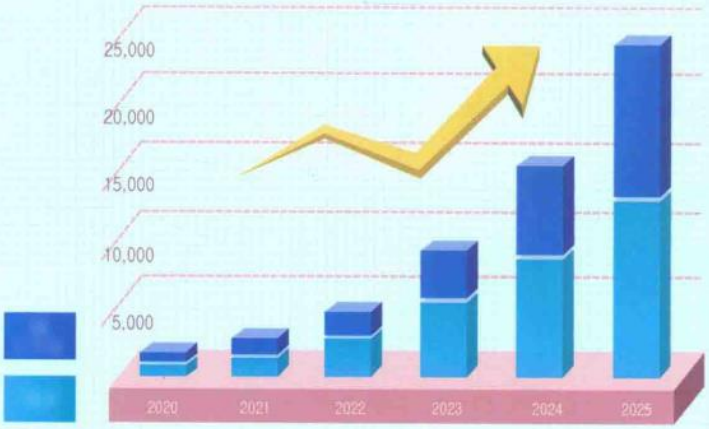
Data Analysis with Excel

# Excel数据分析

主 编 樊玲 曹聪




北京邮电大学出版社
www.buptpress.com

## 内 容 简 介

本书使用通俗易懂的语言、丰富的实例、简洁的图表和紧凑的数学公式，讲述了 Excel 在数据分析方面的应用。既介绍了相关的数理统计知识，又与 Excel 操作结合起来，借助可视化图表，为读者展示了 Excel 强大的数据分析能力。

本书内容包括 Excel 2019 简介、随机变量、抽样、参数估计、假设检验、方差分析、相关分析与回归分析、时间序列分析、聚类分析与判别分析。

本书可作为高等院校研究生、本科生及大学预科生的数据分析教材，也可供对数据分析感兴趣的 IT 人员、数据分析师、管理人员阅读参考。

**图书在版编目(CIP)数据**

Excel 数据分析 / 樊玲，曹聪主编. -- 北京：北京邮电大学出版社，2021.2

ISBN 978-7-5635-6252-7

Ⅰ. ①E… Ⅱ. ①樊… ②曹… Ⅲ. ①表处理软件 Ⅳ. ①TP391.13

中国版本图书馆 CIP 数据核字(2020)第 210325 号

**策划编辑**：刘纳新　姚　顺　　**责任编辑**：刘春棠　　**封面设计**：七星博纳

**出版发行**：北京邮电大学出版社

**社　　址**：北京市海淀区西土城路 10 号

**邮政编码**：100876

**发 行 部**：电话：010-62282185　传真：010-62283578

**E-mail**：publish@bupt.edu.cn

**经　　销**：各地新华书店

**印　　刷**：北京鑫丰华彩印有限公司

**开　　本**：787 mm×1 092 mm　1/16

**印　　张**：15.25

**字　　数**：401 千字

**版　　次**：2021 年 2 月第 1 版

**印　　次**：2021 年 2 月第 1 次印刷

ISBN 978-7-5635-6252-7　　定价：39.00 元

# 前　言

Excel 是 Office 的基本组件，大家可能会经常用到，但大部分人对 Excel 知之甚少，甚至很多人只会用 Excel 画表格。本书就是想告诉大家，如何利用最常见的软件 Excel 来做听起来不那么常见的数据分析工作。

本书不是一本 Excel 初级教程，它介绍的是 Excel 数据分析工具和统计函数的应用，因此建议读者有一定的 Excel 基础。为了方便 Excel 初级读者学习，本书的第 1 章对 Excel 的基础知识以及与统计相关的函数和方法进行了简单介绍。本书也不是一本数理统计的教材。和很多工科读者的习惯一样，我们不对数理统计学的数学定理进行详细的证明和分析，仅学习定理并应用结论。霍金曾经说过，多写一个公式，就会少一半的读者。虽然我们一省再省，最后还是列出来一些基本的统计公式，方便大家查阅。

需要提醒读者的是，实验的过程中，软件能帮助我们快速地得到结论，但是在真正的实验中，并不能完全地依赖软件，一个优质的实验结果更来自你对实验目的的充分理解、对实验过程的独特设计。如果你希望这是一个傻瓜程序，不需要你思考，那么本书或许达不到你的要求。作为一本实验书，建议读者把对应的案例实际操作一遍。为了方便读者更好地学习和应用，本书实际案例的截图很多，阅读的时候需要对照图表和前文的公式或方法才能读懂，我们确实还没有找到更好的说明方法帮助大家理解。

数据是一种资源，可以重复使用、不断产生新的价值。数据到大数据，不仅是量的积累，更是质的飞跃。原本孤立的数据通过被整合、分析变得互相联通，然而大数据并不能被直接拿来使用，统计学依然是数据分析的灵魂。首先，大数据告知信息但不解释信息；其次，事物本身在不断地发展和变化，数据也随之发生着变化。通过数据分析，既研究如何从数据中把信息和规律提取出来，寻找最优化的方案，也研究如何把数据当中的不确定性量化出来。大数据的特点确实对数据分析提出了全新挑战。许多传统统计方法应用到大数据上，巨大的计算量和存储量往往使其难以承受；对结构复杂、来源多样的数据，如何建立有效的统计学模型也需要新的探索和尝试。对于新时代的数据科学而言，这些挑战同时也意味着巨大的机遇，也有可能会产生新的思想、方法和技巧。

本书参考了《概率论与数理统计(第 4 版)》(盛骤、谢式千、潘承毅编,高等教育出版社)、《Excel 统计分析典型实例》(马俊编著,清华大学出版社)、《Excel 函数与公式应用大全》(Excel Home 编著,北京大学出版社),同时高诗琪同学参与了本书的编写、整理和统稿工作,在此一并表示感谢。

# 目　录

# 第 1 章　Excel 简介

Microsoft Excel 是世界上应用最广泛的电子表格程序之一，同时也是 Microsoft Office 套件的一部分。Excel 可以进行各种数据的处理、统计分析和辅助决策操作、数据排序和筛选、自定义公式和文本输入等。Excel 中有大量的公式函数可以应用选择，使用 Microsoft Excel 可以执行计算、分析信息并管理电子表格或网页中的数据信息列表以及数据资料图表制作，支持 Visual Basic For Application 编程，以执行特定功能或重复性高的操作。目前，Microsoft Excel 被广泛应用于管理、统计财经、金融等众多领域。

Excel 的魅力在于它的通用性。虽然 Excel 的优势是进行数字计算，但它对于非数字应用也是非常有用的。Excel 的主要用途包括以下几个方面。

- 数值处理：创建预算、分析调查结果和实施所能想到的任何类型的财务分析。
- 创建图表：创建多种完全可自定义的图表。
- 组织列表：使用行—列布局有效地存储列表。
- 访问其他数据：从多种数据源导入数据。
- 创建图形：使用“形状”和全新的 Smart Art 创建具有专业观感的图表。
- 自动化复杂的任务：借助 Excel 的宏功能，通过单击一次鼠标，执行多次重复任务。

Excel 2019 是目前微软公司提供的 Excel 系列中最新的一个版本，能够以新的、直观的方式查看数据。

- 使用建议的图表和选择预测功能预览趋势。
- 使用 PowerPivot 快速关联表格、运行复杂计算。
- 查询、整理与合并企业及云数据源的数据。
- 使用 Tree Map(树状图)和 Waterfall(瀑布图)等新型图表实现数据可视化。
- 使用 Tell Me(告诉我)搜索栏获取 Excel 即时帮助。

Excel 2019 系统要求如表 1-1 所示。

**表 1-1　Excel 2019 系统要求**

| | |
|---|---|
| 处理器要求 | 1 000 MHz 或更快的 x86 或 x64 处理器，采用 SSE2 指令集 |
| 操作系统要求 | Windows 7 或更高版本、Windows 10 Server、Windows Server 2012 R2、Windows Server 2008 R2 或 Windows Server 2012 |
| 内存要求 | 1 GB RAM(32 位)、2 GB RAM(64 位) |
| 硬盘空间要求 | 3.0 GB 可用磁盘空间 |
| 显示要求 | 1 024 像素×768 像素分辨率 |
| 图形 | 图形硬件加速需要 DirectX 10 图形卡 |
| 多点触控 | 需要支持触控的设备才能使用任何多点触控功能。不过，所有功能始终可以通过键盘、鼠标、其他标准或无障碍输入设备来使用 |

本书所有案例均在 Excel 2019 环境下进行操作，除了 Excel 2019 新添的个别函数（如 IFS 等）以外，其余案例均可在 Excel 的其他版本环境下运行。

## 1.1 Excel 界面简介

在 Excel 中所做的工作是在一个工作簿文件中执行的，打开该文件后有其自己的窗口。可以根据需要打开任意多个工作簿。默认状态下，Excel 工作簿（2007 以后版本）使用 .xlsx 作为文件扩展名。本书均以 Excel 2019 为例，进行操作。

每个工作簿由一个或多个工作表组成，每个工作表由独立的单元格组成。每个单元格包括值、公式或文本。工作表也有不可见的绘图层，该层包含图表、图像和图形。通过单击工作簿窗口底部的标签可访问工作簿中的每个工作表。除此之外，工作簿还可以存储图表。“图表”显示一个单独的图，也可通过单击标签进行访问。Excel 2019 的界面介绍如图 1-1 所示。

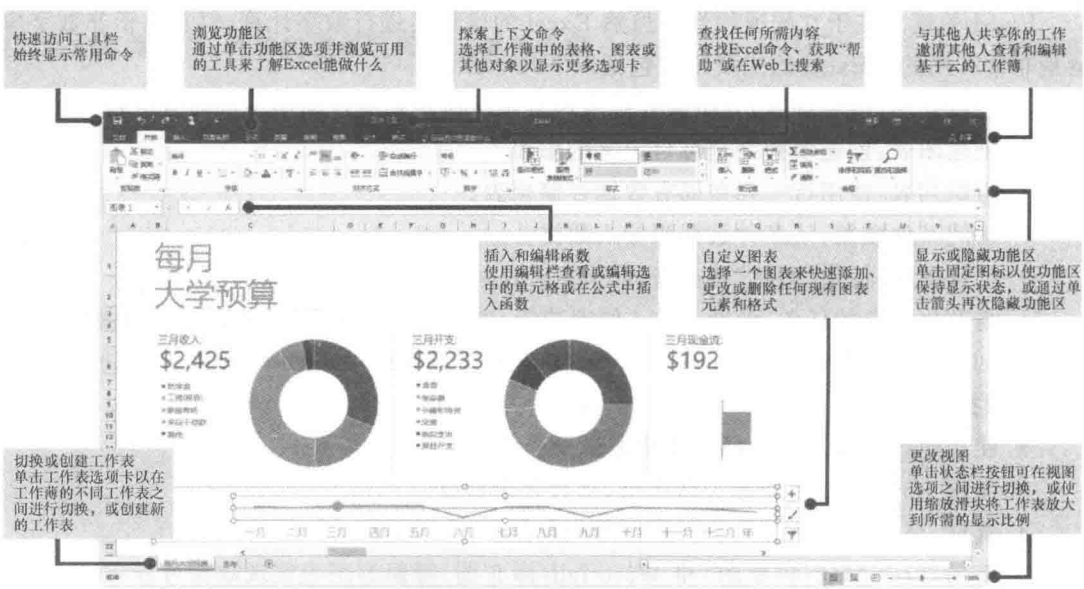



图 1-1 Excel 2019 界面介绍

表 1-2 对图中显示的各个项目进行简要的说明。

**表 1-2 需了解的 Excel 界面的各个部分**

| 名 称 | 描 述 |
|---|---|
| 活动单元格指示符 | 黑色的轮廓表示当前活动的单元格 |
| 应用程序关闭按钮 | 单击该按钮关闭 Excel |
| 窗口关闭按钮 | 单击该按钮关闭活动中的工作簿 |
| 列字母 | 字母范围从 A 到 XFD 列——对应于工作表中 16 384 列中的每一列。可单击列标题选择单元格的整列 |
| 文件按钮 | 该按钮可引出编辑文档时的很多选项或者通用的 Excel 选项 |

续表

| 名　称 | 描　述 |
| --- | --- |
| 公式栏 | 将信息或者公式输入 Excel 时，它们会出现在该栏中 |
| 水平滚动条 | 可水平滚动工作表 |
| 最大化/还原按钮 | 单击该按钮可以增加工作簿窗口的尺寸直到填满工作簿的整个工作区，如果窗口已经最大化，单击该按钮可以还原 Excel 窗口，使其不再填满整个屏幕 |
| 最小化程序按钮 | 单击该按钮最小化 Excel 窗口 |
| 窗口最小化按钮 | 单击该按钮最小化工作簿窗口 |
| 名称框 | 显示活动单元格地址或所选单元格名称、范围或对象 |
| 页面视图按钮 | 通过单击其中一个按钮可以更改工作簿的显示方式 |
| 快速访问工具栏 | 通过自定义来显示常用命令的工作栏 |
| 功能区 | 查找 Excel 命令的主位置，单击选项卡列表中的项目更改显示的功能区 |
| 行号 | 号码从 1 到 1 048 576——每个数字对应工作表中的一行。可以单击行号选择单元格中的整行 |
| 工作表标签 | 每一个类似笔记本的标签代表工作簿中的一个不同的表，一个工作簿中可以有任意数量的表，每个表的名字都显示在标签上。Excel 2019 在默认状态下，每个新建的工作簿含有 1 个表，可以通过单击“插入工作表”按钮添加一个新表 |
| 工作表标签滚动按钮 | 可滚动工作表标签以显示不可见的标签 |
| 选项卡列表 | 显示不同的功能区命令，类似于菜单 |
| 标题栏 | 显示程序的名称及当前工作簿的名称 |
| 垂直滚动条 | 可垂直滚动工作表 |
| 缩放(Zoom)控件 | 可任意放大或缩小工作表 |

Excel 2019 同样取消了传统的菜单操作方式，而使用各种功能区。在 Excel 2019 窗口上方看起来像菜单的名称其实是功能区的名称，当单击这些名称时并不会打开菜单，而是切换到与之相对应的功能区面板。每个功能区根据功能的不同又分为若干个组。

根据所选择的选项卡不同，功能区中可用的命令将有所不同。功能区按照一组相关的命令进行排列。下面简要说明 Excel 的选项卡。

(1) “开始”选项卡中包括剪贴板、字体、对齐方式、数字、样式、单元格和编辑几个组，如图 1-2 所示。此功能主要用于对 Excel 2019 表格进行文字编辑和单元格的格式设置，是最常用的功能区。


图 1-2　“开始”选项卡

(2) “插入”选项卡中包括表格、插图、加载项、图表、演示、迷你图、筛选器、链接、文本和符号几个组，如图 1-3 所示。当需要在工作表中插入各种对象时选择该选项卡。

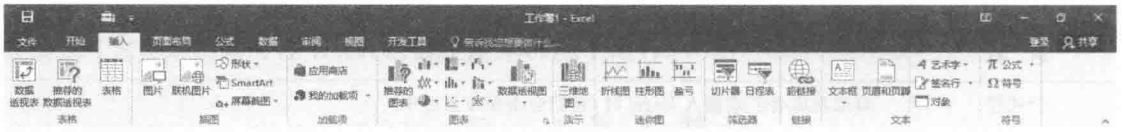


图 1-3 “插入”选项卡

(3)“页面布局”选项卡中包括主题、页面设置、调整为合适大小、工作表选项、排列几个组,如图 1-4 所示。此功能用于帮助用户设置 Excel 2019 工作簿页面样式,包括影响整个工作表外观的命令,也包含打印设置。

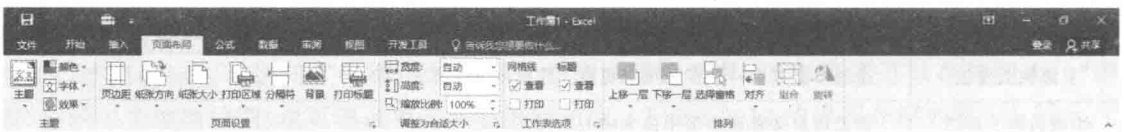


图 1-4 “页面布局”选项卡

(4)“公式”选项卡包括函数库、定义的名称、公式审核和计算几个组,如图 1-5 所示。此功能主要用于在 Excel 2019 表格中进行各种数据计算。

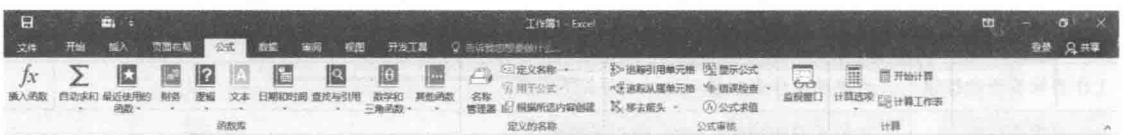


图 1-5 “公式”选项卡

(5)“数据”选项卡包括获取外部数据、获取和转换、连接、排序和筛选、数据工具、预测、分级显示和分析几个分组,如图 1-6 所示。此功能主要用于在 Excel 2019 表格中进行数据处理相关的操作。

图 1-6 “数据”选项卡

(6)“审阅”选项卡包括校对、中文简繁转换、见解、语言、批注和更改几个组,如图 1-7 所示。此功能主要用于对 Excel 2019 表格进行校对和修订等操作,也包含拼写检查、翻译文字、添加批注或保护工作表工具。

图 1-7 “审阅”选项卡

(7)“视图”选项卡包括工作簿视图、显示、显示比例、窗口和宏几个分组,如图 1-8 所示。

此功能主要用于设置 Excel 2019 表格窗口的视图类型。

图 1-8　“视图”选项卡

(8)“开发工具”选项卡包括代码、加载项、控件和 XML 几个组，如图 1-9 所示。此功能主要是方便程序员使用，利用它可以进行一些编程工作，插入表单控件和 Activex 控件等。

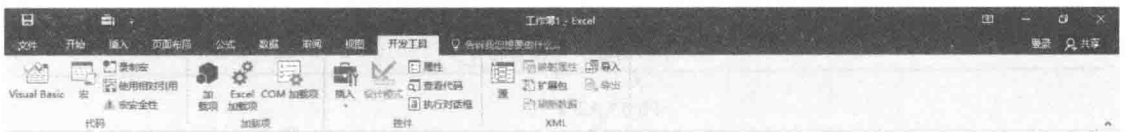


图 1-9　“开发工具”选项卡

**注意**：“开发工具”选项卡在默认状态下是不可见的，需要用户选择添加显示。要显示“开发工具”选项卡，选择“Excel 按钮”菜单中的“选项”命令，并选择“自定义功能区”，勾选“主选项卡”中的“开发工具”复选框，如图 1-10 所示。


图 1-10　添加“开发工具”选项卡

## 1.2 Excel 公式与函数初步

Excel 不仅是一个可在列或行中输入数字的网格，还可以使用 Excel 求出一列或一行数字的总和，也可根据自己插入的变量计算抵押贷款付款、解答数学或工程问题、找到最佳情况方案。

Excel 公式是 Excel 工作表中进行数值计算的等式，Excel 公式用于对工作表中的数据执行计算或其他操作。Excel 公式的组成包括一个等号“=”和一个或者多个运算码。运算码包含下列所有内容或其中之一：函数、引用、运算符和常量，如图 1-11 所示。

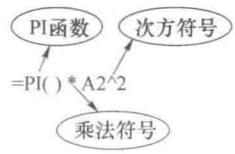



图 1-11 Excel 公式

- 函数：PI( ) 函数返回 PI 值：3.14159……
- 引用：A2 返回单元格 A2 中的值。
- 常量：直接输入公式中的数字或文本值，如 2。
- 运算符：*（星号）运算符表示数字的乘积，而^（脱字号）运算符表示数字的乘方。

**1. 运算符**

运算符可指定要对公式元素执行的计算类型。Excel 遵循常规数学规则进行计算，即括号、指数、加减乘除，或首字母缩写 PEMDAS（Please Excuse My Dear Aunt Sally）。可使用括号更改计算次序。

计算运算符分为 4 种不同类型：算术、比较、文本连接和引用。

（1）算术运算符

若要进行基本的数学运算（如加法、减法、乘法或除法）、合并数字以及生成数值结果，可使用表 1-3 所示的算术运算符。

**表 1-3 算术运算符**

| 算术运算符 | 含 义 | 示 例 |
|---|---|---|
| +（加号） | 加法 | =3+3 |
| -（减号） | 减法<br>负数 | =3-3<br>=-3 |
| *（星号） | 乘法 | =3*3 |
| /（正斜杠） | 除法 | =3/3 |
| %（百分号） | 百分比 | 30% |
| ^（脱字号） | 乘方 | =3^3 |

（2）比较运算符

可使用表 1-4 所示的运算符比较两个值。使用这些运算符比较两个值时，结果为逻辑值

TRUE 或 FALSE。

**表 1-4　比较运算符**

| 比较运算符 | 含　义 | 示　例 |
| --- | --- | --- |
| =(等号) | 等于 | =A1=B1 |
| >(大于号) | 大于 | =A1>B1 |
| <(小于号) | 小于 | =A1<B1 |
| >=(大于或等于号) | 大于或等于 | =A1>=B1 |
| <=(小于或等于号) | 小于或等于 | =A1<=B1 |
| <>(不等号) | 不等于 | =A1<>B1 |

(3) 文本连接运算符

可以使用与号 (&) 连接(联接)一个或多个文本字符串,以生成一段文本,如表 1-5 所示。

**表 1-5　文本连接运算符**

| 文本连接运算符 | 含　义 | 示　例 |
| --- | --- | --- |
| &(与号) | 将两个值连接(或串联)起来产生一个连续的文本值 | (1) ="North"&"wind" 的结果为 "Northwind"<br>(2) A1 代表 "Last name",B1 代表 "First name",则 =A1&", "&B1 的结果为 "Last name, First name" |

(4) 引用运算符

可以使用表 1-6 所示的引用运算符对单元格区域进行合并计算。

**表 1-6　引用运算符**

| 引用运算符 | 含　义 | 示　例 |
| --- | --- | --- |
| : | 区域运算符,生成一个对两个引用之间所有单元格的引用(包括这两个引用) | B5:B15 |
| , | 联合运算符,将多个引用合并为一个引用 | =SUM(B5:B15,D5:D15) |
| (空格) | 交集运算符,生成一个对两个引用中共有单元格的引用 | B7:D7 C6:C8 |

**2. 公式**

公式的录入有三种方法。

(1) 直接输入

例如,要计算算式 2×3+5,那么在相应单元格内输入“=2*3+5”,如图 1-12 所示。

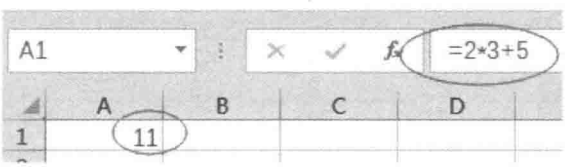



图 1-12　直接输入数据

输出单元格显示结果“11”，而编辑框内显示公式“=2*3+5”。

如果在公式中使用常量而不是对单元格的引用(例如=2*3+5)，则仅在修改公式时结果才会变化。通常，为了轻松查找和更改常量，会将常量放置在指定单元格内，然后在公式中引用这些单元格。

(2) 输入计算式

有时候我们需要对指定单元格的值进行计算，而不是具体的数值计算，这时计算结果会随着指定单元格值的变化而变化。

**例 1.1** 在图 1-13 中，将单元格 A1、B1 和 C1 中的值相加。

| | A | B | C |
|---|---|---|---|
| 1 | 7 | 96 | 42 |
| 2 | 23 | 3 | 23 |
| 3 | 27 | 29 | 22 |
| 4 | 71 | 47 | 52 |
| 5 | 70 | 95 | 6 |
| 6 | 65 | 36 | 92 |

图 1-13 求和数据

要将单元格 A1、B1 和 C1 中的值相加，那么在单元格内输入“=A1+B1+C1”，如图 1-14 所示。

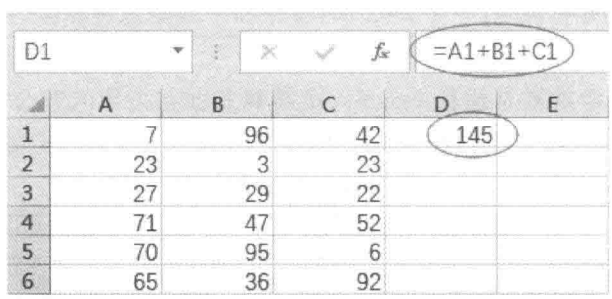



图 1-14 输入计算式

输出单元格显示结果“145”，而编辑框内显示公式“=A1+B1+C1”。

(3) 调用函数输入

例 1.1 也可以调用函数完成求和运算，在相应单元格中输入 “=SUM(A1:C1)”，如图 1-15 所示。

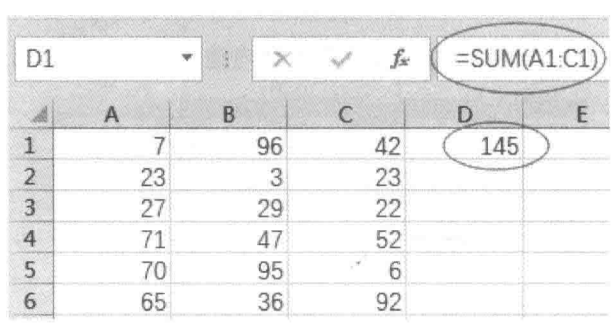



图 1-15 调用函数输入

### 3. 函数

表 1-7 为最常用的 10 个 Excel 函数。

**表 1-7　Excel 中常用的 10 个函数**

| 函　数 | 说　明 |
| --- | --- |
| SUM | 此函数用于对单元格中的值求和 |
| IF | 此函数用于在条件为真时返回一个值，条件为假时返回另一个值 |
| LOOKUP | 需要查询一行或一列并查找另一行或列中相同位置的值时，使用此函数 |
| VLOOKUP | 如果需要按行查找表或区域中的内容，使用此函数。例如，按员工号查找某位员工的姓氏，或通过查找员工的姓氏查找该员工的电话号码(就像使用电话簿) |
| MATCH | 此函数用于在单元格区域中搜索某项，然后返回该项在单元格区域中的相对位置 |
| CHOOSE | 此函数用于根据索引号从最多 254 个数值中选择一个 |
| DATE | 此函数用于返回代表特定日期的连续序列号。此函数在公式而非单元格引用提供年、月和日的情况中非常有用 |
| DAYS | 此函数用于返回两个日期之间的天数 |
| FIND、FINDB | 函数 FIND 和 FINDB 用于在第二个文本串中定位第一个文本串。这两个函数返回第一个文本串的起始位置的值，该值从第二个文本串的第一个字符算起 |
| INDEX | 此函数用于返回表格或区域中的值或值的引用 |

下面以最常见的 IF 函数为例，对 Excel 中的函数进行介绍。IF 函数主要的功能是对结果值和期待值进行逻辑比较，判断是否满足某个条件，如果满足该条件则返回一个值，如果不满足则返回另一个值。

IF 函数语法为

```
IF(Logical_test,Value_if_true,Value_if_false)
```

IF 函数参数如图 1-16 所示。

- Logical_test：必需，表示判断的条件。
- Value_if_true：必需，表示如果判断条件为真时显示的值。
- Value_if_false：必需，表示如果判断条件为假时显示的值。

因此，IF 语句可能有两个结果。第一个结果是比较结果为 True，第二个结果是比较结果为 False。


图 1-16　IF 函数参数