

第 1 章

数据分析的基本原理和工具

本章将简要介绍数据分析的基本原理，使读者在对数据开始真正分析之前，从整体上对数据分析的基本概念和基本流程有一个系统的了解。本章还将从整体上介绍 Excel 提供的多种分析工具的功能和特点，使读者对这些工具有大致的了解，本书后续章节会对这些分析工具的使用做详细介绍。

1.1 了解数据分析

本节将介绍数据分析的一些基本信息，使读者对数据分析有初步的了解。

1.1.1 数据分析是什么

简单来说，数据分析是通过人眼观察或相关分析工具，从错综复杂的数据中挖掘出有实际意义的信息，这些信息可能是现有数据的分布规律，也可能是对未来趋势的预测，它们都将为商业决策或个人行为提供帮助。

“数据分析”针对的对象是“数据”，要对“数据”做的事情是“分析”。不同的数据，其来源和复杂程度并不相同。无论数据的形式如何，分析数据时，首先需要对数据分类，分类后的数据才有进一步分析的意义。“分析”实际上是对数据做“比较”，只有对同类数据做比较，才能得出有价值的结果。

例如，在一个销售明细表中包含苹果、蓝莓、芒果 3 种产品在 3 个月内的销量，如图 1-1 所示。如果单独观察任意一种产品 3 个月的总销量，例如 E2 单元格中的 600，通过这个数字只能获悉苹果这种产品 3 个月的总销量是 600，再也得不到其他结论。

	A	B	C	D	E
1		1月	2月	3月	总计
2	苹果	100	200	300	600
3	蓝莓	300	300	200	800
4	芒果	200	500	300	1000
5	总计	600	1000	800	

图 1-1 3 种产品的销量

如果对比该产品每个月的销量，通过这些数字可以发现更多有价值的信息，即哪个月的销

量最多，哪个月的销量最少。如果对比 3 种产品在同一个月的销量，则可以得出“哪种产品的销量最多、哪种产品的销量最少”的结论，如图 1-2 所示。

	A	B	C	D
1		1月	2月	3月
2	苹果	100	200	300

	A	B
1		1月
2	苹果	100
3	蓝莓	300
4	芒果	200

图 1-2 分类后的数字放在一起做比较提供有意义的信息

根据数据比较的结果，可以进一步分析导致销量最多和销量最少的原因，是季节因素还是其他因素，是产品质量好还是受众人群众。这个示例说明了“分类”和“比较”在数据分析中的意义。

在数据分析的过程中，通常需要使用大量的统计学概念和工具。在图 1-2 的示例中，对比各个产品在各个月份的销量后得出的结论，在统计学中属于描述统计学。如需根据现有的销售数据预测产品在未来的销售趋势，在统计学中属于推断统计学。描述统计学主要用于汇总从过去发生的事件中搜集到的数据，推断统计学主要用于预测某个事件在未来发生的可能性。

1.1.2 为什么要进行数据分析

1.1.1 小节其实已经给出了进行数据分析的原因——了解现状和预测未来。了解现状是为了对当前阶段进行总结，弄清楚哪些方面做得好，哪些方面做得不好，然后及时调整和改善。例如，总结企业现阶段的整体运营和财政收支情况，从而衡量企业的整体发展形势和状况。

预测未来是为了让企业更好的发展而制定的策略和计划。数据分析虽然面对的是大量枯燥乏味的数字，但是这项工作对企业的长远发展至关重要。从更广泛的角度来说，进行数据分析的原因有以下 3 点：

- 数字化信息随处可见，无法避开数据分析。
- 影响个人和企业的行为都依赖于数据分析。
- 可以帮助个人和企业做出更明智的决策。

1.1.3 数据分析的基本流程

为了在分析过程中减少错误，最后得到正确的分析结果，在分析数据时需要遵循以下基本流程：

创建数据⇨整理数据⇨分析数据⇨展现数据

1. 构建数据

“数据分析”针对的对象是“数据”，所以首先要有数据。数据主要有两种来源，一种是将数据手动输入到数据分析软件中，例如 Excel。另一种是将其他程序产生的数据导入到数据分析软件中。无论使用哪种方法，都要确保数据的格式符合数据分析软件的格式规范，以便数据分析软件能够正确识别和处理这些数据。

2. 整理数据

正如在上一个环节中提到的，无论是手动输入的数据，还是从外部程序导入的数据，通常都存在格式不规范、内容有错误等问题，所以需要数据必要的“清洗”工作。“清洗”是指对数据本身涵盖的内容及其格式进行规范化处理，涉及的操作主要包括对数据的转换、提取、拆分、合并等。

3. 建模数据

业务复杂的数据通常存储在多个表中，当需要对业务涉及的所有数据进行分析时，需要先对整个数据建模。“建模”就是在所有表之间建立关系，将它们在逻辑上联接为一个整体，使分析工具通过表间关系在各个表中的数据之间无障碍地穿梭。为所有表建立关系后，就可以定义新的计算指标来丰富数据模型。创建有效的数据模型可以更准确地传达数据本身蕴含的信息，使用户可以更快、更有效地了解和使用数据，以后也更容易维护数据。

4. 展现数据

分析数据的最终目的是使分析结果易于被人理解，所以如何以易于识别和理解的形式展现数据就变得格外重要。图表是展现数据的有用工具，将数据以线条和形状等图形化的形式呈现出来，可以使晦涩难懂的数字变成直观醒目的图形，从而更容易洞察隐藏在数据背后的信息。

1.2 数据分析的基本概念

数据分析与统计学密不可分，实际上，数据分析所依赖的正是统计学的概念和理论。例如，计算某个年度的销售总额、计算每个月的平均销量、计算加权后的平均工资、根据去年的销售额预测未来的销售趋势等，所有这些都要用到统计学方面的知识。本节将简要介绍统计学方面的基本概念，以便将其更好地运用到数据分析中。

1.2.1 定性数据和定量数据

分析数据前，首先需要明确区分数据的类型，因为不同类型的数据所适用的分析方法并不相同。如果对不同类型的数据运用同一种分析方法，可能会得到截然不同的结果，有些结果是毫无意义甚至是错误的，这会给商业决策带来隐患或严重问题。简单来说，当了解数据的不同类型及其特性之后，就会明白哪些数据适合计算总和，哪些数据只适合统计个数。

根据数据是否是数值型的，可以将数据分为定性数据和定量数据两种类型。所有非数值型的数据都是定性数据，性别、籍贯、发色、手机品牌等都是定性数据的示例。所有数值型数据都是定量数据，年龄、员工人数、销售额、销量等都是定量数据的示例。

虽然大多数数据都遵循上述划分原则，但是某些数值型的数据实际上也是定性数据。以学生的学号为例，第一个学生的学号是1，第二个学生的学号是2，其他学号以此类推。虽然学号是数值型数据，但是对它进行数学运算并无实际意义，学号只是使用非重复数字对每名学生进行标识，它并不具备数学意义上的计算。

对于定量数据，可将其细分为离散型和连续型两种。离散型定量数据的各个取值之间存在一定的“间隙”。“人数”是离散型定量数据的一个示例，人数可能是3个，也可能是5个，还可能是10个，但是绝对不可能是5.6个。离散型定量数据通常是通过计数得到的。

连续型定量数据的取值可以是指定范围内的任意值。“身高”是连续型定量数据的一个示例，身高可能是1.8米，也可能是1.78米，还可能是1.75米。连续型定量数据通常是通过测量得到的。

1.2.2 数据的测量尺度

划分数据类型的另一种方式是依据“测量尺度”。数据的测量尺度可分为名义尺度、顺序尺度、区间尺度、比率尺度4种，它们的级别由低到高依次排列，名义尺度的级别最低，比率尺度的级别最高。级别的高低决定了可对数据进行计算和分析的方式。

1. 名义尺度

名义尺度数据没有顺序之分，通常只对它们进行分类和计数。“性别”是名义尺度的一个示例，性别分为“男”和“女”两类，可以先统计公司男员工的人数，也可以先统计女员工的人数，男员工和女员工没有先后顺序之分。“商品类别”也是一个名义尺度，可以统计超市提供的所有商品类别的总数，而商品类别中的各个分类之间没有先后顺序，可以将“饮料酒水”排在第一位，也可以将“米面粮油”排在第一位。

2. 顺序尺度

比名义尺度高一个级别的数据是顺序尺度。与名义尺度相比，顺序尺度多了一个“顺序”的特性，可以对顺序尺度数据进行排序。例如，根据销售业绩对员工评级，分为优、良、中、差 4 个级别，此处的“级别”就是一个顺序尺度，因为各个级别之间有好坏之分，“好坏”意味着它们是有先后顺序的。最好的级别是“优”，比“优”差一点的是“良”，比“良”差一点的是“中”，最差的级别是“差”。但是无法知道“优”到底比“良”好多少，也就是说，对顺序尺度数据之间的差异无法通过数学运算得到某个具体的值，它们仅具有顺序关系。

3. 区间尺度

比顺序尺度高一个级别的数据是区间尺度，该尺度具有顺序尺度的所有特性，同时还具有一个新的特性——数值之间的差值可被量化，且是固定不变的。温度是区间尺度的一个典型示例，两个温度之差表示相对温差，但两个温度相乘或相除却没有任何意义。温度中的 0 不代表没有温度，而也只是作为温度中的一个普通温度——0°。

4. 比率尺度

比率尺度是位于最高级别的数据，该尺度具有区间尺度的所有特性，同时，数值之间的差异还能以比值的形式进行量化，而且 0 具有特殊的意义。销售量是比率尺度的一个示例，两个销售量的差值可以说明销售量之差，它们的比值也可以反映销售量相差的倍数。例如，如果 1 月份牛奶卖了 500 箱，2 月份卖了 1000 箱，通过计算它们的比值，可以知道 2 月份牛奶的销售量是 1 月份的 2 倍。如果牛奶的销售量为 0，则表示一箱牛奶也没卖出去。实际上，几乎所有的定量数据都是比率尺度，只不过对某些定量数据计算比值没有实际意义。

1.2.3 描述数据位置的指标

“位置度量”实际上就是人们所熟知的“平均数”，它用于确定一组数值的中心位置。算术平均数、加权平均数、几何平均数、中位数、众数等都是描述数据位置的指标。

1. 算术平均数

最常见的一种平均数是算术平均数，其计算方法是使用所有数值的总和除以所有数值的数量。例如，A 型号手机售价是 1000 元，B 型号手机的售价是 1500 元，C 型号手机的售价是 2000 元，如需计算这 3 种型号手机的平均售价，可以使用以下公式：

$$(1000+1500+2000) \div 3=1500$$

由于算术平均数计算的是所有给定数值，所以当其中存在特别大或特别小的值时，计算得到的平均数将不能准确地表示该组数据的平均值。

2. 加权平均数

加权平均数是算术平均数的一种特殊情况。当一组数据中有几个相同的值时，算术平均数

就变成了加权平均数。例如，某种奶茶的容量分为大、中、小3种，大杯奶茶的价格是12元，中杯奶茶的价格是9元，小杯奶茶的价格是6元。奶茶店在上午卖出了10杯奶茶，其中有2个大杯、5个中杯、3个小杯。如需计算10杯奶茶的平均价格，可以使用以下公式：

$$(2 \times 12 + 5 \times 9 + 3 \times 6) \div (2 + 5 + 3) = 8.7$$

在上面的公式中，权重就是不同容量奶茶售出的杯数，即大杯奶茶的权重是2，中杯奶茶的权重是5，小杯奶茶的权重是3。计算加权平均数时，分母永远都是所有权重的总和。

3. 几何平均数

当计算随时间推移的数据的平均值时，将会用到几何平均数。“随时间推移”是几何平均数的关键因素，因为在随时间推移过程中产生的每个值，都是基于上一个时间点的值进行计算的，也就是说，这将涉及累积值的问题。

例如，刚入职时的工资是6000元，今年工资上涨了10%，明年工资上涨了20%，在这两年之中，工资的平均年增长率是多少？如果使用算术平均数的计算方法，可能会得到15%的结果，但这是错误的，因为它将每年工资增长率的基础工资都设定为最初的工资6000元，而忽略了第二年工资的增长率是在上一年上涨后的工资基础之上的。

第一年工资上涨了10%，上涨后的工资为：

$$6000 \times 10\% + 6000$$

即

$$6000 \times (1 + 10\%) = 6600$$

第二年工资上涨了20%，它所基于的起始工资是上一年上涨10%之后的工资6600元，而非最初的6000元。所以，第二年工资上涨20%后的工资为：

$$6600 \times (1 + 20\%) = 7920$$

如需计算工资的平均年增长率，需要分别计算每年增长率与1之和，然后将计算结果相乘，最后再计算该结果的2次方根。此处开2次方是因为计算的是两年，如果计算的是3年，则开3次方根。基于该计算方法，计算本例中的平均年增长率时，首先计算以下公式的结果为1.32。

$$(1 + 10\%) \times (1 + 20\%)$$

然后对1.32开2次方后再减1，结果为0.1489，转换成百分比值为14.89%，该结果就是工资的平均年增长率。

4. 中位数

由于算术平均数会受到特别大或特别小的值的影响，所以在这种情况下，中位数变得非常有用，它只考虑数据的位置关系，而不会让数据受到数值大小的影响。中位数只对可排序数据有效，这意味着名义尺度数据没有中位数，因为名义尺度数据没有顺序之分，也就不存在中间位置的概念了。

中位数是一组数据中位于中间位置的那个数据，这种情况针对的是有奇数个数据的情况。如果包含偶数个数据，则中位数是位于中间两个数据的算术平均值。无论数据的个数是奇数还是偶数，在取出中间位置的数据之前，必须先将这组数据升序或降序排列。升序是指按照数值从小到大的顺序排列，降序是指按照数值从大到小的顺序排列。

在1、5、2、4、3这组数字中，中位数是3。首先将这组数字从小到大排列，得到的是1、

2、3、4、5，由于3位于这组数字的中间位置，所以该数字就是这组数字的中位数。

如需找出1、5、6、2、4、5这组数字的中位数，首先仍需将它们从小到大排列，得到的是1、2、3、4、5、6。所有数字的个数是6，由于它是一个偶数，没有绝对位于中间位置的数字，所以需要位于相对中间位置的两个数字3和4求和，再除以2，得到的3.5就是这组数字的中位数。

5. 众数

众数是一组数据中出现次数最多的那一个。无论哪种类型的数据，只要存在出现次数最多的数据，就存在众数。由于名义尺度数据没有中位数，所以众数对于该类型的数据更加有用。

在1、2、2、3、3和3这组数字中，由于数字3出现的次数最多，所以3是这组数字的众数。如果出现次数最多的数据不止一个，则将存在多个众数。在1、2、2、3和3这组数字中，由于2和3都出现两次，所以2和3都是众数。如果一组数据中的每个值的出现次数相同，可以认为这组数据没有众数或者每个数据都是众数。

1.2.4 描述数据离散程度的指标

如果在分析数据时只考虑平均数，则可能得出一些错误的结论。例如，城市A住房的平均售价为200万元，城市B住房的平均售价也是200万元。如果仅通过平均售价，可能得出“两个城市的住房平均售价的价格分布相同”的结论，实际情况果真如此吗？

如果仔细研究两个城市住房售价的具体价格范围，就会发现上述结论是错误的。例如，城市A的住房售价的价格范围在100万元~300万元，而城市B的住房售价的价格范围在50万元~350万元。虽然两个城市的住房平均售价相同，但是住房售价的价格分布并不相同。

为了在分析数据时得到正确的结论，除了考虑数据的平均数之外，还需要考虑数据的离散程度（也称为离中趋势或变异程度）。极差、平均离差、方差、标准差等都是用于描述数据离散程度的指标。

1. 极差

极差的计算方式最简单，它使用一组数据中的最大值减去最小值得到。在1、2、3、4、5这组数字中，极差是4，使用最大值5减去最小值1得到。

2. 平均离差

平均离差描述的是数据相对于分布的均值或中心的相对接近或相对聚集程度。在计算平均离差之前，需要先理解什么是离差。离差是一组数据中的每个值与该组数据的算术平均数之间的差值。在1、2、3、4、5这组数字中，算术平均数是3，这组数字的离差分别是以下几个：

```
1-3=-2
2-3=-1
3-3=0
4-3=1
5-3=2
```

可能已经发现，一组数据中各个值的离差之和等于0。为了计算平均离差，要对所有值的离差求和，然后除以所有值的个数。由于各个值的离差之和等于0，所以在对所有值的离差求和时，需要将每个离差转换为绝对值，以确保最后的求和结果不是0，这样才有实际意义。计算上面5个数字的平均离差的公式如下：

```
(|-2|+|-1|+0+1+2)÷5=1.2
```

3. 方差

方差可用来比较两组或多组数据的离散程度。方差的计算方法与平均离差相似，但是在计算方差时，不是计算各个值与算术平均值之差的绝对值，而是计算各个值与算术平均值之差的平方，这样得到的必然是正数。然后对得到的所有平方求和，最后将求和结果除以所有值的个数。仍以前面示例中的 1、2、3、4、5 这组数字为例，计算该组数字的方差的公式如下：

$$((1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2) \div 5 = 2$$

在统计学中经常会遇到两个术语——总体和样本。总体是所研究的个人、事物或者关于个人和事物的度量值的全部集合，样本是总体的一部分。上面公式计算的是总体方差，如需计算样本方差，需要将公式中用作分母的值减 1，所以计算 1、2、3、4、5 这组数字的样本方差的公式如下：

$$((1-3)^2 + (2-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2) \div (5-1) = 2.5$$

4. 标准差

方差的平方根就是标准差。与方差类似，标准差也分为总体标准差和样本标准差。由于前面计算可知，1、2、3、4、5 这组数字的总体方差是 2，其总体标准差就是 2 的平方根，结果为 1.414。1、2、3、4、5 这组数字的样本方差是 2.5，其样本标准差就是 2.5 的平方根，结果为 1.581。

较大的标准差说明一组数据中的各个值比较分散地分布在均值周围，而较小的标准差说明一组数据中的各个值紧密地聚集在均值附近。

1.2.5 概率的基本术语和概念

概率是学习推断统计学的基础，此处将简要介绍概率的一些基本概念和术语。概率表示某件事发生的可能性有多大。概率是一个 0~1 的数值，包括 0 和 1 两个值，概率越接近 1，表示事件发生的可能性越大；概率越接近 0，表示事件发生的可能性越小。概率是 1 表示事件必定发生，概率是 0 表示事件绝对不会发生。

上面提到的“事件”是统计学中的一个术语，“试验”和“结果”是另外两个常用术语，这 3 个术语紧密相关。“试验”是执行的一项操作，试验的结果可能有多个，但是每次进行试验时只能得到其中一个结果。

“抛硬币”是介绍概率基本术语的一个易于理解的示例。“抛硬币”就是一项试验，抛一次硬币是一项试验，连续抛三次硬币也是一项试验。每次抛硬币后肯定有一面朝上，可能是正面，也可能是反面，“正面”和“反面”就是这项试验的两个结果，但是在抛起硬币时无法确定将会出现哪种结果。

无论每次试验的结果是正面还是反面，只要进行一次试验，就会产生一个事件。例如，抛一次硬币，结果是正面，这是一个事件。接下来连续抛三次硬币，结果分别是正面、反面、正面，这又是一个事件。

判断一个事件发生的概率有 3 种方法：古典概率、经验概率和主观概率。古典概率是通过希望的结果的个数除以所有可能结果的个数得到的。例如，抛一枚硬币出现正面的概率是多少？此处希望的结果是正面朝上，而结果的个数是 1，而抛硬币一共有“正面”和“反面”两种可能的结果。因此，通过古典概率推断抛一枚硬币出现正面概率的公式如下：

$$1 \div 2 = 0.5$$

经验概率是根据事件已发生的次数占已知试验次数比例的一种推断概率的方法。仍以抛硬

币为例，如果只抛一次，则出现正面的经验概率可能是 0 或者 1。如果不断增加试验次数，出现正面的经验概率将会逐渐接近 0.5。

主观概率是通过个人的经验主观推断出事件发生的概率。下面是两个主观概率的示例：

- 根据今天的天气状况，估计明天下雨的概率。
- 根据晚饭后的饱腹感，估计半夜吃夜宵的概率。

1.3 数据分析模型

如今已经存在大量的数据分析模型，这些模型为数据分析提供框架结构方面的指导，使数据分析过程能够有条不紊地进行。本节将介绍几种常见的数据分析模型。

1.3.1 5W2H 分析模型

5W2H 分析模型主要对用户行为、营销活动、业务问题等进行分析。5W2H 中的 5W 是指 Why（为什么）、What（什么事）、Who（谁）、When（什么时候）和 Where（什么地方），2H 是指 How（如何做）和 How much（什么价格）。下面是 5W2H 在分析用户购买行为时的用法：

- Why：用户为什么购买？产品的吸引力在哪儿？
- What：产品能提供什么功能？
- Who：用户群是什么？这个群体有什么特点？
- When：用户购买的频次是多少？
- Where：产品在哪儿销售出去的？在哪儿最受欢迎？
- How：用户如何购买？购买方式是什么？
- How much：用户购买的金钱成本是多少？时间成本是多少？

1.3.2 4P 分析模型

4P 是营销领域常用的分析模型，4P 是指 Product（产品）、Price（价格）、Place（渠道）和 Promotion（推广）。将产品、价格、渠道和推广四者完美融合，从而提高企业的市场份额，为企业带来更多的利润。

- 产品：是指提供给市场、被人们使用和消费的任何东西。
- 价格：是指人们购买产品时的价格，包括基本价格、折扣价格、支付期限等。影响价格主要有成本、需求和竞争 3 个因素。
- 渠道：是指产品从生产端到达用户端所经历的全过程涉及的各个环节。
- 推广：是指企业通过改变销售方式来刺激并加大用户消费，广告、宣传推广、人员推销、销售促进是推广的几种主要方式。

1.3.3 SWOT 分析模型

SWOT 分析模型主要用于确定企业自身的内部优势、劣势和外部的机会和威胁等。SWOT 中的 S（strengths）是优势、W（weaknesses）是劣势，O（opportunities）是机会、T（threats）是威胁或风险。使用 SWOT 分析模型可以对研究对象所处的情景进行全面、系统、准确的研究，从而将公司的战略和公司内部资源、外部环境完美结合在一起。

1.3.4 PEST 分析模型

PEST 主要用于行业分析，从各个方面把握宏观环境的现状和变化趋势。PEST 是指 Politics

(政治)、Economy (经济)、Society (社会) 和 Technology (技术)。

- 政治环境：包括政治体制、经济体制、财政政策、税收政策、产业政策、投资政策等。
- 经济环境：包括 GDP 及增长率、进出口总额及增长率、利率、汇率、通货膨胀率、消费价格指数、居民可支配收入、失业率、劳动生产率等。
- 社会环境：包括人口规模、性别比例、年龄结构、生活方式、购买习惯、城市特点等。
- 技术环境：包括折旧和报废速度、技术更新和传播速度、技术商品化速度等。

1.3.5 波士顿分析模型

波士顿分析模型最开始用于时间管理，按照紧急、不紧急、重要、不重要排列组合分成四个象限，从而可以有效地管理时间。在市场营销中使用波士顿分析模型，可以合理搭配产品，优化产品组合，促进产品销售稳步增长并加大市场占有率，使企业利润最大化。可以将销售额和利润两个指标作为横、纵坐标轴分为四个象限，然后将产品分为以下 4 种类型。

- 明星类：销售增长率和市场占有率都高，用户需求量大并且有很大发展空间的产品。
- 金牛类：销售增长率低但是市场占有率高，收益大但是没有太大发展空间的产品。
- 问题类：销售增长率高但是市场占有率低，用户需求量大但是可能存在问题的产品。
- 瘦狗类：销售增长率和市场占有率都低，没有任何发展空间的失败产品。

1.3.6 漏斗分析模型

漏斗分析模型主要用于分析业务流程中各个环节的转化率，从而改善业务状况，提高利润。漏斗分析模型的一个常见应用场景是对顾客购物流程的分析，从顾客将商品加入购物车到成功完成订单交易，对整个流程中的每一步的转化率进行分析，通过异常的数据指标找出有问题的环节并解决问题，从而提升购买转化率，为企业带来更多的利润。

1.4 Excel 中的数据分析工具

分析数据的目的是从大量数据中发现有价值的信息，最终形成有效的观点或结论，为个人或企业做出明智的决策提供帮助。Excel 提供了大量的分析工具，使用这些工具可以对数据进行不同类型的分析。为了便于描述和记忆，本节对 Excel 中的分析工具进行了分类，并简要介绍它们的功能和特点，以及在 Excel 界面中的位置，本书后续章节会详细介绍这些工具的用法。

1.4.1 常用分析工具——排序、筛选、分类汇总和数据透视表

Excel 中的基本分析工具包括排序、筛选、分类汇总和数据透视表，将它们称为“基本分析工具”是因为这些工具简单易用，使用它们无须具备统计学方面的知识，即可快速获得分析结果。另一个原因是使用这些工具只能完成数据的基本分析工作。

1. 排序

“排序”工具位于 Excel 功能区的“数据”选项卡中，如图 1-3 所示。使用该工具可以对数据按升序、降序或自定义顺序排列，以便了解数据的分布规律，快速找出数据中的最大值和最小值。“排序”工具可以在普通数据区域和数据透视表中使用。



图 1-3 “排序”工具

2. 筛选

“筛选”工具位于 Excel 功能区的“数据”选项卡中，如图 1-4 所示。使用该工具可以快速找出符合条件的数据。无论哪种类型的数据，它们都有一些相同的筛选操作。此外，不同类型的数据也有其特有的筛选操作。“筛选”工具可以在普通数据区域和数据透视表中使用。



图 1-4 “筛选”工具

提示: 在“开始”选项卡中也可以找到“排序”和“筛选”两个工具，只需在该选项卡中单击“排序和筛选”按钮，如图 1-5 所示。

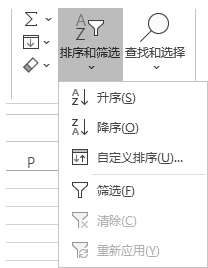


图 1-5 “开始”选项卡中的“排序”和“筛选”工具

3. 分类汇总

“分类汇总”工具位于 Excel 功能区的“数据”选项卡中，如图 1-6 所示。使用该工具可以对数据按照一个或多个特定的类别划分，并自动计算分类数据的数值总和、个数、均值、最大值和最小值等。“分类汇总”工具可以在普通数据区域中使用。



图 1-6 “分类汇总”工具

4. 数据透视表

“数据透视表”工具位于 Excel 功能区的“插入”选项卡中，如图 1-7 所示。使用该工具可以在不编写任何公式的情况下，快速完成大量数据的汇总工作。只需调整字段在数据透视表中的位置，即可快速创建出具有不同含义的分析报表。



图 1-7 “数据透视表”工具

创建数据透视表后，将在功能区中显示“数据透视表分析”和“设计”两个选项卡，其中包含与数据透视表操作相关的选项，如图 1-8 所示。



图 1-8 与数据透视表操作相关的选项

第5章将详细介绍排序、筛选、分类汇总和数据透视表的使用。

1.4.2 专业分析工具——模拟分析、单变量求解、规划求解和分析工具库

Excel 中的高级分析工具包括模拟分析、单变量求解、规划求解和分析工具库。与基本分析工具相比,高级分析工具的针对性更强,有些工具需要用户具备统计学方面的知识才能正确使用。

1. 模拟分析

“模拟分析”工具位于 Excel 功能区的“数据”选项卡中，如图 1-9 所示。使用该工具可以基于现有的计算模型，对影响最终结果的多种因素进行预测和分析，从而得到最接近目标的方案。



图 1-9 “模拟分析”工具

2. 单变量求解

“单变量求解”工具位于 Excel 功能区的“数据”选项卡中，如图 1-10 所示。使用该工具可以进行与模拟分析相反方向的分析。



图 1-10 “单变量求解”工具

3. 规划求解

“规划求解”工具位于 Excel 功能区的“数据”选项卡中，如图 1-11 所示，使用该工具前需要先安装“规划求解”加载项。使用“规划求解”工具可以为多个变量设置约束条件，通过不断调整这些变量的值来得到期望的结果。在经营决策、生产管理等需要对资源和产品进行合理规划的应用场景中，“规划求解”工具能够帮助企业获得最佳的经济效益，例如利润最大、产量最高、成本最小等。



图 1-11 “规划求解”工具

4. 分析工具库

“分析工具库”工具位于 Excel 功能区的“数据”选项卡中，如图 1-12 所示，使用该工具前需要先安装“分析工具库”加载项。使用“分析工具库”工具可以对数据进行统计分析和工程计算。



图 1-12 “分析工具库”工具

第 5 章将详细介绍模拟分析、单变量求解、规划求解和分析工具库的用法。

1.4.3 可视化工具——图表、迷你图和条件格式

图表是展现数据的有力工具，将难于记忆和理解的数字以线条和形状的方式绘制出来，可以快速发现隐藏在数据背后的信息。Excel 提供了近 20 种图表类型，每种图表类型又细分为一个或多个子类型。

迷你图是从 Excel 2010 开始支持的新功能，可以将其看作是“图表”工具的微缩版。使用“迷你图”工具，可以在单元格中创建微型图表，用于显示特定的数据点或表示一系列数据的变化趋势。迷你图只能显示一个数据系列，且不具备普通图表所拥有的一些图表元素。

“图表”和“迷你图”两个工具位于 Excel 功能区的“插入”选项卡中，如图 1-13 所示。



图 1-13 “图表”工具和“迷你图”工具

创建图表后，将在功能区中显示“图表设计”和“格式”两个选项卡，其中包含与图表操作相关的选项，如图 1-14 所示。



图 1-14 与图表操作相关的选项

创建迷你图后，将在功能区中显示“迷你图”选项卡，其中包含与迷你图操作相关的选项，如图 1-15 所示。

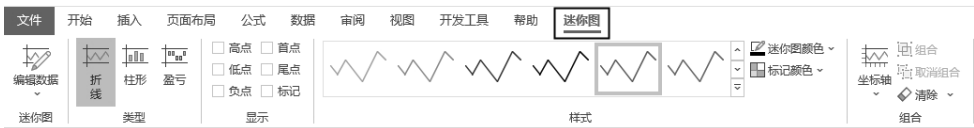


图 1-15 与迷你图操作相关的选项

虽然条件格式与图表和迷你图的功能截然不同，但是由于条件格式也能从外观上改变数据的显示效果，所以也可以将其看作是一种数据可视化工具。使用“条件格式”工具可以满足条件的数据自动设置字体、边框、填充色等格式，以达到突出显示特定数据的目的。“条件格式”工具位于 Excel 功能区的“开始”选项卡中，如图 1-16 所示。

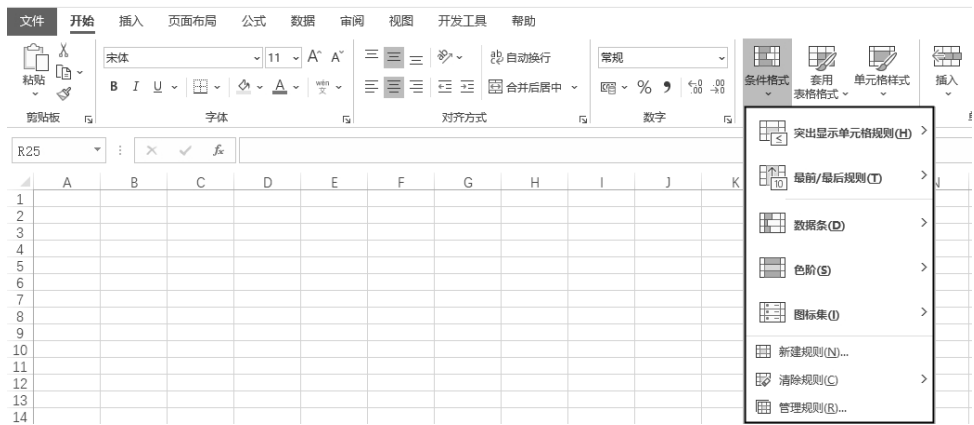


图 1-16 “条件格式”工具

第 6 章将详细介绍图表的用法。

1.4.4 商业智能分析工具——Power BI

目前最受欢迎的两个商业智能分析工具是 Power BI 和 Tableau，前者由微软创建，也是本书介绍的重点。两种工具的整体功能大同小异，细节之处却各有千秋。

Power BI 工具有两种使用方式，第一种是在计算机中安装名为 Power BI Desktop 的独立程序，该工具包括数据的导入、整理、建模、分析、展示等数据分析涉及的每个阶段所需使用的工具。

如果不想单独安装 Power BI Desktop，也可以在 Excel 中使用 Power BI 的功能，为此需要在 Excel 中安装几个名称以 Power 开头的加载项的名称。在 Excel 中使用的 Power BI 功能称为 Power BI for Excel。本书的后半部分将详细介绍 Power BI Desktop 和 Power BI for Excel 两个工具的用法。

提示：对 Tableau 感兴趣的读者可参阅本书笔者编写的《Tableau 数据可视化分析从新手到高手》一书，该书已于 2023 年 10 月由清华大学出版社有限公司出版。

1.4.5 公式和函数

公式和函数是 Excel 执行数据计算的核心功能。Excel 内置了 400 多个函数，使用这些函数可以快速完成不同领域的计算和统计工作。除了通过编写公式直接计算单元格中的数据之外，还可以在条件格式、数据验证、图表中嵌入公式和函数，从而增强数据的控制和显示效果。

第 4 章将详细介绍公式的基本概念和操作，以及使用不同类型的函数计算和统计数据的方法。

第 2 章

在 Excel 中构建数据

为了使用 Excel 分析数据，首先需要在 Excel 中创建数据。在 Excel 中创建数据有手动输入和从外部程序导入两种方式，本章将介绍这两种方式的操作方法，以及与构建和编辑数据相关的一些常用操作。

2.1 输入和导入数据

在 Excel 中可以输入不同类型的数据，所有类型的数据共同遵循基本的输入方法，然而，每类数据也有其特定的输入规则。只有按照规则输入数据，Excel 才能正确识别和处理它们。本节将介绍输入数据的基本方法和特定数据类型的输入规则，还将介绍一些提高输入效率的技巧，最后以文本文件和 Access 数据库为例，介绍在 Excel 中导入外部数据的方法。

2.1.1 Excel 中的数据类型

Excel 支持 5 种数据类型：数值、文本、日期和时间、逻辑值、错误值。数据在 Excel 中的存储和处理方式由数据的类型决定。在单元格中输入数据后，不同的数据类型在单元格中具有不同的默认位置：文本在单元格中左对齐，数值、日期和时间在单元格中右对齐，逻辑值和错误值在单元格中居中对齐，如图 2-1 所示。

	A	B	C	D	E
1	文本	数值	日期和时间	逻辑值	错误值
2	Excel	666	2025/3	TRUE	#NUM!
3	数据分析	888	2025年3月1日	FALSE	#VALUE!

图 2-1 不同的数据类型在单元格中具有不同的默认位置

1. 文本

任何作为描述性的内容都可以存储为文本，例如姓名、性别、商品名称、产品编号等，文本都是定性数据。一些不需要参与计算的数字，也可以存储为文本，例如电话号码和身份证号码。在 Excel 中可以比较文本的大小，但是不能对文本执行数学运算。

文本可以由任意字符组成。虽然一个单元格最多可以容纳 32767 个字符，但是在单元格中

最多只能显示 1024 个字符，多出的字符可以在 Excel 功能区下方的编辑栏中查看。

2. 数值

数学和数值是两个不同的概念，数字是由 0 ~ 9 这十个数字任意组合而成的单纯的数，数值是具有特定用途或含义的数字，例如金额、销售量、员工人数、体重、身高等，数值都是定量数据。

Excel 也会将一些带有特殊符号的数字识别为数值，这些符号包括百分号(%)、货币符号(例如¥)、千位分隔符(,)、科学计数符号(E)等。在 Excel 中可以对数值执行数学计算。

Excel 支持的最大正数约为 $9E+307$ ，最小正数约为 $2E-308$ ，最大负数与最小负数与这两个数字相同，只是需要在数字开头添加负号。虽然 Excel 支持很大的数字范围，但是如果在单元格中输入一个 15 位以上的整数，则从第 16 位开始会自动变成 0。例如，111222333444555666 会变成 111222333444555000。如果输入的是一个 15 位以上的小数，则会自动将第 15 位之后的所有数字删除。

如需在单元格中输入 15 位以上的整数或小数，需要将数字以文本格式输入，即将数值变成文本，才能使数字完整地保留下来，具体方法请参考 2.1.4 小节。

在单元格中输入数字时，如果数字的长度超过单元格的宽度，Excel 会自动增加列宽以完全容纳整个数字。如果整数位数超过 11 位，则自动以科学计数形式显示。如果数值的小数位数较多，且超过单元格的宽度，Excel 会自动对超出宽度的第一个小数位进行四舍五入，并删除其后的小数位。无论是以科学计数形式显示的整数，还是对小数位四舍五入后显示的小数，都可以在编辑栏中查看它们的完整形式，如图 2-2 所示。

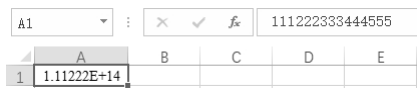


图 2-2 在编辑栏中显示数字的完整形式

3. 日期和时间

由于日期和时间存在先后次序，因此，日期和时间在 Excel 中以“序列值”的形式存储，序列值的范围是 1~2958465，第一个日期的序列值是 1，最后一个日期的序列值是 2958465，每个日期对应一个序列值。

在 Windows 操作系统的 Excel 版本中，1900 年 1 月 1 日是第一个日期，其序列值为 1，1900 年 1 月 2 日是第二个日期，其序列值为 2，其他日期与序列值的对应关系以此类推，最后一个日期是 9999 年 12 月 31 日，其序列值为 2958465。在 Macintosh 计算机的 Excel 版本中，第一个日期是 1904 年 1 月 1 日，其序列值为 1，其他日期以此类推。

Excel 允许用户在这两种日期系统之间转换。只需打开“Excel 选项”对话框，在“高级”选项卡中选择所需的日期系统，如图 2-3 所示。勾选“使用 1904 日期系统”复选框将起始日期指定为 1904 年 1 月 1 日，不勾选该复选框时的起始日期为 1900 年 1 月 1 日。

如前所述，每个日期都用一个整数来表示。无论日期表示的是哪一天，每一天日期的数值单位都是 1，一天有 24 个小时，使用 1 除以 24 得到的就是每个小时的数值表示方式。继续使用类似的方法，可以得到 1 分钟和 1 秒的数值表示方式：

1 分钟： $1 \div 24 \div 60$
1 秒： $1 \div 24 \div 60 \div 60$

中午 12 点可以表示为：

$1 \div 24 \times 12$

如果表示日期的数字大于 1，Excel 会将整数部分换算为日期，将小数部分换算为时间。例如，是一个表示日期的序列值 45717.75 表示 2025 年 3 月 1 日 18 点。

如需查看一个日期对应的序列值，可以先在单元格中输入该日期，然后将其数字格式设置为“常规”。如需查看一个序列值对应的日期，可以在单元格中输入该序列值，然后将其数字格式设置为一种日期格式，如图 2-4 所示。该设置位于功能区“开始”选项卡的“数字”组中。

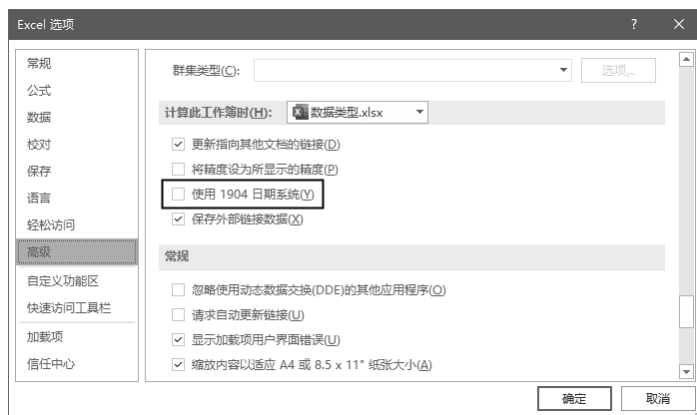


图 2-3 转换日期系统

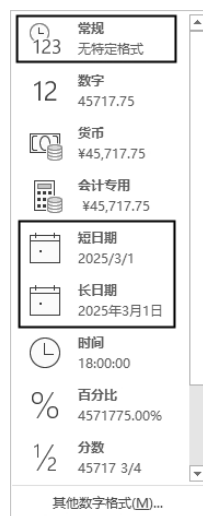


图 2-4 在日期和序列值之间转换

4. 逻辑值

逻辑值只包含 TRUE（真）和 FALSE（假）两个值，该类型的数据主要在条件判断中使用。当判断结果为 TRUE 时执行一种计算，当判断结果为 FALSE 时则执行另一种计算，这样就可以根据条件的不同判断结果自动执行指定的计算，使公式能够应对不同的情况。

当对逻辑值执行四则运算时，逻辑值 TRUE 等价于 1，逻辑值 FALSE 等价于 0。

```
TRUE+TRUE=2
TRUE+FALSE=1
FALSE+FALSE=0
TRUE×TRUE=1
TRUE×FALSE=0
FALSE×FALSE=0
```

当在条件判断中使用逻辑值时，任何非 0 数字等价于逻辑值 TRUE，0 等价于逻辑值 FALSE。

5. 错误值

与其他几种数据类型不同，在 Excel 中通常不会主动输入“错误值”类型的数据，而是由公式遇到错误后自动创建的一类信息，用于提醒用户公式出错的类型和可能原因。Excel 中的错误值有 7 个，每一个错误值的名称都以 # 号开头，如表 2-1 所示。

表 2-1 Excel 中的 7 种错误值

错 误 值	出 错 原 因
#DIV/0!	在公式中使用 0 作除数
#NUM!	在公式中使用无效的数值
#VALUE!	在公式中使用错误的参数类型
#REF!	在公式中引用不存在的单元格或工作表
#NAME?	无法将公式中的文本识别为有效的内容
#N/A	使用公式无法找到目标数据
#NULL!	在公式引用没有交集的两个区间的重叠部分

2.1.2 行、列和单元格的表示方法

在 Excel 中，列的编号使用 A、B、C 等英文大写字母表示，通常将列的编号称为列标。行的编号使用 1、2、3 等阿拉伯数字表示。每一行和每一列的交叉点都是单元格，使用列名在前、行号在后的方式标识每个单元格，这种方式类似于点的坐标。

例如，A1 表示 A 列与第一行交叉处的单元格，B3 表示 B 列与第 3 行交叉处的单元格。将用于标识单元格的文本称为单元格地址。A1:B6 表示由以 A1 单元格为矩形的左上角顶点、以 B6 单元格为矩形的右下角顶点所组成的单元格区域。A:A 表示 A 列中的所有单元格。B:E 表示 B~E 列中的所有单元格。1:3 表示 1~3 行中的所有单元格。

提示：如果发现列标变成了阿拉伯数字，则可以打开“Excel 选项”对话框，在“公式”选项卡中取消勾选“R1C1 引用样式”复选框。

“活动单元格”是输入数据时需要了解的一个重要概念。在 Excel 中新建一个工作表，工作表中默认被选中的那个单元格就是活动单元格，输入的内容会自动填入活动单元格中。如果单击其他任意一个单元格，则活动单元格将变成刚单击的这个单元格。如果同时选择多个单元格，则第一个被选中的单元格就是活动单元格，其背景色为白色，选中的其他单元格的背景色为暗色。如图 2-5 所示，当前同时选中了 6 个单元格，其中的 B2 单元格是活动单元格。

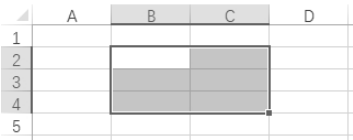


图 2-5 B2 单元格是活动单元格

2.1.3 输入数据的基本方法

无论输入哪种类型的数据，都要遵循基本的输入方法。输入数据前需要先选择一个单元格，然后开始输入所需的内容。输入过程中会在单元格中显示一条闪烁的竖线，将其称为“插入点”，它指明输入内容的当前位置，如图 2-6 所示。

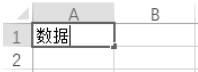


图 2-6 插入点指明输入内容的当前位置

输入完成后，按 **Enter** 键或单击编辑栏中的  按钮，输入的内容将同时显示在单元格和编辑栏中。这两种方法的区别是：

- 按 **Enter** 键：使正下方的单元格成为活动单元格。
- 单击  按钮：不改变活动单元格的位置。

技巧：在“Excel 选项”对话框的“高级”选项卡中，可以勾选“按 Enter 键后移动所选内容”复选框，然后在“方向”下拉列表中选择按 Enter 键时激活单元格的方向，如图 2-7 所示。



图 2-7 设置按 Enter 键时激活单元格的方向

如果在输入过程中想要取消本次输入，则可以按 **Esc** 键或单击编辑栏中的  按钮，会立刻清除当前已输入到单元格中的内容，使单元格恢复到输入前的状态。

输入数据时，在 Excel 窗口底部的状态栏左侧会显示当前的输入模式，分为“输入”“编辑”“点”3 种。

1. “输入”模式

选择一个单元格后开始输入内容，或者直接双击空单元格，都将进入“输入”模式，此时会在状态栏的左侧显示“输入”二字，如图 2-8 所示。在该模式下，插入点会随输入的内容自动右移，只能从左向右依次输入。输入时如果按方向键，则会立刻结束输入并退出“输入”模式。

2. “编辑”模式

选择一个单元格后，单击编辑栏或按 **F2** 键，将进入“编辑”模式，此时会在状态栏的左侧显示“编辑”二字，如图 2-9 所示。在该模式下，可以使用方向键或单击来改变插入点的位置。修改单元格中的内容时需要使用“编辑”模式。

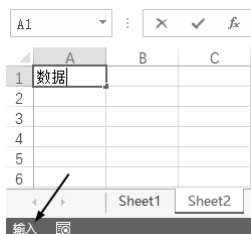


图 2-8 “输入”模式

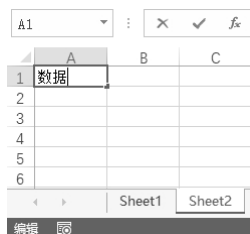


图 2-9 “编辑”模式

3. “点” 模式

只有在单元格中输入公式时，才会用到“点”模式。在公式中输入等号或运算符后，按方向键或单击任意一个单元格，将进入“点”模式，此时会在状态栏的左侧显示“点”字，如图 2-10 所示。在该模式下，当前选中的单元格的边框将变为虚线，该单元格的地址会被自动添加到公式中。

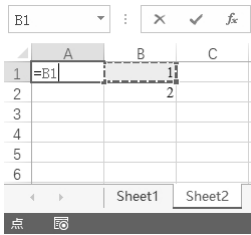


图 2-10 “点” 模式

2.1.4 输入 15 位以上的数字

正如在 2.1.1 小节中介绍的，当在 Excel 中输入 15 位以上的数字时，从第 16 位开始的所有位数都会变成 0，这种结果是不可逆的。例如，在单元格中输入 18 位身份证号码 110101199001016666，将显示为 110101199001016000，即最后 3 位都变为 0，如图 2-11 所示。

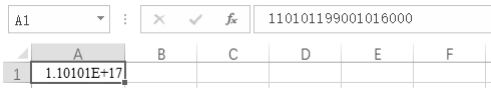


图 2-11 超过 15 位的数字都会变为 0

为了正确显示 15 位以上的数字，需要以文本形式输入数字，有以下两种方法：

- 选择一个单元格，在功能区的“开始”选项卡的“数字格式”下拉列表中选择“文本”选项，然后在该单元格中输入 15 位以上的数字，如图 2-12 所示。
- 直接在单元格中输入数字，但是需要在数字的开头先输入一个英文半角单引号（'）。只能在编辑栏中看到单引号，而不会显示在单元格中，如图 2-13 所示。



图 2-12 将单元格的数字格式设置为文本

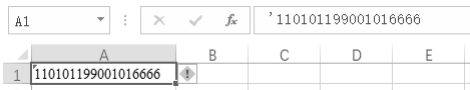


图 2-13 在数字的开头添加一个英文单引号

提示：使用第二种方法输入数字后，会在单元格的左上角显示一个蓝色标记，并在单元格右侧显示一个黄色叹号。这是因为 Excel 发现单元格包含错误而给出的警告信息，解决方法请参考第 4 章。

2.1.5 输入具有指定间隔的一系列数字

如需经常输入一系列具有指定间隔的数字，可以使用 Excel 中的“填充”功能来提高操作效率。“填充”是指使用鼠标拖动单元格右下角的填充柄，在拖动过的每个单元格中将自动填入数据。“填充柄”是指选中的单元格右下角的小方块。将鼠标指针指向填充柄时，鼠标指针会变为十字形，此时拖动鼠标即可执行填充操作，如图 2-14 所示。

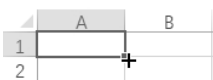


图 2-14 单元格右下角的填充柄

注意：如果鼠标指针没有变为十字形，说明当前无法使用“填充”功能，此时需要在“Excel 选项”对话框的“高级”选项卡中勾选“启用填充柄和单元格拖放功能”复选框，如图 2-15 所示。



图 2-15 启用填充功能

“填充”功能适用于在水平的行或垂直的列中快速输入一系列具有指定间隔的数字，行和列的数量可以是一个或多个。下面以列为例，介绍填充功能的几种用法，在行中填充数据的方法与此类似。

1. 从输入前两个数字开始

在一列的两个连续单元格中输入起始的两个数字，例如 1 和 2。选择这两个单元格，然后向下拖动位于下方的那个单元格右下角的填充柄，拖动过程中会在鼠标指针附近显示当前值。显示目标值时，释放鼠标左键即可，如图 2-16 所示。

2. 从只输入起始数字开始

在一列的某个单元格中输入起始数字，例如 1，然后向下拖动该单元格右下角的填充柄，拖动过程中需要一直按住 Ctrl 键。显示目标值时，释放鼠标按键和 Ctrl 键，释放顺序不分先后。

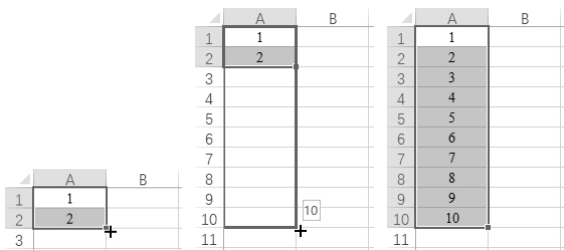


图 2-16 输入前两个数字后开始填充

3. 使用“序列”对话框执行填充

如果要填充的数字个数很多，可以使用“序列”对话框加快填充速度，操作步骤如下：

- (1) 在一个单元格中输入起始数字，选择该单元格。
- (2) 在功能区的“开始”选项卡中单击“填充”按钮，然后在弹出的快捷菜单中选择“序列”命令，如图 2-17 所示。
- (3) 打开“序列”对话框，如图 2-18 所示，进行以下几项设置：
 - 在“序列产生在”类别中选择填充方向。如需在一行中进行填充，需要选择“行”选项。选择“列”选项将在一列中进行填充。
 - 在“类型”类别中选择填充的数据值特征，“等差序列”选项是指两个相邻数字之间的差值都相同，例如 1、2、3、4、5，相邻数字之间的差值都是 1。“等比序列”是指两个相邻数字之间的比值都相同，例如 1、2、4、8、16，相邻数字之间的比值都是 2。
 - 步长值：该选项用于设置两个相邻数字之间的递增量。
 - 终止值：该选项用于设置数字序列的值范围的上限，序列中的最后一个数字小于或等于终止值，具体是哪种情况取决于步长值。

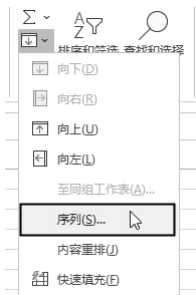


图 2-17 选择“序列”命令



图 2-18 “序列”对话框

按照图 2-18 中的设置，将得到一个 1~10 的数字序列，该序列填充在一列中。如果打开“序列”对话框之前，在 A1 单元格中输入了 1，则该数字序列将填充在 A1、A2 一直到 A10 单元格中。如果在 B2 单元格中输入了 1，则会将该数字序列填充在 B2、B3 一直到 B11 单元格中。无论将起始数字 1 输入到哪个单元格中，都会以该单元格为起点，向下一一直填充到第 10 个单元格中。

4. 双击填充柄自动完成填充

如果在要填充的列的相邻列中包含数据，则可以直接双击单元格的填充柄快速在该列中填充数据。

如图 2-19 所示，A 列包含大写英文字母 A~G，如需在 B 列中输入这些字母的 ASCII 码，可以先输入前两个字母的 ASCII 码。字母 A 的 ASCII 码是 65，字母 B 的 ASCII 码是 66。

选择字母 A 和 B 所在的单元格，本例为 A2 和 A3。然后双击 A3 单元格右下角的填充柄，即可自动将 B 列中其他字母的 ASCII 码填入 A 列中，如图 2-20 所示。

	A	B
1	字母	ASCII 码
2	A	65
3	B	66
4	C	
5	D	
6	E	
7	F	
8	G	

图 2-19 先输入前两个字母的 ASCII 码

	A	B
1	字母	ASCII 码
2	A	65
3	B	66
4	C	67
5	D	68
6	E	69
7	F	70
8	G	71

图 2-20 双击填充柄执行填充操作

注意：B 列中的数据之间不能存在空白单元格，否则无法填充完整的数据。

5. 填充由非数字字符和数字组成的编号

示例文件/第2章/填充由非数字字符和数字组成的编号.xlsx

如果编号由非数字字符和数字组成，则只需输入第一个编号，然后直接拖动该编号所在单元格的填充柄，即可自动在其他单元格中填入连续的编号，其中的非数字字符保持不变，数字部分会持续递增，如图 2-21 所示。

	A	B	C
1	字母	ASCII 码	
2	A	ASCII-65	
3	B		
4	C		
5	D		
6	E		
7	F		
8	G		

	A	B	C
1	字母	ASCII 码	
2	A	ASCII-65	
3	B		
4	C		
5	D		
6	E		
7	F		
8	G		

	A	B
1	字母	ASCII 码
2	A	ASCII-65
3	B	ASCII-66
4	C	ASCII-67
5	D	ASCII-68
6	E	ASCII-69
7	F	ASCII-70
8	G	ASCII-71

图 2-21 填充由非数字字符和数字组成的编号

提示：编号中的数字部分不一定非要位于结尾，也可以位于开头或非数字字符之间。

2.1.6 输入日期

在 Excel 中输入日期和时间时，除了遵循 2.1.3 小节介绍的基本输入方法之外，还需要按照正确的格式来输入，使 Excel 能够将输入的内容识别为日期和时间。在 Windows 操作系统的 Excel 版本中，需要按照以下格式输入日期：

- 使用“-”或“/”符号分隔表示年、月、日的数字，在一个日期中可以混合使用这两种符号，例如 2025-5-29、2025/5/29、2025-5/29。
- 在表示年、月、日的数字之后添加“年”“月”“日”等文字，例如 2025 年 5 月 29 日。

注意：如果省略表示“年”的数字，则默认为操作系统当前的年份；如果省略表示“日”的数字，则将日默认为所输入的月份的第一天，例如 2025-5 等同于 2025-5-1。

输入时间时，需要使用冒号分隔表示小时、分钟和秒的数字。时间有 12 小时制和 24 小时制两种，如需使用 12 小时制表示时间，需要在表示凌晨和上午的时间结尾添加 AM，在表示下午和晚上的时间结尾添加 PM，需要在 AM 和 PM 与时间结尾之间保留一个空格。如果时间结尾没有 AM 或 PM，则将使用 24 小时制的时间格式。

例如，“8:30 AM”表示上午 8 点 30 分，“8:30 PM”表示晚上 8 点 30 分。如果把时间结尾的 AM 和 PM 去掉，此时“8:30”仍然表示上午 8 点 30 分，而晚上 8 点 30 分需要使用“20:30”表示。

注意：输入的时间可以省略“秒”部分，但是必须包含“小时”和“分钟”两个部分。

如需输入一系列连续的日期，可以先输入第一个日期，然后使用鼠标拖动填充柄，将自动在单元格中填入其他日期，如图 2-22 所示。

如果使用鼠标右键拖动填充柄，则可以在弹出的快捷菜单中选择以不同方式填充日期，如图 2-23 所示。还可以选择“序列”，然后使用“序列”对话框设置填充日期的选项。

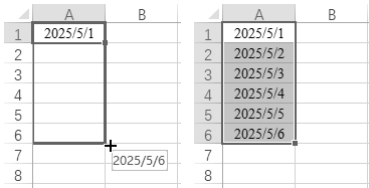


图 2-22 使用“填充”功能快速输入一系列日期

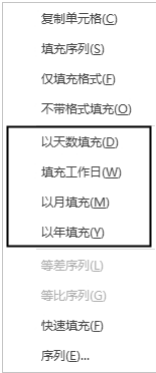


图 2-23 填充日期的不同方式

2.1.7 在一个单元格中输入多行内容

当在单元格中输入的内容超过单元格的宽度时，使用“自动换行”功能可以自动将超出宽度的内容移动到单元格中的下一行继续显示。如需使用“自动换行”功能，需要先选择一个或多个单元格，然后在功能区的“开始”选项卡中单击“自动换行”按钮。如图 2-24 所示为 A1 单元格中的内容自动换行前后的效果。

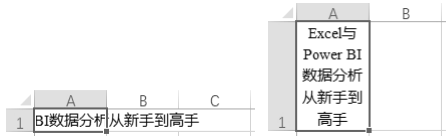


图 2-24 自动换行前后的效果

有时可能需要在指定的位置换行，而不是由 Excel 自动决定换行的位置。如需在指定的位置换行，可以进入“编辑”模式，将插入点定位到要换行的位置，然后按 Alt+Enter 组合键。如图 2-25 所示为对 A1 单元格中的内容手动换行后的结果，在编辑栏中也会显示手动换行后的格式。

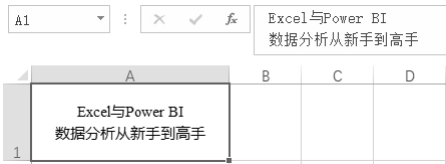


图 2-25 在指定位置手动换行

提示：为内容手动换行后，如果看不到换行的效果，可以增大单元格的宽度。

2.1.8 一次性在多个单元格中输入数据

如需在多个单元格中输入相同的内容，可以先选择这些单元格，然后输入所需的内容，最后按 **Ctrl+Enter** 组合键，如图 2-26 所示。

	A	B	C	D	E
1					
2		数据分析	数据分析	数据分析	
3		数据分析	数据分析	数据分析	
4		数据分析	数据分析	数据分析	
5					

图 2-26 一次性在多个单元格中输入数据

选择多个单元格有以下几种方法：

- 单击一个单元格，然后按住鼠标左键在工作表中拖动，到达另一个单元格时释放鼠标左键，将选中以这两个单元格为左上角和右下角的矩形范围内的所有单元格。
- 单击一个单元格，然后按住 **Shift** 键，再单击另一个单元格。
- 单击一个单元格，然后按 **F8** 键，进入扩展选择模式，此时单击另一个单元格，选择效果与上一种方法相同。在扩展选择模式下按 **F8** 键或 **Esc** 键将退出该模式。
- 单击一个单元格，然后按住 **Ctrl** 键，多次单击其他单元格，将选中所有单击过的单元格。这种方法适合于选择不相邻的多个单元格。

2.1.9 导入文本文件中的数据

在很多情况下，要分析的数据可能是由其他程序创建的，此时可以将这些数据导入到 Excel 中。Excel 支持多种来源的数据，例如文本文件、Access 数据库、SQL Server 数据库以及 OLAP 多维数据集等。

文本文件是一种跨平台的通用文件格式，便于在不同的操作系统和程序之间交换数据，在 Excel 中可以轻松导入文本文件中的数据。如图 2-27 所示的文本文件中有 5 列数据，各列数据之间以制表符分隔。

示例文件/第2章/商品销售明细.txt
示例文件/第2章/导入文本文件中的数据.xlsx

将该文本文件中的数据导入到 Excel 中的操作步骤如下：

(1) 新建或打开一个 Excel 工作簿，在功能区的“数据”选项卡中单击“从文本/CSV”按钮，如图 2-28 所示。

商品编号	名称	产地	单价	销量
SP001	面包	北京	6	200
SP002	牛奶	天津	2	100
SP003	酸奶	上海	3	300
SP004	饼干	山东	5	600
SP005	香油	江苏	15	400
SP006	酱油	广东	10	500

图 2-27 数据以制表符分隔



图 2-28 单击“从文本/CSV”按钮

提示：如果使用的是 Excel 2019 之前的 Excel 版本，则需要单击“数据”选项卡中的“自文本”按钮。

(2) 打开“导入数据”对话框，进入“商品销售明细.txt”文件所在的文件夹，然后双击该文件，如图 2-29 所示。

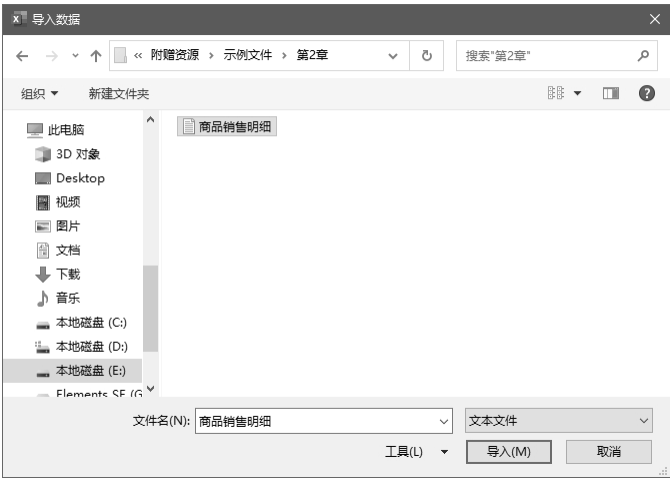


图 2-29 双击要导入的文本文件

(3) 打开如图 2-30 所示的对话框，Excel 会自动检测文本文件中的数据格式，并自动选择合适的选项，并显示数据的预览视图。如果视图中的数据 displays 有误，则可以更改上方的选项。完成后单击“加载”按钮。

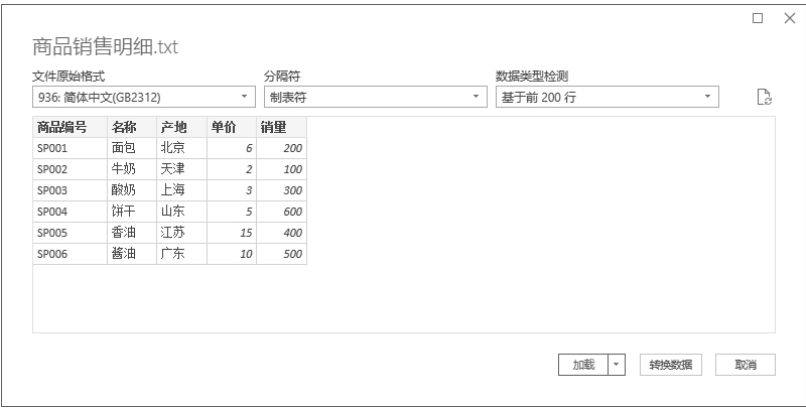


图 2-30 设置与数据格式相匹配的选项

提示：如果使用的是 Excel 2019 之前的 Excel 版本，打开的将是“文本导入向导”对话框，按照向导进行操作即可。

(4) Excel 将在当前工作簿中新建一个工作表，并将数据以“表格”的形式导入到该工作表中，如图 2-31 所示。

如果修改了文本文件中的数据，为了使导入到 Excel 中的数据也做同步修改，可以右击数据区域中的任意一个单元格，然后在弹出的快捷菜单中选择“刷新”命令，如图 2-32 所示。

提示：“表格”是 Excel 提供了一种动态管理数据的功能，它可以自动扩展数据区域，还可以在不输入公式的情况下自动完成求和、最大值、最小值、平均值等常规运算。如果不想保留这种动态功能，也可以随时将其转换为普通的数据区域。

	A	B	C	D	E
1	商品编号	名称	产地	单价	销量
2	SP001	面包	北京	6	200
3	SP002	牛奶	天津	2	100
4	SP003	酸奶	上海	3	300
5	SP004	饼干	山东	5	600
6	SP005	香油	江苏	15	400
7	SP006	酱油	广东	10	500

图 2-31 以“表格”形式导入数据

图 2-32 刷新数据以保持同步更新

2.1.10 导入 Access 数据库中的数据

Access 与 Excel 同为微软 Office 中的成员，与 Excel 处理数据的方式不同，Access 是专门用于处理大量相关数据的关系数据库程序。Access 中的数据存储在多个表中，通过在多个相关表之间创建关系，使这些表中的数据在逻辑上联接为一个整体，使用这些数据时就好像它们都存在于同一个表中。Excel 也支持导入由 Access 创建的关系数据库中的数据。

示例文件/第2章/商品销售管理系统.accdb

示例文件/第2章/导入Access数据库中的数据.xlsx

如图 2-33 所示的 Access 数据库中有 3 个表，此处要将名为“商品销售明细”的表中的数据导入到 Excel 中，操作步骤如下：

所有 Access 对象 <<

搜索...

表

客户订单信息

商品基本信息

商品销售明细

商品销售明细

商品编号	名称	产地	单价	销量
SP001	面包	北京	6	200
SP002	牛奶	天津	2	100
SP003	酸奶	上海	3	300
SP004	饼干	山东	5	600
SP005	香油	江苏	15	400
SP006	酱油	广东	10	500
*				

图 2-33 Access 数据库中的数据

(1) 新建或打开一个工作簿，在功能区的“数据”选项卡中单击“获取数据”按钮，然后在弹出的快捷菜单中选择“来自数据库”→“从 Microsoft Access 数据库”命令，如图 2-34 所示。



图 2-34 选择“从 Microsoft Access 数据库”命令

提示：如果使用的是 Excel 2019 之前的 Excel 版本，则需要单击“数据”选项卡中的“自 Access”按钮。

(2) 打开“导入数据”对话框，双击要导入的“商品销售管理系统 .accdb”文件。

(3) 打开如图 2-35 所示的对话框，在左侧选择要导入的数据所属的表，本例为“商品销售明细”表，右侧会显示该表中的数据预览视图。确认无误后，单击“加载”按钮。



图 2-35 选择要导入的数据所属的表

注意：如果使用的是 Excel 2019 之前的 Excel 版本，打开的将是“选择表格”对话框和“导入数据”对话框，选择要导入的 Access 表和放置表的位置即可。

提示：如果勾选“选择多项”复选框，则可以同时选择多个表，以便将它们一次性导入到 Excel 中。

(4) 与导入文本文件类似，Excel 也会在当前工作簿中新建一个工作表，并将导入的 Access 数据以“表格”形式导入到该工作表中。

2.2 编辑数据

将数据输入或导入到 Excel 中后，通常需要对数据进行一些编辑。例如，修改有错误的文字、修正不正确的数值或格式、移动数据的位置，以及通过复制操作快速获得数据的副本。

2.2.1 修改数据

如需修改单元格中的数据，需要先进入“编辑”模式，有以下几种方法：

- 双击单元格。
- 单击单元格，然后按 F2 键。
- 单击单元格，然后单击编辑栏。

进入“编辑”模式后，使用鼠标单击或者按键盘上的方向键，将插入点定位到所需的位置，然后输入新的内容或删除插入点两侧的内容。也可以拖动鼠标选择单元格中的部分内容，然后输入新的内容，直接替换选中的内容。完成修改后，按 Enter 键或单击其他单元格。

注意：如果不进入“编辑”模式而直接输入内容，则新输入的内容会直接替换单元格中原有的所有内容。

如果要修改的内容出现在多个单元格中，则可以使用“替换”功能一次性完成修改。在功能区的“开始”选项卡中单击“查找和选择”按钮，然后在弹出的快捷菜单中选择“替换”命令，

打开“查找和替换”对话框的“替换”选项卡，如图 2-36 所示。在“查找内容”文本框中输入要修改的原始内容，在“替换为”文本框中输入修改后的内容，最后单击“全部替换”按钮。



图 2-36 “查找和替换”对话框中的“替换”选项卡

提示：如果在“替换为”文本框中不输入任何内容，则单击“全部替换”按钮后，将会删除所有匹配的内容。

如需精确控制匹配内容的方式，可以在“查找和替换”对话框的“替换”选项卡中单击“选项”按钮，展开后的对话框如图 2-37 所示，各个选项的功能如下。



图 2-37 精确控制匹配内容的方式

- **范围：**查找数据的位置，包含“工作表”和“工作簿”两项，“工作表”是指当前工作簿，“工作簿”是指当前工作簿中的所有工作表。
- **搜索：**查找数据的行列次序，包含“按行”和“按列”两项，“按行”是指按先行后列的顺序查找，“按列”是指按先列后行的顺序查找，默认为“按行”。
- **查找范围：**指定内容的类型，包含“公式”“值”和“批注”三项。“公式”包括输入的数据和组成公式的字符，但是不包括公式的计算结果；“值”包括输入的数据和公式的计算结果，但是不包括构成公式的字符；“批注”只包括批注内容。
- **区分大小写：**是否区分英文字母的大小写形式。
- **单元格匹配：**是否查找完全相同的内容。勾选该复选框，将严格按照“查找内容”文本框中的内容查找匹配的单元格。例如，如果在“查找内容”文本框中输入“数字”并勾选“单元格匹配”复选框，则将查找只包含“数字”二字的单元格，所以“数字值”不会被匹配，因为它比“数字”多了一个“值”字。
- **区分全/半角：**是否区分字符的全角和半角。

如果单元格中的内容不再有用，可以使用以下几种方法将数据删除：

- 选择单元格，然后按 **Delete** 键。
- 右击单元格，在弹出的快捷菜单中选择“清除内容”命令。

如果为单元格设置了字体、背景色等格式，则使用上面的方法只能删除单元格中的内容。

如需同时删除单元格中的内容和格式，可以在功能区的“开始”选项卡中单击“清除”按钮，然后在弹出的快捷菜单中选择“全部清除”命令，如图 2-38 所示。

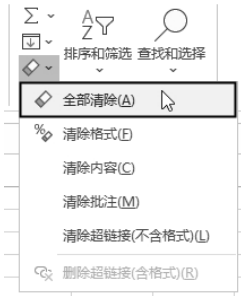


图 2-38 使用“全部清除”命令同时删除内容和格式

2.2.2 移动和复制数据

需要移动和复制的数据可以位于单元格或单元格区域中，只能移动连续单元格区域中的数据，复制的单元格区域可以是位于同行或同列的连续或不连续的单元格区域。如图 2-39 所示的两个选中的单元格区域可以执行复制操作，虽然它们的列数不同，但是它们都包含第 1~3 行。如图 2-40 所示的两个选中的单元格区域不能执行复制操作，因为第一个选区包含第 1~3 行，第二个选区只包含第 1~2 行。

	A	B	C	D	E
1	1		1	1	
2	2		2	2	
3	3		3	3	
4	4		4		
5	5				
6					

图 2-39 可以同时复制的两个单元格区域

	A	B	C	D	E
1	1		1	1	
2	2		2	2	
3	3		3	3	
4	4		4		
5	5				
6					

图 2-40 不能同时复制的两个单元格区域

移动和复制数据可以使用很多方法，包括拖动鼠标、功能区命令、鼠标快捷菜单和键盘快捷键，这些方法的效果等同。

1. 拖动鼠标

移动数据：将鼠标指针指向单元格的边框，当鼠标指针变为十字箭头时，按住鼠标左键并拖动到目标单元格。

复制数据：与移动数据的方法类似，但是需要在拖动鼠标的过程中按住 **Ctrl** 键。到达目标单元格后，释放鼠标左键和 **Ctrl** 键，释放顺序不分先后。

无论移动还是复制数据，如果目标单元格包含数据，都会显示如图 2-41 所示的提示信息，单击“确定”按钮，将使用当前正在移动或复制的数据替换目标单元格中的数据。单击“取消”按钮，将取消当前的移动或复制操作，不会对目标单元格中的数据产生影响。

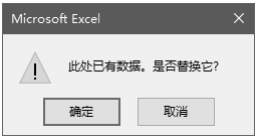


图 2-41 是否替换目标单元格中的数据

2. 功能区命令

移动数据：选择数据所属的单元格，在功能区的“开始”选项卡中单击“剪切”按钮，然后选择目标单元格，再在功能区的“开始”选项卡中单击“粘贴”按钮，如图 2-42 所示。

复制数据：与移动数据的方法类似，只需将单击“剪切”按钮改为单击“复制”按钮，其他操作相同。

3. 鼠标快捷菜单命令

移动数据：右击数据所属的单元格，在弹出的快捷菜单中选择“剪切”命令，然后右击目标单元格，在弹出的快捷菜单中选择“粘贴选项”中的“粘贴”命令，如图 2-43 所示。

复制数据：与移动数据的方法类似，只需将选择“剪切”命令改为选择“复制”命令，其他操作相同。



图 2-42 使用功能区命令移动和复制数据



图 2-43 选择“粘贴选项”中的“粘贴”命令

4. 键盘快捷键

移动数据：选择数据所属的单元格，按 **Ctrl+X** 组合键执行与前几种方法相同的剪切操作，然后选择目标单元格，按 **Ctrl+V** 组合键或 **Enter** 键执行与前几种方法相同的粘贴操作。

复制数据：与移动数据的方法类似，只需将按 **Ctrl+X** 组合键改为按 **Ctrl+C** 组合键，其他操作相同。

提示：除了第一种方法之外，当使用其他几种方法对单元格执行剪切操作后，单元格的边框都会变为虚线。如图 2-44 所示，表示当前处于“剪切复制”模式。在该模式下可以执行多次粘贴操作，这意味着可以反复执行粘贴操作，以便将同一个单元格中的数据复制到多个不同位置。如果是通过按 **Enter** 键执行的粘贴操作，则在粘贴后会立刻退出“剪切复制”模式。进入“剪切复制”模式后，可以按 **Esc** 键直接退出该模式。



图 2-44 “剪切复制”模式下的单元格边框变为虚线

2.2.3 以指定方式粘贴数据

在 2.2.2 小节介绍移动和复制数据的方法时，其中一种方法是使用鼠标快捷菜单中的“粘贴选项”功能。除了其中的“粘贴”命令之外还有很多命令，使用它们可以实现不同的粘贴效果。

粘贴数据时，Excel 会自动将单元格中的数据、格式、数据验证规则、批注等所有内容都粘贴到目标单元格中。有时可能只想粘贴单元格中的数据而不包含格式，或者只粘贴单元格中的格式而不包含数据，此时就需要使用“粘贴选项”功能。

对数据执行剪切或复制操作后，右击目标单元格，在弹出的快捷菜单的“粘贴选项”类别中包含几个命令提供了几种常用的粘贴方式，如图 2-45 所示。例如，如果希望去除数据本身带有的格式，而以纯文本格式的方式粘贴数据，则可以选择“粘贴选项”中的“值”命令，如图 2-46 所示。



图 2-45 粘贴选项



图 2-46 以“值”的方式粘贴数据

如需查看完整的粘贴选项，可以选择如图 2-45 所示的菜单中的“选择性粘贴”命令，将打开“选择性粘贴”对话框，从中选择所需的粘贴方式，如图 2-47 所示。



图 2-47 “选择性粘贴”对话框

2.2.4 快速转换数据的行和列

示例文件/第2章/快速转换数据的行和列.xlsx

如图 2-48 所示是 2.1.1 小节介绍 Excel 数据类型的示例，借助“选择性粘贴”功能，可以快速将数据中的行变成列，将数据中的列变成行。

	A	B	C	D	E
1	文本	数值	日期和时间	逻辑值	错误值
2	Excel	666	2025/3	TRUE	#NUM!
3	数据分析	888	2025年3月1日	FALSE	#VALUE!

图 2-48 待转换的数据

操作步骤如下：

- (1) 选择数据区域 A1:E3，然后按 Ctrl+C 组合键，复制选中的数据。

（2）右击当前工作表或其他工作表中的任意一个空白单元格，在弹出的快捷菜单中选择“粘贴选项”的“转置”命令，如图 2-49 所示。

转换数据的行和列之后的结果如图 2-50 所示，原来的 3 行转换后变成了现在的 3 列，原来的 5 列转换后变成了现在的 5 行。



图 2-49 选择“转置”命令

	A	B	C
1	文本	Excel	数据分析
2	数值	666	888
3	日期和时间	2025/3	2025年3月1日
4	逻辑值	TRUE	FALSE
5	错误值	#NUM!	#VALUE!

图 2-50 转换行和列后的数据

2.3 保存数据

在 Excel 中输入或导入数据后，为了便于以后反复使用，需要将数据以文件的形式存储到计算机中。Excel 支持将数据保存为多种文件格式，可以指定保存文件时的默认格式。为了减少由于意外关闭 Excel 程序而导致数据损失，还可以设置创建用于恢复工作簿的临时备份文件的时间间隔。

2.3.1 Excel 支持的工作簿类型

“工作簿”是对 Excel 文件的特殊称呼，每个工作簿就是一个 Excel 文件，它们存储在计算机磁盘中。Excel 支持的工作簿类型分为普通工作簿、模板、加载宏 3 类。普通工作簿是平时最常用的工作簿类型，在其中可以输入数据、为数据设置格式、计算和分析数据、使用图表展现数据等。模板用于快速创建具有相同或相似格式和内容的多个工作簿。加载宏是一种扩展和增强 Excel 功能的工作簿，其中包含编写好的 VBA 代码，将加载宏工作簿安装到 Excel 中后，可以在每个打开的工作簿中使用加载宏工作簿中的功能。

以上 3 类工作簿都有两种不同的文件格式，以 Excel 2007 作为分界线。以普通工作簿为例，Excel 2007 之前的工作簿的文件扩展名为 .xls，而从 Excel 2007 开始的工作簿的文件扩展名改为 .xlsx。

前两类工作簿还可细分为包含 VBA 代码和不包含 VBA 代码两种文件版本。在 Excel 2003 中，无论工作簿是否包含 VBA 代码，都使用同一种文件格式存储数据。在 Excel 2007 以及更高版本中，文件格式将根据工作簿是否包含 VBA 代码而有所不同。表 2-2 列出了 Excel 支持的工作簿类型。

表 2-2 Excel 支持的工作簿类型

工作簿类型	是否可以存储 VBA 代码	文件扩展名
Excel 工作簿	不可以	.xlsx
Excel 启用宏的工作簿	可以	.xlsm
Excel 97-2003 工作簿	可以	.xls
Excel 模板	不可以	.xltx
Excel 启用宏的模板	可以	.xltm
Excel 97-2003 模板	可以	.xlt
Excel 加载宏	可以	.xlam
Excel 97-2003 加载宏	可以	.xla

2.3.2 将数据保存为 Excel 工作簿

为了随时查看和编辑工作簿，需要将工作簿中的内容保存到计算机中，有以下几种方法：

- 单击快速访问工具栏中的“保存”命令。
- 单击“文件”按钮，然后选择“保存”命令。
- 按 Ctrl+S 组合键。

无论使用哪种方法，都会显示“另存为”界面，如图 2-51 所示。选择一个保存位置，将打开“另存为”对话框，如图 2-52 所示，输入工作簿的名称并单击“保存”按钮，即可保存工作簿。



图 2-51 选择保存位置

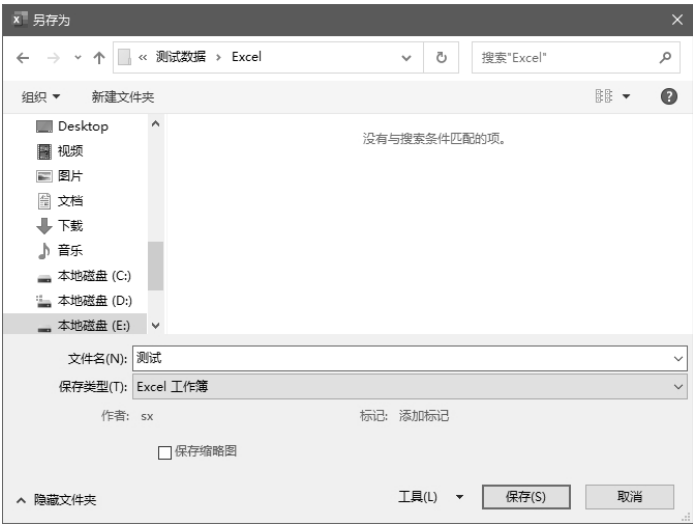


图 2-52 “另存为”对话框

提示：如果已将工作簿保存到计算机中，则在执行保存操作时，会将上次保存后的最新修改直接保存到当前工作簿中，而不会显示“另存为”界面和“另存为”对话框。

2.3.3 设置保存工作簿的默认文件格式

在 Excel 2007 以及更高版本中，每次保存工作簿时的默认格式为“Excel 工作簿”，文件扩展名为 .xlsx。如果经常需要将工作簿保存为其他类型的文件格式，可以更改工作簿的默认保存文件格式，操作步骤如下：

(1) 单击“文件”按钮，然后选择“更多”→“选项”命令，如图 2-53 所示。



图 2-53 选择“选项”命令

(2) 打开“Excel 选项”对话框，在“保存”选项卡的“将文件保存为此格式”下拉列表中选择所需的文件格式，然后单击“确定”按钮，如图 2-54 所示。

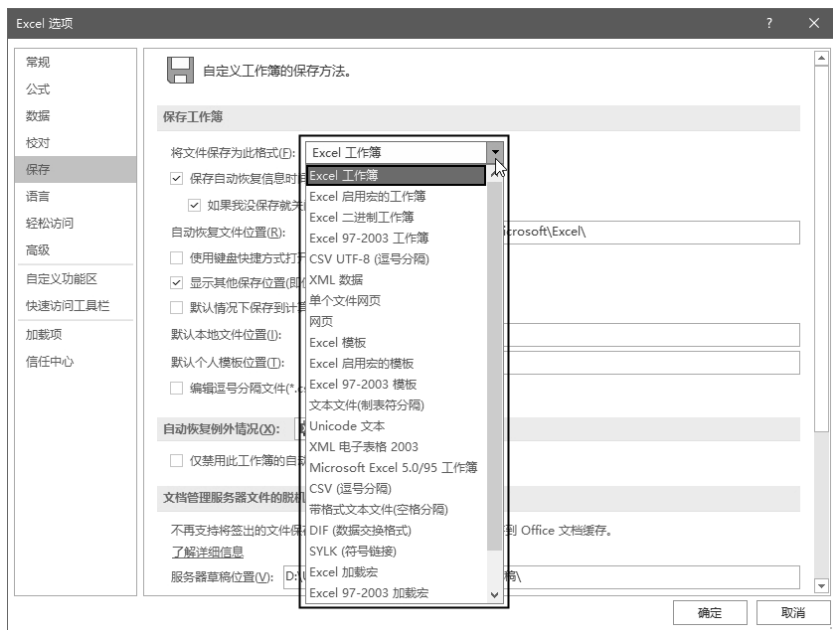


图 2-54 设置保存工作簿的默认文件格式

2.3.4 设置自动恢复工作簿的保存时间间隔和位置

默认情况下，Excel 会每隔 10 分钟自动保存当前打开工作簿的一个临时备份，一旦 Excel 意外关闭，可以在下次启动 Excel 时使用临时备份文件恢复上次意外关闭 Excel 时处于打开状态的工作簿。为了减少数据损失，可以将保存临时备份文件的时间间隔缩短，操作步骤如下：

(1) 单击“文件”按钮，然后选择“更多”→“选项”命令。

(2) 打开“Excel 选项”对话框，在“保存”选项卡中勾选“保存自动恢复信息时间间隔”复选框，然后在右侧的文本框中输入以“分钟”为单位的数字，表示保存临时备份文件的时间间隔。在下方的“自动恢复文件位置”文本框中可以设置保存临时备份文件的位置，如图 2-55 所示，完成后单击“确定”按钮。

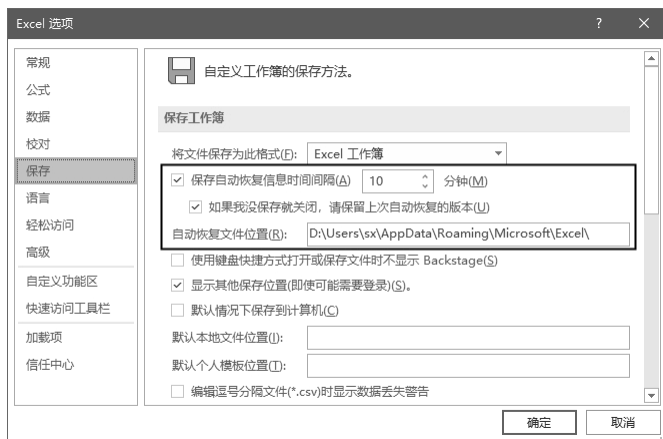


图 2-55 设置自动恢复工作簿的保存时间间隔和位置