一次测试 * 学号	20	100.0%	0	0.0%	20	100.0%
第二次测试 * 学号	20	100.0%	0	0.0%	20	100.0%

图 5-14 案例摘要

OLAP 立方体分析结果								
学号: 总计								
	合计	N	均值	标准差	总和的 %	合计 N 的 %	最小值	极大值
第一次测试	786	20	39.30	10.393	100.0%	100.0%	23	64
第二次测试	941	20	47.03	8.621	100.0%	100.0%	30	60

图 5-15 常用统计量

在图 5-14 中，列出了【第一次测试】和【第二次测试】两个变量的摘要结果。在图 5-15 中，

列出了【第一次测试】和【第二次测试】两个变量的常用统计量，包括均值、标准差、极大值、极小值等。

用户可以看到，第二次测试的均值明显高于第一次测试的均值，而且方差较小，说明在第二次测试中，大部分学生的成绩有了明显的提高。

(7) 在【输出窗口】的输出结果图上右击鼠标，执行【编辑内容】|【在阅读器中】命令。此时，学号选项变为可以选择的下拉列表，单击

下拉按钮，选择不同的学号内容，即可实时显示指定学号的学生的成绩。图 5-16 显示的是学号 970312515 的学生的成绩。

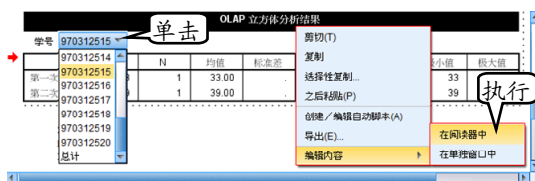


图 5-16 实时显示指定学号学生的成绩

5.3 个案汇总

个案汇总主要用来对指定数量的个案情况进行汇总分析，并能够在同一张表格内呈现对多个变量的组合分析情况。

个案汇总分析过程，可以将数据编辑窗口中的数据，按照指定要求在输出窗口里显示出来，以方便用户查看或用于打印。除此之外，个案汇总还可以对数据做一些简单的统计描述。

5.3.1 个案汇总的应用

个案汇总过程为一个或多个分组变量类别中的变量计算子组统计量。所有级别的分组变量要进行交叉制表。用户可以选择显示统计量的顺序，还可以将显示所有类别的每个变量的摘要统计，并设置每个类别中的数据值是否列出等。而对于大型数据集，可以选择只列出前 n 个个案。

根据实际需要，用户可以对个案汇总进行以下操作：

- 选择一个或多个分组变量以将数据划分

成子组。

- 更改输出标题，在输出下面添加题注，或排除具有缺失值的个案。
- 获取可选的统计量。
- 选择显示个案以列出每个子组中的个案。默认情况下，系统只列出文件中的前 100 个个案。用户可以升高或降低“将个案限制到前 n 个”中的 n 值，也可以取消选择该复选框以列出所有个案。

5.3.2 个案汇总的操作过程

在统计数据较多时，往往需要制作统计表格。本例使用【个案排序】的方法制作满足用户特定需求的统计表。要求统计出十个城市油价的差值以及涨价后十个城市油价的平均值。

(1) 打开油价调整.sav 文件，执行【分析】|【报告】|【个案汇总】命令，如图 5-17 所示。

(2) 在弹出的【摘要个案】对话框中，选择【基础价位】变量，按 Shift 键，选择【第六次调整】变量，单击【变量】中的【添加】按钮。然后，选择【城市】变量，单击【分组变量】中的【添加】按钮。最后，禁用【显示个案】复

选框，如图 5-18 所示。

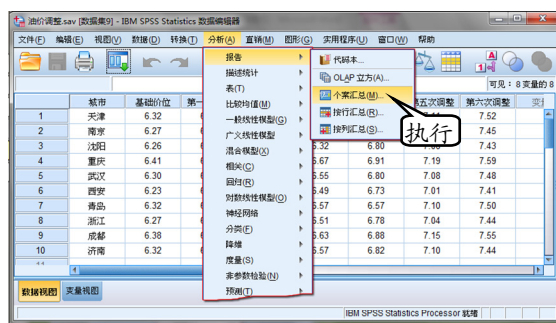


图 5-17 执行命令

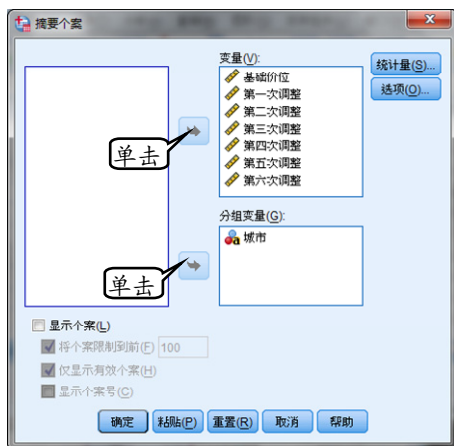


图 5-18 设置变量

在【摘要个案】对话框中，各选项设置如下。

- **【变量】** 用于选择对数据进行分类汇总的变量。
- **【分组变量】** 用于选择对数据进行汇总的分类变量。
- **【显示个案】** 用于选择显示个案以列出每个子组中的个案。默认情况下，系统只列出文件中的前 100 个个案。用户可以升高或降低【将个案限制到前 n 个】中的 n 值，也可以取消选择该项以列出所有个案。

(3) 单击【统计量】按钮，弹出【摘要报告：统计量】对话框。选择【个案数】变量，单击【单元格统计量】中的【删除】按钮；选择【范围】变量，单击【单元格统计量】中的【添加】按钮；选择【均值】变量，单击【单元格统计量】中的【添加】按钮。最后，单击【确定】按钮，返回【摘要个案】对话框，如图 5-19 所示。

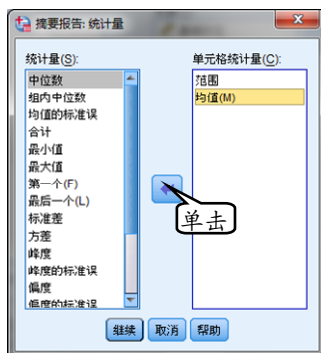


图 5-19 设置统计量

(4) 单击【选项】按钮，在弹出的【选项】对话框中，禁用【总计副标题】复选框，单击【继

续】按钮，如图 5-20 所示。返回【摘要个案】对话框，单击【确定】按钮。

提示

在【摘要报告：统计量】对话框中，其中统计量的含义与 OLAP 立方体中的统计量设置一样。

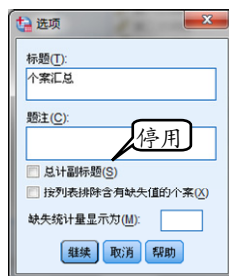


图 5-20 设置选项

摘要【选项】对话框中，允许用户更改输出的标题或者添加题注，并显示在输出表下面。

用户还可以选择显示或不显示总计的子标题，以及包含或排除在任何分析中使用的任何变量具有缺失值的个案。通常需要在输出中用句号或星号表示缺失个案。用于输入要在值缺失的情况下显示的字符、短语或代码，否则将不会对输出中的缺失个案应用特殊处理。

(5) 用户通过对数据进行个案汇总之后，则得出的计算结果如 5-21 和图 5-22 所示。

案例处理摘要							
		案例				总计	
		已包含		已排除			
		N	百分比	N	百分比		
基础价位 * 城市		10	100.0%	0	0.0%	10	100.0%
第一次调整 * 城市		10	100.0%	0	0.0%	10	100.0%
第二次调整 * 城市		10	100.0%	0	0.0%	10	100.0%
第三次调整 * 城市		10	100.0%	0	0.0%	10	100.0%
第四次调整 * 城市		10	100.0%	0	0.0%	10	100.0%
第五次调整 * 城市		10	100.0%	0	0.0%	10	100.0%
第六次调整 * 城市		10	100.0%	0	0.0%	10	100.0%

图 5-21 案例摘要结果

个案汇总								
城市		基础价位	第一次调整	第二次调整	第三次调整	第四次调整	第五次调整	第六次调整
成都	全距	.00	.00	.00	.00	.00	.00	.00
	均值	6.3800	6.6300	6.4500	6.6300	6.8800	7.1500	7.5500
济南	全距	.00	.00	.00	.00	.00	.00	.00
	均值	6.3200	6.5700	6.5700	6.5700	6.8200	7.1000	7.4400
南京	全距	.00	.00	.00	.00	.00	.00	.00
	均值	6.2700	6.5300	6.5300	6.5300	6.5300	7.0500	7.4500
青岛	全距	.00	.00	.00	.00	.00	.00	.00
	均值	6.3200	6.5700	6.5700	6.5700	6.5700	7.1000	7.5000
沈阳	全距	.00	.00	.00	.00	.00	.00	.00
	均值	6.2600	6.5200	6.1800	6.3200	6.8000	7.0800	7.4300
天津	全距	.00	.00	.00	.00	.00	.00	.00
	均值	6.3200	6.5800	6.5800	6.5800	6.8300	7.1100	7.5200
武汉	全距	.00	.00	.00	.00	.00	.00	.00
	均值	6.3000	6.5500	6.3500	6.5500	6.8000	7.0800	7.4800
西安	全距	.00	.00	.00	.00	.00	.00	.00
	均值	6.2300	6.4900	6.3000	6.4900	6.7300	7.0100	7.4100
浙江	全距	.00	.00	.00	.00	.00	.00	.00
	均值	6.2700	6.5300	6.3500	6.5100	6.7800	7.0400	7.4400
重庆	全距	.00	.00	.00	.00	.00	.00	.00
	均值	6.4100	6.6700	6.4800	6.6700	6.9100	7.1900	7.5900
全距	全距	.18	.18	.40	.35	.38	.18	.18
	均值	6.3080	6.5640	6.4360	6.5400	6.7650	7.0910	7.4810

图 5-22 个案汇总

在图 5-21【案例处理摘要】结果中,【已包含】值为 10,【已排除】值为 0,说明在此次统计所有值都参与了统计。

在图 5-22【个案汇总】中,列出了每次调

价后各城市的价格最高值与最低值的差异。同时,图中也列出了每次调价后各城市油价的平均值。观察最后一行的统计结果,每次调价后的油价平均值呈增涨的趋势。

5.4

按行汇总

如果用户通过“按行汇总”功能,则对数据进行汇总操作。则该功能的特点是为系统提供了更多的选项,以便对输出的表格进行更为精细的定义。另外,通过“按行汇总”功能所得到的结果,并不是表格格式,而是文本格式。用户可以随意地修改和编辑结果数据,并将其复制粘贴到其他文档中。

而【按行汇总】与上一节【个案汇总】操作一样。用户也可以在汇总时,对数据进行简单的统计描述,但【按行汇总】给出更为复杂的报告形式,输出格式的设置也更为详细。其次,用户对【按行汇总】操作的方法,与【按行汇总】操作方法基本相同。而两者之间的区别是不能列出原始数据,输出结果略有差异。在此仅以【按行汇总】为例,讲解相应的操作方法。

按行汇总的操作步骤

(1) 打开【促销效果.sav】文件,执行【分析】|【报告】|【按行汇总】命令,如图 5-23 所示。

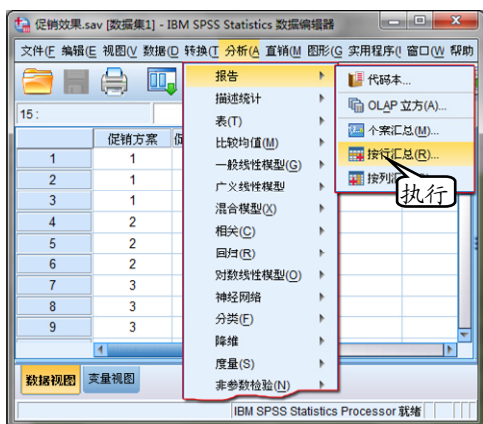


图 5-23 执行命令

(2) 弹出【报告:行摘要】对话框,选择【促销前销售量】变量,单击【数据列变量】中的【添加】按钮;选择【促销后销售量】变量,单击【数据列变量】中的【添加】按钮;选择【促销方案】变量,单击【终止列变量】中的【添加】按钮,如图 5-24 所示。

在【报告:行摘要】对话框中,各组成部分的含义如下。

□【数据列】栏 列出需要个案列表或摘要统计的报告变量,并控制数据列的显示格式。

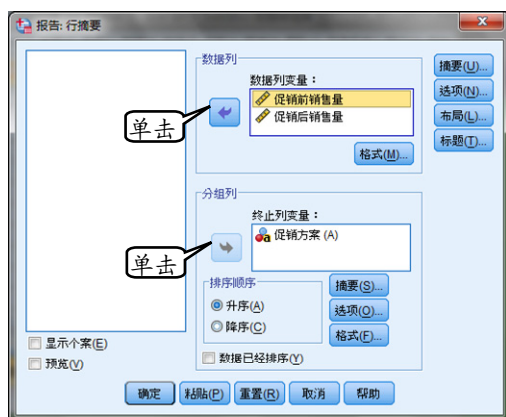


图 5-24 设置变量

□【分组列】栏 列出将报告划分为各个组的可选中变量,并控制摘要统计以及中断列的显示格式。

对于多个中断变量而言,在列表中先前中断变量的类别中,每个中断变量的每个类别均有一个单独的组。中断变量应是离散型分类变量,它们将个案划分为有限个有具体含义的类别。每个中断变量的个别值排好顺序之后,作为单独一列

出现在所有数据列的左边。

□ **【显示个案】复选框** 启用复选框为每个个案显示数据列变量的实际值（或值标签）。这样会生成列表报告，并且该报告可能比摘要报告要长得多。

□ **【预览】复选框** 只显示报告的第一页。此选项对于预览报告的格式不处理，但整个报告很有用。

□ **【数据已排序】栏** 对于具有中断变量的报告，在生成报告之前必须按中断变量值对数据文件排序。如果数据文件已经按中断变量的值排序，则可以通过选择此选项来节省处理时间。在运行预览报告之后，此选项非常有用。

(3) 单击【报告：行摘要】对话框中的【摘要】按钮，在弹出的【报告：最终摘要行】对话框中，启用【值的和】复选框。单击【继续】按钮，返回【报告：行摘要】对话框，如图 5-25 所示。

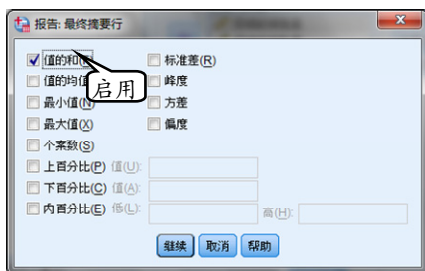


图 5-25 设置摘要

在【报告：最终摘要行】对话框中，包括摘要统计量有值的和、值的均值、最小值、最大值、个案数、标准差、峰度、方差、偏度等，具体的参数设置与【OLAP 立方体】中的参数设置一样。

而【上百分比】表示在结果中输出当前分组变量大于指定值的个案比例，在其后的输入框中指定临界值。【下百分比】与【内百分比】的含义与此类似。

(4) 单击【报告：行摘要】对话框中的【选项】按钮，在弹出的【报告：选项】对话框中，启用【按列表排除含有缺失值的个案】复选框。单击【继续】按钮，返回【报告：行摘要】对话框，如图 5-26 所示。

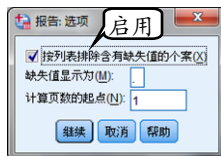


图 5-26 设置选项

在【报告：选项】对话框中，用于报告中缺失值的处理。其中，各选项的含义如下。

□ **【按列表排除含有缺失值的个案】复选框** 启用该复选框，则从报告中排除任何报告变量具有缺失值的所有个案。

□ **【缺失值显示为】** 允许用户指定在数据文件中代表缺失值的符号。该符号只能是一个字符，用于表示系统缺失值和用户缺失值。

□ **【计算页数的起点】** 允许用户为报告的第一页指定页码。

(5) 单击【报告：行摘要】对话框中的【布局】按钮，在弹出的【报告：布局】对话框中，选择【对齐】列表中的【居中】选项。单击【继续】按钮，返回【报告：行摘要】对话框，如图 5-27 所示。

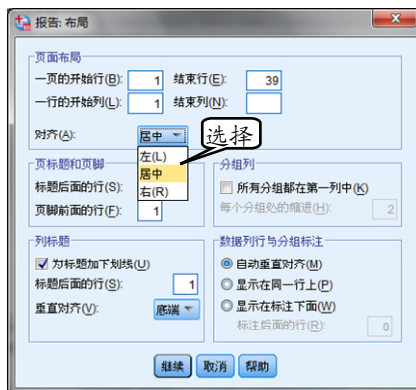


图 5-27 设置布局

在【报告：布局】对话框中，用于控制每个报告页的宽度和长度、页面上报告的放置以及空白行和标签的插入。

□ **【页面布局】** 控制以行数（顶部和底部）和字符数（左和右）表示的页边距以及页边距内的报告对齐方式。

□ **【页标题和页脚】** 控制将页标题和页脚与报告正文分隔的行数。

□ **【分组列】** 控制分组列的显示。如果指定了多个分组变量，则它们可以位于分开的列或位于第一列。将所有分组变量放置在第一列会生成较窄的报告。

□ **【列标题】** 控制列标题的显示，包括标题下划线、标题和报告正文之间的间隔以及列标题的垂直对齐方式。

□ **【数据列行与分组标注】** 控制数据列信息（数据值和/或摘要统计）相对于每个中断类别开始处的中断标签的放置。数据列信息的第一行可以开始于与中断类别标签相同的行，也可以从中断类别标签之后指定的行数处开始。（不适用于列摘要报告。）

(6) 单击 **【报告：行摘要】** 对话框中的 **【标题】** 按钮，在弹出的 **【报告：标题】** 对话框中，选择 **【特殊变量】** 栏中的 **DATE** 函数；再单击 **【第 1 页标题行，共 1 行】** 栏中 **【左】** 的 **【添加】** 按钮 。单击 **【继续】** 按钮，返回 **【报告：行摘要】** 对话框，如图 5-28 所示。



图 5-28 设置标题

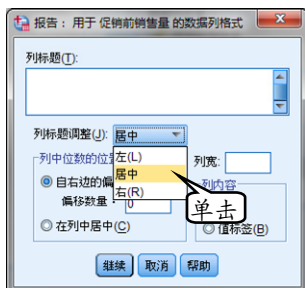
在 **【报告：标题】** 对话框，用于控制报告标题和页脚的内容和放置。其最多可指定 10 行页面标题和 10 行页脚，并带有可控制各行行为左对齐、居中和右对齐的组件。

如果将变量插入到标题或页脚中，则变量的当前值标签或值会显示在标题或页脚中。在标题中，会显示与页开始处的变量的值对应的值标签。在页脚中，会显示与页结束处的变量的值对

应的值标签。如果没有值标签，则会显示实际值。

在 **【特殊变量】** 列表框中包含 **DATE** 和 **PAGE** 变量，允许用户将当前日期或页码插入到报告页眉或页脚的任意一行。如果数据文件包含名为 **DATE** 或 **PAGE** 的变量，则不能在报告标题或页脚中使用这些变量。

(7) 选择 **【数据列变量】** 中的 **【促销前销售量】** 变量，单击 **【数据列】** 中的 **【格式】** 按钮，弹出 **【报告：用于促销前销售量的数据列格式】** 对话框。单击 **【列标题调整】** 的下拉按钮，选择 **【居中】** 选项。单击 **【继续】** 按钮，返回 **【报告：行摘要】** 对话框，如图 5-29 所示。


图 5-29 设置 **【数据列】** 变量的格式

(8) 用同样的方法，用户可以设置 **【促销后销售量】** 变量的 **【列标题调整】** 的格式为居中。

在 **【报告：用于促销前销售量的数据列格式】** 对话框中，用于控制列标题、列宽、文本对齐方式以及数据值或值标签的显示。其中，各组成部分包含如下。

□ **【列标题】** 对于选定的变量，控制列标题。长标题在列中自动换行。使用 **Enter** 键可手工在想要标题换行的地方插入换行符。

□ **【列中位数的位置】** 对于选定的变量，控制列中数据值或值标签的对齐方式。值或标签的对齐方式不影响列标题的对齐方式。用户可以使列内容缩进指定的字符数或者使内容居中。

□ **【列内容】** 对于选定变量，控制数据值或定义的值标签的显示。对于没有定义的值标签的任何值，始终显示数据值。（不适用于列摘要报告中的数据列。）

(9) 选择 **【终止列变量】** 中的 **【促销方案】** 变量，单击 **【分组列】** 中的 **【摘要】** 按钮。在弹

出的【报告：摘要行促销方案】对话框中，启用【值的均值】复选框，单击【继续】按钮，返回【报告：行摘要】对话框，如图 5-30 所示。

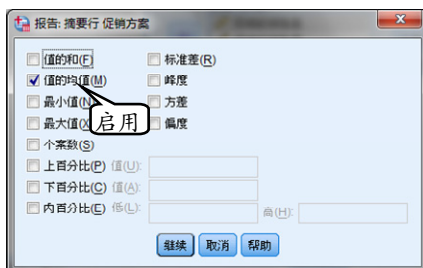


图 5-30 设置【分组列】变量的摘要

在【摘要行】对话框中，用于控制中断组的摘要统计以及整个报告的摘要统计的显示。可用的摘要统计量有值的和、值的均值、最小值、最大值、个案号、指定值以上或以下的个案百分比、指定值范围内的个案百分比、标准差、峰度、方差和偏度等。各选项的含义与【OLAP 立方体】的一样，在此不再赘述。

(10) 单击【分组列】中的【选项】按钮，在弹出的【报告：中断选项】对话框中，选择【分页之前跳过的行数】选项。然后，单击【继续】按钮，返回【报告：行摘要】对话框，如图 5-31 所示。

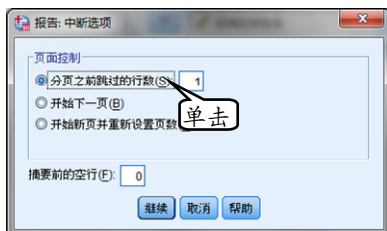


图 5-31 设置【分组列】变量的选项

在【报告：中断选项】对话框中，用于控制中断类别信息的间距和页编号。其中，各选项的含义如下。

- 【页面控制】为选定中断变量的类别控制间距和分页。用户可以在中断类别之间指定多个空白行，或者在新的页面上开始每个中断类别。
- 【摘要前的空行】控制中断类别标签之间，或者数据与摘要统计之间的空白行数。对于既包含个别个案列表又包含中断类别的摘要统计的组合报告，这非常

有用；在这些报告中，用户可以在个案列表之间和摘要统计之间插入空格。

(11) 单击【分组列】中的【格式】按钮，弹出【报告：中断格式】对话框。单击【列标题调整】的下拉按钮，选择【居中】选项。然后，单击【继续】按钮，如图 5-32 所示。返回【报告：行摘要】对话框，单击【继续】按钮。

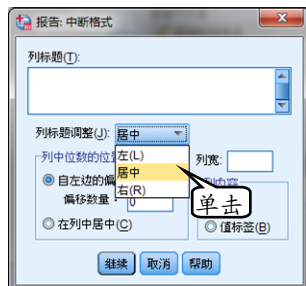


图 5-32 设置【分组列】变量的格式

在【报告：中断格式】对话框中，用于控制列标题、列宽、文本对齐方式以及数据值或值标签的显示。

- 【列标题】对于选定的变量，控制列标题。长标题在列中自动换行。使用 Enter 键可手工在想要标题换行的地方插入换行符。
- 【列中位数的位置】对于选定的变量，控制列中数据值或值标签的对齐方式。值或标签的对齐方式不影响列标题的对齐方式。用户可以使列内容缩进指定的字符数或者使内容居中。
- 【列内容】对于选定变量，控制数据值或定义的值标签的显示。对于没有定义的值标签的任何值，始终显示数据值。
(不适用于列摘要报告中的数据列。)

(12) 用户通过对数据进行按行汇总后，则得出的结果如图 5-33 所示。

24 May 13		页	1
促销方案	促销前销售量	促销后销售量	
1			
均值(M)	83	88	
2			
均值(M)	86	89	
3			
均值(M)	98	111	
值的和	802	866	

图 5-33 按行汇总的报告结果

在图 5-33 中,按照用户的设置,生成了按行汇总的报告。在该报告中列出了不同促销方案在促销前和促销后的均值汇总和各种方案的销

量总和。同时,在该报告中,列出了报告生成时间和页码等标题信息。用户可以根据需要直接打印输出报表。

5.5

Bootstrap 简介

Bootstrap 广泛应用于频数分析、描述性分析、假设检验、回归分析等过程中。此功能提供了近年来比较流行的一种非参数统计方法。该方法不需要对总体分布作假设或事先推导估计量的解析式。因此,在分布假设不满足或者解析式很难推导时,就可以尝试使用 Bootstrap 法。它开辟了解决问题的创新思路。

5.5.1 Bootstrap 计算原理

所谓 Bootstrap 方法,也称为自助方法,是一种通过估计统计量方差,进而进行区间估计的统计方法。

在对小样本数据进行统计分析时,通过 Bootstrap 方法效果很好,并且提供了另一种寻找样本分布的办法。

Bootstrap 方法的基本思想是采用重抽样技术,在原始样本中重复 N 次抽取给定数量的样本。根据抽取的 N 个样本计算出 N 个给定的统

计量,再计算 N 个统计量的样本方差。而通过方差的估计就可以构造相应的置信区间、导出标准误的估计值等。尤其,在假设检验过程中,当参数估计不满足相应的假设条件时,或者解析式较为复杂时,都可以考虑使用 Bootstrap 方法来替代原先的分析方法。

关于非参数检验的方法,在本书后面内容中有详细的介绍。在此,只针对 Bootstrap 方法的选项设置及使用方法进行讲解。

5.5.2 Bootstrap 的使用方法

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

Bootstrap 方法还可用于构建假设检验。当参数估计方法的假设存在疑问(例如,异方差残差拟合较小样本的回归模型),参数推论无法执行或需要非常复杂的标准误计算公式(如为中位数、四分位数和其他百分位数计算置信区间)时,Bootstrap 是最好的替代选项。

Bootstrap 在实际操作中主要用于以下内容。

- ❑ 运用重抽样(允许重复)的方法从原始样本中抽取一定数据的样本。
- ❑ 根据抽取样本的数据计算统计量 T (次)。
- ❑ 重复抽样 N 次($N > 1000$),得到 T 的 N 个样本。
- ❑ 计算 T 的样本方差,得出 T 的方差估计值。

5.5.3 Bootstrap 的操作步骤

在对某项指标进行分析时,往往不能只停留在某一个具体的数字上,尤其是处理较为庞大的数据。因为,一个微小的误差经过若干次放大后会造成非常大的影响,所以需要更精确的统计数字或者在某一置信区间下统计数字的范围。

在本实例中,使用 Bootstrap 重复抽样的方法,统计客户的平均年龄。但是,需要注意,在

启用 Bootstrap 选项后,再次做其他分析时会默认打开此项。如果不需要使用 Bootstrap,用户在做其他分析时首先需要关闭此项。

(1) 打开电信客户消费情况.sav 文件,执行【分析】|【描述统计】|【描述】命令,如图 5-34 所示。

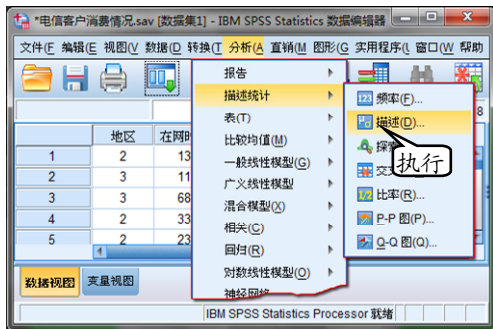


图 5-34 执行命令

(2) 在弹出的【描述性】对话框中, 选择【年龄】变量, 单击【变量】中的【添加】按钮, 如图 5-35 所示。

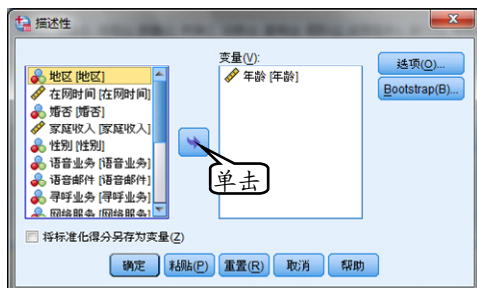


图 5-35 添加变量

(3) 单击【选项】按钮, 弹出【描述: 选项】对话框。然后, 在该对话框中, 禁用【离散】栏中的所有复选框, 单击【继续】按钮, 如图 5-36 所示。返回【描述性】对话框, 单击【确定】按钮。



图 5-36 设置选项

(4) 通过对数据进行描述分析时, 在不启用 Bootstrap 的情况下, 得出的结果如图 5-37 所示。

描述统计量		
	N	均值
年龄	1000	41.68
有效的 N (列表状态)	1000	

图 5-37 未启用 Bootstrap 时平均值计算结果

(5) 而启用 Bootstrap 选项, 计算【年龄】变量的平均值。例如, 再次执行【分析】|【描述统计】|【描述】命令, 弹出【描述性】对话框。然后, 单击【Bootstrap】按钮, 弹出【Bootstrap】对话框。

(6) 在【Bootstrap】对话框中, 启用【执行 bootstrap】复选框, 在【置信区间】栏中, 输入【水平】为 99, 单击【继续】按钮, 如图 5-38 所示。返回【描述性】对话框, 单击【确定】按钮。



图 5-38 设置 Bootstrap 选项

(7) 启用 Bootstrap 选项后, 当对数据进行描述分析 (时间较长) 后, 得出的结果如图 5-39 和图 5-40 所示。

Bootstrap 指定		
采样方法	简单箱图	
样本数		1000
置信区间度		99.0%
置信区间类型	百分位	

图 5-39 Bootstrap 运行参数

描述统计量						
		Bootstrap ^a				
		统计量	偏差	标准误	99% 置信区间	
年龄		N	1000	0	1000	1000
		均值	41.68	.00	40.66	42.75
有效的 N (列表状态)		N	1000	0	1000	1000

a. 除非另行注明, bootstrap 结果将基于 1000 bootstrap samples

图 5-40 启用 Bootstrap 选项后平均值计算结果

比较启用 Bootstrap 选项前后的结果可以发现,统计结果中的均值一致,都是 41.68。但是应用 Bootstrap 后给出了更为详细的结果,从图 5-39 和图 5-40 中得知,此次分析结果的置信区间是 99%,1000 次重复抽样均有效,标准误差较小,且没有偏差。最终【年龄】变量的在 99%

置信区间下的取值范围为 40.66 ~ 42.75。

提示

在统计分析中,误差是测量值与真值之间的差值。一般用误差衡量测量结果的准确度,用偏差衡量测量结果的精密度;误差是以真实值为标准,偏差是以多次测量结果的平均值为标准。

5.6

实验指导：制作摘要报表

在对数据分析之间,了解变量的属性信息十分必要,尤其是在选择分析方法时。例如,本例使用报表的形式分析数据的属性信息,并对数据的趋势进行估计,以方便为下一步分析选择合适的方法。

1. 实验目的:

- ☐ 代码分析
- ☐ 摘要结果

2. 代码分析

(1) 打开药物测试.sav 文件,执行【分析】|【报告】|【代码本】命令,如图 5-41 所示。

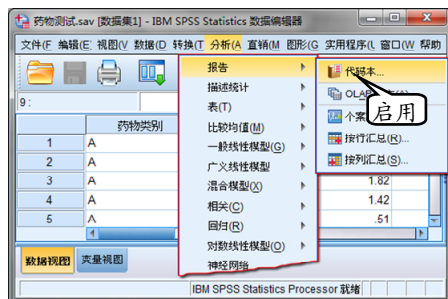


图 5-41 执行命令

(2) 弹出【代码本】对话框,选择【药物类别】变量,按 Shift 键,选择【半衰期】变量,单击【确定】按钮,如图 5-42 所示。

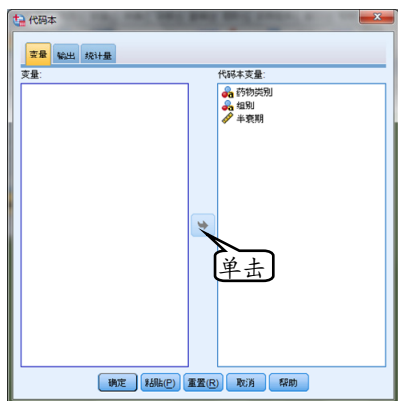


图 5-42 设置变量

(3) 通过对数据进行代码分析后,得出的结果如图 5-43,图 5-44 和图 5-45 所示。

药物类别			
标准属性	位置	值	计数
标签		<无>	1
类型		字符串	
格式		A255	
测量		标称	
角色		输入	
自定义属性	\$ODBC.Name	药物类别	
	\$ODBC.Table	Sheet1\$	
	\$ODBC.Size	255	
	\$ODBC.Type	12	
有效值	A		5
	B		5
	C		5
			33.3%

图 5-43 药物类别代码信息

组别			
标准属性	位置	值	计数
标签		<无>	2
类型		字符串	
格式		A8	
测量		标称	
角色		输入	
自定义属性	\$ODBC.Name	组别	
	\$ODBC.Table	Sheet1\$	
	\$ODBC.Size	15	
	\$ODBC.Type	8	
有效值	1		3
	2		3
	3		3
	4		3
	5		3
			20.0%

图 5-44 组别代码信息

半衰期			
标准属性	位置	值	计数
标签		<无>	3
类型		数字	
格式		F8.2	
测量		标度	
角色		输入	
自定义属性	\$ODBC.Name	半衰期	
	\$ODBC.Table	Sheet1\$	
	\$ODBC.Size	15	
	\$ODBC.Type	8	
N	有效		15
	缺失		0
集中趋势和离散	均值		2.6867
	标准差		1.67731
	百分位 25		1.3000
	百分位 50		2.1100
	百分位 75		4.8100

图 5-45 半衰期代码信息

代码结果中,列出了各个变量的基本属性信息。对于数据型【半衰期】变量,列出了较为详细的属性信息,包括均值、标准差、百分位等。

3. 摘要结果

(1) 在窗口中,用户可以执行【分析】|【报告】|【OLAP 立方】命令,如图 5-46 所示。

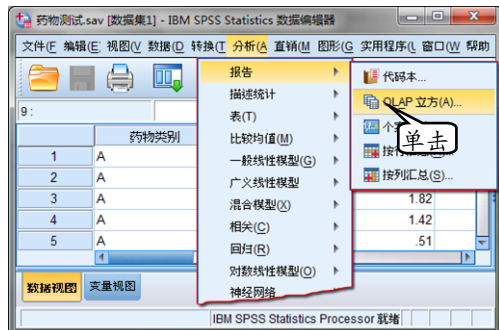


图 5-46 执行命令

(2) 弹出【OLAP 立方体】对话框,选择【半衰期】变量,单击【摘要变量】中的【添加】按钮;选择【组别】变量,单击【分组变量】中的【添加】按钮;选择【药物类别】变量,单击【分组变量】中的【添加】按钮。单击【确定】按钮,执行命令,如图 5-47 所示。

(3) 通过对数据,进行 OLAP 立方体分析

后,得出的结果如图 5-48 和图 5-49 所示。

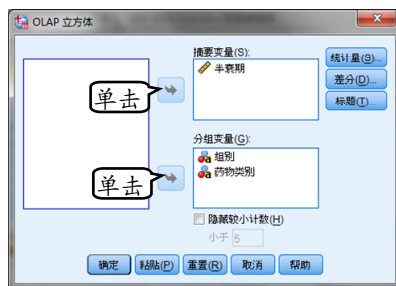


图 5-47 设置变量

案例处理摘要					
	已包含		已排除		总计
	N	百分比	N	百分比	N 百分比
半衰期 * 组别 * 药物类别	15	100.0%	0	0.0%	15 100.0%

图 5-48 案例摘要结果

OLAP 立方体						
组别: 总计						
	合计	N	均值	标准差	总和的 %	合计 N 的 %
半衰期	40.30	15	2.6867	1.67731	100.0%	100.0%

图 5-49 立方体结果

在 OLAP 立方体结果中,得出了关于数值型【半衰期】变量更为详细的信息。如果用户需要,可以在【OLAP 立方体】对话框中,单击【统计量】按钮,设置需要的统计信息。

5.7 实验指导: 应用重复抽样精确分析致病因素

对于较多或较为详细的统计数据,不同的抽样方法会产生不同的分析结果。应用 SPSS 内置的 Bootstrap 方法进行重复抽样,可以精确的控制抽样,如置信区间、抽样次数等。本例以病人丢失情况数据为例,分析患胆固醇和肥胖症之间是否存在某种关系。

1. 实验目的:

- ☐ 描述性分析
- ☐ 相关分析

2. 描述性分析

(1) 打开病人丢失情况.sav 文件,执行【分析】|【描述统计】|【描述】命令,如图 5-50 所示。

(2) 在弹出的【描述性】对话框中,选择【胆固醇】变量,单击【变量】中的【添加】按钮;选择【肥胖症】变量,单击【变量】中的【添加】按钮,如图 5-51 所示。

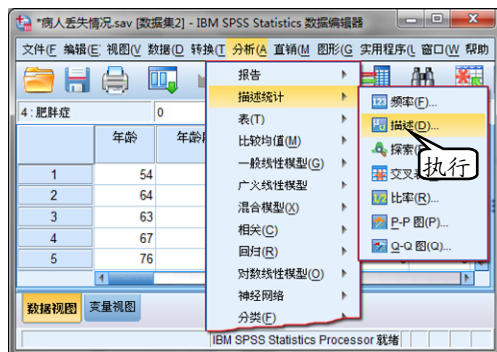


图 5-50 执行命令

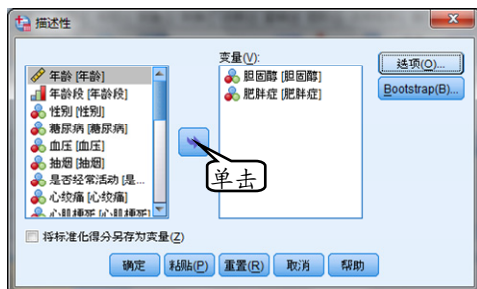


图 5-51 设置变量

(3) 单击【选项】按钮，弹出【描述：选项】对话框。然后，在对话框中，启用【均值】和【合计】复选框，再启用【离散】栏中所有的复选框，单击【继续】按钮，如图 5-52 所示。



图 5-52 设置选项

(4) 单击【Bootstrap】按钮，弹出【Bootstrap】对话框。然后，在弹出的对话框中，启用【执行 Bootstrap】复选框，在【置信区间】中【水平】文本框中，输入 99，单击【继续】按钮，如图 5-53 所示。返回【描述性】对话框，单击【确定】按钮。



图 5-53 设置 Bootstrap 选项

(5) 通过在描述分析中，使用 Bootstrap 方法对数据进行分析后，得出的结果如图 5-54 和图 5-55 所示。

Bootstrap 指定	
采样方法	简单随机
样本数	1000
置信区间度	99.0%
置信区间类型	百分位

图 5-54 Bootstrap 抽样设置

		统计量		标准误		Bootstrap ^a	
						99% 置信区间	
						下限	上限
胆固醇	N	100				100	100
	全距	1					
	最小值	0					
	极大值	1					
	和	41					
	均值	.41	.049	.00	.05	.29	.55
	标准差	.494		-.003	.010	.456	.503
肥胖症	N	100				100	100
	全距	1					
	最小值	0					
	极大值	1					
	和	20					
	均值	.20	.040	.00	.04	.10	.31
	标准差	.402		-.003	.030	.302	.465
有效的 N (列表状态)	N	100				100	100

a. 除非另行注明，bootstrap 结果将基于 1000 bootstrap samples

图 5-55 描述性结果

图 5-54 是抽样方法报告，图 5-55 是描述性统计结果。观察图中数据可以发现，【胆固醇】变量和【肥胖症】变量的误差都较小，偏差为 0。用户可以这样认为，这两组数据在使用 Bootstrap 方法抽样时准确度是较高的，从而认为分析结果是可靠的。

进一步观察两组数据的均值，相比其他统计量，两组数据的均值是比较接近的，说明两组数据可能存在某种关系。所以在下面的分析过程中，重点分析这两个变量的相关关系。

3. 相关分析

(1) 在窗口中，用户可以执行【分析】|【相关】|【双变量】命令，如图 5-56 所示。

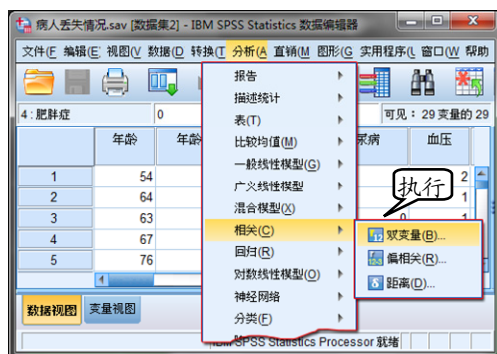


图 5-56 执行命令

(2) 在弹出的【双变量相关】对话框中, 选择【胆固醇】变量, 单击【变量】中的【添加】按钮; 选择【肥胖症】变量, 单击【变量】中的【添加】按钮。然后, 启用【相关系数】栏中的【Kendall 的 tau-b】和【Spearman】复选框, 如图 5-57 所示。



图 5-57 设置变量

(3) 单击【Bootstrap】按钮, 并在弹出的【Bootstrap】对话框中, 启用【执行 Bootstrap】复选框; 在【置信区间】中【水平】文本框中, 输入 99, 单击【继续】按钮, 如图 5-58 所示。返回【描述性】对话框, 单击【确定】按钮。



图 5-58 设置 Bootstrap 选项

(4) 通过对数据进行相关分析后, 得出的结果如图 5-59 和图 5-60 所示。

Bootstrap 指定	
采样方法	简单箱图
样本数	1000
置信区间度	99.0%
置信区间类型	百分位

图 5-59 Bootstrap 抽样设置

相关系数			
Kendall 的 tau-b	胆固醇	相关系数	1.000
		Sig. (双侧)	.266
		N	100
		Bootstrap ^c	
		偏差	.002
		标准误差	.093
		99% 置信区间	
		下限	1.000
		上限	1.000
	肥胖症	相关系数	-.112
Spearman 的 rho	胆固醇	相关系数	1.000
		Sig. (双侧)	.268
		N	100
		Bootstrap ^c	
		偏差	.002
		标准误差	.093
		99% 置信区间	
		下限	1.000
		上限	1.000
	肥胖症	相关系数	-.112

*. 在置信度 (双侧) 为 0.05 时, 相关性是显著的。

**. 在置信度 (双侧) 为 0.01 时, 相关性是显著的。

c. 除非另行注明, bootstrap 结果将基于 1000 bootstrap samples

图 5-60 相关系数结果

在相关系数结果中, 无论采用【Kendall 的 tau-b】方法, 还是采用【Spearman 的 rho】方法, Sig 值 (双侧) 都大于 0.05, 说明患【胆固醇】的患者与【肥胖症】之间没有必然的关系。

5.8

思考与练习

一、填空题

- _____是指采用表格、图表等形式动态地呈现数据的统计情况。
- OLAP 是_____ (Online Analytical Processing) 的缩写, 使用该功能可以在指定的范围

内对一个或多个变量进行_____统计分析,

- _____主要用来对指定数量的个案情况进行汇总分析。
- Bootstrap 法, 是一种通过估计统计量_____进而进行_____的统计方法。

二、选择题

1. 回归分析的主要目的不包括_____?
 A. 代码本
 B. 描述性
 C. 个案汇总
 D. 按行汇总
2. 下列关于代码本的描述中，_____是不正确的?
 A. 代码本功能是 SPSS 新增的一个选项
 B. 代码本使用方便，简明清晰
 C. 代码本能统计所有类型的变量
 D. 代码本只能针对特定属性的变量进行分析
3. 下列关于 OLAP 的说法不正确的是_____?
 A. OLAP 使用快捷、灵活多样

- B. OLAP 的分析结果简洁明了，便于理解
- C. OLAP 对于多维数据资料，可以从不同的角度给出分析报告
- D. OLAP 输出的表格是静态的
4. 下列关于按行汇总的说法中，哪个是不准确的?
 A. 按行汇总可以对表格进行更为精细的定义
 B. 按行汇总使用该菜单所得到的结果并非表格格式的
 C. 用户可以修改和编辑按行汇总的结果
 D. 按行汇总与按行汇总基本相同

三、简答题

1. 简述不同的报告过程?
2. 简述 Bootstrap 的使用方法?
3. 简述 OLAP 分析的过程?