

● 普通高等教育“十三五”规划教材
(计算机专业群)

Excel 2010 数据统计分析 实用案例教程

主 编 苏志军 楼 骏 汤巧英
副主编 景秀眉 王昌云 徐骏骅

普通高等教育“十三五”规划教材（计算机专业群）

Excel 2010 数据统计分析

实用案例教程

主 编 苏志军 楼 骏 汤巧英

副主编 景秀眉 王昌云 徐骏骅



中国水利水电出版社
www.waterpub.com.cn

• 北京 •

内 容 提 要

本书通过众多实际案例系统地介绍如何利用 Excel 2010 工具解决工作中的数据分析问题,帮助读者从日常错综复杂的数据中分析出有意义的结论。书中精心安排了案例的分析、实施、总结等多个过程,让读者对解决问题能有清晰的思路。

全书共 11 章:前 10 章系统讲解各类数据分析问题,包括描述性统计、概率分布、假设检验、参数估计、方差分析、相关分析、回归分析、时间序列分析、规划求解和决策树,选取了众多日常工作中遇到的数据分析问题进行详细讲解,再给出具体的操作过程,并且每一章后面都附有习题,方便读者巩固练习;第 11 章是专门的综合案例分析,运用大型实际案例帮助读者进一步认识数据统计分析。全书突出 3 个特点:道理简单明了、思路清晰透彻、案例新颖实用。

本书可作为普通高校金融、物流、电商、工程等涉及数据分析相关专业的教材,也可作为数据分析初学者进行 Excel 学习的辅导用书,还可作为各类相关培训班的培训教材。

图书在版编目(CIP)数据

Excel 2010 数据统计分析实用案例教程 / 苏志军, 楼骏, 汤巧英主编. — 北京: 中国水利水电出版社, 2018.1

普通高等教育“十三五”规划教材. 计算机专业群
ISBN 978-7-5170-6253-0

I. ①E… II. ①苏… ②楼… ③汤… III. ①表处理
软件—应用—统计分析—高等学校—教材 IV. ①C819

中国版本图书馆CIP数据核字(2018)第011325号

策划编辑: 石永峰

责任编辑: 周益丹 高 辉

封面设计: 李 佳

书 名	普通高等教育“十三五”规划教材(计算机专业群) Excel 2010 数据统计分析实用案例教程 EXCEL 2010 SHUJU TONGJI FENXI SHIYONG ANLI JIAOCHENG
作 者	主 编 苏志军 楼 骏 汤巧英 副主编 景秀眉 王昌云 徐骏骅
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路 1 号 D 座 100038) 网址: www.waterpub.com.cn E-mail: mchannel@263.net (万水) sales@waterpub.com.cn 电话: (010) 68367658 (营销中心)、82562819 (万水)
经 售	全国各地新华书店和相关出版物销售网点
排 版	北京万水电子信息有限公司
印 刷	三河市鑫金马印装有限公司
规 格	184mm×260mm 16 开本 12.25 印张 295 千字
版 次	2018 年 1 月第 1 版 2018 年 1 月第 1 次印刷
印 数	0001—3000 册
定 价	29.00 元

凡购买我社图书,如有缺页、倒页、脱页的,本社营销中心负责调换

版权所有·侵权必究

前 言

曾经有一位男性顾客投诉一家塔吉特店给他还在读书的女儿寄婴儿用品的优惠券。这家美国第二大零售商会搞出如此大的乌龙？经过这位父亲与女儿进一步沟通后，才发现自己的女儿真的已经怀孕了。原来每位顾客初次到塔吉特刷卡消费时，都会获得一组顾客识别编号，内含顾客姓名、信用卡卡号及电子邮件等个人资料。日后凡是顾客在塔吉特消费，计算机系统就会自动记录其消费内容、时间等信息。再加上从其他渠道取得的统计资料，塔吉特便能形成一个庞大的数据库。塔吉特的统计师们通过对顾客的消费习惯进行一次次的测试和数据分析，基本上可以判断出哪些顾客是孕妇，甚至还可以进一步估算出她们的预产期，在最恰当的时候给她们寄去最符合她们需要的优惠券，满足她们最实际的需求。这家成立于 1961 年的零售商能有今天的成功，数据分析功不可没。

上述例子就是零售企业数据分析的一个典型应用。当前，“大数据”这个词语在网络上喷涌而出。大数据掀起的风暴已席卷到生活的各个角落，大数据使我们的生活变得更加美好。那么什么是大数据？相信对很多人来说这还只是个模糊的概念，但是当前大数据已经广泛应用于日常生活的诸多领域。例如，必应搜索通过集成以往的飞机票价预测出未来票价走势；谷歌利用用户搜索记录判断出美国流感疫情的现状，在时间上比美国疾控中心的预测还快两周；对冲基金通过剖析社交网络推特的数据信息预测股市的表现；交通部门通过大数据分析出实时路况；交友网站利用大数据分析给需要的人匹配合适的对象以帮助其寻找爱情；利用穿戴装备（如智能手表、智能手环等）生成最新的数据，可以根据自身热量的消耗以及睡眠模式追踪身体健康情况等。这些无一不体现出大数据与我们的生活紧密相连。

大数据虽然无处不在，但是分析大数据对于大部分人来说还是个难题（无论是具备的软硬件还是知识水平都远远达不到要求）。我们只是享受了大数据的成果，获取了大数据带来的好处，自己却无法掌握它的原理和实际应用。因此，能不能让数据分析走到普通人的身边，在日常生活工作中，直接为其服务，而不再是专业人士的“专利”？本书编写的目的就在于此。

本书搜罗了大量来自我们日常生活工作的实际案例，有的来自社会领域、有的来自建筑工程领域、有的来自财经投资领域、有的来自水利行业、有的来自制造行业、有的来自生产销售行业、有的来自教育行业，可以说是包罗万象。目的是让读者从认识身边的小数据开始，逐渐来认识大数据。日常生活工作中遇到的数据可以说是无处不在，利用简单易学的 Excel 工具，就可以从数据中分析出一个个有意义的结论。在编写过程中，编者对每一个案例都进行了实证检验，并进行了通俗易懂的解读。本书重点突出了 3 个特点：道理简单明了、思路清晰透彻、案例新颖实用。

2015 年，马云在云栖大会上表示，人类已经由 IT 时代进入 DT 时代，数据取代石油成为最核心的资源，在未来，数据会成为像水、电、石油一样的公共资源。数据也成为了一种重要的商业资本，可以创造新的经济利益和商业模式。本书或许可以帮助读者开启一扇认识数据、分析数据、利用数据的知识之门。

一本书要体现作者个人的写作风格，书中文字是作者个性情感的流露，倾注了作者的思

想和灵魂。写一本书，尤其是写一本好书，必然有作者的思想表达和感情流露在里面。诚然，理工类的书没有文艺类的书显得那么明显，但还是可以察觉得到作者的用心所在。

本书的编写工作有很多老师参与，每一章节的文字、每一个案例都经过编者本人的解读、核实、修改、校验，使全书风格尽量保持统一。

编写这本书时得到了很多人的帮助和建议，感谢参与编写工作的老师以及众多网友（如关文忠、星空颖雪、阮一峰等），他们提供的许多经典案例给了我很大启发，让本书得以集众人之所长。另外，感谢我的家属在艰苦的岁月里鼓励我坚持完成本书的编写，也感谢中国水利水电出版社在本书出版过程中给予的大力支持。

由于时间仓促及编者水平有限，书中的错误和不妥之处在所难免，恳请专家和读者批评指正。

苏志军

2017年12月

目 录

前言

第1章 描述性统计	1	1.5.4 案例实施	16
1.1 频数和算术平均值	1	1.5.5 总结	17
1.1.1 案例提出	1	1.5.6 拓展练习	17
1.1.2 相关知识点	2	第2章 概率分布	18
1.1.3 案例分析	3	2.1 二项分布	18
1.1.4 案例实施	3	2.1.1 案例提出	18
1.1.5 总结	4	2.1.2 相关知识点	19
1.1.6 拓展练习	4	2.1.3 案例分析	20
1.2 几何平均值和调和平均值	5	2.1.4 案例实施	20
1.2.1 案例提出	5	2.1.5 总结	22
1.2.2 相关知识点	5	2.1.6 拓展练习	22
1.2.3 案例分析	6	2.2 泊松分布	23
1.2.4 案例实施	6	2.2.1 案例提出	23
1.2.5 总结	7	2.2.2 相关知识点	23
1.2.6 拓展练习	7	2.2.3 案例分析	24
1.3 众数和中位数	7	2.2.4 案例实施	24
1.3.1 案例提出	7	2.2.5 总结	25
1.3.2 相关知识点	8	2.2.6 拓展练习	25
1.3.3 案例分析	8	2.3 正态分布	25
1.3.4 案例实施	9	2.3.1 案例提出	25
1.3.5 总结	10	2.3.2 相关知识点	26
1.3.6 拓展练习	10	2.3.3 案例分析	26
1.4 偏度和峰度	11	2.3.4 案例实施	27
1.4.1 案例提出	11	2.3.5 总结	28
1.4.2 相关知识点	12	2.3.6 拓展练习	29
1.4.3 案例分析	12	第3章 假设检验	30
1.4.4 案例实施	13	3.1 z-检验和 t-检验	30
1.4.5 总结	14	3.1.1 案例提出	31
1.4.6 拓展练习	14	3.1.2 相关知识点	33
1.5 方差和标准差	14	3.1.3 案例分析	34
1.5.1 案例提出	14	3.1.4 案例实施	34
1.5.2 相关知识点	15	3.1.5 总结	37
1.5.3 案例分析	15	3.1.6 拓展练习	37

3.2 F-检验	39	5.1.2 相关知识点	57
3.2.1 案例提出	39	5.1.3 案例分析	57
3.2.2 相关知识点	39	5.1.4 案例实施	58
3.2.3 案例分析	40	5.1.5 总结	59
3.2.4 案例实施	41	5.1.6 拓展练习	60
3.2.5 总结	42	5.2 双因素方差分析	61
3.2.6 拓展练习	42	5.2.1 案例提出	61
3.3 卡方检验	42	5.2.2 相关知识点	62
3.3.1 案例提出	43	5.2.3 案例分析	62
3.3.2 相关知识点	43	5.2.4 案例实施	63
3.3.3 案例分析	44	5.2.5 总结	65
3.3.4 案例实施	44	5.2.6 拓展练习	66
3.3.5 总结	46	第6章 相关分析	67
3.3.6 拓展练习	46	6.1 双变量相关分析	67
第4章 参数估计	48	6.1.1 案例提出	67
4.1 总体均值估计	48	6.1.2 相关知识点	68
4.1.1 案例提出	48	6.1.3 案例分析	69
4.1.2 相关知识点	49	6.1.4 案例实施	69
4.1.3 案例分析	49	6.1.5 总结	72
4.1.4 案例实施	49	6.1.6 拓展练习	72
4.1.5 总结	50	6.2 多重相关和偏相关分析	73
4.1.6 拓展练习	50	6.2.1 案例提出	73
4.2 总体比例估计	51	6.2.2 相关知识点	74
4.2.1 案例提出	51	6.2.3 案例分析	74
4.2.2 相关知识点	51	6.2.4 案例实施	75
4.2.3 案例分析	51	6.2.5 总结	76
4.2.4 案例实施	51	6.2.6 拓展练习	77
4.2.5 总结	52	第7章 回归分析	79
4.2.6 拓展练习	52	7.1 一元线性回归分析	79
4.3 样本容量估计	52	7.1.1 案例提出	79
4.3.1 案例提出	53	7.1.2 相关知识点	80
4.3.2 相关知识点	53	7.1.3 案例分析	80
4.3.3 案例分析	54	7.1.4 案例实施	80
4.3.4 案例实施	54	7.1.5 总结	84
4.3.5 总结	55	7.1.6 拓展练习	85
4.3.6 拓展练习	55	7.2 多元线性回归分析	85
第5章 方差分析	56	7.2.1 案例提出	85
5.1 单因素方差分析	56	7.2.2 相关知识点	86
5.1.1 案例提出	56	7.2.3 案例分析	87

7.2.4 案例实施	87	9.2.2 案例实施	121
7.2.5 总结	89	9.2.3 分析结论	123
7.2.6 拓展练习	89	9.3 销售利润最大化	124
7.3 非线性回归分析	91	9.3.1 案例分析	124
7.3.1 案例提出	91	9.3.2 案例实施	124
7.3.2 相关知识点	92	9.3.3 分析结论	125
7.3.3 案例分析	92	9.4 最佳生产方案	126
7.3.4 案例实施	92	9.4.1 案例分析	127
7.3.5 总结	94	9.4.2 案例实施	127
7.3.6 拓展练习	94	9.4.3 分析结论	128
第8章 时间序列分析	96	9.5 总结	128
8.1 移动平均法	96	9.6 拓展练习	128
8.1.1 案例提出	97	第10章 决策树	131
8.1.2 相关知识点	98	10.1 开工方案决策	132
8.1.3 案例分析	99	10.1.1 案例分析	132
8.1.4 案例实施	99	10.1.2 案例实施	133
8.1.5 总结	103	10.1.3 分析结论	134
8.1.6 拓展练习	103	10.2 投资理财风险决策	134
8.2 指数平滑法	104	10.2.1 案例分析	135
8.2.1 案例提出	104	10.2.2 案例实施	135
8.2.2 相关知识点	105	10.2.3 分析结论	136
8.2.3 案例分析	105	10.3 电视机厂投产决策	137
8.2.4 案例实施	105	10.3.1 案例分析	137
8.2.5 总结	110	10.3.2 案例实施	137
8.2.6 拓展练习	110	10.3.3 分析结论	138
8.3 季节变动预测法	110	10.4 企业投资咨询决策	139
8.3.1 案例提出	111	10.4.1 案例分析	139
8.3.2 相关知识点	112	10.4.2 案例实施	140
8.3.3 案例分析	112	10.4.3 分析结论	145
8.3.4 案例实施	113	10.5 总结	145
8.3.5 总结	116	10.6 拓展练习	146
8.3.6 拓展练习	116	第11章 综合案例分析	148
第9章 规划求解	118	11.1 我国家庭收入数据分析	148
9.1 淘宝“双11”活动凑单	119	11.1.1 案例分析	148
9.1.1 案例分析	120	11.1.2 案例实施	148
9.1.2 案例实施	120	11.1.3 分析结论	151
9.1.3 分析结论	121	11.2 美国枪击案数据分析	152
9.2 巧解九宫格数学题	121	11.2.1 案例分析	152
9.2.1 案例分析	121	11.2.2 案例实施	153

11.2.3 分析结论	154	11.7.2 案例实施	167
11.3 游戏玩家调查分析	154	11.7.3 分析结论	170
11.3.1 案例分析	154	11.8 制鞋厂经营安全性分析	170
11.3.2 案例实施	155	11.8.1 案例分析	170
11.3.3 分析结论	156	11.8.2 案例实施	171
11.4 房地产评估	156	11.8.3 分析结论	175
11.4.1 案例分析	157	11.9 钢板最佳切割方案	176
11.4.2 案例实施	157	11.9.1 案例分析	176
11.4.3 分析结论	159	11.9.2 案例实施	177
11.5 股票投资风险分析	160	11.9.3 分析结论	178
11.5.1 案例分析	160	11.10 研究院的决策难题	179
11.5.2 案例实施	161	11.10.1 案例分析	179
11.5.3 分析结论	161	11.10.2 案例实施	179
11.6 股市图形分析	162	11.10.3 分析结论	181
11.6.1 案例分析	162	11.11 案例实践	182
11.6.2 案例实施	163	11.11.1 项目问卷	182
11.6.3 分析结论	165	11.11.2 信心指数的计算方法	183
11.7 蒙特卡洛模拟	166	11.11.3 研究目标及思路	184
11.7.1 案例分析	166	参考文献	185

第 1 章 描述性统计

描述性统计是指将调查样本中包含的大量数据资料进行整理、概括和计算，是推断性统计的基础。

描述性统计是以揭示数据分布特性的方式汇总并表达定量数据的方法。主要包括数据频数分析、数据集中趋势分析、数据离散程度分析、数据的分布以及一些基本的统计图形。

常见的描述性统计方法可分为三类：一是用数据的统计量来描述，如均值、标准差等；二是用图示技术来描述，如直方图、散布图、趋势图、排列图、条形图和饼图等；三是用语言文字分析和描述，如统计分析表、分层、因果图、亲和图和流程图等。描述性统计的应用范围广，它适用于能够收集到定量数据的所有领域。

我们研究大量数据发现：多数情况下，数据出现的频数会呈现出一种钟形分布。即各个变量值与中间位置的距离越近，出现的次数越多，反之越少。从而形成了一种以中间值为中心的集中趋势，这个集中趋势是现象规律性的体现。

1.1 频数和算术平均值

频数也称为次数，就是对数据按某种标准分组，统计各组内包含个体的数目。例如，一月内某河道水位 1.5 米~2 米的天数为 13 天，1 米~1.5 米的天数为 15 天，13 和 15 即为频数。

算术平均值大家就比较熟悉了，一般也称为算术平均数，比如计算某地区的平均日降雨量、某课程的平均成绩、某城市建筑的平均高度等。

1.1.1 案例提出

【例 1.1】某研究部门对某城市 CBD 区域的建筑高度做了个调查统计，统计的建筑及其高度如图 1-1 所示，现在需要统计各个高度区间内建筑物的数量。

A		B	C	D	E
1	某市高层建筑统计				
2	大楼编号	高度（单位：米）	区间	数量	
3	A01	120	<=50		
4	A02	49	50~100		
5	A03	86	101~200		
6	A04	88	201~300		
7	A05	150	300~400		
8	A06	220	400~500		
9	A07	360	>=500		
10	A08	480			
11	A09	510			
12	A10	251			
13	A11	78			
14	A12	68			
15	A13	95			
16	A14	101			
17	A15	452			
18	A16	480			
19	A17	380			
20	A18	110			
21	A19	182			
22	A20	68			

图 1-1 某市高层建筑统计

【例 1.2】淘宝同类店铺很多，有的卖家月营业额是几万元，而有的卖家月营业额只有几千元。如图 1-2 所示，对出售某儿童用品的淘宝店月营业额进行了抽样统计，A 列按照营业额分为 7 个等级，C 列为抽样店铺数量，B 列为对应等级店铺月营业额的平均数，现在需要计算所有店铺的平均营业额。

	A	B	C
1	出售某儿童用品的淘宝店月营业额		
2	按高低分类	月营业额(元/店)	抽样店铺数
3	最低营业额	2972.38	420
4	低营业额	4718.44	396
5	中下营业额	6517.99	765
6	中等营业额	8989.65	725
7	中上营业额	12002.25	658
8	高营业额	17092.36	321
9	最高营业额	26988.25	218

图 1-2 出售某儿童用品的淘宝店月营业额

1.1.2 相关知识点

1. 计算频数

我们可以用 Excel 2010 中的 FREQUENCY() 函数来实现，FREQUENCY() 函数有 2 个参数，如图 1-3 所示。

函数功能：返回一组数据的频数分布。

参数说明：Data_array：用来计算频数的数组。

Bins_array：计算频数的分段点，注意统计的是数据值 ≤ 分段点的频数。

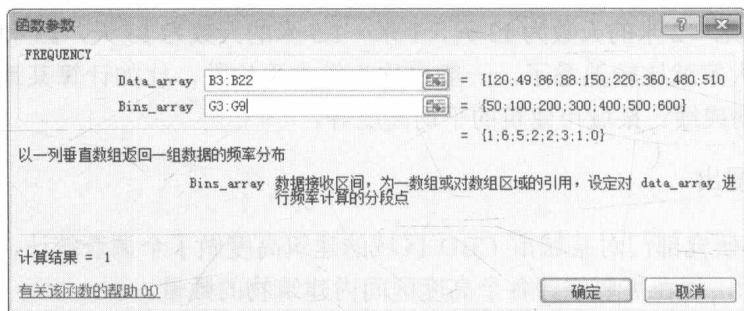


图 1-3 FREQUENCY() 函数参数

2. 计算算术平均值

通常用我们熟知的 AVERAGE() 函数即可，但是诸如例 1.2 之类的涉及组数据的算术平均值就做不到，这时我们可以采用 SUMPRODUCT() 函数求得乘积之和再来算平均值，如图 1-4 所示。

函数功能：返回相应数组或区域乘积和。

参数说明：Array1：需要乘积的第一组数区域。

Array2：需要乘积的第二组数区域。

如图 1-4 中 Array1、Array2、Array3 所选区域，那么 SUMPRODUCT() 函数的值实际就是以下三组单元格数字的成绩之和。表达式如下：

$$\text{SUMPRODUCT}(B3:B9,C3:C9,D3:D9)=B3*C3*D3+B4*C4*D4+\dots+B9*C9*D9$$

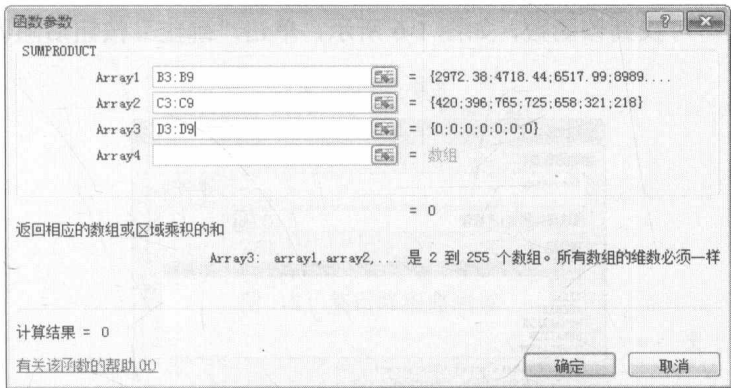


图 1-4 SUMPRODUCT()函数系数

1.1.3 案例分析

【例 1.1】统计各个高度区间内建筑物的数量显然是计算频数的问题，根据题意需要构造分段点。

【例 1.2】如果我们计算出每一组总营业额，即每一组的月营业额乘以对应的店铺数，然后将得到的数字相加再除以总的店铺数即可得到所有店铺的平均营业额，但是如果涉及数据较多会比较麻烦，采用 SUMPRODUCT()函数来求乘积和会非常方便。

1.1.4 案例实施

- 【例 1.1】实施步骤：
- (1) 根据题意对建筑物高度的区间构造分段点，见图 1-5 中 Excel 表格中的 G 列数据。
注意：分段点要从小到大排列，次序不可错乱。

	A	B	C	D	E	F	G	H	I	J
1	某市高层建筑统计									
2	大楼编号	高度（单位：米）		区间	数量		数据分段点			
3	A01	120		<=50	G3:G9)		50			
4	A02	49		50~100			100			
5	A03	86		101~200			200			
6	A04	88		201~300			300			
7	A05	150		300~400			400			
8	A06	220		400~500			500			
9	A07	360		>=500						
10	A08	480								
11	A09	510								
12	A10	251								
13	A11	78								
14	A12	68								
15	A13	95								
16	A14	101								
17	A15	452								
18	A16	480								
19	A17	380								
20	A18	110								
21	A19	182								
22	A20	68								
23										

函数参数

FREQUENCY

Data_array B3:E22 = {120;49;86;88;150;220;360;480;510}

Bins_array G3:G9 = {50;100;200;300;400;500;600}

= {1;6;5;2;2;3;1;0}

以一列垂直数组返回一组数据的频率分布

Data_array 用来计算频率的数组，或对数组单元区域的引用（空格及字符串忽略）

计算结果 = 1

有关该函数的帮助(H)

确定 取消

图 1-5 FREQUENCY()函数参数及相关数据

- (2) 选择 E3:E9 单元格，单击公式编辑框边上的 f_x 按钮，搜索函数文本框输入 frequency，

单击“转到”按钮即可找到该函数，如图 1-6 所示，单击“确定”按钮后随即跳出该函数的参数对话框。

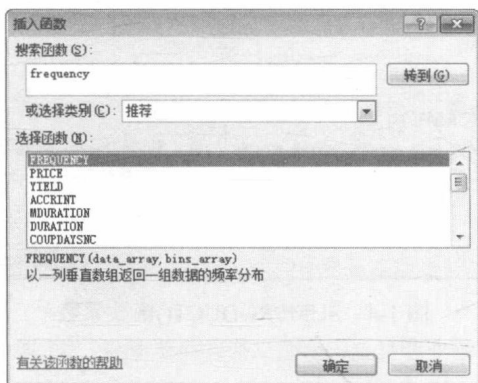


图 1-6 “插入函数”对话框

(3) 如图 1-5 所示，在参数 Data_array 中选择需要计算频数的数据区域，即 B3:B22；在参数 Bins_array 中选择分段点的数据区域，即 G3:G8；然后单击“确定”按钮。

(4) 将光标点到公式编辑栏，按下 Ctrl+Alt+Enter 组合键，即可得到计算结果。

(注：末尾出现的一个数据表示为高度 $\geq 500\text{m}$ 的建筑物数量)

【例 1.2】实施步骤：

(1) 我们需要求出抽样店铺的总数和总营业额，店铺总数用 SUM() 函数轻松可得。

(2) 对于总营业额的计算需要用 SUMPRODUCT() 函数，如图 1-7 所示。输入参数后单击“确定”按钮即可求得。

店铺总数	3503
总营业额	=SUMPRODUCT(B3:B9, C3:C9)
店铺平均营业额	

图 1-7 计算总营业额

(3) 店铺的平均营业额也轻松可得：店铺平均营业额=总营业额/店铺总数。

1.1.5 总结

(1) 使用 FREQUENCY() 函数可以实现数据频数的计算，这里要注意的是使用前需要构造合理的分段点，而且计算的是统计数据值 $\leq$ 分段点的频数。另外，因为是组数据计算，在操作上需要使用 Ctrl+Alt+Enter 组合键。

(2) SUMPRODUCT() 函数求的是几组数据的乘积之和，各组数据的个数要一致，否则计算要出错。

1.1.6 拓展练习

(1) 请自行登录当地的气象网站，摘录每日气温，前后共计 3 个月时间。请构造合适的分段点，统计温度在各个分段区间内出现的天数。

(2) 某市在各个地区设立了降雨监测点，某天突降大雨，各个监测点测量的数据如图 1-8

所示。请计算该市平均降雨量。

某市降雨监测点	
测得降雨量（单位：mm）	监测点数目
231.2	3
256.6	1
236.5	2
225.7	5
222.1	4

图 1-8 某市降雨监测点

1.2 几何平均值和调和平均值

除了我们日常熟悉的算术平均值之外，还有几何平均值和调和平均值，它们在不同的领域有各自的应用。

1.2.1 案例提出

【例 1.3】众所周知，居民的幸福指数跟居民收入有很大联系，一般情况下居民收入越高幸福指数越高，但是不成正比关系，因为幸福指数的参考不仅仅是收入。随机调查了两个地区居民收入，现在要根据两地的人均收入（没有其他更多数据）来估算两地居民的幸福指数，如图 1-9 所示。

	A	B	C	D	E
1	A地区			B地区	
2	人员1	收入		人员1	收入
3	人员2	50000		人员2	32000
4	人员3	48000		人员3	200000
5	人员4	39000		人员4	42000
6	人员5	52000		人员5	36000
7	人员6	50000		人员6	250000
8	人员7	48000		人员7	30000
9	人员8	53000		人员8	25000
10	人员9	49000		人员9	15000
11	人员10	36000		人员10	36000

图 1-9 A 地区和 B 地区的居民收入数据

【例 1.4】有车一族都知道汽车轮胎经过一定里程的行驶就会有磨损，磨损达到一定程度就需要换轮胎。由于一般汽车发动机都是前置，前轮为驱动轮，驱动轮的磨损比较严重。为了物尽所用，一般行驶到一定里程后 4S 店会建议前后轮胎互换，然后接着行驶，最后等到轮胎需要报废时 4 个轮胎正好达到一样的磨损程度。根据测算，某个品牌的车前轮行驶 6 万 km 就必须报废，后轮行驶 9 万 km 才报废。请问该品牌车辆最多可以行驶多少 km，行驶多少 km 的时候需要前后轮互换？

1.2.2 相关知识

几何平均值是日常常见的度量平均值的统计方法，常用于增长率、收益率等计算，比较不受极端值影响，但数据不能为 0 或者负数。计算公式如下：

$$\overline{X_G} = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n}$$

(公式 1.1)

调和平均值表示所有数据倒数的算术平均值的倒数, 公式如下:

$$\overline{X_H} = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{X_i}} = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}} \quad (\text{公式 1.2})$$

在 Excel 2010 中, 可以使用 GEOMEAN() 函数和 HARMEAN() 函数来实现几何平均值与调和平均值的计算。

1.2.3 案例分析

【例 1.3】幸福指数跟收入紧密相关, 但是年收入千万的人未必比年收入十万的人幸福 100 倍。如图 1-9 所示, 显然 A 地区的人均收入比较均衡, B 地区有 2 个特别高收入的人士, 极大地拉高了人均收入, 如果用算术平均值来比对幸福指数, 无法获得 B 地区大多数人的认同, 所以就可以考虑用几何平均值来压低特别高收入的个人对平均幸福指数的拔高效应。

【例 1.4】假定前后轮待磨损量各为 1, 那么每行驶 1 万 km, 前轮磨损量为 $1/6$, 后轮磨损量为 $1/9$, 单位 (1/万)。假设一共行驶 x 万 km 后, 前后轮正好同时报废, 那么前后轮行驶 x 万 km 后, 它们的磨损量正好达到了 2, 有如下方程:

$$x \times \left(\frac{1}{6} + \frac{1}{9} \right) = 2 \quad (\text{方程 1.1})$$

求得

$$x = \frac{2}{\frac{1}{6} + \frac{1}{9}}$$

求得的 x 就是车辆总的行驶里程, 就是两个数字倒数的算术平均值的倒数, 即为调和平均值。

对于第二个问题, 先假定行驶 y 万 km 后需要前后轮调换, 调换时前轮剩余磨损量为 $1-y/6$, 然后前轮拿到后轮去磨损还可以行驶 $x-y$ 万 km, 有如下方程:

$$1 - \frac{y}{6} = (x - y) \times \frac{1}{9} \quad (\text{方程 1.2})$$

求出的 y 即为需要前后轮调换时的行驶里程。在 Excel 2010 中无需通过公式计算, 只要案例是调和平均的问题, 直接采用 HARMEAN() 函数即可完成。

1.2.4 案例实施

【例 1.3】实施步骤:

- (1) 如图 1-9 所示, 在 Excel 表格中输入原始数据。
- (2) 选择 A15 单元格, 在公式编辑框中输入=GEOMEAN(B3:B11), 即可求得 A 地区的几何平均收入。
- (3) 选择 A14 单元格, 在公式编辑框中输入=AVERAGE(B3:B11), 即可求得 A 地区的算术平均收入。
- (4) 同理可求得 B 地区的几何平均收入和算术平均收入, 如图 1-10 所示。

13	A地区收入		B地区收入		比对
14	算术平均值	47222.22	74000		0.638138
15	几何平均值	46872.52	46310.64		1.012133

图 1-10 A 地区和 B 地区的居民收入平均值

分析结论：按照算术平均收入来看，A地区的幸福指数只有B地区的0.64倍，显然B地区绝大多数居民认为自己的收入比A地区要低，幸福指数也应该更低才对，统计的结论与多数人的观念相反。然而按照几何平均收入来看，A地区的幸福指数是B地区的1.01倍，显得比较客观。

【例 1.4】实施步骤：

如图 1-11 所示，用 HARMEAN()函数直接可以求得数据的调和平均值。

里程（单位：万km）	
前轮可行使	6
后轮可行使	9
调和平均值	=HARMEAN(B28:B29)

图 1-11 使用 HARMEAN()函数计算调和平均值

分析结论：求出的调和平均值为 7.2，说明汽车的前后轮一共可以行驶 7.2 万 km。将 7.2 代入方程 1.2 中的 x ，即可求得 y 为 3.6，说明在行驶到总里程一半的时候前后轮胎调换最佳。

小知识：如果我们不知道前后轮能行驶多少 km（如新车到手后），这时我们可以先测量轮胎的胎纹厚度，比如是 1.5cm，等行驶了 3 万 km 后再测量下前后轮的胎纹厚度，比如前轮剩下 0.8cm，后轮剩下 1.2cm，通过计算前后轮的磨损程度来计算前后轮什么时候可以调换。

1.2.5 总结

（1）几何平均值在某些情况下可以一定程度上压低算术平均带来的极端数据的影响，使得统计更加接近真实情况，例 1.3 中的幸福指数比对即为一例。

（2）我们发现调和平均值的计算创造了最大化的经济价值，在经济学、工程学中有广泛应用。

1.2.6 拓展练习

（1）轮船从 A 到 B 顺流速度是 50km/h，从 B 到 A 逆流速度是 30km/h，请问它的平均速度是多少？

（2）某工业机器上有 3 个部位可以用到相同的零件，已知这 3 个部位零件的使用寿命不一样，A 部位只能使用 20 天，B 部位可以使用 30 天，C 部位可以使用 50 天，请问这 3 个零件如何使用可以使得使用时间达到最大化？

1.3 众数和中位数

众数是一组数据中出现次数最多的那个数，常用 M_o 表示，通常我们用众数来反映社会经济现象的一般水平。例如， $M_o=85$ 反映了某次考试学生成绩最集中的水平是 85 分；中国成年女性鞋子尺码的众数为 37，鞋企在生产时需要考虑将女鞋 37 码的产量设成最大；某地区居民最普遍的生活水平日均消费 25 元人民币等。

中位数是数据按大小排列后处于中间位置的数值，常用 M_e 表示。表明其数据分成相等的两份，一半大于中位数，一半小于中位数，中位数是中间位置代表值，不受极端数据影响，所以通常用来衡量地区的收入水平。

1.3.1 案例提出

【例 1.5】小王专门研究某银行股票，他统计了该股票去年 180 个交易日的价格变化区间，

如图 1-12 所示, 求该股票价格的众数和中位数。

	A	B
1	某银行股票价格区间	
2	价格区间	出现天数
3	2~2.5	23
4	2.5~3	15
5	3~3.5	27
6	3.5~4	54
7	4~4.5	13
8	4.5~5	32
9	5~5.5	16
10		180

图 1-12 某银行股票价格区间

1.3.2 相关知识

非组数据的众数和中位数的计算比较简单, 用 Excel 2010 中的 MODE() 函数和 MEDIAN() 函数即可求得, 对于组数据的计算需要用到相关公式。

众数计算公式:

$$M_o = L + \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})} \times c \quad (\text{公式 1.3})$$

说明: L 为众数所在组的下限; f_m 为众数所在组的频数; f_{m-1} 和 f_{m+1} 分别表示众数所在组的前一组和后一组的频数; c 为众数所在组的间距宽度。

提示: 频数最大的组即为众数所在组。

中位数计算公式:

$$M_e = L + \frac{\frac{\sum_{i=1}^n f_i}{2} - s_{m-1}}{f_m} \times i \quad (\text{公式 1.4})$$

说明: L 为中位数所在组的下限; f_m 为中位数所在组的频数; s_{m-1} 为中位数前一组的累积频数; i 为中位数所在组的组距; $\sum_{i=1}^n f_i$ 就是各组频数之和, 即总频数。

提示: 累积频数刚好超过总频数一半的组即为中位数所在组。

1.3.3 案例分析

对于例 1.5 中的众数计算, 由于是组数据, 不能直接用函数来求, 需要用公式 1.3 来求。首先可以通过 MAX() 函数轻易获得频数最大的组是价格在 3.5~4 所在的组 (频数为 54), 即为众数所在组, 下限为 3.5; 众数所在组的前一组和后一组的频数不难得出是 27 和 13, 众数所在组的组距为 0.5。

例中组数据中位数的计算需要用公式 1.4 来求, 首先需要求得各组的累积频数, 可以通过公式得到。例如, 价格 2.5~3 组的累积频率为 23+15=38, 价格 3~3.5 组的累积频率为 23+15+27=65, 当累积频数正好超过总频数一半时的那一组即为中位数所在组。经过测算, 中位数所在组为价格 3.5~4 组。