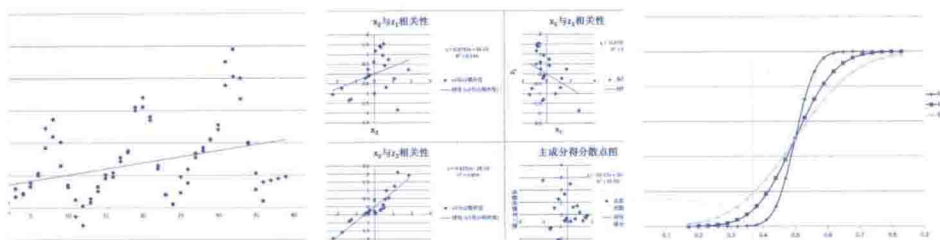
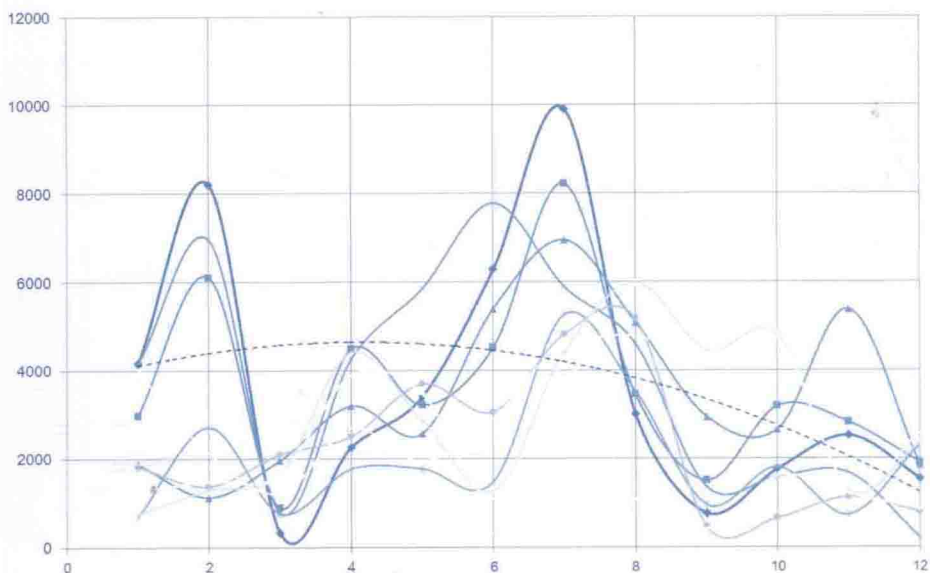


- 用Excel揭开统计学冰冷的面纱
- 全程用心记录每一步操作
- 让你的数据分析更专业
- 让你的受众更懂你



# 精通 Excel 数据统计与分析

蒲括 邵朋 编著

# 精通 Excel 数据统计与分析

蒲括 邵朋 编著

人民邮电出版社  
北京

## 图书在版编目 (C I P) 数据

精通Excel数据统计与分析 / 蒲括, 邵朋编著. —  
北京: 人民邮电出版社, 2014. 8  
ISBN 978-7-115-33103-8

I. ①精… II. ①蒲… ②邵… III. ①表处理软件  
IV. ①TP391.13

中国版本图书馆CIP数据核字(2013)第216110号

## 内 容 提 要

本书由简单到复杂, 详细介绍了各种统计分析功能在 Excel 中的实现过程。通过对统计理论的回顾和操作实例的讲解, 实现以 Excel 为工具解决统计分析问题, 帮助读者在掌握统计学原理的基础上, 熟练运用 Excel 进行统计分析。

全书共 11 章, 系统讲解统计分析中的各类问题, 包含描述性统计、统计图绘制、假设检验、方差分析、回归分析、相关分析、判别分析、时间序列分析、马尔可夫链分析, 聚类分析, 以及因子分析与主成分分析等。书中选取有针对性的例子讲解具体操作, 个别章节还介绍如何用 VBA 编程的方法来解决统计分析问题。全书知识全面, 实例丰富, 翔实, 操作具体详尽, 可使读者轻松入门并快速提高。

本书可供从事数据统计与分析的人员以及想深入了解 Excel 统计功能的读者使用, 也可作为 Excel 统计学参考书供高校以及各类相关培训班使用。

- 
- ◆ 编 著 蒲 括 邵 朋  
责任编辑 贾鸿飞  
责任印制 杨林杰
  - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号  
邮编 100164 电子邮件 315@ptpress.com.cn  
网址 <http://www.ptpress.com.cn>  
北京鑫正大印刷有限公司印刷
  - ◆ 开本: 787×1092 1/16  
印张: 16.75  
字数: 400 千字  
印数: 1—3 000 册
- 

定价: 49.00 元

读者服务热线: (010)81055410 印装质量热线: (010)81055316

反盗版热线: (010)81055315

# 前 言

1980年，著名未来学家阿尔文·托夫勒在其出版的经典著作《第三次浪潮》中预言：如果说IBM的主机拉开了信息化革命的大幕，那么“大数据”则是第三次浪潮的华彩乐章。

2011年5月，麦肯锡公司发布了《大数据：创新、竞争和生产力的下一个前沿领域》报告，陈述了“大数据”时代的到来，指出“数据已经渗透到每一个行业和业务职能领域，逐渐成为重要的生产因素”。

2012年，Twitter上每天发布超过4亿条微博，Facebook上每天更新的照片超过1000万张，Farecast公司用将近10万亿条价格记录来预测机票价格。

关于数据在我们这个时代所起的作用，其实也无需赘言。但如何把隐没在一大批看起来杂乱无章的数据里的信息、规律、关键点萃取出来，却是很多人都应该了解，甚至是应该深入研究的学问。对很多在工作中需要跟数据打交道的人，或者就是以数据统计、分析为工作的人来说，不仅仅是需要一定的统计学知识，而且还需要用统计分析工具将结果呈现出来。

毋庸置疑，对大多数人来讲，用得最多的可以用来进行数据统计与分析的工具就是Excel——这不仅是因为Excel相对容易获得，还因为Excel具备丰富的统计与分析功能，可以满足大多数人的大部分需求。

而这些需要用Excel进行数据统计与分析的人，大致可以分为这么几类：普通的办公一族，经常需要对数据进行统计与分析，而且想让自己的分析更加专业；从事数据统计与分析工作，使用专业的数据分析软件，也以Excel作为备用工具；统计学及相关专业的学生；想从事数据分析相关工作的人。

本书在对用Excel进行数据统计与分析的讲解上，紧扣这部分人的学习需求——先给出实例，然后讲解需要应用的分析原理，最后详细给出用Excel实现的方法与步骤。

需要说明的是，本书默认读者已经掌握以下内容：创建工作簿、插入工作表、保存文件等基本操作；在各个工作簿之间浏览；使用Windows基本功能，例如文件管理、复制和粘贴等操作。

## 本书的组织结构

本书分为两个部分。

第一部分：基础知识。

这一部分包括第1章和第2章。这一部分首先总体介绍了统计分析中需要用到的统计工具和统计图形等。目的是让读者对Excel在统计分析中的应用有一个基本的了解。

第二部分：统计分析。

第3~11章，分别介绍了假设检验、方差分析、回归分析、相关分析、判别分析、时间序列分析、马尔可夫链分析、聚类分析以及主成分分析等内容。对这些分析方法，本书给出了详细的例子及操作过程，以供读者参考。

## 章节结构

本书在介绍每种分析方法时也分为两个部分。

第一部分：分析原理。

本书在每章都会先介绍分析方法的原理，方便读者理解每步操作的实际作用，进而学会理解得到的结果的意义，以及如何对结果进行判定。

第二部分：分析实例。

为了方便读者学习如何在 Excel 中实现分析方法，每种分析方法都会分别以统计函数、趋势线和数学分析工具等不同的方式来实现。对于不同的实例，最简洁方便的分析方法不可能完全一样，所以需要在熟练使用各种实现方式之后，才能在实际中灵活使用，提高效率。

## 致 谢

在成书过程中，需要感谢门春杰、张铮、王杉、马宏和李广鹏等参与了本书的部分编写工作，感谢杜强和王命达等参与本书的审校对并对本书策划提出了宝贵意见。同时，也向关心本书的同仁和朋友表示诚挚的感谢。

尽管本书编者尽了最大努力，仍难免有不尽如人意之处，恳请广大读者提出宝贵意见和建议。欢迎登录 <http://www.book95.com> 与我们联系，也可以发邮件至 [jiahongfei@ptpress.com.cn](mailto:jiahongfei@ptpress.com.cn) 与图书编辑联系。

编 者  
2013 年夏

请在最左侧的本

# 目 录

第 1 章 描述性统计	1	3.2 t-检验	87
1.1 列联表	1	3.2.1 t-检验原理	87
1.2 数据频数分析	5	3.2.2 t-检验实例	87
1.2.1 频数分布函数	5	3.3 F-检验	95
1.2.2 直方图分析工具	7	3.3.1 F-检验原理	95
1.3 数据集中趋势分析	11	3.3.2 F-检验实例	95
1.3.1 算术平均值	11	3.4 Z-检验	99
1.3.2 几何平均值	16	3.4.1 Z-检验原理	99
1.3.3 调和平均值	17	3.4.2 Z-检验实例	99
1.3.4 众数	18	3.5 小结	105
1.3.5 中位数	20	3.6 习题	105
1.4 数据的离散程度分析	22	第 4 章 方差分析	107
1.4.1 极差	22	4.1 方差分析简介	107
1.4.2 四分位差	23	4.2 单因素方差分析	107
1.4.3 方差和标准差	27	4.2.1 单因素方差分析原理	107
1.4.4 变异系数	30	4.2.2 单因素方差分析实例	108
1.5 分布形态测定及分析	32	4.3 双因素方差分析	112
1.5.1 偏度	32	4.3.1 双因素方差分析原理	112
1.5.2 峰度	38	4.3.2 无重复双因素方差分析实例	115
1.6 描述统计工具	42	4.3.3 有重复双因素方差分析实例	119
1.7 小结	44	4.4 小结	122
1.8 习题	44	4.5 习题	122
第 2 章 统计图绘制	47	第 5 章 回归分析	124
2.1 直条图	48	5.1 回归分析简介	124
2.1.1 单式直条图	48	5.2 一元线性回归分析	124
2.1.2 复式直条图	52	5.2.1 一元回归分析简介	124
2.1.3 分段条图	55	5.2.2 一元回归分析实例	125
2.1.4 误差条图	58	5.3 多元线性回归分析	132
2.2 线图	61	5.3.1 多元线性回归分析简介	133
2.3 散点图	65	5.3.2 多元线性回归分析实例	134
2.4 箱图	68	5.4 非线性回归分析	137
2.5 正态概率分布图	72	5.4.1 非线性回归分析原理	137
2.6 时间序列图	80	5.4.2 非线性回归分析实例	138
2.7 小结	83		
2.8 习题	83		
第 3 章 假设检验	86		
3.1 假设检验简介	86		



5.5	小结	150	8.3	指数平滑法预测	207
5.6	习题	150	8.3.1	指数平滑法原理	208
<b>第 6 章</b>	<b>相关分析</b>	152	8.3.2	指数平滑法实例	208
6.1	相关分析简介	152	8.4	季节变动预测	210
6.2	双变量相关分析	153	8.4.1	季节变动预测原理	211
6.2.1	散点图图表法	153	8.4.2	季节变动预测实例	212
6.2.2	相关系数公式	158	8.5	小结	217
6.2.3	相关系数分析工具	160	8.6	习题	217
6.3	Spearman 秩相关	162	<b>第 9 章</b>	<b>马尔可夫链分析</b>	220
6.3.1	Spearman 秩相关简介	162	9.1	马尔可夫链简介	220
6.3.2	Spearman 秩相关实例分析	163	9.2	马尔可夫链分析实例	221
6.4	多重相关及偏相关分析	166	9.3	VBA 编程解决马尔可夫链分析	226
6.4.1	多重相关及偏相关分析简介	166	9.4	小结	230
6.4.2	多重相关与偏相关分析实例	166	9.5	习题	230
6.5	小结	170	<b>第 10 章</b>	<b>聚类分析</b>	231
6.6	习题	170	10.1	聚类分析原理简介	231
<b>第 7 章</b>	<b>判别分析</b>	172	10.2	水域污染情况聚类分析及 Excel 编程解决	234
7.1	判别分析简介	172	10.3	小结	243
7.2	判别分析实例	175	10.4	习题	243
7.3	Excel 操作及 VBA 编程解决	183	<b>第 11 章</b>	<b>主成分分析与因子分析</b>	245
7.4	小结	191	11.1	主成分分析	245
7.5	习题	191	11.1.1	主成分分析原理	245
<b>第 8 章</b>	<b>时间序列预测</b>	193	11.1.2	主成分分析实例	247
8.1	移动平均法预测	193	11.2	因子分析	254
8.1.1	移动平均法原理	193	11.2.1	因子分析原理	254
8.1.2	移动平均法实例	195	11.2.2	因子分析实例	256
8.2	回归趋势线预测	199	11.3	小结	259
8.2.1	趋势线预测方法原理	199	11.4	习题	259
8.2.2	趋势线预测方法实例	200	<b>参考文献</b>		261

# 第 1 章 描述性统计

描述性统计 (Descriptive Statistic) 是通过图表或数学方法, 对统计数据进行处理、分析, 并对数据的分布状态、数字特征和随机变量之间的关系进行估计和描述的方法。描述性统计的任务就是描述随机变量的统计规律。

要完整地描述随机变量的统计特性需要用分布函数, 但求随机变量的分布函数并不容易。实际上, 对于一些问题也不需要去全面考察随机变量的变化规律, 而只需知道随机变量的某些特征。

例如, 研究某一地区居民的消费水平, 只需知道该地区的平均消费水平即可; 但如检查一批灯泡的质量时, 则既需要注意灯泡的平均寿命, 又需要注意灯泡寿命与平均寿命的偏离程度。尽管这些数值不能完整地描述随机变量, 但能描述随机变量在某些方面的重要特征, 这些数字特征在理论和实践上都具有重要的意义。

因此, 在分析数据时, 一般首先要对数据进行描述性统计分析, 以发现其内在的规律, 再选择进一步分析的方法。在描述性统计中, 主要使用集中趋势、离散程度、偏度度量、峰度度量等方法来描述数据的集中性、分散性、对称性和尖端性, 以归纳数据的统计特性。常用的描述统计量有众数、中位数、算术平均数、调和平均数、几何平均数、四分位差、标准差、方差、变异系数等。

集中趋势测度: 算术平均值、几何平均数、调和平均数、中位数、众数。

离散程度测度: 极差 (全距)、标准差、方差、四分位差、变异系数。

数据分布测度: 偏度、峰度。

数值统计: 最小值、最大值、总和、总个数。

Excel 中用于计算描述统计量的方法有两种: 函数方法和描述统计工具的方法。本章将首先介绍列联表的使用以及数据的频数分析, 然后详细介绍如何使用 Excel 2007 中给出的统计函数来求解各种统计量和使用描述统计工具来实现对统计数据的描述性统计。

## 1.1 列 联 表

列联表是观测数据按两个或更多属性 (定性变量) 进行交叉分类时所列出的频数表。列联表分析常用来判断同一个调查对象的两个特性之间是否存在明显相关性。例如, 房地产商常常设计相关列联表问卷, 调查顾客的职业类型和顾客所选房子的户型是否有明显的相关性。同样, 列联表分析也可以在 Excel 2007 中实现。

一个实际频数  $f_{ij}$  的期望频数  $e_{ij}$ , 是总频数的个数  $n$  乘以该实际频数  $f_{ij}$  落入第  $i$  行和第  $j$  列的概率, 即



$$e_{ij} = n \cdot \left(\frac{r_i}{n}\right) \cdot \left(\frac{c_j}{n}\right) = \frac{r_i c_j}{n}$$

$\chi^2$  统计量的计算公式为

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

其自由度为  $(r-1)(c-1)$ 。 $\chi^2$  独立性检验可以检验列联表中行变量与列变量之间的相关性。根据显著性水平  $\alpha$  和自由度  $(r-1)(c-1)$  查出临界值  $\chi_\alpha^2$ ，若  $\chi^2 \geq \chi_\alpha^2$ ，则行变量与列变量之间是相关的；若  $\chi^2 < \chi_\alpha^2$ ，则行变量与列变量之间独立。

Excel 提供函数 CHITEST 计算  $\chi^2$  统计量的概率，提供函数 CHIINV 计算临界值  $\chi_\alpha^2$ 。

**函数语法：**CHITEST(actual\_range,expected\_range)

CHITEST 函数语法具有以下参数。

- actual\_range 为包含观察值的数据区域，将对期望值作检验。
- expected\_range 为包含行列汇总的乘积与总计值之比率的数据区域。

**函数语法：**CHIINV(probability,degrees\_freedom)

CHIINV 函数语法具有以下参数。

- probability 为与  $\chi^2$  分布相关的概率。
- degrees\_freedom 为自由度的数值。

## 例 1.1 顾客所在地区和所选房子地板类型之间的相关性分析

下面用一个具体例子说明列联表相关性分析。表 1.1 是某装修公司的调查报告数据表，用列联表分析方法分析顾客所在地区与所选房子地板类型之间是否存在明显的相关性。

表 1.1 某装修公司的调查报告数据表

	地区 1	地区 2	地区 3	地区 4	行总数
大理石地板	7	10	14	19	50
钢砖地板	26	10	16	33	85
木质地板	72	8	12	23	115
列总数	105	28	42	75	250

新建工作表“例 1.1 装修公司的调查报告数据.xlsx”，输入表 1.1 中的调查报告数据，如图 1.1 所示。

	地区 1	地区 2	地区 3	地区 4	行总数
大理石地板	7	10	14	19	50
钢砖地板	26	10	16	33	85
木质地板	72	8	12	23	115
列总数	105	28	42	75	250

图 1.1 装修公司的调查报告数据

下面使用 Excel 2007 进行相关性分析，具体操作步骤如下。

Step 01: 打开“例 1.1 装修公司的调查报告数据.xlsx”，如图 1.2 所示，先在 A8:F12 单元格范围建立期望频数表的框架。

	A	B	C	D	E	F
8	期望频数	地区1	地区2	地区3	地区4	行总数
9	大理石地板					
10	钢砖地板					
11	木质地板					
12	列总数					

图 1.2 装修公司的调查报告数据图

Step 02: 单击 B9 单元格，在编辑栏中输入公式“=B\$6\*\$F3/\$F\$6”，然后按回车键结束；再单击 B9 单元格，将鼠标指针移动至 B9 单元格右下角，当鼠标指针变为小黑色十字形状时按下鼠标左键拖曳至 E11 单元格，求出 B9:E11 各单元格值。

Step 03: 利用 Excel 的求和函数 SUM 计算行总数。单击 F9 单元格，在编辑栏中输入“=SUM(B9:E9)”，按回车键；再单击 F9 单元格，将鼠标指针移动至 F9 单元格右下角，当鼠标指针变为黑色十字形状时，按下鼠标左键拖曳至 F12 单元格，利用自动填充功能求出各行总数。

Step 04: 计算列总数。单击 B12，在编辑栏中输入“=SUM(B9:B12)”，按回车键；然后单击 B12 单元格，将鼠标指针移动至 B12 单元格右下角，当鼠标指针变为黑色十字形状时，按下鼠标左键并拖曳至 F12 单元格，求出各列总数，从而建立期望频数表，如图 1.3 所示。

	A	B	C	D	E	F
8	期望频数	地区1	地区2	地区3	地区4	行总数
9	大理石地板	21	5.6	8.4	15	50
10	钢砖地板	35.7	9.52	14.28	25.5	85
11	木质地板	48.3	12.88	19.32	34.5	115
12	列总数	105	28	42	75	250

图 1.3 期望频数

Step 05: 在 A14 单元格输入标志项“卡方概率值”，先点击 B14 单元格，单击菜单栏【公式】/【插入函数】命令，弹出【插入函数】对话框，在【或选择类别】一项选择【统计】；在【选择函数】中选择【CHITEST】函数，如图 1.4 所示。

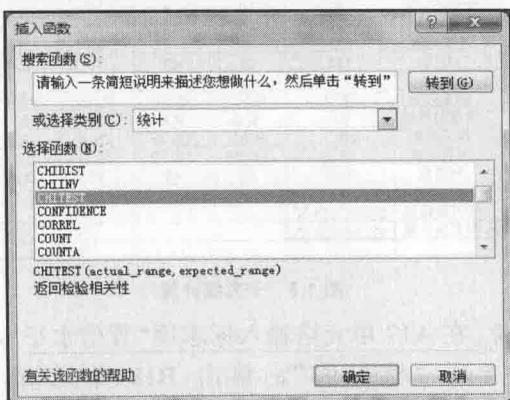


图 1.4 插入函数对话框

Step 06: 单击【插入函数】对话框【确定】按钮，弹出【函数参数】对话框；单击【Actual\_range】后的折叠按钮，选择 B3:E5 单元格区域；单击打开折叠按钮，返回【函数参数】对话框；单击【Expected\_range】后的折叠按钮，选择 B9:E11 单元格区域；单击打开折叠按钮，

返回【函数参数】对话框，如图 1.5 所示。最后单击【确定】按钮，即可得到卡方概率值 1.30821E-07，如图 1.6 所示。

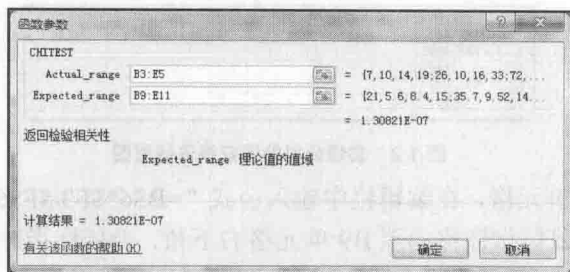


图 1.5 函数参数对话框

	A	B	C	D	E	F
1	列联表分析					
2	真实频数	地区1	地区2	地区3	地区4	行总数
3	大理石地板	7	10	14	19	50
4	钢砖地板	26	10	16	33	85
5	木质地板	72	8	12	23	115
6	列总数	105	28	42	75	250
7						
8	期望频数	地区1	地区2	地区3	地区4	行总数
9	大理石地板	21	5.6	8.4	15	50
10	钢砖地板	35.7	9.52	14.28	25.5	85
11	木质地板	48.3	12.88	19.32	34.5	115
12	列总数	105	28	42	75	250
13						
14	卡方概率值	1.30821E-07				
15						

图 1.6 卡方概率值计算结果

Step 07: 求  $\chi^2$  统计量。在 A15 单元格输入标志项“卡方统计量”，单击 B15 单元格，在编辑栏中输入公式“=SUM((B3:E5-B9:E11)^2/B9:E11)”，完成后按 Ctrl+Shift+Enter 组合键结束，结果如图 1.7 所示。

	A	B	C	D	E	F
1	列联表分析					
2	真实频数	地区1	地区2	地区3	地区4	行总数
3	大理石地板	7	10	14	19	50
4	钢砖地板	26	10	16	33	85
5	木质地板	72	8	12	23	115
6	列总数	105	28	42	75	250
7						
8	期望频数	地区1	地区2	地区3	地区4	行总数
9	大理石地板	21	5.6	8.4	15	50
10	钢砖地板	35.7	9.52	14.28	25.5	85
11	木质地板	48.3	12.88	19.32	34.5	115
12	列总数	105	28	42	75	250
13						
14	卡方概率值	1.30821E-07				
15	卡方统计量	42.74819145				
16						

图 1.7 卡方统计量

Step 08: 进行假设检验。在 A17 单元格输入标志项“置信水平”，在 B17 单元格输入 0.01；在 A18 单元格输入标志项“临界值”，单击 B18 单元格，在编辑栏中输入公式“=CHIINV(B17,6)”，按回车键；在 A19 单元格输入标志项“检验结果”，单击 B19 单元格，在编辑栏中输入公式“=IF(B15>B18,"拒绝两种属性不相关的假设","接受两种属性不相关的假设)”，按回车键。结果如图 1.8 所示。

【注意】CHIINV 函数的自由度=(第一类属性的分类数-1)×(第二类属性的分类数-1)，即  $(r-1)(c-1)=(3-1) \times (4-1)=6$ 。

	A	B	C	D	E	F
1	列联表分析					
2	真实频数	地区1	地区2	地区3	地区4	行总数
3	大理石地板	7	10	14	19	50
4	钢砖地板	26	10	16	33	85
5	木质地板	72	8	12	23	115
6	列总数	105	28	42	75	250
7						
8	期望频数	地区1	地区2	地区3	地区4	行总数
9	大理石地板	21	5.6	8.4	15	50
10	钢砖地板	35.7	9.52	14.28	25.5	85
11	木质地板	48.3	12.88	19.32	34.5	115
12	列总数	105	28	42	75	250
13						
14	卡方概率值	1.30821E-07				
15	卡方统计量	42.74819145				
16						
17	置信水平	0.01				
18	临界值	16.81189383				
19	检验结果	拒绝两种属性不相关的假设				
20						

图 1.8 列联表分析结果

**【结论】**

以上的操作步骤即完成对整个列联表的分析,从图 1.8 所示中可以看出, B14 单元格的卡方概率值与 B15 单元格的卡方统计量是表格的两个重要计算结果。其中卡方概率值等于 1.30821E-07,表明如果总体的两类属性,即所在地区和所选地板类型,是不相关的,那么得到以上观察的样本的概率是 0.000000130821,这个概率几乎等于 0,所以可以认为总体的这两个属性是显著相关的。

## 1.2 数据频数分析

频数也称“次数”,是对总数据按某种标准进行分组,统计出各个组内含个体的个数。通过统计调查得到的数据往往是杂乱,没有规则的,因此,必须对得到的大量原始数据进行加工整理,经过数据分析得出科学结论。对于一组数据,考察不同数值出现的频数,或者是数据落入指定区域内的频数,可以了解数据的分布状况。在 Excel 2007 中,数据频数分析主要通过频数分布函数与直方图分析工具等来进行。

### 1.2.1 频数分布函数

#### 例 1.2 居民购买消费品支出频数分析

某县城统计部门抽样调查 50 户居民购买消费品支出,支出资料如图 1.9 所示(单位:元)。

	A	B	C	D	E
1	880	1180	1100	860	1420
2	1210	1260	1180	1250	1510
3	1050	1230	1420	1510	1180
4	1630	1250	1410	1420	1200
5	1170	1100	1510	1160	1070
6	1010	1210	1140	1030	1270
7	1190	1370	1420	1230	1370
8	1080	1070	1160	830	960
9	1320	1460	930	1190	810
10	1050	1230	860	1340	870
11					

图 1.9 居民购买消费品月支出

对其按 800~900、900~1000、1000~1100、1100~1200、1200~1300、1300~1400、1400~1500、1500~1600、1600 以上分为 9 个组进行频数分析。

Excel 提供了一个专门用于统计分组的频数分布函数 FREQUENCY，它以一列垂直数组返回某个区域中的数据分布，描述数据分布状态。

### 函数语法：FREQUENCY(data\_array,bins\_array)

FREQUENCY 函数语法具有以下参数。

- data\_array 是一个数组或对一组数值的引用，要为数组或数值计算频率。如果 data\_array 中不包含任何数值，函数 FREQUENCY 将返回一个零数组。
- bins\_array 是一个区间数组或对区间的引用，该区间用于对 data\_array 中的数值进行分组。如果 bins\_array 中不包含任何数值，函数 FREQUENCY 返回的值与 data\_array 中的元素个数相等。

【注意】在选择了用于显示返回分布结果的相邻单元格区域后，函数 FREQUENCY 应以数组公式的形式输入。返回数组中的元素个数比 bins\_array 中的元素个数多 1 个。多出来的元素表示最高区间之上的数值个数。函数 FREQUENCY 将忽略空白单元格和文本。

在使用此函数时，先将样本数据排成一列。新建工作表“例 1.2 居民购买消费品支出资料.xlsx”。将图 1.9 所示的数据排成一列，本例中为 A1:A50。

利用频数分布函数进行统计分组和计算频数，具体操作步骤如下。

Step 01: 打开工作表“居民购买消费品支出资料.xlsx”。选定单元格区域，本例中选定的区域为 D1:D9，单击菜单栏【公式】/【函数库】/【插入函数】命令，弹出【插入函数】对话框，如图 1.10 所示。

Step 02: 在【选择类别】中选择【统计】，在【选择函数】中选择【FREQUENCY】，如图 1.11 所示。

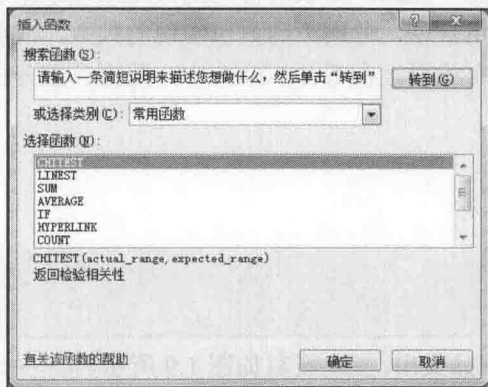


图 1.10 插入函数对话框

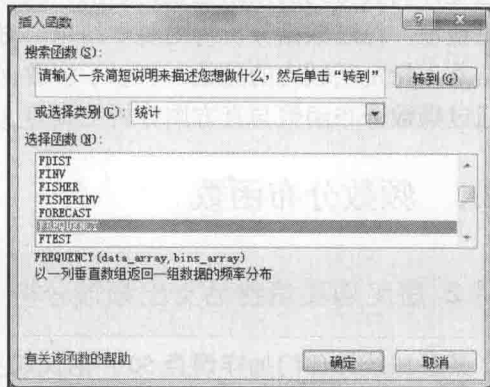


图 1.11 选择“FREQUENCY”对话框

Step 03: 单击图 1.11 中【确定】按钮，弹出 FREQUENCY【函数参数】对话框；单击【Data\_array】后的折叠按钮，选择 A1:A50 单元格区域；单击打开折叠按钮，返回【函数参数】对话框；在【Bins\_array】栏中填写“{899;999;1099;1199;1299;1399;1499;1599;1699}”，如图 1.12 所示。

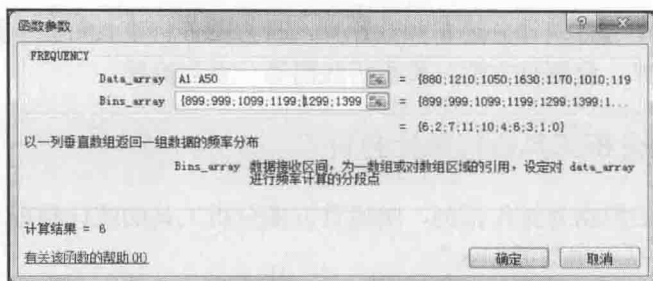


图 1.12 FREQUENCY【函数参数】对话框

【注意】data\_array 用于计算频率的数组，或对数组单元区域的引用。本例中为 A1:A50。bins\_array 是数据接受区间，为一组数或对数组区间的引用，设定对 data\_array 进行频率计算的分段点。本例中为 899、999、1099、1199、1299、1399、1499、1599、1699。频数分布函数要求按组距的上限分组，不接受非数值的分组标志（如“不足××”或“××以上”等）。在输入的数据两端必须加大括号，各数据之间用分号隔开。输入完成后，由于频数分布是数组操作，所以不能单击【确定】按钮。

Step 04: 按“Ctrl+Shift+Enter”组合键，在最初选定单元格区域 D1:D9 内得到频数分布结果，如图 1.13 所示。至此，频数分布函数进行统计分析的功能就全部操作完成了。

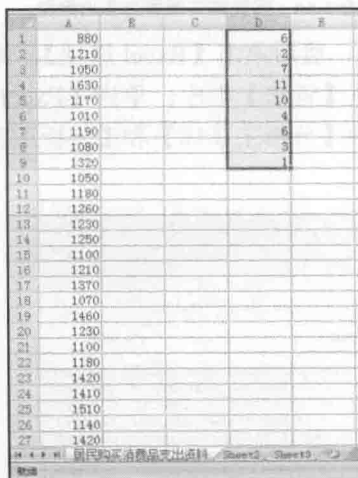


图 1.13 频数分布结果

### 【结论】

通过以上分析步骤可看出，使用 Excel 2007 中提供的 FREQUENCY 函数，可以很方便地求出频数分布。从图 1.13 所示的结果中，可以看到居民购买消费品的花销在不同分段点内的分布状况。

## 1.2.2 直方图分析工具

直方图是将所收集的测定值、特性值或结果值，分为几个相等的区间作为横轴，并将各区间内所测定值依据所出现的次数累积而成的面积，用“柱子”排起来的图形。因此，也叫作柱形图。



频数分布函数只能进行统计分组和频数计算，直方图分析工具可完成数据的分组、频数分布与累积频数的计算、绘制直方图与累积折线图等一系列操作。

### 例 1.3 利用直方图分析工具进行统计分析

这里仍以例 1.2 的数据为操作范例，阐述直方图分析工具的统计整理功能，新建工作表“例 1.3 居民购买消费品支出数据.xlsx”。

利用直方图分析工具进行分析，具体操作步骤如下。

Step 01: 在 G1 单元格输入分组标志，在 G2:G10 单元格区域输入 899、999、1099、1199、1299、1399、1499、1599、1699，如图 1.14 所示。

	A	B	C	D	E	F	G
1		消费品支出					分组
2		880	1180	1100	860	1420	899
3		1210	1260	1180	1250	1510	999
4		1050	1230	1420	1510	1180	1099
5		1630	1250	1410	1420	1200	1199
6		1170	1100	1510	1160	1070	1299
7		1010	1210	1140	1030	1270	1399
8		1190	1370	1420	1230	1370	1499
9		1080	1070	1160	830	960	1599
10		1320	1460	930	1190	810	1699
11		1050	1230	860	1340	870	
12							

图 1.14 居民购买消费品支出数据


Step 02: 单击 Office 按钮 ，然后单击【Excel 选项】。弹出【Excel 选项】对话框，单击【加载项】，如图 1.15 所示。在【管理】框中，单击【Excel 加载项】，然后单击【转到】，弹出【加载宏】对话框；单击选中【分析工具库】和【分析工具库-VBA】复选框，如图 1.16 所示，完成后单击【确定】按钮。



图 1.15 Excel 选项对话框

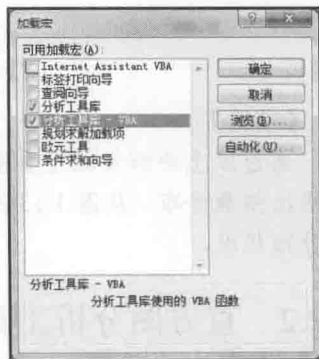






图 1.16 加载宏对话框

【注意】一般在安装 Excel 后都有加载项，如果在【Excel 加载项】中没有需要的加载项，则可能需要安装该加载项。若要安装通常随 Excel 一起安装的加载项（例如规划求解或分析

工具库), 请运行 Excel 或 Microsoft Office 的安装程序, 并选择【更改】选项以安装加载项。重新启动 Excel 之后, 加载项显示在【可用加载项】框中。

☞ Step 03: 单击菜单栏【数据】/【分析】/【数据分析】命令, 弹出【数据分析】对话框, 从【分析工具】列表框中选择【直方图】选项, 如图 1.17 所示。

☞ Step 04: 单击【确定】按钮, 弹出【直方图】对话框(如图 1.18 所示); 单击【输入区域】后的折叠按钮 , 将对话框折叠, 选择 A2:E11 单元格区域; 单击打开折叠按钮 , 返回【直方图】对话框; 单击【接收区域】后的折叠按钮 , 将对话框折叠, 选择接收区域对应的 G2:G10 单元格; 单击打开折叠按钮 , 返回【直方图】对话框。

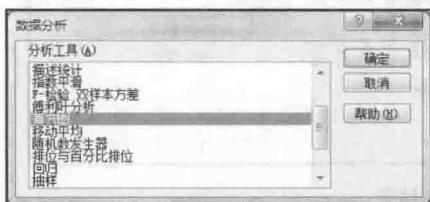


图 1.17 数据分析对话框

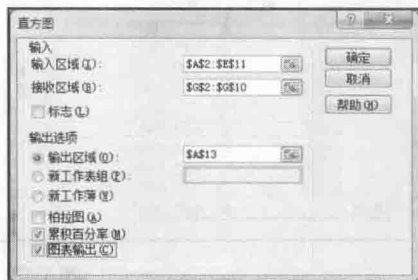




图 1.18 直方图对话框

☞ Step 05: 单击【输出区域】单选按钮, 单击【输出区域】后的折叠按钮 , 选择 A13 单元格区域, 单击打开折叠按钮 , 返回【直方图】对话框; 单击选中【累积百分率】和【图表输出】复选框, 如图 1.18 所示。

### 说明: 【直方图】对话框。

【输入区域】: 输入待分析数据区域的单元格引用, 本例中输入区域为 \$A\$2:\$E\$11。

【接收区域】: 输入接收区域的单元格引用, 该框为空, 则系统自动利用输入区域中的最小值和最大值建立平均分布的区间间隔的分组, 本例中接收区域为 \$G\$2:\$G\$10。

【标志】: 若输入区域有标志项, 则选中【标志】复选框, 否则, 系统自动生成数据标志。

【输出区域】: 在此选项中可选择输出去向, 输入在对输出表左上角单元格的引用。本例中选择【输出区域】为 \$A\$13。

【新工作表组】: 单击此选项可在当前工作簿中插入新工作表, 并从新工作表的 A1 单元格开始粘贴计算结果。若要为新工作表命名, 请在其右侧的框中输入名称。

【新工作簿】: 单击此选项, 可创建一新工作簿, 并在新工作簿的新工作表中粘贴计算结果。

【柏拉图】: 可以在输出表中同时按降序排列频数数据。

【累积百分率】: 可在输出表中增加一系列累积百分比数值, 并绘制一条累积百分比曲线。

【图表输出】: 可生成一个嵌入式直方图。

☞ Step 06: 单击【确定】按钮, 在输出区域单元格可得到频数分布, 如图 1.19 所示。

☞ Step 07: 左键单击条形图的任一直条, 再单击右键, 在快捷菜单中选取【设置数据系列格式】, 弹出【设置数据系列格式】对话框; 在【系列选项】中将【分类间距】滑块拖动到最左边, 即复选框中显示“0%”, 如图 1.20 所示, 单击【关闭】按钮即可, 从而将条形图转换

成标准直方图，如图 1.21 所示。



图 1.19 直方图频数分布结果

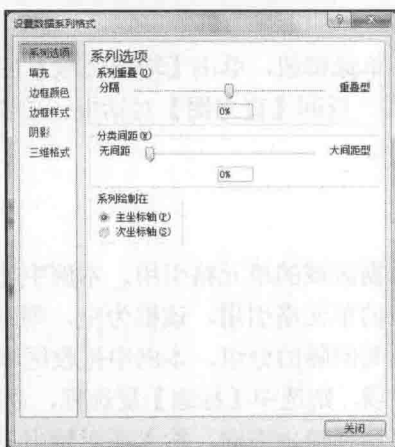


图 1.20 设置数据系列格式对话框

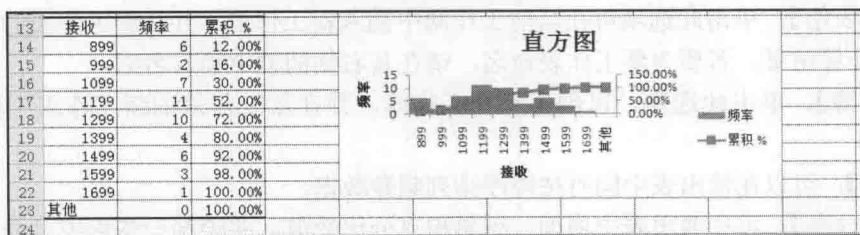


图 1.21 标准直方图

### 【结论】

从图 1.21 所示中可以观察不同分组段的频率，简洁直观。直方图分析工具不仅仅能计算出各组频率大小，还可以绘制出直方图和累积折线图，使数据的频数分析更加方便直观，一目了然。