

第 2 章 统计数据的采集和整理



本章主要讲解统计的基本含义及内容, 以及 Excel 在统计数据采集及整理中的具体应用方法。通过本章的学习, 读者应掌握以下内容:

- 统计的基本含义及统计分析的过程和方法
- 统计数据的构成、种类及 Excel 数据清单的操作
- 统计数据整理的内容、方法及 Excel 的数据整理功能

2.1 统计理论概述

2.1.1 统计的意义

“统计”一词, 英语为 **Statistics**, 包括 3 层含义: 统计工作、统计资料和统计学。这三者之间存在着密切的联系, 统计资料是统计工作的成果, 统计学来源于统计工作。

通常所说的统计主要指统计工作, 即统计活动。统计活动是人们认识客观世界的一种认识活动。客观世界的任何事物都有质和量两个方面, 统计活动就是实际做调查研究, 占有大量资料, 从量的方面探讨事物的联系, 从而达到对质的认识。通过统计活动, 可以准确、及时、全面、系统地反映国民经济和社会发展情况, 为制定各种政策和计划提供依据。

统计工作包括 3 个主要阶段: 统计调查、统计整理和统计分析。统计调查是根据调查方案从客观调查对象取得实际统计数据; 统计整理是对统计调查所得的原始数据进行汇总、加工, 使之系统化、条理化; 统计分析则是以条理化的数据为基础, 运用科学的分析方法, 定量分析和定性分析相结合, 对事物的本质和规律作出说明, 从而达到反映和监督社会经济的目的。

2.1.2 统计分析的过程和方法

统计工作的 3 个主要阶段中, 统计分析是最主要、最核心的阶段。统计分析是一个从感性认识到理性认识的抽象思维活动, 是人类认识活动的重要组成部分。

1. 统计分析过程

统计分析的基本过程大致可以分解为以下几个阶段:

(1) 数据准备。这是统计分析的前提工作, 没有充分的、合乎实际的数据, 是无法进行分析的。

(2) 方法选择。无论进行什么样的统计分析, 都必须以适当的统计分析方法支持。

(3) 数据运算。这一阶段实际上是前两个阶段的结合, 其实质是根据一定的目的, 按照一定的方法, 将数据进行一系列的运算, 最后得到能产生某种分析结论的计算结果。随着统计分析水平的提高, 数据运算量越来越大, 数据运算的复杂程度也越来越高, 这使得计算机的应

用成为必要。

(4) 判断预测。这一阶段以数据运算结果为依据,对客观现象的发展规律作出判断,并对现象的发展趋势作出预测。

2. 统计分析方法

统计分析方法的选择决定于对客观事物的数量方面进行分析的需要,而各种方法又用来分析说明客观事物的总体数量特征。客观事物从不同方面、不同角度分析有不同的特征,相应就产生了多种分析方法。统计分析的基本方法主要有以下几类:

(1) 统计描述方法。主要用于特征分析,即通过一些概括性指标来反映数据的全貌和特征。用来描述数据分布特征的概括性指标主要有:①描述集中趋势的指标,如均值、中位数、众数;②反映离散程度的指标,如全距、平均差、方差、标准差;③反映数据在分布中地位的指标,如百分位分数、百分等级分数等;④相关分析指标,当事物之间存在联系但又不能直接做出因果关系的解释时,可用一些合理的指标对相关事物的观测值进行相关分析,其相关程度用相关系数表示,如积差相关、等级相关等。

(2) 统计推论方法。在无法直接估计总体参数的情况下,需要采用抽样方式对样本进行研究,并由样本统计量对事物的总体量做出统计推论和估计。它包括两个方面内容:①总体参数估计,根据样本的数字特征推断总体的相应的数字特征,它又有点估计和区间估计之分;②假设检验,在许多研究中(比如投资决策研究、农作物推广研究等),首先需要提出一个假设(比如谁比谁在什么状态下好或者差或其他),这一假设合理或正确与否,需要抽取样本用其统计量进行检验。通常根据总体是否服从正态分布,将其分为参数检验和非参数检验,前者有Z检验、t检验、 χ^2 检验、方差分析等;后者有中数检验、符号检验、符号秩次检验、U检验、秩次方差分析等。

(3) 多元统计方法。由于客观事物的因素不是单一的,而是多方面的、多层次的、多特征的,因而要分析这些因素之间的各种关系需要用多元统计方法。多元统计分析的基本方法主要有:

1) 回归分析。对于两个具有不确定关系的变量,相关分析可以对这两个变量是否相关做出定性描述,对其相关程度做出总的定量描述,但是如何通过自变量的值去估计和预测因变量的发展变化,相关分析无能为力,这时需要用回归分析。它一般分为一元线性回归和多元线性回归两种。

2) 因素分析。当影响事物性质的变量比较多时,常常需要从中提取出几个主要的因素进行分析,这时就需要使用因素分析法。

3) 聚类分析。也称分类分析或数值分类。即依据变量指标的定量分析对变量实施分类(如果类别已经清楚,只需归类;如果事先并不清楚类别,这时就是寻求一种规则进行新的恰当分类),使同类的变量比较均质,而不同种类的变量差异比较大。

2.2 统计数据的采集

统计是从数量上认识客观世界的科学,数据的质量是统计的生命。统计数据的质量高低直接影响着统计分析的结果。

2.2.1 统计数据的构成

统计中所用的“数”与数学中所用的“数”有很大差别,最显著的差别是:统计中的数有具体的客观事物与之相对应,是有实际意义的;而数学中的数则是抽象的数,不会对应于某

一个具体的客观事物。例如“0”这个数,在数学中它可以说只是一个非正非负的整数,而无实际意义;而如果要将这个数用到统计中,就需要明确表明其实际含义,比如,某个时间的气温为 0°C ,某班学生某星期的迟到人数是0人等。如果数值无实际意义,就不能成为统计的研究对象,也就无法进行统计分析。

统计数据按其来源可以分为原始资料和次级资料(二手资料)两类。原始资料来自直接的调查和科学试验,次级资料来自别人的调查和试验。一般说来,统计数据包括5个组成部分:

(1) 数据名称。

它表明客观事物某一方面的特征。如某企业的“年产值”、“劳动生产率”、“销售利润率”、“库存商品额”、“某职工的性别”等。数据名称必须具备一个条件,即能表现为一定的结果,如年产值为500万元,销售利润率为15%等。如果不能表现为一定的结果,就不能称之为数据名称。

(2) 数据值。

数据值是数据名称的结果表现,如500、15%等。需要指出的是:数据值并不都是数字,也可以是文字,如职工的性别为“男”或“女”。统计分析在大多数情况下处理的是数字数据,但也避免不了处理文字数据。

(3) 计量单位。

计量单位可分为两大类:名数和无名数。名数就是指计量单位有具体的名称,包括实物计量单位(如台、件、人、千克、米等)、货币计量单位(如元、万元、美元等)和劳动计量单位(如工时、工日、台时等)。无名数只有抽象的名称或无名称,通常采用的有系数、倍数、成数、百分数、千分数等。

(4) 时间范围。

任何数据都表明客观事物在特定时间条件下的特征,离开了时间限制,统计数据就无任何说服力。比如某高中学生的升学率为90%,而另一高中学生的升学率为50%,若要比两所学校的教学质量,就必须明确升学率是哪一年的,如果没有时间上的统一,就会出现不可比的情况。

根据现象所属的时间不同,数据有时期数据和时点数据之分。时期数据表明某一段时期(如年、月)内容观现象的数量表现,如某企业2011年的生产总值为1000万元,某商场1月份的销售额为500万元等。时期数据值的大小与时期长短有直接关系,并且具有时间上的可累加性,如某企业两年的产值要大于一年的产值,而一年的产值等于各月产值之和。时点数据表明某一时间点(如年末、月初)上客观现象的数量表现,如某地区2011年末人口总数为2000万人,某企业2011年6月末产成品库存量为20万吨等。时点数据值的大小与时间长短没有直接关系,并且不具有时间上的可累加性,如某企业年末产成品库存量不见得比每个月末的库存量高,也不等于各月末库存量之和。

(5) 空间范围。

为保证统计数据意义的完整,空间限制必不可少。如“2011年国民收入50亿元”,这是一个县的国民收入呢?还是一个省呢?必须加以说明。

一般情况下,一个完整的统计数据应具备以上5个组成部分。如“大华公司2011年的销售总额为3000万元”这一数据,“大华公司”是空间范围,“2011年”是时间范围,“销售总额”是数据名称,“3000”是数据值,“万元”是计量单位。但是,文字数据并不一定全部具备上述5个条件,比如某职工性别为“男”,就没有时间和计量单位。

2.2.2 统计数据的种类

在采集数据的过程中,会有各种各样的数据,如产值、销售额、班级人数、年龄、性别、民族、宗教等。对于统计数据,可以有不同的分类方式。

1. 按照数据值的表现形式不同分类

(1) 属性数据(定性数据)。属性数据说明现象属性方面的特征,如性别、民族、宗教等。属性数据的数据值用文字表示,如某人的性别是男、民族是汉族等。

(2) 数值数据(定量数据)。数值数据说明现象数量方面的特征,如产值、销售额、班级人数、年龄等。数值数据的数据值用数字表示,如某人的身高是170cm、体重是70kg等。

数值数据又可以分为离散数据和连续数据两种。离散数据是指数据值之间都是以整数断开的,两个相邻的数据值之间不可能有小数连接的数据,如班级人数;连续数据是指数据值可以连续不断,两个相邻的整数之间可以用无限个小数连接起来的数据,如产值、销售额、年龄等。所以从表现形式上看,离散数据不能用小数表示,而连续数据则可以用小数表示。

2. 按照取得数据时的计量方式不同分类

(1) 测量值数据。凡用量具测量得出的数据称为测量值数据。如人的身高、体重,土地的面积、大气湿度等。这类数据具有连续数据的特征,即相邻整数之间可以有无限个小数连接。

(2) 计数值数据。凡是以清点方式得出的数据称为计数值数据。如某班学生人数、某医院的病床数、某企业拥有的机器设备台数等。这类数据具有与离散数据相同的特征,即相邻整数之间不能有小数,只能用整数表示。

(3) 分类数据。凡是通过现象分类所获得的数据称为分类数据。分类数据只能说明现象所属的类别,它没有数量大小,也没有先后顺序,如性别、民族、宗教等都是按照分类方式取得的数据。在具体应用时,虽然可以将类别用不同的数字符号来表示,如用“1”表示男性,用“2”表示女性,但这些数字仅仅是一个符号,没有大小和顺序的差别。

(4) 排序数据。凡是通过排序方式获得的数据称为排序数据。排序数据只能用顺序等次来说明现象的特征,它没有数量大小。排序数据有顺序的意义,如考试成绩名次、顾客对某商品的满意度等。排序数据虽然有排序的意义,但无法精确说明数据的差距,如第一名的成绩是95分,第二名的成绩是90分,第三名的成绩是80分,第一、二、三名只能说明成绩的好坏,但不能说明名次之间的具体分数差距。

2.2.3 数据清单

在Excel中,数据是以数据清单的形式存储的。数据清单由工作表中一系列连续单元格的数据集合组成,实际上也是一个工作表,但有不同于一般工作表的特点。数据清单中的数据既可以通过直接输入产生,也可以通过调用数据文件取得,还可以由公式或一般数据快速填充方式产生新的数据。在数据清单中,可以实现数据的复制、移动、插入、删除、排序、筛选等灵活操作。

1. 数据清单结构

创建数据清单时,以每一列为一个字段(Field),以每一行为一个记录(Record),一般以数据清单的第一行创建字段(变量)名。在一张工作表中,只能建立一个数据清单。数据清单中的某些命令,如筛选、排序等,每次只能在一个数据清单中使用。数据清单中应避免出现空

白行或列。数据清单的形式如图 2-1 所示。

	A	B	C	D	E	F	G	H
1	学号	姓名	性别	年龄	财务会计	审计原理	统计学	体育
2	62301	高蕊	女	19	90	85	64	90
3	62302	李毅	男	19	65	60	85	95
4	62303	蒋霄亮	男	19	65	70	80	65
5	62304	王东洲	男	19	82	70	92	70
6	62305	董琳	女	18	50	90	78	88
7	62306	冯妍	女	18	60	70	80	88
8	62307	张萍	女	19	88	70	85	88
9	62308	余敏	女	18	65	80	82	80
10	62309	邱美芳	女	19	85	90	95	75
11	62310	邱万雄	男	18	80	80	86	85
12	62311	林静	男	18	75	90	60	80
13	62312	张媛	女	19	75	70	78	65
14	62313	董玉锋	男	19	90	60	75	50

图 2-1 数据清单

2. 录入数据

数据的录入有 3 种基本方法:

- (1) 单击目标单元格直接输入数据。
- (2) 双击目标单元格输入数据。这种方法多用于修改单元格中的数据。
- (3) 单击目标单元格, 在编辑栏中输入数据。在 Excel 的编辑栏中, 可以复制、粘贴进行编辑或修改数据。如果一个单元格需要录入很多文字或需要输入公式时, 常采用这种方式。

3. 数据自动填充

输入数据或公式时, 如果输入的数据或公式具有一定的规律性, 可以不必一项一项手工输入, 而是利用 Excel 的自动填充功能完成批量输入。这既减少了数据输入的工作量, 又减少了错误输入的概率。数据的自动填充可以采用“填充柄”(见第 1 章内容)、Ctrl+Enter 组合键或“填充”命令来完成。

例 2-1 利用 Ctrl+Enter 组合键在单元格区域 A1:A8、B1:F1、C3:D8、F4:F8 中输入数据 100。

- (1) 选定单元格区域 A1:A8、B1:F1、C3:D8、F4:F8。首先选定单元格 A1, 在按住 Ctrl 键继续选取其他单元格区域。
- (2) 输入数据 100。
- (3) 按 Ctrl+Enter 组合键, 结果如图 2-2 所示。

	A	B	C	D	E	F
1	100	100	100	100	100	100
2	100					
3	100		100	100		
4	100		100	100		100
5	100		100	100		100
6	100		100	100		100
7	100		100	100		100
8	100		100	100		100

图 2-2 按 Ctrl+Enter 组合键自动填充

利用“填充”命令进行自动填充时, 首先要在一个单元格中输入第一个数据, 然后选定一个单元格区域, 使输入的第一个数据位于单元格区域的顶行、底行、最左边或最右边, 在“开始”选项卡中的“编辑”组中单击“填充”按钮, 在下拉菜单中选择“向下”、“向上”、“向左”或“向右”命令, 可以完成行或列的填充。以上操作也可以通过选取“填充”→“系列”命令,

在弹出的“序列”对话框中选中“自动填充”单选按钮来完成。通过“序列”对话框，还可以完成等差数列、等比数列的填充。

例 2-2 利用“序列”对话框在 A1:F1 区域填充数据，初始值为 1，比值为 2。

(1) 在单元格 A1 中输入 1，选取单元格区域 A1:F1。

(2) 在“开始”选项卡中“编辑”组中单击“填充”按钮，在下拉菜单中选择“系列”命令，弹出“序列”对话框，如图 2-3 所示。

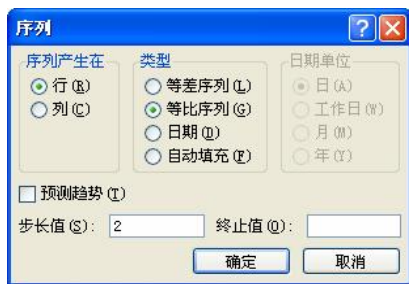


图 2-3 “序列”对话框


(3) 在其中的“类型”选项区中选中“等比序列”单选按钮，“步长值”设为 2，单击“确定”按钮，结果如图 2-4 所示。

	A	B	C	D	E	F
1	1	2	4	8	16	32
2						
3						
4						
5						

图 2-4 利用“序列”对话框产生等比数列

4. 利用“记录单”编辑数据

比较简单的数据清单，数据的编辑可以直接在数据清单中操作。如果数据清单中的数据非常复杂，就可以利用“记录单”数据的操作。数据“记录单”具有浏览记录、添加记录、修改记录和删除记录等功能。使用“记录单”命令，可以在数据清单中一次输入、显示、查找或删除一行完整记录。

在打开的 Excel 工作簿中单击“文件”按钮打开后台视图，然后单击“选项”按钮。在打开的“Excel 选项”对话框中切换到“快速访问工具栏”选项卡。在“从下列位置选择命令”下拉列表框中选择“不在功能区中的命令”选项，随后找到“记录单”命令将其添加到“自定义快速访问工具栏”中。此时就可以在“快速访问工具栏”中找到“记录单”按钮.

当需要输入数据时，单击记录单按钮，在打开的“数据清单”对话框中即可轻松地输入数据，如图 2-5 所示。单击“新建”按钮在相应的文本框中输入数据可以添加新的记录，还可以删除查找逐条浏览相应记录。

对话框的右上角的分数表示当前记录在数据清单中的位置以及数据清单总共记录数。如 1/13 表示该数据清单中共有 13 条记录，当前记录为第 1 个记录。在“数据清单”对话框中，可以完成以下操作：

(1) 浏览记录。通过单击“上一条”、“下一条”按钮或拖动对话框中间的垂直滚动条，可以快速浏览记录内容。

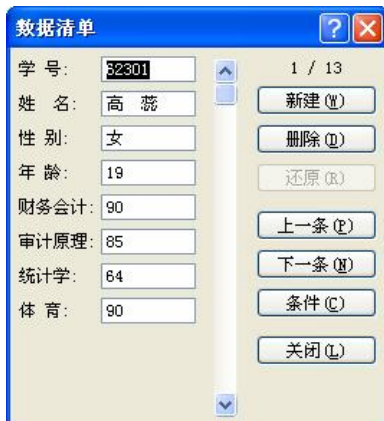


图 2-5 “数据清单”对话框

(2) 浏览符合条件的记录。单击对话框中的“条件”按钮，记录框数据区域将变成空白。如果这时在字段框内输入相应条件，就可以查找符合条件的记录。条件既可以是一个，也可以是多个；既可以在一个字段框中输入条件，也可以同时在多个字段框中输入多个条件。例如，查找学号为 62307，性别为女的记录，首先单击“条件”按钮，在“学号”文本框内输入 62307，在“性别”文本框内输入“女”，按 Enter 键后会显示符合条件的记录。

(3) 添加记录。单击“新建”按钮，对话框中的文本框里面变成空白，逐一输入相应内容后按 Enter 键完成一条记录的添加。在添加记录的过程中，如果单击“还原”按钮，则可将当前录入的内容全部清除。

(4) 修改记录。在浏览记录内容时，可以对字段内的记录进行修改，修改完毕后单击“新建”、“关闭”按钮或按 Enter 键，系统将接受修改，单击其他按钮将放弃所作的修改。

(5) 删除记录。在浏览记录时，单击“删除”按钮，系统会提示是否删除，单击“确定”按钮将删除正在显示的当前记录。

2.3 统计数据的整理

2.3.1 统计数据整理的内容

统计数据的整理是对调查搜集到的大量原始数据进行科学的分类、汇总，使之成为系统化、条理化、标准化，能反映总体现象特征的综合统计资料的工作过程。通过各种方式取得的存储在数据清单中的原始数据是杂乱、无序排放的，不能反映现象的本质与规律性，这就需要对其采用专门的方法进行整理。

一般来讲，统计整理的内容通常包括以下几个方面：

- 根据分析的目的对数据进行排序和分类（组）。
- 对分类后的数据进行汇总，计算各类（组）及总体的指标。
- 通过统计表或统计图描述汇总的结果。

Excel 提供了多种数据整理工具，主要有：



- 数据排序和筛选。
- 频数分布函数。

- 数据透视表。
- 统计图。
- 直方图分析工具。

2.3.2 统计数据的排序与筛选

1. 统计数据排序

数据的排序是以数据清单中的一个或几个字段为关键字，对整个数据清单的所有个体进行重新排列。排序可以按升序，也可以按降序。对于数字型字段，排序是按数值的大小；对于字符型字段，排序是按 ASCII 码大小；中文字段按拼音或笔画排序。通过排序，可以清楚地反映数据之间的大小关系。

(1) 按单字段排序。数据的排序，可以用“数据”功能区“排序和筛选”组中的“升序排序”按钮  和“降序排序”按钮  来进行。要进行数据清单中某一字段数据的排序，首先单击该字段下的任一单元格（不能单击该字段的列标，否则将只对该列数据排序），再单击升序排序按钮或降序排序按钮即可完成本字段的升序排列或降序排列。这种排序方式具有快捷方便的优点。

(2) 按多字段排序。使用升降序按钮排序虽然操作简便，但只能按单个字段的内容进行排序，不能满足复杂的排序要求。如果需要按多个字段进行排序，就需要采用“数据”功能区“排序和筛选”组中的“排序”按钮进行。具体操作步骤如下：

1) 选定数据清单中的任一单元格为当前单元格。

2) 单击“数据”功能区“排序和筛选”组中的“排序”按钮，弹出“排序”对话框，如图 2-6 所示。



图 2-6 “排序”对话框

3) 在“排序”对话框中可以按多个变量进行多字段排序。整个数据清单要先按主要关键字排序，关键字相同者排在一起；若指定了次要关键字，则主要关键字相同者再按次要关键字排序；若指定了第三关键字，依此类推。单击“添加条件”按钮，在出现的次要关键字中选择次要关键字，若需要添加多个次要关键字，则继续单击“添加条件”按钮。

例 2-3 6 个企业 2000 年的主要财务指标如图 2-7 所示。

1) 以“主营业务收入为关键字”，按升序排列以上数据。选中“主营业务收入”字段下的任一单元格，再单击工具栏上的“升序排序”按钮，排序结果如图 2-8 所示。

2) 以“主营业务收入”为主要关键字，“主营业务利润”为次要关键字，“净利润”为第三关键字，对图 2-7 中的数据全部进行降序排序。

	A	B	C	D	E	F
1	编号	企业名称	主营业务收入	主营业务利润	管理费用	净利润
2	1001	同仁堂	102439	46408	13384	14665
3	1002	西藏药业	8312	6926	1325	3572
4	1003	云南白药	80641	23838	7171	4922
5	2001	首钢	1191658	131116	17526	75873
6	2002	宝钢	3094053	673436	132559	299210
7	2003	邯钢	585668	99292	4590	75261
8						

图 2-7 企业财务数据

	A	B	C	D	E	F
1	编号	企业名称	主营业务收入	主营业务利润	管理费用	净利润
2	1002	西藏药业	8312	6926	1325	3572
3	1003	云南白药	80641	23838	7171	4922
4	1001	同仁堂	102439	46408	13384	14665
5	2003	邯钢	585668	99292	4590	75261
6	2001	首钢	1191658	131116	17526	75873
7	2002	宝钢	3094053	673436	132559	299210
8						

图 2-8 数据按“主营业务收入”排序

选中数据区中的任一单元格，在“数据”功能区的“排序和筛选”组中单击“排序”按钮，弹出“排序”对话框，单击“主要关键字”右面的下拉按钮，在下拉列表框中选择“主营业务收入”，“排序依据”选择“数值”，“次序”选择“降序”；单击“添加条件”按钮，在出现的次要关键字中选择“主营业务利润”，“排序依据”选择“数值”，“次序”选择“降序”；再单击“添加条件”按钮，在出现的次要关键字中选择“净利润”，“排序依据”选择“数值”，“次序”选择“降序”，如图 2-9 所示。



图 2-9 设置关键字及排序方式

如果需要进一步设置，可单击“序列”对话框中的“选项”按钮，在弹出的“排序选项”对话框中进行详细设置，如图 2-10 所示。



图 2-10 “排序选项”对话框

在“排序”对话框中，单击“确定”按钮完成数据排序，排序结果如图 2-11 所示。

	A	B	C	D	E	F
1	编号	企业名称	主营业务收入	主营业务利润	管理费用	净利润
2	2002	宝钢	3094053	673436	132559	299210
3	2001	首钢	1191658	131116	17526	75873
4	2003	邯钢	585668	99292	4590	75261
5	1001	同仁堂	102439	46408	13384	14665
6	1003	云南白药	80641	23838	7171	4922
7	1002	西藏药业	8312	6926	1325	3572
8						

图 2-11 数据按多字段排序

2. 统计数据的筛选

利用 Excel 提供的筛选功能可以把符合要求的数据集中在一起，把不符合要求的数据隐藏起来。数据的筛选包括自动筛选和高级筛选两项功能。

(1) 自动筛选。自动筛选是一种快速的筛选方法，它可以方便地将满足条件的数据显示在工作表上，将不满足条件的数据隐藏起来。使用自动筛选的步骤如下：

- 1) 在数据清单中选择任一单元格为当前单元格。
- 2) 单击“数据”功能区“排序和筛选”组中的“筛选”按钮，会看到在数据清单中的每一列字段名旁都会出现一个下拉箭头按钮，如图 2-12 所示。

	A	B	C	D	E	F
1	编号	企业名称	主营业务收入	主营业务利润	管理费用	净利润
2	1001	同仁堂	102439	46408	13384	14665
3	1002	西藏药业	8312	6926	1325	3572
4	1003	云南白药	80641	23838	7171	4922
5	2001	首钢	1191658	131116	17526	75873
6	2002	宝钢	3094053	673436	132559	299210
7	2003	邯钢	585668	99292	4590	75261
8						

图 2-12 打开“自动筛选”功能

3) 单击某一个下拉按钮，在下拉列表框中选定筛选的条件。筛选条件根据单元格中的数据类型有“文本筛选”和“数字筛选”。“文本筛选”条件包括“等于...”、“不等于...”、“开头是...”、“结尾是...”、“包含...”、“不包含...”、“自定义筛选...”，如图 2-13 所示；“数字筛选”条件包括“等于...”、“不等于...”、“大于...”、“大于或等于...”、“小于...”、“小于或等于...”、“介于...”、“10 个最大的值...”、“高于平均值”、“低于平均值”、“自定义筛选...”，如图 2-14 所示。



图 2-13 “文本筛选”条件



图 2-14 “数字筛选条件”

“数字筛选”条件中,“10 个最大的值...”筛选的数据不一定是最大的数字,也可以是最小的数字;数字个数不一定只是 10 个,可以根据需要设定需要筛选的数据个数。方法是单击“前 10 个...”选项,弹出“自动筛选前 10 个”对话框,如图 2-15 所示。在对话框中可以设定筛选的条件为最大或最小,以及筛选值的项数(或百分比)。



图 2-15 “自动筛选前 10 个”对话框

“自定义筛选...”选项适用于较为复杂的筛选条件。单击“自定义筛选...”选项可以弹出“自定义自动筛选方式”对话框,如图 2-16 所示。在对话框中可以设置两组筛选条件。每一组筛选条件由关系运算符构成条件表达式,其中左侧运算符有等于、不等于、大于、大于或等于、小于、小于或等于、开头是、开头不是、结尾是、结尾不是、包含、不包含等。在对话框右侧,可以单击下拉按钮选择值,也可以直接输入数据。两组条件之间可以是“与”或者是“或”的关系。



图 2-16 “自定义自动筛选方式”对话框

4) 设定好筛选条件后,单击对话框中的“确定”按钮,Excel 将在工作表中显示符合条件的记录。

例 2-4 以例 2-3 的资料,显示“主营业务收入”大于 100000 小于 1000000 的数据。

- 1) 选中数据清单中的任一单元格。
- 2) 单击“数据”功能区“排序和筛选”组中的“筛选”按钮。
- 3) 单击“主营业务收入”字段名右边的下拉按钮,在弹出的下拉列表框中选定“自定义筛选”选项。
- 4) 在“自定义自动筛选方式”对话框中,第一个筛选条件设定为大于 100000,第二个筛选条件设定为小于 1000000 元,两者的关系是“与”,如图 2-17 所示。

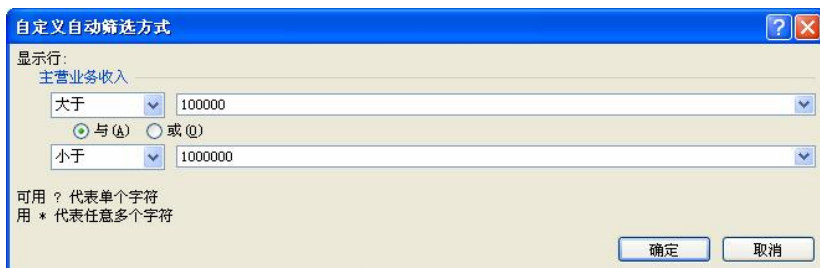


图 2-17 设定筛选条件

5) 单击“确定”按钮,工作表中将显示筛选后的数据,如图 2-18 所示。

完成自动筛选后,单击“数据”功能区“排序和筛选”组中的“筛选”按钮,将退出自

动筛选状态，字段名旁的下拉按钮同时消失。

	A	B	C	D	E	F
1	编 号	企业名称	主营业务收入	主营业务利润	管理费用	净利润
2	1001	同仁堂	102439	46408	13384	14665
7	2003	邯钢	585668	99292	4590	75261
8						
9						
10						
11						

图 2-18 自定义筛选结果

(2) 高级筛选。自动筛选只能适用于比较简单的条件，如果需要指定的筛选条件比较多，就需要使用 Excel 的高级筛选功能。

高级筛选的关键是条件区域的设定。通常是将条件区域放在整个数据清单的下边（以防止被筛选隐含，并且不改动数据清单的位置），至少要用一个空行隔开。条件区域的第 1 行为字段名，第 2 行及以下各行为条件值。同一行条件之间为“与”的关系，不同行条件之间为“或”的关系，可采用的条件符号有 >、<、≥、≤。

设置好条件区域后，选定数据清单中的任一单元格，单击“数据”功能区“排序和筛选”组中的“高级”按钮，在弹出的“高级筛选”对话框中根据需要选择显示筛选结果的方式，并分别指定数据清单（列表区域）和条件区域所在单元格的位置，单击“确定”按钮即完成高级筛选，如图 2-19 所示。

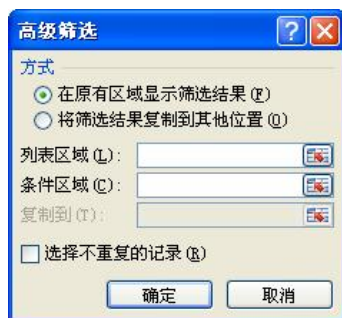


图 2-19 “高级筛选”对话框

例 2-5 以例 2-3 的资料，查找“主营业务收入”大于 1000000，“主营业务利润”大于 100000，“净利润”大于 10000 的数据。

1) 设定条件区域。将单元格区域 A1:F1 中的数据复制到单元格区域 A9:F9，在单元格 C10、D10、F10 中分别输入条件“>1000000”、“>100000”、“>10000”，如图 2-20 所示。

	A	B	C	D	E	F
1	编 号	企业名称	主营业务收入	主营业务利润	管理费用	净利润
2	1001	同仁堂	102439	46408	13384	14665
3	1002	西藏药业	8312	6926	1325	3572
4	1003	云南白药	80641	23838	7171	4922
5	2001	首钢	1191658	131116	17526	75873
6	2002	宝钢	3094053	673436	132559	299210
7	2003	邯钢	585668	99292	4590	75261
8						
9	编 号	企业名称	主营业务收入	主营业务利润	管理费用	净利润
10			>1000000	>100000		>10000

图 2-20 设定条件区域

2) 单击“数据”功能区“排序和筛选”组中的“高级”按钮,弹出“高级筛选”对话框。

3) 在“列表区域”中输入\$A\$1:\$F\$7,在“条件区域”中输入\$A\$9:\$F\$10,“方式”选择“在原有区域显示筛选结果”,如图 2-21 所示。



图 2-21 输入列表区域和条件区域

4) 单击“确定”按钮,显示筛选结果如图 2-22 所示。

	A	B	C	D	E	F
1	编号	企业名称	主营业务收入	主营业务利润	管理费用	净利润
5	2001	首钢	1191658	131116	17526	75873
6	2002	宝钢	3094053	673436	132559	299210
8						
9	编号	企业名称	主营业务收入	主营业务利润	管理费用	净利润
10			>1000000	>100000		>10000

图 2-22 高级筛选结果

如果想将筛选的结果复制到其他位置,可以在“高级筛选”对话框中的“方式”中选择“将筛选结果复制到其他位置”,激活“条件区域”下面的“复制到”选项,输入指定位置左上角的单元格行列号,单击“确定”按钮。

如果对重复记录不想全部显示,可以选中对话框中的“选择不重复的记录”复选框,则如果有重复记录,就只显示一个了。

2.3.3 统计数据分类汇总

为了使数据清单的内容更加清晰、明确,可以利用 Excel 的分类汇总功能将数据归类,并进行求和、均值等计算,并将计算结果显示出来,以便对数据作进一步分析。

例 2-6 某班 10 个学生的“成本会计”课程考试成绩如图 2-23 所示,要求统计不同分数学生的人数。

	A	B	C
1	学生编号	成绩	
2	1	82	
3	2	95	
4	3	91	
5	4	65	
6	5	74	
7	6	65	
8	7	78	
9	8	85	
10	9	51	
11	10	74	

图 2-23 学生成绩

(1) 对“成绩”字段按升序进行排序，如图 2-24 所示。

	A	B	C
1	学生编号	成绩	
2	9	51	
3	4	65	
4	6	65	
5	5	74	
6	10	74	
7	7	78	
8	1	82	
9	8	85	
10	3	91	
11	2	95	

图 2-24 按“成绩”的升序排序

(2) 单击数据清单中的任一单元格，在“数据”功能区的“分级显示”组中选择“分类汇总”命令，弹出“分类汇总”对话框。

(3) 在“分类汇总”对话框中，“分类字段”选择“成绩”，“汇总方式”选择“计数”，“选定汇总项”选择“学生编号”，单击“确定”按钮。可以看到系统已经对分数进行了计数汇总，如图 2-25 所示。

1	2	3	A	B	C
	1		学生编号	成绩	
+	3		1	51	计数
+	6		2	65	计数
+	9		2	74	计数
+	11		1	78	计数
+	13		1	82	计数
+	15		1	85	计数
+	17		1	91	计数
+	19		1	95	计数
-	20		10		总计数

图 2-25 对“成绩”分类汇总

可以看出，成绩为 65 分和 74 分的学生各有两人，其余分数各有一人。

单纯利用排序与分类汇总还不能很好地描述数据的分布状态，为此，Excel 提供了一个频数分布函数 (FREQUENCY)，利用它可以对数据进行分组，建立频数分布，从而更好地描述数据分布状态。该函数以一系列垂直数组返回某个区域中数据的频率分布。例如，使用函数 FREQUENCY 可以计算在给定的分数范围内测验分数的个数。

语法：FREQUENCY(data_array,bins_array)

其中，data_array 为一数组或对一组数值的引用，用来计算频率。如果 data_array 中不包含任何数值，函数 FREQUENCY 返回零数组。bins_array 为间隔的数组或对间隔的引用，该间隔用于对 data_array 中的数值进行分组。如果 bins_array 中不包含任何数值，函数 FREQUENCY 返回 data_array 中元素的个数。


在选定相邻单元格区域（该区域用于显示返回的分布结果）后，函数 FREQUENCY 应以数组公式的形式输入。

返回的数组中的元素个数比 bins_array（数组）中的元素个数多 1。返回的数组中所多出来的元素表示超出最高间隔的数值个数。例如，如果要计算输入到 3 个单元格中的 3 个数值区间（间隔），一定要在 4 个单元格中输入 FREQUENCY 函数计算的结果。多出来的单元格将返回 data_array 中大于第三个间隔值的数值个数。

例 2-7 利用例 2-6 中排序后的资料 (图 2-24), 分别统计 60 分以下、60~69 分、70~79 分、80~89 分、90 分 (含) 以上的学生人数。

(1) 分别在单元格 C1、D1 中输入“分组”、“频数”字样。

(2) 在单元格区域 C2:C6 中分别输入 60、70、80、90、100, 分别表示分数在 60 分以下、60 分以上但在 70 分以下, ……。

(3) 选定单元格区域 D2:D6, 单击编辑栏左边的“插入函数”按钮 , 弹出“插入函数”对话框, 在“函数分类”列表中选择“统计”, 在“函数名”列表中选择 FREQUENCY, 单击“确定”按钮, 弹出“函数参数”对话框, 如图 2-26 所示。

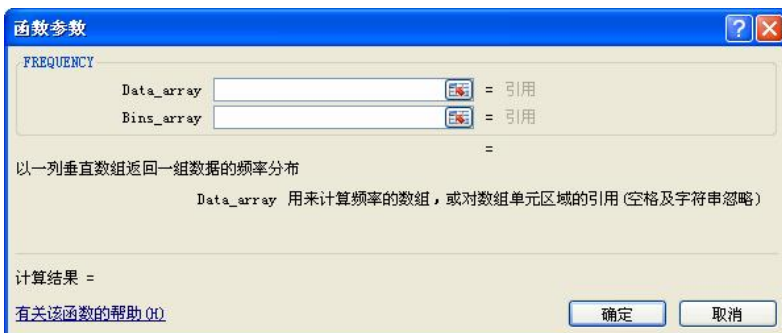


图 2-26 “函数参数”对话框

(4) 在数据区域 data_array 中输入 B2:B11, 在数据接受区域 bins_array 中输入 C2:C6。由于频数分布是数组操作, 所以不能直接单击“确定”按钮, 而应按 Ctrl+Shift 组合键, 同时按 Enter 键, 得到频数分布结果, 如图 2-27 所示。

	A	B	C	D	E
1	学生编号	成绩	分组	频数	
2	9	51	60	1	
3	4	65	70	2	
4	6	65	80	3	
5	5	74	90	2	
6	10	74	100	2	
7	7	78			
8	1	82			
9	8	85			
10	3	91			
11	2	95			

图 2-27 频数分布结果

可以看到, 各分数段的人数分别为 1、2、3、2、2。

2.3.4 统计数据的透视分析

Excel 2010 提供了数据透视表和数据透视图功能。数据透视表将排序、筛选和分类汇总功能结合起来, 对数据清单或外来的数据重新组织和计算, 并以多种不同的形式显示出来。数据透视图是另一种数据表现形式, 与数据透视表不同的是, 它利用适当的图表和多种色彩来描述数据的特征。利用数据透视图, 可以更直观地显示数据。与标准图表一样, 数据透视图显示数据系列、类别、数据标记和坐标轴。用户还可以更改图表类型及其他选项, 如标题、图例、位置、数据标签和图表位置等。

1. 数据透视表

例 2-8 某地区 3 个商场 6 月份彩电和空调的销售数据清单如图 2-28 所示，通过数据透视表比较不同商场及不同厂家的不同产品的销售情况。

	A	B	C	D
1	商场	商品种类	生产厂家	销售量(台)
2	明珠	彩电	长虹	420
3	明珠	彩电	海尔	380
4	明珠	彩电	TCL	320
5	明珠	空调	长虹	450
6	明珠	空调	海尔	410
7	明珠	空调	TCL	390
8	京客隆	彩电	长虹	340
9	京客隆	彩电	海尔	300
10	京客隆	彩电	TCL	320
11	京客隆	空调	长虹	150
12	京客隆	空调	海尔	210
13	京客隆	空调	TCL	120
14	华联	彩电	长虹	240
15	华联	彩电	海尔	230
16	华联	彩电	TCL	210
17	华联	空调	长虹	110
18	华联	空调	海尔	230
19	华联	空调	TCL	130

图 2-28 商品销售资料

(1) 在工作表中，选取任一单元格，在“插入”选项卡的“表格”组中单击“数据透视表”按钮，在展开的下拉菜单中选择“数据透视表”命令，弹出“创建数据透视表”对话框，选择默认选项，单击“确定”按钮，生成空白数据透视表，如图 2-29 所示。

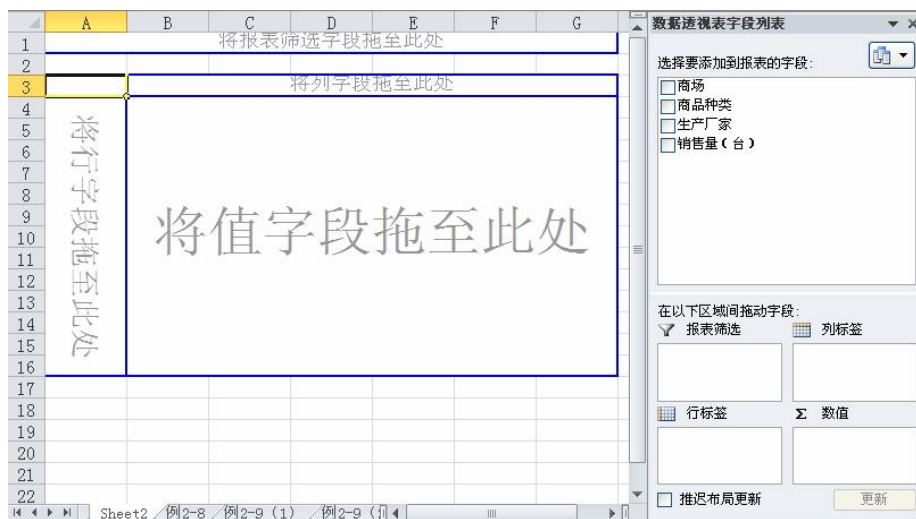


图 2-29 空白数据透视表

(2) 在“数据透视表字段列表”中选“商场”、“商品种类”并用鼠标拖到“在以下区域间拖动字段”区域的“行标签”中，把“生产厂家”拖到“列标签”中，把“销售量(台)”拖到“数值”中。生成的数据透视表如图 2-30 所示。

为使数据透视表更加美观，可以使用“数据透视表工具”——“设计”选项卡中的功能对透视表进行修饰。“数据透视表工具”——“设计”选项卡的“数据透视表样式”组中提供了多种样式，用户可以选择合适的格式修饰数据透视表。例如，对图 2-30 所示的数据透视表，

在“数据透视表样式”组中选择“中等深浅3”，生成的格式如图2-31所示。当然，用户也可以通过操作单元格设计自己喜欢的格式，这样更能满足用户的某些特殊需要，并且符合使用习惯。

求和项:销售量(台)		生产厂家			
商场	商品种类	TCL	长虹	海尔	总计
华联	彩电	210	240	230	680
	空调	130	110	230	470
华联 汇总		340	350	460	1150
京客隆	彩电	320	340	300	960
	空调	120	150	210	480
京客隆 汇总		440	490	510	1440
明珠	彩电	320	420	380	1120
	空调	390	450	410	1250
明珠 汇总		710	870	790	2370
总计		1490	1710	1760	4960

图 2-30 生成的数据透视表

求和项:销售量(台)		生产厂家			
商场	商品种类	TCL	长虹	海尔	总计
华联	彩电	210	240	230	680
	空调	130	110	230	470
华联 汇总		340	350	460	1150
京客隆	彩电	320	340	300	960
	空调	120	150	210	480
京客隆 汇总		440	490	510	1440
明珠	彩电	320	420	380	1120
	空调	390	450	410	1250
明珠 汇总		710	870	790	2370
总计		1490	1710	1760	4960

图 2-31 改变数据透视表样式

数据透视表中的数据是数据清单中的数据汇总得来的，如果发现表中有错误，不能直接进行修改，而必须先数据清单中修改数据，再单击“数据透视表工具”—“选项”选项卡“数据”组中的“刷新”按钮进行更新。

2. 数据透视图

为了更直观地反映数据透视表的分析结果，可以采用数据透视图绘制统计图形。有两种方法可以创建数据透视图。

第一种方法：如果之前已经创建了数据透视表，可以首先选中数据透视表（如图2-30所示的数据透视表）中的任一单元格，在“数据透视表工具”—“选项”选项卡的“工具”组中单击“数据透视图”按钮，弹出“插入图表”对话框，在其中选择图表类型后单击“确定”按钮，如图2-32所示。同时页面上会显示“数据透视图工具”选项卡，如图2-33所示。

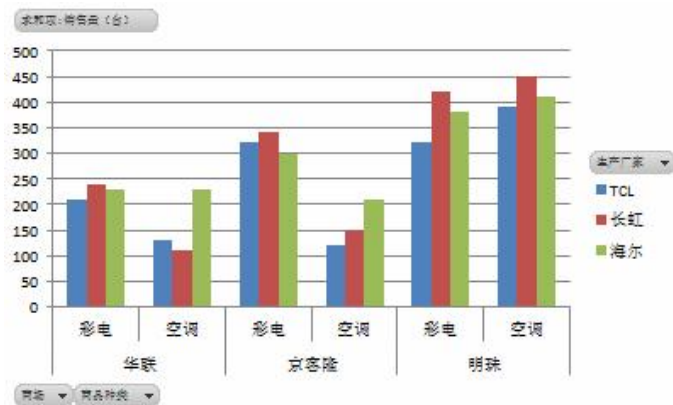


图 2-32 数据透视图（簇状柱形图）

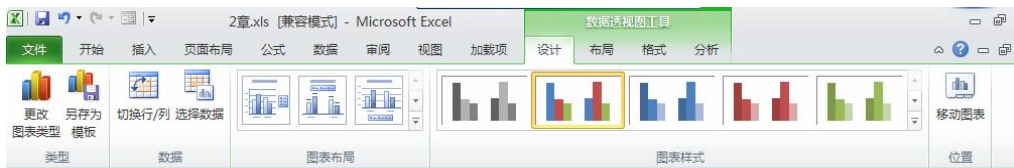


图 2-33 “数据透视图工具”选项卡

如果想改变数据透视图的图表类型，可在“数据透视图工具”—“设计”选项卡的“类型”组中单击“更改图表类型”按钮，在弹出的“更改图表类型”对话框中选择需要的图表类型，如选择“折线图”，单击“确定”按钮后如图 2-34 所示。



图 2-34 数据透视图改变类型（折线图）

如果想改变数据透视图的图表样式，可在“数据透视图工具”—“设计”选项卡的“图表样式”组中进行选择。如在“数据透视图工具”—“设计”选项卡的“图表样式”组中选择“样式 26”，则数据透视图变为图 2-35 所示的样式。

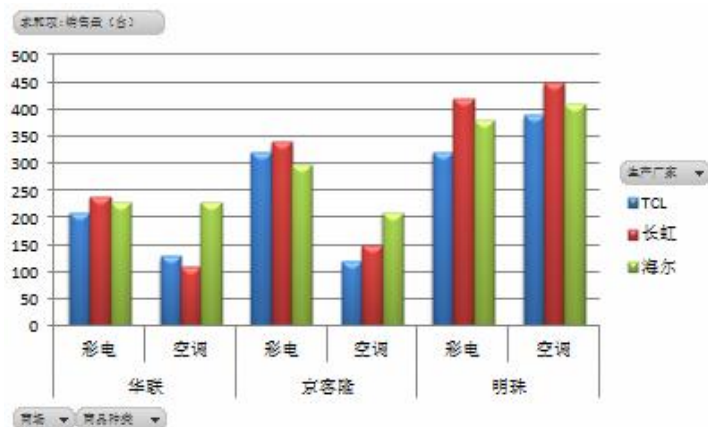


图 2-35 数据透视图改变样式（簇状柱形图）

第二种方法：如果之前还没有创建数据透视表，可以单击“插入”选项卡“表格”组中的“数据透视表”按钮，在弹出的下拉菜单中选择“数据透视图”命令。

例 2-9 以例 2-8 的数据资料创建数据透视图。

(1) 单击“插入”选项卡“表格”组中的“数据透视表”按钮，在弹出的下拉菜单中选择“数据透视图”命令，弹出“创建数据透视表及数据透视图”对话框，其中的选项均保持默认，单击“确定”按钮，形成空白数据透视图，如图 2-36 所示。



图 2-36 空白数据透视图

(2) 在“数据透视表字段列表”面板中选择“商场”、“商品种类”并用鼠标拖到“在以下区域间拖动字段”区域的“行标签”中，把“生产厂家”拖到“列标签”中，把“销售量(台)”拖到“数值”中，生成的数据透视表如图 2-32 所示。

以上是初步创建的数据透视图，不一定会达到理想状态，这时可以单击“数据透视图工具”——“布局”(或“数据透视图工具”——“格式”)功能区“当前所选内容”组中的“设置所选内容格式”按钮，在弹出的对话框中进行设置。

2.3.5 统计图与直方图分析

1. 统计图

使用统计图可以更加直观、形象地描述大量统计数据所体现出现象的发展规律与发展趋势，以及现象之间的相互关系和现象内部的结构。关于统计图的绘制方法，在第 1 章中已做过讲述，这里不再重复。

例 2-10 某大学共有教师 600 人，其中助教 180 人、讲师 240 人、副教授 120 人、教授 60 人，计算教师的职称结构，并绘制职称结构的饼形图。

(1) 建立“教师职称.xls”工作表，录入相关数据，如图 2-37 所示。

	A	B	C
1	职称	人数	结构(%)
2	助教	180	
3	讲师	240	
4	副教授	120	
5	教授	60	
6	合计		
7			

图 2-37 教师结构工作表

(2) 计算职工总数。选中单元格 B6，在编辑栏内输入公式“=SUM(B2:B5)”，按 Enter 键后显示 600。

(3) 计算教师比例。选中单元格 C2，在编辑栏内输入公式“=B2/B6”，按 Enter 键后显

示 30%（要把单元格的数字格式设成“百分比”），照此方法依次计算单元格 C3、C4、C5 中的数据。选中单元格 C6，在编辑栏内输入公式“=SUM(C2:C5)”，按 Enter 键后显示 100%，如图 2-38 所示。

	A	B	C
1	职称	人数	结构 (%)
2	助教	180	30.00%
3	讲师	240	40.00%
4	副教授	120	20.00%
5	教授	60	10.00%
6	合计	600	100.00%

图 2-38 计算合计与比例

(4) 从工作表中选择单元格区域 A1:A5、C1:C5，在“插入”选项卡的“图表”组中单击“饼图”按钮，在列表中选择“三维饼图”，形成统计图如图 2-39 所示。

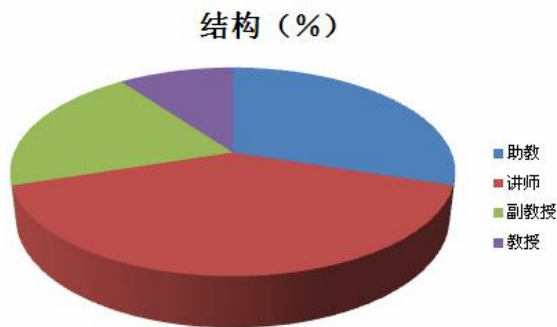


图 2-39 教师职称结构饼形图

(5) 单击图表标题位置，在文本框中输入“教师职称结构 (%)”字样；鼠标指向饼图并右击，在弹出的快捷菜单中选择“添加数据标签”命令，则统计图的形式变成如图 2-40 所示。

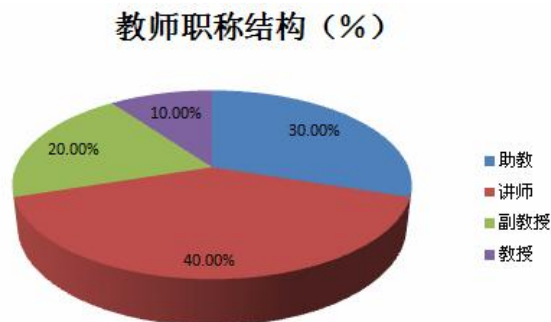


图 2-40 修改标题文字、增加数据标签后的图形

(6) 可以进一步对图中的标题文字、图例文字及数据标签格式进行设置。鼠标指向标题文字并右击，选择“字体”命令，在弹出的“字体”工具栏中设置字体为“隶书、常规、14号”。用同样的方法设置图例的字体为“楷体、常规、10号”。鼠标指向饼图并右击，在弹出的快捷菜单中选择“设置数据标签格式”命令，弹出“设置数据标签格式”对话框，如图 2-41 所示。在“标签位置”选项区中选中“数据标签外”单选按钮，单击“关闭”按钮。修饰后的图形如图 2-42 所示。



图 2-41 “设置数据标签格式”对话框

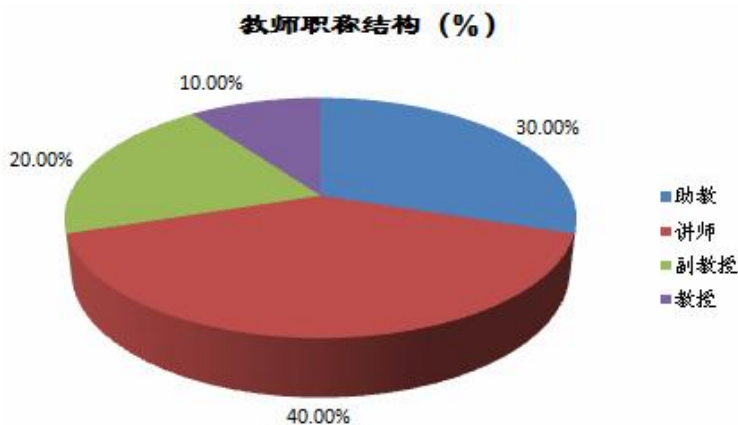


图 2-42 进一步修饰后的图形

2. 直方图分析

直方图是基于一组数据，并根据数据分布于若干称为区间内的个数（频率）绘制的图表（通常是简单的柱形图）。Excel 提供了一个直方图分析工具，可用于确定数据的频数分布、累计频数分布等。直方图分析工具是 Excel 分析工具库中的一种，利用分析工具库中的分析工具可以进行更加复杂的统计计算及数据分析。

分析工具库在默认情况下是不随 Excel 的安装而一同安装的，所以需要单独安装，方法是在“文件”选项卡中单击“选项”按钮，在弹出的“Excel 选项”对话框的左侧项目中单击“加载项”，在右侧下方单击“转到”按钮，弹出“加载宏”对话框，如图 2-43 所示。

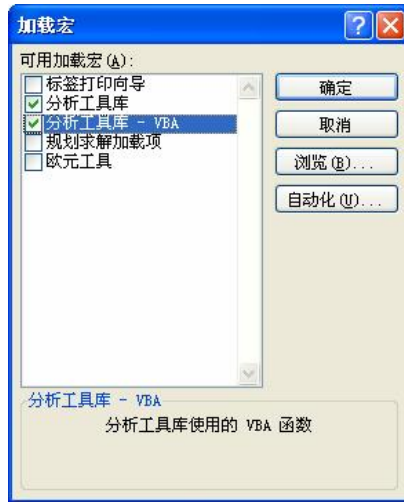


图 2-43 “加载宏”对话框

在“加载宏”对话框中，选中“分析工具库”和“分析工具库—VBA”复选框，单击“确定”按钮，系统会进行安装。安装完毕后，在“数据”选项卡中会出现“分析”组，“分析”组中有“数据分析”按钮。单击“数据分析”按钮，弹出“数据分析”对话框，如图 2-44 所示。



图 2-44 “数据分析”对话框

在“数据分析”对话框的“分析工具”列表框中选择“直方图”选项，单击“确定”按钮，弹出“直方图”对话框，如图 2-45 所示。



图 2-45 “直方图”对话框

在“直方图”对话框中，“输入区域”组合框内应输入待分析数据区域的单元格引用；“接

收区域”组合框内输入接收区域的单元格引用。“接收区域”组合框可以保留为空,“直方图”工具随后会自动使用输入区域中的最小值和最大值作为起点和终点来创建平均分布的区间间隔;如果输入区域的第一行或第一列包含标志项,则可选中“标志”复选框,如果输入区域没有标志项,则不选“标志”复选框,Excel将在输出表中自动生成数据标志;如选中“输出区域”单选按钮,应指定输出表左上角单元格的引用。如果输出表将覆盖已有的数据,Excel会自动确定输出区域的大小并显示信息;如选中“新工作表组”单选按钮,则可在当前工作簿中插入新的工作表,并从新工作表的A1单元格开始粘贴计算结果。如果需要给新工作表命名,则在右侧的编辑框中输入工作表名称;若选中“新工作簿”单选按钮,则可以创建一个新工作簿,并在新工作簿的新工作表中粘贴计算结果;选中“柏拉图”复选框可以在输出表中同时按降序排列频数数据,如果不选,则只按升序排列数据;选中“累积百分率”复选框可以在输出表中添加一系列累积百分比数值,并同时直方图表中添加累积百分比折线;选中“图表输出”复选框则可以在输出表中同时生成一个嵌入式直方图。

例 2-11 以例 2-6 的资料,用直方图分析工具对学生成绩进行分析。

(1) 打开工作表,将成绩按升序排序。

(2) 在单元格 D1 中输入文字“分组”,在单元格区域 D2:D6 中分别输入 59、69、79、89、99,表示分组区间的间隔点。

(3) 单击“数据”选项卡“分析”组中的“数据分析”按钮,在“数据分析”对话框中选择“直方图”,单击“确定”按钮。

(4) 在“直方图”对话框的“输入区域”组合框中输入单元格区域\$B\$1:\$B\$11,在“接收区域”组合框中输入\$D\$1:\$D\$6,选中“标志”复选框,在“输出区域”组合框中输入\$E\$1,选中“图表输出”复选框,如图 2-46 所示。



图 2-46 “直方图”对话框设置

(5) 单击“确定”按钮,Excel会同时生成一个频率分布表和一个图表,分别如图 2-47 和图 2-48 所示。

E	F
分组	频率
59	1
69	2
79	3
89	2
99	2
其他	0

图 2-47 频率分布表

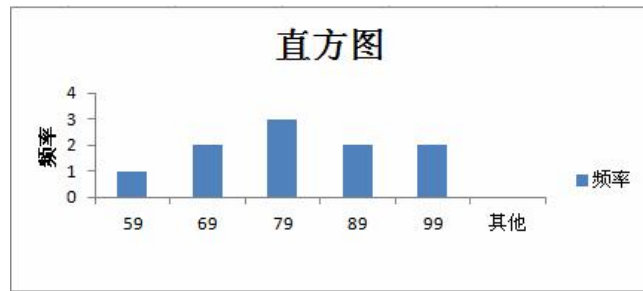


图 2-48 直方图表

注意：直方图显示等于或大于区间值并小于下一区间值的输入值的数量，并将其显示于“频率”列中。表中的最后一个值表示等于或大于最后一个区间值的输入值的数量。

习题二

- 2-1 统计分析的主要过程和主要分析方法有哪些？
- 2-2 如何建立数据清单？如何向数据清单中输入数据？如何编辑数据清单？
- 2-3 如何进行数据的排序，筛选和分类汇总？
- 2-4 FREQUENCY 函数如何运用？
- 2-5 如何建立数据透视表和数据透视图？
- 2-6 Excel 的直方图分析工具如何运用？