

第 3 章

SPSS基本统计分析

在SPSS中对数据进行统计分析时，常常以各种报表的形式呈现统计结果，以获得变量的相关信息。SPSS统计报表可以根据用户的需求，以各种列表的形式输出数据的统计量。本章将对各种统计报表以及统计分析方法进行详细介绍。

3.1 统计报告

常用的统计报表包括个案摘要报告、行形式摘要报告、列形式摘要报告和在线分析处理报告。

3.1.1 个案摘要报告

个案摘要报告可以将数据按用户的要求进行整理和分析, 对一个或者多个分组类别中的变量计算其直属统计量, 也可以分别计算这些变量的子组统计量。

SPSS 的个案摘要报告功能可以将数据按要求进行整理和报告, 即程序可以为一个或多个分组变量类别中的变量计算子组统计量, 并将各级别的统计量进行列表, 以形成个案摘要报告。所有级别的分组变量要进行交叉制表, 因此研究者可以选择显示统计的顺序。个案摘要报告将显示跨所有类别的每个变量的汇总统计。每个类别中的数据值可以列出, 也可以不列出, 大型数据的个案摘要报告可以只列出部分个案的分析结果。

个案摘要报告可在“查看器”窗口生成, 方便用户浏览或打印, 或者对数据进行简单的描述统计。打开要分析的数据文件, 可以在“数据编辑器”窗口进行个案摘要报告的生成操作。

动手练 个案摘要报告案例



下面以“员工信息”数据文件为例说明个案摘要报告的生成步骤。本例需要考察不同“学历”和不同“职称”在“工资”上的不同之处。具体操作步骤如下。

Step 01 打开“员工信息”数据文件, 执行“分析”|“报告”|“个案摘要”命令, 如图3-1所示。

Step 02 打开“个案摘要”对话框, 单击按钮, 将“基本工资 [工资]”选入“变量”列表中, 如图3-2所示。

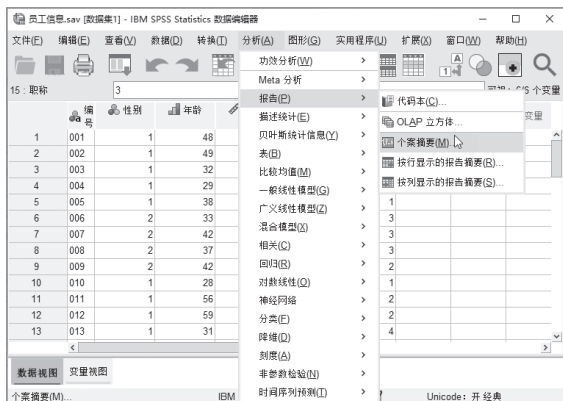


图 3-1

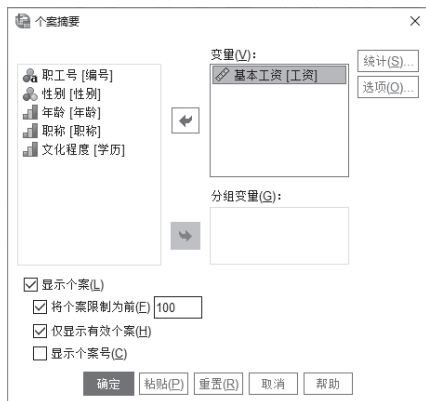


图 3-2

Step 03 将“职称”和“文化程度”变量选入“分组变量”列表中，单击“统计”按钮，如图3-3所示。

Step 04 打开“摘要报告：统计”对话框，将“个案数”“均值”“标准差”选入“单元格统计”列表中，单击“继续”按钮，如图3-4所示。

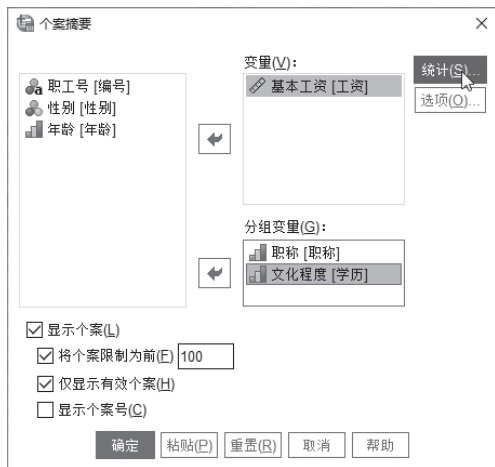


图 3-3

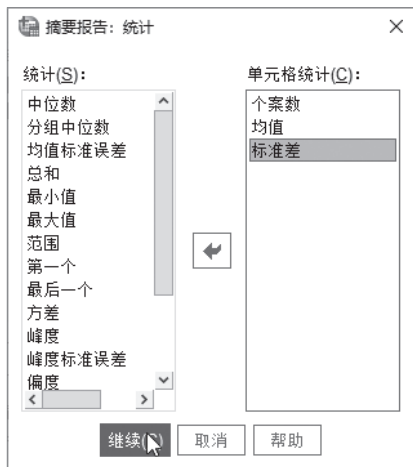


图 3-4

Step 05 返回“个案摘要”对话框，单击“选项”按钮，如图3-5所示。

Step 06 打开“选项”对话框，定义标题为“基本工资个案汇总表”，其他选项保持默认，单击“继续”按钮，如图3-6所示。

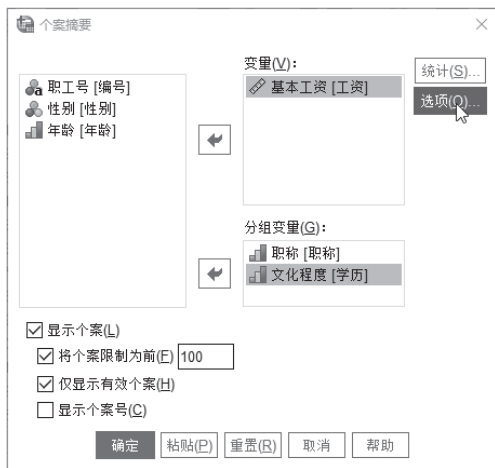


图 3-5

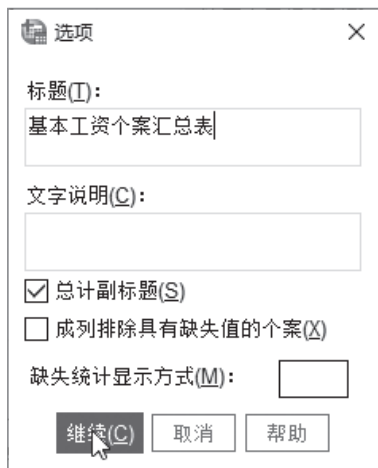


图 3-6

Step 07 返回上一级对话框，单击“确定”按钮，完成操作。“查看器”窗口生成个案摘要报告，如图3-7所示。

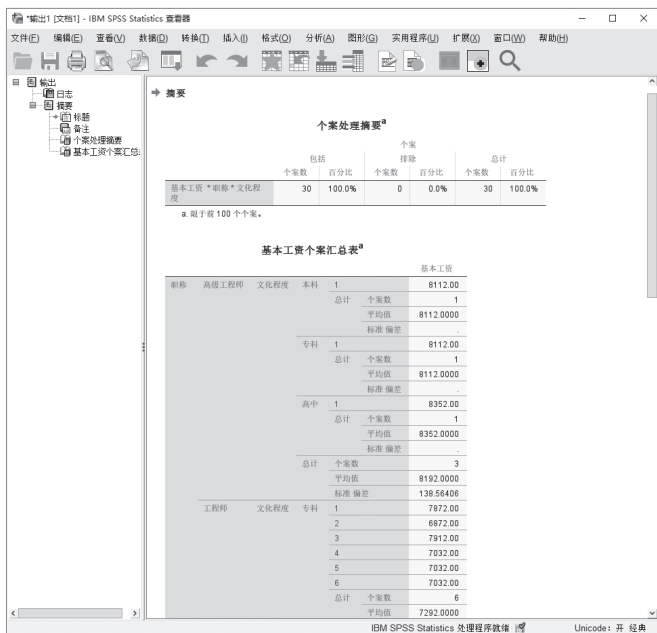


图 3-7

3.1.2 行形式摘要报告

行形式摘要报告的生成操作和个案摘要报告的生成操作类似，不同之处在于行形式摘要报告可以生成更为复杂的摘要形式。

1. 行形式摘要报告概述

行形式摘要报告的分析过程就是把数据重新进行组合，按照指定的要求显示在“查看器”窗口，此外，还可以进行相关数据的统计分析，并给出相应的统计量报告，以方便用户浏览、打印或对数据进行简单的描述统计。

行形式摘要报告的分析生成过程与个案摘要报告类似，只是更为复杂，输出的格式设置也更为详细。打开要分析的数据文件，可以在“数据编辑器”窗口进行行形式摘要报告的生成操作。

动手练 行形式摘要报告案例

下面仍以“员工信息”数据文件为例，说明行形式摘要报告的生成操作步骤，输出“工资”在“职称”和“学历”上的差异的行形式摘要报告。

Step 01 打开“员工信息”数据文件，执行“分析”|“报告”|“按行显示的报告摘要”命令，如图3-8所示。

Step 02 打开“报告：行摘要”对话框，将“基本工资”变量选入“数据列变量”列表中；将“职称”和“文化程度”变量选入“分界列变量”列表中。随后单击“数据列变量”列表下方的“格式”按钮，如图3-9所示。

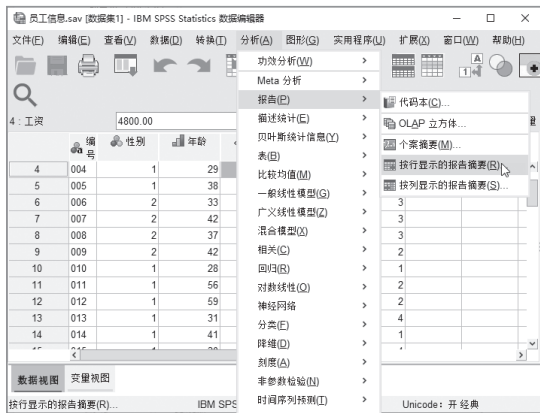


图 3-8

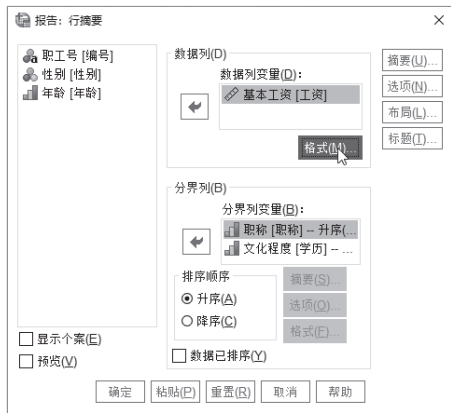


图 3-9

Step 03 打开“报告：工资的数据列格式”对话框，输入列标题，单击“继续”按钮返回上一级对话框，如图3-10所示。

Step 04 单击“摘要”按钮，如图3-11所示。

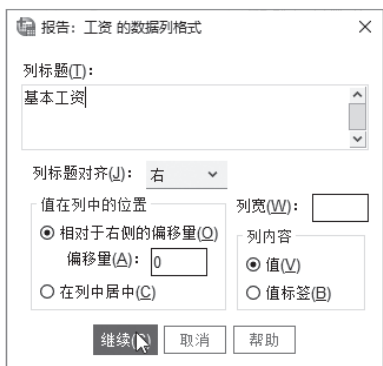


图 3-10

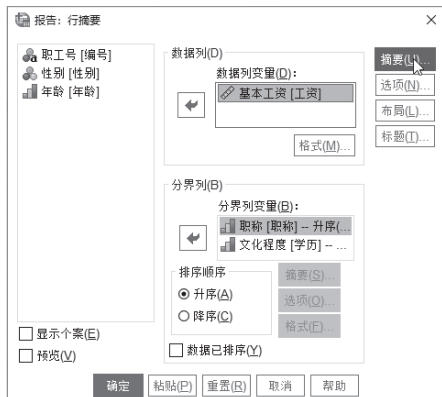


图 3-11

Step 05 打开“报告：最终摘要行”对话框，勾选“值的均值”“标准差”和“个案数”复选框，单击“继续”按钮返回上一级对话框，如图3-12所示。

Step 06 单击“选项”按钮，打开“报告：选项”对话框，勾选“成列排除具有缺失值的个案”复选框，单击“继续”按钮返回上一级对话框，如图3-13所示。



图 3-12

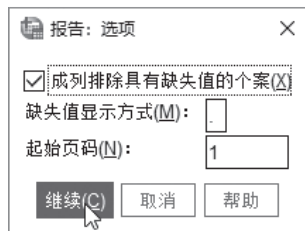


图 3-13

Step 07 单击“布局”按钮，打开“报告：布局”对话框，在该对话框中可以设置输出报告的整体布局。此处选择系统默认设置，单击“继续”按钮返回上一级对话框，如图3-14所示。

Step 08 单击“标题”按钮，打开“报告：标题”对话框，在该对话框中可以不为报告设置标题行和脚注，也可以自行输入或从源变量列表中选择变量作为标题或脚注，此处选择系统默认设置，单击“继续”按钮返回上一级对话框，如图3-15所示。



图 3-14

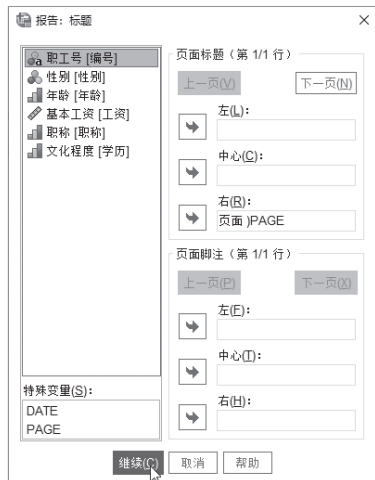


图 3-15

知识点拨

“分界列变量”的属性设置和“数据列变量”的设置基本相同。在“分界列”区域下方分别单击“摘要”“选项”和“格式”按钮，即可设置终止列变量的统计量、页面参数、列标题和列的属性等，此处不再赘述。

Step 09 所有操作完成后，单击“确定”按钮，如图3-16所示。

Step 10 “查看器”窗口中随即生成行形式摘要报告，如图3-17所示。

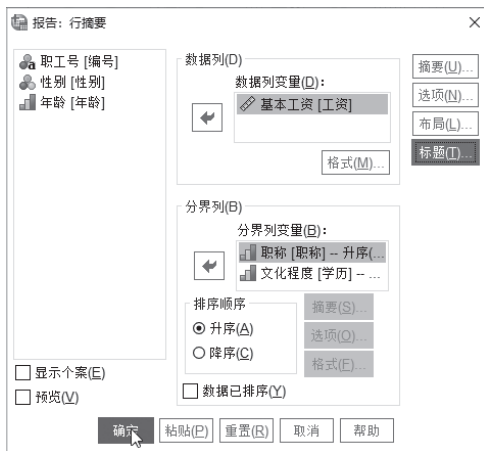


图 3-16

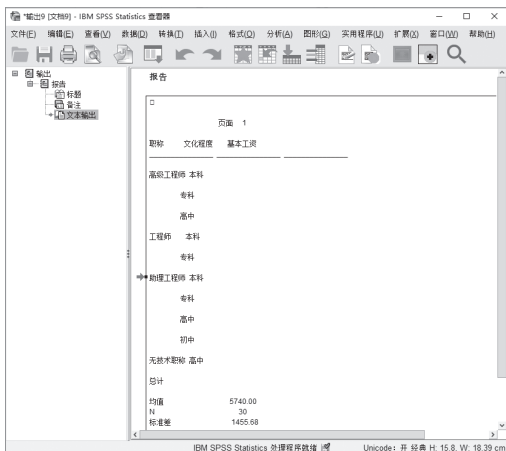


图 3-17

3.1.3 列形式摘要报告

列形式摘要报告的生成步骤和行形式摘要报告的生成步骤类似，基本功能也比较接近，只是输出的报告形式有所差异。

1. 列形式摘要报告的操作概述

和行形式摘要报告一样，列形式摘要报告的分析过程也是把数据重新进行组合，按照研究者指定的要求罗列在“查看器”窗口，方便用户浏览、打印，同样，列形式摘要报告也可以进行相关数据的统计分析，并给出相应的统计量报告，以便于研究者对数据进行简单的描述统计。列形式摘要报告的分析生成过程与行形式摘要报告类似，只是输出的报告形式有所差异。

动手练 列形式摘要报告案例应用

下面对“员工信息”数据文件进行列形式摘要报告输出的操作，以观察不同职称和学历的基本工资和奖金情况。

Step 01 打开“员工信息”数据文件，执行“分析”|“报告”|“按列显示的报告摘要”命令，如图3-18所示。

Step 02 打开“报告：列摘要”对话框。将“工资”和“奖金”变量选入“数据列变量”列表中，变量将显示为“工资:sum”和“奖金:sum”，因为系统默认“数据列变量”列表中的变量显示为“变量:sum”，如图3-19所示。



图 3-18

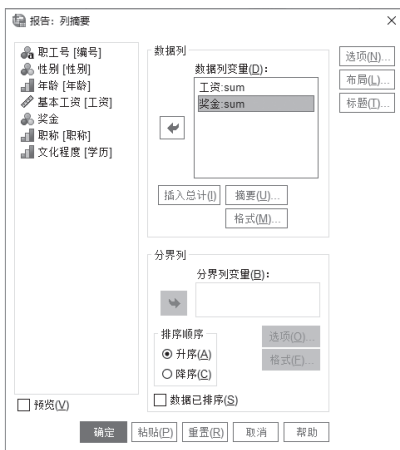


图 3-19

Step 03 将“职称”和“文化程度”依次选入“分界列变量”列表中，选中“工资:sum”变量，单击“摘要”按钮，如图3-20所示。

Step 04 打开“报告：工资的摘要行”对话框，选中“值的均值”单选按钮，单击“继续”按钮，如图3-21所示。

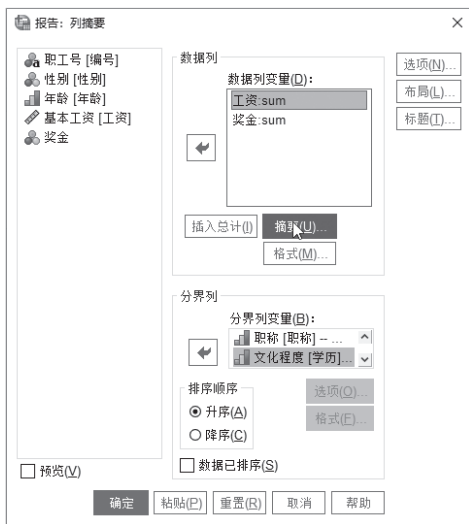


图 3-20

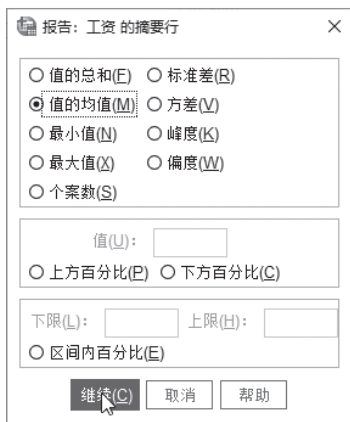


图 3-21

Step 05 “工资:sum”将变为“工资:均值”。重复以上操作，完成对“奖金:均值”的统计量设置。随后选中“工资:均值”统计量，单击“格式”按钮，如图3-22所示。

Step 06 打开“报告:工资的数据列格式”对话框，定义列标题为“工资平均值”，单击“继续”按钮返回上一级对话框，如图3-23所示。重复上述操作，定义“奖金:均值”。

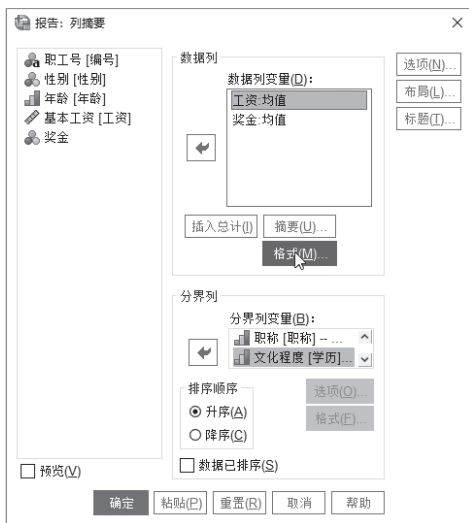


图 3-22

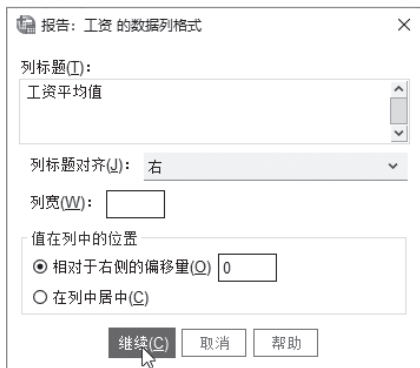


图 3-23

Step 07 单击“插入总计”按钮，在“数据列变量”列表中插入新变量“总计”，选中“总计”变量，单击“摘要”按钮，如图3-24所示。

Step 08 打开“报告:摘要列”对话框，将“数据列”列表框中的变量全部选入“摘要列”列表中，“摘要函数”使用默认的“各个列的总和”选项，单击“继续”按钮，如图3-25所示。

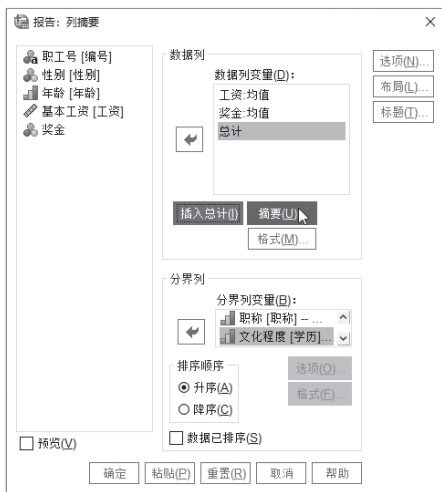


图 3-24



图 3-25

Step 09 单击“确定”按钮完成设置，如图3-26所示。

Step 10 “查看器”窗口随即输出员工工资的列形式摘要报告，如图3-27所示。

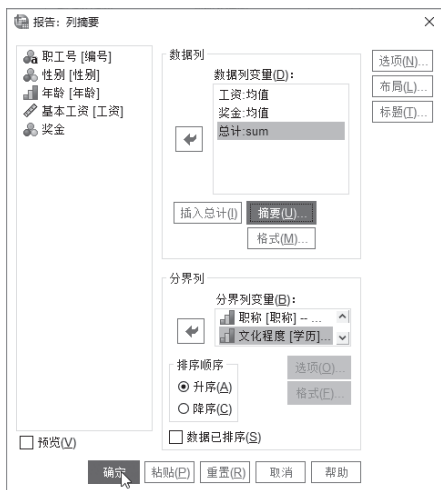


图 3-26

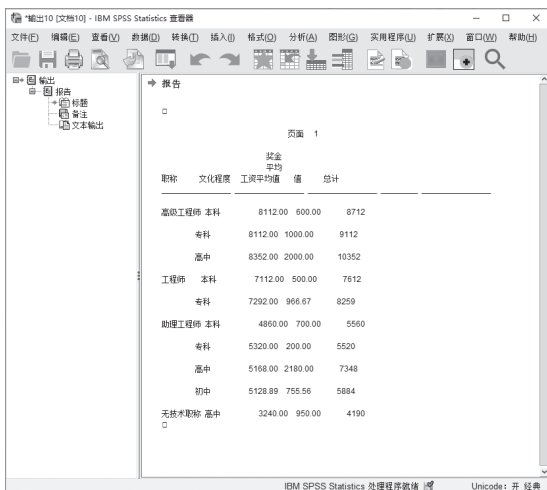


图 3-27

3.1.4 在线分析处理报告

在线分析处理（On-Line Analytical Processing, OLAP）报告是一套以多维度方式进行资料分析，并能呈现整合性决策信息的方法，多用于决策支持系统、商务智能或数据仓库，其主要功能在于方便大规模数据分析及统计计算，对决策提供参考和支持。

1. 在线分析处理报告的概述

20世纪60年代，关系数据库之父E. F. Codd提出了关系模型，促进了在线事务处理（OLTP）的发展（数据以表格的形式而非文件方式存储）。然而，随着数据库技术的发

展和应用,数据库存储的数据量也成几何倍数迅猛增长,OLTP已不能满足终端用户对数据库查询分析的需要,于是E. F. Codd便在1993年提出了多维数据库和多维分析的概念,即在线分析处理。OLAP这类软件能够使分析人员、管理人员或执行人员多角度地对出自原始数据且能够真正为用户所理解并真实反映变量特性的信息进行快速、一致、交互地存取,从而获得对数据更深入的了解。它的目标是满足决策支持或多维环境特定的查询和报表需求,其技术核心是“维”(即人们观察数据的特定角度,是考虑问题时的一类属性,如时间维、空间维等),因此OLAP也可以说是多维数据分析工具的集合。同时,其还具有分析功能灵活、数据操作直观和分析结果可视化等优点,从而使基于大规模复杂数据的分析变得轻松而高效,有利于决策者迅速作出正确判断。

OLAP多维数据分析可以计算一个或多个分组变量中连续摘要变量的总和、平均值和其他单变量统计,并在表格中为每个分组变量的每个类别创建单独的层。其优势在于能够生成多层表格,用户可以按照需要对分组变量的某个特定水平组合进行结果输出。其最大特点是交互性强,用户可自行选择报告的内容和形式。打开要分析的数据文件,可以在“数据编辑器”窗口进行在线分析处理报告的生成操作。

动手练 在线分析处理报告案例

下面以“员工信息”数据文件为例,说明生成在线分析处理报告的操作步骤。本次数据分析的目的是根据“年龄”和“学历”对工资进行对比。

Step 01 打开“员工信息”数据文件,执行“分析”|“报告”|“OLAP立方体”命令,如图3-28所示。

Step 02 打开“OLAP立方体”对话框,将“年龄”和“文化程度”选入“摘要变量”下的列表中;将“基本工资”选入“分组变量”下的列表中。单击“统计”按钮,如图3-29所示。



图 3-28

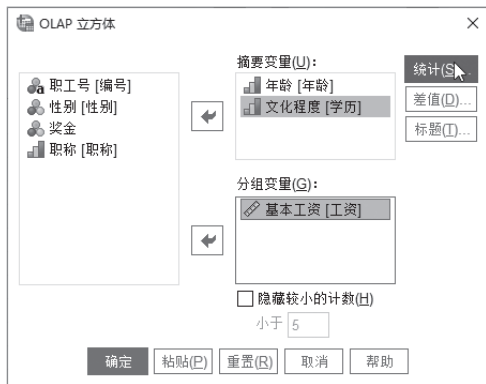


图 3-29

Step 03 打开“OLAP立方体：统计”对话框，依次将“统计”列表中的“总和”“个案数”“均值”“标准差”“最小值”和“最大值”选入“单元格统计”列表中，单击“继续”按钮，如图3-30所示。

Step 04 返回“OLAP立方体”对话框，单击“差值”按钮，打开“OLAP立方体：差值”对话框，进行差值设置，本例选择默认设置。单击“继续”按钮，如图3-31所示。

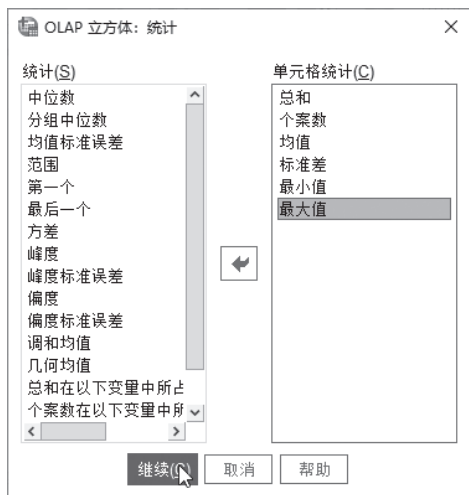


图 3-30

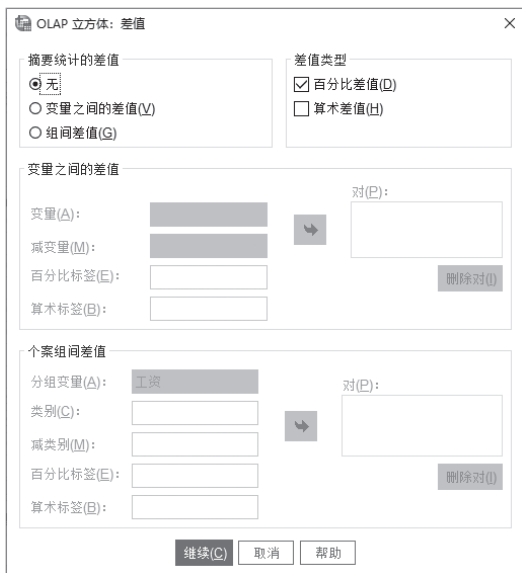


图 3-31

知识 点拨

“OLAP立方体：统计”对话框中的“统计”列表中列举了SPSS在线处理分析报告可以计算的各种统计量，选入“单元格统计”列表中的统计量将显示在最后的输出表格内。“统计”列表中的统计量的含义如下。

- **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，那么中位数是个案在以升序或降序排列的情况下中间的两个个案的平均值。中位数是对集中趋势的测量，但对于远离中心的值不敏感（这与平均值不同，平均值容易受到少数多个非常大或非常小的值的影响）。
- **分组中位数**：针对编码到组中的数据计算的中位数。分组中位数是由已编码的数据计算得出的。
- **均值标准误差**：取自同一分布的样本与样本之间的平均值之差的测量，可以粗略地将观察到的平均值与假设值进行比较（即如果差与标准误差的比值小于-2或大于2，那么可以断定两个值不同）。
- **最小值**：数值变量的最小值。
- **最大值**：数值变量的最大值。
- **范围**：数值变量最大值和最小值之间的差，即最大值减去最小值。
- **第一个**：数据文件中的第一个数据值。
- **最后一个**：数据文件中的最后一个数据值。

- **方差**: 是对围绕平均值的离差的测量, 其测量值等于各变量值与平均值的差的平方和除以个案数减1。度量方差的单位是变量本身的单位的平方。
- **峰度**: 是对观察值聚集在众数周围的程度的测量。对于标准正态分布, 峰度系数值为0。峰度系数大于0为高峰度, 高峰度表示相对于标准正态分布, 观察值聚集于众数周围的更多; 峰度系数小于0为低峰度, 低峰度表示相对于标准正态分布, 观察值聚集于众数周围的更少。
- **峰度标准误差**: 峰度与其标准误差 (0.662) 的比值可用于正态性检验 (即如果比值小于-2或大于2, 就可以拒绝正态性)。
- **偏度**: 用于分布的不对称性测量。正态分布是对称的, 偏度值为0。具有显著的正偏度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。当偏度值超过偏度标准误差的两倍时, 认为不具有对称性。
- **偏度标准误差**: 偏度与其标准误差 (0.337) 的比值可用作正态性检验 (即如果比值小于-2或大于2, 就可以拒绝正态性)。
- **调和均值**: 在组中样本大小不等的情况下用来估计平均值大小。调和平均值是样本总数除以样本大小的倒数总和。
- **几何均值**: 即数据值的乘积的 n 次根, 其中 n 代表个案数目。
- **总和**: 即所有带有非缺失值的个案的值的合计。
- **总和在以下变量中所占百分比**: 其他分组变量的类别内指定分组变量的总和的百分比。如果只有一个分组变量, 那么此值与总和百分比相同。
- **个案数在以下变量中所占的百分比**: 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, 那么此值与总个案数百分比相同。
- **标准差**: 是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倍标准差范围内, 95%的个案在平均值的两倍标准差范围内。
- **均值**: 是对集中趋势的测量, 是算术平均, 即总和除以个案数。
- **在总和中所占的百分比**: 表示每个类别中的总和百分比。
- **个案数**: 个案 (观察值或记录) 的数目。
- **在总个案数中所占的百分比**: 每个类别中的个案总数的百分比。

Step 05 返回“OLAP立方体”对话框, 单击“标题”按钮, 如图3-32所示。打开“OLAP立方体: 标题”对话框。该对话框包括“标题”和“文字说明”两个文本框, 分别用于定义输出表格的标题和对表格的说明。此处, 设置标题为“工资对比”, 单击“继续”按钮返回上一级对话框, 如图3-33所示。

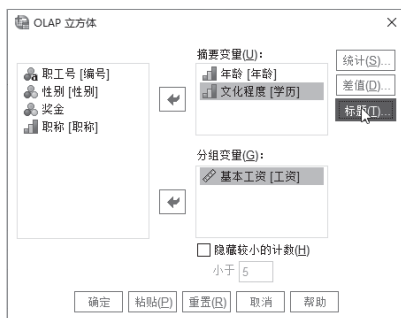


图 3-32

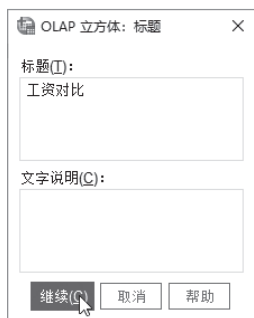


图 3-33

Step 06 完成操作后，SPSS“查看器”窗口中将自动成在线分析处理报告，如图3-34所示。

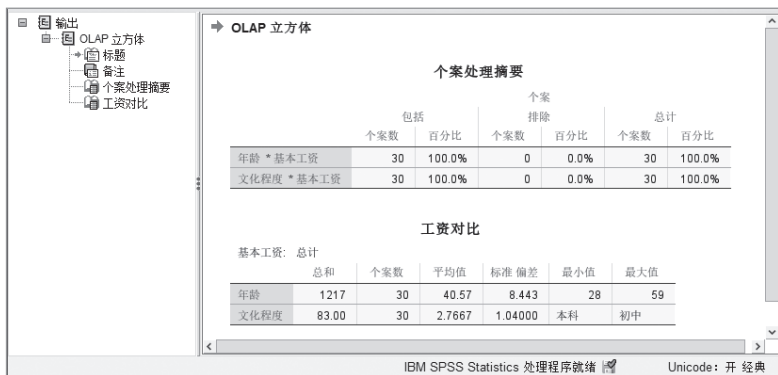


图 3-34

知识点拨

“OLAP立方体：差值”对话框提供四个选项组。

- **摘要统计的差值：**该组包括“无”“变量之间的差值”和“组间差值”三个单选按钮，分别表示不进行差值计算、计算变量之间的差值和计算分组间的差值。如果选择后两项，将分别激活“变量之间的差值”和“个案组间差值”选项组。
- **差值类型：**该选项组包括“百分比差值”和“算术差值”两个复选框，用于选择要计算的差值统计量的类型。
- **变量之间的差值：**该项用于计算变量之间的差值，即每一对变量中第一个变量的汇总统计值减去第二个变量（减去的变量）的汇总统计值。因此计算变量对之间的差值至少需要两个摘要变量。该选项组包括“变量”“减变量”两个下拉列表和“百分比标签”“算术标签”两个文本框，分别用于设置对比的变量和差异形式。设置完毕后单击，将其选入“对”文本框，即可用于多种变量之间的对比。也可以单击“删除对”按钮，取消“变量对”的比较。
- **个案组间差值：**该选项组用于计算由分组变量定义的组间的差值，即每一对中第一个类别的汇总统计值减去第二个类别（减去的类别）的汇总统计值。个案组之间的差值至少需要一个分组变量。该选项组中包括一个“分组变量”下拉列表和“类别”“减类别”“百分比标签”和“算术标签”四个文本框，分别用于设置分组变量、比较的各类别取值和差异方式。

S 3.2 频率分布

对数据的描述统计往往是从最基本的频率分析开始的。通过频率分析，用户可以了解变量的取值状况，以进一步把握数据的分布特征。

3.2.1 频率分布概述

频率是指某一观察值在一组数据中出现的次数，把各个观察值及其相应的频率全部罗列出来就是“频率分布”，SPSS中有专门用于生成频率分布表的分析模块——“频

率”过程。通过频率分布表可以对数据进行分组归类整理。该分析模块还可以生成条形图、饼图和直方图等统计图，以直观展现频率的分布形态。

1. 频率分布表

生成频率分布表是进行频率分析的第一个基本任务，其主要组成部分如下。

- **频率**：指变量值位于某个区间（或类别）的次数。
- **百分比**：指各频率占总样本数的百分比。
- **有效百分比**：指各频率占总有效样本的百分比。
- **累计百分比**：指百分比逐级累加的结果，最终取值为100%。

2. 频率分析的常用统计图

频率分析的另一功能是绘制统计图。统计图是一种非常直观的数据指示方法，能清晰地展示变量的取值分布状况。

- **条形图**：用相同宽度的条形的高度指示频率分布变化的图形，适用于定序和定类变量的统计分析。条形图的Y轴可以使用频率或百分比，X轴通常为定类变量。条形图包括简单条形图和集群条形图。
- **饼图**：用圆形及扇形的面积指示频率百分比变化的图形。圆内扇形面积可以表示频率，也可以表示百分比。一般条形图与饼图适用于类别变量。
- **直方图**：用矩形的面积表示频率分布变化的图形，适用于连续变量的分析研究。还可以在直方图上附加正态分布曲线，便于将频率分布与正态分布进行比较。

3. 频率分布的特征

频率分布具有集中趋势和离散两个特征。集中趋势用于描述频率分布位置的统计量，包括平均值、中位数、众数和所有值的总和。离散用于测量数据中变异和扩散的统计量，包括标准差、方差、范围、最小值、最大值和平均值标准误差。

3.2.2 生成频率分布分析报告



下面对“初中生学习压力调查”问卷的回收数据进行分析，以了解初中生学习压力的整体现状。

Step 01 打开“初中生学习压力调查”数据文件，执行“分析”|“描述统计”|“频率”命令，如图3-35所示。



图 3-35

Step 02 打开“频率”对话框。从对话框左侧变量列表中选择“压力总分”并将其选入“变量”下的列表中，如图3-36所示。

Step 03 在“频率”对话框中单击“统计”按钮，打开“频率：统计”对话框，选择需要输出的统计量。该对话框包含四个统计量选项组，分别是“百分位值”“集中趋势”“离散”和“分布”，选择后单击“继续”按钮，返回“频率”对话框，如图3-37所示。



图 3-36

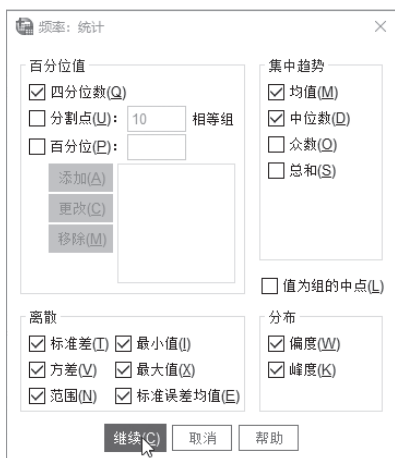


图 3-37

Step 04 单击“图表”按钮，打开“频率：图表”对话框。该对话框有“图表类型”和“图表值”两个选项组，“图表类型”提供“无”“条形图”“饼图”和“直方图”四种格式。本案例选中“直方图”单选按钮，并勾选“在直方图中显示正态曲线”复选框；“图表值”选项选择程序的默认设置“频率”，单击“继续”按钮，返回“频率”对话框，如图3-38所示。

Step 05 单击“格式”按钮，打开“频率：格式”对话框。在“排序方式”选项组中选择频率分析输出结果的排列顺序；在“多个变量”选项组中选择有多个变量时分析结果的输出方式。如果勾选“禁止显示具有多个类别的表”复选框，表示变量有太多的类别时不输出频率分布表。系统默认最大类别数为10。本例选择默认设置“按值的升序排序”和“比较变量”。单击“继续”按钮，返回“频率”对话框，如图3-39所示。

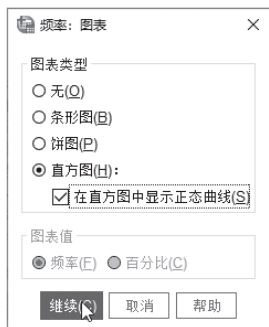


图 3-38

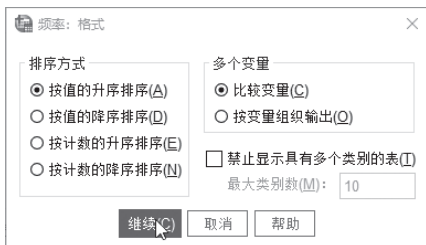


图 3-39

Step 06 单击“样式”按钮，打开“表样式”对话框，设置输出表格的各种属性。“表样式”对话框用于指定根据具体条件自动更改透视表属性的条件。单击“继续”按钮，返回“频率”对话框，如图3-40所示。



图 3-40

Step 07 单击“确定”按钮，输出频率分布的分析结果，如图3-41~图3-43所示。

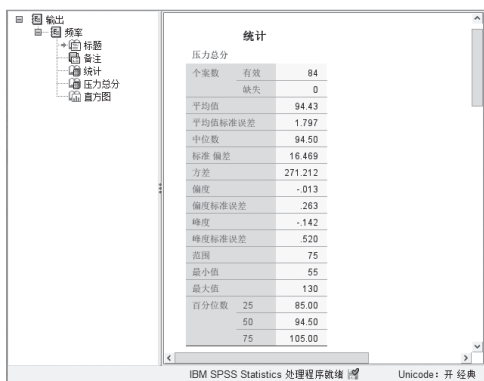


图 3-41

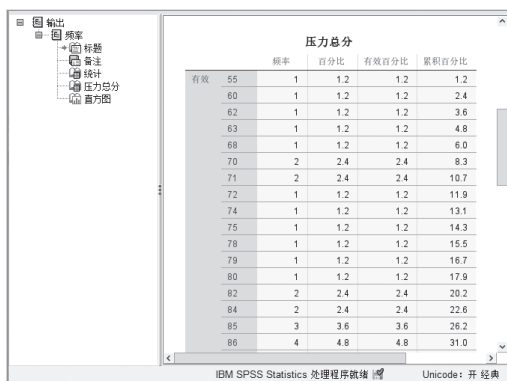


图 3-42

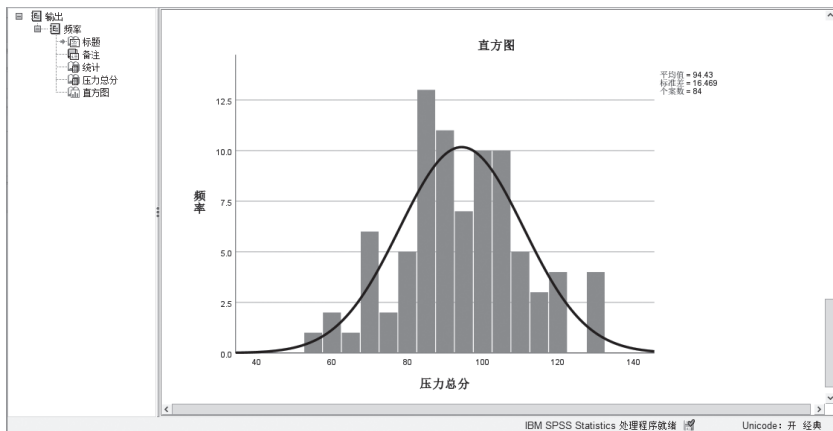


图 3-43

S 3.3 描述性统计

描述性统计主要是指对调查问卷等方式所采集的数据进行归纳、整理和整体性描述。

3.3.1 描述性统计概述

描述性统计主要用于对定距变量的数据分布状况的分析。通常，描述性统计可以通过平均值和平均值的标准误差展示定距数据的集中趋势；通过标准差、方差、最大值、最小值和范围等统计量描述数据的离散程度；通过峰度和偏度指示数据的分布形态。描述性统计的分析步骤与频率分布分析的操作类似，所涉及统计量的含义与频率分布分析的统计量相同，故不再一一赘述。

3.3.2 描述性统计案例讲解

本节以“初中生学习压力调查”数据文件为例，分析初中生对学习压力的感受。本例中的问卷由“姓名”“年级”以及28个问题和“压力总分”构成，对每个问题的判别方式使用5级利克特量表来表达从“很同意”（用1表示）到“很不同意”（用5表示）的态度变化。这样的量表类似于定距变量，因此可以进行描述性统计分析。具体分析步骤如下。

Step 01 打开“初中生学习压力调查”数据文件，执行“分析”|“描述统计”|“描述”命令，如图3-44所示。



图 3-44

Step 02 打开“描述”对话框。从对话框左侧变量列表中选择所有问题变量，并依次将这些变量选入“变量”列表中。该对话框下方还有“将标准化值另存为变量”复选框，用以将普通分数变成标准分数——Z分数。本例无须勾选此复选框，如图3-45所示。

Step 03 单击“选项”按钮，打开“描述：选项”对话框，选择需要的统计量，在“显示顺序”选项组选中“按均值的升序排序”单选按钮，以便于直观展示态度变化的趋势。设置完成后，单击“继续”按钮，返回“描述”对话框，如图3-46所示。

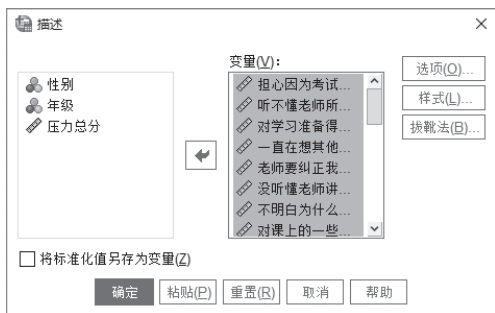


图 3-45

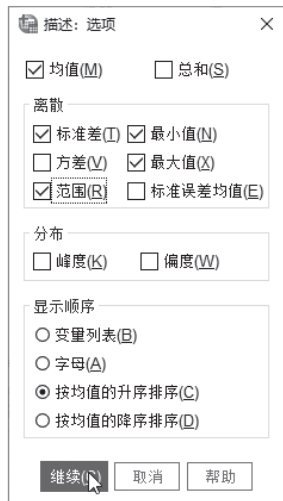


图 3-46

Step 04 单击“确定”按钮，完成操作，如图3-47所示。“查看器”窗口中随即输出分析报告表，如图3-48所示。

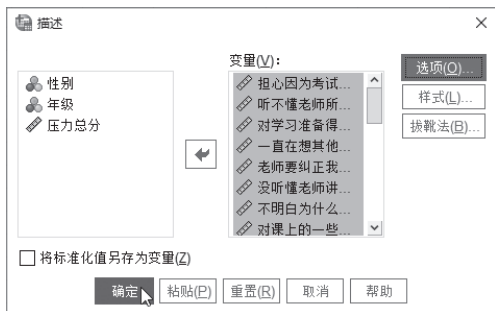


图 3-47

描述统计						
	N	范围	最小值	最大值	平均值	标准差
担心因为考试不及格带来的后果	84	4	1	5	1.85	1.114
课上不愿自愿发言	84	4	1	5	2.15	.988
听不懂老师讲话时会很不自在	84	4	1	5	2.17	1.062
老师问事先没有准备的问题时感到紧张	84	4	1	5	2.50	1.092
觉得其他同学的成绩比自己好	84	4	1	5	2.60	1.121
课上做没有准备的发言时感到恐慌	84	4	1	5	2.64	1.158
课上很紧张以致知道的东西都忘了	84	4	1	5	2.65	1.237
一直在想其他同学的成绩比自己好	84	4	1	5	2.68	1.234
上课时在想一些和课堂内容无关的事	84	4	1	5	2.73	1.112
对课上的一些小测验感到紧张	84	4	1	5	2.77	1.165
没听懂老师讲课内容会感到害怕	84	4	1	5	2.83	1.139
老师点到名字时感到心跳得很厉害	84	4	1	5	2.85	1.177
听不懂老师说什么感到很不安	84	4	1	5	2.86	1.153
那么多要求背诵的内容使人害怕	84	4	1	5	2.88	1.176
如果学习不好担心别同学取笑	84	4	1	5	2.93	1.170
不明白为什么有些人对学习这么害怕	84	4	1	5	2.94	1.264
对学习准备得很充分,还是感到焦虑	84	4	1	5	2.95	1.171
未做好准备会觉得有压力	84	4	1	5	3.04	1.145
上课使我感到烦恼	84	4	1	5	3.05	1.260
课上发言时感到紧张不安	84	4	1	5	3.05	1.221
希望最好不用去上课	83	4	1	5	3.06	1.391
课程进展很快,担心跟不上	84	4	1	5	3.07	1.149
在课上发言很自信	84	4	1	5	3.14	1.132
去上课的路上感到很轻松	84	4	1	5	3.18	1.020
上课使我感到紧张和不安	84	4	1	5	3.26	1.152
考试准备得越多越觉得没戏	84	4	1	5	3.29	1.168
课上老师叫我会发抖	83	4	1	5	3.41	1.169
老师要纠正我错误时很害怕	84	4	1	5	3.50	1.156
有效个案数(成列)	82					

图 3-48

知识点拨

为调查初中生学习压力使用态度调查描述统计表。在一个5级量表中，中位数为2.5。高于2.5表示否定态度，低于2.5表示肯定态度。从平均值的排序来看，“担心因为考试不及格带来的后果”的统计量平均值最低，为1.85，说明在该项上受访者的态度是最肯定的。可以看出，受访者对前3项基本都持否定态度。另一方面，第四项之后的项的平均值都大于2.5，说明中学生在这些选项上持否定态度。后11项的平均值相对较高，说明学生对这些问题持否定态度的占比更大。另外，根据标准差也能判断出各项上的个体差异，标准差越小，说明个体差异越小，标准差越大，则说明个体差异越大。

S 3.4 探索性分析

所谓探索性分析，即用户对数据分布的特点尚不太了解，需要对数据进行试探性的考察或探索，计算常用统计量并绘制统计图，便于为下一步采用何种统计方式进行数据分析提供参考。因此，探索性分析是统计分析中的一个重要环节。

3.4.1 探索性分析概述

探索性分析基于对数据的基本特征有所了解，但仍需对数据进行更为细致和深入的描述性观察分析。因此，不仅包括数据分析的一般性描述，还增加了对数据的文字或图表的描述，使得数据分析更深入、细致和全面。探索性分析主要包括以下几项。

(1) 寻找数据中的奇异值

在数据的整理输入过程中，难免会有所疏漏，使得数据中产生某些影响分析结果的奇异值。寻找这些奇异值，并分析其产生的原因，以决定对这些数值删除或保留。

(2) 正态分布检验

通过探索性分析，可以进一步对数据的分布状况进行深入分析，以验证是否符合正态分布，进而确定能否使用正态分布数据适用的分析方法对数据进行分析。常用的正态分布检验是Q-Q概率图。

(3) 方差齐性检验

探索性分析程序通过方差齐性检验来比较各组数据之间的方差是否相等，以判定数据的离散程度是否存在差异。如果方差齐性检验得到的显著性水平小于0.05，就可以拒绝方差相同的假设。

3.4.2 探索性分析结果的图形描述

探索性分析除了能够进行一般的简单数据描述外，还增加了以图形的方式对数据的分布予以直观呈现的功能。这些图形包括茎叶图、直方图、箱图和Q-Q概率图。

(1) 茎叶图

茎叶图是用于描述连续变量的一种手法，主要由三部分组成，即频率、茎和叶。其中，茎和叶分别表示数据的整数部分和小数部分。茎对应观测值的十位数，叶对应观测值的个位数。一个个位数代表一个观测值，每一行的左边的频率就是该行对应的个案数。每个茎叶图的底部还注明茎宽和每叶代表的个案数。数据的值即为茎叶组成的数值再乘以茎宽。茎叶图既保留了数据的频率分布，也保留的原始数据，是探索性分析的常用方法之一，如图3-49所示。



图 3-49

(2) 直方图

直方图用于对连续变量数据的表示，将连续变量数据分为若干连续区间，然后计算观测值落在各个区间的频率。和条形图类似，直方图也是以区间作为水平轴，以各个区间的频率作为相应条块的高度来绘制统计图。从直方图中可以非常直观地看到数据的分布状况，如数据分布是否对称、偏左还是偏右、众数是什么，等等，还可以大体判断数据分布是否服从正态分布，如图3-50所示。

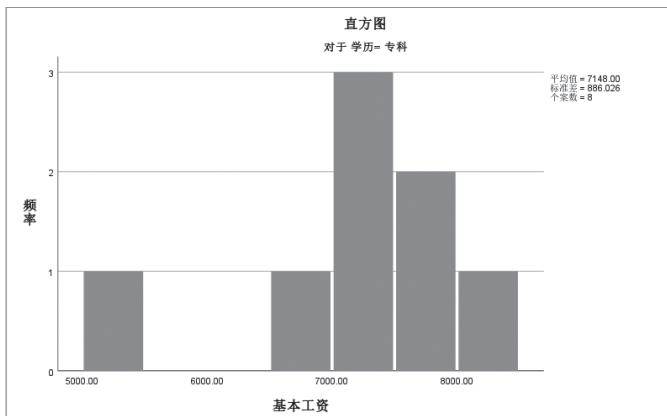


图 3-50

(3) 箱图

箱图是表现五数（最小值、最大值、中位数、第一个四分位数、第三个四分位数）的图形形式，其中的矩形为箱图的主体，箱体的上下边线差为箱长，也称为“内四分位限”。箱体部分包含全体数据约50%的数值，箱体的上、中、下三条平行线分别表示

75%、50%（即中位数）和25%分位数。纵贯箱体中间的竖线称为触须线，触须线上下两端的横线代表该组变量数值的最大值（97.5%）和最小值（2.5%）。箱图在比较两个或多个变量时尤其有用，还可以用于判断极端值的存在。如果箱图中有异常值，用“。”表示，如果有极端异常值，则用“*”表示，如图3-51所示。

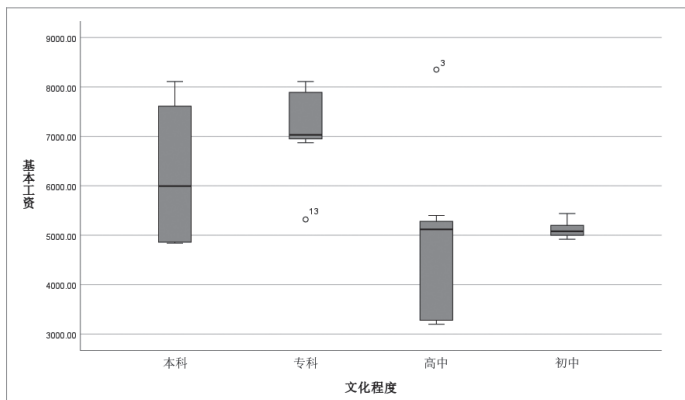


图 3-51

(4) Q-Q概率图

Q-Q概率图是散点图，用于验证数据分布的正态性，其有正态Q-Q概率图和去势（离散）正态Q-Q概率图两种表现形式。正态Q-Q概率图如图3-52所示，是以变量的观测值为 X 轴、以期望值的正态分布为 Y 轴组成的图形。图形中还有一条直线作为理论的正态累积概率分布线，该直线的斜率为标准差，截距为平均值。如果构成实际概率分布线的散点大致呈直线，且均匀地分布于理论正态累积概率分布线两侧附近，那么数据的分布就近似呈现正态分布。去势（离散）正态Q-Q概率图如图3-53所示，是以变量的观测值为 X 轴坐标，以该变量各个观测值的 Z 分数与正态分布期望值的偏差为纵坐标。水平直线为理论正态累积概率分布线，若散点均匀地分布在该线上下，甚至在线上，说明该点符合正态分布。图3-52和图3-53中的散点都较为均匀地分布于理论正态累积概率分布线上下两侧，说明男生的英语成绩基本呈正态分布状态。

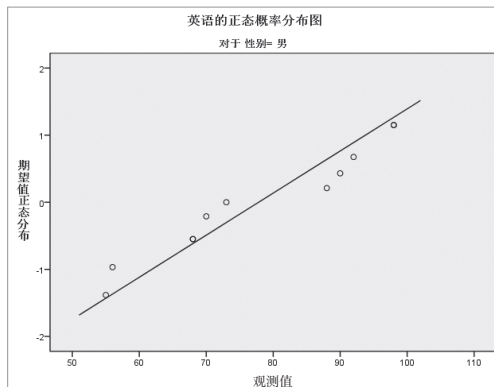


图 3-52

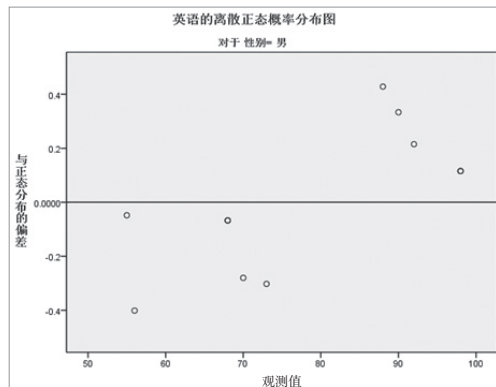


图 3-53

S 3.5 列联表分析

通过频率分析可以获得单个变量的数据分布状况，但是实际生活中，往往还要对两个或多个分类变量同时进行描述及推断的统计分析。列联表分析可以对多个分类变量进行频率分布分析，并通过“交叉表格”过程输出二维或多维的列联表，同时还可以对非数值型变量进行独立性或相关性检验。

3.5.1 列联表分析概述

列联表是指一个频率对应两个变量的表（一个变量用于对行分类，一个变量用于对列分类）。在实际分析中，除了需要对单个变量的数据分布情况进行分析外，还需要掌握多个变量在不同取值情况下的数据分布情况，从而进一步深入分析变量之间的相互影响和关系，这种分析称为列联表分析，经常用来分析问卷调查的数据。列联表分析可以较好地反映出两个因素之间有无关联性，两个因素与现象之间的相关关系。

1. 数据交叉列联表分析的基本任务

(1) 根据收集的样本数据产生二维或多维交叉列联表

列联表是两个或两个以上的变量交叉分组后形成的频率分布表，由表头、行、列、排序、计算和求百分比等部分构成。SPSS列联表分析程序能对两个或多个分类变量进行联合描述，可产生二维甚至 n 维的列联表，并计算相应的行、列合计百分比和行、列汇总指标。

(2) 在交叉列联表的基础上对两两变量间是否存在关联性进行检验

列联表的频率分析结果不能直接用来确定行变量和列变量之间的关系及关系的强弱。要想获得变量之间的关联性的信息，仅仅依靠描述统计的数据是不够的，还需借助一些变量间相关程度的统计量和一些非参数检验方法。SPSS提供多种适用于不同类型数据的相关系数的表达式。这些相关性检验的零假设都是行和列变量之间相互独立，不存在显著的相关关系。根据SPSS检验后得出的相伴概率（Concomitant Significance）判断是否存在相关关系。如果相伴概率 $P \leq 0.05$ ，那么拒绝零假设，行列变量之间彼此相关；如果相伴概率 $P > 0.05$ ，那么接受原假设，行列变量之间彼此独立。

2. 列联表分析中相关性检验的方法

(1) 卡方 (χ^2) 检验

卡方检验常用于检验行列变量之间是否相关。卡方检验首先假设行、列变量之间是独立的，并得到期望频率，再通过比较所有期望频率和实际观测频率的差异构造一个卡方统计量，如果卡方统计量大于临界值，则说明差异过大，因而假设不成立，行变量和列变量不相互独立；反之，则认为行、列变量相互独立。卡方检验的计算公式为：

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - f_{ij}^e)^2}{f_{ij}^e}$$

式中 f_{ij} 和 f_{ij}^e 分别表示实际观测频率和期望频率， r 和 c 分别代表行、列变量的取值个数。统计量综合了所有实际与期望的差异。因此，统计量的大小可以反映行、列变量的独立性。统计量越大，说明实际与期望频率的差异越大，此时行、列变量独立性越弱，在统计上，统计量近似服从自由度为 $(r-1) \times (c-1)$ 的卡方分布。

SPSS在进行运算之后，会给出相应的相伴概率 P 值，或者称作“观测到的显著水平”。如果 P 小于给定的显著性水平临界值，则拒绝原假设，认为变量间不独立；反之，则认为变量间独立。

(2) 列联系数

基于卡方检验的相关性测量，常用于名义变量之间的相关系数计算。值的范围为0~1，其中0表示行变量和列变量之间不相关，而接近1的值表示变量之间的相关度很高。可能的极大值取决于表中的行数和列数。列联系数的计算方法有很多种，最常用的是皮尔逊（Pearson）定义的列联系数，简称 C 系数：

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}}$$

其中， χ^2 是列联数据资料的检验统计量； n 是样本容量。

当双变量的测量数据被整理成次数分布表后，也可用列联相关系数表示两变量的相关程度。

3.5.2 列联表分析案例

本节使用列联表分析职称和工资之间有无必然相关性。

Step 01 打开“员工信息”数据文件，执行“分析”|“描述统计”|“交叉表”命令，如图3-54所示。

Step 02 打开“交叉表”对话框。将左侧列表框内变量“职称”选入“行”列表框，并将“基本工资”选入“列”列表框，如图3-55所示。“交叉表”对话框下部还有“显示簇状条形图”和“禁止显示表”两个复选框。前者表示显示各个变量不同交叉取值下关于频率分布的柱形图，后者表示不输出列联表的具体表格，而直接显示交叉列联表分析过程中的统计量。如果没有选中统计量，则不产生任何结果，所以，一般情况下，只有在分析行列变量间关系时勾选此复选框。



图 3-54

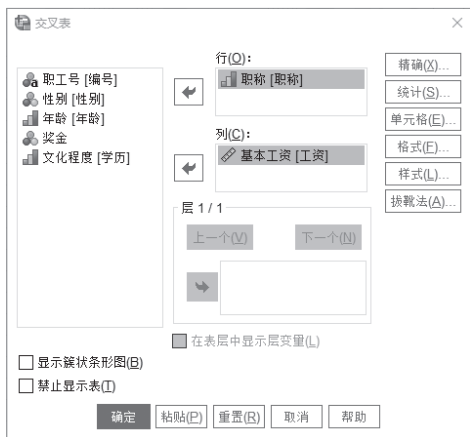


图 3-55

Step 03 单击图3-55中的“精确”按钮，打开“精确检验”对话框，该对话框提供了三种用于不同条件的检验方式检验行、列变量的相关性，如图3-56所示。用户可任意选择一种检验方式。

Step 04 单击图3-55中的“统计”按钮，打开“交叉表：统计”对话框。选择输出合适的统计检验统计量，如图3-57所示。

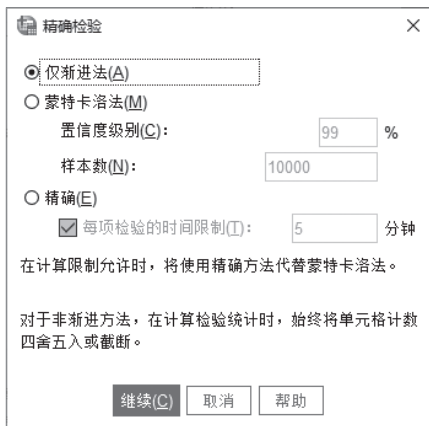


图 3-56



图 3-57

Step 05 单击图3-55中的“单元格”按钮，打开“交叉表：单元格显示”对话框，对单元格显示内容进行设置，如图3-58所示。

Step 06 单击图3-55中的“格式”按钮，打开“交叉表：表格式”对话框。在该对话框中，可以指定列联表的输出排列顺序。本例选择系统默认的升序排列。完成设置后单击“继续”按钮返回“交叉表”对话框，如图3-59所示。图3-59中各选项的具体含义如下。

- **升序**：系统默认选项，升序显示各变量值。
- **降序**：降序显示各变量值。

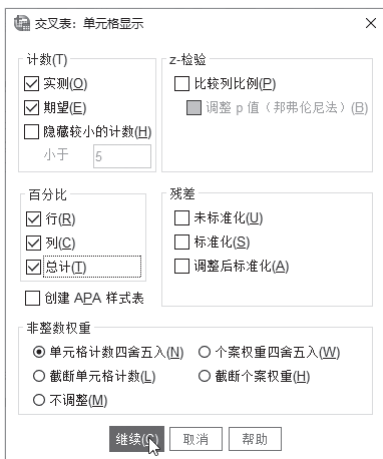


图 3-58



图 3-59

Step 07 单击图3-55中的“确定”按钮，完成所有设置。在“查看器”窗口输出所有图表。

(1) 个案处理摘要表

从图3-60中的个案处理摘要表可以看出，本例列联表分析有效数据为30人，缺失数据为0，符合分析要求。

	有效		个案 缺失		总计	
	N	百分比	N	百分比	N	百分比
职称 * 基本工资	30	100.0%	0	0.0%	30	100.0%

图 3-60

(2) 职称*基本工资交叉表

图3-61中的职称*基本工资交叉表罗列了所有职称在不同基本工资水平上的百分比以及总体上各自的比例。

		基本工资										
		3200.00	3240.00	3280.00	4800.00	4840.00	4880.00	4920.00	4960.00	5000.00	5040.00	5080.00
高级工程师	计数	0	0	0	0	0	0	0	0	0	0	0
	期望计数	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1
	占职称的百分比	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	占基本工资的百分比	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
工程师	计数	0	0	0	0	0	0	0	0	0	0	0
	期望计数	.2	.2	.2	.2	.2	.2	.2	.2	.2	.2	.2
	占职称的百分比	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	占基本工资的百分比	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
助理工程师	计数	0	0	0	1	1	1	1	1	1	1	1
	期望计数	.6	.6	.6	.6	.6	.6	.6	.6	.6	.6	.6
	占职称的百分比	0.0%	0.0%	0.0%	5.9%	5.9%	5.9%	5.9%	5.9%	5.9%	5.9%	5.9%
	占基本工资的百分比	0.0%	0.0%	0.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
无技术职称	计数	1	1	1	0	0	0	0	0	0	0	0
	期望计数	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1
	占职称的百分比	33.3%	33.3%	33.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	占基本工资的百分比	100.0%	100.0%	100.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
总计	计数	1	1	1	1	1	1	1	1	1	1	1
	期望计数	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	占职称的百分比	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%
	占基本工资的百分比	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
占总计的百分比		3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%

图 3-61

(3) 卡方检验表

图3-62中的卡方检验表给出了各种统计量的值和相应的渐进显著性 P 值。其中线性关联的 P 值小于0.05, 达到了显著水平, 表明职称和基本工资之间存在显著相关性。

(4) 对称测量表

图3-63中的对称测量表的分析结果也有效地支持了卡方检验的结果, 即职称和基本工资之间存在很强的显著性。

	值	自由度	渐进显著性 (双侧)
皮尔逊卡方	90.000 ^a	78	.166
似然比	67.316	78	.801
线性关联	27.358	1	<.001
有效个案数	30		

a. 108 个单元格 (100.0%) 的期望计数小于 5。最小期望计数为 .10。

图 3-62

	值	渐进标准误差 ^a	近似 T ^b	渐进显著性
区间到区间 皮尔逊 R	-.971	.010	-21.602	<.001 ^c
有序到有序 斯皮尔曼相关性	-.897	.044	-10.755	<.001 ^c
有效个案数	30			

a. 未假定原假设。
b. 在假定原假设的情况下使用渐进标准误差。
c. 基于正态近似。

图 3-63



案例实战：使用探索性分析生成男女体重差异报告

本案例对训练前后男生和女生的体重变化进行探索性分析, 以了解男生和女生在体重上的数据分布情况。

1. 具体操作过程

Step 01 打开“训练前后男女体重检测”数据文件, 执行“分析”|“描述统计”|“探索”命令, 如图3-64所示。

Step 02 打开“探索”对话框。将左侧列表框中的变量“训练前体重”和“训练后体重”选入“因变量列表”列表框。选择“性别”, 将其选入“因子列表”列表框, 另外该对话框还有一个“个案标注依据”列表框, 用于选择某一变量来对个案进行区分和标注, 如图3-65所示。



图 3-64



图 3-65

Step 03 单击图3-65中的“统计”按钮，打开“探索：统计”对话框。勾选“描述”“M-估计量”和“离群值”复选框。单击“继续”按钮返回“探索”对话框，如图3-66所示。

Step 04 单击图3-65中的“图”按钮，打开“探索：图”对话框。该对话框包含“箱图”“描述图”和“含莱文检验的分布-水平图”三个选项组，以及“含检验的正态图”复选框。本例选中“因子级别并置”和“无”单选按钮，并勾选“茎叶图”“直方图”“含检验的正态图”复选框。单击“继续”按钮返回“探索”对话框，如图3-67所示。

Step 05 单击图3-65中的“选项”按钮，打开“探索：选项”对话框。程序默认选中“成列排除个案”单选按钮，单击“继续”按钮返回“探索”对话框，如图3-68所示。

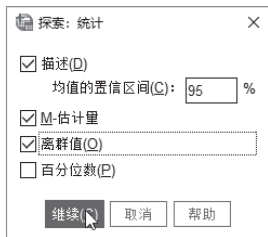


图 3-66



图 3-67

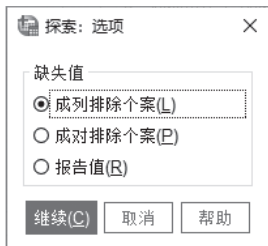


图 3-68

知识点拨

“探索：选项”对话框中各选项的含义如下。

- **成列排除个案**：从所有分析中排除任何因变量或因变量具有缺失值的个案，这是默认选项。
- **成对排除个案**：在该组的分析中包含组（单元格）中变量不具有缺失值的个案。该个案可能 在其他组中使用的变量中有缺失值。
- **报告值**：因子变量的缺失值被视为单独类别，而在结果中生成一个附加类别。

Step 06 单击图3-65中的“确定”按钮完成操作，“查看器”窗口输出分析报告。

2. 报告解读

(1) 个案处理摘要

从图3-69所示的个案处理摘要可以看出，本次探索分析有效数据为，训练前男生24人、女生27人，训练后男生24人、女生27人，无数据缺失。

	性别	有效		个案缺失		总计	
		个案数	百分比	个案数	百分比	个案数	百分比
训练前体重	男	24	100.0%	0	0.0%	24	100.0%
	女	27	100.0%	0	0.0%	27	100.0%
训练后体重	男	24	100.0%	0	0.0%	24	100.0%
	女	27	100.0%	0	0.0%	27	100.0%

图 3-69

(2) 描述性统计数据

如图3-70所示的男生和女生训练前体重的平均值分别为77.458和65.741，中位数分别为78.00和66.00，标准偏差分别为4.1701和3.7987，说明受试者的体重离散程度稍高。双方的最大值和最小值的差距，即范围较大，也说明了数据的离散程度稍高。图3-71为男生和女生训练后体重的各项统计描述。

描述					
性别		统计	标准错误		
训练前体重	男	平均值	77.458	.8512	
		平均值的 95% 置信区间			
		下限	75.697		
		上限	79.219		
		5% 剪除后平均值	77.815		
		中位数	78.000		
		方差	17.389		
		标准偏差	4.1701		
		最小值	64.0		
		最大值	84.0		
		全距	20.0		
		四分位距	4.5		
		偏度	-1.684	.472	
		峰度	3.910	.918	
	女	女	平均值	65.741	.7311
			平均值的 95% 置信区间		
		下限	64.238		
		上限	67.243		
		5% 剪除后平均值	65.479		
		中位数	66.000		
		方差	14.430		
		标准偏差	3.7987		
		最小值	59.0		
		最大值	79.0		
		全距	20.0		
		四分位距	3.0		
		偏度	1.316	.448	
		峰度	5.101	.872	

图 3-70

训练后体重	男	平均值	75.438	.8488	
		平均值的 95% 置信区间			
			下限	73.682	
			上限	77.193	
			5% 剪除后平均值	75.718	
			中位数	76.000	
			方差	17.289	
			标准偏差	4.1581	
		最小值	65.0		
		最大值	81.0		
		全距	16.0		
		四分位距	4.8		
		偏度	-1.344	.472	
		峰度	1.694	.918	
	女	女	平均值	63.111	.6368
		平均值的 95% 置信区间			
		下限	61.802		
		上限	64.420		
		5% 剪除后平均值	62.807		
		中位数	63.000		
		方差	10.949		
		标准偏差	3.3089		
	最小值	58.0			
	最大值	75.0			
	全距	17.0			
	四分位距	4.0			
	偏度	1.738	.448		
	峰度	5.487	.872		

图 3-71

(3) M-估计量表

图3-72中有a、b、c、d四种加权常量，该表的结果是使用这四种加权常量计算出的M-估计量。通过比较男生和女生的M-估计量，可以判定男生和女生的体重差异较大。

M-估计量					
性别		休伯 M 估计量 ^a	图基双权 ^b	汉佩尔 M 估计量 ^c	安德鲁波 ^d
训练前体重	男	78.190	78.527	78.232	78.546
	女	65.763	65.960	65.843	65.970
训练后体重	男	76.190	76.617	76.307	76.615
	女	62.808	62.719	62.712	62.719

a. 加权常量为 1.339。
b. 加权常量为 4.685。
c. 加权常量为 1.700、3.400 和 8.500
d. 加权常量为 1.340*pi。

图 3-72

(4) 最值

图3-73和图3-74分别列出了5个男生和5个女生体重的最大值和最小值，从中可以看出最大值和最小值的差异。

极值					
		性别	个案号		值
训练前体重	男	最高	1	2	84.0
			2	38	81.0
			3	48	81.0
			4	8	80.0
			5	13	80.0 ^a
	最小值	1	12	64.0	
		2	20	70.0	
		3	1	72.0	
		4	23	75.0	
		5	14	75.0 ^b	
	女	最高	1	17	79.0
			2	39	71.0
			3	22	68.0
			4	35	68.0
			5	36	68.0 ^c
最小值		1	7	59.0	
		2	27	60.0	
		3	19	60.0	
		4	31	62.0	
		5	30	62.0	

图 3-73

训练后体重	男	最高	1	13	81.0
			2	2	80.0
			3	6	79.0
			4	18	79.0
			5	28	79.0 ^d
		最小值	1	20	65.0
			2	12	65.0
			3	1	70.0
			4	23	71.0
			5	47	74.0 ^e
	女	最高	1	17	75.0
			2	39	68.0
			3	5	66.0
			4	11	65.0
			5	22	65.0 ^f
最小值	1	7	58.0		
	2	34	60.0		
	3	27	60.0		
	4	19	60.0		
	5	15	60.0 ^g		

a. 在较大极值的表中, 仅显示了不完整的个案列表 (这些个案的值为 80.0)。
b. 在较小极值的表中, 仅显示了不完整的个案列表 (这些个案的值为 75.0)。
c. 在较大极值的表中, 仅显示了不完整的个案列表 (这些个案的值为 68.0)。
d. 在较大极值的表中, 仅显示了不完整的个案列表 (这些个案的值为 79.0)。
e. 在较小极值的表中, 仅显示了不完整的个案列表 (这些个案的值为 74.0)。
f. 在较大极值的表中, 仅显示了不完整的个案列表 (这些个案的值为 65.0)。
g. 在较小极值的表中, 仅显示了不完整的个案列表 (这些个案的值为 60.0)。

图 3-74

(5) 正态性检验

图3-75所示提供了“柯尔莫戈洛夫-斯米诺夫”和“夏皮洛-威尔克”两种正态性检验的统计量。如果统计量的显著性概率值 $P>0.05$, 未达到0.05的显著水平, 则接受正态分布的虚无假设, 数据分布呈正态; 反之, 如果 $P<0.05$, 达到0.05的显著水平, 则拒绝正态分布的虚无假设, 数据分布违反正态性。在正态性假设检验中, 如果观测值的总数为大样本, 一般采用“柯尔莫戈洛夫-斯米诺夫”统计量检验, 反之则用“夏皮洛-威尔克”统计量检验。本例参考“柯尔莫戈洛夫-斯米诺夫”统计量检验的结果: 训练前的男生和女生体重显著性概率值均小于0.05, 拒绝虚无假设。而训练后男生和女生的体重均大于0.05, 接受虚无假设, 表明训练后体重的分布呈现正态分布。

正态性检验							
		柯尔莫戈洛夫-斯米诺夫 ^a			夏皮洛-威尔克		
性别		统计	自由度	显著性	统计	自由度	显著性
训练前体重	男	.218	24	.004	.842	24	.002
	女	.214	27	.003	.848	27	.001
训练后体重	男	.198	24	.016	.869	24	.005
	女	.173	27	.037	.848	27	.001

a. 里利氏显著性修正

图 3-75

(6) 体重直方图

图3-76和图3-77是男生训练前后的体重直方图。从二者的比较可以看出,训练后的平均体重低于训练前的平均体重,训练后数据分布整体呈左偏移趋势,分布呈正态性。图3-78和图3-79是女生训练前后的体重直方图。经过对比可以发现,训练后体重的平均值低于训练前,且标准差更小。

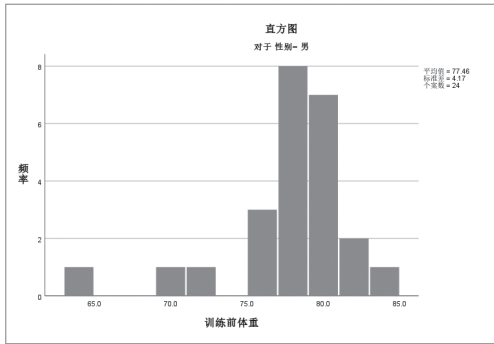


图 3-76

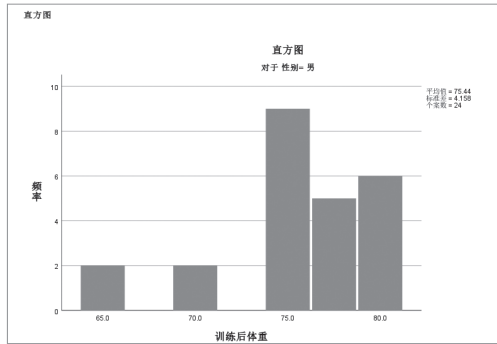


图 3-77

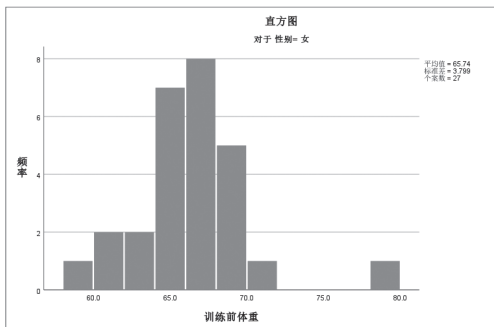


图 3-78

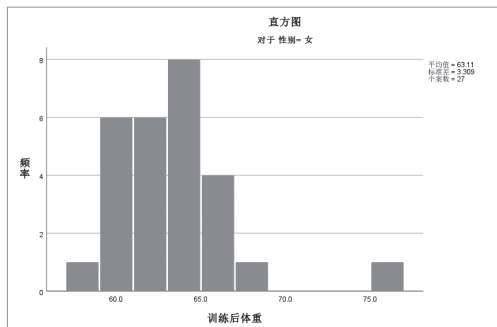


图 3-79

(7) 体重茎叶图

图3-80和图3-81分别为男生训练前和训练后体重茎叶图。主干宽度均为10.0,每片叶代表一个个案。由于茎叶图保留了原始数据,因此可以将其看作一种变形的、被放倒的直方图,从中可以大概看出数据的分布是否呈正态。图3-82和图3-83为女生训练前、后体重茎叶图。在这两张图中分别出现了2个(≥ 71.0),以及1个(≥ 75)的极值。



图 3-80

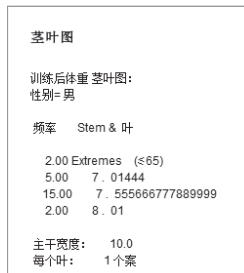


图 3-81

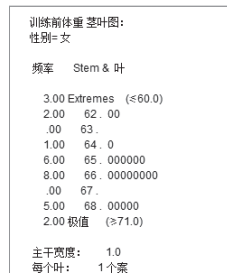


图 3-82

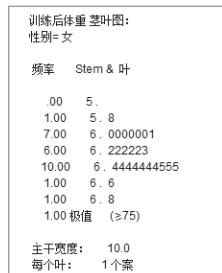


图 3-83

(8) 体重正态Q-Q概率图

图3-84~图3-87分别为男生和女生训练前、后体重正态Q-Q概率图。可以看出，女生训练前后体重数据所组成的实际概率分布线与理论正态累积概率直线重合度更高，呈现较好的正态性，同时也看到，有1个数值明显远离理论正态累积概率直线，故判断为异常值。相较而言，男生组的实际概率分布线重合度要低一些，因而数据分布略呈偏态。这一结论也可以由随后的训练前、后体重的去势正态Q-Q概率图加以佐证。

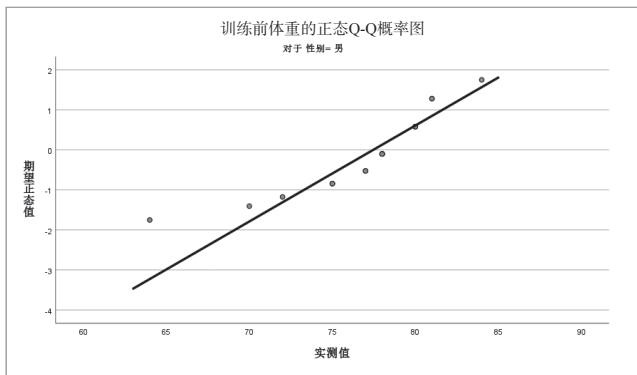


图 3-84

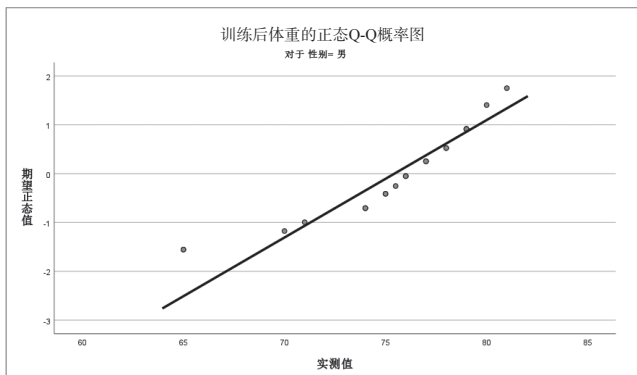


图 3-85

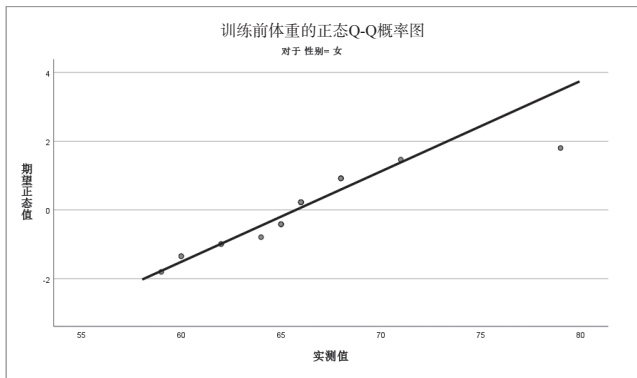


图 3-86

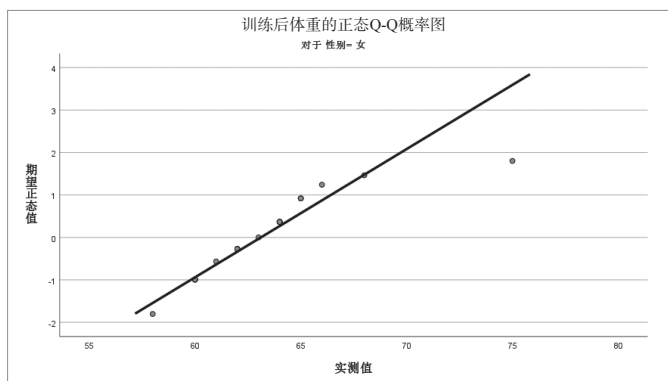


图 3-87

(9) 去势正态Q-Q概率图

图3-88~图3-91是训练前、后体重的去势正态Q-Q概率图，在男生体重去势正态Q-Q概率图中，有较多的数值远离中间的直线而使整体数据分布呈现偏态；相反，女生体重去势正态Q-Q概率图中则有较多的数据分布于直线附近，所以女生体重分布更趋正态，但是仍然有1个数值远远脱离了理论正态累积概率直线而成为异常值。

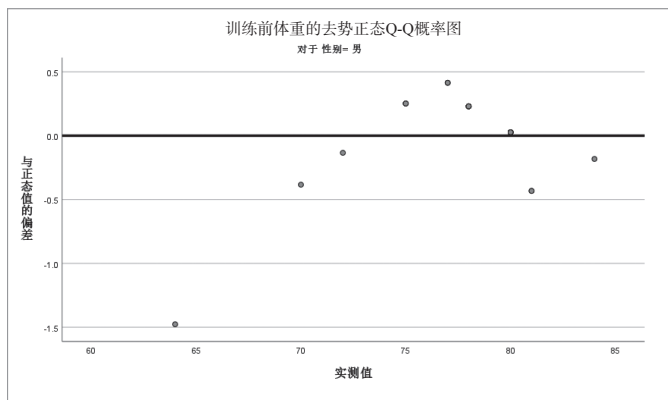


图 3-88

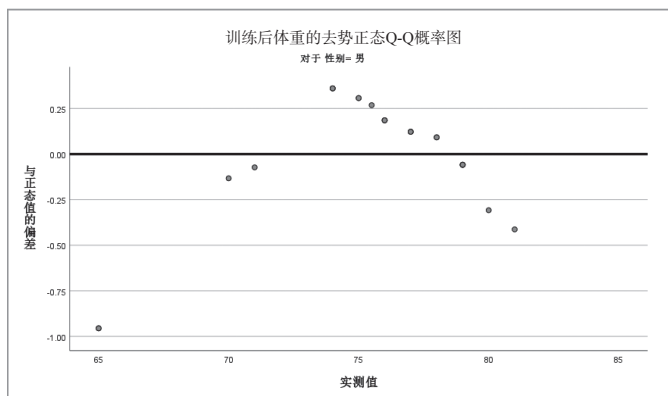


图 3-89

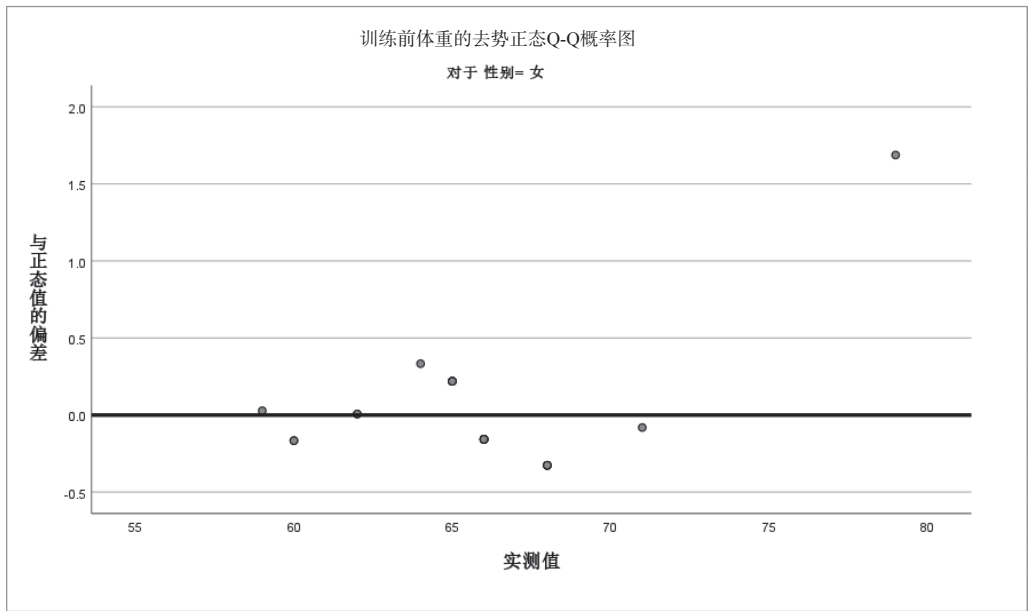


图 3-90

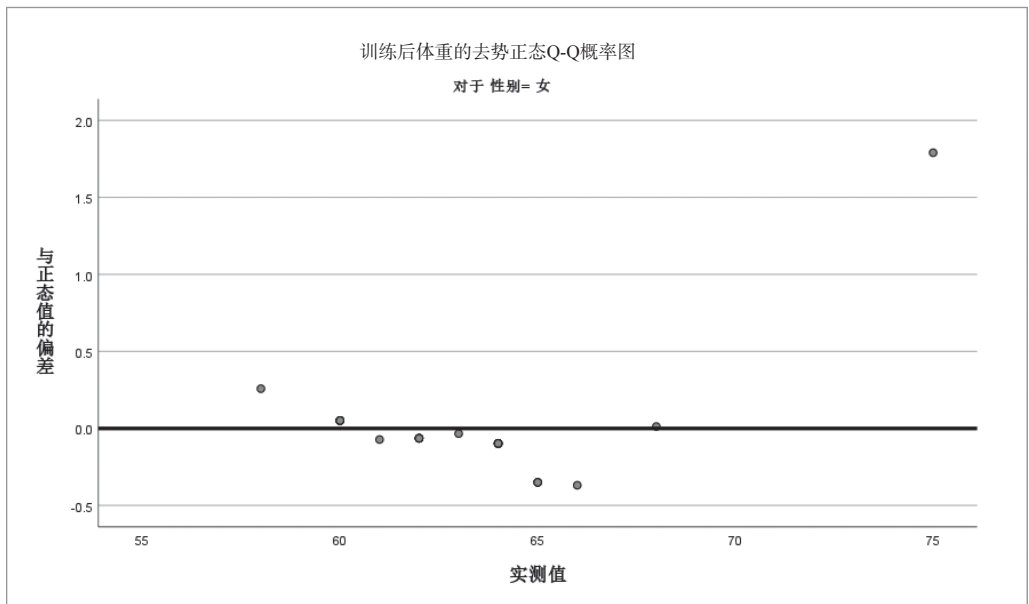


图 3-91

(10) 箱图

在如图3-92和图3-93所示的箱图中可以清楚地看出男生、女生在训练前、后体重数据分布的对比情况。训练前二者的中位数基本位于箱图的中间，但训练后女性的中位数明显偏高，且训练前、后女性最大值和最小值差距较大。

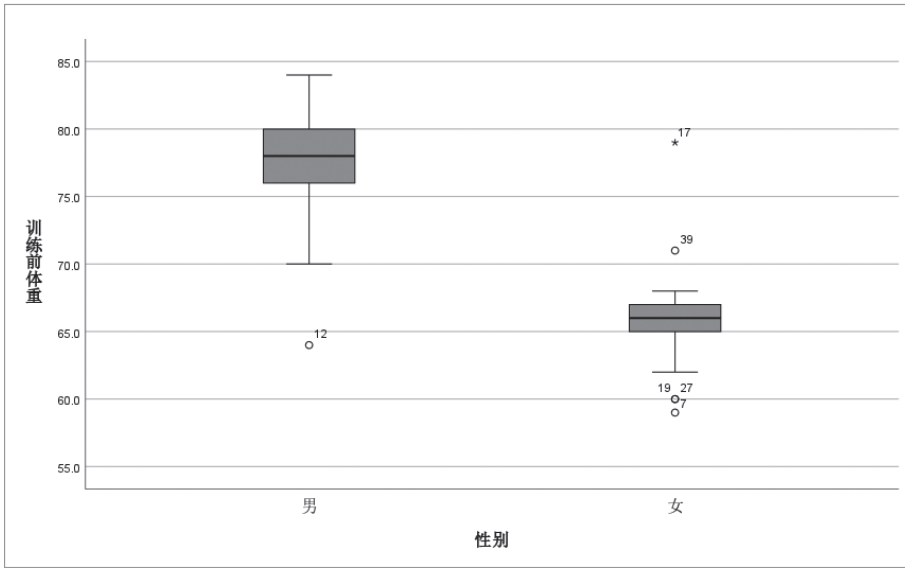


图 3-92

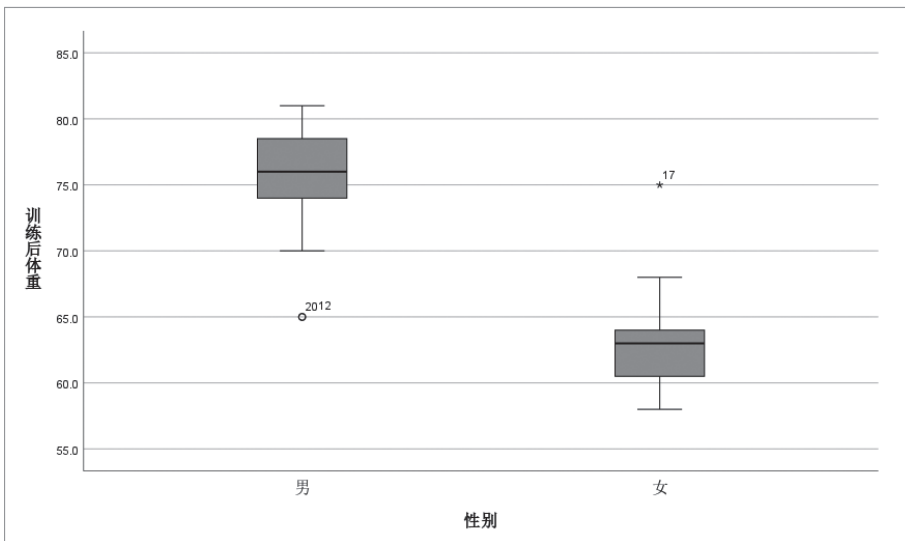


图 3-93