



第4章

数据处理

任务1 采集数据

主编 | 傅连仲 等

目 录

Contents

- 4.1.1 了解常用数据处理软件
- 4.1.2 掌握数据采集的基本方法
- 4.1.3 格式化数据和表格

采集数据

数据处理的前提是数据本身，因此，收集和保存数据是一切数据处理和数据分析的基础。本节通过对常用数据处理软件的介绍，使读者掌握有关数据采集、存储及数据类型的处理和格式化的相关技能。



任务情景

为了学习数据处理的基本技能，小华按照老师布置的实训任务，创建如图所示的某图书销售公司的销售数据情况表。

	A	B	C	D
1	某图书销售	公司销售情况表		
2	经销部门	图书类别	数量(册)	销售额(元)
3	第1分部	计算机类	1596	111720
4	第2分部	计算机类	1290	90300
5	第3分部	计算机类	1540	107800
6	第1分部	社科类	1615	80750
7	第2分部	社科类	993	49650
8	第3分部	社科类	1232	56760
9	第1分部	少儿类	2126	63780
10	第2分部	少儿类	1497	44910
11	第3分部	少儿类	1492	44760

任务分析

开始数据处理的前提是启动数据处理软件，输入数据信息、设置表格格式是制作电子表格的基础，可靠、有效地保存表格数据是完成电子表格制作的关键，在老师的指导下，小华将工作分解为以下三个任务：

任务一：运行相应的数据处理软件，通过数据处理软件创建如图4-11所示的数据表格，并保存在磁盘指定位置；

任务二：观察表格中的数据，了解数据类型的基本知识及不同类型数据的特点；

任务三：根据需求对表格进行必要的格式化设置。

4.1.1 了解常用数据处理软件

常用数据处理软件包括Excel、WPS表格、SQL Server、SAS、SPSS、Tableau、R等，其中WPS表格、Excel适用于一般的办公环境，SQL Server、SAS、SPSS、Tableau、R常用于专业领域。

软件名称	功能和特点
WPS表格	WPS表格是WPS Office套装中的一个组件，与Excel的应用领域和功能类似。此外，还有自身的特点：快速输入已输入过的内容、插入复杂的流程图、带计数的筛选面板、应用XML数据交换技术，无障碍兼容.xlsx文件格式，可以方便地实现与Excel相互编辑
Excel	Excel作为桌面办公软件，广泛地应用于管理、金融等众多领域，它也是应用广泛的分析软件工具，其主要功能包括：制作表格、处理数据、分析数据、数据可视化等功能，再结合VBA编程，可以提高数据自动处理程度
SQL Server	SQL Server 是一款关系型数据库管理系统，是适合中小型企业的数据管理和分析平台，包括电子商务、数据仓库和业务流程解决方案所需的基本功能

软件名称	功能和特点
SAS	SAS (Statistical Analysis System) 是大型应用软件系统, 适用于商业智能分析。使用SAS进行制造业分析, 能够调整生产运营以实现最低成本和风险, 同时利用数据资产, 越来越高效的来生产和运输商品
SPSS	SPSS在调查统计行业、市场研究行业、研究机构中广泛用于数据分析, 主要功能包括描述性分析、差异性分析、多元回归分析等。SPSS提供了数据管理、统计分析、统计绘图和统计报表等功能
Tableau	Tableau 是一款较为简单的商业智能分析软件。使用Tableau能够分析处理连锁店的实际销售数据, 实现对员工的生产力进行分析。Tableau的核心本质其实就是excel的数据透视表和数据透视图, 但是可视化效果更为优秀。Tableau的功能包括: 大数据处理、快速分析、智能仪表盘

操作步骤

1. 启动WPS表格

- ① 单击Windows的“开始”按钮，打开“开始”菜单。
- ② 选择“开始”→“所有程序”→“WPS表格 专业版”→“WPS表格”命令，即可启动WPS表格。

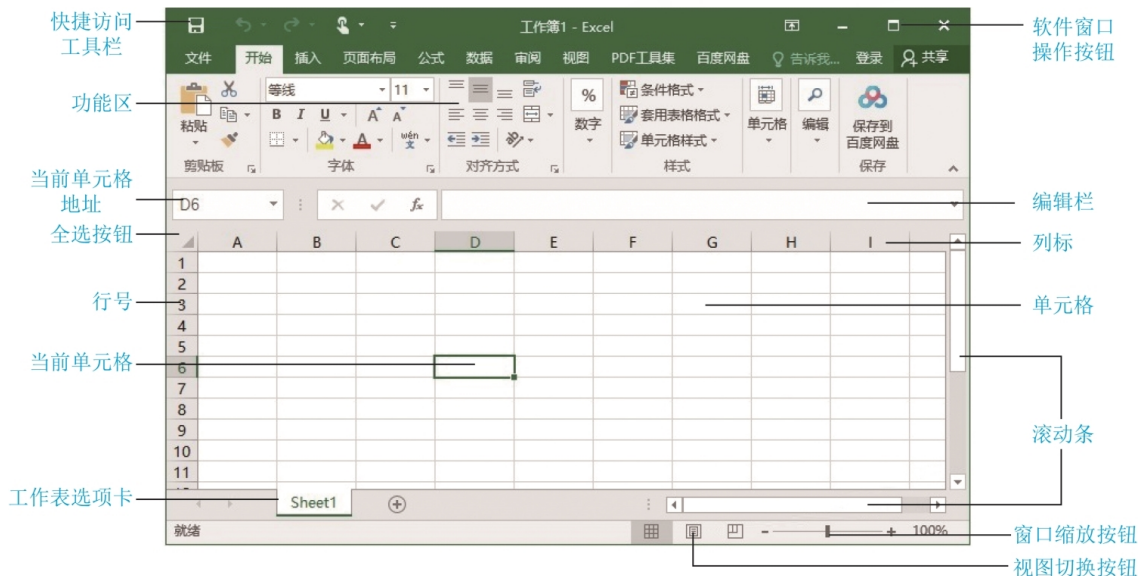
2. 启动Excel

- ① 单击Windows的“开始”按钮，打开“开始”菜单。
- ② 选择“开始”→“所有程序”→“Microsoft Office”→“Microsoft Office Excel”命令，启动Excel。

操作步骤

3. 认识软件界面

WPS表格和Excel启动后界面大致相似，下面以Excel为例，介绍软件界面的基本操作功能，如图所示。



(1) 当前单元格地址框。

当前单元格框用于显示当前单元格或单元格区域的名称或地址，可以在当前单元格框输入单元格名称或地址。

(2) 编辑栏。

编辑栏用于编辑单元格的数据和运算表达式，光标定位在编辑栏后可以从键盘输入文字、数字和运算表达式等。

(3) 全选按钮。

全选按钮用于选中工作表中的所有单元格，单击“全选”按钮可选中整个表格，在任意位置单击时则取消全选。

(4) 行号。

行号是用阿拉伯数字从上到下表示单元格的行坐标，共有1048576行。在行号上单击，可以选中整行。

(5) 列标。

列标是用大写英文字母从左到右表示单元格的列坐标，共有16384列。在列标上单击，可以选中整列。

(6) 单元格。

单元格是Excel中存放数据的最小单位，由列标和行号来唯一确定。单击鼠标左键可以选中单元格。

(7) 工作表选项卡。

用于不同工作表之间的显示切换，由工作表标签和工作表区域构成。单击工作表标签可以切换工作表。

(8) 功能区。

功能区存放各种操作命令按钮，单击命令按钮，即可完成相应操作。

4. 相关概念介绍

(1) 单元格。

单元格是WPS表格和Excel中存放数据的最小单位，在界面上表现为一个个长方形格子，输入的数据保存在这些单元格中。单元格的地址由列标和行号来唯一确定。单击可以选中某个单元格，在这个被选中的单元格中可以直接输入数据，这个被选中的单元格被称为当前单元格，当前单元格的地址显示在当前单元格地址框内。

(2) 工作表。

工作表由单元格组成，工作表的标签名为Sheet1、Sheet2、Sheet3……可以通过单击工作表的标签在不同的工作表之间进行切换，也可以通过双击工作表标签修改工作表的名称。

(3) 工作簿。

工作簿是存储数据的文件，也就是将表格数据保存在磁盘上的文件名。每个工作簿可以包含多个工作表（Sheet），如图4-12所示的Sheet1，用户可以通过单击右侧的“+”或者右击工作表的标签名，插入新的工作表。

说一说

你还曾经使用过哪些数据
处理工具?



4.1.2 掌握数据采集的基本方法

启动数据处理软件后，创建和采集表格的数据内容是后续数据处理的基础，有效而可靠地保存表格数据是完成电子表格制作的关键。数据采集的方法有很多，包括人工录入、外部导入和利用工具软件采集等。在对任务进行简单分析后，小华利用人工录入的方式，按照样表的内容输入数据，同时了解WPS表格和Excel中各种数据类型及其特点、功能，并进一步尝试了外部导入和利用工具软件采集数据的基本方法。

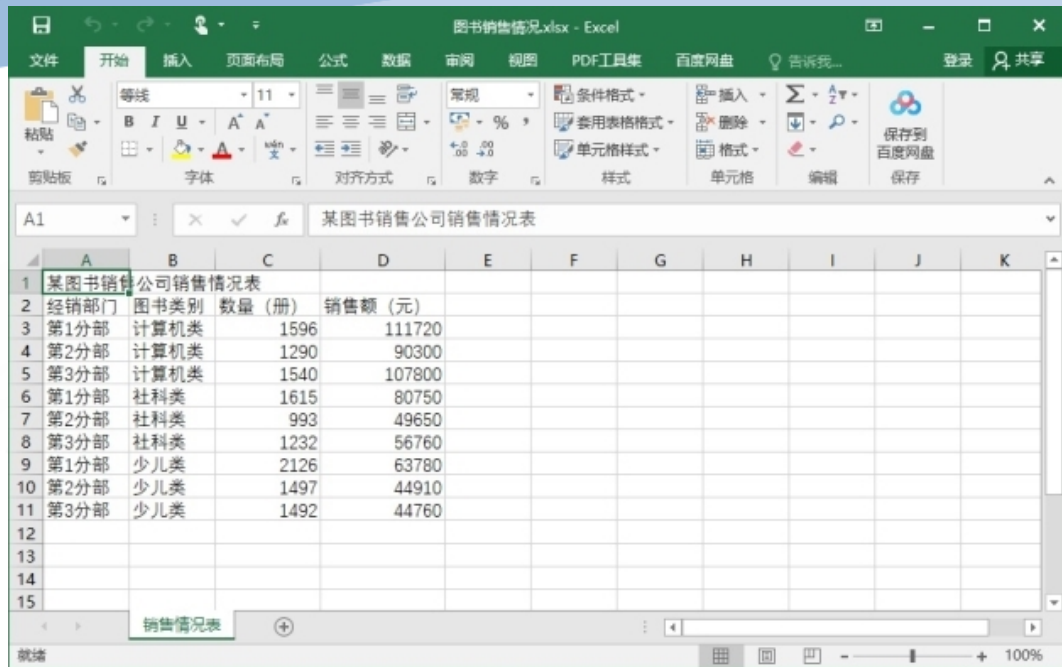
操作步骤

1. 人工录入数据

① 单击相应单元格，输入相应内容后按【Enter】键，直至所有数据录入完毕。

② 双击工作表标签，修改工作表Sheet1名称为“销售情况表”后按【Enter】键。

③ 单击“保存”按钮，选择保存位置，输入工作簿文件名为“图书销售情况”后单击“保存”按钮，如图所示。



(1) 常用数据类型。

数据处理过程中，最为常见的数据类型有以下几种。

◆字符型数据。

字符型数据包括汉字、英文字母、数字、空格等，默认情况下，字符型数据自动沿单元格左边对齐。当输入的字符串超出了当前单元格的宽度时，如果右边相邻单元格中没有数据，那么字符串会往右延伸显示；如果右边单元格中有数据，超出的那部分数据就会被隐藏起来，只有把单元格的宽度调大后才能显示出来。

字符型数据也可以全部由阿拉伯数字组成，如果要输入的字符串全部由数字组成，如产品型号、汽车牌号、存折账号等，特别是需要以“0”开头的数字，为了避免WPS表格和Excel将它按数值型数据处理，在输入时可以先输一个单引号“'”（英文符号），再接着输入具体的数字。例如，要在单元格中输入某产品型号“006401”，则先输入单引号“'”（英文符号），再输入“006401”，然后按【Enter】键，出现在单元格里的就是“006401”，并自动左对齐。

◆数值型数据。

数值型数据包括0~9中的数字，以及含有正号、负号、货币符号、百分号等任一种符号的数据。默认情况下，数值型数据自动沿单元格的右边对齐。

◆日期型数据和时间型数据。

在日常操作中，经常需要录入一些日期型数据和时间型数据，在录入过程中要注意以下几点。

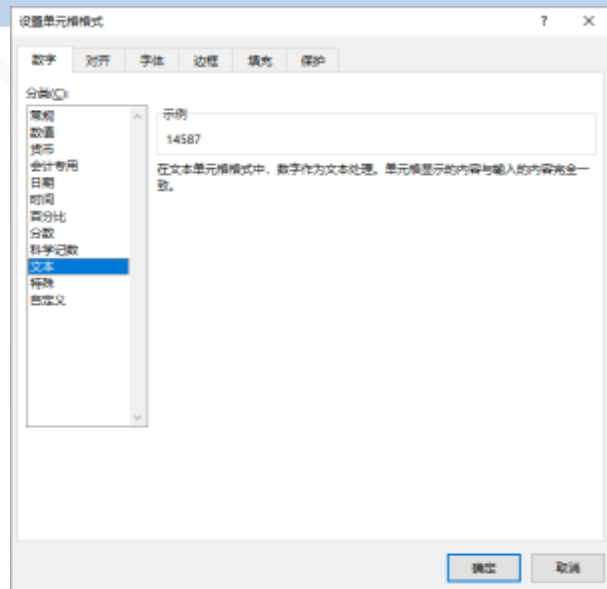
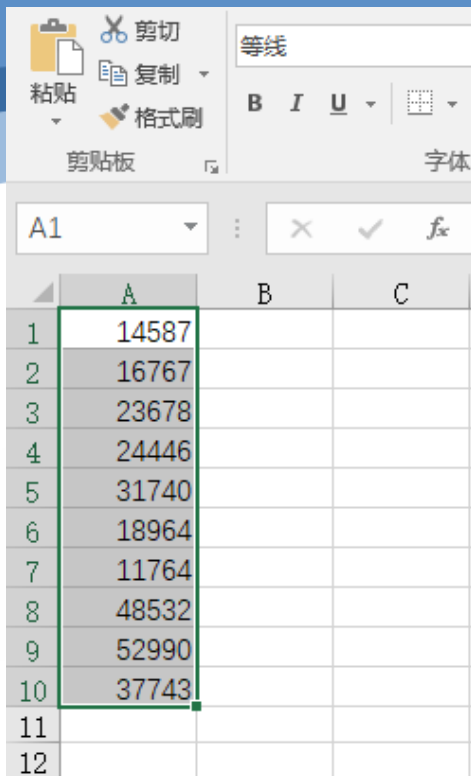
- ① 输入日期时，年、月、日之间要用“/”或“-”隔开，如“2021/3/25”“2021-3-25”。
- ② 输入时间时，时、分、秒之间要用冒号“:”（英文符号）隔开，如“15:30”“3:30PM”“15:30:27”。
- ③ 若要在单元格中同时输入日期和时间，日期和时间之间应该用空格隔开，如“2021/3/25 15:30”。

(2) 分数的录入。

默认情况下，WPS表格和Excel都以小数点的形式显示所录入的非整数数据，但也可以用分数的形式录入。要想在单元格中录入分数形式的数据，应先在单元格中输入“0”和一个空格，然后输入分数。例如，要在单元格中输入分数“ $\frac{2}{3}$ ”，首先在单元格中输入“0”和一个空格，然后输入“2/3”，按【Enter】键后单元格中就会出现分数“ $\frac{2}{3}$ ”；要在单元格中输入分数“ $\frac{5}{3}$ ”，首先在单元格中输入“0”和一个空格，然后输入“5/3”，按【Enter】键后单元格中就会出现分数“ $\frac{5}{3}$ ”。

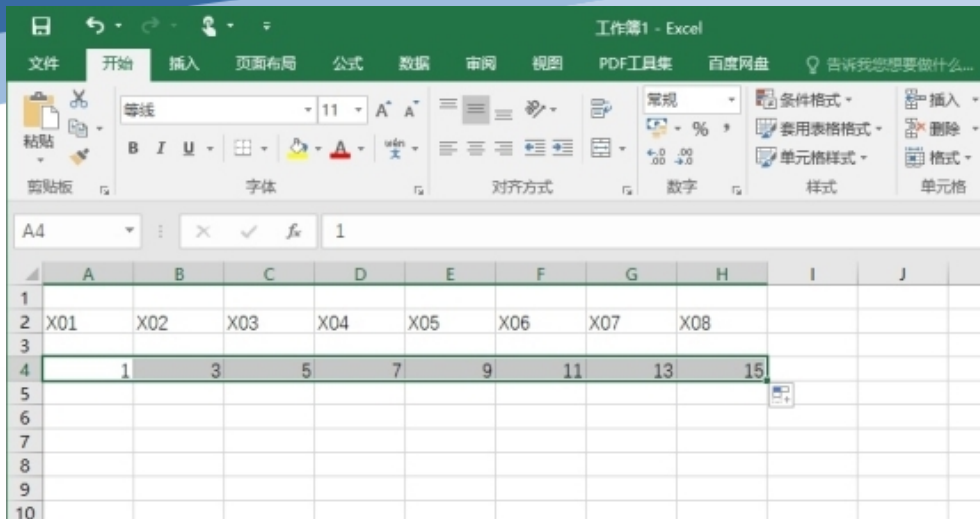
(3) 数据的类型转换。

在实际使用过程中，也可以将已经存储在表格中的数据进行数据类型的转换。例如，需要将图4-14中的数据转换为字符型，则只需选中该区域数据，单击“开始”→“单元格”→“格式”按钮，在下拉菜单中选择“设置单元格格式”命令，打开“设置单元格格式”对话框，然后选择“数字”选项卡，再选择“分类”列表框中的“文本”选项，最后单击“确定”按钮即可。

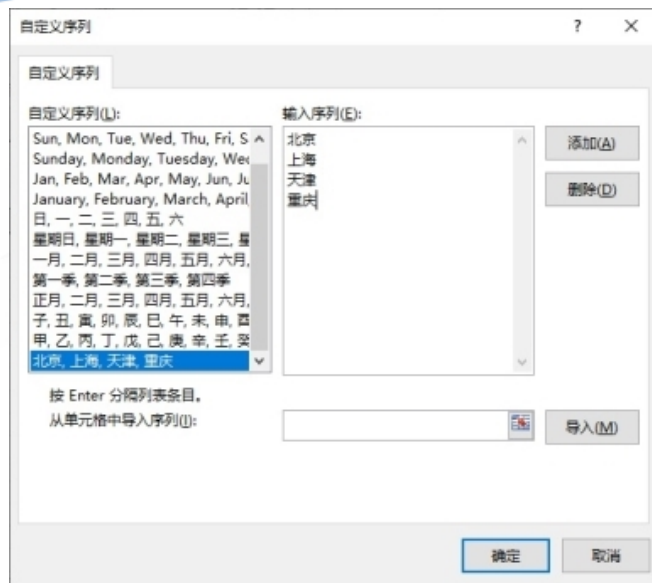


(4) 自动填充功能。

在输入“编号”等具有连续性的数据或有规律变化的数据时，我们可以利用软件提供的自动填充功能来实现快速输入。例如：在A2单元格内输入“X01”，然后选中A2单元格，再将鼠标移至A2单元格右下角的“填充句柄”（实心小方块），这时鼠标光标变成“+”形状，然后按下鼠标左键并拖动到E8单元格，可实现“X01~X08”系列数据的自动填充。利用该方法填充有规律的数字时，如输入“1, 3, 5, 7, 9, ...”序列时，需要先输入1和3，然后同时选中这两个单元格，将鼠标移到第二个单元格右下角“填充句柄”处拖动。



利用“自定义序列”对话框也可以填充序列。选择“文件”→“选项”→“高级”选项，然后找到并单击“编辑自定义列表”按钮，打开“自定义序列”对话框。在“输入序列”窗格中输入清单，每行一项，输完后单击“添加”按钮，如图4-17所示。也可以单击“导入”按钮导入工作表里已有的数据。序列建立完毕后，即可在任意一个单元格中输入序列中的任何项，然后拖动“填充柄”，便可用序列中的其他项去填充单元格。



(5) 数据和表格的编辑与修改。

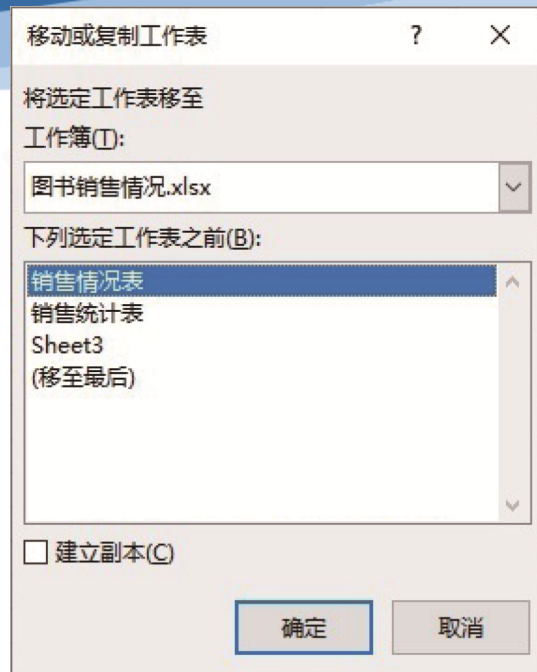
在输入表格数据时，可能会出现输入错误或漏掉数据等问题，这就需要对表格数据进行必要的编辑。在WPS表格和Excel中，编辑电子表格包括对工作表、行、列及单元格的相关操作，也包括对数据的编辑修改、数据的查找与替换、数据的保护等操作。

◆重命名。

双击工作表标签可以进行工作表名称的重命名，完成后按【Enter】键即可。

◆移动、复制工作表操作。

右击工作表标签，在弹出的快捷菜单中选择“移动或复制”命令，打开“移动或复制工作表”对话框，如图所示，选定要复制的工作表。



在此对话框中可以将工作表移动到目标位置。选择“建立副本”复选框则执行复制操作，单击“确定”按钮完成操作。

◆插入行和列。

选中某行，单击“开始”→“单元格”→“插入”按钮，在弹出的下拉列表中选择“插入工作表行”命令，选中的行的上方将产生一个空白行。

同样，选中某列，单击“开始”→“单元格”→“插入”按钮，在弹出的下拉列表中选择“插入工作表列”命令，使选中的列的左侧产生一个空白列。

◆删除数据行或数据列。

选中待删除的行（或列），单击“开始”→“单元格”→“删除”按钮，在弹出的下拉列表中选择“删除工作表行”（或“删除工作表列”）命令即可删除选中的行（或列）。

◆移动、复制行或列的操作。

以行操作为例，右击待移动行的行号，在弹出的快捷菜单中选中“剪切”或“复制”命令，然后右击目标位置，在弹出的快捷菜单中选择“插入剪切的单元格”或“插入复制的单元格”命令。列操作与之类似，此处不再赘述。

◆调整行高或列宽。

选中待调整的行（或列），单击“开始”→“单元格”→“格式”按钮，在下拉菜单中选择“行高”（或“列宽”）命令，在出现的对话框中输入数值即可调整行高（或列宽）。

◆查找与替换数据。

将光标移到待查找与替换的工作表中，单击“开始”→“编辑”→“查找和选择”按钮，在下拉菜单中选择“查找”或“替换”命令，打开“查找和替换”对话框。若是查找，单击“查找”选项卡，直接在“查找内容”文本框中输入要查找的内容即可；若是替换，则单击“替换”选项卡，在“查找内容”文本框中输入要查找的内容，在“替换为”文本框中输入替换后的内容，单击“全部替换”按钮即可完成全部替换工作。

2. 导入外部数据

生成和收集数据的方法有很多，可以是手工录

入，也可以在WPS表格和Excel中通过“数据”选项

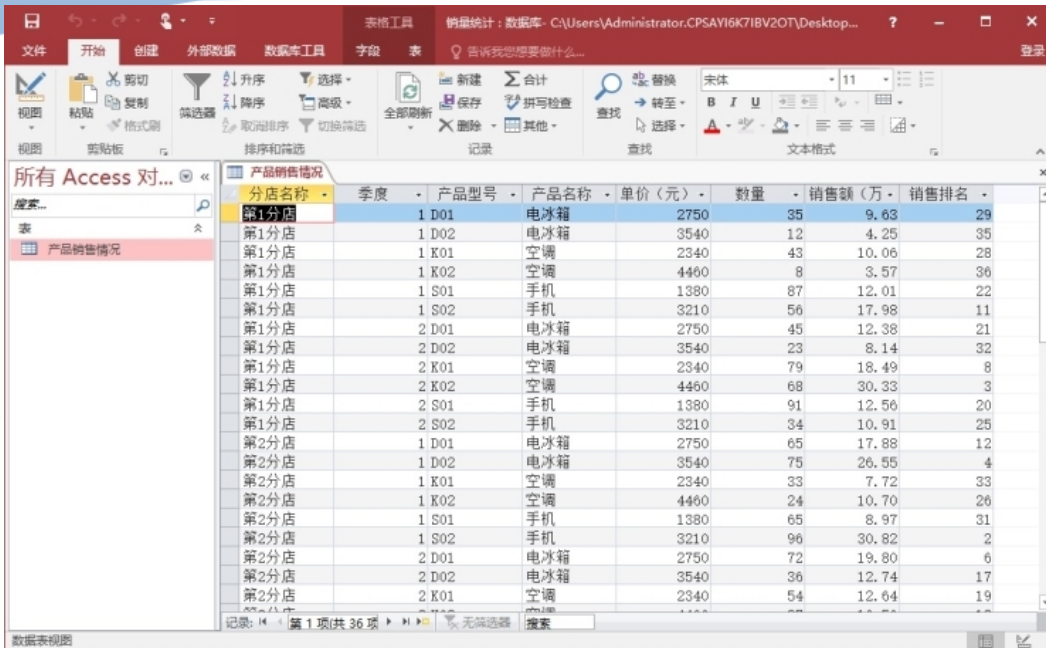
卡的“获取外部数据”功能区中的多种方法一次性导

入外部数据。

案例：现有一Microsoft Access数据库文件“销

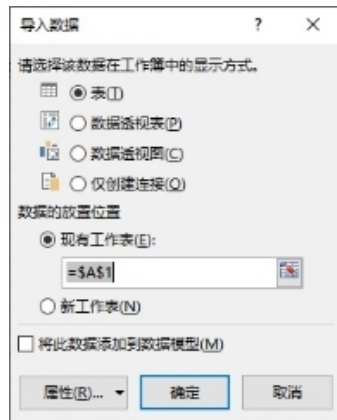
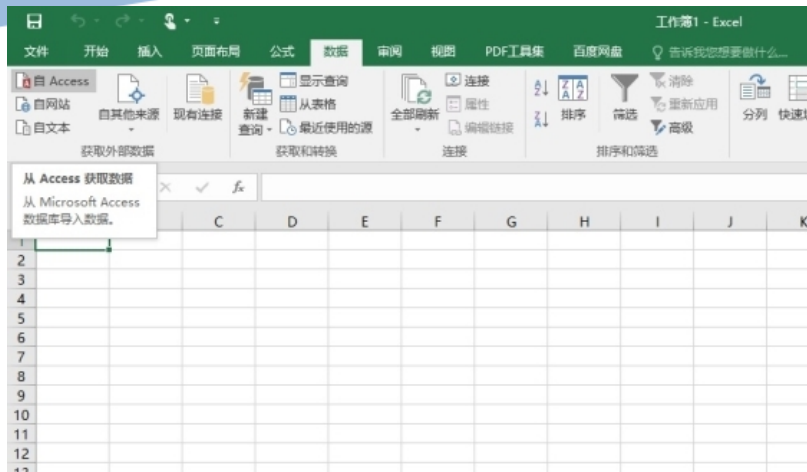
量统计.accdb”，其中表“产品销售情况”内容如图

所示，现将此表一次性导入电子表格软件中。

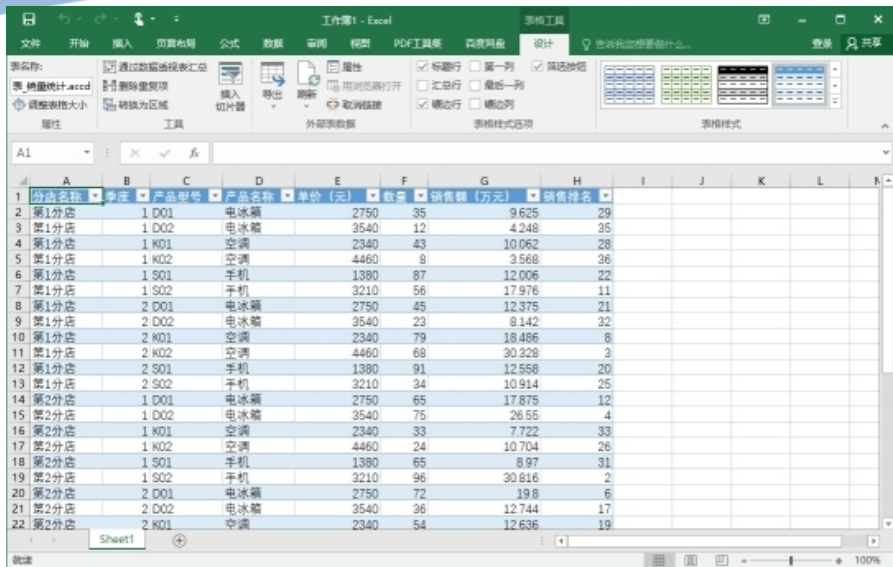


分店名称	季度	产品型号	产品名称	单价(元)	数量	销售额(万)	销售排名
第1分店	1	D01	电冰箱	2750	35	9.63	29
第1分店	1	D02	电冰箱	3540	12	4.25	35
第1分店	1	K01	空调	2340	43	10.06	28
第1分店	1	K02	空调	4460	8	3.57	36
第1分店	1	S01	手机	1380	87	12.01	22
第1分店	1	S02	手机	3210	56	17.98	11
第1分店	2	D01	电冰箱	2750	45	12.38	21
第1分店	2	D02	电冰箱	3540	23	8.14	32
第1分店	2	K01	空调	2340	79	18.49	8
第1分店	2	K02	空调	4460	68	30.33	3
第1分店	2	S01	手机	1380	91	12.56	20
第1分店	2	S02	手机	3210	34	10.91	25
第2分店	1	D01	电冰箱	2750	65	17.88	12
第2分店	1	D02	电冰箱	3540	75	26.55	4
第2分店	1	K01	空调	2340	33	7.72	33
第2分店	1	K02	空调	4460	24	10.70	26
第2分店	1	S01	手机	1380	65	8.97	31
第2分店	1	S02	手机	3210	96	30.82	2
第2分店	2	D01	电冰箱	2750	72	19.80	6
第2分店	2	D02	电冰箱	3540	36	12.74	17
第2分店	2	K01	空调	2340	54	12.64	19

- ① 新建一个空白的电子表格文件，选择“数据”→“获取外部数据”→“自Access”按钮。
- ② 在出现的“选取数据源对话框”中选择需要导入的“销量统计.accdb”Access数据库文件，单击“打开”按钮后弹出“导入数据”对话框。



③ 在“导入数据”对话框中选择数据的显示方式及数据的放置位置后，单击“确定”按钮，即可将Access数据库文件中“产品销售情况”表的内容导入Excel中



分店名称	季度	产品型号	产品名称	单价(元)	数量	销售额(万元)	销售排名
第1分店	1	D001	电冰箱	2750	35	9.625	29
第1分店	1	D002	电冰箱	3540	12	4.248	35
第1分店	1	K01	空调	2340	43	10.062	28
第1分店	1	K02	空调	4460	8	3.568	36
第1分店	1	S01	手机	1380	87	12.006	22
第1分店	1	S02	手机	3210	56	17.976	11
第1分店	2	D01	电冰箱	2750	45	12.375	21
第1分店	2	D02	电冰箱	3540	23	8.142	32
第1分店	2	K01	空调	2340	79	18.486	8
第1分店	2	K02	空调	4460	68	30.228	3
第1分店	2	S01	手机	1380	91	12.558	20
第1分店	2	S02	手机	3210	34	10.914	25
第2分店	1	D01	电冰箱	2750	65	17.875	12
第2分店	1	D02	电冰箱	3540	75	26.55	4
第2分店	1	K01	空调	2340	33	7.722	33
第2分店	1	K02	空调	4460	24	10.704	26
第2分店	1	S01	手机	1380	65	8.97	31
第2分店	1	S02	手机	3210	96	30.816	2
第2分店	2	D01	电冰箱	2750	72	19.8	6
第2分店	2	D02	电冰箱	3540	36	12.744	17
第2分店	2	K01	空调	2340	54	12.636	19

3. 利用工具软件收集数据

除自行输入数据和从外部导入数据外，还可以利用互联网工具收集数据，这其中最典型的应用就是网络调查，又称在线问卷调查。在线问卷调查是指通过互联网及其调查系统把传统的调查形式、分析方法在线化、智能化。市面上在线问卷调查工具并不少，这里以“51调查”为例为大家梳理如何做一份网络调查。

调查问卷的结构一般包括三个部分：问卷标题、问卷正文和问卷收集。

(1) 问卷标题。

首先是问候语，并向被调查对象简要说明调查的宗旨、目的和对问题回答的要求等，引起被调查者的兴趣。



(2) 问卷正文。

该部分是问卷的主体部分，主要包括：被调查者信息、调查项目、调查者信息等部分，是调查问卷的核心内容，是组织单位将所要调查了解的内容，主要以各类问卷题目（选择题、填空题、打分排序等）的形式呈现，具体化为一些问题和备选答案，如图4-23所示。

(3) 问卷收集。

通过“发布及分享”功能，可以利用网页、邮件、微信等多种回收渠道，结合丰富的推荐模式，延伸答卷数据范围，在短时间内生成和收集到大量高质量的答卷数据。

说一说

你所知道的其他数据采集方法还有哪些？



4.1.3 格式化数据和表格

原始的数据表格建立后，小华决定对表格进行简单的美化，WPS表格和Excel中“设置单元格格式”等格式化处理工具是美化表格和表格中数据的重要手段。

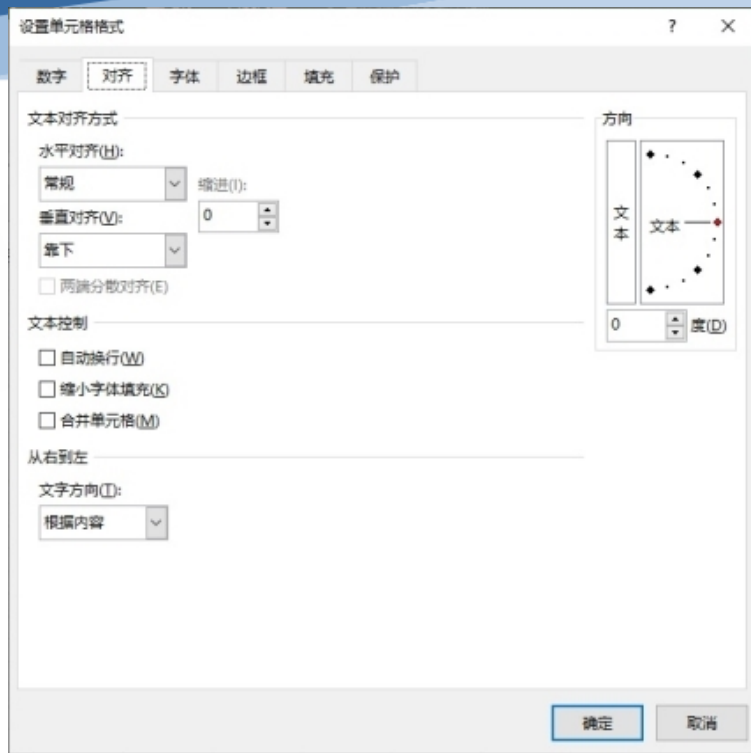
操作步骤

使用WPS表格和Excel的默认格式编辑工作表，操作相对简单，但是样式可能仍然不够美观，用户可以通过设置样式和效果美化电子表格。本操作是从单元格的格式、工作表的格式化及条件格式化等方面介绍电子表格的美化操作。

1. 单元格的格式化

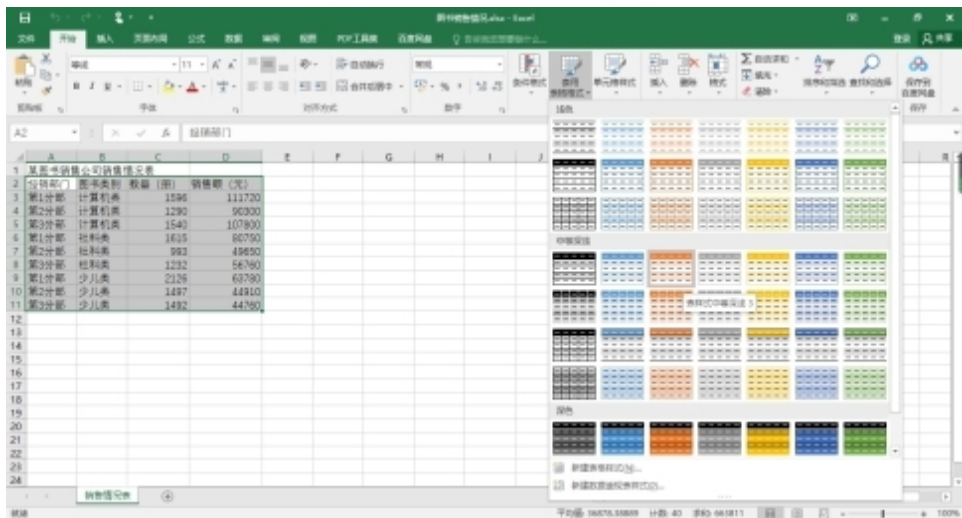
(1) 选中待设置格式的单元格或单元格区域，单击“开始”→“单元格”→“格式”按钮，在下拉菜单中选择“设置单元格格式”命令，打开“设置单元格格式”对话框，如图所示。

(2) 在“设置单元格格式”对话框中，有6个选项卡，可进行不同格式的设置操作。其中，“数字”选项卡可进行数字类型的设置；“对齐”选项卡用于设置单元格内容的对齐方式、文本控制和文字方向；“字体”选项卡用于设置字体、字号、字形、颜色及特殊效果等；“边框”选项卡用于设置单元格是否加边框及边框的线条样式、颜色等；“填充”（WPS表格中为“图案”）选项卡用于设置单元格的底纹及图案，也可以设置填充效果；“保护”选项卡用于工作表的锁定和隐藏，以保护工作表不受破坏。



2. 工作表的格式化

对工作表进行格式化的方式很多，可以根据工作的实际需求自由选择软件中各种格式设置功能。其中，通过选择软件中预定义的表格格式，如“套用表格格式”，可以简单、快速设置一组单元格或者整张工作表的格式。具体操作如下：选择需要设置格式的单元格区域，单击“开始”选项卡的“样式”组中的“套用表格格式”按钮，在“套用表格格式”下拉列表中，单击想要的“套用格式”图标即可。



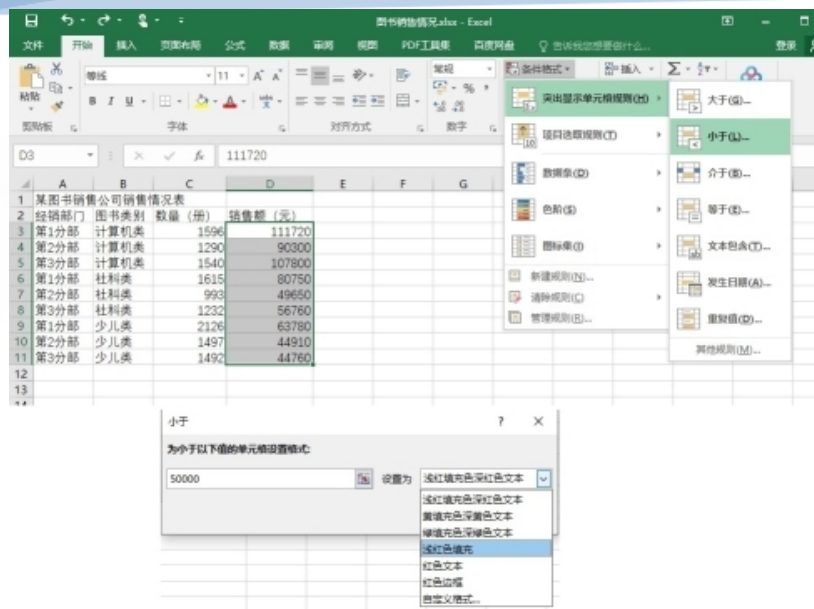
3. 设置条件格式

WPS表格和Excel提供了条件格式功能，条件格式功能可以对单元格应用某种条件来决定数据的显示格式，使用数据条、色阶和图标集，以突出显示单元格，强调异常值，以及实现数据的可视化效果。例如打开“图书销售情况.xlsx”工作簿，选择“销售情况表”标签，对销售额小于50000元的单元格设置为浅红色填充，具体步骤如下。

(1) 选择D3:D11单元格区域，选择“开始”选项卡，在“样式”组中单击“条件格式”按钮，在下拉列表中然后选择“突出显示单元格规则”中的“小于”命令，打开“小于”对话框。

(2) 在数值文本框中输入“50000”，在“设置为”下拉列表中选择“浅红色填充”选项，单击“确定”按钮。

选定单元格或单元格区域，单击“开始”选项卡的“样式”组中的“条件格式”按钮，在弹出的下拉菜单中选择相应的命令选项，再根据对话框内容进行相关设置即可。



4. 相关操作

(1) 单元格区域的选定。

选择单个单元格：单击相应的单元格。

选择连续的单元格区域：将鼠标指针定位在需要选中区域的左上角单元格，然后拖动鼠标至左下角单元格。

选择不相连的单元格区域：按住【Ctrl】键的同时依次选中单元格或单元格区域。

(2) 行和列的选定。

选择单行（列）：单击行号（或列标）。

选择连续行（列）：在需要选中的行号（或列标）拖动鼠标。

选择不连续行（列）：先选中一行（列），然后按住【Ctrl】键的同时单击行号（或列标）。

(3) 工作表的选定。

选择单张工作表：单击工作表标签。

选择连续的多张工作表：单击第一个工作表标签，然后按住【Shift】键的同时单击最后一个工作表标签。

选择不连续的多张工作表：单击第一个工作表标签，然后按住【Ctrl】键的同时单击其他工作表标签。

说一说

从职业素养的角度谈谈美化

数据表格有哪些好处？







第4章

数据处理

任务2 加工数据

主编 | 傅连仲 等

目 录

Contents

- 4.2.1 使用运算表达式
- 4.2.2 使用函数
- 4.2.3 整理数据

加工数据

在完成原始数据的录入后，通过运用工具对数据进行适当的加工和计算，可以从原始数据中得到更加准确和有用的信息。本节通过对WPS表格和Excel中相关数据处理工具的介绍，帮助读者掌握基本的数据处理技能。



任务情景

根据老师安排的实训要求，小华需要对创建的图书销售情况表进行数据处理，统计不同经销部门及不同图书类别的相关销售情况，并对图书销售情况表进行排序、筛选和分类汇总。

任务分析

要利用运算表达式和函数对数据进行计算和分析，首先必须掌握WPS表格和Excel运算表达式的创建方法，特别是常用函数的参数如何使用，因此，本任务可以分解为如下三个子任务。

- (1) 创建运算表达式，对图书销售情况表进行简单的计算处理。
- (2) 使用函数对图书销售情况表进行统计运算，获得不同经销部门及不同图书类别的相关销售情况统计。
- (3) 根据需求，对图书销售情况表进行排序、筛选和分类汇总，分析表格数据。

4.2.1 使用运算表达式

WPS表格和Excel都具有强大的数据运算功能，使用运算表达式与函数，可以灵活地对数据进行整理、计算、汇总、查询、分析等，自动得出所期望的结果，帮助用户建立数据处理和分析模型，化解工作中的许多棘手问题。小华利用自己创建的图书销售情况表，根据老师提出的案例要求，通过实际操作来体验WPS表格和Excel的数据计算功能。

操作步骤

1. 计算单册平均价格

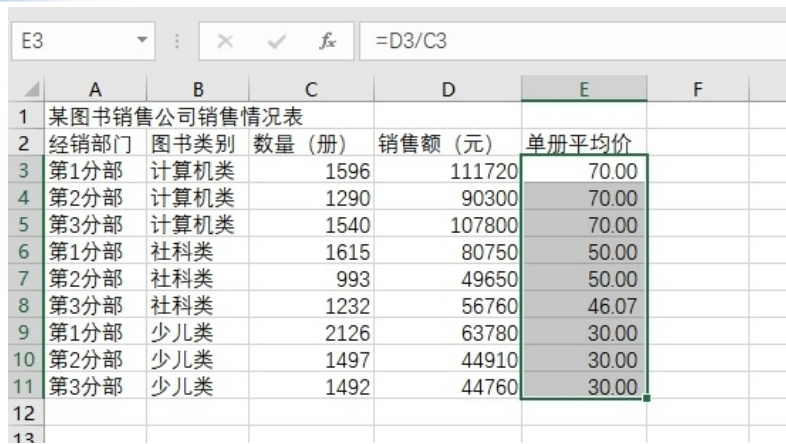
打开“图书销售情况”工作簿，在工作表“销售情况表”的E2单元格输入“单册平均价”，利用运算表达式计算“图书销售情况表”中各分部售出的各类图书的单册平均价格，并保存在E列相应位置，具体步骤如下。

(1) 单击选择E2单元格，输入“单册平均价”。

(2) 单击选择E3单元格，输入运算表达式“=D3/C3”，然后按【Enter】键。

(3) 单击选择E3单元格，然后打开“开始”选项卡的“数字”功能区，将E3单元格格式设置为保留小数点后2位。

(4) 单击选择E3单元格，拖动“填充句柄”至E11单元格。



The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F
1	某图书销售公司销售情况表					
2	经销部门	图书类别	数量(册)	销售额(元)	单册平均价	
3	第1分部	计算机类	1596	111720	70.00	
4	第2分部	计算机类	1290	90300	70.00	
5	第3分部	计算机类	1540	107800	70.00	
6	第1分部	社科类	1615	80750	50.00	
7	第2分部	社科类	993	49650	50.00	
8	第3分部	社科类	1232	56760	46.07	
9	第1分部	少儿类	2126	63780	30.00	
10	第2分部	少儿类	1497	44910	30.00	
11	第3分部	少儿类	1492	44760	30.00	
12						
13						

The formula bar at the top shows the formula `=D3/C3` for cell E3.

类别	运算符	含义	应用示例
算术运算符	+(加号)	加	1+2
	-(减号)	减	2-1
	-(负号)	负数	-1
	*(星号)	乘	2*3
	/ (斜杠)	除	4/2
	^(乘方)	乘幂	3^2
比较运算符	= (等于号)	等于	A1=A2
	> (大于号)	大于	A1>A2
	< (小于号)	小于	A1<A2
	>= (大于等于号)	大于等于	A1>=A2
	<= (小于等于号)	小于等于	A1<=A2
	<> (不等号)	不等于	A1<>A2
文本运算符	& (连字符)	将两个文本连接起来产生连续的文本	"2013" & "年"
引用运算符	: (冒号)	区域运算符, 两个引用单元格之间的区域引用	A1:D4
	, (逗号)	联合运算符, 将多个引用合并为一个引用	SUM(A1:D1, A2:C2)
	(空格)	交集运算符, 两个引用中所共有的单元格的引用	A1:D1 A1:B4

2. 计算售出的各类图书金额占总销售金额的百分比

打开“图书销售情况”工作簿，在工作表“销售情况表”的F2单元格内输入“销售额占比”，利用运算表达式计算“销售情况表”中各分部售出的各类图书金额占总销售金额的百分比（保留小数点后1位），并保存在F列相应位置，具体步骤如下。

- (1) 单击选择F2单元格，输入“销售额占比”。
- (2) 在单元格D12输入运算表达式“=SUM(D3:D11)”（求销售额总和），然后按【Enter】键。
- (3) 单击选择F3单元格，输入运算表达式“=D3/\$D\$12”，然后按【Enter】键。
- (4) 单击选择F3单元格，然后打开“开始”选项卡的“数字”功能区，将F3单元格格式设置为百分比格式，小数点后保留1位。
- (5) 单击选择F3单元格，拖动填充句柄至F11单元格。



	A	B	C	D	E	F	G	
1	某图书销售公司销售情况表							
2	经销部门	图书类别	数量(册)	销售额(元)	单册平均价	销售额占比		
3	第1分部	计算机类	1596	111720	70.00	17.2%		
4	第2分部	计算机类	1290	90300	70.00	13.9%		
5	第3分部	计算机类	1540	107800	70.00	16.6%		
6	第1分部	社科类	1615	80750	50.00	12.4%		
7	第2分部	社科类	993	49650	50.00	7.6%		
8	第3分部	社科类	1232	56760	46.07	8.7%		
9	第1分部	少儿类	2126	63780	30.00	9.8%		
10	第2分部	少儿类	1497	44910	30.00	6.9%		
11	第3分部	少儿类	1492	44760	30.00	6.9%		
12				650430				
13								

说一说

电子表格中的运算符与我们平时书写的运算符有哪些区别？



4.2.2 使用函数

在WPS表格和Excel中，函数是一种预置的运算表达式，它在得到输入值后就会执行运算，完成指定的操作任务，然后返回结果值，其目的是简化和缩短工作表中的运算表达式，特别适用于执行复杂的运算表达式。在掌握运算表达式的基础上，小华开始学习使用函数来进行更加复杂、更加实用的数据计算方法。

函数作为特殊的运算表达式，由三部分组成，分别为函数名、参数和返回值，表现形式为“函数名(参数)”。

函数通过运算后，得到一个或几个运算的结果，返回给用户或运算表达式。如果提供的参数不合理，函数运算后得到一个错误的结果，这时函数将返回一个错误值。例如，返回错误值为“#VALUE! ”，则表示此时使用的参数或运算操作符与数据项不匹配。

序号	函数名及格式	功能	参数说明
1	MAX(number1,[number2],...)	计算一组数值中的最大值	参数number1、number2.....代表需要求最大值的数值或引用的单元格（或区域）地址
2	MIN(number1,[number2],...)	计算一组数值中的最小值	参数number1、number2.....代表需要求最小值的数值或引用的单元格（或区域）地址
3	SUM(number1, [number2],...)	计算所有参数所代表数值的和	参数number1、number2.....代表要求和的值，可以是具体的数值、引用的单元格（或区域）、逻辑值等
4	AVERAGE(number1,[number2],...)	计算所有参数所代表数值的算术平均值	参数number1、number2.....代表需要求平均值的数值或引用的单元格（或区域）地址
5	MODE(number1,[number2],...)	返回数据集中出现最多的数值	参数number1、number2.....代表数据集中的数值或引用单元格（或区域）地址
6	RANK(number,ref,[order])	返回某一数值在一列数值中相对于其他数值的排位	参数number代表需要排序的数值；参数ref代表排序数值所处的单元格区域；参数order代表排序方式参数〔如果为“0”或者忽略，则按降序排序，即数值越大，排名结果数值越小；如果为非“0”值（一般为1），则按升序排序，即数值越大，排名结果数值越大〕
7	COUNT(value1, [value 2],...)	计算参数列表中数值的个数	参数value1、value2.....代表包含数值的单元格地址及参数列表中的数值。另外，还有COUNTA函数，用于求“非空”单元格个数；COUNTBLANK函数，用于求“空”单元格个数
8	IF(logical_test,[value_if_true],[value_if_false])	根据逻辑判断的真假结果，返回相对应的内容	参数logical_test代表逻辑判断表达式；参数value_if_true表示当判断条件为逻辑“真”（True）时的显示内容，如果忽略则返回“True”；参数value_if_false表示当判断条件为逻辑“假”（False）时的显示内容，如果忽略则返回“False”
9	COUNTIF(range,criteria)	对指定区域中符合指定条件的单元格进行计数	参数range代表要统计的单元格区域；参数criteria表示产生计数的条件表达式
10	SUMIF(range,criteria,[sum_range])	计算符合指定条件的单元格区域内的数值之和	参数range代表用于条件判断的单元格区域；参数criteria为指定的条件表达式；参数sum_range代表要求和的数值所在的单元格区域
11	ROUND(number, num_digits)	将数值四舍五入到指定小数位数	参数number代表需要四舍五入的数值型数字或引用的单元格地址，num_digits代表四舍五入后需要保留的小数位数

案例一：打开“近三年月平均气温统计表.xlsx”文件，其中sheet1工作表的内容，计算近三年各月平均气温的最高值和最低值置于“最高值”行和“最低值”行的相应单元格内，具体步骤如下。

- (1) 单击选择B7单元格，输入“=MAX(B3:B5)”，然后按【Enter】键。
- (2) 单击选择B7单元格，拖动“填充句柄”至M7单元格，即可求出近三年各月气温的最高值。
- (3) 单击选择B8单元格，输入“=MIN(B3:B5)”，然后按【Enter】键。
- (4) 单击选择B8单元格，拖动“填充句柄”至M8单元格，即可求出近三年各月气温的最低值。

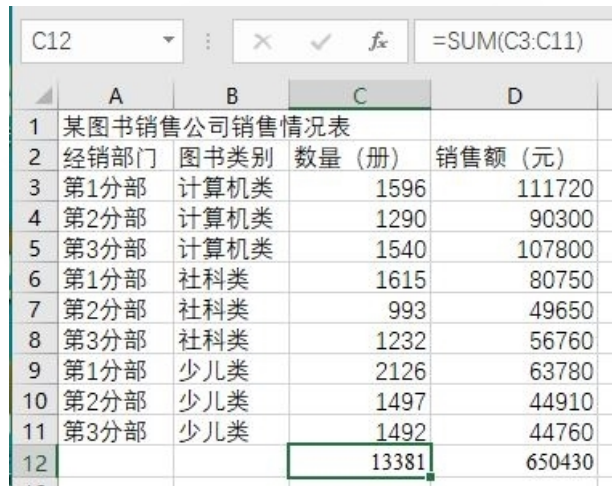
	A	B	C	D	E	F	G	H	I	J	K	L	M
1	某地区近	三年月平均气温统计表(单位:度)											
2	月份	一月	二月	三月	四月	五月	六月	七月	八月	九月	十月	十一月	十二月
3	2018年	3.1	5.1	10.2	15.9	24.3	27.1	29.3	33.2	26.5	21.8	17.2	9.8
4	2019年	2.5	5.6	10.9	15.1	24.7	29.6	32.1	34.1	27.3	22.1	15.8	11.4
5	2020年	2.2	5.2	10.3	15.3	25.1	28.5	31.9	32.9	24.8	22.4	16.9	10.7
6													
7	最高值												
8	最低值												

案例二：打开“图书销售情况”工作簿，利用函数计算“销售情况表”工作表中各分部售出的所有图书的总册数及总金额，并保存在C12和D12单元格内，具体步骤如下。

(1) 单击选择C12单元格，输入“=SUM(C3:C11)”，然后按【Enter】键。

(2) 单击选择C12单元格，拖动填充句柄至D12单元格。

同样的方法可以计算“销售情况表”工作表中各分部售出的所有图书的平均册数及平均销售额，只需将函数换成求平均值的AVERAGE()函数即可。



The screenshot shows an Excel spreadsheet with a formula bar at the top displaying "=SUM(C3:C11)". The spreadsheet contains a table with the following data:

	A	B	C	D
1	某图书销售公司销售情况表			
2	经销部门	图书类别	数量(册)	销售额(元)
3	第1分部	计算机类	1596	111720
4	第2分部	计算机类	1290	90300
5	第3分部	计算机类	1540	107800
6	第1分部	社科类	1615	80750
7	第2分部	社科类	993	49650
8	第3分部	社科类	1232	56760
9	第1分部	少儿类	2126	63780
10	第2分部	少儿类	1497	44910
11	第3分部	少儿类	1492	44760
12			13381	650430

案例三：打开“员工调薪统计表.xlsx”文件，计算sheet1工作表“调薪后工资（元）”列的内容（调薪后工资=现工资+现工资×调薪系数），利用MODE()函数计算现工资和调薪后工资的普遍工资金额，分别置于B18和D18单元格内，具体步骤如下。

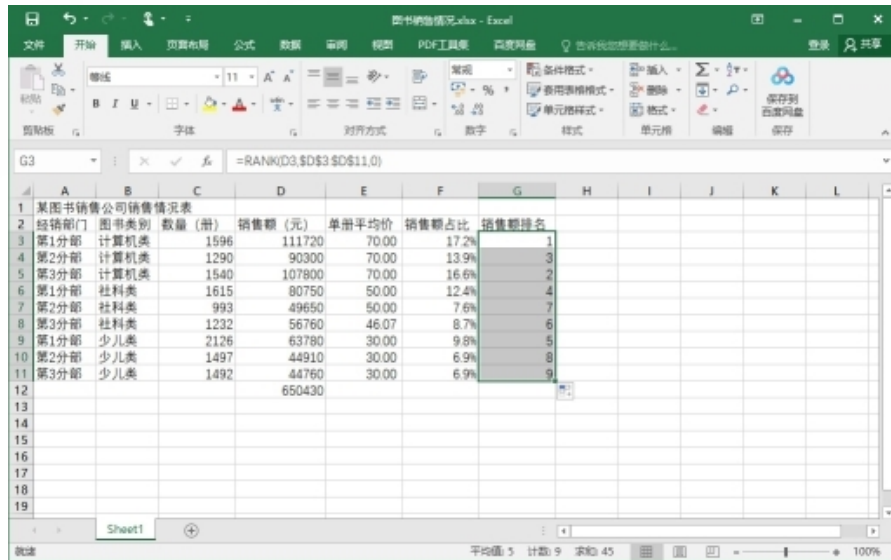
- (1) 单击选择D3单元格，输入运算表达式“=B3+B3*C3”，然后按【Enter】键。
- (2) 单击选择D3单元格，向下拖动填充句柄至D17单元格，即可求出所有员工调薪后的工资。
- (3) 单击选择B18单元格，输入“=MODE (B3:B17)”，然后按【Enter】键，即可求出现工资的普遍金额。
- (4) 同样，单击选择D18单元格，输入“=MODE (D3:D17)”，然后按【Enter】键，即可求出调薪后工资的普遍金额。

	A	B	C	D
1	某公司员工调薪统计表			
2	员工号	现工资(元)	调薪系数	调薪后工资(元)
3	AS1	2400	35%	
4	AS2	2200	20%	
5	AS3	2000	35%	
6	AS4	2500	40%	
7	AS5	2700	25%	
8	AS6	2200	25%	
9	AS7	2800	35%	
10	AS8	3800	20%	
11	AS9	2500	20%	
12	AS10	2200	30%	
13	AS11	2500	35%	
14	AS12	2500	40%	
15	AS13	2800	30%	
16	AS14	2200	35%	
17	AS15	2500	40%	
18	普遍工资			

D18				
	A	B	C	D
1	某公司员工调薪统计表			
2	员工号	现工资(元)	调薪系数	调薪后工资(元)
3	AS1	2400	35%	3240
4	AS2	2200	20%	2640
5	AS3	2000	35%	2700
6	AS4	2500	40%	3500
7	AS5	2700	25%	3375
8	AS6	2200	25%	2750
9	AS7	2800	35%	3780
10	AS8	3800	20%	4560
11	AS9	2500	20%	3000
12	AS10	2200	30%	2860
13	AS11	2500	35%	3375
14	AS12	2500	40%	3500
15	AS13	2800	30%	3640
16	AS14	2200	35%	2970
17	AS15	2500	40%	3500
18	普遍工资	2500		3500

案例四：打开“图书销售情况”工作簿，在工作表“销售情况表”的G2单元格内输入“销售额排名”，利用函数计算“销售情况表”工作表中各分部售出的各类图书的销售额按降序排列，并保存在G列相应位置，具体步骤如下。

- (1) 单击选择G2单元格，输入“销售额排名”。
- (2) 单击选择G3单元格，输入运算表达式“=RANK(D3,\$D\$3:\$D\$11,0)”，然后按【Enter】键。
- (3) 单击选择G3单元格，拖动“填充句柄”至G11单元格。



The screenshot shows an Excel spreadsheet titled "图书销售情况.xlsx - Excel". The active cell is G3, containing the formula `=RANK(D3,D3:D11,0)`. The spreadsheet data is as follows:

1	A	B	C	D	E	F	G	H	I	J	K	L
2	某图书销售公司销售情况表											
3	经销部门	图书类别	数量(册)	销售额(元)	单册平均价	销售额占比	销售额排名					
4	第1分部	计算机类	1596	111720	70.00	17.2%	1					
5	第2分部	计算机类	1290	90300	70.00	13.9%	3					
6	第3分部	计算机类	1540	107800	70.00	16.6%	2					
7	第1分部	社科类	1615	80750	50.00	12.4%	4					
8	第2分部	社科类	993	49650	50.00	7.6%	7					
9	第3分部	社科类	1232	56760	46.07	8.7%	6					
10	第1分部	少儿类	2126	63780	30.00	9.8%	5					
11	第2分部	少儿类	1497	44910	30.00	6.9%	8					
12	第3分部	少儿类	1492	44760	30.00	6.9%	9					
13				650430								
14												
15												
16												
17												
18												
19												

案例五：打开“产品2018年销量统计表.xlsx”文件，首先计算Sheet1工作表中“全年总量”行的数据，置于B15单元格内；再计算“所占百分比”列的数据（所占百分比 = 月销售量/全年销量，百分比型，小数点后保留两位）；如果“所占百分比”列内容大于或等于8%，在“备注”列内显示文字“良好”，否则在“备注”列内显示文字“一般”，具体步骤如下。

- (1) 单击选择B15单元格，输入表达式“=SUM(B3:B14)”，然后按【Enter】键。
- (2) 单击选择C3单元格，输入表达式“=B3/\$B\$15”，然后按【Enter】键，再次选择C3单元格，设置单元格格式为百分比型、小数点后保留两位，即可求出1月销量占全年销量的百分比。
- (3) 单击选择C3单元格，向下拖动填充句柄至C14单元格，即可求出各月销量占全年销量的百分比。
- (4) 单击选择D3单元格，输入“=IF(C3>=8%,"良好","一般)”，然后再次选择D3单元格，向下拖动填充句柄至D14单元格。

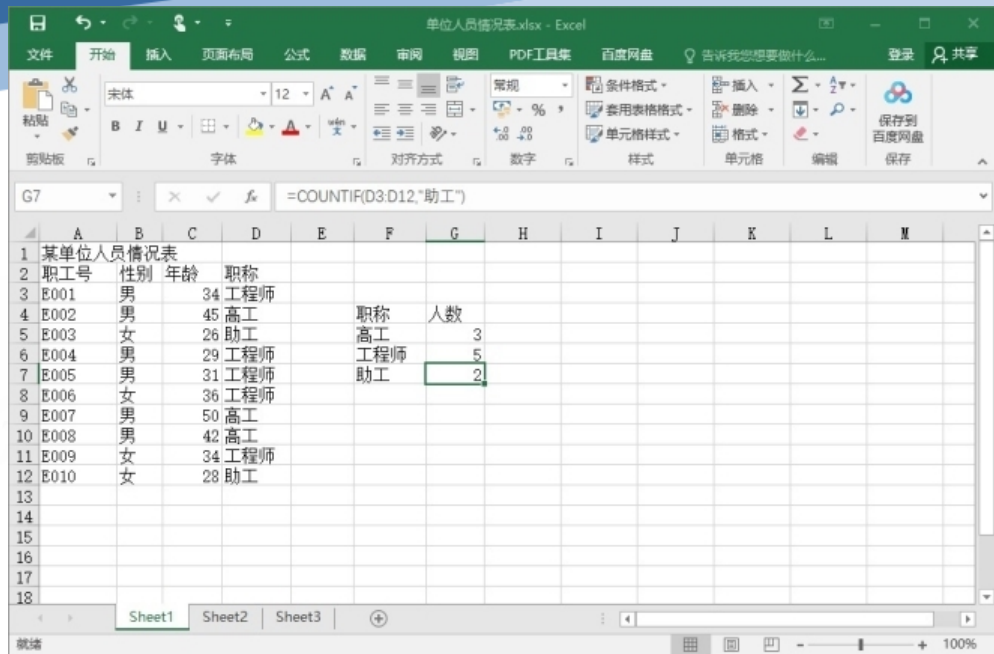
The screenshot shows the Microsoft Excel interface with a spreadsheet titled "产品18年销量统计表.xlsx". The formula bar displays the formula $=IF(C3>=8\%, "良好", "一般")$. The spreadsheet contains the following data:

1	某产品2018年销量统计表(单位:个)			
2	月份	18年	所占百分比	备注
3	1月	332	6.72%	一般
4	2月	156	3.16%	一般
5	3月	180	3.65%	一般
6	4月	421	8.53%	良好
7	5月	679	13.75%	良好
8	6月	934	18.91%	良好
9	7月	631	12.78%	良好
10	8月	388	7.86%	一般
11	9月	464	9.40%	良好
12	10月	290	5.87%	一般
13	11月	288	5.83%	一般
14	12月	175	3.54%	一般
15	全年销量	4938		
16				

案例六：打开“单位人员情况表.xlsx”文件，利用COUNTIF()函数计算Sheet1工作表中职称分别为高工、工程师和助工的人数，分别置于G5到G7单元格区域，具体步骤如下。

(1) 单击选择G5单元格，输入表达式“=COUNTIF(D3:D12,"高工")”，然后按【Enter】键。

(2) 然后分别选择G6、G7单元格，输入表达式“=COUNTIF(D3:D12,"工程师")”和“=COUNTIF(D3:D12,"助工")”后按【Enter】键，即可求出各职称人数。



案例七：打开“学生成绩表.xlsx”文件，首先计算Sheet1工作表中学生的“平均成绩”列的数据（设置为数值型，小数点后保留2位）；然后利用COUNTIF函数计算第一组学生的人数，置于G3单元格内；最后，利用SUMIF函数计算第一组学生的平均成绩，置于G5单元格内，具体步骤如下。

- (1) 单击选择F3单元格，输入表达式“=AVERAGE(C3:E3)”后按【Enter】键，并设置数字格式为数值型，小数点后保留2位。
- (2) 单击选择F3单元格，向下拖动“填充句柄”至F12单元格，即可求出所有学生的三门课平均成绩。
- (3) 单击选择G3单元格，输入表达式“=COUNTIF(B3:B12,"第一组")”后按【Enter】键，可求出第一组人数。
- (4) 单击选择G5单元格，输入“=SUMIF(B3:B12,"第一组",F3:F12)/G3”，即可求出第一组的平均成绩，如图所示。

The screenshot shows an Excel spreadsheet titled "学生成绩表.xlsx". The formula bar for cell G5 displays the formula `=SUMIF(B3:B12,"第一组",F3:F12)/G3`. The spreadsheet contains the following data:

学号	组别	数学	语文	英语	平均成绩	第一组人数
A1	第一组	112	98	106	105.33	6
A2	第一组	98	103	109	103.33	第一组平均成绩
A3	第一组	117	99	99	105.00	103.39
A4	第二组	115	112	108	111.67	
A5	第一组	104	96	90	96.67	
A6	第二组	101	110	105	105.33	
A7	第一组	93	109	107	103.00	
A8	第二组	95	102	106	101.00	
A9	第一组	114	103	104	107.00	
A10	第二组	89	106	116	103.67	

在上述案例中，数值格式的设置可以使用数据处理软件中的“单元格格式”进行设置，也可以使用ROUND()函数进行设置，函数可以嵌套使用，函数的嵌套是指一个函数可以作为另一个函数的参数使用。例如在案例七中，F3单元格中的运算表达式可输入“=ROUND(AVERAGE(C3:E3),2)”，便直接可以将计算结果保留小数点后2位。其中，ROUND()作为一级函数，AVERAGE()作为二级函数，先执行AVERAGE()函数，再执行ROUND()函数。

说一说

你还知道哪些函数？



4.2.3 整理数据

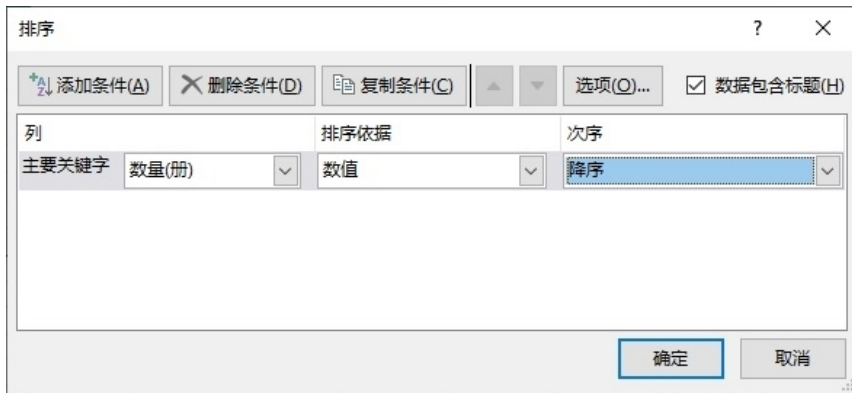
WPS表格和Excel不仅具有计算、处理数据的功能，还具有强大的数据管理分析功能。小华利用创建的“图书销售情况表”，按照老师的要求，通过WPS表格和Excel的排序、筛选和分类汇总等数据管理功能，提高表格数据的管理能力。

操作步骤

1. 数据排序

打开“图书销售情况”工作簿，对工作表“销售情况表”内数据以“数量（册）”为关键字，按降序对表格进行排序，具体步骤如下。

- (1) 单击数据清单中的任一单元格。
- (2) 单击“数据”选项卡的“排序和筛选”功能区中的“排序”按钮，打开“排序”对话框。
- (3) 在Excel的“排序”对话框中，“主要关键字”选择“数量（册）”，“次序”选择“降序”。在WPS表格的“排序”对话框中，“主要关键字”选择“数星（册）”，并选择“降序”单选钮。
- (4) 单击“确定”按钮，表格数据按设置进行排序。



2. 自动筛选

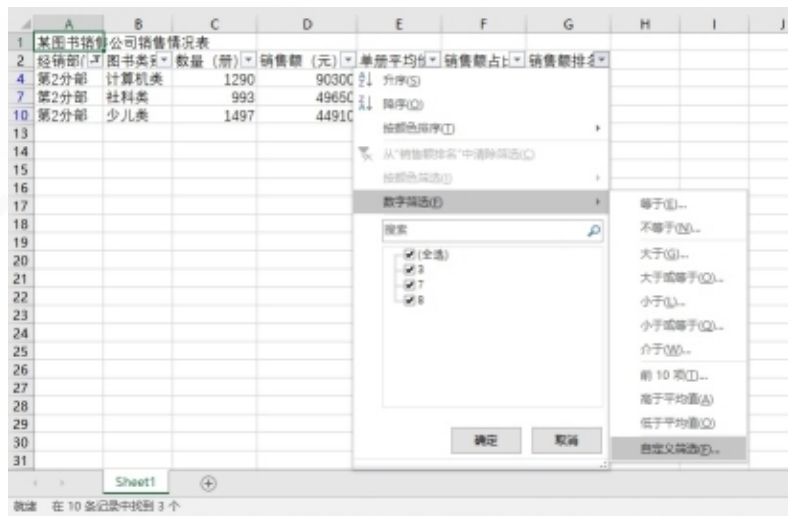
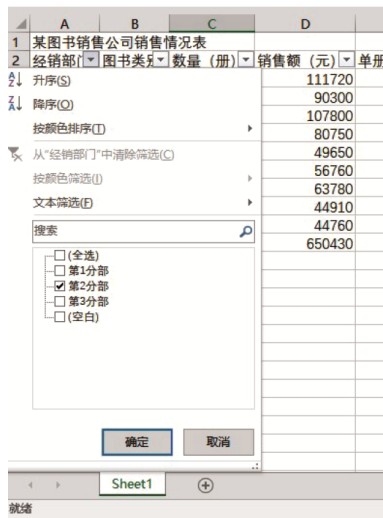
打开“图书销售情况”工作簿，对工作表“销售情况表”内的数据清单的内容进行自动筛选，筛选条件为“第2分部”且销售额排名在前五名，具体步骤如下。

- (1) 单击数据清单中的任一单元格。
- (2) 单击“数据”选项卡的“排序和筛选”功能区中的“筛选”按钮。

(3) 单击A2单元格右侧下拉按钮，仅选择“第2分部”。

(4) 单击G2单元格右侧下拉按钮，选择“数字筛选”展开菜单中的“自定义筛选”命令。弹出的在“自定义筛选方式”对话框的“销售额排名”第一个下拉列表框中选择“小于或等于”，右侧输入框中输入“5”，然后单击“确定”按钮。

若要取消筛选状态，再次单击“数据”选项卡的“排序和筛选”功能区中的“筛选”按钮，数据清单便恢复到筛选前的状态。



3. 分类汇总

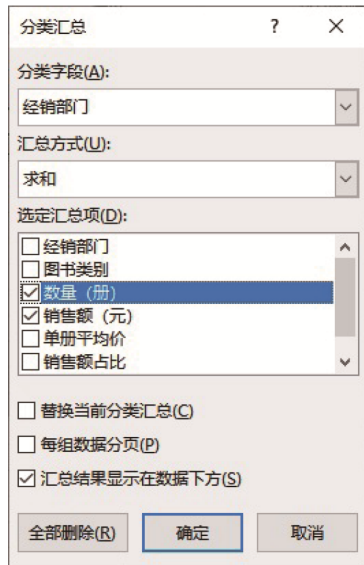
打开“图书销售情况”工作簿，对工作表“销售情况表”内的数据清单的内容进行分类汇总，汇总计算各分部图书销售的总册数及总金额，具体步骤如下。

(1) 以“经销部门”为主要关键字对数据清单进行排序。

(2) 选择A2:H11单元格区域，单击“数据”选项卡的“分级显示”功能区中的“分类汇总”按钮，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“经销部门”，“汇总方式”选择“求和”，“选定汇总项”选择“数量（册）”和“销售额（元）”，选择“汇总结果显示在数据下方”复选框，单击“确定”按钮。

若要取消分类汇总状态，单击“数据”选项卡的“分级显示”功能区中的“分类汇总”按钮，弹出“分类汇总”对话框，单击对话框中“全部删除”按钮，数据清单便恢复到分类汇总前的状态。



4. 高级筛选

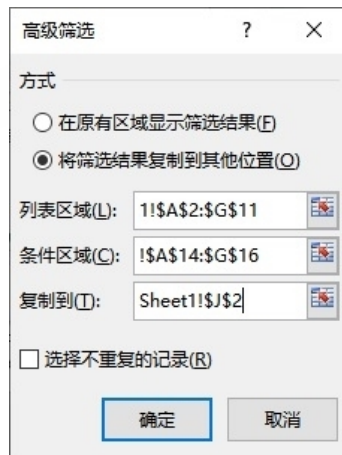
打开“图书销售情况”工作簿，对工作表“销售情况表”内的数据清单的内容进行高级筛选操作，筛选出全部社科类图书和销售额排名在前五位的图书，并将筛选结果复制到A20开始的单元格区域中。在Excel中的操作过程如下。

(1) 在工作表“销售情况表”中的A14:E16区域进行筛选条件的建立和编辑。

(2) 选择A2:E11区域，单击“数据”选项卡的“排序和筛选”功能区中的“高级”按钮。

(3) 在弹出的“高级筛选”对话框中，设置“方式”为“将筛选结果复制到其他位置”。利用切换按钮在数据清单中选择“列表区域”为“1!\$A\$2:\$G\$11”，“条件区域”为“!\$A\$14:\$G\$16”，“复制到”为“Sheet1!\$J\$2”。

14	经销部门	图书类别	数量(册)	销售额(元)	单册平均价	销售额占比	销售额排名
15		社科类					
16							<=5



(4) 单击“确定”按钮，完成筛选，筛选后的结果如图所示。

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	某图书销售公司销售情况表																
2	经销部门	图书类别	数量 (册)	销售额 (元)	单册平均价	销售额占比	销售额排名				经销部门	图书类别	数量 (册)	销售额 (元)	单册平均价	销售额占比	销售额排名
3	第1分部	计算机类	1596	111720	70.00	17.2%	1			第1分部	计算机类	1596	111720	70.00	17.2%	1	
4	第1分部	社科类	1615	80750	50.00	12.4%	4			第1分部	社科类	1615	80750	50.00	12.4%	4	
5	第1分部	少儿类	2126	63780	30.00	9.8%	5			第1分部	少儿类	2126	63780	30.00	9.8%	5	
6	第2分部	计算机类	1290	90300	70.00	13.9%	3			第2分部	计算机类	1290	90300	70.00	13.9%	3	
7	第2分部	社科类	993	49650	50.00	7.6%	7			第2分部	社科类	993	49650	50.00	7.6%	7	
8	第2分部	少儿类	1497	44910	30.00	6.9%	8			第3分部	计算机类	1540	107800	70.00	16.6%	2	
9	第3分部	计算机类	1540	107800	70.00	16.6%	2			第3分部	社科类	1232	56760	46.07	8.7%	6	
10	第3分部	社科类	1232	56760	46.07	8.7%	6										
11	第3分部	少儿类	1492	44760	30.00	6.9%	9										
12				650430													
13																	
14	经销部门	图书类别	数量 (册)	销售额 (元)	单册平均价	销售额占比	销售额排名										
15		社科类															
16							<=5										
17																	

说一说

细节决定成败。在进行数据整理时，你遇到了哪些问题？你是怎么解决的？你的体会是什么？







第4章

数据处理

任务3 分析数据

主编 | 傅连仲 等

目 录

Contents

4.3.1 简单分析数据

4.3.2 制作简单数据图表

分析数据

数据查询和分析是整个数据处理过程的最终目的，WPS表格和Excel两款软件都提供了一定的数据查询和分析功能。关于数据查询，可以通过软件的“查找”功能来简单完成，也可以利用LOOKUP、VLOOKUP等函数进行，还可以利用“数据”选项卡中的多种功能性工具建立相应的查询。

简单分析数据

分析数据

制作简单数据图表

任务情景

根据老师提出的进一步的数据分析要求，小华利用WPS表格和Excel中的“数据透视表”工具，对不同类别之间的数据进行比较和分析，并使用“图表”功能将表格中的数据进行了更具比较性的可视化的展现。

任务分析

根据任务需求，小华需要分别使用“数据透视表”和“图表”功能，进行数据分析。现将任务分解成以下过程：

- (1) 利用“数据透视表”工具，获得不同经销部门及不同图书类别的相关销售情况统计信息。
- (2) 根据需求和数据清单的数据特点，创建出更具比较性和直观性的图表类型。
- (3) 对所创建的图表进行编辑和美化。

4.3.1 简单分析数据

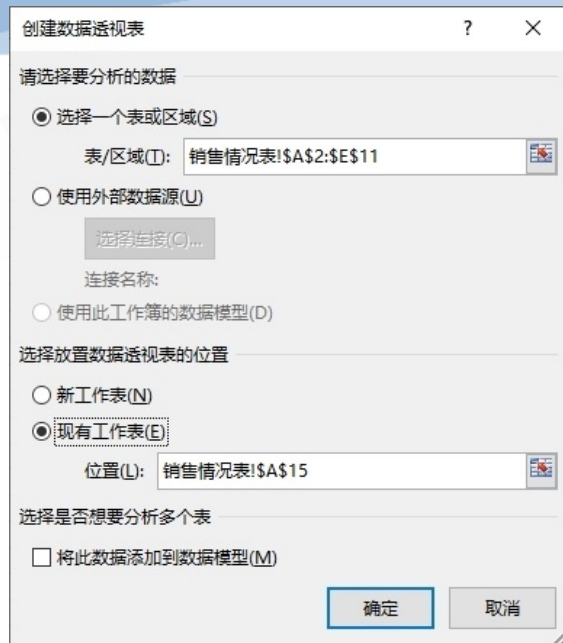
在WPS表格和Excel中，数据透视表是一种对大量数据进行快速汇总和建立交叉比较的交互式表格。使用数据透视表可以从多种不同的角度深入分析数据，可在汇总数量较大的数字列表并对各种数据进行多种比较时使用。小华在掌握数据运算的基础上，开始利用获取的数据进行分析、比较，通过对图书销售情况表创建数据透视表的过程，掌握利用“数据透视表”工具对复杂数据进行相关分析的技巧。

操作步骤

(1) 打开最初创建的“图书销售情况”工作簿，选择工作表“销售情况表”。

(2) 单击选择数据清单内任一单元格。

(3) 单击“插入”选项卡“表格”功能区中的“数据透视表”按钮，打开“创建数据透视表”对话框，如图4-47所示。



(4) 在“创建数据透视表”对话框中，首先选择用于创建数据透视表的数据清单，然后选择数据透视表的放置位置（此处选择“现有工作表”，“位置”选择“销售情况表!\$A\$15”起始的区域），然后再单击“确定”按钮，进入数据透视表创建界面。

(5) 在界面右侧“选择要添加到报表的字段”区域中，将“经销部门”字段拖到“行”标签区域；将“图书类别”字段拖到“列”标签区域；将“销售额（元）”字段拖到“值”标签区域。与此同时，在指定存放位置会出现数据透视表。

The screenshot shows the Excel interface with the following data table:

经销部门	图书类别	数量(册)	销售额(元)	
第1分部	计算机类	1596	111720	
第2分部	计算机类	1290	90300	
第3分部	计算机类	1540	107800	
第1分部	社科类	1615	80750	
第2分部	社科类	993	49650	
第3分部	社科类	1232	56760	
第1分部	少儿类	2126	63780	
第2分部	少儿类	1497	44910	
第3分部	少儿类	1492	44760	
总计		309820	153450	187160

The PivotTable task pane on the right shows the following configuration:

- 选择要添加到报表的字段: 搜索
- 已选字段:
 - 经销部门
 - 图书类别
 - 数量(册)
 - 销售额(元)
- 更多表格...
- 在以下区域间拖动字段:
 - 筛选器
 - 列: 图书类别
 - 行: 经销部门
 - 值: 求和项: 销售额...
- 推迟布局更新
- 更新

说一说

你还知道哪些简单分析数据的方法?



4.3.2 制作简单数据图表

图表是数据的一种可视化表现形式。通过使用柱形图或折线图等图表，可按照图形格式显示系列数值数据，形象、直观地表示数值大小及其变化趋势，让数据与图形联系起来。

在老师的指引下，小华在已经完成的数据清单基础上，学习创建图形化、直观性更好的数据图表的基本方法。

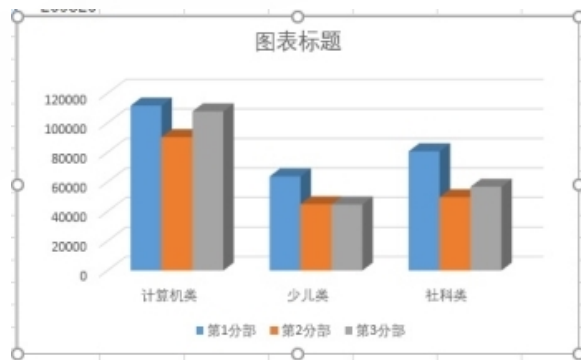
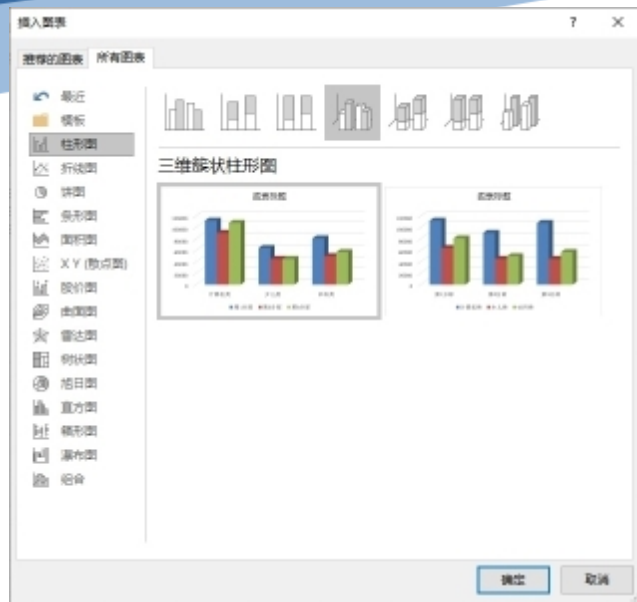
操作步骤

1. 生成三维簇状柱形图

- (1) 打开“图书销售情况”工作簿，新建一个工作表并在其中创建如图4-49所示数据内容，将此工作表命名为“销售统计表”。
- (2) 选中“销售部门”“计算机类”“少儿类”和“社科类”所在列的数据区域，确定图表所需的数据源。

	A	B	C	D	E
1	销售部门	计算机类	少儿类	社科类	总计
2	第1分部	111720	63780	80750	256250
3	第2分部	90300	44910	49650	184860
4	第3分部	107800	44760	56760	209320

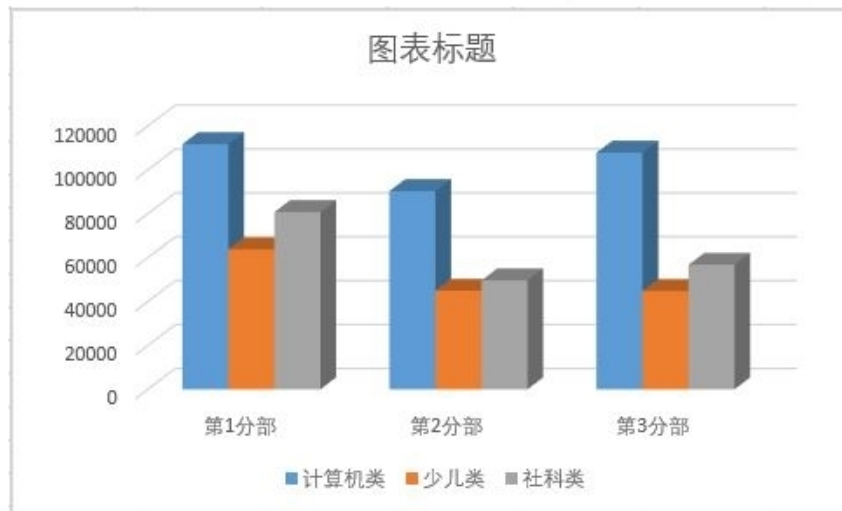
(3) 单击“插入”选项卡“图表”组中的“查看所有图表”按钮，打开如图4-50所示“插入图表”对话框，在左侧窗框中单击“柱形图”，选择“三维簇状柱形图”，单击“确定”按钮即可生成图表，如图4-51所示。



2. 图表组成元素编辑

(1) 更改图表行列。

如果需要更改图表中数据序列的顺序，可在选中图表后，在“图表工具/设计”选项卡“数据”组内单击“切换行/列”按钮即可，可以满足从不同角度分析比较数据的要求。图4-52即为对图所示的图表进行“切换行/列”操作的结果。



(2) 图表标签设置。

①设置图表标题。

选中图表后，打开“图表工具/设计”选项卡，单击“图表布局”组中的“添加图表元素”按钮，在下拉选项中选择“图表标题/图表上方”，使图表区上方出现“图表标题”字样。选中这个对象，可将其修改为“图书销售统计”。

②设置坐标轴标题。

选中图表后，打开“图表工具/设计”选项卡，单击“图表布局”组中的“添加图表元素”按钮，在下拉选项中选择“轴标题/主要横坐标轴”，即可添加相应的坐标轴标题，如图4-53所示。

③调整图例位置。

选中图表后，打开“图表工具/设计”选项卡，单击“图表布局”组中的“添加图表元素”按钮，在下拉选项中选择“图例”，即可决定图表中放置图例的位置。

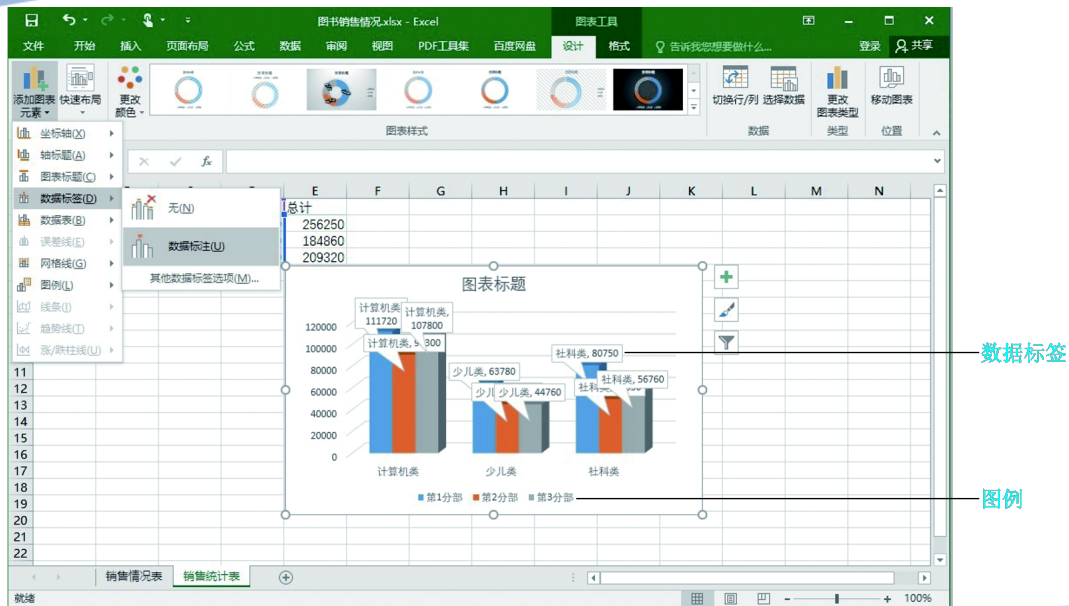
④添加数据标签。

选中图表后，打开“图表工具/设计”选项卡，单击“图表布局”组中的“添加图表元素”按钮，在下拉选项中选择“数据标签/数据标注”，此时在图表区中显示数据标签。



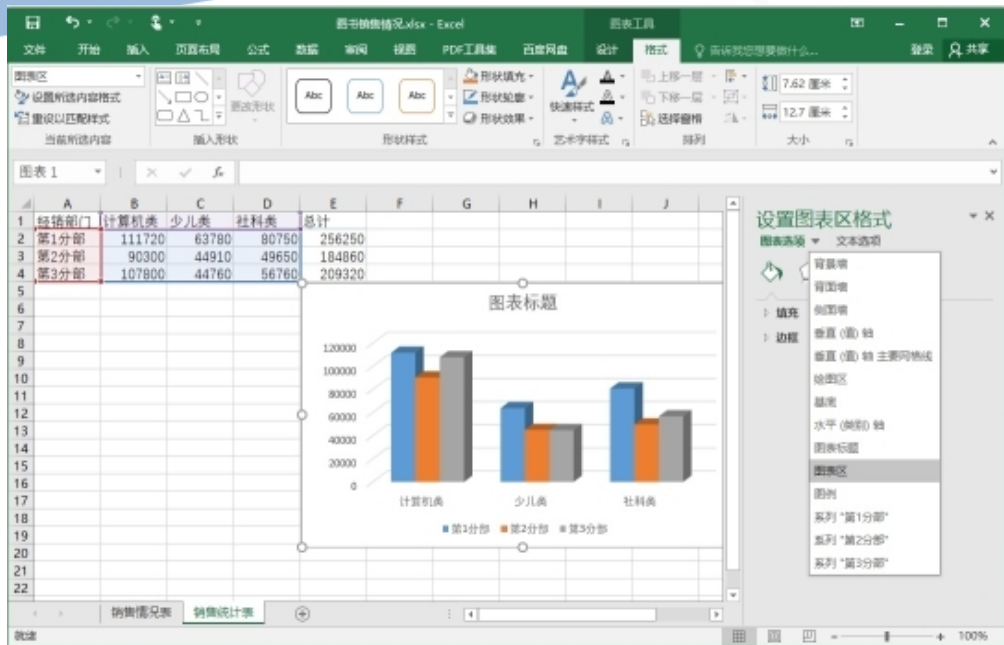
(3) 坐标轴设置。

选中图表后，打开“图表工具/设计”选项卡，单击“图表布局”组中的“添加图表元素”按钮，在下拉选项中选择“坐标轴”或“网格线”，就可以分别对坐标轴或相应的网格线格式进行设置。



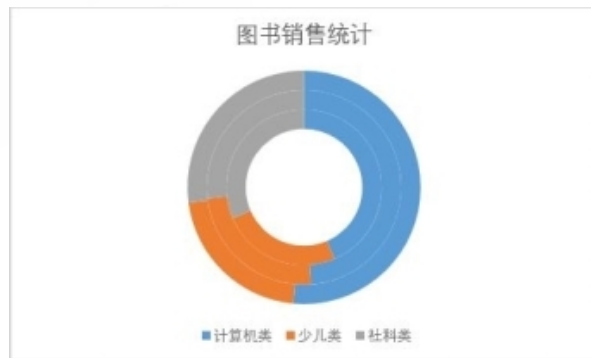
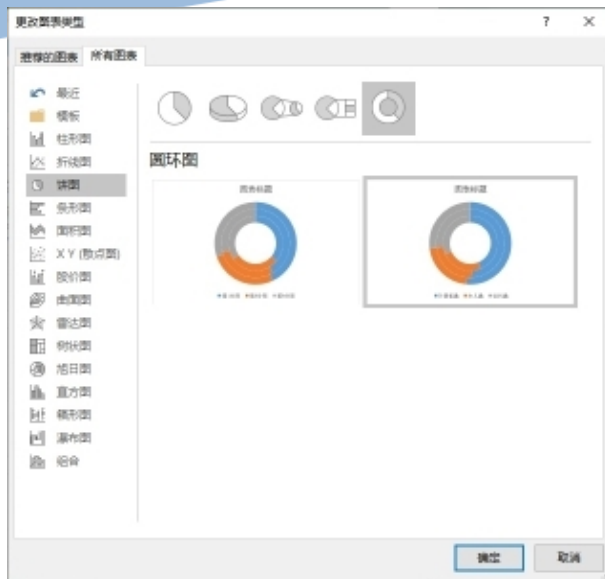
(4) 设置图表区格式。

选中图表后，打开“图表工具/格式”选项卡，单击“当前所选内容”组中的“设置所选内容格式”按钮，在工作表右侧区域显示“设置图表区格式”区域，单击“图表选项”右侧下拉按钮，即可对图表中各元素的填充、边框、文本等进行设置，如图所示。



(5) 更改图表类型。

选中图表后，打开“图表工具/设计”选项卡，单击“类型”组中的“更改图表类型”按钮，弹出“更改图表类型”对话框，可根据实际需求，选择更为直观的图表类型，此处选择圆环图示例，如图所示。



说一说

你还知道怎样合理地选用图
表类型吗?







第4章

数据处理

任务4 初识大数据

主编 | 傅连仲 等

目 录

Contents

4.4.1 大数据基础知识

4.4.2 大数据采集与分析方法

初识大数据

现代社会是一个高速发展、科技发达、信息快速流通的社会，人们之间的交流越来越密切，大数据也应运而生。随着云时代的来临，大数据（Big Data）一词越来越多地被人们提及，也吸引了越来越多的关注。

当前，全社会信息量爆炸式增长，数量巨大、来源分散、格式多样的大数据对人们提出了新的挑战，也带来了新的机遇。大数据的应用越来越彰显其优势，所涉及的领域也越来越多，无人驾驶、智慧城市、生态监测……大数据几乎无所不在，并正在改变着人们生活的方方面面。大数据正在助推企业不断发展新业务、创新运营模式。



任务情景

小华学习了基本的数据处理知识和方法后，对数据处理产生了浓厚的兴趣，并在老师的引导下，开始了解有关大数据的基本知识。

1. 大数据的定义

“大数据”研究机构Gartner将“大数据”定义为：“大数据”是无法在一定时间范围内用常规软件工具进行捕捉，需要新处理模式才能具有更强的决策力、洞察发现力和流程优化力来适应的海量、高增长和多样化的信息资产。

麦肯锡将“大数据”定义为：一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合，具有海量的数据规模、快速的数据流转、多样的数据类型和价值密度低四大特征。

维基百科将“大数据”定义为：大数据是指一些使用传统数据库管理工具或数据处理应用很难处理的大型而复杂的数据集。

2. 大数据的产生历史

2005年，Hadoop项目诞生。Hadoop本身不是一个产品，而是由多个软件产品组成的一个生态系统，这些软件产品共同实现功能全面和灵活的大数据分析。从技术上看，Hadoop由两项关键服务构成：采用Hadoop分布式文件系统（HDFS）的可靠数据存储服务和利用一种叫作MapReduce技术的高性能并行数据处理服务。这两项服务的共同目标是，提供一个使结构化和复杂数据的快速、可靠分析变为现实的基础。

2008年末，计算社区联盟（Computing Community Consortium）发表了一份有影响力的白皮书《大数据计算：在商务、科学和社会领域创建革命性突破》。它使人们的思维不仅局限于数据处理的机器，并提出：大数据真正重要的是新用途和新见解，而非数据本身。

2. 大数据的产生历史

2010年2月，大数据专题报告—《数据，无所不在的数据》在《经济学人》上发表。报告中提到：“世界上有着无法想象的巨量数字信息，并以极快的速度增长。”从经济界到科学界，从政府部门到艺术领域，很多方面都已经感受到了这种海量信息的影响。科学家和计算机工程师已经为这个现象创造了一个新词汇—大数据。

2011年5月，全球知名咨询公司麦肯锡全球研究院（MGI）发布了一份报告—《大数据：创新、竞争和生产力的下一个新领域》，自此大数据开始备受关注。

2011年11月，我国工业和信息化部发布的《物联网“十二五”发展规划》中，将信息处理技术作为四项关键技术创新工程之一提出来，其中包括了海量数据存储、数据挖掘、图像视频智能分析，这些都是大数据的重要组成部分。

2. 大数据的产生历史

2015年8月，国务院印发《促进大数据发展行动纲要》（以下简称《纲要》）。《纲要》的发布，标志着我国在顶层设计上对大数据的实践与实施做出了总体部署。

2016年12月，我国工业和信息化部印发《大数据产业发展规划（2016 - 2020年）》，全面部署“十三五”时期大数据产业发展工作，加快建设数据强国，为实现制造强国和网络强国提供强大的产业支撑。

2018年6月，国家市场监督管理总局、中国国家标准化管理委员会发布中国首个国家大数据交易标准《信息技术 数据交易服务平台 交易数据描述》（GB/T 36343—2018），该标准由中国电子技术标准化研究院、贵阳大数据交易所等单位起草，从2019年1月1日正式实施。

2. 大数据的产生历史

2020年12月，中国（上海）大数据产业创新峰会成功举办，会上发布了一批公共数据开放应用试点项目及大数据联合创新实验室建设成果；10个部门获颁公共数据开放应用成效突出部门；成立了“上海国际数据港产业合作共同体”；举行了开放数据赛事联盟各赛事颁奖仪式；揭幕了一批数智创新载体。至此，上海大数据产业发展和创新应用“十三五”完成收官，大数据赋能城市数字化转型全新起步。

2021年6月，《中华人民共和国数据安全法》由中华人民共和国第十三届全国人民代表大会常务委员会第二十九次会议通过，自2021年9月1日起施行。

3. 大数据的特征 (5V)

大数据主要有五大特征，包括体量大 (Volume)、多样化 (Variety)、速度快 (Velocity)、真实性 (Veracity) 和价值密度低 (Value)。大数据的“5V”特征表明大数据不仅仅是数据海量，对于大数据的分析和处理将更加复杂、速度更快且更注重时效。

体量大

多样化

速度快

真实性

价值密度低

4. 大数据的作用

- (1) 新一代信息技术融合应用的关键在于对大数据的处理和分析。
- (2) 大数据成为信息产业不断发展的新途径。
- (3) 大数据成为提升核心竞争力的关键因素。
- (4) 大数据时代科学研究方法也会出现相应变化。

5. 大数据的处理流程

目前，大数据领域每年都会涌现出大量新的技术，成为大数据获取、存储、处理分析和可视化的有效手段。大数据技术能够将大规模数据中隐藏的有价值信息和知识挖掘出来，为人类社会经济活动提供依据，提高各个领域的运行效率，乃至整个社会经济的集约化程度。在大数据环境下，数据来源非常丰富且数据类型多样，任何完整的大数据平台，其数据的处理过程一般都包括数据采集、数据清洗、数据存储、数据挖掘和数据展现五个过程。

说一说

现实生活中你所了解的大
数据应用实例有哪些？



4.4.2 大数据采集与分析方法

在了解了大数据的基础知识后，小华理解到，在大数据处理过程中，数据采集是基础，数据分析是关键，数据安全是保障。

大数据的采集方法一般有以下几种。

- (1) 离线采集。
- (2) 实时采集。
- (3) 互联网采集。
- (4) 其他数据采集方法。

2. 大数据的分析方法

大数据的意义不在于掌握多大量级的数据信息，而在于如何处理这些数据信息得到想要的结果。也就是说，大数据价值的关键在于对数据的加工能力和分析能力。对数据进行深度挖掘，可以解决实际问题，实现其价值。

数据分析是大数据管理的一大挑战。由于数据量较大，一般的数据分析应用程序无法很好地进行处理。大数据分析从技术手段上采用了最新的数据分析模型,通过数据之间特有的相关关系可以产生出许多有关联、有价值的结论。大数据分析在许多领域发挥了巨大的作用。

大数据分析主要包括以下五个基本方面，它们共同作用，决定了最终的大数据分析结果。

数据质量和数据管理

预测性分析

数据挖掘算法

可视化分析

语义引擎

3. 数据安全

大数据在成为社会竞争新焦点的同时，不仅带来了人类社会发展的新机遇，同时也带来了更多的数据安全风险，对人们提出了更高的数据安全防范要求。随着数据发掘的不断深入和在各行各业应用的不断推进，大数据安全的“脆弱性”逐渐凸显，国内外数据泄露事件频发，用户隐私受到极大挑战。在数据驱动环境下，网络攻击也更多地转向存储重要敏感信息的信息化系统，大数据安全防护已成为大数据应用发展的一项重要课题。

大数据安全的定义

大数据的安全风险

大数据的安全需求

大数据的安全策略

4. 大数据的发展趋势

(1) 数据资源化。

所谓资源化，是指大数据成为企业和社会关注的重要战略资源，并已成为大家争相抢夺的新焦点。因而，企业必须提前制订大数据营销战略计划，抢占市场先机。

(2) 与云计算的深度融合。

大数据离不开云计算，云计算为大数据提供了弹性可拓展的基础设备，是产生大数据的平台之一。自2013年开始，大数据技术已开始和云计算技术紧密结合，预计未来两者关系将更为密切。除此之外，物联网、移动互联网等新兴计算形态，也将一起助力大数据发展，让大数据营销发挥出更大的影响力。

(3) 与人工智能的深度结合。

人工智能通过数据采集、处理、分析，从各行各业的海量数据中，获得有价值的信息，为更高级的算法提供素材。人工智能其实就是以大量的数据为基础，让可以通过机器来做判别的问题最终转化为数据问题。人工智能的飞速发展，背后离不开大数据的支持。而在大数据的发展过程中，人工智能的加入也使得更多类型、更大体量的数据能够得到迅速处理与分析。

(4) 科学理论的突破。

随着大数据的快速发展，就像计算机和互联网一样，大数据很有可能引发新一轮的技术革命。随之兴起的数据挖掘、机器学习和人工智能等相关技术，可能会改变数据世界里的很多算法和基础理论，实现科学技术上的新突破。

未来，数据科学将成为一门专门的学科，被越来越多的人所认知。各大高校将设立专门的数据科学类专业，也会催生一批与之相关的新的就业岗位。与此同时，基于数据这个基础平台，也将建立起跨领域的数据共享平台，之后，数据共享将扩展到企业层面，并且成为未来产业的核心一环。

另外，大数据作为一种重要的战略资产，已经不同程度地渗透到每个行业领域和部门，其深度应用不仅有助于企业经营活动，还有利于推动国民经济发展。它对于推动信息产业创新、大数据存储管理挑战、改变经济社会管理面貌等方面意义重大。同时，合法地获取和使用数据也是用户应该努力培养的基本信息素养。

说一说

在信息社会应该树立怎样的大数据观念？



