

行业畅销书升级版

← 告诉你职场图表背后的故事 →

“谁说菜鸟不会 数据分析”

张文霖 刘夏璐 狄松 编著

入门篇



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>



专家评论

当谈到用数据解决问题时，我经常用这样的语言去诠释：“如果你不能量化它，你就不能理解它，如果不理解就不能控制它，不能控制也就不能改变它”。数据无处不在，信息时代的最主要特征就是“数据处理”，数据分析正以我们从未想象过的方式影响着日常生活。

在知识经济与信息技术时代，每个人都面临着如何有效地吸收、理解和利用信息的挑战。那些能够有效利用工具从数据中提炼信息、发现知识的人，最终往往成为各行各业的强者！

这本书向我们清晰又友好地介绍了数据分析方法、技巧与工具，强烈推荐读一读这本书，它或许会给你带来意想不到的收获！

沈浩

中国传媒大学电视与新闻学院，教授；调查统计研究所，副所长
数据挖掘研发中心，主任；GIPSOS公司，首席技术顾问

统计学是一门很难，但是很有趣，更是很有用的工具学科。懂得如何使用它的人总是乐在其中，而尚未入门的人则畏之如虎。国内讲述统计学理论，以及讲述统计软件操作的书籍可谓汗牛充栋，但是多数流于理论，疏于应用和实践指导。很大一部分读者需求未被满足。

近年来随着信息技术的普及，各行各业的业务数据自动化趋势愈来愈明显，使得数据分析的需求开始从统计专业人士向各行业人员全面扩展。在此背景之下，出版一本能够深入浅出，从实际应用的角度介绍统计分析基础知识的书就变得很有必要。

这本书在理论和实践的平衡方面做了很有价值的尝试，基于最为普及的Excel、5W2H、PEST等数据分析方法论，深入浅出地介绍了如何满足具体工作中的常见统计分析需求，对于需要应用统计分析，但是又未接受过这方面系统培训的读者来说，本书应当是一本非常合适的数据分析入门教材。

张文彤 博士

英德知联恒市场咨询（上海）有限公司 新事业开发部，全国技术总监

新浪微博：

@博文视点Broadview



策划编辑：张月萍

责任编辑：许艳

封面设计：李玲

上架建议：计算机>办公软件

ISBN 978-7-121-18780-3



9 787121 187803 >

定价：49.00元

谁说菜鸟不会 数据分析

张文霖 刘夏璐 狄松 编著

(入门篇)



数据分析



NLIC2970848131

电子工业出版社
Publishing House of Electronics Industry
北京•BEIJING

内容简介

这是一本有趣的数据分析书！

本书基于通用的Excel工具，加上必知必会的数据分析概念，以小说般通俗易懂的方式讲解。

本书基于职场三人行来构建内容，完全按照数据分析工作的完整流程来讲解。全书共8章，依次讲解数据分析必知必会知识、确定数据分析的结构化思维、数据处理技巧、数据展现的技术、通过专业化的视角来提升图表之美以及专业分析报告的撰写等内容。

本书有足够的魅力让你一口气读下去，在无形之中掌握数据分析的技能，提升职场竞争能力。

本书能有效帮助职场新人提升职场竞争力，也能帮助市场营销、金融、财务、人力资源管理人员及产品经理解决实际问题，还能帮助从事咨询、研究、分析行业的人士，各级管理人士提高专业水平。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目（CIP）数据

谁说菜鸟不会数据分析·入门篇 / 张文霖, 刘夏璐, 狄松编著. —北京：电子工业出版社，2013.1

ISBN 978-7-121-18780-3

I. ①谁… II. ①张… ②刘… ③狄… III. ①表处理软件 IV. ①TP391.13

中国版本图书馆CIP数据核字（2012）第250460号

策划编辑：张月萍

责任编辑：许 艳

印 刷：中国电影出版社印刷厂

装 订：三河市皇庄路通装订厂

出版发行：电子工业出版社

北京市海淀区万寿路173信箱 邮编：100036

开 本：787×980 1/16 印张：15.5 字数：366千字

印 次：2013年1月第2次印刷

印 数：15001~25000册 定价：49.00元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至z1ts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

服务热线：（010）88258888。

前　　言

经常有朋友询问：数据分析该怎么做？有什么分析技巧？这些数据怎么处理分析？

因为大量问题具有通用性，而且“懒”得挨个答复类似的问题，于是就结合大家关心的问题，编写这本通俗易懂的数据分析书。市面上数据分析的书大部分还停留在大雅的范畴，要么就是高深的统计学理论，要么就是专业的统计分析软件，给人感觉门槛非常高。而且，所讲解的案例大部分来自科研一线，让人看了摸不着北。这无形间在学习者与数据分析之间建起了鸿沟。

其实，通过多年的数据分析实践来看，数据分析还是件很有乐趣的事情。我们需要做的是：基于通用的工具Excel，加上必知必会的数据分析概念，采用通俗易懂的讲解方式。这样数据分析就不那么晦涩了，而且故事化的情境设计，让我们有一口气读下去的勇气，天堑也变通途了。

虽然积累了多年的数据分析实战经验，但是要上升到一本书还是花费了近1年的时间。她的第1章、第8章由狄松完成，第2章、第5章、第6章由张文霖完成，第3章、第4章、第7章由刘夏璐完成。这个创作过程是艰辛的，但也是很有成就感的。我们努力讲好数据分析的故事，同时把这个故事尽量展现得美丽动人。请允许我们以“她”来称呼这本与众不同的数据分析书籍，很多人翻开这本书的时候，可能会有大量疑惑，但，请耐着性子慢慢读下去，您将会有莫大的收获。

如果你觉得她看起来很轻松，千万别误以为她是一本小说，她其实是一本讲述数据分析的书籍

她抛开复杂的数学或者统计学原理，只和你讲必知必会的要点，关注解决实际问题；

她不去探究科班的学术问题，只和你耐心地分享职场中的实战案例；

她不板起脸和你讲大道理，只和你娓娓道来切身的趣味故事；

她天生丽质，图表漂亮绝伦；

她多姿多彩，还有卡通漫画风。

可能你会觉得她肤浅……

但是，当你揭开她华丽的外衣时，你会惊艳；

也会被她通俗而不庸俗，美丽而又深刻的本质所吸引。

把她珍藏起来吧，因为：

她会循循善诱地把你领进数据分析的大门；

她会让你的简历更加具有吸引力；

她会让老板对你刮目相看；

她值得在你的书架上长期逗留，会让你的书架也增加色彩。

她讲述了职场三人行的故事，她的故事还会让你偷着笑

牛董，关键词：私企董事、要求严格、为人苛刻。

小白，关键词：应届毕业生，刚入职场的伪白骨精（白领+骨干+精英）、牛董助手、单身

>> 谁说菜鸟不会数据分析（入门篇）

女白领、爱臆想。

Mr.林，关键词：小白的同事、数据分析达人、成熟男士、乐于助人、做事严谨。

哪些人会对她的故事有阅读兴趣呢

- ★ 需要提升自身竞争力的职场新人。
- ★ 在市场营销、金融、财务、人力资源、产品设计等工作中需要做数据分析的人士。
- ★ 经常阅读经营分析、市场研究报告的各级管理人员。
- ★ 从事咨询、研究、分析等工作的专业人士。

其实她还有效率助手

她的效率助手是一些常用的插件工具，此外她的配书数据文件可以到www.dozan.cn或<http://blog.sina.com.cn/xiaowenzi22>下载。拥有了这些实用的插件，可以让我们的数据分析如鱼得水。

致谢

感谢笔者的好朋友李治的鼓励和支持，让笔者下定决心写这本书。在此要衷心感谢成都道然科技有限责任公司的姚新军先生，感谢他的提议和在写作过程中的支持。感谢参与本书优化的朋友：王斌、李伟、张强林、万雷、李平、王晓、景小燕、余松。非常感谢本书的插画师王馨的辛苦劳动，您的作品也让本书增色了不少。

感谢邓凯、黄成明、李双、刘晓霞、刘云锋、欧维平、石军、沈浩、张文彤、张立良、张志成、郑来轶、祝迎春、王雍、伍昊等书评作者，感谢他们在百忙之中抽空阅读书稿，撰写书评，并提出宝贵意见。

最后，要感谢三位作者的家人，感谢他们默默无闻的付出，没有他们的理解与支持，同样也没有本书。

尽管我们对书稿进行了多次修改，仍然不可避免地会有疏漏和不足之处，敬请广大读者批评指正，我们会在适当的时间进行修订，以满足更多人的需要。

升级版说明

《谁说菜鸟不会数据分析》自2011年7月出版以来得到广大读者朋友的大力支持，而且很荣幸获得中国书刊发行业协会颁发的“2011年度全行业优秀畅销书品种”称号。这个荣誉的取得与广大读者的大力支持是分不开的。为了让这本书精益求精，在征集了大量的读者反馈意见后，我们进行了本次的升级。

与故事的作者联系：

博客：<http://blog.sina.com.cn/xiaowenzi22>

邮箱：xiaowenzi22@qq.com

新浪微博：[@小蚊子乐园](#)

与故事的策划者联系：

邮箱：info@dozan.cn

新浪微博：[@长颈鹿27](#)

业内人士的推荐（排名不分先后，以姓氏拼音排序）

本书将看似“浮云”的数据分析知识，蕴于商业化的场景之中，生动形象地让读者了解到“给力”的数据分析师是如何炼成的！本书引导非专业人士从数据的角度，认识、剖析、解决商业问题；对专业人士而言，亦能提供一次梳理知识和提高技能的机会。

邓凯

数据挖掘与数据分析博主，资深数据分析师

这是一本适合普通大众的“专业”数据分析图书，由浅入深，富有体系。让人有一口气读完的冲动，想马上找一台电脑试一试这些“新奇”的分析方法，更想拿一些数据来分析其中的规律。

读完本书，你会发现数据分析的乐趣，它并不是那么枯燥，数据背后的故事简直太有意思了。从此你将发现：无论新闻媒体，还是企业报表中的数字都有生命，因为它们在那里用自己的语言和你交流！

祝愿大家早日练就一颗数据分析的“心”！

黄成明

数据化管理顾问及培训师，零售及服装企业数据化管理咨询顾问

本书内容实用，语言简洁生动，通俗易懂。通过富有逻辑的路径式图示引导，对复杂步骤的图文分解，让读者快速掌握用Excel进行数据分析的各项实用技巧，令人耳目一新。不仅便于学习，也便于上机操作。

李双

数据分析与挖掘交流站，站长

数据分析的门槛有多高？可以很高，也可以不高。小蚊子的这本书可以给初学者一些实战性的指引。书中介绍的多个常用数据分析方法，对于初学者甚至是具有一定经验的数据分析师都有很好的启发。

刘晓霞

慧聪邓白氏研究e-Eyes事业部，副总经理

>> 谁说菜鸟不会数据分析（入门篇）

本书是市面上少见的一本系统讲解数据分析的书籍。它没有针对高级分析方法和统计函数的介绍，而是针对职场人士日常工作中遇到的问题提出解决方案。书中通过小白跟师父学艺的整个过程，生动形象地描述和解释了什么是数据分析以及如何有效地进行数据分析，通俗易懂，趣味性非常强，是一本非常适合初学者的书籍。

刘云峰

北京简博市场咨询有限公司，高级经理

本书有四大亮点。其一，抛开了烦琐的统计公式与数理推论，完全以实践应用为导向，十分适合于没有统计背景的普通白领；其二，本书基本是小蚊子实际工作经验的总结，与大部分作者言之不尽的陋习不同的是，小蚊子在著作中对自己掌握的知识倾囊相授，这也是小蚊子一贯的品性；其三，作品除了教会你使用Excel简单处理日常工作涉及的数据分析，还在统计分析图表的展示方面为读者提供了重要的指导；其四，本书写作方式有新意，如小说一般的写作，使枯燥的数据分析过程兴趣盎然。

如果你正在为复杂的统计公式而头痛，如果你正在为学习统计软件而烦恼，如果你正在为如何将数据分析的结果进行专业的呈现而伤神，那么，选择这本书，也许你就找到了终南捷径。

欧维平

广州市准诚市场咨询有限公司，研究总监

数据分析理论、公式和方法对部分初学者来说是枯燥、乏味的，或坠入云山雾海中不得其道。本书最大的特点是使用幽默风趣的语言，结合工作中典型案例加以分析、解读，是一本值得数据分析工作者一读的好书。

石军

安徽同微信息技术有限公司，总经理

当谈到用数据解决问题时，我经常用这样的语言去诠释：“如果你不能量化它，你就不能理解它，如果不理解就不能控制它，不能控制也就不能改变它”。数据无处不在，信息时代的最主要特征就是“数据处理”，数据分析正以我们从未想象过的方式影响着日常生活。

在知识经济与信息技术时代，每个人都面临着如何有效地吸收、理解和利用信息的挑战。那些能够有效利用工具从数据中提炼信息、发现知识的人，最终往往成为各行各业的强者！

这本书向我们清晰又友好地介绍了数据分析方法、技巧与工具，强烈推荐读一读这本书，它或许会给你带来意想不到的收获！

沈浩

中国传媒大学电视与新闻学院，教授

调查统计研究所，副所长

数据挖掘研发中心，主任

IPSOS公司，首席技术顾问

对我们财务人员来说，每个月都要写经济活动分析，但总是列出一堆干巴巴的数字，领导不爱看，自己也脸上无光，而这本书却能改变这一切。不懂数据库？不会Excel？不知如何做图表？没关系，这本书充分考虑了初学者的知识背景，让你从入门到精通。更难能可贵的是，本书设计的场景风趣幽默，让人仿佛是在看小说，把枯燥无味的学习变得生动有趣。

张立良

Excel必备工具箱，开发者

统计学是一门很难，但是很有趣，更是很有用的工具学科。懂得如何使用它的人总是乐在其中，而尚未入门的人则畏之如虎。国内讲述统计学理论，以及讲述统计软件操作的书籍可谓汗牛充栋，但是多数流于理论，疏于应用和实践指导。很大一部分读者需求未被满足。

近年来随着信息技术的普及，各行各业的业务数据自动化趋势愈来愈明显，使得数据分析的需求开始从统计专业人士向各行业人员全面扩展。在此背景之下，出版一本能够深入浅出，从实际应用的角度介绍统计分析基础知识的书就变得很有必要。

本书在理论和实践的平衡方面做了很有价值的尝试，基于最为普及的Excel、SWOT、PEST等数据分析方法论，深入浅出地介绍了如何满足具体工作中的常见统计分析需求，对于需要应用统计分析，但是又未接受过这方面系统培训的读者来说，本书应当是一本非常合适的数据分析入门教材。

张文彤博士

英德知联恒市场咨询（上海）有限公司 新事业部，全国技术总监

这是一本真正介绍数据分析而不是介绍数据分析工具的书，全书内容就是按照数据分析流程而组织的，每一章节均通过案例来讲解，语言生动有趣。更加可贵的是，案例的“剧情”大多都

>> 谁说菜鸟不会数据分析（入门篇）

是作者多年来在现实中遇到的，因此实用性较高！在宏观结构上采用的经典结构能够带领读者按照正确的顺序稳步前进，在微观上采用的叙述方式极大增强了“渴”读性，使得应用技术更加直观。本书还是一本经典案例大全，内容涵盖人力资源等各方面的应用。因此，本书适合所有工作中需要做数据分析的朋友阅读！

张志威

<http://blog.soufun.com/site>, 选址分析师

数据分析是一种能力，更是一种思想。本书结构有层次、内容全面、通俗易懂，一步步带你走进数据分析的世界，让数据分析变得既简单又有趣。

郑来扶

中国统计网站长，资深数据分析师

这是一本简单易懂，但又不缺乏深度的数据分析图书。该书作者常年从事数据分析工作，熟悉日常数据分析工作常见的问题和解决方案，所以该书积累了大量数据分析的实用性方法与技巧，需要细致的深入其境的学习，最好直接跟着书中内容实际操作，边做边学边领悟，这样可以达到事半功倍的效果。数据分析需要不断在工作中实践，这是一本入门性的书籍，最终还是需要读者靠自己的意志力克服畏难情绪去学习。付出才会有收获，学习任何东西都是如此。

祝迎春

高等学校教材《SPSS统计分析高级教程》，合作者

数据分析圈的朋友应该都知道小蚊子或小蚊子乐园这个博客，本书是小蚊子同学多年数据分析工作的积累总结，是一本简单实用的书，是数据分析技巧与工作实例充分结合的一本书。全书通过幽默的对话勾勒出数据分析的全景，彻底打开了以往数据分析相关专业书籍单一枯燥的局面。

王雍

数据元素博主，资深数据分析师

你们想想，带着这本书出了城，吃着火锅还唱着歌，突然就学会数据分析了……

伍昊

五号咨询，Excel首席培训师

目 录

第1章 数据分析那些事儿

- 1.1 数据分析是“神马” /14
 - 1.1.1 何谓数据分析 /15
 - 1.1.2 数据分析的作用 /16
- 1.2 数据分析六步曲 /17
 - 1.2.1 明确分析目的和思路 /18
 - 1.2.2 数据收集 /20
 - 1.2.3 数据处理 /21
 - 1.2.4 数据分析 /21
 - 1.2.5 数据展现 /22
 - 1.2.6 报告撰写 /23
- 1.3 数据分析的三大误区 /24
- 1.4 数据分析师的职业发展 /25
 - 1.4.1 数据分析的广阔前景 /25
 - 1.4.2 数据分析师的职业要求 /27
 - 1.4.3 数据分析师的基本素质 /28
- 1.5 几个常用指标和术语 /32
- 1.6 本章小结 /36

第2章 结构为王——确定分析思路

- 2.1 数据分析方法论 /38
 - 2.1.1 数据分析方法论与数据分析法的区别 /38
 - 2.1.2 数据分析方法论的重要性 /39
- 2.2 常用的数据分析方法论 /40
 - 2.2.1 PEST分析法 /40
 - 2.2.2 5W2H分析法 /43
 - 2.2.3 逻辑树分析法 /44
 - 2.2.4 4P营销理论 /45
 - 2.2.5 用户行为理论 /47
- 2.3 本章小结 /48

第3章 无米难为巧妇——数据准备

- 3.1 理解数据 /50

>> 谁说菜鸟不会数据分析（入门篇）

- 3.1.1 字段与记录 /51
- 3.1.2 数据类型 /52
- 3.1.3 数据表 /53
- 3.2 数据来源 /57**
 - 3.2.1 导入数据 /57
 - 3.2.2 问卷录入要求 /63
- 3.3 本章小结 /65**

第4章 三心二意——数据处理

- 4.1 何为数据处理 /67**
 - 4.1.1 “三心二意”处理数据 /67
 - 4.1.2 数据处理的内容 /69
- 4.2 数据清洗 /70**
 - 4.2.1 重复数据的处理 /71
 - 4.2.2 缺失数据处理 /76
 - 4.2.3 检查数据逻辑错误 /80
- 4.3 数据加工 /82**
 - 4.3.1 数据抽取 /82
 - 4.3.2 数据计算 /87
 - 4.3.3 数据分组 /91
 - 4.3.4 数据转换 /92
- 4.4 数据抽样 /97**
- 4.5 本章小结 /98**

第5章 工欲善其事必先利其器——数据分析

- 5.1 数据分析方法 /101**
 - 5.1.1 对比分析法 /102
 - 5.1.2 分组分析法 /105
 - 5.1.3 结构分析法 /106
 - 5.1.4 平均分析法 /107
 - 5.1.5 交叉分析法 /108
 - 5.1.6 综合评价分析法 /109
 - 5.1.7 杜邦分析法 /113
 - 5.1.8 漏斗图分析法 /114
 - 5.1.9 矩阵关联分析法 /115
 - 5.1.10 高级数据分析方法 /120
- 5.2 数据分析工具 /121**
 - 5.2.1 初识数据透视表 /121

- 5.2.2 创建数据透视表的三步法 /122
- 5.2.3 数据透视表分析实践 /124
- 5.2.4 数据透视表小技巧 /130
- 5.2.5 多选题分析 /132

5.3 本章小结 /135

第6章 给数据量体裁衣——数据展现

- 6.1 揭开图表的真面目 /138
 - 6.1.1 图表的作用 /138
 - 6.1.2 经济适用图表有哪些 /139
 - 6.1.3 通过关系选择图表 /140
 - 6.1.4 图表制作五步法 /145
- 6.2 表格也疯狂 /146
 - 6.2.1 突出显示单元格 /146
 - 6.2.2 项目选取 /147
 - 6.2.3 数据条 /148
 - 6.2.4 图标集 /149
 - 6.2.5 迷你图 /150
- 6.3 给图表换装 /151
 - 6.3.1 平均线图 /152
 - 6.3.2 双坐标图 /153
 - 6.3.3 竖形折线图 /156
 - 6.3.4 瀑布图 /159
 - 6.3.5 帕累托图 /160
 - 6.3.6 旋风图 /165
 - 6.3.7 人口金字塔图 /169
 - 6.3.8 漏斗图 /171
 - 6.3.9 矩阵图（散点图） /173
 - 6.3.10 发展矩阵图 /176
 - 6.3.11 改进难易矩阵（气泡图） /178
- 6.4 本章小结 /180

第7章 专业化生存——图表可以更美的

- 7.1 别让图表犯错 /183
 - 7.1.1 让图表“五脏俱全” /183
 - 7.1.2 要注意的条条框框 /185
 - 7.1.3 图表会说谎 /196
- 7.2 浓妆淡抹总相宜——图表美化 /200

>> 谁说菜鸟不会数据分析（入门篇）

- 7.2.1 图表美化的三原则 /200
- 7.2.2 略施粉黛——美化图表的技巧 /203
- 7.2.3 图表也好“色” /209
- 7.3 如虎添翼的招儿 /213**
 - 7.3.1 我的图表模板 /214
 - 7.3.2 快速制图 /215
 - 7.3.3 添加标签小工具 /216
 - 7.3.4 修剪超大值 /218
- 7.4 本章小结 /220**

第8章 专业的报告——体现你的职场价值

- 8.1 初识数据分析报告 /222**
 - 8.1.1 数据分析报告是什么 /222
 - 8.1.2 数据分析报告的写作原则 /222
 - 8.1.3 数据分析报告的作用 /224
 - 8.1.4 数据分析报告的种类 /225
- 8.2 数据分析报告的结构 /228**
 - 8.2.1 标题页 /228
 - 8.2.2 目录 /230
 - 8.2.3 前言 /231
 - 8.2.4 正文 /233
 - 8.2.5 结论与建议 /234
 - 8.2.6 附录 /234
- 8.3 撰写报告时的注意事项 /235
- 8.4 报告范例 /236
- 8.5 本章小结 /244**

写在后面的Q/A

附录A 网络学习资源

第 1 章

数据分析那些事儿

数据分析是“神马”

数据分析六步曲

数据分析的三大误区

数据分析师的职业发展

几个常用指标和术语



>> 谁说菜鸟不会数据分析（入门篇）

小白过五关斩六将，通过严格的面试，最终从众多优秀毕业生中脱颖而出，成为公司的一员。在报到的第一天，公司HR向小白介绍了她的职位——公司牛董的助理，负责文秘工作，可能需要做一些数据分析之类的活儿。小白一听到数据分析这个词，就感觉头皮发麻，这时，她的脑子里幻想出一些穿白大褂的科研人员在实验室的计算机前不断忙碌的场景。虽然在上大学时也使用过Excel，但是如果要做数据分析工作，她还真的有些不知如何下手。无数个问号涌到她的脑海中：数据分析到底要做什么呢？我要怎么做数据分析？老板想要看什么样的结果呢？……唉，只好边走边干了。

HR看出小白的心事，说道：关于数据分析你不用太担心，如果遇到难题，你可以请教我们公司的Mr.林，他在这方面可是专家喔！

小白一听有救星，立马兴奋起来，好像抓到救命稻草一样，想赶紧找到这位大师级人物，然后一股脑把疑问全倒出来。

HR：小白，你跟我来吧，我给你引荐下Mr.林。

小白：好的。

说着小白跟HR来到了Mr.林的办公桌旁，**HR说道：**Mr.林，这是我们公司新来的同事，叫小白，现担任牛董的助理，其中部分工作涉及数据分析，到时候有问题还要麻烦您多多指点。

小白紧跟着说道：Mr.林，您好，请多多指教。

Mr.林：别客气，有问题直接来找我就可以了。

小白趁机说道：我现在就有问题，您现在是否有时间帮我解答下？数据分析是干什么的，具体要怎么做？

Mr.林听完笑了起来：你还真是不客气呀。好吧，你刚进公司，我就先给你做个简单的培训，带你入个门吧，以后的修行可就靠你自己努力了。

小白用力点着头，HR见这“师徒”二人颇有一见如故之感，大致交待一番后就离开了。

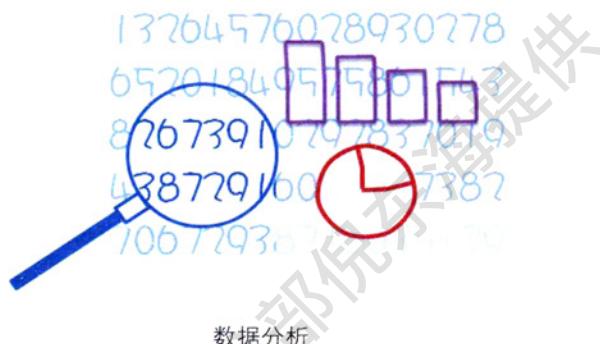
1.1 数据分析是“神马”

Mr.林接着说道：小白，既然要学习数据分析，借用现在网络流行语，你能说说数据分析是“神马”么？

小白：呵呵，数据分析肯定不是“浮云”。从字面上理解，数据分析就是对数据进行分析。Mr.林，我只能理解到这一层面，专业的解释就需要您来指教了。

1.1.1 何谓数据分析

Mr.林: 简单来说, 数据分析就是你刚才讲的——对数据进行分析。较为专业的说法, 数据分析是指用适当的统计分析方法对收集来的大量数据进行分析, 将它们加以汇总和理解并消化, 以求最大化地开发数据的功能, 发挥数据的作用。数据分析是为了提取有用信息和形成结论而对数据加以详细研究和概括总结的过程。这里的数据也称观测值, 是通过实验、测量、观察、调查等方式获取的结果, 常常以数量的形式展现出来。



数据分析

数据分析的目的是把隐藏在一大批看似杂乱无章的数据背后的信息集中和提炼出来, 总结出所研究对象的内在规律。在实际工作中, 数据分析能够帮助管理者进行判断和决策, 以便采取适当策略与行动。例如, 企业的高层希望通过市场分析和研究, 把握当前产品的市场动向, 从而制订合理的产品研发和销售计划, 这就必须依赖数据分析才能完成。

在统计学领域, 有些学者将数据分析划分为描述性数据分析、探索性数据分析以及验证性数据分析(参见图1-1)。其中, 探索性数据分析侧重于在数据之中发现新的特征, 而验证性数据分析则侧重于验证已有假设的真伪证明。

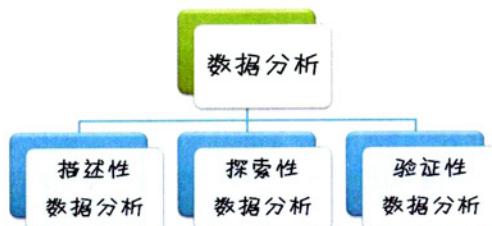


图1-1 数据分析类别

从另一个角度看, 描述性数据分析属于初级数据分析, 常见的分析方法有对比分析法、平均分析法、交叉分析法等。而探索性数据分析以及验证性数据分析属于高级数据分析, 常见的分析

方法有相关分析、因子分析、回归分析等。我们日常学习和工作中涉及的数据分析主要是描述性数据分析，也就是大家常用的初级数据分析。

1.1.2 数据分析的作用

Mr.林：了解了数据分析是“神马”后，那么小白，你能想到数据分析在对企业日常经营状况的分析工作中具体有哪些作用，体现在哪几方面吗？

小白：您刚才说过，数据分析就是把隐藏在一大批看似杂乱无章的数据背后的信息集中和提炼出来，总结出所研究对象的内在规律，帮助管理者进行有效的判断和决策。我觉得它就好比是从矿山中挖掘出金子。

Mr.林：没错，但刚才说的是数据分析的最终目的，要达到这些目的，我们在日常工作中该做些什么呢？比如日常通报、专题分析等，这些都是数据分析具体工作的体现。而什么时候做通报，什么时候该开展专题分析，则需要我们根据实际情况做出选择。很多人经常做这些工作，但不知为何而做，只是为做而做，没有想清楚做的目的，所以常常不得要领被数据所淹没。只有当你对数据分析目的及工作有了足够清晰的认识，开展数据分析时才会如鱼得水，游刃有余。

数据分析在企业的日常经营分析中主要有三大作用，如图1-2所示。



图1-2 数据分析三大作用

◎ 现状分析

简单来说就是告诉你过去发生了什么。具体体现在：

第一，告诉你企业现阶段的整体运营情况，通过各个经营指标的完成情况来衡量企业的运营状态，以说明企业整体运营是好了还是坏了，好的程度如何，坏的程度又到哪里。

第二，告诉你企业各项业务的构成，让你了解企业各项业务的发展及变动情况，对企业运营状况有更深入的了解。

现状分析一般通过日常通报来完成，如日报、周报、月报等形式。

◎ 原因分析

简单来说就是告诉你某一现状为什么发生。

经过第一阶段的现状分析，我们对企业的运营情况有了基本了解，但不知道运营情况具体好在哪里，差在哪里，是什么原因引起的。这时就需要开展原因分析，以进一步确定业务变动的具体原因。例如2012年2月运营收入环比下降5%，是什么原因导致的呢？是各项业务收入都出现下降，还是个别业务收入下降引起的？是各个地区业务收入都出现下降，还是个别地区业务收入下降引起的？这就需要我们开展原因分析，进一步确定收入下降的具体原因，对运营策略做出调整与优化。

原因分析一般通过专题分析来完成，根据企业运营情况选择针对某一现状进行原因分析。

◎ 预测分析

简单来说就是告诉你将来会发生什么。

在了解企业运营现状后，有时还需要对企业未来发展趋势作出预测，为制订企业运营目标及策略提供有效的参考与决策依据，以保证企业的可持续健康发展。

预测分析一般通过专题分析来完成，通常在制订企业季度、年度等计划时进行，其开展的频率没有现状分析及原因分析高。

Mr.林：小白，现在对数据分析有了进一步认识了吧。什么时候开展什么样的数据分析，需要根据你的需求及目的来确定。

小白：是的，Mr.林，那么请问数据分析具体如何开展呢？有哪些准备工作要做呢？

1.2 数据分析六步曲

Mr.林：小白，你的问题问得好，现在我们就来看看如何开展数据分析工作。

数据分析过程主要包括6个既相对独立又互有联系的阶段。它们是：明确分析目的和思路、数据收集、数据处理、数据分析、数据展现、报告撰写等6步（参见图1-3）。

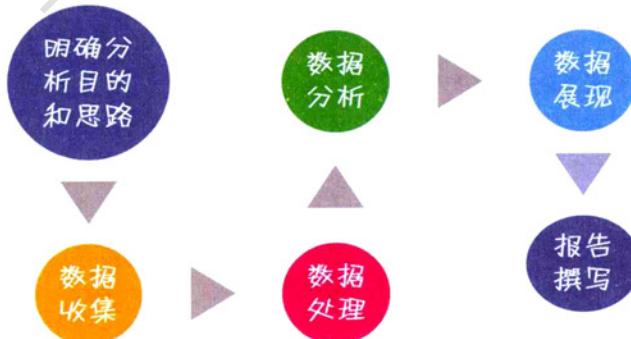


图1-3 数据分析六步曲

1.2.1 明确分析目的和思路

◎ 明确分析目的

Mr.林：做任何事都要有个目标，数据分析也不例外。小白，我们先来看看菜鸟与数据分析师之间的差别。刚好我这里有一张菜鸟与数据分析师想法对比图（见图1-4），你可以对比一下，看看自己在哪些方面想法还属于菜鸟级别，在哪些方面想法已经达到了数据分析师的水平。从他们思考问题的方式，你就能发现为什么一个是菜鸟，而另外一个是数据分析师了。

小白：好的，我来看一看菜鸟与数据分析师各是什么样的。

菜鸟与数据分析师的区别	
菜鸟会想	分析师会想
这张曲线图真好看，怎么作的？	数据变化的背后真相是什么？
这些数据可以做什么样的分析？	从哪些角度分析数据才系统？
高级的分析方法在这里能用吗？	用什么分析方法最有效？
要作多少张图表？	图表是否表达出有效的观点？
除了为数据添加文字说明还需要什么？	数据分析的目的达到了吗？
数据分析报告要写多少页？	数据分析报告有说服力吗？
.....

图1-4 菜鸟与数据分析师的区别

Mr.林：小白，从菜鸟与数据分析师想法对比图中，你发现菜鸟与数据分析师的区别了吗？

小白：我发现，菜鸟很迷茫，目的不明确，而且一味追求高级的分析方法。而数据分析师的目的就很明确，一切都是以解决问题为中心。

Mr.林：说得好！其实他们最主要的区别就在于目的是否明确，如果目的明确，所有问题就自然迎刃而解了。例如，分析师是不会考虑“需要多少张图表”这样的问题的，而是思考这张图表是否有效表达了观点？如果没有，需要怎样调整？如果目的不明确，菜鸟自然会在用什么样的分析方法、作多少张图表、需要多少文字说明、报告要写多少页等这些问题上纠结。

小白：我明白了，我不要做菜鸟，我要朝着数据分析师的方向努力。

Mr.林：菜鸟与数据分析师的区别就在于菜鸟做分析时目的不明确，从而导致分析过程非常盲目。所以在开展数据分析之前，要想想：为什么要开展数据分析？通过这次数据分析我要解决什么问题？只有明确数据分析的目标，数据分析才不会偏离方向，否则得出的数据分析结果不仅没有指导意义，甚至可能将决策者引入歧途，后果严重。

◎ 确定分析思路

Mr.林：当分析目的明确后，我们就要梳理分析思路，并搭建分析框架，把分析目的分解成

若干个不同的分析要点，即如何具体开展数据分析，需要从哪几个角度进行分析，采用哪些分析指标。

只有明确了分析目的，分析框架才能跟着确定下来，最后还要确保分析框架的体系化，使分析结果具有说服力。

小白：Mr.林，前面说的角度与指标我都理解，但分析体系化应该如何理解呢？

Mr.林：体系化也就是逻辑化，简单来说就是先分析什么，后分析什么，使得各个分析点之间具有逻辑联系。这也是很多人常常感到困扰的问题，比如经常不知从哪方面入手，分析的内容和指标常常被质疑是否合理、完整，而自己也说不出个所以然来，所以体系化就是为了让你的分析框架具有说服力。

小白：如何使分析框架体系化呢？

Mr.林：问得好！根据我的经验，就是以营销、管理等理论为指导，结合实际业务情况，搭建分析框架，这样才能确保数据分析维度的完整性，分析结果的有效性及正确性。

营销方面的理论模型有4P、用户使用行为、STP理论、SWOT等，而管理方面的理论模型有PEST、5W2H、时间管理、生命周期、逻辑树、金字塔、SMART原则等。这些都是经典的营销、管理方面的理论，需要在工作中不断实践应用，你才能体会其强大的作用。如图1-5所示，就是以PEST分析理论为指导，搭建的互联网行业PEST分析框架，故而使数据分析变得有血有肉有脉络，真正做到理论指导实践。其他具体应用我在后面会为你介绍。



图1-5 互联网行业PEST分析框架

明确数据分析目的以及确定分析思路，是确保数据分析过程有效进行的先决条件，它可以为数据的收集、处理及分析提供清晰的指引方向。

小白：嗯，在做数据分析前，先想想到底为什么要做数据分析，明确数据分析方向，才能保证不迷失在数据的丛林中。

1.2.2 数据收集

Mr.林：数据收集是按照确定的数据分析框架，收集相关数据的过程，它为数据分析提供了素材和依据。这里所说的数据包括第一手数据与第二手数据，第一手数据主要指可直接获取的数据，第二手数据主要指经过加工整理后得到的数据。一般数据来源主要有以下几种方式，如图1-6所示。

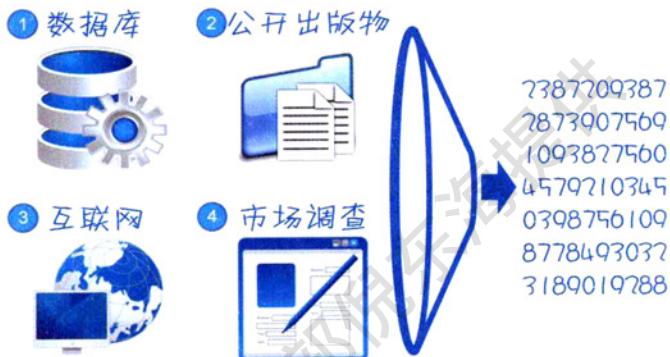


图1-6 数据的来源

① 数据库

每个公司都有自己的业务数据库，存放从公司成立以来产生的相关业务数据。这个业务数据库就是一个庞大的数据资源，需要有效地利用起来。

② 公开出版物

可以用于收集数据的公开出版物包括《中国统计年鉴》《中国社会统计年鉴》《中国人口统计年鉴》《世界经济年鉴》《世界发展报告》等统计年鉴或报告。

③ 互联网

随着互联网的发展，网络上发布的数据越来越多，特别是搜索引擎可以帮助我们快速找到所需要的数据，例如国家及地方统计局网站、行业组织网站、政府机构网站、传播媒体网站、大型综合门户网站等上面都可能有我们需要的数据。

④ 市场调查

进行数据分析时，需要了解用户的想法与需求，但是通过以上三种方式获得此类数据会比较困难，因此可以尝试使用市场调查的方法收集用户的想法和需求数据。市场调查就是指运用科学的方法，有目的、有系统地收集、记录、整理有关市场营销的信息和资料，分析市场情况，了解

市场现状及其发展趋势，为市场预测和营销决策提供客观、正确的数据资料。市场调查可以弥补其他数据收集方式的不足，但进行市场调查所需的费用较高，而且会存在一定的误差，故仅作参考之用。

小白：看来我以后做数据分析时，不用担心没有数据来源啦。

1.2.3 数据处理

Mr.林：数据处理是指对收集到的数据进行加工整理，形成适合数据分析的样式，它是数据分析前必不可少的阶段。数据处理的基本目的是从大量的、杂乱无章、难以理解的数据中，抽取并推导出对解决问题有价值、有意义的数据。

数据处理主要包括数据清洗、数据转化、数据提取、数据计算等处理方法（参见图1-7）。一般拿到手的数据都需要进行一定的处理才能用于后续的数据分析工作，即使再“干净”的原始数据也需要先进行一定的处理才能使用。



图1-7 数据处理

小白：哦，数据处理是数据分析的前提，对有效数据的分析才是有意义的，明白啦！

1.2.4 数据分析

Mr.林：数据分析是指用适当的分析方法及工具，对处理过的数据进行分析，提取有价值的信息，形成有效结论的过程。



数据分析

在确定数据分析思路的阶段，数据分析师应当为需要分析的内容确定适合的数据分析方法，等到真正进入进行数据分析阶段时，就能够驾驭数据，从容地进行分析和研究了。

由于数据分析多是通过软件来完成的，这就要求数据分析师不仅要掌握各种数据分析方法，还要熟悉主流数据分析软件的操作。一般的数据分析我们可以通过Excel完成，后面也将重点为你介绍，而高级的数据分析就要采用专业的分析软件进行了，如数据分析工具SPSS Statistics等。

小白：好的，期待您后续精彩的数据分析讲解。

◎ 数据分析与数据处理的区别

小白：Mr.林，我有个问题。我经常听到数据处理和数据分析这两个词，它们之间有何关系呢？

Mr.林：从前面介绍数据处理和分析的过程不难看出，数据处理是数据分析的基础。通过数据处理，将收集到的原始数据转换为可以分析的形式，并且保证数据的一致性和有效性。如果数据本身存在错误，那么即使采用最先进的数据分析方法，得到的结果也是错误的，不具备任何参考价值，甚至还会误导决策。

◎ 数据分析与数据挖掘的关系

小白：还有个问题，我常听到数据挖掘，那么数据挖掘是什么呢？它与数据分析有什么关系呢？

Mr.林：这个问题问得好！数据挖掘其实是一种高级的数据分析方法。

数据挖掘就是从大量的数据中挖掘出有用的信息，它是根据用户的特定要求，从浩如烟海的数据中找出所需的信息，以满足用户的特定需求。数据挖掘技术是人们长期对数据库技术进行研究和开发的结果。一般来说，数据挖掘侧重解决四类数据分析问题：分类、聚类、关联和预测，重点在寻找模式与规律。

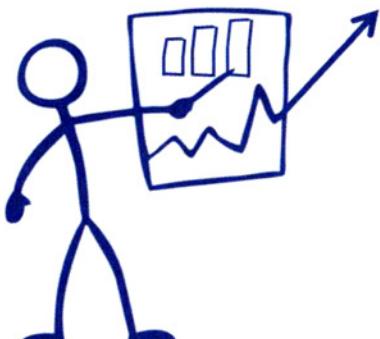
数据分析与数据挖掘的本质是一样的，都是从数据里面发现关于业务的知识。

数据挖掘现在还不是你学习的重点，如果你有兴趣，可在掌握一定的数据分析知识后，再进一步查阅相关资料学习。

1.2.5 数据展现

小白：您刚才讲通过分析，隐藏在数据内部的关系和规律就会逐渐浮现出来，那么通过什么方式展现出这些关系和规律，才能让人一目了然呢？

Mr.林：不错，孺子可教也！这就是我接下来要说的数据展现了。众所周知，每个人看待事物都有自己的理解方式，所以数据分析师在展现结果的时候一定要保证绝大部分人的理解是一致的。



数据展现

一般情况下，数据是通过表格和图形的方式来呈现的，我们常说用图表说话就是这个意思。常用的数据图表包括饼图、柱形图、条形图、折线图、散点图、雷达图等，当然可以对这些图表进一步整理加工，使之变为我们所需要的图形，例如金字塔图、矩阵图、漏斗图、帕雷托图等。

大多数情况下，人们更愿意接受图形这种数据展现方式，因为它能更加有效、直观地传递出分析师所要表达的观点。记住，在一般情况下，能用图说明问题的就不用表格，能用表格说明问题的就不用文字。

小白：原来图表的运用还这么有讲究呀，我一定好好学习如何用图表说话！

1.2.6 报告撰写

Mr.林：数据分析报告其实是对整个数据分析过程的一个总结与呈现。通过报告，把数据分析的起因、过程、结果及建议完整地呈现出来，供决策者参考。所以，数据分析报告是通过对数据全方位的科学分析来评估企业运营质量，为决策者提供科学、严谨的决策依据，以降低企业运营风险，提高企业核心竞争力。

一份好的数据分析报告，首先需要有一个好的分析框架，并且图文并茂，层次明晰，能够让阅读者一目了然。结构清晰、主次分明可以使阅读者正确理解报告内容；图文并茂，可以令数据更加生动活泼，提高视觉冲击力，有助于阅读者更形象、直观地看清楚问题和结论，从而产生思考。

另外，数据分析报告需要有明确的结论，没有明确结论的分析称不上分析，同时也失去了报告的意义，因为我们最初就是为寻找或者求证一个结论才进行分析的，所以千万不要舍本求末。

最后，好的分析报告一定要有建议或解决方案。作为决策者，需要的不仅仅是找出问题，更重要的是建议或解决方案，以便他们在决策时作参考。所以，数据分析师不仅需要掌握数据分析方法，而且还要了解和熟悉业务，这样才能根据发现的业务问题，提出具有可行性的建议或解决方案。下面的图1-8给出了一个分析报告的基本框架示例。

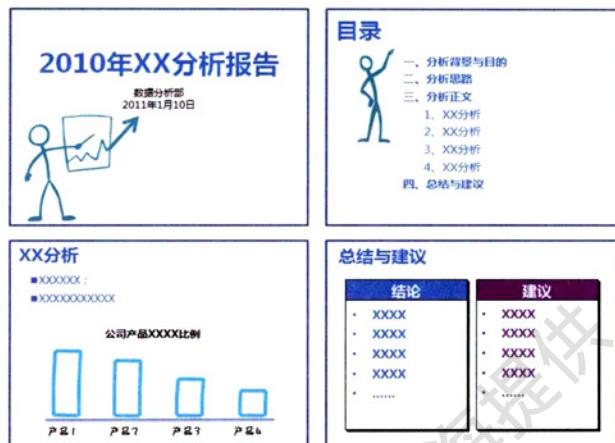


图1-8 数据分析报告示例

Mr.林：以上是数据分析的6个关键步骤，今天只是简要介绍，以后将逐一为你讲解。说了这么多，给你讲个段子轻松一下，顺便回顾下刚讲的内容。

小白一听有段子听，立马精神起来：好啊！好啊！

Mr.林：如何判断数据分析师正处于分析流程的哪个阶段？

- ★ 手托腮帮，沉思痛苦状——在思考分析思路。
- ★ 手放键盘上不动，表情呆滞——在数据处理。
- ★ 鼠标在飞快移动——在用数据透视表做分析。
- ★ 不断交替点击鼠标左右键——在画图表。
- ★ 断断续续敲击键盘，时而移动鼠标——在写PPT分析报告。

小白：还真形象啊！Mr.林，您该不会说的就是您自己吧？呵呵！

Mr.林：哈哈，数据分析师差不多都是这样的状态。

1.3 数据分析的三大误区

Mr.林：小白，刚才对数据分析进行了简要介绍，而在实际的学习、工作中，常常有数据分析人员陷入一些误区，现在我来讲一讲这些误区，你日后一定要注意。

小白：好的，请Mr.林指教。

① 分析目的不明确，为分析而分析

Mr.林：经常有人问：要用多少图？除了摆数据，还需要说些什么？在此我想说的是，数据分析不应为了分析而分析，而是应该围绕你的分析目的（了解现状、找出业务变动原因、预测发

展等)而进行分析。

只有对自己的目的有清晰的认识,你才知道要怎样去实现这个目的,需要通过哪些图表展现,才会知道这些图表是否能反映问题,自然而然地进行相应的问题分析,而不是连该说些什么都不知道。

◎ 缺乏业务知识,分析结果偏离实际

Mr.林: 目前现有的数据分析师大多是统计学、计算机、数学等专业出身,他们大多缺乏从事营销、管理方面的工作经验,对业务的理解相对较浅,对数据的分析偏重于数据分析方法的使用,如回归分析、相关分析等。

有的公司老板抱怨手下的数据分析师每天给他看几十个零散数据,虽然做出的报告很专业,图表也很漂亮,但所作的分析忽视了业务逻辑上的关联性,得不到全面、综合性的结论。

在企业中所作的数据分析不是纯数据分析,而是需要多从业务方面进行分析,不应停留在数据表面,要思考数据背后的事实与真相,使得分析结果更加切合实际,为老板的决策提供有力的支撑,否则就是纸上谈兵。

所以说,数据分析师的任务不是单纯做数学题,数据分析师还必须懂营销,懂管理,更要懂策略。

◎一味追求使用高级分析方法,热衷研究模型

Mr.林: 在进行数据分析时,相当一部分人都喜欢用回归分析、因子分析等高级分析方法,总认为有分析模型就是专业的,只有这样才能体现专业性,结果才是可信的。其实不然,高级的数据分析方法不一定是最好的,能够简单有效解决问题的方法才是最好的。

我们坚信,仅有分析模型远远不够,围绕业务发现问题并解决问题才是数据分析的最终目的!不论高级的分析方法还是简单的分析方法,只要能够解决业务问题,就是好方法,正如我们常说的“不论黑猫还是白猫,只要能抓老鼠就是好猫”。

小白: 好的,我在以后的工作中会注意这些的,做好一个数据分析师并不是一件容易的事,我要努力向Mr.林靠齐。

1.4 数据分析师的职业发展

1.4.1 数据分析的广阔前景

Mr.林: 数据分析作为一个新的行业领域正在全球迅速发展,它开辟了人类获取知识的新途径。

>> 谁说菜鸟不会数据分析（入门篇）

目前，数据库技术、软件工具、各种硬件设备飞速发展，在这些软硬件技术与设备的支持下，信息技术的应用已在各行各业全面展开，尤其是对通信、互联网、金融等行业的发展做出了巨大贡献，并且经过长期的应用积累了大量丰富的数据。但大部分企业对其存储信息的利用率极低。庞大的历史数据是否有价值？有何价值？是否可以综合利用分析？是否能够为领导决策提供参考依据？

回答是肯定的，数据分析这一项工作越来越受到决策层的重视，借助数据分析的各种工具，从海量的历史数据中提取、挖掘对业务发展有价值的、潜在的知识，找出趋势，为决策层的决策提供有力的依据，对产品或服务的发展方向起到积极作用，有力推动企业的科学化、信息化管理。

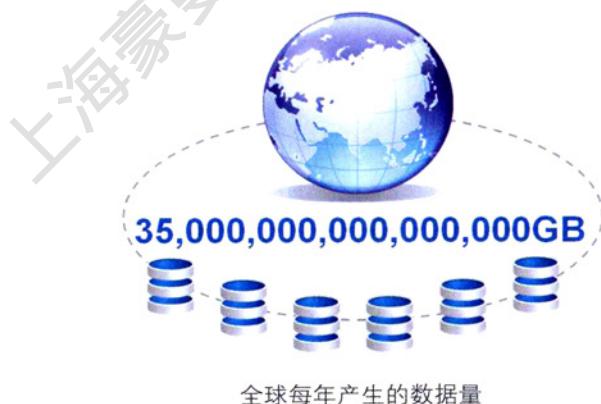
从20世纪90年代起，欧美国家开始大量培养数据分析师，直到现在，对数据分析师的需求仍然长盛不衰，而且还有扩张之势。根据美国劳工部预测，到2018年，数据分析师的需求量将增长20%。就算你不是数据分析师，但数据分析技能也是未来必不可少的工作技能之一。

小白：数据分析师如此抢手的原因何在呢？

Mr.林：一个简单的原因就是社会越发达，人们对数据的依赖就越多。无论政府决策还是公司运营，科学研究还是媒体宣传，都需要数据支持。那么，对数据有如此大的依赖，就必然导致对数据分析的大量需求。因此，将数据转化为知识、结论和规律，就是数据分析的作用和价值。

小白：那数据究竟会庞大到什么地步呢？

Mr.林：据国际知名咨询公司估计，到2020年，全球每年产生的数据量将达到3500万亿GB，打个比方，就是用普通的DVD一张一张地摞起来，可以从地球摞到月球两次。



面对这样庞大的数据，对数据分析师的要求就不仅仅是单纯做分析了，更重要的是与相关部门进行合作，将数据真正应用到业务中，根据实际的业务发展情况识别哪些数据可用，哪些不适用，而不是孤立地在“真空环境”下进行分析。这就要求数据分析师不仅具备洞察数据的能力，还要对相关业务的背景有深入的了解，明白客户或业务部门的需求，从而将数据信息化、可

视化，最后转化为生产力，帮助企业获得利润，这就是整个数据“供应链”。当然，数据分析师也需要理解这个“供应链”。

1.4.2 数据分析师的职业要求

小白：那么怎样才能成为一名优秀的数据分析师呢？数据分析师的职业要求又有哪些？

Mr.林：你的问题可以用“五懂”来回答，即懂业务、懂管理、懂分析、懂工具，还要懂设计。

◎ 懂业务

从事数据分析工作的前提就是需要懂业务，即熟悉行业知识、公司业务及流程，最好有自己独到见解，若脱离行业认知和公司业务背景，分析的结果只会是脱了线的风筝，没有太大的实用价值。

例如公司2011年的运营收入是1000万元，那么不熟悉业务的数据分析师看到的只是1000万这个数字，而熟悉业务的数据分析师，还能看到这个数字背后隐藏的信息，如1000万元是由哪几个业务收入构成，哪个业务收入占主要部分，哪个业务收入是最小占比，最高业务收入的地区又是哪个地区等。

这就是懂业务与不懂业务的数据分析师之间的区别。

从另外一个角度来说，懂业务也是数据敏感的体现。不懂业务的数据分析师，看到的只是一一个个数字；懂业务的数据分析师，则看到的不仅仅是数字，他明白数字代表什么意义，知道数字是大了还是小了，心中有数，这才是真正意义的数据敏感性。

◎ 懂管理

懂管理，一方面是搭建数据分析框架的要求，比如刚才介绍的数据分析六步曲的第一步确定分析思路就需要用到营销、管理等理论知识来指导，如果不熟悉管理理论，那你如何指导数据分析框架的搭建，以及开展后续的数据分析呢？

懂管理另一方面的作用是针对数据分析结论提出有指导意义的分析建议，如果没有管理理论的支撑，就难以确保分析建议的有效性。

所以数据分析师需要掌握一定的管理理论知识。

◎ 懂分析

懂分析是指掌握数据分析的基本原理与一些有效的数据分析方法，并能灵活运用到实践工作中，以便有效地开展数据分析。

基本的分析方法有：对比分析法、分组分析法、交叉分析法、结构分析法、漏斗图分析法、综合评价分析法、因素分析法、矩阵关联分析法等。

>> 谁说菜鸟不会数据分析（入门篇）

高级的分析方法有：相关分析法、回归分析法、聚类分析法、判别分析法、主成分分析法、因子分析法、对应分析法、时间序列等。

不论简单还是复杂的分析方法，只要能解决问题的方法就是好方法。

◎ 懂工具

懂工具是指掌握数据分析相关的常用工具。数据分析工具就是实现数据分析方法理论的工具，面对越来越庞大的数据，依靠计算器进行分析是不现实的，必须利用强大的数据分析工具完成数据分析工作。

常用的数据分析工具有Excel、Access、SPSS、SAS，建议先用好Excel分析工具。Excel就是一款非常实用的数据处理、分析工具，它能解决、满足你工作中80%，甚至100%的要求。有兴趣、时间、需要的话，再学习SPSS、SAS等统计分析工具。

同样，应该根据研究的问题选择合适的工具，只要能解决问题的工具就是好工具。

◎ 懂设计

懂设计是指运用图表有效表达数据分析师的分析观点，使分析结果一目了然。

图表的设计是门大学问，如图形的选择、版式的设计、颜色的搭配等，都需要掌握一定的设计原则。

小白：颜色搭配都有要求？不能根据我自己的喜好选择么？

Mr.林：当然有要求，根据喜好选择也要在符合设计原则的基础上选择，例如公司的VI设计是以蓝色为主色调，属冷色系，那么你选的图表颜色就要与公司的VI设计相吻合，尽量避免使用红色、黄色等暖色系的颜色，这方面知识我会在讲到数据展现部分再为你详细介绍。

小白：好的，看来要成为一名合格的数据分析师还有很长的路要走啊。

1.4.3 数据分析师的基本素质

Mr.林接着说道：要成为一名优秀的数据分析师并非一件容易的事。虽然大学的专业与数据分析不相关，但你可以通过工作中的实践来学习数据分析，这需要付出大量的时间和精力，不经一番寒彻骨，怎得梅花扑鼻香？

小白点点头：嗯，我会努力的。

Mr.林：刚才介绍的是数据分析师的职业要求，现在再介绍数据分析师的五大基本能力和素质。

◎ 态度严谨负责

严谨负责是数据分析师的必备素质之一，只有本着严谨负责的态度，才能保证数据的客观、准确。

数据分析师可以说是企业的医生，他们通过对企业运营数据的分析，为企业寻找症结及潜在

问题。一名合格的数据分析师，应具有严谨负责的态度，保持中立立场，客观评价企业在发展过程中存在的问题，为决策层提供有效的参考依据。数据分析师不应受其他因素影响而更改数据，隐瞒企业存在的问题，这样做对企业发展是非常不利的，甚至会造成严重的后果。而且，对数据分析师自身来说，其职业生涯也会受影响，从此以后所做的数据分析结果都将受到质疑，因为你不再是可信赖的人，在同事、领导、客户面前已经失去了信任。

所以，作为一名数据分析师就必须持有严谨负责的态度，这也是最基本的职业道德。



严谨负责

◎ 好奇心强烈

好奇心人皆有之，但是作为数据分析师，这份好奇心就应该更强烈，应该把自己当成数据福尔摩斯，要积极主动地发现和挖掘隐藏在数据内部的真相。



好奇心强烈

在数据分析师的脑子里，应该充满着无数个“为什么”：为什么是这样的结果，为什么不是那样的结果，导致这个结果的原因是什么，为什么结果不是预期的那样等等，只有这样才有突破点。

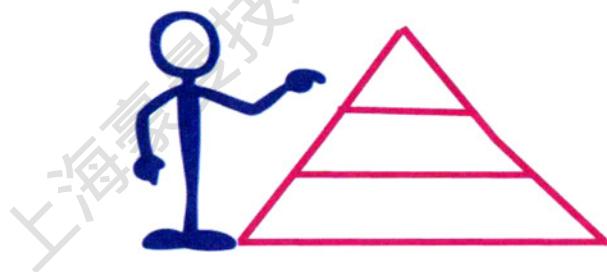
这一系列问题都要在进行数据分析时提出来，并且通过数据分析，给自己一个满意的答案。越是优秀的数据分析师，好奇心越不容易满足，回答了一个问题，又会抛出一个新的问题，继续研究下去。

只有拥有了这样一种刨根问底的精神，才会对数据和结论保持敏感，继而顺藤摸瓜，找出数据背后的真相。当然，你也会从中获取成就感。

◎ 逻辑思维清晰

除了一颗探索真相的好奇心，数据分析师还需要具备缜密的思维和清晰的逻辑推理能力。我记得有位大师说过：结构为王。何谓结构，结构就是我们常说的逻辑，不论说话还是写文章的时候都要有条理、有目的地做，不可眉毛胡子一把抓，不分主次。

通常，从事数据分析时所面对的商业问题都是较为复杂的，我们要思考错综复杂的成因，分析所面对的各种复杂的环境因素，并在问题的若干发展可能性中选择一个最优的方向。这就需要我们对事实有足够的了解，同时也需要我们能真正理清问题的整体以及局部的结构，在深度思考后，理清结构中相互的逻辑关系，只有这样才能真正客观地、科学地找到商业问题的答案。



逻辑思维清晰

◎ 擅长模仿学习

在做数据分析时，有自己的想法固然重要，但是“前车之鉴”也是非常有必要学习的，它能帮助数据分析师迅速地成长。因此，模仿学习是快速提高学习成果的有效方法。这里说的模仿主要是指参考、借鉴他人优秀的分析思路和方法，而非直接“照搬”。

成功的模仿需要领会他人方法的精髓，理解其分析原理，透过表面达到实质。万变不离其宗，要善于将这些精华转化为自己的知识，否则，只能是“一直在模仿，从未超越过”。

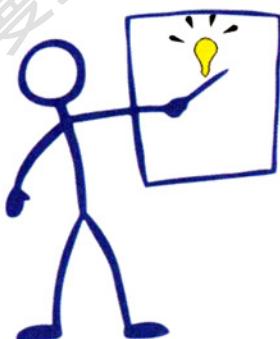


擅长模仿学习

◎ 勇于创新

通过模仿可以借鉴他人的成功经验，但模仿的时间不宜太长，并且建议每次模仿后都要进行总结，提出可以改进的地方，甚至要有所创新，不断总结分析方法、分析思路、分析流程，在总结中前行。

创新是一个优秀数据分析师应具备的精神，只有不断的创新，才能提高自己的分析水平，使自己站在更高的角度来分析问题，为整个研究领域乃至社会带来更多的价值。现在的分析方法和研究课题千变万化，墨守成规是无法很好地解决所面临的新问题的。



创新

听到这里，小白就掰着手指头算自己有几条符合优秀数据分析师的素质和能力。

Mr.林继续说道：学习数据分析需要时间和经验的积累，这些素质与能力不是说有就有的，需要慢慢培养形成，不能一蹴而就。在工作中运用不同的分析方法对数据进行分析，并与业务部门的同事积极沟通，加深自己对整个行业或研究内容的理解，相信在两到三年内，你就能成为一名合格的数据分析师。

小白：好的，我会努力学习的。

1.5 几个常用指标和术语

Mr.林：一个优秀的数据分析师要有扎实的数据解读功底，因为在进行数据分析时，经常会遇到一些分析指标或术语，对这些指标或术语的理解不够充分就没法开展工作。还有点时间，我先给你简单介绍一些基础的分析指标和术语。

小白：太好了！拿支笔先~~

◎ 平均数

Mr.林：我们在日常生活中提到的平均数，一般是指算术平均数，就是一组数据的算术平均值，即全部数据累加后除以数据个数。算术平均数是非常重要的基础性指标，它的特点是将总体内各单位的数量差异抽象化，代表总体的一般水平，掩盖了总体内各单位的差异。

例如现有某学期学生的数学考试成绩，通过计算成绩平均数，可得如图1-9所示的结果。将每位同学的数学成绩与平均数相比较，就能发现哪些同学的数学成绩高于平均数，需要保持；哪些同学的成绩低于平均数，需要继续努力。

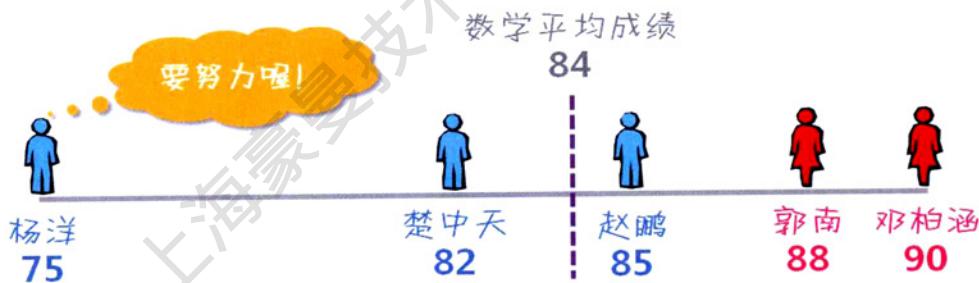


图1-9 数学平均成绩

当然，在平均数这个指标中，除算术平均数以外，还有其他平均数，如调和平均数和几何平均数。

小白：一个小小的平均数都有这么多学问呀。

◎ 绝对数与相对数

Mr.林：绝对数是反映客观现象总体在一定时间、地点条件下的总规模、总水平的综合性指标，也是数据分析中常用的指标，如GDP、总人口等。此外，绝对数也可以表现为在一定时间、地点条件下数量增减变化的绝对数，比如A国人口比B国人口多1000万人。

相对数是指由两个有联系的指标对比计算而得到的数值，用以反映客观现象之间数量联系程

度的综合指标。计算相对数的基本公式是：

$$\text{相对数} = \frac{\text{比较数值(比数)}}{\text{基础数值(基数)}}$$

分母是用做对比标准的指标数值，简称基数；分子是用做与基数对比的指标数值，简称比数。相对数一般以倍数、成数、百分数等表示，它反映了客观现象之间数量联系的程度。

使用相对数时需要注意指标的可比性，同时要与总量指标（绝对数）结合使用。关于绝对数与相对数，我举几个例子，你一看就明白了，如图1-10所示。

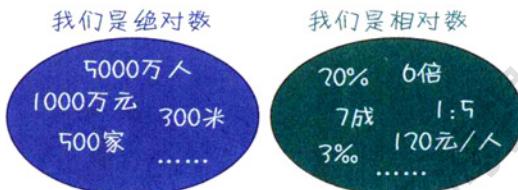


图1-10 绝对数与相对数

◎ 百分比与百分点

Mr.林：百分比是相对数中的一种，它表示一个数是另一个数的百分之几，也称百分率或百分数。百分比通常采用百分号（%）来表示，如8%，50%，168%等。由于百分比的分母都是100，也就是都以1%作为度量单位（如图1-11所示），因此便于比较，在数据分析中的应用非常广泛。

$$\frac{10}{100} = 10\%$$

图1-11 百分比的表示形式

百分点是指不同时期以百分数的形式表示的相对指标的变动幅度，1个百分点=1%。例如，图1-12中某公司发言人就混淆了百分比与百分点的概念。表示构成的变动幅度不宜用百分数，而应该用百分点。因此，这位公司发言人正确的说法应该是“公司今年的利润是45%，比去年的28%提高了17个百分点”。



图1-12 百分比与百分点

◎ 频数与频率

Mr.林：频数是指一组数据中个别数据重复出现的次数。如图1-13的左图所示，某校A班共50名学生，按性别进行分组，分为男与女两个组别，男同学的频数为30，女同学的频数为20。

频率是每组类别次数与总次数的比值，它代表某类别在总体中出现的频繁程度，一般采用百分数表示，所有组的频率加总等于100%。还是用某校A班的例子，30个男同学在50个同学中出现的频率为60%，即 $(30 \div 50) \times 100\%$ ；而20个女同学在50个同学中出现的频率为40%，即 $(20 \div 50) \times 100\%$ ，如图1-13的右图所示。

小白：所以频数是绝对数，频率是相对数。

Mr.林赞许地说：对，活学活用，你领悟得很快嘛。

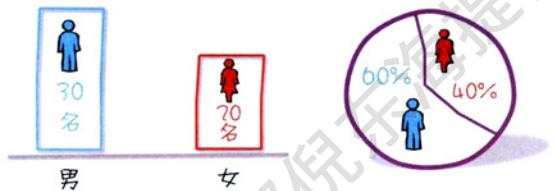


图1-13 频数与频率

◎ 比例与比率

Mr.林：比例与比率都属于相对数。

比例是指在总体中各部分的数值占全部数值的比重，通常反映总体的构成和结构。比如A班共有学生50人，男生30人，女生20人，则男生的比例是30 : 50，女生的比例是20 : 50。由此可以看出，比例的基数（也就是分母）都是全体学生人数，即为同一个基数。

比率是指不同类别数值的对比，它反映的不是部分与整体之间的关系，而是一个整体中各部分之间的关系。比如刚才的例子，男生30人，女生20人，则男生与女生的比率是30 : 20，如图1-14所示。这一指标经常会用在社会经济领域，比如我国的人口性别比就是用每100名女性数量相对的男性数量来表示的。

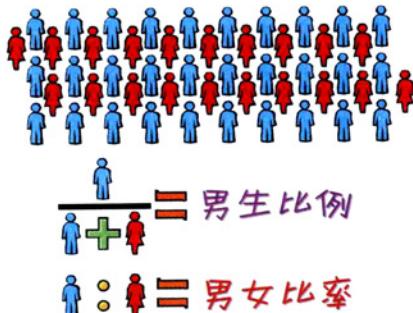


图1-14 比例与比率

◎ 倍数与番数

Mr.林：倍数与番数同样属于相对数，但使用时容易混淆。倍数是一个数除以另一个数所得的商。比如 $A \div B = C$ ，就是说A是B的C倍。需要注意的是，倍数一般表示数量的增长或上升幅度，而不适用于表示数量的减少或下降。

番数是指原来数量的2的N次方倍。比如翻一番为原来数量的2倍（ 2^1 ），翻两番为4倍（ 2^2 ）。如图1-15所示，这位公司发言人在讲话中就混淆了倍数与番数的概念。正确的说法应该是“公司产品销量翻一番（ $6.4 = 3.2 \times 2^1$ ），从去年的3.2万件提高到今年的6.4万件。此外，我们成本控制也很好，由20万元下降了50%，今年成本为10万元”。

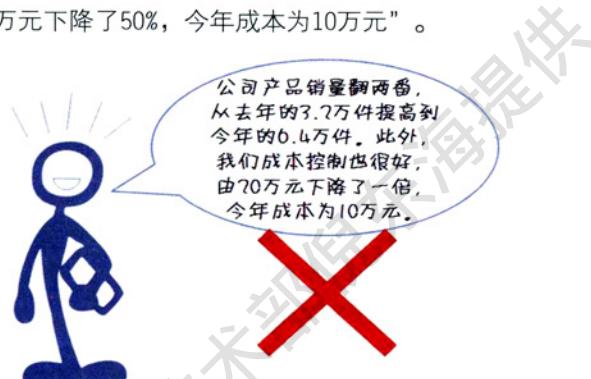


图1-15 倍数与番数

◎ 同比与环比

Mr.林：同比是指与历史同时期进行比较得到的数值，该指标主要反映的是事物发展的相对情况。例如2010年12月与2009年12月相比，如图1-16的左图所示。

环比是指与前一个统计期进行比较得到的数值，该指标主要反映的是事物逐期发展的情况。例如2010年12月与2010年11月相比，如图1-16的右图所示。

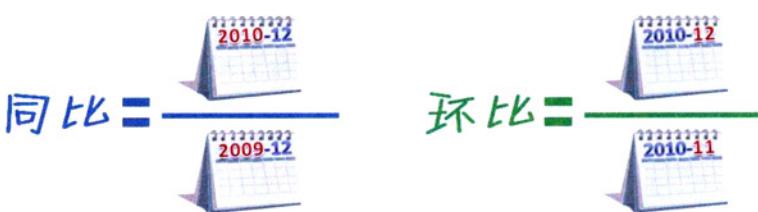


图1-16 同比与环比

Mr.林：讲了这么多，我估计你都听晕了。没关系，这些内容现在听起来虽然枯燥，但能够为你以后做数据分析打下坚实的基础。其中一些内容需要通过实际操作才能有更深刻的体会。

>> 谁说菜鸟不会数据分析（入门篇）

小白：是啊。我明白您讲的这些都是基本功，回去后我一定认真复习、加深理解，绝不辜负您的教导。

1.6 本章小结

Mr.林：今天讲了不少内容，估计你有点应接不暇了，我带着你回顾下这些内容。

- ★ 什么是数据分析以及数据分析的三大作用。
- ★ 数据分析的六步曲：首先是明确分析目的和思路，然后是数据收集，第三步是将收集回来的数据进行处理，第四步是根据分析目的和思路进行数据分析，第五步是将分析出的结果通过图表的方式展现出来，最后一步是撰写数据分析报告。
- ★ 认识常见的数据分析误区，明白数据分析师的要求与基本素质，了解如何成为一个合格的数据分析师。
- ★ 数据分析中经常用到的指标和术语，除掌握最基本的平均数、百分比等概念以外，还需要将容易混淆的指标进行明确区分，避免在以后阐述分析结果时出现错误。

最后送数据分析心法给你，也叫做数据分析三字经。

- ① 学习：先了解，后深入；先记录，后记忆；先理论，后实践；先模仿，后创新；
- ② 方法：先思路，后方法；先框架，后细化；先方法，后工具；先思考，后动手；
- ③ 分析：先业务，后数据；先假设，后验证；先总体，后局部；先总结，后建议。

当然你现在看这个还不会有太深刻的理解，没关系，正如三字经里所说，“先了解，后深入；先记录，后记忆”，等到积累一定经验，再回头看看你就会深有感触。

小白：谢谢Mr.林的指点，您说的我都记下了，我会加倍努力，早日出师。

第2章

结构为王——确定分析思路

数据分析方法论

常用的数据分析方法论

PEST 5W2H 4P

数据分析 = 



.....



第二天早上一上班，小白就揣着笔记本来到Mr.林办公桌前：Mr.林，早上好呀！我又来了，今天学什么呢？

Mr.林：早上好！小白，你好勤奋呀！学习态度真不错。

小白：那当然，有您这样的好老师，我不勤奋点，就太对不起您啦！

Mr.林：那我们开始今天的课程吧！昨天已经向你介绍了什么是数据分析、数据分析的作用、数据分析六步曲等知识，让你对数据分析有个整体认识，今天我们就来学习如何开展数据分析。首先要学的就是如何确定分析思路，这是非常关键的一步，如果分析思路不明确或者错误，那么后续的数据分析工作也无从开展了。

小白：好的。

2.1 数据分析方法论

Mr.林：昨天提到确定分析思路需要以营销、管理等理论为指导，我们把这些跟数据分析相关的营销、管理等理论统称为数据分析方法论。今天我们先学习数据分析方法论。

小白：方法论？听起来好高深哦！

Mr.林：小白，不要一听到方法论就觉得它深不可测，我先跟你讲一件亲身经历的事。

在我刚开始从事数据分析工作时，经常需要做一些专题分析，例如对公司的某个业务问题进行专项研究。每当完成专题分析向老板汇报分析结果时，老板首先就问：“你的分析方法论是什么？说来听听，我看分析报告就先看你的分析方法论，如果分析方法论不正确或不合理，后面的分析结果也就没有必要看了，在一个不正确或不合理的方法论的指导下，得到的分析结果是不可能正确的。”

当年老板的这番话对我触动很深，在后续的专题分析中，我养成了在分析方法论指导下开展分析的习惯，以确保分析结果具有指导意义。

小白点点头：明白了。方法论就是指南针，南辕北辙是很难到达目的地的。

2.1.1 数据分析方法论与数据分析法的区别

Mr.林：刚开始我对方法论的概念也比较模糊，学校里老师没教过，相关的书籍中也没介绍过，只好自己琢磨研究。经过一段时间的思考、研究，以及老板的点拨，我终于参透了其中的奥秘：数据分析方法论主要用来指导数据分析师进行一次完整的数据分析，它更多的是指数据分析思路，比如主要从哪几方面开展数据分析？各方面包含什么内容和指标？

所以，数据分析方法论主要从宏观角度指导如何进行数据分析，它就像是一个数据分析的前期规划，指导着后期数据分析工作的开展。而数据分析法则是指具体的分析方法，例如我们常见

的对比分析、交叉分析、相关分析、回归分析、聚类分析等数据分析法。数据分析法主要从微观角度指导如何进行数据分析。

Mr.林看小白似懂非懂的样子，就想到结合小白的喜好，以吸引她的兴趣及注意力：这样，我给你打个比方，数据分析方法论与数据分析法的关系就好比服装制作。

小白还没等Mr.林说完就抢道：数据分析跟服装制作都能扯上关系？有意思，Mr.林，你继续讲。

Mr.林假装生气地说：我正要说呢，明明是你把话抢过去了，你叫我如何往下说？

小白笑嘻嘻地说道：Mr.林别生气，我这不是一听到做衣服就激动嘛。您大人不计小人过咯！

Mr.林见小白已经认错了：好吧，谁让我心太软呢。

服装制作首先需要进行服装设计，设计服装样式，并画出服装设计图，然后裁缝师傅就按照设计好的服装设计图制作服装。

服装制作期间涉及如裁剪、缝纫、熨烫等工序，都需要用到不同的专业工具及技术。

- ★ 专业工具有剪刀、缝纫机、电熨斗等。
- ★ 裁剪专业的技术有平面裁剪、立体裁剪等。
- ★ 缝纫专业的技术有合缝、包缝、骑缝等。
- ★ 熨烫专业的技术有压烫、吸烫、坐烫等。

数据分析方法论好比服装设计图，它为我们的数据分析工作指引方向，数据分析法好比服装制作的技术，它为完成各个数据的分析提供技术保障与支持，如图2-1所示。

	数据分析	服装制作
方法论	SWOT、4P、逻辑树等分析思路	服装设计图
工具	EXCEL、SPSS、SAS等	剪刀、缝纫机、电熨斗等
技术	交叉分析、相关分析、回归分析、聚类分析等	平面、立体裁剪等 合缝、包缝、骑缝等 压烫、吸烫、坐烫等

图2-1 数据分析与服装制作对比图

小白：Mr.林，经您这么打比方，我对这个数据分析方法论有了更进一步的理解了，基本上已明白是怎么回事，这个比方真是恰如其分。

2.1.2 数据分析方法论的重要性

Mr.林：很多人在做数据分析时，经常遇到这几个问题：不知从哪方面入手开展分析；分析的内容和指标常常被质疑是否合理、完整，而自己也说不出个所以然来。对这些问题常常感到

困扰。

这就是我为什么强调数据分析方法论的原因，刚才也说过数据分析方法论主要用来指导数据分析师进行一次完整的数据分析，而只有在营销、管理等方法和理论的指导下，结合实际业务情况，才能确保数据分析维度的完整性，分析结果的有效性及正确性。

数据分析方法论主要有以下几个作用。

- ★ 理顺分析思路，确保数据分析结构体系化。
- ★ 把问题分解成相关联的部分，并显示它们之间的关系。
- ★ 为后续数据分析的开展指引方向。
- ★ 确保分析结果的有效性及正确性。

如果没有数据分析方法论的指导，整个数据分析报告虽然各方面都涵盖到，但会给人感觉还缺点什么。其实就是报告主线不明，各部分的分析逻辑不清。

在第一节课的时候我就跟你说过，一份好的分析报告，首先需要有好的分析框架，并且图文并茂，层次明晰，能够让阅读者一目了然。结构清晰、主次分明可以使阅读者正确理解报告内容；图文并茂，可以令数据更加生动活泼，提高视觉冲击力，有助于阅读者更形象、直观地看清楚问题和结论，从而产生思考。

小白：确实是这样的，就跟我们写毕业论文一个道理，都需要结构清晰，从提出问题、分析问题到解决问题，始终围绕核心问题展开论述。

Mr.林：对，一样的道理。我们接下来重点介绍数据分析方法论，也就是如何用数据分析方法论指导我们确定分析思路，进而确定需要分析的内容或指标。

2.2 常用的数据分析方法论

2.2.1 PEST分析法

Mr.林：小白，首先要考考你，你知道有哪些营销管理模型吗？

小白为难地说道：不好意思，我不太了解这方面的知识。

Mr.林：好吧，那我来告诉你几个重要的理论。营销方面的理论模型有4P、用户使用行为、STP理论、SWOT等，而管理方面的理论模型有PEST、5W2H、时间管理、生命周期、逻辑树、金字塔、SMART原则等。这些都是经典的营销、管理方面的理论，需要在工作中不断实践应用，你才能体会其强大的作用。

我们就挑选其中的PEST、5W2H、逻辑树、4P、用户使用行为这五个比较经典实用的理论，谈

谈如何在建立数据分析框架时应用它们作指导。

Mr.林: 首先介绍的是PEST分析法。

PEST分析法用于对宏观环境的分析。宏观环境又称一般环境，是指影响一切行业和企业的各种宏观力量。对宏观环境因素作分析时，由于不同行业和企业有其自身特点和经营需要，分析的具体内容会有差异，但一般都应对政治（Political）、经济（Economic）、技术（Technological）和社会（Social）这四大类影响企业的主要外部环境因素进行分析，如图2-2所示，这种方法简称为PEST分析法。

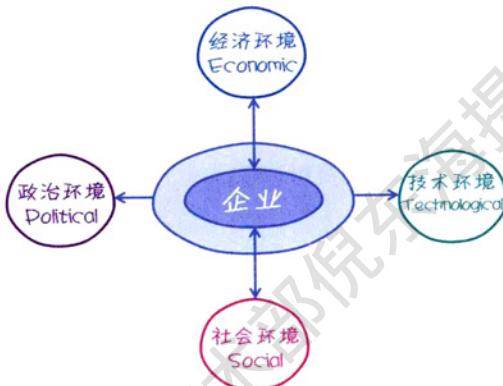


图2-2 PEST分析法示例

◎ 政治环境

政治环境包括一个国家的社会制度，执政党的性质，政府的方针、政策、法令等。不同的国家有不同的社会性质，不同的社会制度对组织活动有不同的限制和要求。即使社会制度不变的同一国家，在不同时期，由于执政党的不同，其政府的方针特点、政策倾向对组织活动的态度和影响也是不断变化的。

构成政治环境的关键指标有：政治体制、经济体制、财政政策、税收政策、产业政策、投资政策、专利数量、国防开支水平、政府补贴水平、民众对政治的参与度等。

◎ 经济环境

经济环境主要包括宏观和微观两个方面的内容。宏观经济环境主要指一个国家的国民收入、国民生产总值及其变化情况，以及通过这些指标反映的国民经济发展水平和发展速度。

微观经济环境主要指企业所在地区或所服务地区的消费者的收入水平、消费偏好、储蓄情况、就业程度等因素，这些因素直接决定着企业目前及未来的市场大小。

构成经济环境的关键指标有：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

社会环境包括一个国家或地区的居民受教育程度和文化水平、宗教信仰、风俗习惯、审美观点、价值观念等。文化水平会影响居民的需求层次；宗教信仰和风俗习惯会禁止或抵制某些活动的进行；价值观念会影响居民对组织目标、组织活动以及组织存在本身的认可；审美观点则会影响人们对组织活动内容、活动方式以及活动成果的态度。

构成社会文化环境的关键指标有：人口规模、性别比例、年龄结构、出生率、死亡率、种族结构、妇女生育率、生活方式、购买习惯、教育状况、城市特点、宗教信仰状况等因素。

◎ 技术环境

技术环境除了要考察与企业所处领域直接相关的技术手段的发展变化外，还应及时了解：

- ★ 国家对科技开发的投资和支持重点。
- ★ 该领域技术发展动态和研究开发费用总额。
- ★ 技术转移和技术商品化速度。
- ★ 专利及其保护情况等。

构成技术环境的关键指标有：新技术的发明和进展、折旧和报废速度、技术更新速度、技术传播速度、技术商品化速度、国家重点支持项目、国家投入的研发费用、专利个数、专利保护情况等因素。

现在我们以中国互联网行业分析为例，采用PEST分析法整理分析思路，构建中国互联网行业分析框架。如图2-3所示，根据PEST分析法列出了进行中国互联网行业分析所需要了解的一些背景，如中国网民与中国公民在人口规模、性别比例、年龄结构、人口分布、生活方式、购买习惯、教育状况、城市特征、宗教信仰状况等方面是否有区别。此处仅为方法使用示例，并不代表互联网行业分析只需要作这几方面的分析，还可根据实际情况进一步调整和细化相关分析指标。

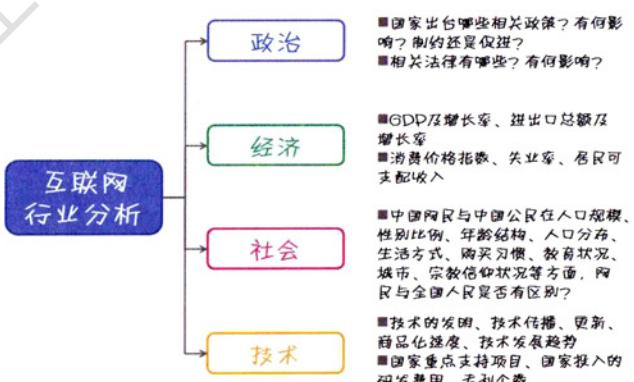


图2-3 用PEST分析法对互联网行业进行分析示例

2.2.2 5W2H分析法

Mr.林: 第二个分析理论就是5W2H分析法。

在职场中，有很多人逻辑思路很清晰，但是也有些人说话老是抓不住重点，费口舌讲了半天，大家还是听不懂他在说什么，这种人就是缺乏逻辑训练。在所有逻辑思考方法中，5W2H分析法可以说是最容易学习和操作的方法之一。

5W2H分析法是以五个W开头的英语单词和两个H开头的英语单词进行提问，从回答中发现解决问题的线索，即何因（Why）、何事（What）、何人（Who）、何时（When）、何地（Where）、如何做（How）、何价（How much），这就构成了5W2H分析法的总框架，如图2-4所示。



图2-4 5W2H分析法示例

该方法简单、方便，易于理解和使用，富有启发意义，广泛用于企业营销、管理活动，对于决策和执行性的活动措施非常有帮助，也有助于弥补考虑问题的疏漏。其实对任何事情都可以从这七大方面去思考，对于不善分析问题的人，只要多练习即可上手，所以同样它也适用于指导建立数据分析框架。

现在以用户购买行为分析为例，来学习5W2H分析法。例如我们需要了解公司产品的用户购买行为是怎样的。这时可在5W2H分析法的指导下整理分析用户购买行为的思路，建立用户购买行为分析框架。如图2-5所示，根据5W2H分析法列出了对用户购买行为的分析所需要了解的一些情况，比如用户购买的目的是什么，公司产品在什么方面吸引了用户等问题。

确定了分析框架后，我们再根据分析框架中的这些问题形成可量化的指标进行衡量和评价，例如月均购买次数、人均购买量、再次购买平均间隔时长等。

Mr.林: 小白，5W2H分析法是不是让你觉得分析某个问题变得简单多了？

小白: 是的，这个方法在分析问题的时候真好用，有了它的指导，我就知道要从哪几个方面思考、分析问题了。



图2-5 5W2H分析法在用户购买行为分析上的应用

Mr.林：同PEST分析法一样，这里举的例子并不代表用户购买行为就只有图2-5所示的这些内容，要做到具体问题具体分析，在使用的时候根据实际情况进行应用。此处举例只是起到抛砖引玉的作用，开拓你的分析思路。

2.2.3 逻辑树分析法

Mr.林：逻辑树又称问题树、演绎树或分解树等。

逻辑树是分析问题最常使用的工具之一，它是将问题的所有子问题分层罗列，从最高层开始，并逐步向下扩展。

把一个已知问题当成树干，然后开始考虑这个问题和哪些相关问题有关。每想到一点，就给这个问题所在的树干加一个“树枝”，并标明这个“树枝”代表什么问题，如图2-6所示。

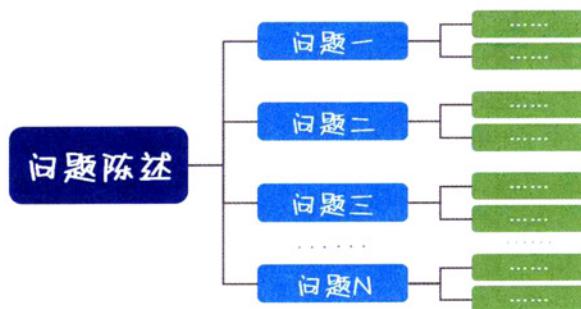


图2-6 逻辑树分析法示例图

辑树的作用主要是帮助你理清自己的思路，避免进行重复和无关的思考。

逻辑树能保证解决问题的过程的完整性，它能将工作细分为便于操作的任务，确定各部分的优先顺序，明确地把责任落实到个人。

逻辑树的使用必须遵循以下三个原则。

- ★ 要素化：把相同问题总结归纳成要素。
- ★ 框架化：将各个要素组织成框架，遵守不重不漏的原则。
- ★ 关联化：框架内的各要素保持必要的相互关系，简单而不孤立。

利用逻辑树分析法，同样可以理清分析思路。例如我们需要进行公司利润下降的专题研究，可采用图2-7所示的框架进行数据分析。当然，要根据自己公司的实际情况进行调整，也就是具体问题具体分析。

不过逻辑树分析法也有它的缺点，就是涉及的相关问题可能有遗漏，虽然可以用头脑风暴法把涉及的问题总结归纳出来，但还是难以避免存在考虑不周全的地方。所以在使用逻辑树的时候，尽量把涉及的问题或要素考虑周全。

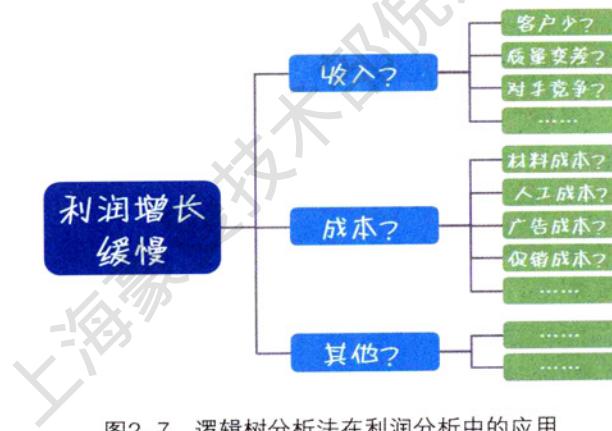


图2-7 逻辑树分析法在利润分析中的应用

2.2.4 4P营销理论

Mr.林：4P营销理论产生于20世纪60年代的美国，它是随着营销组合理论的提出而出现的。营销组合实际上有几个要素，这些要素可以概括为4类：产品（Product）、价格（Price）、渠道（Place）、促销（Promotion），即著名的4P营销理论，如图2-8所示。

- ★ 产品（Product）：从市场营销的角度来看，产品是指能够提供给市场，被人们使用和消费并满足人们某种需要的任何东西，包括有形产品、服务、人员、组织、观念或它们的组合。

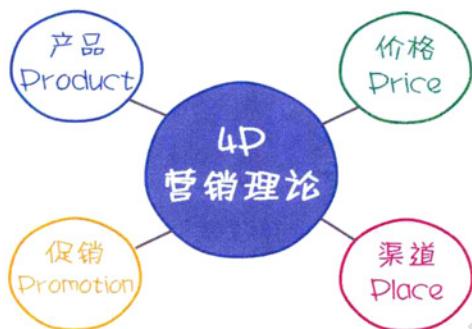


图2-8 4P营销理论示例

- ★ 价格 (Price)：是指顾客购买产品时的价格，包括基本价格、折扣价格、支付期限等。价格或价格决策关系到企业的利润、成本补偿，以及是否有利于产品销售、促销等问题。影响定价的主要因素有三个：需求、成本与竞争。最高价格取决于市场需求，最低价格取决于该产品的成本费用，在最高价格和最低价格的幅度内，企业能把这种产品价格定多高则取决于竞争者的同种产品的价格。
- ★ 渠道 (Place)：是指产品从生产企业流转到用户手上的全过程中所经历的各个环节。
- ★ 促销 (Promotion)：是指企业通过销售行为的改变来刺激用户消费，以短期的行为（比如让利，买一送一，营销现场气氛等等）促成消费的增长，吸引其他品牌的用户或导致提前消费来促进销售的增长。广告、宣传推广、人员推销、销售促进是一个机构促销组合的四大要素。

如果需要了解公司的整体运营情况，就可以采用4P 营销理论对数据分析进行指导，这样就可以较为全面地了解到公司的整体运营情况。现在就以4P 营销理论为指导，搭建公司业务分析框架。

搭建好的公司业务分析框架如图2-9所示。同样，根据这些确定的问题，可再将它们细化为数据分析指标。比如公司提供哪些产品或服务，我们就可用公司产品构成来表示。

Mr.林：小白，还是那句话，具体问题具体分析。不是所有公司业务分析的内容都仅限于图2-9列出的这些，这里只是通过举例说明4P营销理论的指导作用，而在做公司业务分析的时候，需要根据实际业务情况进行调整，灵活运用，切忌生搬硬套。只有深刻理解公司业务的同时才能较好地进行业务方面的数据分析，否则将脱离业务实际，得出无指导意义的结论，犹如纸上谈兵，甚至贻笑大方。

小白连连点头并在职场日记中写下：具体问题具体分析，灵活运用，切忌生搬硬套。



图2-9 4P营销理论在公司业务分析中的应用

2.2.5 用户行为理论

Mr.林：网站分析的发展已经较为成熟，有一套成熟的分析指标。比如IP、PV、页面停留时间、跳出率、回访者、新访问者、回访次数、回访相隔天数、流失率、关键字搜索、转化率、登录率，等等。遇到这么多指标，所有的指标都要采用吗？什么指标该采用？什么指标又不该采用？各指标之间有何联系？哪个指标先分析？哪个指标后分析？

小白做晕菜状：这么多问题！不行了，彻底晕了！

Mr.林：所以我们需要梳理它们之间的逻辑关系，比如利用用户使用行为理论进行梳理。小白，这就是我们要学习的另外一个理论，即用户使用行为理论，也是非常实用的数据分析指导理论之一。

用户使用行为是指用户为获取、使用物品或服务所采取的各种行动，用户对产品首先需要有一个认知、熟悉的过程，然后试用，再决定是否继续消费使用，最后成为忠诚用户。用户使用行为的完整过程，如图2-10所示。

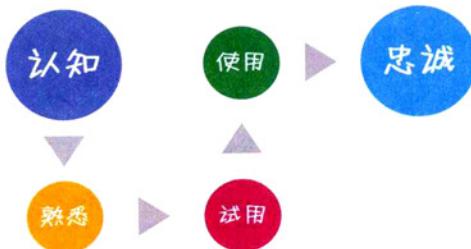


图2-10 用户使用行为的轨迹示例图

现在我们可利用用户使用行为理论，梳理网站分析的各关键指标之间的逻辑关系，构建符合

公司实际业务的网站分析指标体系，如图2-11所示。

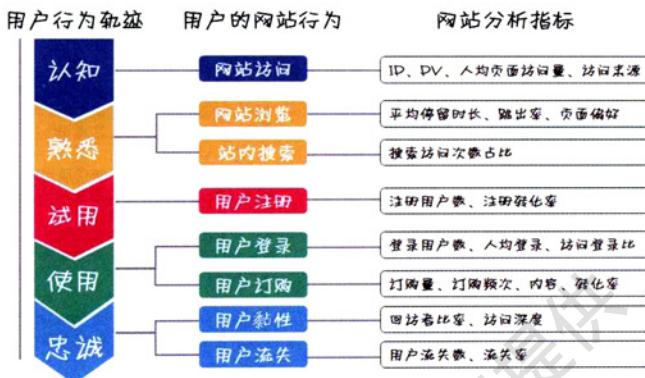


图2-11 用户使用行为理论在网站分析中的应用

Mr.林：小白，这个方法同样需针对具体问题再具体分析，灵活运用，此处就不再赘述了。

小白：好的。

2.3 本章小结

Mr.林缓了缓，接着说：我们前面讲解了的几种数据分析方法论，你可以根据实际情况选择所需的方法论。下面和昨天一样，我们一起作个简要的回顾。

- ★ PEST分析理论主要用于行业分析。
- ★ 4P分析理论主要用于公司整体经营情况分析。
- ★ 逻辑树分析理论可用于业务问题专题分析。
- ★ 用户行为理论的用途较单一，就是用于用户行为研究分析。
- ★ 5W2H分析理论的用途相对广泛，可用于用户行为分析、业务问题专题分析等。

当然，这些方法论也可以相互嵌套使用。例如，用逻辑树分析法搭建分析框架，而下一层级的问题可以从4P的角度分析，也可以用5W2H法分解问题。记住根据实际情况灵活选择使用，切勿生搬硬套。

方法论不限于刚才介绍的几种，除此之外，还有金字塔法、生命周期理论等，对于这些方法的应用，需要你根据自己所在的行业特征在实践中摸索，前提是需要了解行业知识、公司业务，以及相关的营销管理模型，只有把它们有机地结合使用，才能指导数据分析工作有序开展，才能确保数据分析结果具有指导意义。

小白：太棒了，经您这么一介绍，理论指导实践不再那么飘渺虚幻，再多多复习，我就可以自己“设计服装图纸”了。

第3章

无米难为巧妇——数据准备

理解数据

数据来源



这一天小白被牛董责备做事效率低，整理一份报表都要折腾个半天，十分郁闷。她还记得牛董丢下的最后一句话：“现如今是数据、报表、PPT满天飞的商业时代，不懂数据怎么混？有空你去跟Mr.林学学怎么整数据！”

调整好情绪，小白就奔着Mr.林的办公桌来了：今天牛董特意让我来向您请教如何玩转数据。不知道您是否有时间指点指点我？

Mr.林爽快地答应了：没问题呀！不用说“请教”，太客气了。我们就继续一起来聊聊数据吧！俗话说“巧妇难为无米之炊”，小白，你应该听过这句谚语吧？

小白不解：听过啊！怎么啦？这跟我们的数据分析有啥关系？

Mr.林：不仅有关系，而且它们的关系密切着呢！

首先，数据就好比谚语中的“米”。做数据分析必须要有数据吧，从一开始的数据收集、数据处理、数据分析都离不开数据；而做饭呢，同样先要买米，洗米，然后煮米饭，你看看数据和米在数据分析和做饭的流程中作用类似吧！

其次，数据分析师就好比要做饭的小媳妇。数据分析师要对数据进行分析，小媳妇在家就要做饭咯！

最后，数据的处理与分析就好比煮米饭。做饭需要炊具，做数据分析也要有分析工具。我们这里选择最大众的工具——Excel，它是学习和工作中最常见的办公软件之一，而且非常容易掌握，所以下续我介绍的数据处理、分析相关的操作都用它来完成。

小白兴奋地说道：经你这么一说，它们的关系还真是非常密切，现在我对数据分析的了解又进了一步。数据分析就好比婚姻一样，外面的人看着甜蜜、浪漫、风光，可里面的人才知道都是些柴米油盐酱醋茶的琐碎事儿。

Mr.林：对！小白你还没结婚咋就有这么大的感触？呵呵。

小白顿时羞得脸上红云阵阵。

Mr.林接着说道：今天，我们学习以下知识。

- ★ “米”的构造、种类和要求——理解数据。
- ★ “米”从哪里来——数据来源。

小白精神抖擞地说：好！

3.1 理解数据

Mr.林：扎实的数据分析基本功不单是指会使用数据分析工具，更重要的是对数据有深入认识和解读。很多人一开始并不能够清晰地认识到数据分析对数据有什么要求。正是因为如此，从事数据分析相关工作时，才会有比较迷茫、无从下手的感觉。因此，对数据的理解是数据分析的

一个重要前提。

小白: 被你说中了, 我现在还真不知道数据到底长什么模样?

Mr.林: 那好, 千里之行始于足下, 我们现在就来看看数据都长什么模样。

3.1.1 字段与记录

Mr.林: 小白, 你刚从学校走进社会, 就用一个你最熟悉的例子吧。先回想一下学生时代老师手里那份成绩表, 里面不仅能看到自己的成绩, 还能看到其他同学的成绩, 以及整个学期的总分。回想一下这份成绩表是不是如图3-1所示的这样。

小白: 成绩表都差不多这样啦。

学号	姓名	性别	语文	数学	英语	总分	总评
1	赵鹏	男	89	85	84	258	良好
2	郭南	女	90	88	95	273	优秀
3	杨洋	男	75	75	78	228	及格
4	楚中天	男	78	82	70	230	及格
5	邓柏涵	女	83	90	80	253	良好
...

图3-1 某学期学生考试成绩表

Mr.林: 在这份成绩表里, 从横向看, 每一行都是同学的基本情况和成绩; 从纵向看, 每一列都描述了一类数据, 例如第3列是每位同学的性别资料, 第4列是每位同学的语文成绩, 等等。

这样的成绩表从数据分析的角度来看, 就是一个典型的数据库。成绩表最上面的“学号”“姓名”“性别”“总分”等被称为字段, 字段是数据库中的说法, 而每位同学的基本情况和成绩就构成了一条一条的数据记录, 如图3-2所示。



学号	姓名	性别	语文	数学	英语	总分	总评
1	赵鹏	男	89	85	84	258	良好
2	郭南	女	90	88	95	273	优秀
3	杨洋	男	75	75	78	228	及格
4	楚中天	男	78	82	70	230	及格
5	邓柏涵	女	83	90	80	253	良好
...

图3-2 某学期学生考试成绩表(字段与记录)

从数据分析的角度来理解字段和记录的概念:

- ★ 字段是事物或现象的某种特征。比如成绩表中的“学号”“姓名”“总分”等都是字段, 在统计学中称为变量。
- ★ 记录是事物或现象某种特征的具体表现。比如成绩表中的“性别”可以是男或女, “总分”可以是273或者230等, 记录也称为数据或变量值。

小白：明白了，原来数据需要由字段与记录共同组合而成。

3.1.2 数据类型

Mr.林：小白，看完成绩表，我们再来看下一个例子，职工信息表，如图3-3所示。你来看看，表中的数据大概可以分为几种类型？

员工号	姓名	性别	部门	入职日期	工龄
A00006	张三	男	IT部	1970-06-20	17
A00431	赵四	男	市场部	1962-07-14	25
A07520	王二	女	设计部	1979-12-31	18
A01402	周五	女	开发部	1974-12-01	13
A02700	田七	男	销售部	1978-12-29	9
A09943	李九	男	后勤部	1999-03-20	8

图3-3 职工信息表

小白：我瞧瞧，有数字、文字、日期，差不多就这三类吧。

Mr.林：对！我们最常用的数据类型就是你说的这三类，当然还有货币、时间等其他数据类型。有一个方法可以知道数据到底有多少类型，就是在Excel中用单元格格式查看所有的数据类型，操作步骤如下：

- STEP 01** 选择Excel中的任意一个单元格，单击鼠标右键，在弹出的菜单中选择“设置单元格格式”，会出现“设置单元格格式”的对话框，如图3-4所示。
- STEP 02** 在这个对话框中，可以看到各种不同的数据类型，比如：数值、文本、日期、货币、会计专用、时间、百分比、分数、科学计数、特殊等。

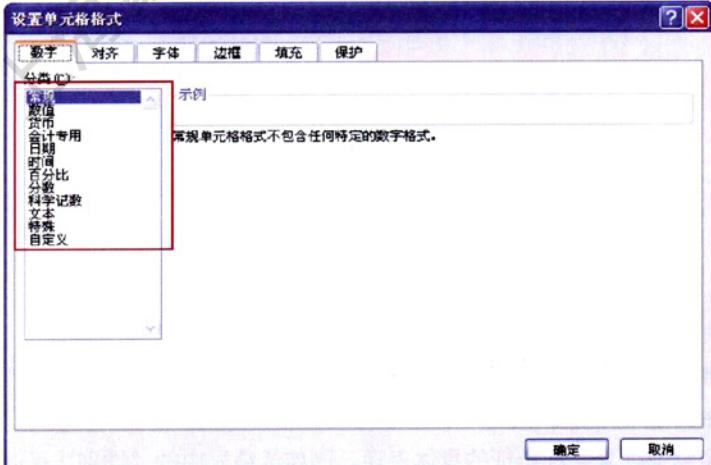


图3-4 “设置单元格格式”对话框

Excel中虽然有这么多数据类型，但最终都可以归结为两大类。

◎ 字符型数据

字符型数据是不具有计算能力的文字数据类型。它包括中文字符、英文字符、数字字符（非数值型）等字符。例如前面成绩表中的“姓名”“性别”“总评”三个变量均为字符型数据，职工信息表中的“员工号”“姓名”“性别”“部门”四个变量均为字符型数据。

◎ 数值型数据

数值型数据是直接使用自然数或度量单位进行计量的数值数据。例如成绩表中的“语文”“数学”“英语”三科成绩汇总即得到“总分”这个字段，而职工信息表中，可按“入职日期”计算工龄，这些数据就是数值型数值。对于数值型数据，我们可直接用算术方法进行汇总和分析，这一点是区分数据是否属于数值型数据的重要依据。

3.1.3 数据表

Mr.林：小白，刚才介绍了字段与记录，还有数据的类型，我们现在就来看看由字段、记录和数据类型构成的数据表。数据分析所需要的数据表，也是有一定要求的。

一张数据表的制作，可以侧面反映出制作者的数据沉淀及应用水平。如果数据表里的合并单元格较多，设计不合理，这基本属于为了满足一时之需的数据表，并不符合数据分析的要求，没有为后续数据分析做长远考虑与规划。

所以千万别忽视基础数据表格的设计，数据表的设计是否合理，关系着后期数据分析的效率及深度。数据表设计具体要求如图3-5所示。

序号	要求
1	数据表由标题行和数据部分组成
2	第一行是表的列标题（字段名），列标题不能重复
3	第二行起是数据部分，数据部分的每一行数据称为一个记录，并且数据部分不允许出现空白行和空白列
4	数据表中不能有合并单元格存在
5	数据表与其他数据之间应该留出至少一个空白行和一个空白列
6	数据表需要以一维的形式存储，但是在实际操作中接的数据往往是以二维表格的形式存在的，此时应将二维表转化为一维表的形式存储数据

图3-5 数据表的设计要求

小白仔细阅读完，问道：什么是一维表，什么是二维表？

>> 谁说菜鸟不会数据分析（入门篇）

Mr.林想了想，快速地在Excel中做了两个表：你看我刚刚绘制的图3-6，是我国五省2006~2008年的国内生产总值（GDP）的一维表与二维表。这两个表有什么区别？

小白分析道：左边的二维表中，北京在2006年的GDP是7861；右边的一维表中“地区”为北京，“年份”为2006，对应的GDP是7861。哦，我知道了！一维表的列标签是字段，而且表中每个指标就对应一个取值。例如图3-6中的一维表的第一行，“地区”对应的是北京，“年份”对应的是2006，GDP对应的是7861。而二维表的列标签是数据：2006年、2007年、2008年，将一维表中的所有年份真实值都放置在列标签里了。

二维表			一维表			
地区	2006年	2007年	2008年	地区	年份	GDP
北京	7861	9353	10488	北京	2006	7861
上海	10366	12189	13698	北京	2007	9353
山东	22077	25966	31072	北京	2008	10488
广东	26160	31084	35696	上海	2006	10366
浙江	15743	18780	21487	上海	2007	12189
				上海	2008	13698
				山东	2006	22077
				山东	2007	25966
				山东	2008	31072
				广东	2006	26160
				广东	2007	31084
				广东	2008	35696
				浙江	2006	15743
				浙江	2007	18780
				浙江	2008	21487

图3-6 一维表与二维表

Mr.林：嗯，你说的差不多，基本上是这样。我来完善一下，这里的“维”指的是分析数据的角度，2006年、2007年、2008年，从数据的角度来说，应该都是“年份”的范畴，是描述各省GDP的一个因素，若要换成一维表，则应该使用同一个字段，将年份单独作为列标签。这是区分左边的“二维表”和右边的“一维表”的关键所在。

小白继续问道：怎样操作才可以将二维表转化为一维表呢？

Mr.林笑道：嘿嘿！我料定你会问这个问题。我们现在就以图3-7中的二维表数据为例，看看如何利用数据透视表进行二维表到一维表的转换。

	A	B	C	D
1	地区	2006年	2007年	2008年
2	北京	7861	9353	10488
3	上海	10366	12189	13698
4	山东	22077	25966	31072
5	广东	26160	31084	35696
6	浙江	15743	18780	21487

图3-7 二维表数据示例

在转换过程中我们要用到数据透视表中的“数据透视表和数据透视图向导”功能，Excel 2007与2010版本并没有在选项卡中直接给出该功能，需要我们自己把它给请出来。采用“添加功

能选项”方式，就能把“数据透视表和数据透视图向导”功能添加至选项卡中。

二维表转换为一维表的具体操作如下。

STEP 01 先添加“数据透视表和数据透视图向导”功能。单击“文件”选项卡→“选项”→“自定义功能区”。在“从下列位置选择命令”中选择“不在功能区中的命令”，找到“数据透视表和数据透视图向导”并选中，在“自定义功能区”选择“主选项卡”，在“数据”主选项卡下面添加“新建组”并选中它，再单击“添加”按钮，即可把“数据透视表和数据透视图向导”添加到“数据”主选项卡的“新建组”中，如图3-8所示。

STEP 02 单击刚添加好的“数据透视表和数据透视图向导”图标，会弹出如图3-9所示的对话框，在数据源类型中选中“多重合并计算数据区域”，并单击“下一步”按钮。



图3-8 “数据透视表和数据透视图向导”图标

图3-9 数据透视表和数据透视图向导——步骤1

其实要打开数据透视表和数据透视图向导，还有更简单快捷的方式——使用快捷键“Alt+D+P”。先按“Alt+D”组合键，然后松开，再按字母键“P”，即可打开“数据透视表和数据透视图向导”对话框。

STEP 03 选中“创建单页字段”，并单击“下一步”按钮，如图3-10所示。

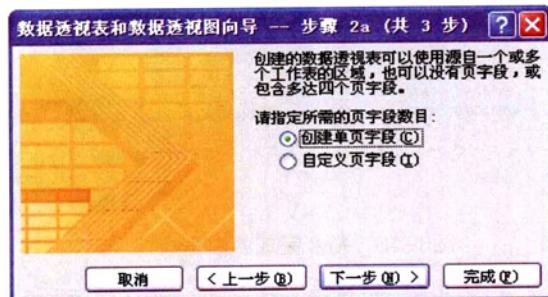


图3-10 数据透视表和数据透视图向导——步骤2a

>> 谁说菜鸟不会数据分析（入门篇）

- STEP 04** 在“选定区域”项中选择整个二维表的数据区域“二维表!\$A\$1:\$D\$6”，单击“添加”按钮，并单击“下一步”按钮，如图3-11所示。

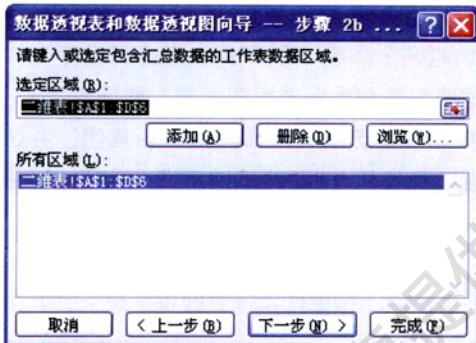


图3-11 数据透视表和数据透视图向导——步骤2b

- STEP 05** 在“数据透视表显示位置”选择“新建工作表”项，然后单击“完成”按钮，如图3-12所示，即可完成数据透视表的创建。创建的数据透视表如图3-13的左边所示。

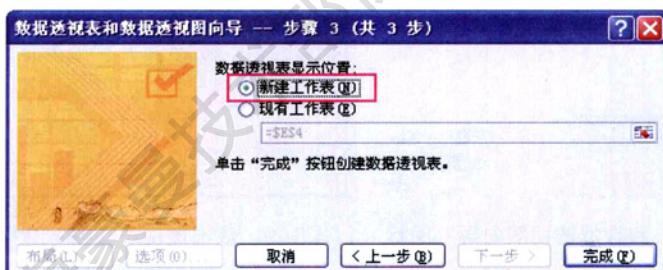


图3-12 数据透视表和数据透视图向导——步骤3

	A	B	C	D	E
1	页1	(全部)			
2					
3	求和项:值	列标签			
4	行标签	2006年	2007年	2008年	总计
5	北京	7861.04	9353.32	10488.03	27702.39
6	广东	26159.52	31084.4	35696.46	92940.38
7	山东	22077.36	25965.91	79115.33	
8	上海	10366.37	12188.85	36253.37	
9	浙江	15742.51	18780.44	56009.87	
10	总计	82206.8	97372.92	1	292021.34
11					
12					
13					
14					

图3-13 初步完成的数据透视表

- STEP 06** 在“数据透视表字段列表”窗格中，取消对“选择要添加到报表的字段”列表中的“列”和“行”字段的勾选，即将“列”和“行”前面的钩号“√”去掉。或用鼠标

标拖动法，将列标签里的“列”及行标签里的“行”拖回至“选择要添加到报表的字段”列表中，如图3-13右边“数据透视表字段列表”窗格中的箭头所示。经过“行”“列”字段移除整理后的数据透视表如图3-14所示。

STEP 07

双击图3-14中所示的B4单元格（唯一的汇总数据），Excel会自动创建一个新工作表，并基于原二维表数据源生成新的一维表，如图3-15所示。

	A	B
1	页1	(全部)
2		
3	求和项:值	汇总
4	汇总	292021.3

图3-14 移除行和列标签的数据透视表

	A	B	C	D
1	行	列	值	页1
2	北京	2006年	7861.04	项1
3	北京	2007年	9353.32	项1
4	北京	2008年	10488	项1
5	上海	2006年	10366.4	项1
6	上海	2007年	12188.9	项1
7	上海	2008年	13698.2	项1
8	山东	2006年	22077.4	项1
9	山东	2007年	25965.9	项1
10	山东	2008年	31072.1	项1
11	广东	2006年	26159.5	项1
12	广东	2007年	31084.4	项1
13	广东	2008年	35696.5	项1
14	浙江	2006年	15742.5	项1
15	浙江	2007年	18780.4	项1
16	浙江	2008年	21486.9	项1

图3-15 由数据透视表生成的一维表

此外，直接双击图3-13所示表中的最后一个单元格E10也能达到同样的效果。最后，把数据表的列标题（字段名）改为相应的字段名称即可。

小白惊讶地说：好神奇呀！

Mr.林笑道：这里先简单介绍数据透视表的二维表到一维表转换功能，关于数据透视表的具体功能等讲到数据分析工具时再为你介绍。

小白：好的，数据透视表真是强大。

3.2 数据来源

Mr.林：小白，刚才我们已经介绍了数据特点及要求，下面就来介绍如何获取数据。获取数据的方式可以分为两种：导入外部数据和自己录入数据，我们先说说导入外部数据吧。

小白：好的。

3.2.1 导入数据

Mr.林：导入的外部数据最常见的来源有两种：文本和网站数据来源。文本数据是比较常见的数据来源，所以，我先教你如何将文本文件导入Excel。

>> 谁说菜鸟不会数据分析（入门篇）

① 导入文本数据

在员工满意度调查中我们有一份文本型的“问卷录入结果”文件，现在以导入这份文件为例，实践一下导入文本数据的具体操作。

STEP 01 鼠标单击“数据”选项卡，选择“自文本”选项，Excel会自动弹出相对应的对话框。

STEP 02 找到保存的“问卷录入结果.txt”文件。

还有一种方式是，在“文件”选项卡中，选择“打开”选项，Excel会自动弹出对话框，在“文件类型”中选择“文本文件”，然后，从要打开的文本文件所在位置打开“问卷录入结果.txt”文件。此时，将会弹出“文本导入向导”对话框，如图3-16所示。

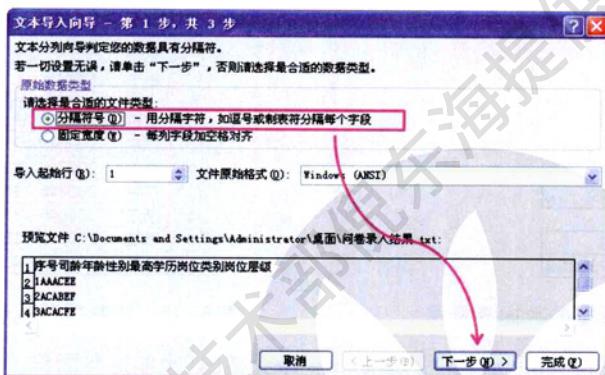


图3-16 文本导入向导-第1步

STEP 03 在图3-16所示的对话框中有两个选项“分隔符号”和“固定宽度”。如果文本文件中的列标签以制表符、冒号、分号、空格或其他字符分隔，则选择“分隔符号”；如果你想自己设定每列分隔的具体位置，则选择“固定宽度”。由于我们的文本数据是以Tab键分隔的，所以这里选择“分隔符号”，单击“下一步”按钮，得到如图3-17所示的对话框。

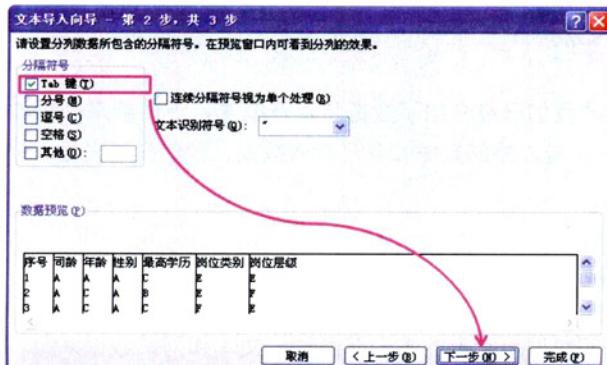


图3-17 文本导入向导-第2步

STEP 04 如图3-17所示，有Tab键、分号、逗号、空格的分隔符可供选择，如果分隔符是其他字符，则选中“其他”复选框，然后在后面的文本框中输入字符。在这里我们选择“Tab键”，单击“下一步”按钮，弹出文本导入向导第3步的对话框，如图3-18所示。

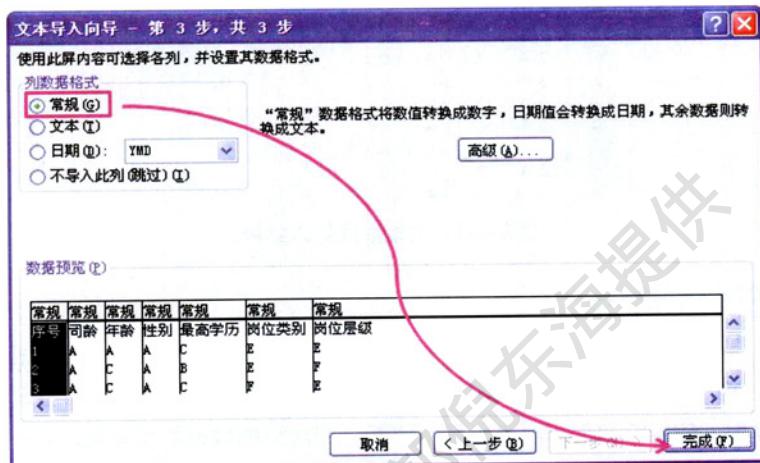


图3-18 文本数据导入向导-第3步

STEP 05 如果不需要将某列导入Excel中，可以在“数据预览”中选择此列，然后勾选“不导入此列（跳过）”，则该列就不会输出在Excel文件里。在这里，我们不需要删除某列，故选择“常规”即可，单击“完成”按钮。

STEP 06 弹出“导入数据”对话框，在其中选择存放数据的位置，单击 OK 按钮，拖动或者缩放单元格区域，再次单击 OK 按钮恢复对话框，最后单击“确定”按钮，如图3-19所示。

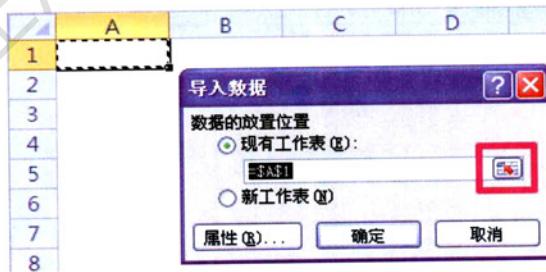


图3-19 导入文本数据

STEP 07 返回工作表，文本文件中的数据就会按所设置的格式自动导入到其中，效果如图3-20所示。

小白念叨道：嗯，文本数据导入步骤我已经记下了，回去我就要动手试一试。

>> 谁说菜鸟不会数据分析(入门篇)

	A	B	C	D	E	F	G
1	序号	司龄	年龄	性别	最高学历	岗位类别	岗位层级
2	1	A	A	A	C	E	E
3	2	A	C	A	B	E	F
4	3	A	C	A	C	F	E
5	4	D	D	A	D	A	F
6	5	A	B	B	C	F	C
7	6	B	C	B	B	F	D
8	7	A	B	A	C	F	D
9	8	B	C	B	C	F	D
10	9	A	B	B	C	E	D
11	10	B	C	A	D	B	G
12	11	A	C	A	B	C	D
13	12	A	B	A	B	F	C
14	13	C	C	A	B	A	C

图3-20 文本数据导入结果

◎ 自动导入网站数据

Mr.林: 除了本地文本数据外, 网络数据源是如今信息时代不可或缺的数据源, 如股票行情、产品报价、销售排行、统计局网站公布的经济数据等。而且, Excel中还设置了“刷新控件”的功能, 即导入的网络数据可以进行即时更新, 当网页的数据发生变化时, 不需要重新导入数据就能获取最新数据。导入网站数据的具体操作步骤如下。

- STEP 01** 单击“数据”选项卡, 选择“自网站”选项, Excel会弹出“新建Web查询”对话框。
- STEP 02** 在对话框的地址栏里输入要导入数据的网址, 这里我们输入国家统计局“消费者信心指数(2010年9月)”数据的网址: http://www.stats.gov.cn/tjsj/jdsj/t20101105_402681367.htm。单击“转到”按钮, 对话框中将显示相应的页面, 单击按钮, 使其图标变为, 在这里单击表格前面的图标, 如图3-21所示。

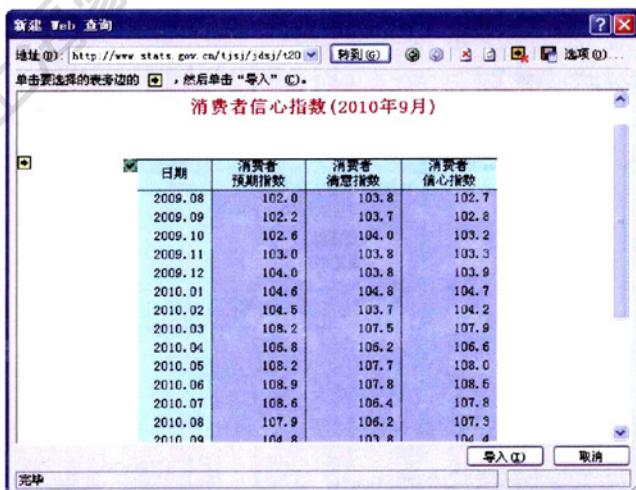


图3-21 导入网站数据

STEP 03 单击“导入”按钮，弹出“导入数据”的对话框，在工作表中选择需要放置数据的区域，这里我们选择默认位置A1，单击“确定”按钮，网站中的数据将自动导入到Excel中，导入后的效果如图3-22所示。

	A	B	C	D
1	日期	消费者	消费者	消费者
2		预期指数	满意指数	信心指数
3	2009.08	102	103.8	102.7
4	2009.09	102.2	103.7	102.8
5	2009.1	102.6	104	103.2
6	2009.11	103	103.8	103.3
7	2009.12	104	103.8	103.9
8	2010.01	104.6	104.8	104.7
9	2010.02	104.5	103.7	104.2
10	2010.03	108.2	107.5	107.9
11	2010.04	106.8	106.2	106.6
12	2010.05	108.2	107.7	108
13	2010.06	108.9	107.8	108.5
14	2010.07	108.6	106.4	107.8
15	2010.08	107.9	106.2	107.3
16	2010.09	104.8	103.8	104.4

图3-22 网站数据导入结果

小白问道：您刚刚说不用打开网页也能自动更新数据，是怎么做到的？

Mr.林：记性真好！其实更新数据的方式有三种，即时刷新、定时刷新或者打开文件时自动刷新。

首先教你怎么实现即时刷新吧！只要单击“数据”主选项卡→“刷新数据”，或者选择导入的外部数据所在区域中的任意一个单元格，然后单击鼠标右键，在弹出的快捷菜单中选择“刷新”命令，如图3-23所示。



图3-23 即时刷新网站数据

>> 谁说菜鸟不会数据分析（入门篇）

怎么实现定时刷新或者打开文件时自动刷新呢？注意，图3-23所示“方法二”的快捷菜单中还有个“数据范围属性”，选择这个选项，即弹出“外部数据区域属性”对话框，如图3-24所示。在此对话框中勾选“刷新频率”复选框，选择刷新的间隔时间，就能实现定时刷新了。

另外，在“外部数据区域属性”对话框中还有“打开文件时刷新数据”复选框，选择这个选项将实现打开文件时数据自动刷新。

小白突然想到了什么，问道：我能不能用先“Ctrl+C”，接着再“Ctrl+V”的方式直接将网页的数据复制粘贴过来呢？这样能实现网页数据自动更新吗？

Mr.林笑了笑：真聪明！这就是我正准备跟你讲的另一种导入网站数据的方法，这种方法比较方便，也能够实现网页数据自动更新，如图3-25所示。

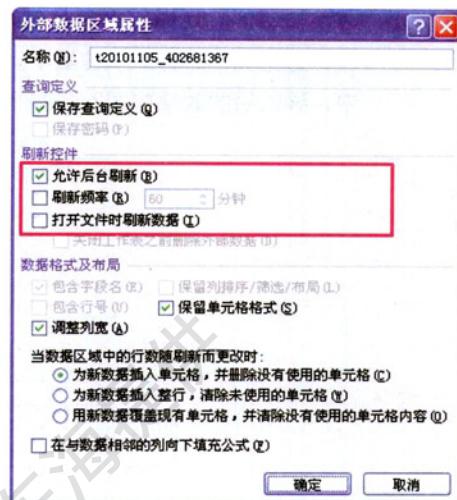


图3-24 外部数据区域属性

A	B	C	D	E	F
日期	消费者预期指数	消费者满意指数	消费者信心指数		
2009.08	102	103.8	102.7		
2009.09	102.2	103.7	102.8		
2009.1	102.6	104	103.2		
2009.11	103	103.8	103.3		
2009.12	104	103.8	103.9		
2010.01	104.6	104.8	104.7		
2010.02	104.5	103.7	104.2		
2010.03	108.2	107.5	107.9		
2010.04	106.8	106.2	106.6		
2010.05	108.2	107.7	108		
2010.06	108.9	107.8	108.5		
2010.07	108.6	106.4	107.8		
2010.08	107.9	106.2	107.3		
2010.09	104.8	103.8	104.4		
17					
18					
19					
20					
21					

图3-25 直接粘贴网站数据

在工作表中粘贴数据后，数据区域的右下角会出现一个“粘贴”按钮 (Ctrl) ▾，如图3-25所示。单击它，选择“可刷新的Web查询”命令，在弹出的对话框中进行相应设置即可。

3.2.2 问卷录入要求

Mr.林：介绍完导入外部数据，我们再来看手工录入，就以问卷录入为例吧。

做数据分析时，除了我刚刚讲的对一般的数据表有特定要求外，我们经常接触到的调查问卷数据，其录入格式也是有讲究的。例如，我们公司最近搞的员工满意度调查，经过问卷的发放、填写、回收、核实和清理之后，就需要把收集回来的选项进行编码、录入。而对于不同类型的问题，也有不同的录入格式要求。小白，我问你问卷题目大致可以分为哪几类呢？

小白机灵地翻出员工满意度调查表，念道：常用的类型主要有单选、多选、排序和开放性文字题这四种类型。

Mr.林：嗯！没错。那么，你知道怎样分辨这几种类型吗？每种类型采取怎样的录入格式呢？

小白：……

Mr.林停顿了一会儿，继续说道：先别急着回答，下面我就以员工满意度调查为例慢慢讲解。

现在我在公司员工满意度问卷中抽取出一份问卷，问卷题目很多，每种题型的题目就分别挑选一道作为示例讲解吧，如图3-26所示。

问卷编号： 405

员工满意度调查问卷

第一类型：数值

3. 您在本公司的入职年月是 2001 年 9 月

第二类型：单选

7. 您认为公司的组织管理框架清晰、分工明确：

A. 完全不符合 B. 不太符合 C. 一般 D. 基本符合 E. 完全符合

第三类型：多选

25. 您认为用什么样的方式能够有效激励您？(最多选 3 项)

A. 提供个人职业发展空间 B. 提供个人能力提升机会 C. 提高薪酬待遇
 D. 及时表彰等精神激励 E. 加强上下级沟通与工作指导 F. 公平透明的管理环境
G. 加强企业文化建设

第四类型：排序

31. 您希望所在的企业有什么样的价值取向，请按重要性排序：

第一重要 () 第二重要 () 第三重要 ()

A. 优胜劣汰 B. 客户至上 C. 团队精神 D. 开拓创新 E. 能者上，庸者下 F. 其他

第五类型：开放性文字题

40. 对公司的其它意见和建议：

建议增加对员工的业务培训

图3-26 公司员工满意度问卷示例

◎ 数值题

在示例问卷中，第3题就是数值题，数值题一般要求被调查者填入相应的数值，或者打分。录入人员只需输入被调查者实际填入的数值即可。

◎ 单选题

单选题的特征就是答案只能有一个选项，所以编码时只需定义一个变量，即给该题留一列进行数据的录入。录入时可采用1、2、3、4分别代表A、B、C、D四个选项，例如选C则录入3。对于示例问卷中的第7题，只需在对该份问卷的记录中对应第7题所在的列位置录入2即可。

◎ 多选题

多选题的特征是答案可以有多个选项，其中又分为项数不定多选和项数限定多选。项数不定多选就是对所选择选项的数目不作限定，项数限定多选有“最多选**项”的要求，如示例问卷中的第25题就对项数有限制。

多选题的录入有两种方式：二分法和多重分类法。

- ★ 二分法：把每一个相应选项定义为一个变量，每一个变量值均作如下定义：“0”代表未选，“1”代表已选，即对于被调查者选中的选项录入1，对未选的选项录入0。比如，示例问卷中被调查者选ACF，则A、B、C、D、E、F、G的选项下分别录入1、0、1、0、0、1、0。
- ★ 多重分类法：事先定义录入的数值，比如1、2、3、4、5、6、7分别代表选项A、B、C、D、E、F、G，并且根据限选的项数确定应录入的变量个数。例如示例问卷第25题限选3项，那么需要设立3个变量，被调查者在该题选ACF，则在3个变量的值分别为1、3、6。

◎ 排序题

对于排序题需要对选项重要性进行排序，比如示例问卷第31题，总共有6个选项，需要按重要程度排出前3名来。排序题的录入与多重分类法类似，先定义录入的数值，1、2、3、4、5、6分别代表选项A、B、C、D、E、F，然后按照被调查者填写的顺序录入选项，因此对于第31题，我们按顺序录入3、4、1。

◎ 开放性文字题

开放性文字题一般都放在问卷的末尾，需要被调查者自己填写一些文字表述观点或建议，例如示例问卷的第40题。对于开放性文字题，如果可能的话可以按照含义相似的答案进行归类编码，转换成为多选题进行分析。如果答案内容较为丰富，不容易归类，就应对这类问题直接做定性分析。

小白忙打断：等等！您就这么走马看花似地说了一遍，我记不住！

Mr.林: 呵呵, 对照这几道题的录入结果, 你就清楚了! 如图3-27所示。

二分法

编号	第3题		第 7 题	第25题							第31题			第40题
	年	月		A	B	C	D	E	F	G	第一 重要	第二 重要	第三 重要	
405	2001	9	2	1	0	1	0	0	1	0	3	4	1	建议增加对员工的业务培训

多重分类法

编号	第3题		第 7 题	第25题			第31题			第40题
	年	月		选项一	选项二	选项三	第一 重要	第二 重要	第三 重要	
405	2001	9	2	1	3	6	3	4	1	建议增加对员工的业务培训

图3-27 问卷录入结果

小白仔仔仔细细地对比了图3-26和图3-27, 然后说道: 哦, 我明白了!

3.3 本章小结

Mr.林缓了缓, 接着说: 巧妇难为无米之炊——数据准备的内容就跟你介绍完了。小白, 现在我们来回顾一下, 今天主要讲了两方面的内容。

- ★ 认识了数据的特点与要求, 它是以字段和记录的形式存储在数据表中, 数据类型主要有字符型和数值型两大类。了解了数据表的六点要求。
- ★ 熟悉了几种数据来源, 学习了数据导入方法与调查问卷常见的五种题型及录入格式。

小白以迅雷不及掩耳盗铃之势接过话: 您介绍的方法我都在本子里记下了!

Mr.林: 嗯, 不错! 不过, 提醒一下, “巧妇难为无米之炊”的下半句是“懒妇不为有米之炊”。做数据分析是个辛苦活儿, 收集一组组数据, 统计一个个指标, 建立一层层关系, 重复一次次检查……都需要我们亲力亲为, 严格把关, 容不得半点疏忽, 这也是之前提到的严谨负责的态度。

所以, 做数据分析首先得克服惰性, 你需要认真贯彻到实践中去, 刚刚所讲的内容都得多操练几遍才能熟练掌握。后面我会详细讲述数据处理的过程, 不过……

Mr.林学着一休哥的语气: 在这之前, 我们先“休息, 休息一会儿”。

第4章

三心二意——数据处理

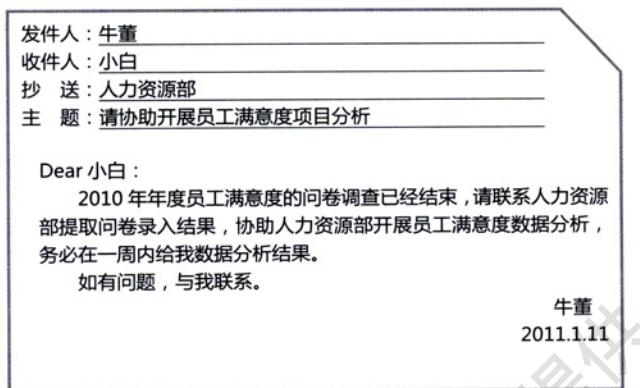
何为数据处理

数据清洗

数据加工

数据抽样





这天刚上班，小白就收到牛董的这封E-mail，没想到任务来得这么快，于是愁眉不展地呆坐了很久。昨天才刚跟Mr.林学了点基础知识，现在任务就来了。数据拿到后该怎么办？要怎么处理？要提炼出哪些信息出来？牛董想看到哪方面的结果？小白觉得一筹莫展，无从下手。没办法，“兵来将挡，水来土掩”，只能再次向Mr.林请教。

4.1 何为数据处理

小白向Mr.林介绍完情况后，Mr.林不慌不忙说道：正好！昨天已经介绍了数据准备，今天准备教你数据处理。呵呵，听完我今天这堂课，完成牛董的任务就是小菜一碟！老规矩，欲练神功先修内功，做数据处理首先要做到“三心二意”。

4.1.1 “三心二意” 处理数据

小白不解：三心二意？开玩笑吧，只听说做事要一心一意，而数据处理怎能三心二意呢？

Mr.林：别急啊！汪涵大叔说，要做个让父母放心，让爱人开心，让领导省心，对女人善意，对男人随意的“三心二意”的好男人。作为一名资深数据分析师，对此我深有体会。做数据处理也要“三心二意”。小白，你要认真听好咯！

◎ 信心

数据分析师都知道，要经过无数次对海量数据的收集整理、逻辑运算、分析汇总和论证检验，才能得出一个简洁的结果。所以，成千上万的数据在数据分析师眼里不过是家常便饭，你首先得有信心，先藐视它们，再干掉它们！

看小白一脸土色，Mr.林继续补充道：我给你讲个老故事压压惊。某小子去少林寺学武，师

傅让他天天拍水缸。小子拍了一周，手掌红肿，师傅说继续；小子拍了一个月，濒临崩溃，师傅还说继续；小子拍了半年，心想你这老头在敷衍我，老子我不练了，拂袖而去。回家，敲门，一掌，门碎。小子哭着回到山中长跪。什么叫信心？信心是指未看见任何未来时，你依然怀抱希望，坚持下去。

Mr.林说的这番话比一罐红牛还奏效，小白顿感精神十足：真给力！唯有相信，才有可能。谢谢师傅！

◎ 细心

Mr.林：数据处理做的可都是细活儿，五大三粗的人干不了！为什么？毫不夸张地说，一个数字能影响一个结果，一个结果能影响一个决策，一个错误的决策能毁掉一个公司！数据分析师有一个公式：

$$1\% \text{ 的错误} = 100\% \text{ 的失败}$$

“千里之堤，溃于蚁穴”在数据分析中的例子比比皆是，所以，我们绝对不能对任何一个细微之处掉以轻心。另一方面，数据分析师还要对差异非常敏感，一个异常值可能就是问题的症结所在。

◎ 平常心

当别人在酒吧喝酒、唱歌，而你还在加班，一遍又一遍地核对数据时；当你辛辛苦苦处理数据，Excel很久都没有响应时；当你废寝忘食制作报表，即将接近尾声，电脑突然死机或断电时……你还能保持淡定吗？

所以，处理数据要有一颗平常心——平淡、平等、平静地对待问题。加班加点时不能急躁焦灼，遇到问题时不能退缩放弃，找出原因时不能敷衍伪造，权威压力之下也不能跟着人云亦云。也许你会说，人吃五谷，岂无六欲？可数据分析师就得修炼一颗平常心，不为感性所牵引，洞悉事物本质，完全实事求是。对于数据分析师来说，只有数据、事实和规律，其他神马都是浮云！

小白：您讲得太对了！除了信心、细心和平常心的“三心”外，还有哪“二意”？

Mr.林：诚意、合意！

◎ 诚意

记得上大学时追一漂亮MM被拒，人家对我说：“你连我性格、喜好都不知道，凭什么说喜欢我？”大学毕业应聘第一份工作被拒，人家对我说：“你连我们公司是干什么的都不知道，凭什么要录你？”这两个经历告诉我：做任何事情，无论是追女孩，找工作，还是学东西，如果只是抱着随便看看的心态，就只能得到随便看看就能得到的东西。

所以，小白，要成为优秀的数据分析师，就得诚心诚意。同时，诚意还包含了诚实、尊重事实的含义，就像我最开始跟你说过的，严谨负责是数据分析师必备素质之一，只有严谨负责的态度，

才能保证数据的客观、准确。数据分析若失去了真实可信的本质，就等于丧失了其尊严和价值。

◎ 合意

合意就是满足需求方（领导、运营部门等）的分析目的与需求。刚入职的朋友经常会遇到这样的情形，自己没日没夜地加班，含辛茹苦写出来的数据分析报告，却因为不符合需求方的目的和要求，被要求重新返工。所以，做分析前，先要了解清楚需求方想要什么，在做的过程中，要不断反馈细节和进度，确保你做的与需求方想的在同一轨道上运行。

小白问道：要是我觉得他们想的不对咋办？

Mr.林：你可以给他们一些建议，但是如果他不采纳你的建议，切记不要一意孤行，自己埋头苦干！

小白：哇，您开场就给我上了一节职场启蒙课！真是受益匪浅，感激不尽！

4.1.2 数据处理的内容

Mr.林：呵呵，谁叫我“好为人师”呢，那现在开始进入正题咯！首先问你，数据准备的工作做得怎么样啦？

小白：我打电话给人力资源部的同事，他们已经将录入结果发给我了，不过是文本形式的，我再用您教我的文本导入方式转换成我们需要的Excel格式，这就给您看看数据。

小白打开转换好的Excel文件，傻眼了：这Excel表格像一群唧唧喳喳的麻雀一样，折腾得小白心里不得安宁，如图4-1所示。

A	B	C	D	E	F	G	H	I	J	K	L	M
873	175	175#ID136#A_B_D	10302	718-408-4184	4273	#####	D	D	D	B	D	B
874	176	176#ID42#B_C_B_E	10314	212-775-3421	4274	#####	E	E	E	D	D	D
875	177	177#ID135#A_B_A	11361	212-340-5502	4275	01/14/08 08:01	D	E	E	D	E	E
876	178	178#ID177#B_C_B	10169	646-354-5674	4276	01/14/08 09:05	E	D	A	E	D	D
877	179	179#ID186#A_B_B	11413	917-363-2471	4277	01/14/08 11:41	E	E	E	D	E	E
878	180	180#ID63#B_C_A_I	11207	347-625-6847	4278	01/14/08 15:10	D	C	A	C	A	C
879	181	181#ID10#A_C_A_I	11366	917-613-6421	4279	01/15/08 08:44	D	D	D	E	D	D
880	182	182#ID90#A_B_A_I	11220	347-346-9229	4280	01/15/08 13:10	E	C	D	D	E	B
881	183	183#ID131#C_C_A	10122	917-785-7273	4281	01/15/08 14:04	E	D	E	D	D	D
882	184	184#ID142#A_B_A	10112	718-245-5292	4282	01/15/08 16:33	D	D	D	D	D	D
883	185	185#ID94#B_C_A_E	10013	646-391-9926	4283	01/16/08 09:51	E	E	E	D	E	E
884	186	186#ID158#D_D_A	11209	917-647-6620	4284	01/16/08 13:54	E	E	E	D	E	E
885	187	187#ID83#C_C_A_E	11418	917-350-9472	4285	01/16/08 14:48	C	D	D	E	E	C
886	188	188#ID189#D_D_B	10028	718-377-3801	4286	01/17/08 09:24	C	D	D	B	D	B
887	189	189#ID84#C_C_A_I	10451	917-612-4762	4287	01/17/08 10:30	E	D	E	D	E	C
888	190	190#ID43#A_B_A_I	10158	646-519-3002	4288	01/17/08 13:20	A	E	A	D	D	D

图4-1 员工满意度调查——初始数据表

Mr.林：你认为它应该是什么样？规规矩矩的一出来就是你想要的样子？告诉你吧，我们工作中遇到的常常是这种杂乱无章、残缺不全的数据。这个时候你得有清洁工的精神，一点一点将它弄得井井有条、干干净净。运用下面我教你的数据处理步骤，你想要它变成啥样，它就能变成啥样！

第一步，数据清洗。顾名思义，数据清洗就是将多余重复的数据筛选清除，将缺失的数据补充完整，将错误的数据纠正或删除。最后的数据状态应该是“多一分则肥，少一分则瘦”。

第二步，数据加工。经过清洗后的数据，并不一定是我们想要的数据，例如身份证号码，其实我们只想抽取里面出生年月的信息。所以，还要对数据字段进行信息提取、计算、分组、转换等加工，让它变成我们想要的数据表。

完成这两步，我们就可以对数据进行分析了，这个过程如图4-2所示。

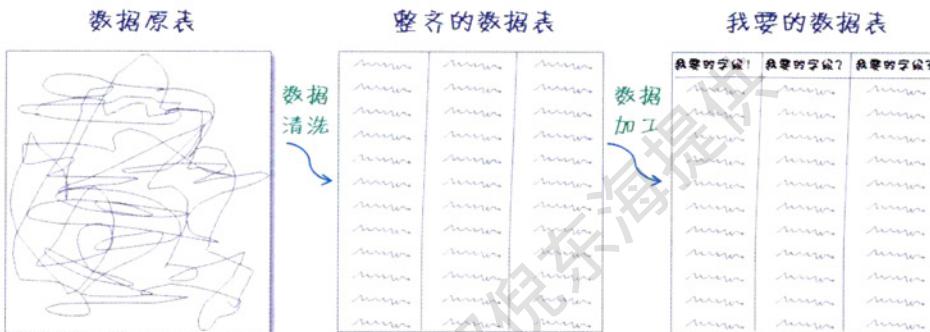


图4-2 数据处理的步骤

Mr.林：小白，理解了吗？

小白：明白！这不就跟炒土豆丝一样吗？第一步，将土豆洗得干干净净，削掉土豆皮，切成土豆丝，这就类似数据清洗咯，可以说是量的变化吧；第二步就是加工，你要做蒸土豆丝，炸土豆丝，还是炒土豆丝，你想做什么就按照不同的方法进行烹饪、加工。

同样，根据你需要的数据，采取不同的手法改造、加工。这感觉像是质的变化了。

Mr.林听完笑道：这都能想到吃，I服了U！

小白有点不好意思：这不是做个类比方便理解，融会贯通嘛。

Mr.林：呵呵，我们来做下总结。数据处理就是根据数据分析的目的，将收集到的数据，用适当的处理方法进行整理加工，形成适合数据分析的要求样式，它是数据分析前必不可少的阶段。

小白：记下了。

4.2 数据清洗

Mr.林笑了笑：既然我们已经有了数据，现在是不是到了洗土豆的时候了？

小白皱了皱眉头：可是我不喜欢洗土豆，脏兮兮的，不知道从何下手。我们拿到的数据也是，一点也不清爽，密密麻麻地挤成一团儿……

Mr.林哭笑不得: 就属你最懒! 所以, 接下来我就要先教你数据清洗的技巧。清洗数据包括三部分: 清除掉不必要的重复数据、填充缺失的数据、检测逻辑错误的数据。做这些工作的目的是为后面的数据加工提供简洁、完整、正确的数据。

4.2.1 重复数据的处理

Mr.林: 先给你讲个真实案例, 这也是我的朋友参加一次应届生招聘面试的经历。他发现大部分应届生都说自己精通Excel、Word、PPT。那到底是不是真的精通呢? 于是, 他问了一道简单的Excel问题“用几种不同的方法可以找出一张表中的重复数据”, 并让所有应聘者都作答。有个小子憋得实在不行, 挤出了一个字“数”!

小白满脸通红: 其实我也不知道……

Mr.林笑笑说: 呵呵, 没关系, 所以我说干什么工作都要有“诚”意。重复数据的处理在我们的数据处理中是非常普遍的, 虽然方法有很多, 但学起来一点都不难。不信, 跟着我来玩玩。

下面是我截取的一列“员工编号”的数据, 如图4-3所示, 教你几个处理重复数据的方法。

A	编号
1	
2	A667708
3	A310882
4	A356517
5	A520304
6	A776477
7	A466074
8	A466074
9	A466074
10	A776477
11	A218912

图4-3 处理重复数据

① 函数法

Mr.林: 首先, 介绍一个用函数识别重复数据的方法, 这里要用到COUNTIF函数。

公式密钥

COUNTIF(range, criteria): 对区域中满足单个指定条件的单元格进行计数。

要计数的单元格范围。

计算条件, 其形式可以为数字、表达式或文本;
例如, 条件可以表示为 32、"32"、>32 或 apples。

图4-4 公式密钥——COUNTIF

利用COUNTIF函数识别重复数据的具体操作如下。

STEP 01 选中B2单元格，然后输入函数公式：=COUNTIF(A:A, A2)。

STEP 02 选中C2单元格，然后输入函数公式：=COUNTIF(A\$2:A2, A2)。

STEP 03 将公式复制到B3:C11的所有单元格，效果如图4-5所示。

	A	B	C	D	E
1	编号	重复 标记	第二次重 复标记	重复项公式	第二次重 复项公式
2	A667708	1	1	=COUNTIF(A:A,A2)	=COUNTIF(\$A\$2:A2,A2)
3	A310882	1	1	=COUNTIF(A:A,A3)	=COUNTIF(\$A\$2:A3,A3)
4	A356517	1	1	=COUNTIF(A:A,A4)	=COUNTIF(\$A\$2:A4,A4)
5	A520304	1	1	=COUNTIF(A:A,A5)	=COUNTIF(\$A\$2:A5,A5)
6	A776477	2	1	=COUNTIF(A:A,A6)	=COUNTIF(\$A\$2:A6,A6)
7	A466074	3	1	=COUNTIF(A:A,A7)	=COUNTIF(\$A\$2:A7,A7)
8	A466074	3	2	=COUNTIF(A:A,A8)	=COUNTIF(\$A\$2:A8,A8)
9	A466074	3	3	=COUNTIF(A:A,A9)	=COUNTIF(\$A\$2:A9,A9)
10	A776477	2	2	=COUNTIF(A:A,A10)	=COUNTIF(\$A\$2:A10,A10)
11	A218912	1	1	=COUNTIF(A:A,A11)	=COUNTIF(\$A\$2:A11,A11)

图4-5 利用COUNTIF函数识别重复值

Mr.林： B列中的结果代表的是每一个员工编号出现的次数，所以B列中大于1的单元格所对应的员工编号即重复的编号。

没等Mr.林说完，小白忙插上一句：为什么还要加上C列？这有什么不同吗？

Mr.林解释道： C列查找的是出现了两次及其以上的重复项，以C9对应的“A466074”为例，结果“3”代表了从A1至A9，A466074是第三次重复出现。因此，筛选出C列中等于1的数即可找出数据中所有非重复项，如果对B列进行筛选，则无法完整找出非重复项。

小白恍然大悟： 哦，原来如此。

◎ 高级筛选法

Mr.林： 提到筛选，我就告诉你另一个找出重复数据的方法。其实在Excel里，可以直接利用筛选功能筛选出非重复值。具体操作如下。

STEP 01 选择数据单元格区域A1:A11。

STEP 02 在“数据”选项卡上的“排序和筛选”组中，单击“高级”按钮，弹出“高级筛选”对话框。

STEP 03 选择“将筛选结果复制到其他位置”选项，在“复制到”文本框输入B1区域，再勾选“选择不重复的记录”复选框，单击“确定”按钮，筛选效果和步骤如图4-6所示。

小白惊呼： 哇，好神奇！这个方法好，太方便了！



图4-6 利用筛选功能处理重复值

① 条件格式法

Mr.林笑道：别急，还有更神奇的方法呢！Excel 2010里面还内设了标识重复项的功能，选择“开始”→“条件格式”→“突出显示单元格规则”→“重复值”，就可以把重复的数据及所在单元格标为不同的颜色，如图4-7所示。

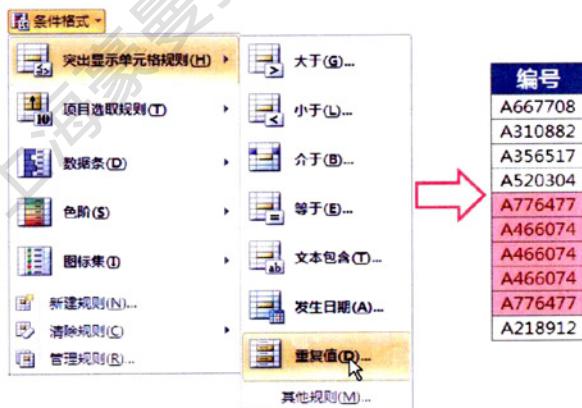


图4-7 用条件格式标记重复值

小白：这个法子也不错，还能标为不同的颜色，这样更容易识别重复数据了。

② 数据透视表法

Mr.林：呵呵，其实数据透视表一样能计算数据重复的频次，而且比COUNTIF函数还来得方

>> 谁说菜鸟不会数据分析（入门篇）

便，只要简单地拖动相应字段即可。

用数据透视表统计各数据出现的频次，出现2次及以上就说明该数据属于重复项；如果统计结果为1，则说明该数据没有重复出现。现我们仍以员工编号为例查找重复数据，具体操作步骤如下。

STEP 01 单击“插入”选项卡，在“表”功能组中，单击“数据透视表”按钮。在弹出的“创建数据透视表”对话框“选择一个表或区域”中选择数据源单元格范围“重复数据!\$A\$1:\$A\$11”，如图4-8所示。

STEP 02 继续在弹出的“创建数据透视表”对话框“选择放置数据透视表的位置”中选择“现有工作表”，并指定位置为“重复数据!\$B\$1”，如图4-8所示。

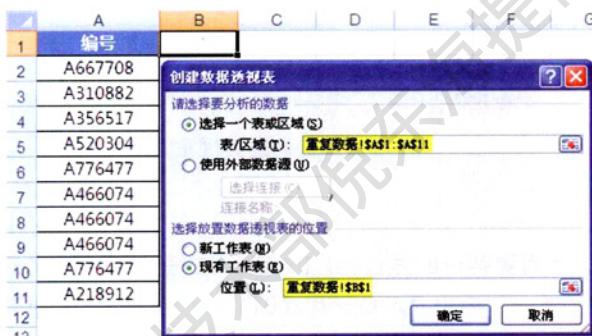


图4-8 重复数据查找操作示例

STEP 03 将“编号”字段拖至行标签，再将“编号”字段拖至数值汇总区域，如图4-9所示。

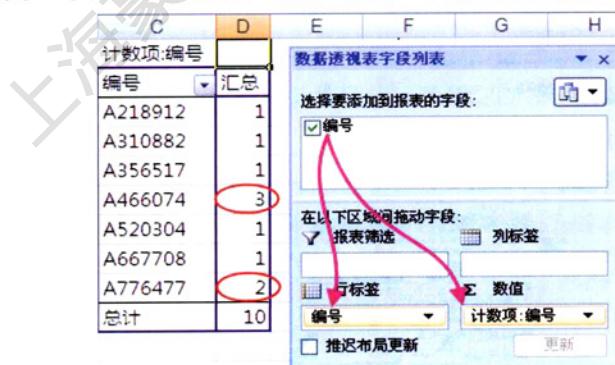


图4-9 重复数据查找结果示例

通过数据透视表的分析，我们可以得知员工编号A466074重复出现3次，A776477重复出现2次。

Mr.林：小白，你看这不就找出重复数据了。

小白：没错，真好用！简单、方便、快捷！

① 删除重复数据

Mr.林喝了口水，继续说道：最后，介绍三种删除重复项的方法。先介绍最简单的方式，直接利用菜单栏里“删除重复项”选项。

方法一：通过菜单操作删除重复项

- STEP 01 选择A1:A11数据区域。
- STEP 02 在“数据”选项卡上的“数据工具”组中，单击“删除重复项”。
- STEP 03 在“列”区域下，选择要删除的列，单击“确定”按钮。
- STEP 04 Excel 将显示一条消息，指出有多少重复值被删除，有多少唯一值被保留。单击“确定”按钮，完成操作，如图4-10所示。

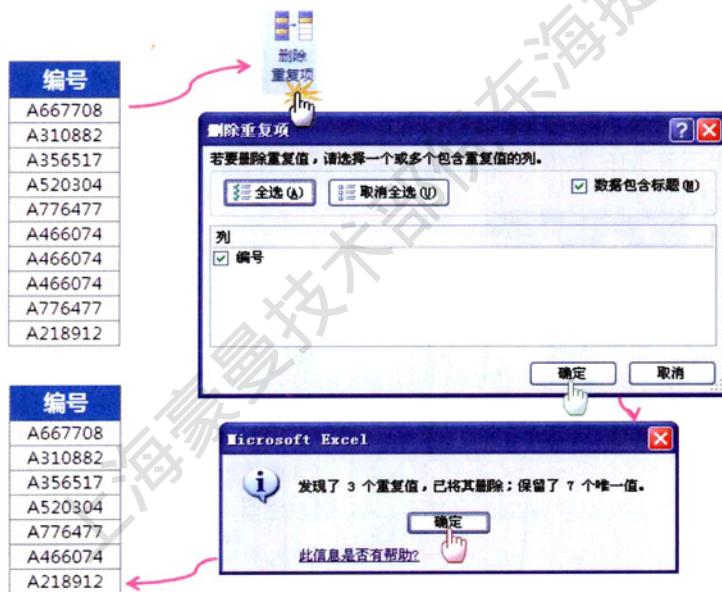


图4-10 通过菜单操作删除重复项

方法二：通过排序删除重复项

Mr.林：之前，我们采用函数识别重复值的方法得到了如图4-5的COUNTIF辅助列。可以利用该辅助列采用排序的方法删除重复数据，如图4-11所示。

- STEP 01 选中“第二次重复项”中任意一个有数据的单元格。
- STEP 02 选择“开始”主选项卡→“编辑部分”功能区→“排序和筛选”→“降序”，于是得到了重新排序的数据，其中前三项是重复项（因为数值大于1）。
- STEP 03 删除前三项所在的行，就OK了！

>> 谁说菜鸟不会数据分析（入门篇）

编号	重复标记	第二次重复项
A466074	3	3
A466074	3	2
A776477	2	2
A667708	1	1
A310882	1	1
A356517	1	1
A520304	1	1
A776477	2	1
A466074	3	1
A218912	1	1

编号	重复标记	第二次重复项
A466074	3	3
A466074	3	2
A776477	2	2
A667708	1	1
A310882	1	1
A356517	1	1
A520304	1	1
A776477	2	1
A466074	3	1
A218912	1	1

图4-11 利用排序功能删除重复项

方法三：通过筛选删除重复项

同样得到图4-5所示的辅助列后，选中“第二次重复项”中任意一个有数据的单元格，然后选择“数据”主选项卡→“排序和筛选”功能区→“筛选”，就会在列标签中出现下拉菜单，单击下拉菜单，挑出不等于1的数值，筛选后就得到非重复项了。具体操作如图4-12所示。



图4-12 利用筛选功能删除重复项

Mr.林：最后，删除筛选出来的行，这个方法稍显复杂些。

4.2.2 缺失数据处理

Mr.林：除了重复数据外，我们还会经常碰到缺失值的问题。如果缺失值过多，说明数据收集过程中存在着严重的问题。可以接受的标准是，缺失值在10%以下。

小白：什么是缺失值呢？什么情况下会造成缺失值？

Mr.林：缺失值是指数据集中某个或某些属性的值是不完全的，这在数据分析中非常常见。

缺失值产生的原因多种多样，主要分为机械原因和人为原因。机械原因是由于数据收集或保存失败造成的数据缺失，比如数据存储的失败、存储器损坏、机械故障导致某段时间数据未能收集等。人为原因是由于人的主观失误、历史局限或有意隐瞒造成的数据缺失，比如，在市场调查中被访人拒绝透露相关问题的答案，或者对问题的回答是无效的，抑或数据录入人员失误漏录了数据。

在数据表里，缺失值最常见的表现形式就是空值或者错误标识符，现在就教你怎样在庞大的数据表里，马上查找出所有缺失值。

◎ 定位输入

看小白在认真听，Mr.林继续说：如果缺失值是以空白单元格形式出现在我们的数据表中，怎样才能一步将所有的空白单元格查找出来呢？

小白摇摇头：不知道，我只会用眼睛看的。

Mr.林：最快捷的方式是采用定位功能。在Excel“开始”主选项卡的“编辑”功能区，单击下拉菜单里的“定位条件”选项，或者直接使用快捷键“Ctrl+G”，则弹出“定位”对话框，如图4-13所示。再选择“定位条件”→“空值”→“确定”，则所有的空值都被一次性选中了。



图4-13 定位查找

小白：果然是一步到位！可是查找到所有缺失值后，需要怎么处理呢？

Mr.林答道：我们一般用以下4种方法处理缺失值。

1. 处理缺失值的四种方法

方法一：用一个样本统计量的值代替缺失值。最典型的做法就是使用该变量的样本平均值代替缺失值。

方法二：用一个统计模型计算出来的值去代替缺失值。常使用的模型有回归模型、判别模型等，不过这得用专业数据分析软件才行。

方法三：将有缺失值的记录删除，不过可能会导致样本量的减少。

方法四：将有缺失值的记录保留，仅在相应的分析中做必要的排除。当调查的样本量比较大，缺失值的数量又不是很多，而且变量之间也不存在高度相关的情况下，采用这种方式处理缺失值比较可行。

在实际操作中，采用样本平均值替代缺失值是比较常见的实用的方法。当样本量较大时，我们可以采用定位查找一次，选出样本里的所有空值，再利用“Ctrl+Enter”快捷键在所有选中单元格中一次性输入样本平均值。

2. “Ctrl+Enter” 快捷键

小白问道：“Ctrl+Enter”键？

Mr.林：对的，“Ctrl+Enter”快捷键，在不连续的区域中同时输入同一个数据或公式时很好用，我用一个例子说明它的具体用法。

STEP 01 选择一个单元格或区域，然后按住Ctrl键不放，再选择第二个、第三个……单元格或区域，直到选中所有要输入数据的区域才松开Ctrl键，如图4-14所示。

	A	B	C
1			
2			
3			
4			

图4-14 “Ctrl+Enter”快捷键——步骤1

STEP 02 松开Ctrl键后，输入要录入的数据，在这里我们以输入“小白”为例，因为最后一个选中的单元格是C4，则C4中出现了我们录入的“小白”，如图4-15所示。

	A	B	C
1			
2			
3			
4			小白

图4-15 “Ctrl+Enter”快捷键——步骤2

STEP 03 别急着按Enter键，这里改成按“Ctrl+Enter”快捷键，则所有选中的单元格都变成“小白”了，如图4-16所示。

	A	B	C
1			
2	小白		小白
3		小白	
4	小白		小白

图4-16 “Ctrl+Enter”快捷键——步骤3

Mr.林: “Ctrl+Enter” 快捷键可以与定位查找搭配使用。当你用F5键或“Ctrl+G”快捷键定位方式定位到空白单元格之后，可以输入任何你想要输入的数据，再按“Ctrl+Enter”快捷键，让所有选中的单元格全部乖乖变成你要的样子。

小白: 哈哈，“Ctrl+Enter”快捷键，我记住了！

① 查找替换

Mr.林: 当缺失值是以错误标识符形式出现的时候，可以采用第二种方法——查找替换，查找出所有出现同一错误标识符的单元格，在“开始”选项卡的“编辑”功能区中，单击“查找和选择”按钮，如图4-17所示。



图4-17 查找替换

同样可以利用快捷键，查找功能的快捷键为“Ctrl+F”，替换功能的快捷键为“Ctrl+H”，在“查找内容”文本框中输入要查找的文本或数字，在“替换为”中，输入要替换的文本或数字，单击“替换”按钮即可。

例如，我要查找错误标识符“#DIV/0!”并将其全部替换成“0”，操作如下。

STEP 01 选中所有数据区域。

STEP 02 按“Ctrl+H”快捷键，弹出如图4-18所示的“查找和替换”对话框。

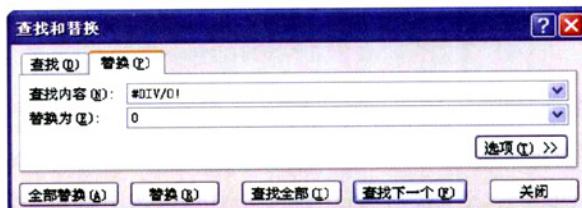


图4-18 “查找和替换”对话框

STEP 03 在“查找内容”中输入要搜索的文本或数字，这里我们输入“#DIV/0!”，在“替换为”右栏内输入“0”，再单击“全部替换”按钮，则所有“#DIV/0!”都替换为“0”了。

4.2.3 检查数据逻辑错误

Mr.林：处理完了重复多余的数据和缺失不完整的数据，我们现在来看怎么对付错误数据。

错误数据一般是怎样出现的呢？以这次的员工满意度调查为例，一般会有两种错误形式。

- ★ 被调查者输入的选项不符合要求。例如“最多选择3个选项”的多选题，答题者选择了4个选项。
- ★ 录入错误。例如二分法的多选题录入时，出现了“0”和“1”之外的数据。

如果我们不对这种数据做些处理，势必会影响分析结果。小白，在这里我要告诉你一个秘密，一般情况下IF函数能解决大部分问题，如果再巧妙地搭配条件格式，绝对“见效快、疗效好、安全可靠、无副作用”。

小白也配合道：嘿嘿，有了IF+条件，错误数据哪里逃！

Mr.林笑道：回归正题，我们来看T1题就有要求“最多选择3个选项”，而T1的数据就出现了刚刚我们讲的两种错误情况：答题者选择的选项超过了3个；录入者录入的数据出现了“0”和“1”之外的数据。

其实这两种错误是数据逻辑错误的典型，我现在就告诉你如何把它们给检验出来，如图4-19所示。首先用IF函数来检查第一类错误，错误的数据返回“错误”，正确的数据返回“正确”；再利用条件格式检查第二类错误，将非“0”和“1”的数据用红色加粗斜体字体突出显示。

The screenshot shows a Microsoft Excel spreadsheet with data in rows 3 to 6 and columns A to L. Row 1 contains column headers: 序号 (Row), T1 (Column), and 检验 (Validation). Row 2 contains column headers: A, B, C, D, E, F, G, H, I, J, K, L. Rows 3 through 6 contain data. Row 3 has values: 1, 0, 0, 1, 1, 0, 1, 0. Row 4 has values: 2, 1, 1, 0, 1, 0, 1, 0. Row 5 has values: 3, 1, 0, 1, 0, 2, 0, 0. Row 6 has values: 4, 1, 10, 1, 0, 0, 0, 0. Column I contains validation results: 正确 (Correct) for row 3, 错误 (Error) for row 5, 正确 (Correct) for row 6, and 正确 (Correct) for row 4. A callout box labeled '检验一 (IF函数)' points to cell I3 with the formula: =I3=IF(COUNTIF(B3:H3,"<>0")>3,"错误","正确"). Another callout box labeled '检验二 (利用条件格式)' points to cell B3 with the formula: =OR(B3=1,B3=0)=FALSE.

图4-19 检查逻辑错误——IF函数

◎ 利用IF函数检查错误

Mr.林接着说：先检查第一类错误，在重复数据处理中，我们讲到过COUNTIF函数，你还记得吗？

小白自信地说道：记得！这很容易记住啊，COUNT在英文中就是“计数”的意思，IF就是“在××的条件下”的意思，COUNTIF就是“对满足指定条件的单元格进行计数的意思”。在这

里 “COUNTIF(B3:H3,”<>0”)” 代表 “对B3:H3区域中不等于0的单元格进行计数”。

Mr.林笑道：看来你私底下功课做得很足，呵呵。你刚刚对 “COUNTIF(B3:H3,”<>0”)” 的理解很正确。这里又要解释一下IF函数了：IF即判断语句，判断逻辑值是真还是假。那么根据什么来判断真假呢？表达式。

所以IF函数有三个组成部分：需要判断的表达式，表达式为真时的显示值，表达式为假时的显示值，如图4-20所示。

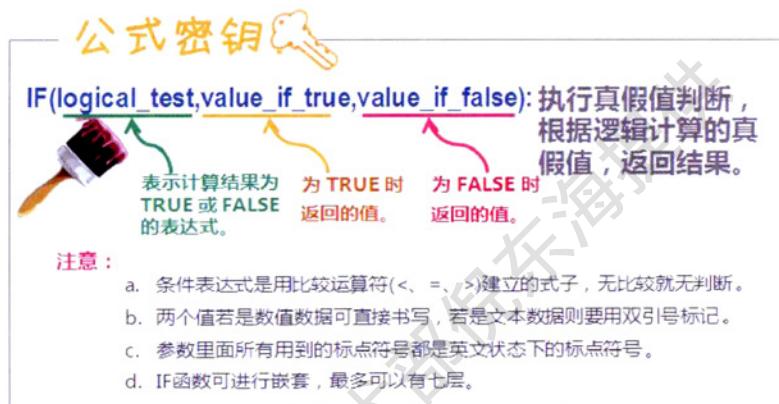


图4-20 公式密钥——IF函数

注意，0代表被调查者未选该选项，所以对非0的数据进行计数，就用你刚刚解释的公式COUNTIF(B3:H3,”<>0”)。那么 “IF(COUNTIF(B3:H3,”<>0“)>3,”错误”,“正确”)” 的意思是：如果录入的选项（非0的数据）超过3个，则单元格显示“错误”；否则，显示“正确”。

Mr.林：小白，现在你明白了吗？

小白夸张地点点头：彻底领悟了。

◎ 利用条件格式标记错误

Mr.林：好的，现在来检查第二类错误。在对重复数据的处理过程中，标识第二次重复值时，我们同样讲过条件格式的用法。在这里我操作给你看看，就不再多加解释了，如果不记得的话可以看看笔记。

选中数据区域B3:H6，再选择“开始”主选项卡→“条件格式”→“突出显示单元格规则”→“其它规则”→“使用公式确定要设置格式的单元格”，在“为符合此公式的值设置格式”文本框中输入 “=OR(B3=1,B3=0)=FALSE”，最后单击“格式”按钮将格式调整为红色加粗倾斜字体，完成设置。OR与AND函数的使用方法如图4-21所示。

在这里解释一下 “=OR(B3=1,B3=0)=FALSE” 的含义，OR代表“或”的意思，所以OR函数代表的意思是：函数的任意一个参数为真时，返回TRUE，否则返回FALSE。

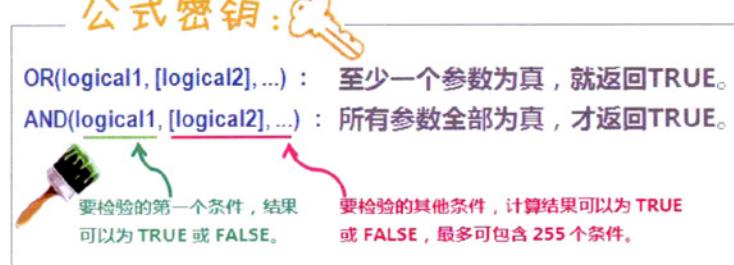


图4-21 公式密钥——OR与AND函数

那么，B3的条件格式规则“=OR(B3=1,B3=0)=FALSE”代表的意思是：如果“B3为1或者为0”的命题是错误的(=FALSE)，即B3既不为1也不为0，则会被条件格式突出标记出来。小白，你明白了吗？

小白眨眨眼：真绕呀！有点晕，不过“实践是检验真理的唯一标准”，我回去自己多操练操练就明白了。

4.3 数据加工

Mr.林：数据清洗完了之后，就是数据处理的重头戏——数据加工了，这是含金量最高的一个活儿，咱们的土豆丝做成什么样也就看这一环节的功夫了。小白，打好精神，坚持就是胜利！

小白兴奋地点点头：好啊！我迫不及待要学呢。

Mr.林：我也只教你最实用的功能与技巧，花费20%的精力解决80%的问题。

为什么要对数据进行加工？因为数据表中现有的数据字段不满足我们的数据分析需求，所以需要对现有字段进行抽取、计算或者转换，形成我们分析所需要的一列新数据字段。

4.3.1 数据抽取

Mr.林：数据抽取，是指保留原数据表中某些字段的部分信息，组合成一个新字段。可以是截取某一字段的部分信息——字段分列；也可以是将某几个字段合并为一个新字段——字段合并；还可以是将原数据表没有但其他数据表中有的字段，有效地匹配过来——字段匹配。

◎ 字段分列

我们先介绍字段分列技巧，以图4-22所示的姓名表为例，我们可以采用两种字段分列的方法将姓和名的信息单独抽取出来。

1. 菜单法

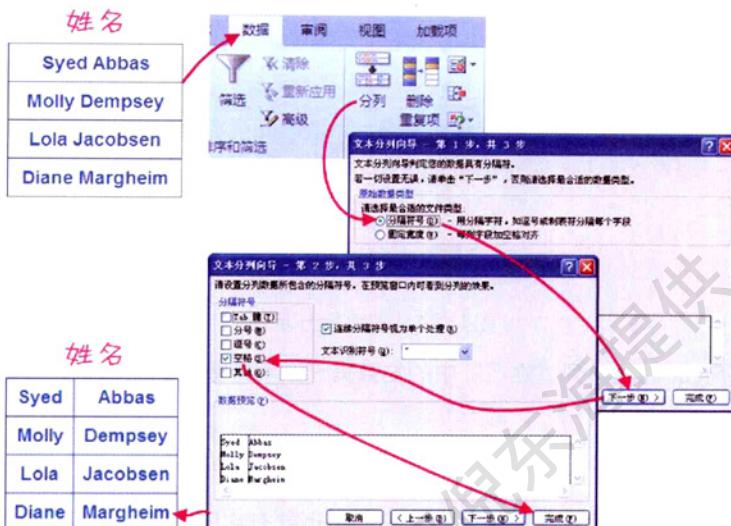


图4-22 文本分列向导

- STEP 01** 选择要转换的数据区域，在“数据”选项卡的“数据工具”组中，单击“分列”按钮。
- STEP 02** 在“文本分列向导-第1步”对话框中，单击“分隔符号”，然后单击“下一步”按钮。
- STEP 03** 在“文本分列向导-第2步”对话框中，根据需要选择分隔符号。在此例中，姓和名是以空格的形式分开的，所以选中“空格”复选框，单击“完成”按钮，字段分列就完成了，效果如图4-22左下角图所示。怎样，简单吧？

小白：嗯，是挺简单的。

2. 函数法

Mr.林：当有特定的分隔符时，采用分列法非常方便快捷。但是有时候，我们需要提取特定的几个字符，或者提取其中的第几个字符，并且没有特定的分隔符。

小白：哇，这么麻烦呀！

Mr.林：呵呵，一点儿也不麻烦，这个时候我们使用LEFT和RIGHT函数就能轻轻松松解决问题。先看看LEFT和RIGHT函数的用法，如图4-23所示。



图4-23 公式密钥——LEFT与RIGHT函数

>> 谁说菜鸟不会数据分析（入门篇）

再举个例子，例如我们要把姓名中的姓氏截取出来，如图4-24所示。

	A	B	C
1	姓名	姓	公式
2	黄雅玲	黄	=LEFT(A2,1)
3	王伟	王	=LEFT(A3,1)
4	谢丽秋	谢	=LEFT(A4,1)
5	王俊元	王	=LEFT(A5,1)
6	孙林	孙	=LEFT(A6,1)
7	王炫皓	王	=LEFT(A7,1)
8	张三丰	张	=LEFT(A8,1)
9	李四光	李	=LEFT(A9,1)
10	王麻子	王	=LEFT(A10,1)
11	赵六儿	赵	=LEFT(A11,1)

图4-24 函数拆分示例

小白：直接用LEFT提取A列“姓名”中的左数第一个字符就达到我们要的效果了，Excel函数真是太好用啦！

◎ 字段合并

Mr.林：是啊，而且还很容易理解。有字段分列也就有字段合并。字段合并是将文字或数字合并成一个单元格，例如A列是“××年”，B列是“××月”，我们可以将这两列的数据合并成C列“××年××月”。假设你要根据几个数据列创建一个句子以便发送大量信函，或者要用文本格式化数字而不影响使用这些数字的公式，就需要用到这个方法。合并文本和数字有两种方式，利用CONCATENATE函数和“&”（逻辑与）运算符。下面我举个例子，如图4-25所示。

A	B
员工	迟到
小白	5
Mr.林	10%
公式	说明(结果)
=A2&"迟到"&B2&"次"	将以上内容合并为一个短语。 (小白迟到 5 次)
=A3&"迟到比例为 "&TEXT(B3,"0%")	将以上内容合并为一个短语。 (Mr.林迟到比例为 10%)
=CONCATENATE(A2,"迟到",B2,"次")	将以上内容合并为一个短语。 (小白迟到 5 次)

如果没有我，10% 就会变成0.1哦~

图4-25 字段合并

Mr.林：小白，你也许注意到公式中用了函数TEXT。它的作用是，在使用连接运算符连接数字与文本字符串时，控制数字的显示方式。如果不TEXT，则默认显示引用单元格中的基本数据，即此例中会显示0.1，而不是在单元格中所看到的具有格式的值（10%），使用函数TEXT可恢复原数字格式。另外，CONCATENATE函数的用法也很简单，如图4-26所示。

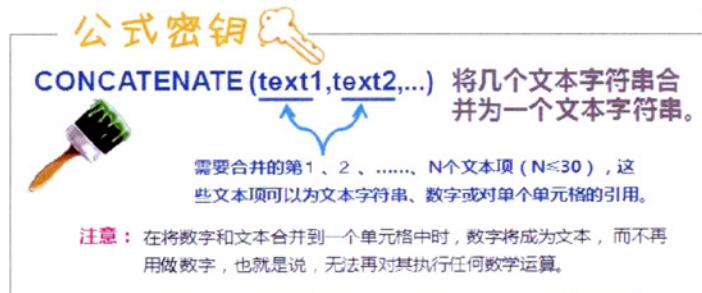


图4-26 公式密钥——CONCATENATE函数

◎ 字段匹配

Mr.林: 前面讲的字段分列和字段合并都是从原数据表中的某些字段提取信息, 但有时候原数据表没有我们需要的字段, 而需要从其他数据表中获取字段, 这时就需要用到字段匹配。举个例子, 我们公司销售部门的员工职位经常发生变动, 图4-27所示是最新的员工职位表, 现在我想截取其职务信息对应到图4-28所示的“员工个人信息(销售部)”表中。

	A 姓名	B 工号	C 部门	D 职务
1	黄雅玲	A776477	销售部	销售代表
2	王伟	A667708	销售部	销售代表
4	谢丽秋	A520304	销售部	销售代表
5	王俊元	A310882	销售部	销售总监
6	孙林	A466074	销售部	销售代表
7	王炫皓	A356517	销售部	销售代表
8	张三丰	A277381	市场部	市场总监
9	李四光	A254382	市场部	市场助理
10	王麻子	A213541	市场部	市场助理
11	赵六儿	A309752	市场部	市场助理

图4-27 员工职位表

	A 姓名	B 工号	C 出生年月	D 性别	E 工龄	F 职务
2	黄雅玲	A776477	12/8/1968	女	37	
3	王俊元	A310882	2/19/1952	男	45	
4	谢丽秋	A520304	8/30/1963	女	28	
5	王炫皓	A356517	9/19/1958	男	33	
6	孙林	A466074	3/4/1955	男	29	
7	王伟	A667708	7/2/1963	男	8	

图4-28 员工个人信息(销售部)

小白: 对哦, 两个不同的表格, 怎么将表中的信息截取过来呢?

Mr.林: 按下面的操作进行。

STEP 01 打开“员工职位表”和“员工个人信息(销售部)”两张表格。

STEP 02 在“员工个人信息(销售部)”表格的F2单元格中输入公式“=VLOOKUP(B2,[员工职位

表.xlsx]Sheet1!\$B\$1:\$D\$11,3,0)" , 按Enter键。注意, 输入VLOOKUP函数的第二个参数时, 不需要手动录入, 直接选中“员工职位表”中B1:D11的区域, 参数将自动录入成 “[员工职位表.xlsx]Sheet1!\$B\$1:\$D\$11”。

STEP 03 复制F2单元格, 并粘贴至F3:F7, 即完成数据提取。为了更方便查看公式, 我们隐藏C、D、E列, 得到的效果如图4-29所示。

	A	B	F	G	公式
1	姓名	工号	职务		
2	黄雅玲	A776477	销售代表	=VLOOKUP(B2,[员工职位表.xlsx]Sheet1!\$B\$1:\$D\$11,3,0)	
3	王俊元	A310882	销售总监	=VLOOKUP(B3,[员工职位表.xlsx]Sheet1!\$B\$1:\$D\$11,3,0)	
4	谢丽秋	A520304	销售代表	=VLOOKUP(B4,[员工职位表.xlsx]Sheet1!\$B\$1:\$D\$11,3,0)	
5	王炫皓	A356517	销售代表	=VLOOKUP(B5,[员工职位表.xlsx]Sheet1!\$B\$1:\$D\$11,3,0)	
6	孙林	A466074	销售代表	=VLOOKUP(B6,[员工职位表.xlsx]Sheet1!\$B\$1:\$D\$11,3,0)	
7	王伟	A667708	销售代表	=VLOOKUP(B7,[员工职位表.xlsx]Sheet1!\$B\$1:\$D\$11,3,0)	

图4-29 完成匹配的“员工个人信息(销售部)”表

Mr.林: 小白, 你也许要问VLOOKUP的含义和用法。VLOOKUP函数在查找与匹配中的应用非常广泛, 它的作用是在表格的首列查找指定的数据, 并返回指定的数据所在行中的指定列处的单元格内容, 如图4-30所示。



图4-30 公式密钥——VLOOKUP函数

注意, table_array第一列的值必须是要查找的值 (lookup_value) , 否则就会出现错误标识符 “#N/A” 。

另外, 还有两种情况会出现错误标识符 “#N/A” :

- ★ 数据存在空格, 此时可以嵌套使用TRIM函数将空格批量删除。
- ★ 数据类型或格式不一致, 此时将类型或格式转为一致即可。

4.3.2 数据计算

◎ 简单计算

Mr.林: 有时候数据表中的字段不能从数据源表字段中直接提取出来，但是可以通过计算来实现我们的需求。小白，看一看图4-31，我们获得的数据源只有产品销售数量和单价，老板需要的是销售额，以便了解公司业绩。这时候，通过简单计算才能达到目的。

没等Mr.林往下讲，小白打断了他：等等，什么是简单计算呀？

Mr.林: 简单计算就是字段通过加、减、乘、除等简单算术运算就能计算出来。先告诉你，在Excel中加、减、乘、除对应的运算符就是键盘中的“+、-、*、/”符号。我们知道销售额、销售数量和单价的关系如下：

$$\text{销售额} = \text{销售数量} \times \text{单价}, \text{总销售额} = \sum \text{各产品销售额}$$

现在以图4-31中所示的数据源为例，教你计算销售额。

	A	B	C	D	E
1	产品名称	销售数量	单价	销售额	公式
2	产品A	200	¥78.00	¥15,600.00	=B2*C2
3	产品B	300	¥88.00	¥26,400.00	=B3*C3
4	产品C	100	¥85.00	¥8,500.00	=B4*C4
5	产品D	50	¥100.00	¥5,000.00	=B5*C5
6	产品E	87	¥68.00	¥5,916.00	=B6*C6
7	合计	737		¥61,416.00	=SUM(D2:D6)

图4-31 计算产品销售额

STEP 01 在D2中输入“=B2*C2”，然后按Enter键完成输入。

STEP 02 将鼠标移动到D2单元格的右下角，直到出现填充柄，再双击填充柄，则D3:D6自动填充了D2的公式，D6的公式是“=B6*C6”。

Mr.林: D7是计算总销售额的单元格，即我们希望D7是D2:D6之和，如何实现呢？

小白: 你都说了“之和”，不就可以用“=D2+D3+D4+D5+D6”实现吗？

Mr.林: 是的，但是当要求和的数据不是D2:D6，而是D2:D60呢？这种公式输入的过程是不是很复杂？所以我教你另一个简单的方法，看第3步。

STEP 03 选中D7单元格，再选择“开始”主选项卡→“编辑”功能组→“自动求和”→“求和”，如图4-32所示，按Enter键，完成输入。

可以看到D7中有公式“=SUM(D2:D6)”，SUM即对多个数值求和。同时我们看到“自动求和”的下拉菜单中还有平均值、计数、最大值、最小值等功能项，当我们想求得这些值时也可利用它们来实现。

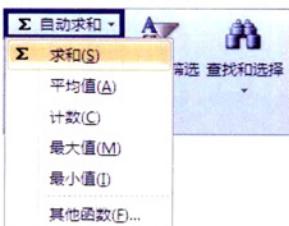


图4-32 自动求和

① 函数计算

Mr.林: 有简单计算，也就有复杂计算，所谓的复杂计算就是指运用到函数的计算。其实平常工作中我们用到的函数并不复杂，接下来我就告诉你几个既简单又实用的函数。

1. 平均值与总和

计算平均值与总和应该是我们最常见的问题了，例如之前介绍的用样本平均值代替缺失值。下面就以计算产品的季度平均销售量和销售总量为例，教你计算平均值与总和，如图4-33所示。

	A	B	C	D	E	F	G	H	I	J
1	产品名称	一季度	二季度	三季度	四季度	季度平均	总销售量			
2	产品A	51	41	56	92	60	240			=AVERAGE(B2:E2)
3	产品B	20	16	22	36	24	94			=AVERAGE(B3:E3)
4	产品C	154	123	169	277	181	724			=AVERAGE(B4:E4)
5	产品D	60	48	66	108	71	282			
6	产品E	96	77	106	173	113	451			
7	产品F	95	76	105	171	112	448			
8										
9										
10										

图4-33 平均值和求和函数举例

求平均值函数为AVERAGE()函数，求和函数为SUM()函数，括号内是需要计算的参数，参数可以为数字、单元格引用、区域或者定义的名称，参数与参数之间用逗号隔开。这两个函数的具体形式如图4-34所示。

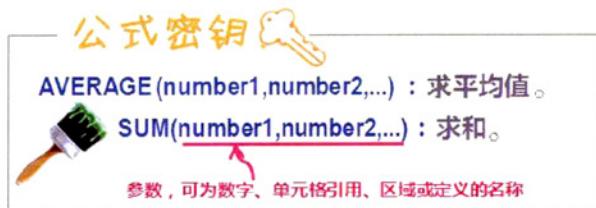


图4-34 公式密钥—AVERAGE与SUM函数

例如，图4-33中F2单元格内的公式“=AVERAGE(B2:E2)”，括号内的参数是B2:E2区域，计算

的是B2、C2、D2和E2的平均值；G5的公式“=SUM(B5:E5)”，计算的是B5、C5、D5和E5的总和。

小白：很简单嘛，我懂啦！

2. 日期的加减法

Mr.林：在工作中，除了数字和文本格式的数据外，日期和时间格式也是我们经常要面对和处理的，所以，我再告诉你一些计算日期的方法。首先，介绍一个快速输入当前日期的方法。

小白有点不屑：输入当前日期？自己写进去不就好了。

Mr.林：嘿嘿，你还在手动输入当前日期“年月日时分秒”吗？太落伍了！使用NOW或者TODAY函数就能轻轻松松搞定，如图4-35所示。

显示	公式	快捷键
2010-11-5	=TODAY()	Ctrl+;
11:31		Ctrl+Shift+;
2010-11-5 11:31	=NOW()	“Ctrl+;(分号)”，再按空格键，接着按“Ctrl+Shift+;”

公式插入的是动态的日期或时间
快捷键插入的是静态的日期或时间

图4-35 输入当前系统时间/日期

我们经常会遇到需要对日期增减的问题，例如我们希望通过添加两周的时间来调整一个项目的计划日期。这时，只需要使用一个简单的加减符就可在原日期的基础上增减天数了。例如，在单元格A1中输入“2011-1-1”，在单元格B1中输入“=A1+5”，B1中就会显示“2011-1-6”的结果。考虑更复杂的情况，如果要为日期添加“年月日”组合的时候，我们可以用DATE()函数，如图4-36所示。

A	B
日期	公式
2011-1-1	=DATE(YEAR(A2)+3,MONTH(A2)+1,DAY(A2)+5)
结果	说明
2014-7-6	给 2011年 1 月 1 日加 3 年、1 个月和 5 天

图4-36 增减日期

DATE()函数有三个参数，依次代表年、月、日，而且这三个参数是不可省略的。B2单元格中的公式“=DATE(YEAR(A2)+3,MONTH(A2)+1,DAY(A2)+5)”中“YEAR(A2)+3”代表年，含义为取A2的年份再加3年；月份为“MONTH(A2)+1”，即取A2的月份再加1个月；天数为“DAY(A2)+5”，即取A2的天数再加5天。这四个函数的具体含义如下图4-37所示。

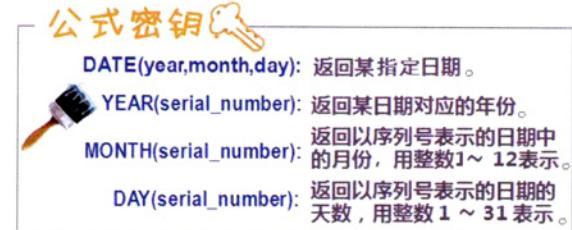


图4-37 公式密钥——DATE, YEAR, MONTH和DAY函数

3. 用函数DATEDIF计算工龄

Mr.林： 我们接着讲最后一个函数——DATEDIF，下面用这个函数来计算工龄。在这次员工满意度调查中，我们可以从不同工龄维度去分析，老板可能想知道是不是工龄越高，员工的归属感越强。在这里我们可以根据人事部档案里面的入职日期来提取工龄信息，需要用到DATEDIF函数。DATEDIF函数就是计算两个日期之间年/月/日的间隔数，详细情况可见图4-38。



图4-38 公式密钥——DATEDIF函数

Mr.林： 小白，你这么冰雪聪明，一定马上领悟了DATEDIF的含义。我要考考你，下面来看看具体应用的实例，如图4-39所示。

	A	B	C
1	入职日期	2006-10-1	
2	现在日期	2010-11-7	
3			
4	工龄（年）	4年	=DATEDIF(B1,B2,"Y")&"年"
5	工龄（年月）	4年1月	=DATEDIF(B1,B2,"Y")&"年" "&DATEDIF(B1,B2,"ym")&"月"

图4-39 利用DATEDIF计算工龄

份之差外还计算了月份之差，“ym”代表只考虑“10”月与“11”月的差距，不考虑年份。

4.3.3 数据分组

小白不禁疲惫地打了个哈欠，Mr.林在脑子里搜罗了一圈，想找点有意思的事情给她提提神，于是问道：看过《命运呼叫转移》吗？

小白：看过呀，我还记得电影里面葛大爷的经典台词，比如：“给老婆打电话是大事。”“大哥，调查一下，你们家GDP是多少？”……

Mr.林：这个我也记得，后来他还给大头汇报对村里果树的统计情况：“根据你的要求，我是这样统计的，一共有48户。40棵树以上的，有5户；30棵树以上的，有10户；20棵树以上的，有20户”。葛大爷的汇报里使用了数据分组，结论清晰明了，如图4-40所示。

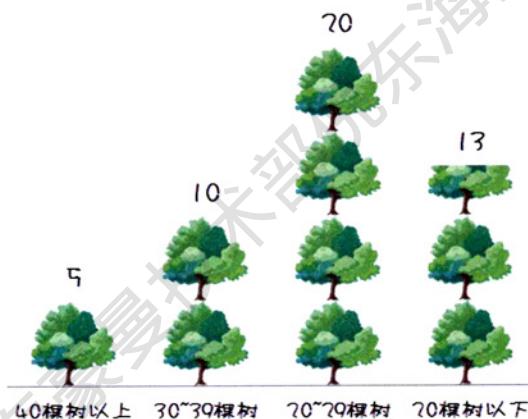


图4-40 某村种植红果树的农户数分布情况

小白：哈哈！没电的山村里的村民都知道怎么去统计，我也不能落后，快教我怎样对数据分组吧！

Mr. 林：下面我们了解下Excel中是如何实现数据分组的，主要用VLOOKUP函数来实现。以图4-41所示的价格数据为例，现在需要对价格进行区间划分，你可能想到用IF函数判断某个价格是否纳入某个区间，但是Excel对函数的嵌套有层数限制，Excel 97-2003版本最多可进行7层的嵌套，而Excel 2007-2010版本对函数嵌套可达64层。当你的价格分组的组数超过Excel函数嵌套层数限制时，用IF函数就不能一步到位了，而VLOOKUP函数却能派上用场。

现在先看一下如何用VLOOKUP函数进行数据分组。

STEP 01 准备一个分组对应表，用来确定分组的范围和标准，如图4-41的右表所示。其中，“阈值”是指每组覆盖的数值范围中的最低值，例如图4-41中第二组5-10元（ $5 \leq X < 10$ ），则阈值设置为5，其他阈值设置以此类推；“分组”记录的是每一组的组名，

例如单元格A3的数值“0.4”对应E列中“0-5元”这组；“备注”记录如何分组，目的是为了方便做数据处理的人员理解和识别。

STEP 02 在B2单元格中输入“=VLOOKUP(A2,\$D\$2:\$E\$12,2)”，并将公式复制到B2:B21的数据区域即可。

	B2	=VLOOKUP(A2,\$D\$2:\$E\$12,2)	C	D	E	F
1	价格	价格分类		阈值	分组	备注
2	0	0-5元		0	0-5元	0≤X<5
3	0.4	0-5元		5	5-10元	5≤X<10
4	2	0-5元		10	10-15元	10≤X<15
5	5.5	5-10元		15	15-20元	15≤X<20
6	4	0-5元		20	20-25元	20≤X<25
7	100	50元及其以上		25	25-30元	25≤X<30
8	50.7	50元及其以上		30	30-35元	30≤X<35
9	47	45-50元		35	35-40元	35≤X<40
10	40.1	40-45元		40	40-45元	40≤X<45
11	40	40-45元		45	45-50元	45≤X<50
12	36	35-40元		50	50元及其以上	50≤X
13	26	25-30元				
14	12	10-15元				
15	4.9	0-5元				
16	5.01	5-10元				
17	15	15-20元				
18	18	15-20元				
19	2.56	0-5元				
20	27	25-30元				
21	19	15-20元				

图4-41 利用VLOOKUP函数进行分组

这里VLOOKUP函数的应用与“数据提取”中有所不同，这里VLOOKUP函数的最后一个参数range_lookup（逻辑值）省略了，即在此处默认的是近似匹配。所以，单元格B2的公式不是在D列中寻找0，而是寻找最接近A2，又不能大于A2的值，找到了单元格D2的值正好等于0，所以返回D2的分组“0-5元”。现在明白了吗？

小白若有所思地回答道：“最接近它又不能大于它”，再让我琢磨琢磨这句话。例如单元格A5（=5.5），阈值中最接近A5又小于或等于A5的值是单元格D2（=5），则对应的E2就是A5的分组“5-10元”。现在明白了，用模糊匹配就能把小于或等于的值给识别出来，VLOOKUP函数可真神奇！呵呵。

4.3.4 数据转换

Mr.林缓口气，接着说：讲完数据抽取、数据计算和数据分组这三大重头戏，数据加工的内容就只剩下数据转换了。这里我主要想介绍两个内容：数据表的行列互换；多选题几种录入方式之间的转换。

◎ 数据表的行列互换

小白: 说到行列互换，我想起每次交给牛董的报表，都需要按照他的意思返工一次。例如上次，就像图4-42上方的表格这样，行列排版形式不合他口味，得重新颠倒来放，有没有快捷一点的方式？我可不想再一个个单元格地粘贴……

Mr.林: 呵呵，这种做法真的很——笨，我来教你一个一步到位的技巧：利用选择性粘贴。选择性粘贴不仅可以解决转置的问题，还可以选择性地粘贴格式、公式等，甚至还能选择数值将它们批量变成负数，或者加/减/乘/除一个固定值。

小白迫不及待: 这么神奇？那赶紧看看吧！

姓氏	黄	王	谢	王	孙
名字	雅玲	俊元	丽秋	炫皓	林
职务	销售代表	销售副总裁	销售代表	销售代表	销售经理
入职年份	2008	1996	2007	2010	2001

姓氏	名字	职务	入职年份
黄	雅玲	销售代表	2008
王	俊元	销售副总裁	1996
谢	丽秋	销售代表	2007
王	炫皓	销售代表	2010
孙	林	销售经理	2001

图4-42 行列互换

Mr.林: 复制好数据区域后，有三种方式可用来进行选择性粘贴。

第一种，在“开始”选项卡上的“剪贴板”组中，单击“粘贴”按钮，然后单击“选择性粘贴”选项。

第二种，你也可以使用键盘快捷方式，按“Ctrl+Alt+V”快捷键会弹出如图4-43所示的对话框，再勾上“转置”复选框，即可实现转置粘贴。

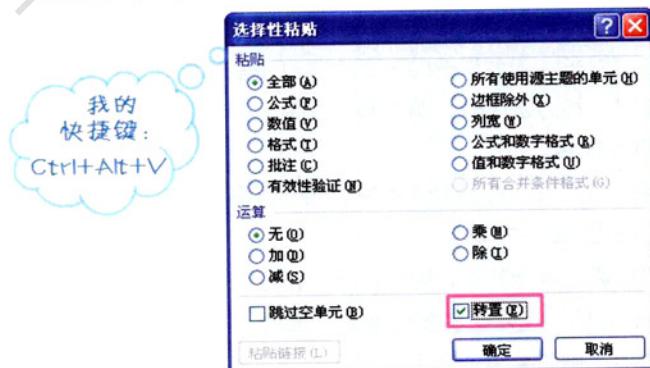


图4-43 选择性粘贴

>> 谁说菜鸟不会数据分析(入门篇)

还有第三种方式，可以直接将数据粘贴在其他区域，这时在粘贴区域的右下角会显示“粘贴”的下拉菜单 Ctrl (Ctrl)，单击下拉菜单，选择 F4 可实现转置粘贴，如图4-44所示。

在这里，我再简单介绍一下其他常用的选择性粘贴功能，如图4-45所示。



图4-44 粘贴下拉菜单

项目	功能
数值	可以去除公式、格式等
公式	只粘贴公式，使用时需注意绝对地址的应用
格式	不用重新设置格式，对复杂格式较为适用，相当于格式刷
运算区域	将复制区域的内容与粘贴区域的内容进行算术结合

图4-45 选择性粘贴功能列表

在“运算区域”里还有“加”“减”“乘”“除”四个选项，选择“加”即将复制区域中的值与粘贴区域中的值相加；“减”即从粘贴区域中的值减去复制区域中的值；“乘”“除”照此类推。例如，我们想在部分单元格数值前加上负号，可以另取任意单元格，输入“-1”，再选择性粘贴“乘”到想变号的数值区域，则该区域数值全部变成了相反数。

◎ 多选题录入数据方式之间的转换

Mr.林：小白，接着来介绍数据转换的另一个方面的内容。还记得多项选择题的两种录入方式吗？我们先来回顾一下。

第一种，各种选项用0和1表示该选项是否被选择录入，就是前面说的二分法。这种方法录入数据的时候比较烦琐，效率低下，而且容易出现错误，但是在数据分析的时候相对简单，所以一般进行数据分析时采用二分法录入的数据表。

第二种，直接录入选项的代码，也不用找对应选项直接录入，就是前面说的多重分类法。这种方式的录入速度会高很多，毕竟现在的人工成本也是蛮高的，而且对出报告的时间与执行时间的要求也蛮紧，但与第一种方法比较，第二种方法最大的缺点就是只能在SPSS里分析数据，若在Excel里分析则相对麻烦很多。

所以在这里介绍一种方法，将多重分类法录入的数据转为二分法录入的数据，这样既能节省数据录入时间，又方便在Excel中分析。

小白故意拉长了声调：两全其美，呵呵，所以，这个必须有！

Mr.林：以图4-46为例，图左边多重分类法下面的“选项一”“选项二”“选项三”代表多选题中选择的三个选项，例如某被调查者选的是A、B、C，则B2:D2的单元格中分别输入1-2-3，也可以输入1-3-2，2-3-1，2-1-3，3-1-2，3-2-1，发现没有？只要保证把1、2、3这三个选

项都塞在“选项一”“选项二”“选项三”这三个空格中就行了，至于你顺着放、倒着放、颠三倒四地放，怎么放都没有关系。

The screenshot shows a spreadsheet with columns labeled A through J. Row 1 contains labels "多重分类法" (Multi-class classification) and "二分法" (Binary classification). Row 2 contains labels "选项一" (Option 1), "选项二" (Option 2), and "选项三" (Option 3). Rows 3, 4, and 5 contain numerical values: (1, 2, 3), (1, 3, 5), and (2, 3, 4) respectively. Row 6 contains the formula `F5=IF(ISNUMBER(HLOOKUP(1,B5:D5,1,0)),1,0)`. Row 7 contains the formula `H5=IF(ISNUMBER(HLOOKUP(3,B5:D5,1,0)),1,0)`. Row 8 contains the formula `J5=IF(ISNUMBER(HLOOKUP(5,B5:D5,1,0)),1,0)`. Arrows point from the labels in row 1 to the corresponding columns in the data rows.

图4-46 将多重分类法录入的数据转换为二分法录入的数据

公式“`F5=IF(ISNUMBER(HLOOKUP(1,B5:D5,1,0)),1,0)`”里面包含两个没有介绍过的函数，一个是HLOOKUP函数，另一个是ISNUMBER函数，我们先看看它们的含义及用法。

小白好奇地问道：HLOOKUP函数怎么跟VLOOKUP长得这么像？

Mr.林：看来你还没忘，HLOOKUP函数和VLOOKUP函数都是查找函数，只不过VLOOKUP是按列查找，而HLOOKUP是按行查找。HLOOKUP中的H代表“行”，它的用法也与VLOOKUP函数非常相近，如图4-47所示。



图4-47 公式密钥——HLOOKUP函数

我来解释一下“HLOOKUP(1,B5:D5,1,0)”的含义：在B5:D5区域的第一行中查找数值1，找到该数值所在列，再返回该列所对应的单元格区域的第一行数据，并要求精确匹配，也就是必须找到1，如果找不到1，则返回“#N/A”。ISNUMBER函数就更简单了，小白，你用英语翻译一下，看看是啥意思？

小白：Is Number，是数值吗？

Mr.林：是的，如果是数值则代表HLOOKUP找到了指定数值，`ISNUMBER(HLOOKUP(1,B5:D5,1,0))`就返回TRUE，否则返回FALSE。还记得IF函数吧，这可是最重要、最常用的函数呢。

小白：记得！“`F5=IF(ISNUMBER(HLOOKUP(1,B5:D5,1,0)),1,0)`”的意思是，如果ISNUMBER返回的是TRUE，则单元格F5返回1，否则返回0。

小白再想了想：哦，我明白了，整体来看公式的意思是这样的，因为选项A的录入编码是1，所以我们在B5:D5区域中查找有没有1，有的话则在A下面标记1，代表选择了A；没有的话标记0，代表A选项没有被选。所以，我们在查找有没有选C的时候需要查找3，即将“`HLOOKUP(1,B5:D5,1,0)`”改成“`HLOOKUP(3,B5:D5,1,0)`”，就像图4-46中H5单元格中的公式那样。

Mr.林：Bingo！看来我都不用解释了。小白，还有一种很普遍的录入方式，就是将多重录入法的“选项一”“选项二”“选项三”全部录入到一个单元格中。

例如某一被调查者选的是ABC，则在一个单元格内输入123，如图4-48中的B2单元格所示。这样做的好处是，省掉了输入Tab键的麻烦，并且也不用预留足够多的选项列。同样，缺点就是不好进行输入错误的检验和数据分析。所以，下面我就介绍将其转化为二分法录入数据的方法。

如图4-48所示，E4单元格的公式“`=IF(ISNUMBER(SEARCH("1",B4)),1,0)`”与我们之前讲的多重分类法转换为二分法的公式只有一个地方不同，就是HLOOKUP函数被SEARCH函数代替。其实，这里很好理解，SEARCH就是查找的意思，`SEARCH("1",B4)`是指在B4单元格的字符串中查找1，如果找到了，则ISNUMBER返回TRUE，E4单元格显示“1”代表A选项被选；否则，显示“0”，代表未选择A。SEARCH的用法如图4-49所示。

	A	B	C	D	E	F	G	H	I
1									
2		题6							
3		123							
4		135							
5		234							
6									
7									
8									
9									
10									
11									

<code>E4=IF(ISNUMBER(SEARCH("1",B4)),1,0)</code>	<code>G4=IF(ISNUMBER(SEARCH("3",B4)),1,0)</code>	<code>I4=IF(ISNUMBER(SEARCH("5",B4)),1,0)</code>
--	--	--

图4-48 另一种多重分类录入法转换为二分录入法

公式密钥 

SEARCH(find_text , within_text , start_num): 返回指定的字符串在原始字符串中首次出现的位置。



要查找的文本字符串。要在哪一个字符串查找。从within_text的第一个字符开始查找。

注意： 在find_text中，可以使用通配符，例如：问号“？”和星号“*”。其中问号“？”代表任何一个字符，而星号“*”可代表任何字符串。如果要查找的字符串就是问号或星号，则必须在这两个符号前加上“~”符号。

图4-49 公式密钥——SEARCH函数

4.4 数据抽样

Mr.林： 关于数据处理的课程接近尾声了，小白，兴奋吧？最后我要讲的内容是数据抽样。

小白： 是不是就是跟问卷调查相关的？做问卷调查可是每个公司都要遇见的麻烦事儿。

Mr.林： 呵呵，是的。我们最常见的调查方式主要有两种，普查和抽样调查。

普查是指对总体中的对象一一进行观察、访问与记录，确定资料。例如，人口普查就是一个典型例子。用脑子想想就知道普查又耗时又耗力还耗钱，所以在企业中一般都用抽样调查。

抽样调查是指从调查对象总体中按照随机原则选取一部分对象作为样本进行调查分析，以此推论总体状况的一种调查方式。在数据抽样中，就只给你介绍一个RAND函数，如图4-50所示。

公式密钥 

RAND()函数：



返回[0,1]的均匀分布随机数，而且每次计算工作表时都将返回一个新的数值。

如果在编辑栏中输入“=Rand()”后，保持编辑状态，按“F9”键，则生成的随机数将永远保存，不再返回新的数值。

图4-50 公式密钥——RAND

小白马上接道： 哈哈！Only One呀！难道这个RAND()函数只能返回0~1之间的数，要是我需要返回大于1的数，例如随机抽取60~70之间的数怎么办呢？

Mr.林： 这个简单！若要产生60~70的随机数可以写成公式“=RAND()*10+60”，要取整的话可以用公式“=INT(RAND()*10+60)”。我解释一下，“RAND()*10”就是将RAND()区间扩大了10倍，即从[0,1]扩大到[0,10]，再加上60，则变成[60,70]了。你不明白也没关系。我用数据学语言表述一下，你记住就行了：

a、b分别代表两个数字，其中a<b，若要生成a与b之间的随机实数，可以用公式“=RAND()×(b-a)+a”。

再以我们这次的员工满意度调查为例，公司总共有5500名员工，如果这次员工满意度调查，牛董想节省成本，只抽取1000名员工进行调查来推测全体员工的满意度，则可进行如下操作。

STEP 01 将员工管理档案中的员工编号提取出来，在这里我们粘贴到B列，鼠标选中B3单元格，单击“视图”→“冻结窗格”。

STEP 02 在A列生成序号，在单元格A1中输入1，A2中输入“=A1+1”，再将A2的公式复制粘贴到A3:A5501区域，则生成了不重复的序号，如图4-51所示。

STEP 03 在D列中随机生成1000个1~5500的序号，在单元格D2中输入公式“=INT(RAND()*5500)”，将D2的公式复制粘贴到D2:D1001的单元格中，即生成1000个随机数，再将生成的随机数复制并选择性粘贴成数值，如图4-52的D列所示。

STEP 04 把D列的随机数看作随机生成的序号，参照A、B列，将随机数所对应的员工编号匹配到E列中。操作方法是在E2中输入公式“=VLOOKUP(D2,\$A:\$B,2,0)”并将公式复制到E3:E1001单元格区域中，即生成随机抽取的员工编号，如图4-52所示。

	A	B	C
1	序号	员工编号	
2	1	A776477	
5493	5492	A651404	
5494	5493	A286867	
5495	5494	A257527	
5496	5495	A596125	
5497	5496	A941753	
5498	5497	A468011	
5499	5498	A955927	
5500	5499	A789052	
5501	5500	A414967	
5502			

图4-51 生成不重复的序号

E2	A	B	C	D	E
	序号	员工编号		随机数	抽样
2	1	A776477		3962	A909476
993	992	A423984		4152	A980860
994	993	A682241		2389	A877997
995	994	A573755		447	A629874
996	995	A735102		1985	A972333
997	996	A991852		2571	A722151
998	997	A244296		1201	A145386
999	998	A680686		3443	A186512
1000	999	A579535		4191	A265405
1001	1000	A306132		3569	A958322
1002	1001	A598547			

图4-52 随机抽取的员工编号

STEP 05 最后只需要对随机抽取出来的员工编号进行去重，再用同样的随机抽样方法，凑足1000名不重复的员工编号，对每一个抽取的编号所对应的员工进行调查即可。

听Mr.林讲完最后一步，小白如释重负：哇，数据处理还真不简单！都快把我最后一点精力榨干了。

4.5 本章小结

Mr.林附和道：也快把我最后一点口水榨干了！小白，还没完呢，我们要修炼一颗平常心，

不能有浮躁敷衍的心理喔。最后，我们再来回想一下今天的内容。

第4章 三心二意——数据处理

- ★ 做数据处理得要“三心二意”——信心、细心、平常心、诚意、合意，以及数据处理中的步骤和技巧，数据要处理得有条理，就得采用数据清洗和数据加工“两步走”战略。
- ★ Excel中对重复数据、缺失数据和错误数据的处理方法，其中，需要重点掌握条件格式和查找/替换/定位的技巧。
- ★ 数据抽取、数据计算、数据分组和数据转换四大数据加工技巧。
- ★ 简单的数据抽样的方法，主要涉及RAND函数的使用。

Mr.林一口气噼里啪啦把清洗数据总结完：今天讲的内容就差不多了，我想你笔记里应该记得很详细，回去记得常翻翻，多动手实践。

小白：好的，今天的内容多而且不好理解，回去我要多花点时间慢慢消化才行。

第 5 章

工欲善其事必先利其器——数据分析

数据分析方法

数据分析工具



小白在Mr.林的悉心指导下，已经学会了如何进行数据处理，并运用在公司员工满意度调查收集到的原始数据处理工作上，完成了员工满意度的前期数据处理工作。接下来，按照牛董的要求，要为他提供一份员工满意度分析结果，但是小白不知道到底要做哪些进一步的数据分析工作，得到怎样的分析结果。眼看着牛董限定的期限就快到了，小白就像热锅上的蚂蚁急得团团转。

因为不是科班出身，面对着一堆数据，小白脑袋里一片空白，一时不知从何入手。正在愁肠百转之际，突然，小白想到Mr.林还没教自己怎么进行数据分析，呃，何不找他进一步指导呀！

于是小白揣着记事本来到Mr.林的办公桌旁，可怜巴巴地说道：Mr.林，您一定要救救我！

Mr.林看着小白垂头丧气的样子：怎么了，小白，不是已经教你怎么处理数据了吗？在哪一步遇到难题了？跟我说说看。

小白：数据处理工作我已经做完了，现在要对处理好的数据进行分析，但我不知道要怎么弄，没搞过，只好求救于您啦！

Mr.林：你已经完成数据处理的工作啦？效率真高啊，我原本预计你没这么快完成的，所以就没告诉你怎么做数据分析。

小白指着自己发黑的眼圈：效率不得不高啊，牛董限定的期限快到了呢，我是晚上加班才提前完成的，你看我都成熊猫了。

Mr.林：小白，多注意身体呀。那我们现在就来学习如何做数据分析吧。

数据分析不单单指会用数据分析工具，你还必须懂数据分析原理，没有理论的指导，就无法知晓该从哪方面入手，要分析哪些关键点。就算做出来了，得到的分析结果也无法完全解答你的问题，不符合你最初分析的目的。

5.1 数据分析方法

Mr.林：学习数据分析，首先要了解数据分析的几种基本方法，掌握了它，后面对各种数据分析方法掌握起来就快了。小白，还记得一开始我介绍的数据分析方法的三大作用吗？

小白：当然记得，主要有现状分析、原因分析、预测分析。

Mr.林：没错，那么这三大作用分别用什么数据分析方法来实现呢？这三大作用基本可分别对应对比、细分、预测三大基本方法，具体如图5-1所示。

小白：经您这么一罗列，果然很清晰明了。要解决什么问题，用什么样的分析方法。

Mr.林：接下来我们就具体看看每个数据分析方法要如何运用。第一个数据分析方法就是对比分析法，它可是数据分析的基本方法之一。

数据分析作用	基本方法	数据分析方法
现状分析	对比	对比分析 平均分析 综合评价分析
原因分析	细分	分组分析 结构分析 交叉分析 杜邦分析 漏斗图分析 矩阵关联分析 聚类分析
预测分析	预测	回归分析 时间序列 决策树 神经网络

图5-1 数据分析作用与对应的分析方法

5.1.1 对比分析法

任何事物都是既有共性特征，又有个性特征的。只有通过对比，才能分辨出事物的性质、变化、发展、与别的事物的同等个性特征，从而更深刻地认识事物的本质和规律。因此，人们历来就把对比作为认识客观世界的基本方法。

凡事都是相对的，没有绝对的，就好比物理力学原理中的参照系，选择不同参照物，物体的状态就不同，有可能是前进、静止，或者后退等状态。

数据分析也是如此，对庞大、复杂的数据单独做分析，通常很难发现其意义。只有通过汇总及对比，数据才会有意义。数据分析中各项数据指标没有好坏之分，就看选什么作为参照物，选择比你好的参照物进行对比，那么你就是相对较差或很差，反之就是相对较好或很好。

① 定义

所谓对比分析法，是指将两个或两个以上的数据进行比较，分析它们的差异，从而揭示这些数据所代表的事物发展变化情况和规律性。对比分析法的特点是：可以非常直观地看出事物某方面的变化或差距，并且可以准确、量化地表示出这种变化或差距是多少。

② 分类

对比分析法可分为静态比较和动态比较两类。

- ★ 静态比较是在同一时间条件下对不同总体指标的比较，如不同部门、不同地区、不同国家的比较，也叫横向比较，简称横比。
- ★ 动态比较是在同一总体条件下对不同时期指标数值的比较，也叫纵向比较，简称纵比。

这两种方法既可单独使用，也可结合使用。进行对比分析时，可以单独使用总量指标、相对指标或平均指标，也可将它们结合起来进行对比。比较的结果可用相对数表示，如百分数、倍数等指标。

◎ 实践运用

目前对比分析常用的有以下几个维度。

1. 与目标对比

实际完成值与目标进行对比，属于横比。例如每个公司每年都有自己的业绩目标或计划，所以首先可将目前的业绩与全年的业绩目标进行对比，看是否完成目标，如图5-2所示。如果一年还未过完，处于某阶段，可把目标按时间拆分再进行对比，或直接计算完成率，再与时间进度（当天为止的累计天数/全年天数）进行对比。

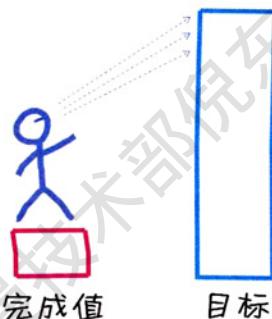


图5-2 完成值与目标对比的示例

2. 不同时期对比

选择不同时期的指标数值作为对比标准，属于纵比。倘若公司未赶上年度业绩目标的时间进度，那么可继续与自身的去年同期及上个月完成情况进行对比，如图5-3所示。与去年同期对比简称同比，与上个月完成情况对比简称环比。

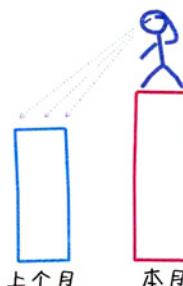


图5-3 不同时期对比的示例

小白：与上个月比还可以理解，但为何选择与去年同期对比呢？为何不选去年的年初或年底？

Mr.林见小主动提问很高兴：小白，你这个问题问得好！选择与去年同期进行对比主要考虑到季节周期性的变化，有淡旺季之分，所以选择去年的同期（同个季节）才具有可比性。当然，如果你要与去年年初或年底的完成值进行对比也是可以的。在对比前，选取对比对象时需要考虑其是否有对比意义。

通过对比自身在不同时间点上的完成情况，就可知道自身是进步还是退步。

3. 同级部门、单位、地区对比

与同级部门、单位、地区进行对比，属于横比，如图5-4所示。这样可了解自身某一方面或各方面的发展水平在公司、集团内部或各地区处于什么样的位置，明确哪些指标是领先的，哪些指标是落后的，进而找出下一步发展的方向和目标。

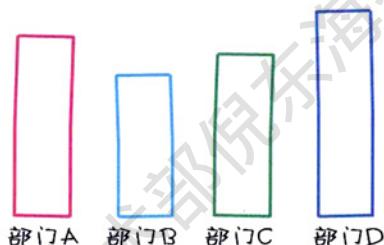


图5-4 同级别对比的示例

4. 行业内对比

与行业中的标杆企业、竞争对手或行业的平均水平进行对比，属于横比，如图5-5所示。通过行业内对比，我们也可了解自身某一方面或各方面的发展水平在行业内处于什么样的位置，明确哪些指标是领先的，哪些指标是落后的，进而找出下一步发展的方向和目标。

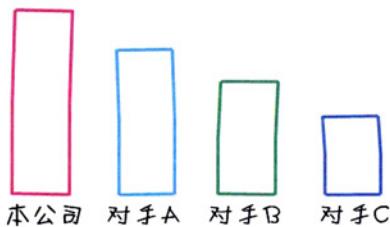


图5-5 同竞争对手对比的示例

5. 活动效果对比

对某项营销活动开展前后进行对比，属纵比，如图5-6所示。做这样的比较可以分析营销活动开展得是否有效果，效果是否明显，还可对企业投放广告的前后业务状况进行对比分析，了解投放的广告是否有效果，如品牌知名度是否提升，产品销量是否有大幅增长等。

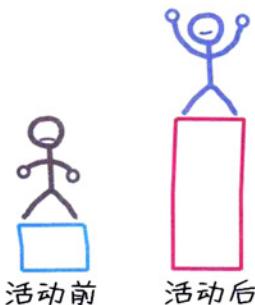


图5-6 活动前后对比示例

同时，我们还可对活动的开展状况进行分组对比，这属于横比。比如对A组（目标用户）开展了优惠活动，而对B组（非目标用户）保持原来的策略发展，通过对两组的业绩进行对比分析，可知活动是否有效果。当然在选择对照组的时候需要考虑如何确保分析结果的准确性。

Mr.林： 对比分析的维度不限于以上5点，这里只是列出常用的5种维度，当然还有其他维度，你可根据自己的实际情况采用不同的维度进行对比分析。

小白： 好的，我先记下，以后在工作实践中在进行学习应用。

◎ 注意事项

Mr.林： 进行对比分析时还要考虑到以下几点因素。

- ★ 指标的口径范围、计算方法、计量单位必须一致，即要用同一种单位或标准去衡量。如果各指标的口径范围不一致，必须进行调整之后才能进行对比。没有统一的标准，就无法比较，或者无法确认比较的结果。例如600美元与3000元人民币就无法直接比较，需要根据当期的汇率进行换算后才可进行比较，否则不具有可比性。
- ★ 对比的对象要有可比性。例如不能拿广州市与华西村、大象与蚂蚁、美国与亚洲进行对比。总之，对比对象之间相似之处越多，越具有可比性。因此，我们在选择和确定对比对象时，一定要分析它们是否具有对比的意义。
- ★ 对比的指标类型必须一致。无论绝对数指标、相对数指标、平均数指标，还是其他不同类型的指标，在进行对比时，双方必须统一。例如，2010年广州的GDP值与2010年深圳GDP增长率是无法进行对比的，因为这两个指标类型不同。

小白： 对比分析确实需要在统一的标准下才能进行，否则就失去了对比的意义。

5.1.2 分组分析法

Mr.林： 刚才讲的是对比分析，接下来介绍分组分析法。还记得前面咱们聊过的《命运呼叫转移》案例，还有我教你的数据分组法吗？

小白：记得啊！采用VLOOKUP函数进行数据分组嘛。

Mr.林：对的，当时只介绍了Excel中如何实现数据分组，但未介绍分组的目的以及如何确定组距，现在就来看看为何要分组。

作数据分析不仅要对总体的数量特征和数量关系进行分析，还要深入总体的内部进行分组分析。分组分析法是一种重要的数据分析方法，这种方法是根据数据分析对象的特征，按照一定的标志（指标），把数据分析对象划分为不同的部分和类型来进行研究，以揭示其内在的联系和规律性。

分组的目的就是为了便于对比，把总体中具有不同性质的对象区分开，把性质相同的对象合并在一起，保持各组内对象属性的一致性、组与组之间属性的差异性，以便进一步运用各种数据分析方法来解构内在的数量关系，因此分组法必须与对比法结合运用。

分组分析法的关键在于确定组数与组距。在数据分组中，各组之间的取值界限称为组限，一个组的最小值称为下限，最大值称为上限；上限与下限的差值称为组距；上限值与下限值的平均数称为组中值，它是一组变量值的代表值。

采用组距分组需要经过以下几个步骤。

STEP 01 确定组数。这个可以由数据分析师决定，根据数据本身的特点（数据的大小）来判断确定。由于分组的目的之一是为了观察数据分布的特征，因此确定的组数应适中。如果组数太少，数据的分布就会过于集中，组数太多，数据的分布就会过于分散，这都不便于观察数据分布的特征和规律。

STEP 02 确定各组的组距。组距是一个组的最大值与最小值之差，可根据全部数据的最大值和最小值及所分的组数来确定，即

$$\text{组距} = (\text{最大值} - \text{最小值}) / \text{组数}$$

STEP 03 根据组距大小，对数据进行分组整理，划归至相应组内。

分好组后，我们就可以进行相应信息的分组汇总分析，从而对比各个组之间的差异以及与总体间的差异情况。

上面所介绍的分组属于等距分组，当然也可以进行不等距分组。采用等距分组还是不等距分组，取决于所分析研究对象的性质特点。在各单位数据变动比较均匀的情况下比较适合采用等距分组；在各单位数据变动很不均匀的情况下比较适合采用不等距分组，此时不等距分组或许更能体现现象的本质特征。数据分析师可根据自己需要选择。

小白：之前只知道怎么进行分组，但不知道为什么要这样分，现在知道了。

5.1.3 结构分析法

Mr.林：结构分析法是指被分析总体内的各部分与总体之间进行对比的分析方法，即总体内各部分占总体的比例，属于相对指标。一般某部分的比例越大，说明其重要程度越高，对总体的

影响越大。例如通过对国民经济的构成分析，可以得到生产、流通、分配和消费各环节占国民经济的比重，或是各部门贡献比重，揭示各部分之间的相互联系及其变化规律。

结构相对指标（比例）的计算公式为：

$$\text{结构相对指标（比例）} = (\text{总体某部分的数值} / \text{总体总量}) \times 100\%$$

结构分析法的优点是简单实用，在实际的企业运营分析中，市场占有率就是一个非常经典的应用。

$$\text{市场占有率} = (\text{某种商品销售量} / \text{该种商品市场销售总量}) \times 100\%$$

市场占有率是分析企业在行业中竞争状况的重要指标，也是衡量企业运营状况的综合经济指标。市场占有率高，表明企业运营状况好，竞争能力强，在市场上占据有利地位；反之，则表明企业运营状态差，竞争能力弱，在市场上处于不利地位。

所以评价一个企业运营状况是否良好，不仅需要了解客户数、收入等绝对数值指标是否增长，而且还要了解其在行业中的比重是否维持稳定或者也在增长。如果在行业中的比重出现下降，那么说明竞争对手增长比你快，相比较而言，企业就是在退步，对此企业要提高警惕，出台相应改进措施。

5.1.4 平均分析法

Mr.林：平均分析法就是运用计算平均数的方法来反映总体在一定时间、地点条件下某一数量特征的一般水平。平均指标可用于同一现象在不同地区、不同部门或单位间的对比，还可用于同一现象在不同时间的对比。

平均分析法的主要作用有两点：

- ★ 利用平均指标对比同类现象在不同地区、不同行业、不同类型单位等之间的差异程度，比用总量指标对比更具有说服力。
- ★ 利用平均指标对比某些现象在不同历史时期的变化，更能说明其发展趋势和规律。

平均指标有算术平均数、调和平均数、几何平均数、众数和中位数等，其中最为常用的是算术平均数，也就是日常所说的平均数或平均值。

算术平均数的计算公式为：

$$\text{算术平均数} = \text{总体各单位数值的总和} / \text{总体单位个数}$$

算术平均数是非常重要的基础性指标。平均数是综合指标，它的特点是将总体内各单位的数量差异抽象化，它只能代表总体的一般水平，掩盖了在平均数后各单位的差异。

平均分析法要结合各种分组和指标对比来进行。比如分析不同行业、地区的平均从业人数、平均营业收入等。总之，对于所有数量指标都可以依据不同的分组用单位数来平均，进行对比与分析。

小白：嗯，平均分析与对比分析结合，真是实用。

5.1.5 交叉分析法

Mr.林: 小白, 现在我要问你几个问题, 看一看图5-7中的数据, 请你告诉我一、二月A地区的所有水果销量是多少?

1	A 月份	B 地区	C 水果	D 销量(吨)
2	一月	A	苹果	48
3	一月	B	苹果	37
4	一月	C	苹果	29
5	一月	A	香蕉	23
6	一月	B	香蕉	35
7	一月	C	香蕉	20
8	一月	A	雪梨	44
9	一月	B	雪梨	24
10	一月	C	雪梨	42
11	二月	A	苹果	25
12	二月	B	苹果	33
13	二月	C	苹果	40
14	二月	A	香蕉	41
15	二月	B	香蕉	28
16	二月	C	香蕉	28
17	二月	A	雪梨	28
18	二月	B	雪梨	32
19	二月	C	雪梨	26

图5-7 一、二月各地区水果销量表

小白听后就马上拿起Mr.林桌上的计算器算起来: 是209吨。

Mr.林: 正确, 那你再算下一、二月B地区的所有水果销量是多少?

小白又继续用计算器啪啪地计算: 是189吨。Mr.林, 需要算C地区的吗?

Mr.林微笑着说道: 不用, 你再计算下一、二月B地区的香蕉销量是多少吧。

小白算了一下: 是63吨。

Mr.林: 你不觉得这样效率很低吗? 而且容易出错。

小白: 我也是这样觉得, 但是也不知道有什么好方法, Mr.林, 您肯定知道, 而且现在打算要教我怎么使用, 对吧?

Mr.林: 聪明。我们现在来学习另一个基本的数据分析方法, 它就是交叉分析法。

交叉分析法通常用于分析两个变量(字段)之间的关系, 即同时将两个有一定联系的变量及其值交叉排列在一张表格内, 使各变量值成为不同变量的交叉结点, 形成交叉表, 从而分析交叉表中变量之间的关系, 所以也叫交叉表分析法。交叉表当然也有二维以上的, 维度越多, 交叉表就越复杂, 所以在选择几个维度的时候需要根据分析的目的决定。下面我主要介绍二维交叉表分析法。

二维交叉表其实就是我在前面介绍的二维表, 例如我们可采用图5-7中的数据进行交叉表分

析，得到交叉表如图5-8所示。

在图5-8的示例中，交叉表中的行沿水平方向延伸（从左侧到右侧），A、B、C地区的数据各占一行。交叉表中的列沿垂直方向延伸（从上到下），苹果、香蕉、雪梨各占一列。汇总字段位于行和列的交叉结点，每个交叉结点的值代表对既满足行条件，又满足列条件的记录的汇总（求和、计数等），如“B”地区和“香蕉”交叉结点的值是63，表示B地区在一、二月的香蕉销量之和为63。

地区	苹果	香蕉	雪梨	行小计
A	73	64	72	209
B	70	63	56	189
C	69	48	68	185
列小计	212	175	196	583

图5-8 一、二月各地区水果销量交叉表示例

通过交叉表分析，我们就很容易了解：

- ★ 一、二月份所有地区所有水果的总销量（总计）。
- ★ 一、二月份不同地区所有水果的销量（行小计）。
- ★ 一、二月份不同水果所有地区的销量（列小计）。
- ★ 一、二月份各个地区不同水果的销量（各交叉结点值）。

所以，刚才我让你计算的那些数据就可以不用啪啪地按计算器计算了，交叉表清晰地呈现了结果，你说是不是很方便？

小白：确实是很方便耶，在这个基础上，可以利用对比分析法分析我需要了解的信息，对吗？

Mr.林：非常正确！

5.1.6 综合评价分析法

Mr.林：小白，前面给你介绍的都是简单分析法，对于分析简单的对象没有问题。但是随着数据分析的广泛开展，分析评价的对象越来越复杂，简单分析法的局限性也越来越明显。经常会出现从这几个指标看甲单位优于乙单位，从那几个指标看乙单位优于丙单位，从其他指标看丙单位又优于甲单位的情况，使分析者难以评价到底孰优孰劣。

因此，人们通过对实践活动的总结，逐步形成了一系列运用多个指标对多个参评单位进行评价的方法，称为多变量综合评价分析方法，简称综合评价分析法。

综合评价分析法的基本思想是将多个指标转化为一个能够反映综合情况的指标来进行分析评价，比如不同国家的经济实力，不同地区的社会发展水平，小康生活水平达标进程，企业经济效益评价等，都可以应用这种方法。

进行综合评价分析，主要有5个步骤，如图5-9所示。



图5-9 综合评价分析的5个步骤

综合评价分析法主要有三大特点，如图5-10所示。

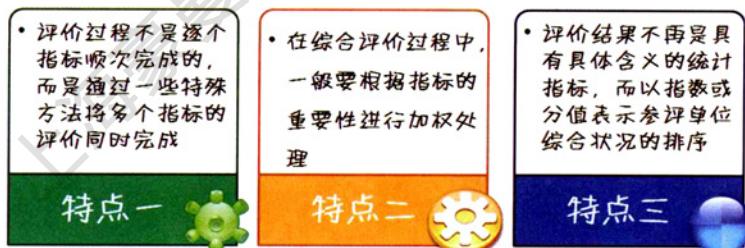


图5-10 综合评价分析法三大特点

Mr.林：小白，下面我们就一起来学习综合评价分析法用到的数据标准化与权重确定方法。

小白：呵呵，我时刻准备着呢！那您能不能讲解下数据标准化的方法呢？

Mr.林：数据的标准化是将数据按比例缩放，使之落入一个特定区间。在比较和评价某些指标时，经常会用到数据的标准化，去除数据的单位限制，将其转化为无量纲的纯数值，便于不同单位或量级的指标能够进行比较和加权。其中最典型的就是0-1标准化和Z标准化，在此就介绍0-1标准化法。

0-1标准化也叫离差标准化，就是对原始数据作线性变换，使结果落到[0,1]区间，如图5-11

所示。作0-1标准化时，对数据的转换公式如下：

第N个经标准化处理的值= (第N个原始值—最小值) / (最大值—最小值)



The screenshot shows a Microsoft Excel spreadsheet titled '标准化.xls'. The formula bar at the top contains the formula: $=\text{A3}-\text{MIN}(\$A\$2:\$A\$19) / (\text{MAX}(\$A\$2:\$A\$19)-\text{MIN}(\$A\$2:\$A\$19))$. A red arrow points from the formula bar to the cell B3, which displays the result 0.61. The table has two columns: '原始数据' (Original Data) and '0-1标准化值' (0-1 Standardized Value). The '原始数据' column contains values from 48 down to 26. The '0-1标准化值' column contains the corresponding standardized values, starting with 1.00 and ending with 0.21.

	A 原始数据	B 0-1标准化值
1	48	1.00
2	37	0.61
3	29	0.32
4	23	0.11
5	35	0.54
6	20	0.00
7	44	0.86
8	24	0.14
9	42	0.79
10	25	0.18
11	33	0.46
12	40	0.71
13	41	0.75
14	28	0.29
15	28	0.29
16	32	0.43
17	26	0.21
18		
19		

图5-11 数据0-1标准化示例

此方法有一个不足之处，就是当有新数据加入时，可能导致最大值和最小值发生变化，需要重新计算。

◎ 权重确定方法

确定指标权重的方法较多，比如专家访谈法、德尔菲法、层次分析法、主成分分析法、因子分析法、回归分析法等，这些方法都较为复杂，操作起来也相对困难，这里介绍一种比较简单的权重确定法，即目标优化矩阵表，如图5-12所示。



The table is a 5x7 matrix for target optimization. The columns are labeled A through G. Row 1 contains the headers: '人才评价' (Human Capital Evaluation), '人品' (Moral Integrity), '动手能力' (Practical Ability), '创新意识' (Innovative Awareness), '教育背景' (Educational Background), '合计' (Total), and '排序' (Ranking). Rows 2 through 5 contain the weights assigned by different individuals for each factor. The weights are represented by the number of 'A's in each cell, indicating the level of importance.

	A 人才评价	B 人品	C 动手能力	D 创新意识	E 教育背景	F 合计	G 排序
1	人才评价						
2	人品						
3	动手能力						
4	创新意识						
5	教育背景						

图5-12 目标优化矩阵表示(一)

目标优化矩阵的工作原理就是把人脑的模糊思维，简化为计算机的1/0式逻辑思维，最后得出量化的结果，这种方法不仅量化准确，而且简单、方便、快捷。目标优化矩阵的用途是非常广泛的，它不但可以用于目标的优化，还可以用于任何项目的排序，如重要性排序等。

对于目标优化矩阵中涉及的权重数值，可以找几个有经验或专业的人士，通过他们的投票表决来确定各项的重要性，从而获知各项目的权重数值。

目标优化矩阵表的用法为：将纵轴上的项目依次与横轴上的项目对比，由专家进行投票表决，如果纵轴上的项目比横轴上的项目重要，那么在两个项目相交的格子中填“1”，否则填“0”，最后将每行数字相加，根据合计的数值进行排序。

例如，假设对人才评价的指标有4个：人品、动手能力、创新意识、教育背景，公司HR需要对每个应试者打分，并计算综合得分，现在需要确定这4个指标的权重，此时我们就可以利用目标优化矩阵表。下面我们就一起来看下如何做。

首先，先将人品、动手能力、创新意识、教育背景4个指标依次填入矩阵表的第1行及第1列，如图5-12所示。

其次，从纵轴的“人品”指标开始，与横轴的“人品”“动手能力”“创新意识”“教育背景”这四个指标逐一进行比较：

- ★ 用“人品”对比“动手能力”，假设“人品”没“动手能力”重要，输入“0”。
- ★ 用“人品”对比“创新意识”，假设“人品”比“创新意识”重要，输入“1”。
- ★ 用“人品”对比“教育背景”，假设“人品”比“教育背景”重要，输入“1”。

“人品”对比完成之后，依此类推，用“动手能力”向右依次对比，再用“创新意识”向右依次对比，最后用“教育背景”依次向右对比。

所有对比完成之后，将所有的分数横向相加，在“合计”列得出各项指标的得分：人品获2分，动手能力获3分，创新意识获1分，教育背景获0分，如图5-13所示。不过，小白同学，这里是假设的数据，不要对号入座哦。

	A	B	C	D	E	F	G
1	人才评价	人品	动手能力	创新意识	教育背景	合计	排序
2	人品		0	1	1	2	2
3	动手能力	1		1	1	3	1
4	创新意识	0	0		1	1	3
5	教育背景	0	0	0		0	4

图5-13 目标优化矩阵表示（二）

小白：嘿，明白。

Mr.林：对比上述得分，各项指标的重要排序结果就出来了：①动手能力；②人品；③创新意识；④教育背景。这几项指标的重要程度依次下降。

我们可利用图5-13中的“合计”项结果来计算权重，由于教育背景为0分，但实际它还是应该占有一定的比重的，所以我们可以每项指标的“合计”的基础上加1，得到新的重要性合计得分，这样在不影响重要性的前提下，我们可以计算其权重：

$$\text{某指标权重} = (\text{某指标新的重要性合计得分}/\text{所有指标新的重要性合计得分}) \times 100\%$$

Mr.林：小白，通过上述公式，我们就可以算出各个指标的权重了，这一招会了吗？很好用的喔，回去好好消化总结下。

小白：太棒了，多谢Mr.林。

5.1.7 杜邦分析法

Mr.林: 小白, 如果牛董问你为什么公司的利润率在下降, 你会怎么回答他?

小白: 这个……我怎么知道为什么下降呀?

Mr.林: 牛董问你的时候可不能这么说啊, 作为他的助理哪能一问三不知呢。我来告诉你怎么查找下降的原因——用杜邦分析法。

杜邦分析法是由美国杜邦公司创造并最先采用的一种综合分析方法, 又称杜邦财务分析体系, 简称杜邦体系。它是利用各主要财务指标间的内在联系, 对企业财务状况及经济效益进行综合分析评价的方法。

该体系以净资产收益率为龙头, 如图5-14所示, 以总资产收益率和权益乘数为核心, 重点揭示企业盈利能力及权益乘数对净资产收益率的影响, 以及各相关指标间的相互影响关系, 为各级管理者优化经营理财状况、提高公司经营效益提供了思路。提高总资产收益率的根本在于扩大销售、节约成本、优化投资配置、加速资金周转、优化资金结构、确定风险意识等。

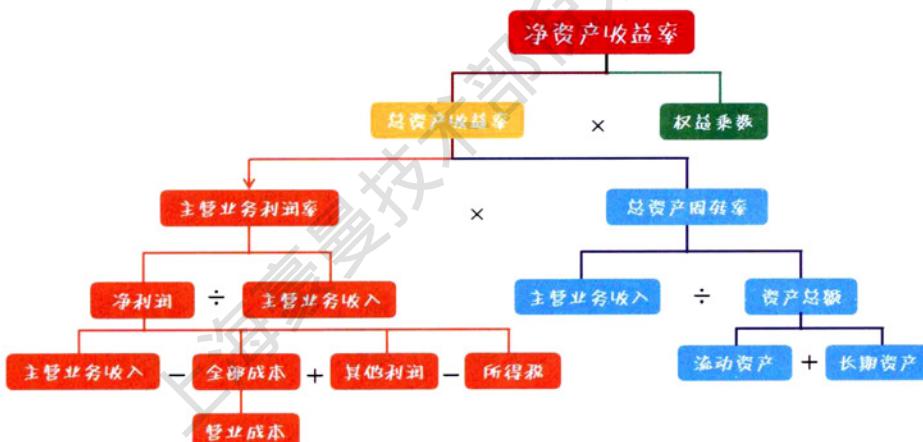


图5-14 杜邦分析体系示例

杜邦分析体系的特点是, 将若干个用以评价企业经营效率和财务状况的比率按其内在联系有机地结合起来, 形成一个完整的指标体系, 并最终通过权益收益率来综合反映。

杜邦分析采用金字塔形结构, 使财务比率分析的层次更清晰、条理更突出, 简洁明了地表达了各财务指标之间的关系。

Mr.林: 小白, 现在你该知道怎么向牛董解释利润率下降的原因了吧!

小白: 知道了, 要是牛董问我的时候, 我也能说出一二三四条了。

Mr.林: 不光是利润率下降的问题, 如果遇到其他类似问题, 都可以采用杜邦分析法。虽然杜邦分析法主要用于财务方面的分析, 但可以利用这种方法结合我们的业务进行分析。例如现有

某行业A公司2010年的市场占有率为67.5%，比2009年下降了7%。但公司的用户规模在增长，老板问你原因何在？这时我们就可以采用杜邦分析方法逐层查找原因，如图5-15所示。

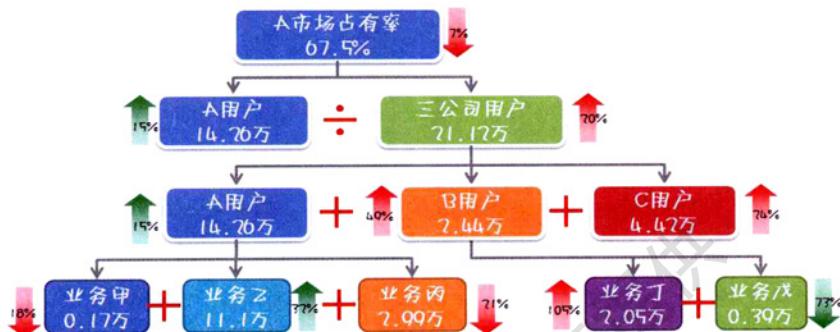


图5-15 市场占有率杜邦分析示例

通过杜邦分析，我们可发现A公司2010年市场占有率为67.5%，比2009年下降了7%的主要原因。

- ★ B公司的业务丁与2009年相比有大幅增长，拉动B公司用户的增长。
- ★ C公司的用户与2009年相比也有一定幅度的增长。
- ★ A公司用户与2009年相比虽然也有15%的增长，但是与B、C公司相比，A公司用户增长幅度相对较小，从而使A公司的市场占有率为67.5%，比2009年下降了7%。
- ★ 我们还可发现A公司业务甲和业务丙与2009年相比都在下降，而A公司的用户增长主要由业务乙拉动。

Mr.林：小白，看到没有，杜邦分析法是不是一个很好用的分析方法呢？不仅可以应用在财务分析上，而且可以用于企业市场、用户等分析上，只要运用得当，就能解决你在学习、工作上的困扰。另外，杜邦分析也充分利用了对比分析原理，将各个指标与2009年进行对比，找出问题所在。

小白：确实是喔，经你这么举例说明，简洁多了。

5.1.8 漏斗图分析法

Mr.林：小白，我们继续数据分析方法的学习。现在要介绍一个非常好用的分析方法——漏斗图分析法，它主要是以漏斗图的形式展现分析过程及结果的。

漏斗图是一个适合业务流程比较规范、周期比较长、各流程环节涉及复杂业务过程比较多的管理分析工具。为什么要在分析业务流程的时候使用漏斗图？因为漏斗图是对业务流程最直观的一种表现形式，并且也最能说明问题的所在。通过漏斗图可以很快发现业务流程中存在问题的环节。

例如漏斗图用于网站中某些关键路径的转化率的分析，不仅能显示用户从进入网站到实现购买的最终转化率，同时还可以展示整个关键路径中每一步的转化率，如图5-16所示。

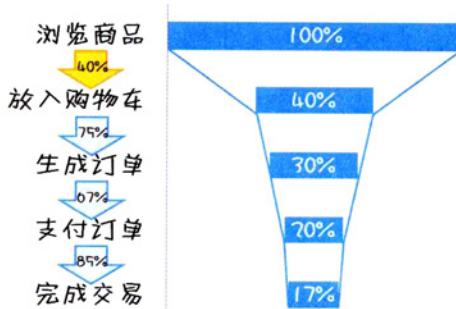


图5-16 网站转化率（漏斗图）

单一的漏斗图无法评价网站某个关键流程中各步骤转化率的好坏。我们可以利用之前介绍的对比分析方法，对同一环节优化前后的效果进行对比分析，或对同一环节不同细分用户群的转化率作比较，或对同行业类似产品的转化率进行对比，等等。

漏斗图不仅能告诉我们用户在业务中的转化率和流失率，还可以告诉我们各种业务在网站中的受欢迎程度或重要程度。通过对不同业务的漏斗图进行对比，可以找出何种业务在网站中更受用户的欢迎或更吸引用户。只要你掌握了之前介绍的对比分析方法，就可以从不同业务角度发现隐藏在其中的业务问题。

5.1.9 矩阵关联分析法

Mr.林：小白，接下来我将介绍一个重量级的数据分析方法，它的功能非常强大，在企业经营分析、市场研究中经常使用，是一种非常实用的分析方法与工具。

还没等Mr.林把话说完，小白就迫不及待地说道：Mr.林，是什么分析方法让您这么大力推荐？赶快告诉我！

Mr.林：我不是正要说嘛，你又把话抢过去了，认真听我继续介绍吧！我现在要介绍的这个方法就是矩阵关联分析法。

◎ 矩阵

矩阵分析法是指根据事物（如产品、服务等）的两个重要属性（指标）作为分析的依据，进行分类关联分析，找出解决问题的一种分析方法，也称为矩阵关联分析法，简称矩阵分析法。

以属性A为横轴，属性B为纵轴，组成一个坐标系，在两坐标轴上分别按某一标准（可取平均值、经验值、行业水平等）进行刻度划分，构成四个象限，将要分析的每个事物对应投射至这四个象限内，进行交叉分类分析，直观地将两个属性的关联性表现出来，进而分析每一个事物在这两个属性上的表现，如图5-17所示，因此它也称为象限图分析法。

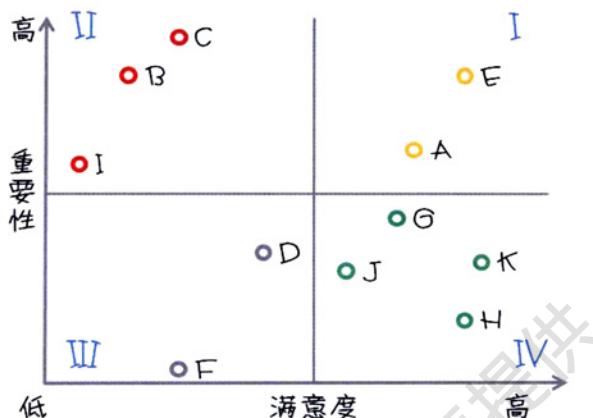


图5-17 2010年某公司用户满意度优先改进矩阵图

矩阵关联分析法在解决问题和资源分配时，为决策者提供重要参考依据。先解决主要矛盾，再解决次要矛盾，有利于提高工作效率，并将资源分配到最能产生绩效的部门、工作中，有利于决策者进行资源优化配置。

下面我就用经典案例——用户满意度研究进行矩阵应用的介绍。如图5-17所示，该矩阵为2010年某公司用户满意度调查情况，通过矩阵能够非常直观地看出公司在各方面竞争的优势和劣势分别是什么，从而合理分配公司有限的资源，有针对性地确定公司在管理方面需要提升的重点。所以在满意度研究中，此矩阵可称为优先改进矩阵或资源配置矩阵。从图5-17中可知：

- ★ 第一象限（高度关注区）属于重要性高、满意度也高的象限。A、E两个服务项目落在这个象限上。它标志着用户对服务项目的满意度与其重要性成比例，即用户对公司提供某方面服务的满意程度与用户所认为此方面服务的重要程度相符合。对这个象限上的两个服务项目，公司应该继续保持并给予支持。
- ★ 第二象限（优先改进区）属于重要性高、但满意度低的象限。B、C、I这三个服务项目落在这个象限上。这个象限标志着改进机会，用户对公司提供某方面服务的满意程度大大低于了他们认为此方面服务的重要程度。公司必须谨慎地确定需要什么类型的改进，用户感觉与事实有时候一致，有时候并不一致，所以必须谨慎地对待。如果确定确实是产品或服务存在问题，则要求进行改进。做好这几项服务项目，可以有效提高用户满意度，为公司赢得竞争优势。
- ★ 第三象限（无关紧要区）属于重要性低、满意度也低的象限。D、F这两个服务项目落在这个象限上。这个象限标志着用户对服务项目的满意度与其重要性成比例，也即用户对公司某方面提供服务的满意程度与他们认为此方面服务的重要程度相符合。对这个象限上的两个服务项目，公司应该进一步的关注用户对其期望值的变化。

★ 第四象限（维持优势区）属于重要性低、满意度高的象限。G、H、J、K这四个服务项目落在这个象限上。这个象限标志着资源过度投入，用户对公司提供某方面服务的满意程度大大超过了他们认为此方面服务的重要程度。公司投入了比用户认可满意的结果更多的时间、资金和资源，如果可能，公司应该把在此区投入的过多的资源转移至其他更重要的产品或服务方面，如第二象限上的B、C、I三个服务项目上。

通过上述分析，我们可得知矩阵关联分析法非常直观清晰、用法也简便，在营销管理活动中应用广泛，对销售管理起到指导、促进、提高的作用，而且在战略定位、市场定位、产品定位、用户细分、满意度研究等方面都有较多的应用。

Mr.林：小白，刚才介绍的四个象限都代表什么意思，知道怎么用了吗？这对你完成牛董布置的员工满意度分析的任务很有帮助喔。

小白第一次见到这样的分析方法，一时还有点懵：还没完全理解，不过我把要点都记下了，回去慢慢理解体会，再应用到我的员工满意度分析中去。

Mr.林：其实这么好的分析方法不用去死记硬背的，而且也行不通，因为每个指标的含义都不一样，有的指标是越大越好，如利润率、市场占有率，而有的指标是越小越好，如用户离网率、折旧率等。切勿生搬硬套，需要理解矩阵分析法的精髓，万变不离其宗。只有掌握了其精髓，才能融会贯通，在任何方面的分析研究也难不倒你。

小白在职场日记中记下：切勿生搬硬套，需要理解各个分析方法的精髓，万变不离其宗。

◎ 发展矩阵

Mr.林：小白，如果现在要了解2008年至2010年，3年来，用户对公司各方面的满意度变化情况，是越来越满意还是越来越不满意。你觉得应该如何来呈现？

小白摇了摇头说道：不知道，请指教。

Mr.林：只需把三年的数据都绘制在一张矩阵图上，并用箭头把每个服务项目在3年中的变化标注出来，如图5-18所示。

此矩阵在简单的矩阵中增加了发展趋势维度，让读者非常直观地了解到之前每个服务项目在用户评价中处于何种位置，现在处于何种位置，将来又如何发展，非常清晰直观，为决策者进行决策提供了强有力的支撑依据。所以我们可以称

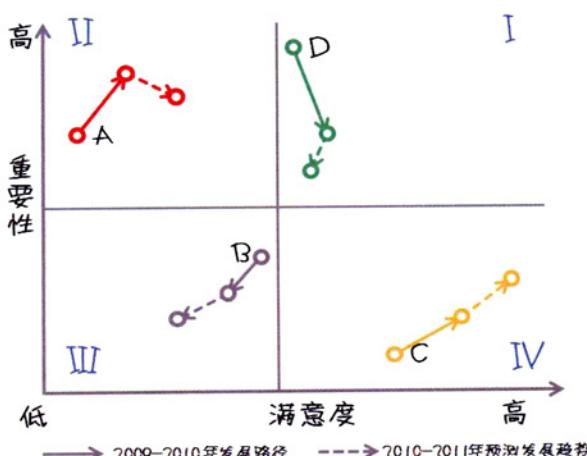


图5-18 发展矩阵示例

此类型矩阵为发展矩阵。

◎ 改进难易矩阵

Mr.林：小白，刚才介绍的矩阵的用法，你都掌握了吗？

小白：基本了解了，不过我回去还要好好复习，消化消化。

Mr.林：对，回去要多复习、多动手实践操作。再问你一个问题，如果企业有较多的短板（需改进的指标）落在图5-17中的第二象限（优先改进区），虽然手心手背都是肉，而企业由于受自身拥有的资源（如人力、物力等）所限，只能先集中有限资源对某个短板进行改进，这时决策者该如何决策把资源投给哪个服务项目呢？

小白：我觉得应该增加一个指标来衡量，不过具体增加什么指标还要再想想。

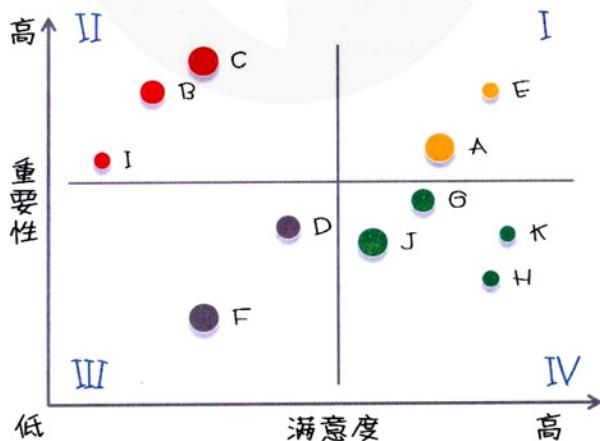
Mr.林：对，我们可在原有两个指标的基础上，增加一个指标维度，例如改进难易程度，即企业可以集中有限的资源与精力先改进对企业来说既重要又比较容易改进的短板，如果有足够的资源，再改进相对较难改进的短板，对短板进行逐一击破，从而有效地进行短板的改进。

小白疑惑不解地问：那这个改进的难易程度我们该怎么确定呢？

Mr.林：这个问题问得好，关于改进的难易程度，这个指标数据并不能直接从用户那里获取，因为用户并不了解，用户只能反映自己对该指标的满意程度。对于这项数据的获取，我们可以采用专家访谈法，获取多位业内专家对各个指标改进难易程度的评价，最后综合各专家的评价以确定最终指标的改进难易程度。另外，也可以用我们之前介绍过的目标优化矩阵来确定难易程度，这与获取权重值的道理一样，通过比较改进的难度来获取其难易程度。

我们可采用气泡图来绘制改良后的矩阵，也称改进难易矩阵，如图5-19所示。

在这个例子里，图中气泡面积的大小代表着改进难度程度，气泡越大，代表着改进程度越难；气泡越小，代表着改进程度越容易。在改进难易矩阵中可快速准确地确定改进的先后次序，为企业进行短板改进提供有效的决策依据。在此例中，很明晰地为决策者指出应先改进I服务项目，其次是B，最后才是C。当然有的决策者希望先挑战难度大的服务项目，那么改进的次序就变为



注：气泡大小代表改进难度，气泡越大，改进难度越大，反之难度越小。

图5-19 改进难易矩阵示例

先改进C，其次是B，最后才是I。

◎ 举一反三

在此再次强调，不能为了学方法而学方法，应该掌握其原理，并应用到学习、工作上。气泡图不光可以用在用户满意度研究上，它同散点矩阵一样，用途广泛，是散点矩阵的延伸。

例如经常有企业用利润率及市场占有率两个关键指标绘制产品矩阵，以衡量企业业绩的好坏，如图5-20所示。通过产品矩阵确实可以衡量业绩的好坏，但是它也存在不足，就是无法体现企业产品的真正贡献，比如公司内部各产品利润哪个最高，分别是多少？

虽A产品市场占有率高，利润率也高，但是其贡献的利润可能不如C、I两个产品，我们可以参考改进难易矩阵，在图5-20产品矩阵的基础上，增加一个产品利润指标维度，构成产品战略发展矩阵，如图5-21所示。

- ★ 第一象限：虽然A、E产品的市场占有率、利润率都相对较高，但利润较小，需要继续关注其发展态势。
- ★ 第二象限：虽然C、I两种产品利润率相对较低，但其利润仍占公司利润一定的比重，需要继续维持。
- ★ 第三象限：虽然G产品市场占有率、利润率都相对较低，但其利润仍占公司利润的一定比重，同样需要继续维

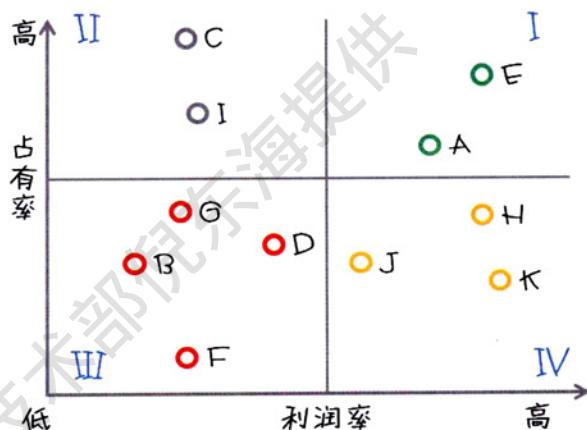
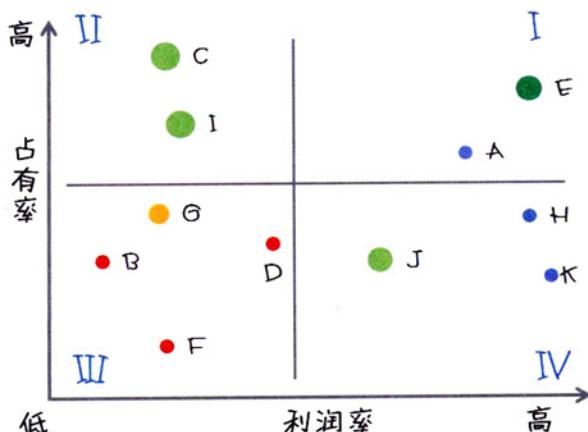


图5-20 某公司产品矩阵示例



注：气泡大小代表利润，气泡越大，利润越大，反之利润越小。

图5-21 产品战略发展矩阵示例

持；而B、F、D三种产品市场占有率、利润率都相对较低，并且利润也较低，可考虑产品战略转移至H、K等产品上。

- ★ 第四象限：虽然H、K产品市场占有率相对较低，但其利润率高，有发展潜力，可提升其市场占有率，以提升公司利润总额。

所以我们在原有二维矩阵的基础上添加一个利润维度，以全面衡量每个产品，找出对公司贡献大的产品、成长中的产品和需淘汰的产品，优化公司产品结构和明确以后的工作方向。

Mr.林：小白，你看我们通过思考、研究，又发现了这样一个新矩阵，简称林式矩阵，如果是你发现的，那就可以简称白式矩阵了，呵呵！

只有通过举一反三，我们才能学以致用，能力才能得到提升，知识得到巩固，从而解决学习、工作中遇到的问题。

小白连连点头，并在职场日记中记下了这样一句话：只有通过举一反三，才能学以致用。

5.1.10 高级数据分析方法

Mr.林：小白，刚才介绍的都是我们常用的基本数据分析方法，当然工作中还会涉及一些高级的数据分析方法以解决业务问题，比如进行市场细分需要用到聚类分析、对应分析等高级的数据分析方法。

因为你在工作中用到这些高级数据分析方法的机会较少，所以就不作具体介绍了，不过我在下面给出各种分析用途可使用的高级分析方法索引供你参考，如图5-22所示，以便需要时有据可依，同时结合查阅相关资料进一步学习。

研究方向	数据分析方法
产品研究	相关分析、对应分析、判别分析、结合分析、多维尺度分析等
品牌研究	相关分析、聚类分析、判别分析、因子分析、对应分析、多维尺度分析等
价格研究	相关分析、PSM价格分析等
市场细分	聚类分析、判别分析、因子分析、对应分析、多维尺度分析、Logistic回归、决策树等
满意度研究	相关分析、回归分析、主成分分析、因子分析、结构方程等
用户研究	相关分析、聚类分析、判别分析、因子分析、对应分析、Logistic回归、决策树、关联规则等
预测决策	回归分析、决策树、神经网络、时间序列、Logistic回归等

图5-22 高级数据分析方法索引表

小白：Mr.林，你考虑得真周到，连“后路”都为我考虑好了，呵呵！

5.2 数据分析工具

Mr.林: 小白, 前面我们已经一起学习了如何确定分析思路、搭建整体的数据分析框架, 以及分析问题的数据分析方法, 现在我们就来看看用数据分析工具如何实现数据分析。

在这里我要向你介绍一个Excel自带的数据分析工具, 它的功能非常强大, 在我看来它相当于万能钥匙, 什么都能干, 这个超级工具就是数据透视表。当然这么说可能夸张了点, 但至少它能让你免去使用那些令人头疼的统计函数的烦恼, 我们日常遇到的绝大多数数据分析问题都可以用它解决。

小白: 嗯, 在之前介绍的二维表转一维表、重复数据处理上, 我已经见识了它的强大功能。

5.2.1 初识数据透视表

Mr.林继续说道: 还记得我们在学习交叉表的时候举的一、二月各地区水果销量的例子吗?

小白: 当然记得啦, 您教我的东西怎么敢忘呢。

Mr.林: 记得就好, 那个交叉表就是用数据透视表做出来的。

小白: 哦, 那到底什么是数据透视表呢?

Mr.林: 数据透视表就是对Excel数据表中的各字段进行快速分类汇总的一种分析工具, 它是一种交互式报表。利用它, 我们可以方便地调整分类汇总的方式, 灵活地以多种不同方式展示数据的特征。

一张数据透视表仅靠鼠标拖动字段位置, 即可变换出各种类型的分析报表。用户只需指定所需分析的字段、数据透视表的组织形式, 以及要计算的类型(求和、计数、平均)。如果原始数据发生更改, 则可以刷新数据透视表来更改汇总结果。

数据透视表有机地综合了数据排序、筛选、分类汇总等数据处理分析功能; 同时, 数据透视表也是解决函数公式速度瓶颈的有效手段之一。因此数据透视表是最常用、功能最全的Excel数据分析工具之一。

下面列出一些数据透视表的相关术语, 如图5-23所示。

术语	内容
轴	数据透视表中的一个维度, 例如行、列或页
数据源	创建数据透视表的数据表、数据库等
字段	数据信息的种类, 相当于数据表中的列
字段标题	描述字段内容的标志, 可通过拖动字段标题对数据透视表进行透视分析
透视	通过改变一个或多个字段的位置来重新安排数据透视表
汇总函数	Excel用来计算表格中数据的值的函数。数值和文本的默认汇总函数分别是求和与计数
刷新	重新计算数据透视表, 以反映目前数据源状态

图5-23 数据透视表相关术语

5.2.2 创建数据透视表的三步法

Mr.林: 现我们以图5-7中的一、二月各地区水果销量数据为例，创建一个简单的数据透视表。

- STEP 01** 选择所要分析的数据源位置，单击“插入”选项卡，在选项卡“表格”功能组中，单击“数据透视表”按钮，在弹出的“创建数据透视表”对话框“选择一个表或区域”中选择数据源单元格范围“交叉表数据!\$A:\$D”，如图5-24所示。

数据透视表的数据源可以选择一个区域，也可以选择几整列数据，如果需要经常更新或添加数据，建议选择几整列，当有新数据增加的时候，只要刷新数据透视表即可，不必重新选择数据源。本例中，选择A至D这4列数据作为数据源。

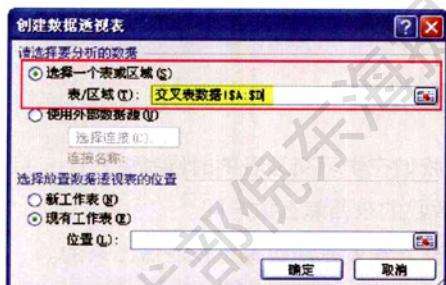


图5-24 创建数据透视表（第一步）

- STEP 02** 选择放置数据透视表的位置。继续在“创建数据透视表”对话框“选择放置数据透视表的位置”中选择“现有工作表”，并指定位置为“交叉表数据!F\$2”，如图5-25所示。

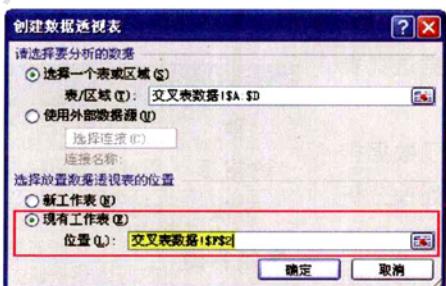


图5-25 创建数据透视表（第二步）

如果数据源较小、维度较少，则可选择现有工作表；如果数据较大，维度较多，则可选择新建工作表，在新的工作表中进行数据透视表的制作。

在确定数据源及数据表放置位置后，即可在指定的工作表位置上创建一个空白的数据透视表框架，同时在其右侧显示一个“数据透视表字段列表”窗格，在数据透视表框架内拖动字段，与

在“数据透视表字段列表”窗格内拖动字段，效果是一样的，如图5-26所示。

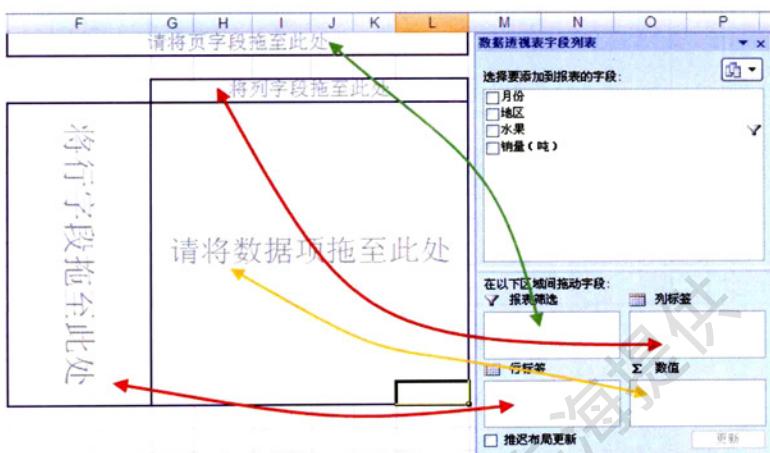


图5-26 创建数据透视表（第二步）结果

- ★ 数据透视表框架中“请将行字段拖至此处”与“数据透视表字段列表”窗格的“行标签”对应，将作为横向分类依据的字段。
- ★ 数据透视表框架中“请将列字段拖至此处”与“数据透视表字段列表”窗格的“列标签”对应，将作为纵向分类依据的字段。
- ★ 数据透视表框架中“请将数据项拖至此处”与“数据透视表字段列表”窗格的“数值”对应，将作为统计汇总依据的字段。汇总的方式有计数、求和、平均值、最大值、最小值、方差等统计指标。
- ★ 数据透视表框架中“请将页字段拖至此处”与“数据透视表字段列表”窗格的“报表筛选”对应，将作为分类显示（筛选）依据的字段。

STEP 03 拖动字段进行数据分析。将需要汇总的字段拖动至相应的“行”“列”“数据”区域，本例我们将“地区”字段拖至行标签，将“水果”字段拖至列标签，将“销量”字段拖至数值汇总区域，如图5-27所示。单击数值汇总区域中“销量”字段的“值字段设置”功能，把汇总方式设置为求和。

Mr.林：最终创建的数据透视表如图5-28所示。小白，你看这不就是我们之前介绍过的各地区水果销量交叉表嘛。当然我们也可以将“地区”与“水果”两个字段的位置进行互换，结果是一样的，只是展示的方式不一样，可根据自己的需求与喜好进行数据透视表布局的设置。

Mr.林：小白，你看，创建一个数据透视表是不是很简单？回去多练几次就能熟练掌握并使用了。

小白：确实没有想象中的那么难，我一定多多练习。

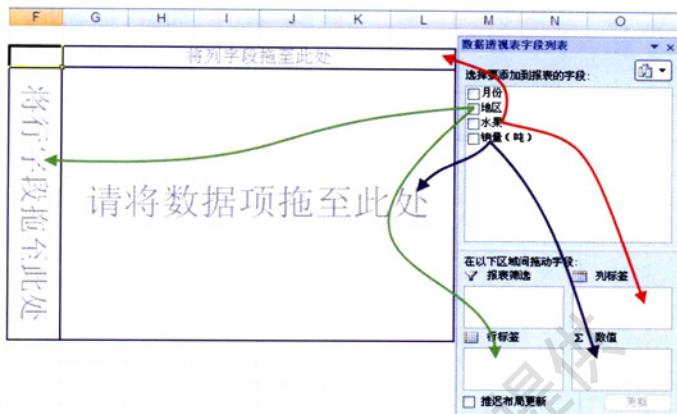


图5-27 创建数据透视表（第三步）

图5-28 数据透视表效果示例

Mr.林: 如果数据字段足够多，我们还可继续把剩余的字段，根据需要汇总的需求，拖动至行标签、列标签或报表筛选处，形成多维交叉表。当然维度越多，透视表的结果也将越复杂，分析起来越困难，所以一般只做到二维交叉表形式。如果再继续添加维度的话，建议根据分析的目的慎重选择，否则会给自己带来不必要的麻烦。

小白: 好的。

5.2.3 数据透视表分析实践

Mr.林: 关于数据透视表的理论基本介绍完了，小白，我们现在就看看具体工作中如何利用数据透视表进行数据分析。以图5-29所示的2010年某公司文具销量明细数据为例，从此表中我们

需要了解以下几个关于公司运营的问题：

- ★ 2010年总销量是多少？总销售额是多少？
- ★ 2010年A、B、C三地区的销量及销售额各是多少？
- ★ 2010年哪种产品销量最好？哪种产品销量最差？
- ★ 2010年各业务员中谁的业绩（销售额）最好？谁的业绩（销售额）最差？
- ★ 2010年公司哪个月的业绩（销售额）最好？哪个月的业绩（销售额）最差？
- ★ 另外，老板需要单独了解2010年B地区业务员王五的钢笔销量是多少？

	A	B	C	D	E	F	G
1	日期	地区	业务员	品名	销量(个)	单价(元)	销售额(元)
2	2010-1-6	A	张三	订书机	95	25	2375
3	2010-1-23	B	王五	钢笔	50	35	1750
4	2010-2-9	B	周六	钢笔	36	35	1260
5	2010-2-26	B	周六	笔记本	360	15	5400
6	2010-3-15	C	田七	订书机	600	25	15000
7	2010-4-1	A	李四	铅笔	930	0.5	465
8	2010-4-18	B	周六	订书机	740	25	18500
9	2010-5-5	B	周六	钢笔	960	35	33600
10	2010-5-22	C	田七	钢笔	530	35	18550
11	2010-6-8	A	李四	笔记本	410	15	6150
12	2010-6-25	B	王五	订书机	940	25	23500
13	2010-7-12	A	张三	铅笔	280	0.5	140
14	2010-7-29	A	张三	订书机	81	25	2025

图5-29 2010年某公司文具销量明细表

Mr.林一脸坏笑地说道：小白，这回换你来给我讲讲如何用数据透视表得出以上六个问题的答案。

小白明显底气不足：好吧，那我只好在公面前耍大刀——献丑了。

第一个问题，在创建好空白数据透视表框架的基础上，将“销量”与“销售额”两个字段拖至数值汇总区域，把“销量”与“销售额”字段汇总方式在它们各自的“值字段设置”功能中都设置为求和，如图5-30所示。所以，2010年总销量为12,146个，总销售额为227,975元，Mr.林，是这样的吗？

Mr.林：没错，非常正确，小白继续回答第二个问题。

经过初次尝试，小白的信心明显增加：第二个问题其实只需在第一个问题的数据透视表中增加一个“地区”维度，也就

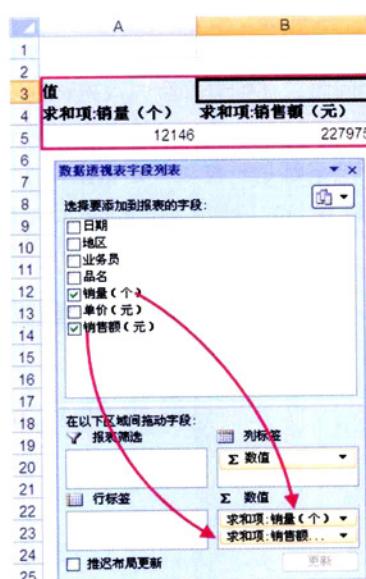


图5-30 2010年某公司文具总销量、销售额计算示例

是将“地区”字段拖至行标签区域，所以，2010年A、B、C三地区文具销量及销售额如图5-31所示。从此数据透视表我们一样可以知道，2010年总销量为12,146个，总销售额为227,975元。

	A	B	C
1	值		
2	行标签	求和项: 销量(个)	求和项: 销售额(元)
3	A	4256	39505
4	B	6481	145915
5	C	1409	42555
6	总计	12146	227975

数据透视表字段列表

选择要添加到报表的字段:

- 日期
- 地区
- 业务员
- 品名
- 销量(个)
- 单价(元)
- 销售额(元)

在以下区域间拖动字段:

报表筛选

行标签	列标签
地区	Σ 数值
品名	Σ 数值
求和项: 销量(个)	求和项: 销售额(元)

推迟布局更新

图5-31 2010年某公司各地区文具总销量、销售额计算示例

Mr.林：很不错啊！继续第三个问题。

经过前两次的数据透视表操作，此时小白己信心十足：OK。

第三个问题，在创建好空白数据透视表框架的基础上，将“销量”字段拖至数值汇总区域，把“销量”字段汇总方式在“值字段设置”功能中设置为求和，将“品名”字段拖至行或列标签区域，计算结果都是一样，只是布局稍微不一样，这里我就选择拖至行标签区域吧，如图5-32所示，2010年铅笔的销量最好，笔记本的销量最差。

	A	B	C	D	E	F
1						
2	行标签	求和项: 销量(个)				
3	笔记本	1318				
4	订书机	3657				
5	钢笔	3281				
6	铅笔	3890				
7	总计	12146				

数据透视表字段列表

选择要添加到报表的字段:

- 日期
- 地区
- 业务员
- 品名
- 销量(个)
- 单价(元)
- 销售额(元)

在以下区域间拖动字段:

报表筛选

行标签	列标签
品名	Σ 数值
	求和项: 销量(个)

推迟布局更新

图5-32 2010年某公司各文具产品销量计算示例

第四个问题同第三个问题原理一样，在创建好空白数据透视表框架的基础上，将“销售额”字段拖至数值汇总区域，把“销售额”字段汇总方式在“值字段设置”功能中设置为求和，将“业务员”字段拖至行标签区域，如图5-33所示，2010年业务员周六的业绩最好，业务员张三的业绩最差。

	A	B	C	D	E	F
1						
2						
3	行标签	求和项:销售额(元)				
4	李四	32940				
5	田七	35450				
6	王五	38015				
7	张三	6565				
8	赵大	7105				
9	周六	107900				
10	总计	227975				
11						
12						
13						
14						
15						
16						

图5-33 2010年某公司业务员业绩计算示例

第五个问题，由于原数据中没有直接给出月份信息，需要新增一个“月份”字段，可利用MONTH函数根据“日期”字段计算出相对应的月份值，图5-34所示。再根据增加月份的新表创建空白数据透视表框架，将“销售额”字段拖至数值汇总区域，把“销售额”字段汇总方式在“值字段设置”功能中设置为求和，将“月份”字段拖至行标签区域。如图5-35所示，2010年公司5月的业绩最好，7月的业绩最差。

	A	B	C	D	E	F	G	H
1	日期	地区	业务员	品名	销量(个)	单价(元)	销售额(元)	月份
2	2010-1-6	A	张三	订书机	95	25	2375	1
3	2010-1-23	B	王五	钢笔	50	35	1750	1
4	2010-2-9	B	周六	钢笔	36	35	1260	2
5	2010-2-26	B	周六	笔记本	360	15	5400	2
6	2010-3-15	C	田七	订书机	600	25	15000	3
7	2010-4-1	A	李四	铅笔	930	0.5	465	4
8	2010-4-18	B	周六	订书机	740	25	18500	4
9	2010-5-5	B	周六	钢笔	960	35	33600	5
10	2010-5-22	C	田七	钢笔	530	35	18550	5
11	2010-6-8	A	李四	笔记本	410	15	6150	6
12	2010-6-25	B	王五	订书机	940	25	23500	6
13	2010-7-12	A	张三	铅笔	280	0.5	140	7

图5-34 在数据表中添加“月份”字段示例

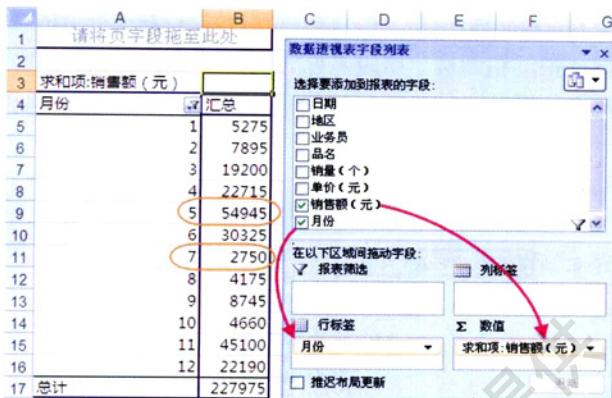


图5-35 2010年某公司各月业绩计算示例

Mr.林: 非常棒！小白，学的数据处理知识马上就用上啦，很不错啊！继续最后一个问题是。

小白有点不好意思了，继续说：至于第六个问题嘛，这个问题可以有两种解答方式。

方式一：用数据透视表做出所有业务员及所有产品的销量情况，然后找出所要的信息。

在创建的空白数据透视表框架中，将“销量”字段拖至数值汇总区域，把“销量”字段汇总方式在“值字段设置”功能中设置为求和，将“地区”字段和“业务员”字段拖至行标签区域（由于本例中一个业务员只对应一个区域，所以地区字段在此处多余，但为了完整呈现，此处依然拖动此字段进行分类汇总），将“品名”字段拖至列标签区域，如图5-36所示，2010年B地区王五的钢笔销量为282支。



图5-36 2010年某公司业务员王五业绩 (方式一操作示例)

第5章 工欲善其事必先利其器——数据分析

能，对地区、业务员、品名进行筛选，选出所要的信息。

在创建的空白数据透视表框架中，将“销量”字段拖至数值汇总区域，把“销量”字段汇总方式在“值字段设置”功能中设置为求和，分别将“地区”“业务员”“品名”字段拖至“报表筛选”区域，如图5-37所示。

A	B
1 地区	(全部) ▾
2 品名	(全部) ▾
3 业务员	(全部) ▾
4	
5 求和项: 销量(个)	汇总
6 汇总	12146
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	

图5-37 2010年某公司业务员王五业绩（方式二操作示例）

在B1单元格“地区”右侧的下拉菜单中选择“B”地区，单击“确定”按钮，如图5-38所示。用同样的方法在“业务员”项中选择“王五”，在“品名”项中选择“钢笔”，最后得到如图5-39所示的结果，很明显2010年B地区王五的钢笔销量为282支。

A	B	
1 地区	(全部) ▾	
2	<input type="checkbox"/> 全部	
3	<input checked="" type="checkbox"/> B	
4	<input type="checkbox"/> C	
5	<input type="checkbox"/> 空白	
6		
7		
8		
9		
10	<input checked="" type="checkbox"/> 选择多项	
11	<input type="button" value="确定"/>	<input type="button" value="取消"/>
12		
13		

A	B
1 地区	B <input checked="" type="checkbox"/>
2 品名	钢笔 <input checked="" type="checkbox"/>
3 业务员	王五 <input checked="" type="checkbox"/>
4	
5 求和项: 销量(个)	汇总
6 汇总	282
7	
8	
9	
10	
11	
12	

图5-38 数据透视表字段筛选示例

图5-39 2010年某公司业务员王五业绩（方式二计算结果）

Mr.林：小白，非常不错。看来你已经基本掌握如何利用数据透视表进行数据分析了，而且还能主动思考通过多种途径解决同一个问题。当然回去后还需要多加练习，熟能生巧。

小白听到Mr.林的称赞，心里美滋滋的，嘴上还是很谦虚：还不是您教导有方，多谢您啦！

>> 谁说菜鸟不会数据分析（入门篇）

5.2.4 数据透视表小技巧

Mr.林: 通过数据透视表可以得知公司业务整体运营状况、个人业绩等具体数值信息，此外我们还需要分析了解公司业务结构以及增长情况，所以我们需要在刚才数据透视表计算得到数据的基础上进一步计算百分比、环比、同比等数据，也就是需要运用结构分析法、对比分析法来分析公司业务运营情况。小白，如果是你，你会怎么做？

小白: 那就直接把数据透视表计算好的结果直接复制出来，再用加减乘除进行计算呗！

Mr.林: 你这个方法是可行，不过有点麻烦，如果更新数据时，需要手工重新进行复制粘贴操作，我教你一个更简单的方法，就是直接在数据透视表中进行计算，如果有数据更新，也可以直接刷新数据透视表，就可得到最新的计算结果了。

小白兴奋地说: 好啊！

◎ 百分比计算

Mr.林: 第一个就是百分比的计算，例如刚才我们已经计算了2010年A、B、C三地区的销量及销售额，下面我们进一步了解三个地区的销售额占比数据，也就是哪个地区贡献大，贡献了多少份额。我们在图5-31的基础上进行操作。

STEP 01 继续将“销售额”字段拖至数值汇总区域，将其汇总方式设置为求和。

STEP 02 用鼠标点击在刚得到“销售额”求和数据范围内的任一单元格，单击右键，选择“值显示方式”→“列汇总的百分比”，将标题名命名为“百分比”，即可得到如图5-40所示的结果。

数据			
地区	求和项 销量(个)	求和项 销售额(元)	百分比
A	4256	39505	17.33%
B	6481	145915	64.00%
C	1409	42555	18.67%
总计	12146	227975	100.00%

图5-40 2010年某公司各地区业务收入占比

Mr.林: “值显示方式”操作还可通过“选项”选项卡“计算”分组中的“值显示方式”进行操作，你可根据自己的操作习惯选择。

◎ 同比、环比计算

Mr.林: 因为这个例子中的数据有限，这里我就介绍环比计算，同比计算与环比计算原理都是一样的。以2010年每月销售额计算为例，如图5-35所示，计算每个月的环比数据。

STEP 01 继续将“销售额”字段拖至数值汇总区域，将其汇总方式设置为求和。

STEP 02 用鼠标点击在刚得到“销售额”求和数据范围内的任一单元格，单击右键，选择“值显示方式”→“差异百分比”。

第5章 工欲善其事必先利其器——数据分析

STEP 03 在弹出的“值显示方式”对话框中（如图5-41所示），设置要计算差异百分比的基本字段、基本项。在本例中基本字段选择“月份”，基本项选择“(上一个)”，也就是环比的意思，单击“确定”，将标题名命名为“环比”，得到结果如图5-42所示。

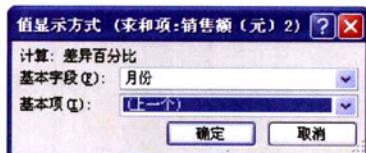


图5-41 数据透视表“值显示方式”对话框

月份	求和项	销售额(元)	环比
1		5275	
2		7895	49.67%
3		19200	143.19%
4		22715	18.31%
5		54945	141.89%
6		30325	-44.81%
7		2750	-90.93%
8		4175	51.82%
9		8745	109.46%
10		4660	-46.71%
11		45100	867.81%
12		22190	-50.80%
总计		227975	

图5-42 2010年某公司各月业绩环比计算示例

Mr.林：如果有两年以上的数据，并且有“年份”字段，可将“年份”与“月份”字段，拖入数据透视表行标签，在“值显示方式”对话框的基本字段中，将出现“年份”与“月份”两个字段供选择，选择“年份”即可计算每月的同比数据。

◎ 数据分组统计

Mr.林：刚才计算每月数据时，采用了函数法增加一个字段，其实在数据透视表中有一个“创建组”功能，它可以对日期型、数值型、文本型数据进行分组。

- ★ 日期型：可分组的维度有年、季度、月、日、小时、分、秒。
- ★ 数值型：按照一定的组距（步长）进行等距分组，但它没有VLOOKUP分组灵活，VLOOKUP可进行不等距分组。
- ★ 文本型：可根据需要自行选择需要分组的项进行分组，需要手工一个个点击，用鼠标右键组合进行分组。

我就以图5-35所示数据中的日期分组为例，讲解数据透视表分组操作。

- STEP 01** 将“月份”字段移开。
- STEP 02** 将“日期”字段拖至行标签。
- STEP 03** 用鼠标点击在刚得到“日期”行分组范围内的任一单元格，单击鼠标右键，选择“创建组”。
- STEP 04** 在弹出的“分组”对话框中（如图5-43所示），确认起始日期、终止日期是否正确，并选择步长为“月”，单击“确定”按钮，即可得到如图5-44所示结果。

小白：Mr.林，您介绍的这几个小技巧都很实用啊！

Mr.林：数据透视表还有“计算字段”“计算项”“切片器”等实用功能，我就不一一介绍了，你可在实际操作中进行摸索研究。

小白：好的，我先记下，等要用到的时候好好琢磨琢磨。

>> 谁说菜鸟不会数据分析(入门篇)

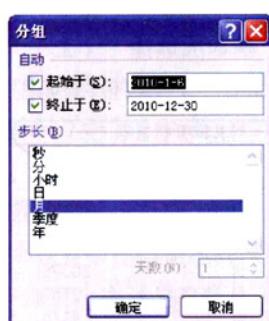


图5-43 数据透视表“分组”对话框

行标签		求和项:销售额(元)
1月		5275
2月		7895
3月		19200
4月		22715
5月		54945
6月		30325
7月		2750
8月		4175
9月		8745
10月		4660
11月		45100
12月		22190
总计		227975

图5-44 数据透视表“分组”计算结果

5.2.5 多选题分析

Mr.林: 小白，现在你知道怎么用数据透视表去完成员工满意度调查分析了吗？

小白: 基本上都会了。不过还有个问题，对于员工满意度调查中的单选题我已经知道怎么分析了，但是多选题要怎么分析？如果按这样的方式操作行吗？

Mr.林: 没错，一般情况下，单选题通过计数方式来统计汇总各个选项被选择的次数。多选题的录入方式决定了其统计方式较为困难，而且会与其他行为、背景等方面的数据脱离联系，如果继续采用单选题的计数汇总方式，只能统计到总体情况，而无法进行交叉分析，即无法对问题进行进一步的分解探寻其真相。

不过我们可变换一下汇总方式，比如用求和的汇总方式进行计算，那么多选题的数据必须是以0、1方式输入的，即选择此项输入“1”，不选则输入“0”。小白，你可以想象一下，如果有20人选了A项，那么就有20个1，加起来就等于20，不就可以统计到选择此项的人数了？这样我们就可以进行多选题的数据分析。

现在就通过实例来学习如何应用数据透视表进行多选题的分析。现有一个用户品牌知名度调查，让用户从A、B、C、D、E五个品牌中选择听说过的品牌，可多选，调查结果如图5-45所示。

	A	B	C	D	E	F	G	H	I	J
1	编号	性别	年龄	学历	A	B	C	D	E	
2	10001	男	35	大学	1	0	1	0	1	
3	10002	女	26	大学	1	1	1	0	0	
4	10003	女	22	中专	0	1	1	1	0	
5	10004	女	28	大学	1	0	1	1	0	
6	10005	男	40	高中	0	1	0	1	1	
7	10006	女	36	高中	1	0	1	0	1	
8	10007	女	41	高中	0	1	0	1	1	
9	10008	女	45	大学	1	0	1	0	1	
10	10009	女	50	中专	1	1	0	1	0	
11	10010	女	52	中专	1	1	1	0	0	
12	10011	女	36	研究生	0	1	1	1	0	
13	10012	女	42	大学	1	0	1	1	0	
14	10013	女	40	高中	0	1	0	1	1	

图5-45 多选题分析数据示例

现在我们要了解以下情况:

- ★ 用户对五个品牌整体的认知度怎样? 哪个品牌用户知名度高? 哪个品牌用户知名度低?
- ★ 不同性别的用户对品牌的认知度是否有差异?
- ★ 不同年龄段的用户对品牌的认知度是否有差异?
- ★ 不同学历的用户对品牌的认知度是否有差异?

Mr.林: 小白, 再给你一次表现的机会, 就用我刚才说的求和方式进行统计汇总。

小白这时对数据透视表的使用已经是轻车熟路了, 爽快地答应: 好的。

第一个问题, 在创建好空白数据透视表框架的基础上, 将A、B、C、D、E五个品牌字段拖至数值汇总区域, 这时我们可以看到统计汇总的结果都为30, 如图5-46所示。这只能说明30行数据都有录入, 但并无法得出每个品牌被选择的次数, 再一次证明Mr.林您刚才说的用计数方式进行多选题的分析是不可行的。

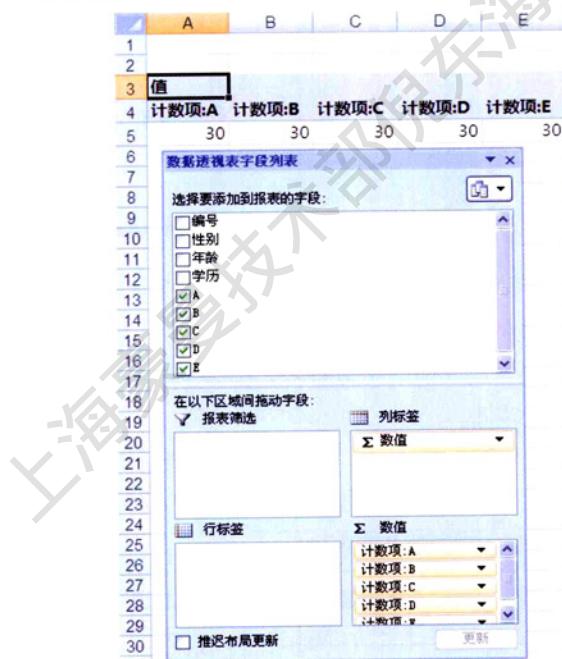


图5-46 多选题分析之整体分析结果(示例1)

我们可继续把A、B、C、D、E五个品牌字段的汇总方式在它们各自的“值字段设置”功能中都设置为求和, 如图5-47所示。

通过数据透视表的分析, 我们可以得知B、C、D三个品牌知名度在用户认知中比A和E高, E品牌知名度相对较低。

第二个问题, 同样只需在第一个问题的数据透视表中增加一个“性别”维度, 也就是将“性

别”字段拖至行标签区域，如图5-48所示。

图5-47 多选题分析之整体分析结果（示例2）

图5-48 多选题分析之性别分析结果示例

通过数据透视表的分析，我们可以得知B、D、E三个品牌的知名度在男性用户中相对较高，而A、C两个品牌知名度相对较低；A、C两个品牌知名度在女性用户中相对较高，而B、D、E三个品牌知名度在女性用户中则相对较低。

第三个问题，由于原数据中没有直接给出年龄段信息，需要新增一个“年龄段”分组字段，可利用VLOOKUP函数根据“年龄”字段进行年龄段的分组，如图5-49所示。

	A	B	C	D	E	F	G	H	I	J	K	L	M
	编号	性别	年龄	学历	A	B	C	D	E	年龄段	分组	值	分组
1	10001	男	35	大学	1	0	1	0	1	中年	青年	0	青年
2	10002	女	26	大学	1	1	1	0	1	青年	中年	30	中年
3	10003	女	22	中专	0	1	1	1	0	青年	中年	50	中年
4	10004	女	28	大学	1	0	1	1	0	青年	中年	0	青年
5	10005	男	40	高中	0	1	0	1	1	中年	青年	30	中年
6	10006	女	36	高中	1	0	1	0	1	中年	青年	50	中年
7	10007	女	41	高中	0	1	0	1	1	中年	青年	0	青年
8	10008	女	45	大学	1	0	1	0	1	中年	青年	30	中年
9	10009	女	50	中专	1	1	0	1	0	老年	青年	50	中年
10	10010	女	52	中专	1	1	1	0	0	老年	青年	0	青年
11	10011	女	36	研究生	0	1	1	1	0	中年	青年	30	中年
12	10012	女	42	大学	1	0	1	1	0	中年	青年	50	中年
13	10013	女	40	高中	0	1	0	1	1	中年	青年	0	青年
14	10014	女	35	研究生	1	0	1	0	1	中年	青年	30	中年
15	10015	女	26	大学	0	1	0	1	1	青年	青年	0	青年

图5-49 多选题分析之年龄段分组分析示例

选择增加“年龄段”字段的新数据，创建空白数据透视表框架。然后，将A、B、C、D、E五个品牌字段拖至数值汇总区域，这几个字段汇总方式在它们各自的“值字段设置”功能中都设置为求和，将“年龄段”字段拖至行标签区域，如图5-50所示。

通过数据透视表的分析，我们可以得知A、B两个品牌知名度在老年用户中相对较高，B、D、E三个品牌知名度在青年用户中相对较高，C品牌知名度在中年用户中相对较高。

第四个问题，同样只需在第一个问题的数据透视表中增加一个“学历”维度，也就是将“学历”字段拖至行标签区域，如图5-51所示。

通过数据透视表的分析，我们可以得知各品牌知名度在不同学历人群间的差异较不明显。

图5-50 多选题分析之年龄段分析结果示例

图5-51 多选题分析之学历分析结果示例

Mr.林微笑着说道：小白，很好，数据处理及数据分析的相关知识都掌握应用得不错。

小白脸红着说：把人家夸得都不好意思了！

5.3 本章小结

Mr.林：小白，关于数据分析方法及工具今天就先介绍这些，现在我们来简要复习下今天学过的知识。

小白：好的。

Mr.林：今天介绍的知识主要有：

- ★ 数据分析三大作用对应的基本分析方法，以及由基本方法发展而来的对比分析、分组分析、结构分析、平均分析、交叉分析、综合评价分析、杜邦分析、漏斗图分析、矩阵关联分析等数据分析法，并且在矩阵分析的基础上进行发散性思考，学会举一反三地学习数据分析法。
- ★ Excel中数据分析工具——数据透视表，并且我通过三个实用的数据透视表小技巧，介绍了如何利用数据透视表实现多选题分析。

小白：好的，今天辛苦Mr.林啦！下班后我一定好好复习，牛董布置的任务我知道要怎样完成了。

Mr.林：OK，有问题再来找我。

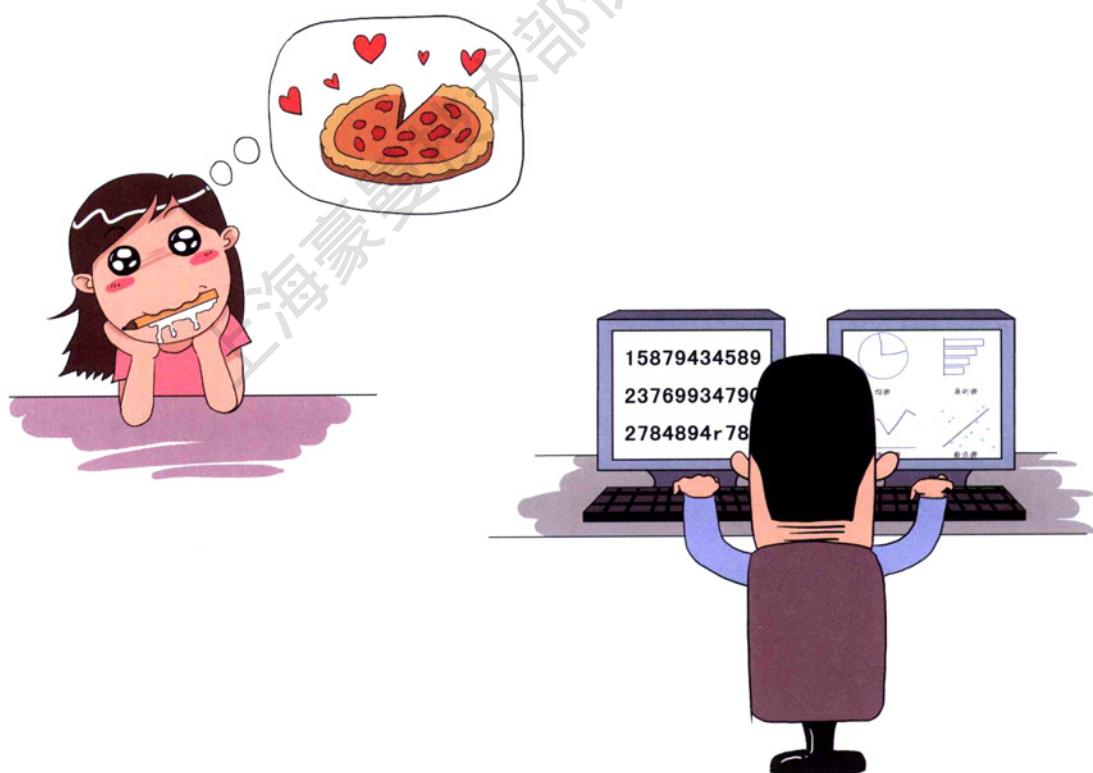
第 6 章

给数据量体裁衣——数据展现

揭开图表的真面目

表格也疯狂

给图表换装



小白按Mr.林传授的数据分析方法及数据透视表分析工具，熬了几天终于弄出员工满意度数据分析的结果，她抑制不住内心的激动，满怀期待地将几十页表格交给了牛董。**不料牛董扫了一眼，龙颜大怒：**“你这是要我在蚂蚁里面挑芝麻吗？给我重新修改！”

小白傻眼了，一筹莫展，只好再回来请教Mr.林。Mr.林首先跟小白介绍了“电梯法则”。

电梯法则：即“麦肯锡30秒电梯理论”，来源于麦肯锡公司一次沉痛的教训。该公司曾经为一家重要的大用户做咨询。咨询结束的时候，麦肯锡的项目负责人在电梯间里遇见了对方的董事长，该董事长问项目负责人：“你能不能说一下现在的结果呢？”由于该项目负责人没有准备，而且即使有准备，也无法在电梯从30层到1层的30秒钟内把结果说清楚。最终，麦肯锡失去了这一重要用户。

于是，小白在职场日记中记下了这样一句话：让老板在30秒钟内读懂你的数据。

Mr.林继续说道：你有没有想过怎样让老板快速理解你的数据报告？你可能会说要用图表说话，那么如何用图表说话？一般常用的图表基本类型都有哪些？它们各有什么特点？一般在什么情况下用？最后，图表的高级应用你知道多少呢？

小白哑口无言，悔恨地想起一句话“书到用时方恨少”，她十分委屈：以前也在图书馆借过《Excel宝典》之类书，但每本都400多页，从单元格开始讲起，就望而生畏了。其实我只想解决几个小问题，拿这么厚一本书砸下来，这不是狗抓刺猬——无从下手吗？

Mr.林：的确，专业书籍里面的内容太细了，有很多知识在工作中并不常用。接下来我带你看看图表的真面目，只需要一个小时就能基本解决你工作上所遇的所有图表问题！

小白两眼放光：好啊！

6.1 揭开图表的真面目

Mr.林：小白，下面我们从图表的作用、类型以及如何选择图表三方面来揭开图表的真面目。

6.1.1 图表的作用

我们先来了解图表在日常的工作、学习中都有些什么作用。使用图表来展示数据主要有三个作用。

◎ 表达形象化

使用图表可以化冗长为简洁，化抽象为具体，化深奥为形象，使读者或听众更容易理解主题和观点。

◎ 突出重点

通过对图表中数据的颜色和字体等信息的特别设置，可以把问题的重点有效地传递给读者或听众。

◎ 体现专业化

恰当、得体的图表传递着制图者专业、敬业、值得信赖的职业形象。专业的图表会极大地提升个人的职场竞争力，为个人发展加分，为成功创造机会。

6.1.2 经济适用图表有哪些

Mr.林：小白，咱们先做个测试，你知道学习和工作中经常用到的图表有哪些吗？

小白想起自己在报刊、杂志、网站上见到的许许多多千奇百怪的图表，回答：我还真留意过很多图表，它们都可以归为几类，比如饼图、柱形图、条形图、折线图等等。

Mr.林：还有吗？

小白歪着头，凝思了片刻：实在想不起来了……

Mr.林：还有散点图。最后不要忘了，还有一个最普通的——表格，我们常说的图表就是图形+表格，如图6-1所示。这些都是我们工作中经常用到的。其实大部分复杂的图表都可以由图6-1中的饼图、条形图、柱形图、折线图、散点图、表格等六类基本图表衍生而来。

简单的往往是最有效的，简单的图表往往更能有效、形象、快速地传递信息。

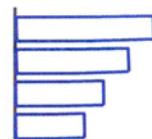
小白：嗯，现在流行的“经济适用男”也是同样的道理，简单实用才是真理，不妨称这六类图表为“经济适用图表”吧，哈哈！

地区	销售额
东部	¥12
西部	¥57
中部	¥123
总计	¥192

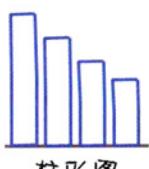
表格



饼图



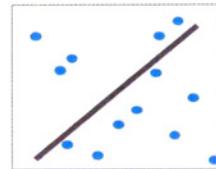
条形图



柱形图



折线图



散点图

图6-1 经济适用图表类型

Mr.林笑道：亏你想得出来，那你知道这些“经济适用图表”都在什么情况下使用吗？只有掌握了这些原则，才可以用图表清晰地表达你所要表达的主题和内容，用图表说话就不像以前一样困难和纠结了。

小白耍了个鬼脸：不知道呀，但我知道您接下来会讲的。

6.1.3 通过关系选择图表

Mr.林有点哭笑不得：下面就教你数据图表化第一招：通过数据间的关系来选择图表。大部分的数据间关系可以归纳为以下六种类型：成分、排序、时间序列、频率分布、相关性、多重数据比较。

◎ 成分

成分也叫做构成，用于表示整体的一部分，一般情况下用饼图表示。

这点好理解，好比一对小情侣吃一个比萨，假如比萨被均匀切成了4块，女生先吃了1块，也就是女生先吃了整个比萨的25%，这就是成分关系，如图6-2所示。这样的部分与总体关系用一个饼图表示最好不过了。

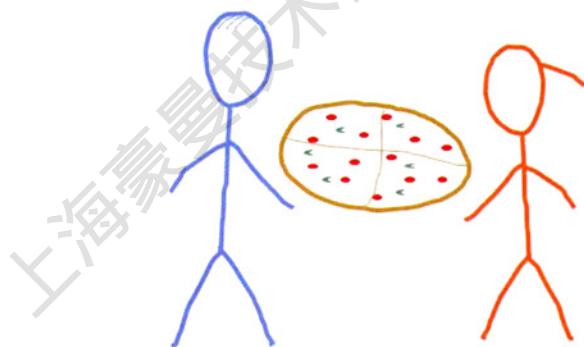


图6-2 某对情侣食用比萨的饼图示例

除了饼图外，还有什么图可以表示成分呢？如图6-3所示，柱形图、条形图、瀑布图都可以表示成分关系，我在后面会陆续介绍。

小白很诧异：柱形图也能表示成分关系？

Mr.林一脸神秘：当然能啦！打开Excel，单击“插入图表”项，查看柱形图所有的图表，你就会发现一个“百分比堆积柱形图”。何谓百分比？不就是表示成分吗？现在还可以用百分比堆积柱形图来表示几对情侣男女生比萨食用情况，如图6-4所示。

要表达的数据和信息	建议采用图形					
	饼图	柱形图	条形图	折线图	气泡图	其他
成分 (整体的一部分)						
排序 (数据间比较)						
时间序列 (走势、趋势)						
频率分布 (数据频次)						
相关性 (数据间关系)						
多重数据比较						

图6-3 常用图表类型与作用

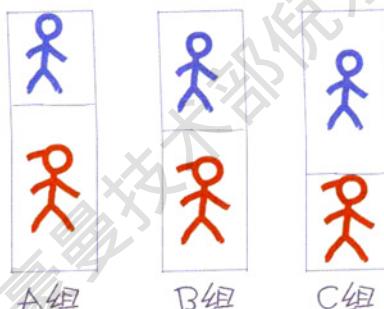


图6-4 三对情侣食用比萨的柱状图示例

Mr.林: 小白，你看哪一组情侣的女生饭量最大了吗？很明显是A组情侣的女生饭量大，有意思吧？

同样的道理，横坐标还可以是时间维度，表示是某对情侣饭量占比随时间变化的情况，可以根据所要表达的主题来选择数据及维度。

Mr.林知道小白最爱吃比萨，果然严重勾起了小白的味蕾兴趣，她现在满脑子的比萨：我能不能用多个饼图来表示几对情侣间各男女比萨食用情况？

Mr.林: 当然可以，但是要考虑到读者是否能快速看懂你的图，而不能让他们花费大量时间寻找两张图的直接关系。咱们这个例子只有男女两个类别，如果再多点分类，那就要把你的读者或老板累坏了。

◎ 排序

Mr.林: 小白，成分关系搞清楚了吗？下面我们要开始介绍第二种数据关系：排序。

排序，顾名思义就是根据需要比较的项目的数值大小进行排列，也就是可以按数值从大到小降序排列，或者从小到大升序排列。还是那句话，根据你所要表达的主题作图。排序可用于不同项目、类别间数据的比较。

例如，对2010年某集团各分公司年度业绩进行降序排序，如图6-5所示，我们就很容易看到该集团2010年业绩前三名是A、C、G三个分公司，当然从图中也能知道最后三名，甚至随意指定某个分公司，其业绩与排名也是非常容易得知的。从图6-3能看到，可排序的图有柱形图、条形图、气泡图、帕累托图，这些都放在后面教你吧。

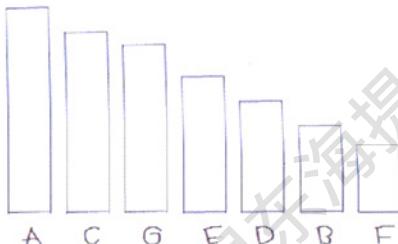


图6-5 2010年某集团各分公司年度业绩排名

① 时间序列

小白点点头表示已经理解，Mr.林继续说：我们接着讲第三种数据关系：时间序列。

时间序列用于表示某事按一定的时间顺序发展的走势、趋势。

例如，我们想知道2010年上半年某集团业绩每月的发展情况是怎样的。如图6-6所示，这时候就不能按业绩大小进行排序，我们可以按照时间的顺序作一个柱形图，这样能看出其发展趋势。当然也可以作折线图，折线图是最直观的趋势图，但是不要用同样的数据在同一张图中既作柱形图又作折线图，柱形图就可以表示趋势，没必要再画蛇添足了。如果你习惯使用折线图，那就使用折线图吧。

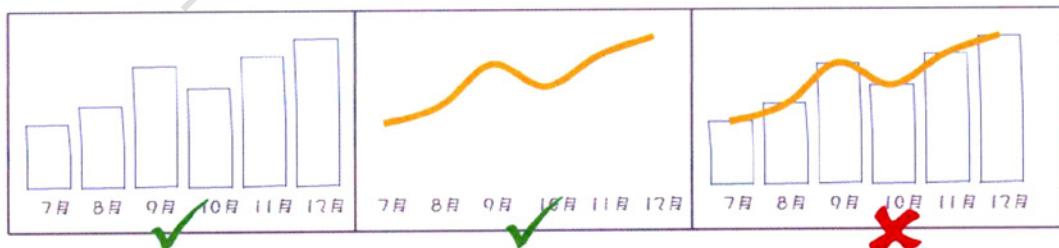


图6-6 2010年上半年某集团业绩发展

Mr.林：可以表示时间序列的还有面积图，如图6-3所示。

小白显然没见过：面积图也可以表示时间序列？

Mr.林: 当然,你看区域上面的线不就是折线吗?那就可以表示趋势了,但是面积图有个缺点,就是上面各数据系列之间可能会出现相互遮挡的情况,难以清晰地看出趋势。

所以表示时间趋势时可首选折线图,其次可考虑使用柱形图。

◎ 频率分布

Mr.林: 小白,刚才说的时间序列关系消化了没有?

小白: 嗯,我都记着呢,现场消化不了的,回去再慢慢消化。

Mr.林: 好的,我们要继续学习第四种数据关系:频率分布。

频率分布与排序一样,都用于表示各项目、类别间的比较,当然这一类比较也可用频数分布表示,只是单位不同,根据需要选择即可。可以说频率分布是一种特殊的排序类图形,因为它只能按指定的横轴排。

例如,我们作图表示2010年某集团产品不同价格区间的销量,那么横轴只能按惯例采用价格从低至高排,而不能按商品的销量从高到低排,如图6-7所示。当然你非要这样排的话也行,只是这样会让价格区间比较乱,不过还是那句话,根据你所要表达的主题进行作图。用于表现频率分布的图表还有柱形图、条形图、折线图。

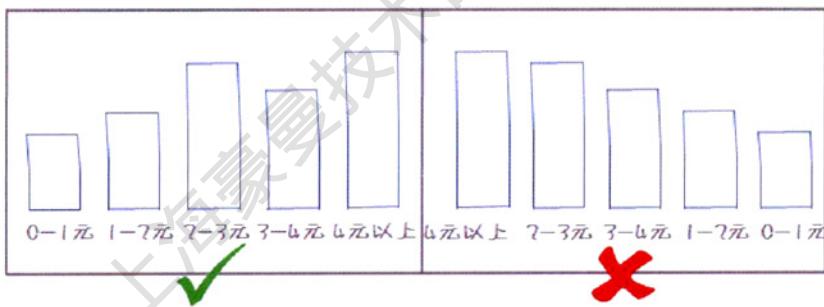


图6-7 2010年某集团各价格区间产品的销量分布

◎ 相关性

怎么解释相关性呢? Mr.林思考片刻,说道:相关性这个词听上去蛮专业的,我用一个身边例子来解释吧。小白,你最近有没有关注生姜价格上涨?随着生姜价格逐步提高,阿姨们每次购买生姜的数量是否有变化?这就是相关性,相关性用于衡量两大类中各项目间的关系,即观察其中一类的项目大小是否随着另一类项目大小有规律地变化。如图6-8所示,我们可以明显看到生姜价格越高销量就越低。

用于衡量相关性的图表还有柱形图、对称条形图(旋风图)、散点图、气泡图。

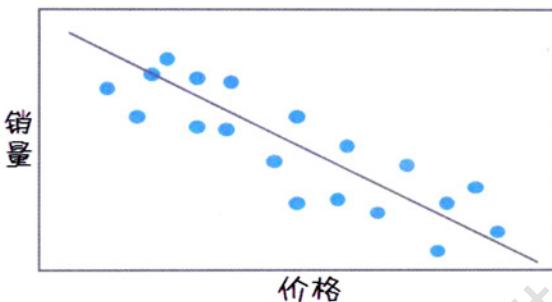


图6-8 生姜价格与销量关系分布

◎ 多重数据比较

就剩下最后一个关系了，Mr.林顿了顿，喝了口水，接着说：多重数据比较就是指数据类型多于2个的数据分析比较。例如要比较A、B、C三种电脑分别在品牌、价格、内存、CPU、硬盘、售后6项指标中的评分情况，可以用簇状柱形图，如图6-9所示。

Mr.林：这幅图给你什么样的感觉？

小白果然是神游去了，才反应过来，回过神想了想：这张图，像一堆柱子挤在一块儿，好臃肿啊！

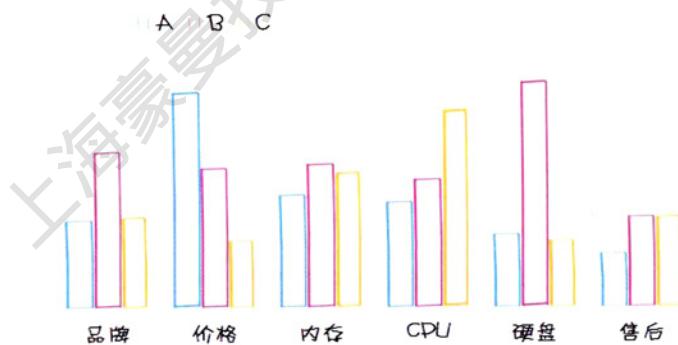


图6-9 A、B、C三类产品在6个指标中的评分（簇状柱状图）

Mr.林：差不多是这个意思。再想想看，现在只有3个类别、6个指标，如果再扩充到4个类别、10个指标，这将是什么样？40根柱子排排站！如果你的报告用Word写，有一个比较野蛮的解决办法，就是把那一页Word文档设置为横向打印。可是我们为什么不用一个更简单的图形表达呢？

小白条件反射地问道：折线图？我觉得折线图和柱形图像一个娘胎里的兄弟，经常在互换角色！

Mr.林：这说明你还没有充分理解折线图。刚才介绍了折线图可以表现趋势或者分布，你这6个指标之间有关系吗？如果没有关系，那么用折线图就会误导读者。

小白一脸茫然：用什么图合适呢？

Mr.林：我们可以用下面的雷达图来表示，如图6-10所示。小白你看，用了雷达图，效果是不是好了很多。

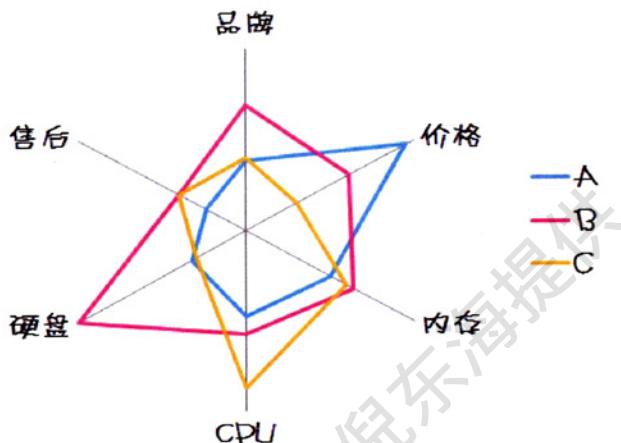


图6-10 A、B、C三类产品在6个指标中的评分（雷达图）

Mr.林：当然，数据的关系及相对应可用的图表不限于刚才介绍的内容，这里主要介绍你工作中常用的图表类型，比如表示成分时使用饼图居多，当然也可以用圆环图来表示。只要能正确地表达你自己所要表达的主题或内容即可。

小白听完茅塞顿开。

6.1.4 图表制作五步法

Mr.林：小白，经过刚才的学习，是不是对什么情况用什么图已经有了清晰的轮廓了？

小白：嗯，思路清晰多了。

Mr.林：确定了所要表达的信息和关系之后，接下来的制图工作就简单了。图6-11给出制图五步法，按这五步依次进行就可以作出所要的图表啦！最重要的就是第1步，如果第1步的主题和目的不明确，后续步骤也会让读者无法准确清楚地理解你所表达的内容。

小白仔细看了看作图五步法，冒出一个念头：怎么觉得这跟我逛街挑衣服一样？

Mr.林：真是本性难移，好端端的作图怎么又跟逛街挑衣服扯上关系了？

小白：你看嘛，下面我画一个对比关系图出来，你就知道我说的有道理了（如图6-11所示）。

就拿我自己做个比方，我走的是知性白领风格，所以直线条、有质感的带领衣服会比较适合我，然后按照这个标准挑选衣服，例如职业套裙就不错，再试穿检查合不合身，能不能突显我的知性风格。

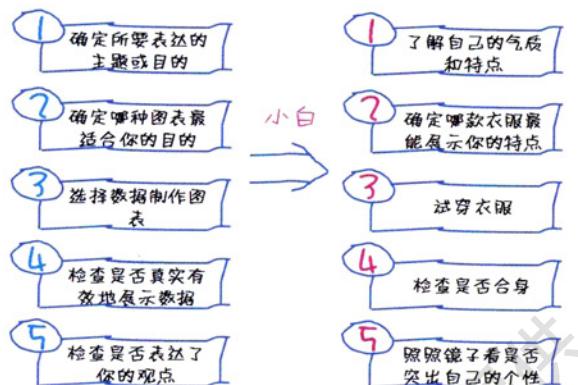


图6-11 图表制作五步法

Mr.林哭笑不得：真拿你没办法，不过也有点道理……

小白倒是一本正经：这叫做触类旁通，这样不是更容易理解和记忆吗？

6.2 表格也疯狂

Mr.林：学会了怎样将数据以图形的形式呈现只是使用图表的第一步，大部分人呈现数据结果时都采用图形，而没有发现表格的精妙之处。下面再告诉你数据图表化第二招：必杀技——用表格来表示“图所不能说的话”。

小白：表格还能解决图不能解决的问题？表格不是地球人都会吗，有什么特别的地方呢？

Mr.林解释道：当需要呈现的数据在3个系列及其以上，尤其是数据间的量纲不同的时候，用表格呈现数据的效果相对较好。Excel 2007或2010版本在表格方面的呈现效果已经比之前的其他版本表现更佳，下面我将介绍条件格式中几个方便好用的功能。它们分别是突出显示单元格、项目选取、数据条、图表集和迷你图。

6.2.1 突出显示单元格

Mr.林边说边打开了Excel演示给小白看：突出显示单元格就是根据指定的规则，把表格中符合条件的单元格用不同颜色背景、字体颜色将数据突出显示出来。可设置的常用规则有：大于、小于、等于、介于、重复值，当然可以另外设置其他规则。如图6-12所示，例如我们知道2004年中国东部有哪些地区的国内生产总值大于10000亿元¹，那么在Excel 2010中操作如下。

- STEP 01** 选取“2004年”这一列的所有单元格数据。
- STEP 02** 单击“开始”选项卡，在“样式”组中，单击“条件格式”。
- STEP 03** 选择“突出显示单元格规则”→“大于”，弹出“条件格式规则管理器”对话框。
- STEP 04** 在输入框中输入“10000”，并设置所需突出显示的单元格样式，在这里我们选择“黄填充色，深黄色文本”，并单击“确定”按钮。

通过上述操作即可把2004年中国东部生产总值大于10000亿元的区域突显出来，如图6-12所示，是不是很简单？

地区	2004年	2005年	2006年	2007年	2008年
北京	6060	6886	7861	9353	10488
天津	3111	3698	4344	5050	6354
河北	8478	10096	11516	13710	16189
辽宁	6672	8009	9214	11023	13462
上海	8073	9154	10366	12189	13698
江苏	15004	18306	21645	25741	30313
浙江	11649	13438	15743	18780	21487
福建	5763	6569	7584	9249	10823
山东	15022	18517	22077	25966	31072
广东	18865	22367	26160	31084	35696
海南	799	895	1032	1223	1459

图6-12 中国东部各地区2004—2008年国内生产总值（单位：亿元）

6.2.2 项目选取

Mr.林：项目选取其实跟突出显示单元格的意思基本一样，等介绍完项目选取再为你演示一下。项目选取同样是根据指定的规则，把表格中符合条件的单元格用不同颜色的背景、字体颜色将数据突出显示出来。区别就在于指定的规则不同，突出显示单元格的规则指定值是与原始数据直接相关的数据，如上文提到的大于10000亿元；而项目选取指定的规则指定值是对原始数据经过计算的数据，如数值最大的10%项、数值最小的10%项、高于平均值、低于平均值等。

如图6-13所示，我们想知道2005年中国东部有哪些地区的国内生产总值高于平均值，那么我们可以在Excel 2010操作如下：

- STEP 01** 选取“2005年”这一列的所有单元格数据。
- STEP 02** 单击“开始”选项卡→“样式”→“条件格式”。
- STEP 03** “项目选取规则”→“高于平均值”，弹出“条件格式规则管理器”对话框。
- STEP 04** 针对选定区域，设置项目选取所需的单元格样式，在这里我们选择“绿填充色，深绿色文本”，并单击“确定”按钮。

通过上述操作即可把2005年中国东部国内生产总值高于平均值的地区选取出来，得到的效果如图6-13所示。

地区	2004年	2005年	2006年	2007年	2008年
北京	6060	6886	7861	9353	10488
天津	3111	3698	4344	5050	6354
河北	8478	10096	11516	13710	16189
辽宁	6672	8009	9214	11023	13462
上海	8073	9154	10366	12189	13698
江苏	15004	18306	21645	25741	30313
浙江	11649	13438	15743	18780	21487
福建	5763	6569	7584	9249	10823
山东	15022	18517	22077	25966	31072
广东	18865	22367	26160	31084	35696
海南	799	895	1032	1223	1459

图6-13 中国东部地区各省2004—2008年国内生产总值(单位:亿元)

6.2.3 数据条

Mr.林接着说：数据条可帮助你查看某个单元格相对于其他单元格的值。数据条的长度代表单元格中的值。数据条越长，表示值越大；数据条越短，表示值越小。在观察海量数据时，如需直观显示数值或数值区间，数据条尤其有用。

如图6-14所示，我们想知道2006年中国东部有哪个地区的国内生产总值最高，哪个地区最低，那么我们可以在Excel 2010按如下步骤操作。

STEP 01 选取“2006年”这一列的所有单元格数据。

STEP 02 单击“开始”选项卡→“样式”组，单击“条件格式”。

STEP 03 选择“数据条”→“蓝色数据条”，当然你也可以根据自己的喜好选择其他颜色的数据条。

通过上述操作即可把2006年中国东部国内生产总值的最高值与最低值的地区选取出来，如图6-14所示，最高为广东，最低为海南。

地区	2004年	2005年	2006年	2007年	2008年
北京	6060	6886	7861	9353	10488
天津	3111	3698	4344	5050	6354
河北	8478	10096	11516	13710	16189
辽宁	6672	8009	9214	11023	13462
上海	8073	9154	10366	12189	13698
江苏	15004	18306	21645	25741	30313
浙江	11649	13438	15743	18780	21487
福建	5763	6569	7584	9249	10823
山东	15022	18517	22077	25966	31072
广东	18865	22367	26160	31084	35696
海南	799	895	1032	1223	1459

图6-14 中国东部各地区2004—2008年国内生产总值(单位:亿元)

当需要呈现的数据在3个系列及其以上，尤其是数据间的量纲不同，将数据在单元格中以数据条的形式用表格呈现，相当于3个微型条形图，效果会比将3个不同量纲的系列数据画在一张图上好很多。

6.2.4 图标集

Mr.林停顿了一下：是不是有些累了，小白？

小白：没有呢，我在听，Go on！

Mr.林点头：好的，还剩下图标集和迷你图没有介绍。先讲图标集，使用图标集可以对数据进行注释，并可以按阈值（就是临界值），将数据分为3~5个类别。每个图标代表一个范围的值。

例如，在三向箭头图标集中，绿色的上箭头代表较高值，黄色的横向箭头代表中间值，红色的下箭头代表较低值。

我们可对2007年中国东部地区的国内生产总值进行分类，国内生产总值在20000亿元及其以上的地区用绿色带钩圆圈表示，大于或等于10000且小于20000亿元的用黄色带感叹号圆圈表示，小于10000亿元用红色带叉圆圈表示，效果如图6-15所示。

地区	2004年	2005年	2006年	2007年	2008年
北京	6060	6886	7861	9353	10488
天津	3111	3698	4344	5050	6354
河北	8478	10096	11518	13710	16189
辽宁	6672	8009	9214	11023	13462
上海	8073	9154	10366	12189	13998
江苏	15004	18306	21645	25741	30313
浙江	11649	13438	15743	18780	21487
福建	5763	6569	7584	9249	10823
山东	15022	18517	22077	25966	31072
广东	18865	22367	26160	31084	35696
海南	799	895	1032	1223	1459

图6-15 中国东部各地区2004—2008年国内生产总值（单位：亿元）

另外，图标集特别适用于企业运营指标发展态势的监控。

如图6-16所示，某集团给下属的A、B、C三个企业2010年第一季度收入目标定为1亿元，而A、B、C三个企业2010年第一季度实际收入分别为1.1亿元、0.95亿元、0.73亿元，那么完成率分别为110%、95%、73%。我们需要对各企业的完成率数据进行处理，让老板或同事清晰快速地知道哪些企业的业绩完成得好，哪些企业完成得差。这时就可以使用图标集功能进行标注。

项目	A企业	B企业	C企业
第一季度收入目标（亿元）	1	1	1
完成值（亿元）	1.1	0.95	0.73
完成率	110%	95%	73%

图6-16 2010年第一季度某集团下属三企业的收入目标及完成情况

图标集功能也在“开始”选项卡“样式”组的“条件格式”中，我们可在图标集规则设置对话框中对完成率进行分类定义，规则设置如图6-17所示：完成率大于或等于100%的企业，用绿色带钩圆圈表示，完成率大于或等于90%且小于100%用黄色带感叹号圆圈表示，小于90%的用红色带叉圆圈表示。这样我们就可以清晰地看到完成业绩目标的企业有哪些，未完成业绩的有哪些，快完成的有哪些。



图6-17 图标集规则设置对话框

当数据越多，使用图标集功能的效果就越明显，还等什么？赶快动手实践一下吧！

小白摩拳擦掌：好啊！现在就做一遍！

6.2.5 迷你图

Mr.林：最后介绍迷你图。顾名思义，迷你图的特点就是小，小到可以塞到每一个单元格中，先给你看看效果，如图6-18所示。



图6-18 迷你图示例

小白一下子两眼放光，精神劲儿来了，惊叹到：哇塞，我可从没见过这么炫的图！

Mr.林：对！迷你图是Excel 2010的新功能，它是工作表单元格中的一个微型图表，可提供对数据的形象表示。使用迷你图可以显示数值系列中的趋势（例如，季节性增加或减少、经济周期），或者突出显示最大值和最小值。在数据旁边放置迷你图可达到最佳效果。

虽然Excel表格行或列中呈现的数据很有用，但很难一眼看出数据的分布形态。通过在数据旁边插入迷你图，可以清晰简明地显示相邻数据的趋势，而且迷你图占用的空间很少。

与Excel工作表上的图表不同，迷你图不是对象，它实际上是单元格背景中的一个微型图表。以图6-18中北京地区在2004—2008年国内生产总值的迷你图为例，在Excel 2010中的操作如下。

- STEP 01** 单击选定所要制作迷你图的单元格G2。
- STEP 02** 单击“插入”选项卡，在“迷你图”组中，选择“迷你折线图”。
- STEP 03** 弹出对话框，选择所需制作迷你图的数据范围B2:F2。
- STEP 04** 单击“确定”按钮，即可生成迷你折线图。

如果天津、河北、辽宁、上海四地区也要制作出同样的迷你折线图，可以直接复制粘贴过去，数据位置也会相应随之发生改变。此外，除了迷你折线图，还可以制作迷你柱形图和迷你盈亏图，在第3步的操作中换成迷你柱形图或迷你盈亏图即可。如图6-18所示，江苏、浙江等6个地区采用了迷你柱形图。

另外还可利用“设计”选项卡中的“标记颜色”功能突出显示最高值（2008年）。

从图6-18的迷你图可以看出，中国东部地区各省份在2004—2008年的省内生产总值都呈增长趋势。

6.3 给图表换装

Mr.林：小白，学了前面两招，工作中80%以上的图表问题你都能解决，当然还会遇到一些棘手的问题。

小白偷笑：区分菜鸟和高手的关键时刻到了！

Mr.林：挺机灵的！对，下面将教你数据图表化第三招——图表换装。

当你遇到比较困难的图表问题时，采用这一招即可奏效。通常，对于复杂一点的问题需使用稍微复杂的图表才能呈现我们要表达的内容，如：双坐标图、矩阵图、旋风图、漏斗图等，这些图表在专业的调研、咨询公司里使用相对较频繁。其实它们也都是由“经济适用图表”变换而成的，经济适用图表经过巧妙地换装，有时会达到意想不到的效果，尤其是矩阵图，下面将一一进行介绍。

6.3.1 平均线图

Mr.林: 平均线图并不是只在图上单独画条平均线，而是在原来的柱形图或折线图的基础上，添加一条平均线，如图6-19所示。

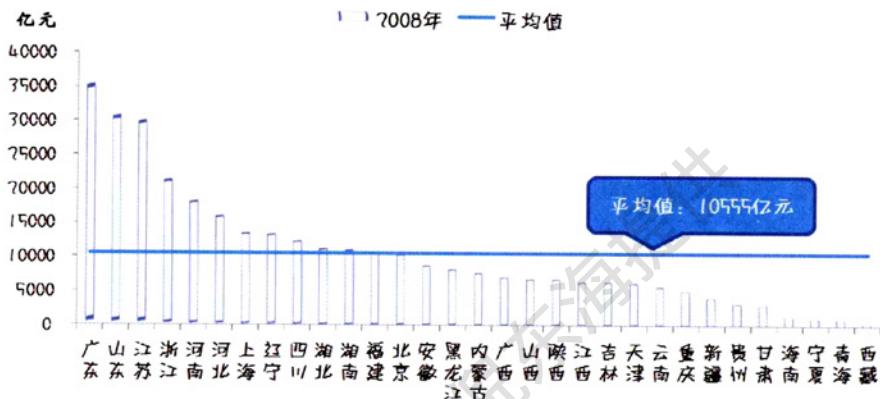


图6-19 2008年中国各省国内生产总值对比完成情况(单位:亿元)

小白的语气明显有些鄙夷：这有什么难的？这个还需要强调吗？

Mr.林：对，大家都觉得平均线图简单，没有引起对它的重视。其实很多数据分析的项目只需要充分贯彻“对比”二字，而平均线图即“对比分析”的典型图例，它可以用来对比图中各项目之间与平均线的差距，有哪些项目处于平均线上方？有哪些项目处于平均线下方？当然这条参考线数据不一定要采用平均线，也可以采用其他值，比如在数据分析原理对比分析中提到的几个比较标准。最简单的方法往往是最有效的，不过它们容易被人忽略，意识不到它的重要性。

小自：怎么绘制平均线图呢？

Mr.林: Excel 2003有线柱图的图表模板可直接套用，但到Excel 2007及2010版本就取消了图表模板，所以需要我们用户绘制。

小自：啊？自己画？

Mr.林: 很简单的,不要怕。现在就来介绍一下在Excel里制作平均线图的方法,看看怎样制作出图6-19这样的平均线图。数据源表格如图6-20所示。

STEP 01 先准备数据，即先算出图6-20中所有数据的平均值，并且在数据表中增加一列，其数据都为平均值。如图6-20的C列所示，增加了一列，数值均为2008年中国各省国内生产总值平均值10555亿元。

STEP 02 选取A1:C32单元格中的数据制作柱形图。

STEP 03 选择柱形图中“平均值”的任一柱条，单击鼠标右键，选择“更改系列图表类型”，把柱形图改为折线图（不带数据标记），到这一步为止平均线图基本制作完成。

为了便于查看图表，我们可以对2008年各省国内生产总值按照降序排序，如图6-19所示，这样就可以一目了然地知道广东、山东、江苏、浙江等13个地区国内生产总值高于全国平均值，而安徽、黑龙江、内蒙古、广西等18个地区国内生产总值低于全国平均值。

为了读者更清晰快速地了解图所要表达的主题，我们还可在图表的平均线上添加一个标注，注明此线为平均线，并且标出平均值为10555亿元，如图6-19所示。

我在前面也提到折线图是可以制作平均线图的，其作法相对简单，选择数据直接作折线图，不需要再对其中的平均线更改图表类型，因为它已经是折线图。其他步骤基本一致，我就不多说了，免得你嫌我啰嗦。

6.3.2 双坐标图

Mr.林：小白，刚才向你介绍了平均线图，现在我们继续来学习下一个图表，双坐标图。你知道双坐标图是什么样的吗？一般在什么情况使用这样的图形呢？

小白睁着大眼睛望着Mr.林：还没有见过，连平均线图我都是第一次在你这见到。不过从字面意思理解，双坐标图应该是比平常的图形多一个坐标轴，Mr.林，是这样的吗？

Mr.林：小白，有长进嘛，确实是这样的。双坐标图比平常的图形多了一个纵坐标轴，我们称之为次纵坐标轴，有次就有主，主坐标轴就是我们常用的左轴，可简称左轴，次纵坐标轴在右轴显示，故也可简称为右轴。

一般在图表中有两个系列及其以上的数据，并且它们的量纲不同或者数据差别很大时，在同一纵坐标轴下就无法很好地展现出数据原本的面貌，这时我们就可以采用双坐标图来绘制。

如图6-21所示，在同一个纵坐标轴下，我们就无法很好地了解2010年该集团的收入情况，因为收入与产品的销量单位不同，且数量差别太大，这时我们可把收入采用次纵坐标轴来展示，做法如下。

图6-19 中国04年-08年各省国内生产总值.xls		
	A	B
1	地区	2008年
2	广东	35696
3	山东	31072
4	江苏	30313
5	浙江	21487
6	河南	18408
7	河北	16189
8	上海	13698
9	辽宁	13462
10	四川	12506
11	湖北	11330
12	湖南	11157
13	福建	10823
14	北京	10488
15	安徽	8874
16	黑龙江	8310
17	内蒙古	7762
18	广西	7172
19	山西	6939
20	陕西	6851
21	江西	6480
22	吉林	6424
23	天津	6354
24	云南	5700
25	重庆	5097
26	新疆	4203
27	贵州	3333
28	甘肃	3178
29	海南	1459
30	宁夏	1099
31	青海	982
32	西藏	393

图6-20 中国2008年各省国内生产总值（单位：亿元）

>> 谁说菜鸟不会数据分析（入门篇）

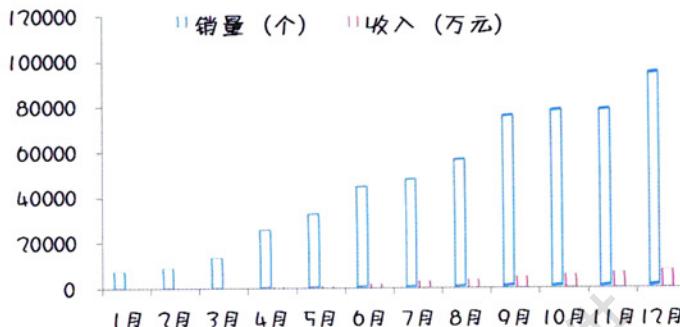


图6-21 2010年每月某集团产品的销量及收入（单坐标轴）

STEP 01 选择柱形图中任一“收入”柱子，单击鼠标右键，选择“设置数据系列格式”。

STEP 02 在弹出的“设置数据系列格式”对话框中，如果未选择“系列选项”，需要单击此项，然后在“系列绘制在”选项下，单击“次坐标轴”，然后单击“关闭”按钮。

Mr.林：小白，看看效果吧。

小白担忧地说道：咦！怎么两个图都重叠在一起了？这样怎么能看清楚啊？**Mr.林**您肯定有办法解决这个问题，快教教我如何解决。

Mr.林：小白，你不会这么快就忘了吧，刚才和你说过的！

小白立马恍然大悟：是跟平均线图作法一样，将“收入”这一系列的柱形图改为折线图，对吗，**Mr.林**？

Mr.林：完全正确，加10分。

STEP 03 选择“收入”这一系列柱形图中任一柱条，单击鼠标右键，选择“更改系列图表类型”，把柱形图改为“带数据标记的折线图”，到这一步为止，双坐标图基本制作完成了。为了更好地区分销量与收入分别对应哪个坐标轴，可以在原数据表的字段标签增加左轴与右轴说明，如图6-22所示。

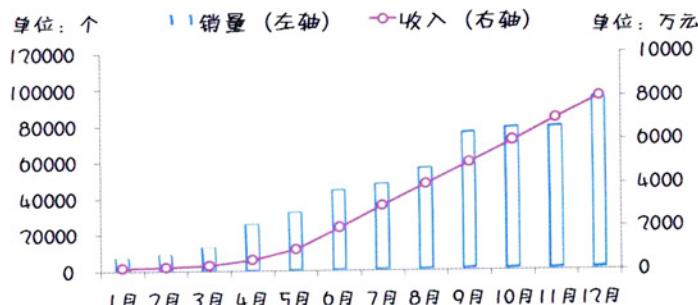


图6-22 2010年每月某集团产品的销量及收入（双坐标轴）

第6章 给数据量体裁衣——数据展现

Mr.林: 小白，你看用双坐标图来展示，效果是不是好很多了？这样的图由柱形图和折线图组成，我们常把它称之为线柱图。

小白快速地反应道: 那么双坐标图就是线柱图？

Mr.林: 当然不是，双坐标图也可以是双柱图、双线图。

小白: 双线图我知道怎么作了，两个折线图不会重叠在一起，除非是两个数据系列的数据完全相同，而双柱图刚才已经作过，重叠在一起了。Mr.林，怎么作双柱图呀？

Mr.林: 小白，你思考下，刚才作的柱图都重叠在一起了，那我们要怎样让它们不重叠呢？

小白: 在它们中间直接插入一个数据把它们隔开？

Mr.林: 思路对了，我们可以在它们中间直接插入数据，并且要插入两个数据，而不是一个。插入什么样的数据呢？首先要理解为何要插入这两个数据，目的就是让它们占住位置，但又不显示出来，让原本我们需要的两个实际数据隔开，我们称之为占位数据。

如图6-23所示，在销量和收入之间增加了两个占位数据。接下来我们就来看看到底如何利用占位数据制作双柱图。

月份	销量	销量占位	收入占位	收入
1月	7600	5000	2000	1520
7月	9235	3000	1500	1850

此处占位数据应为0，但为了方便操作设置，在此先以其他值替代，待完全设置好后再更改回0

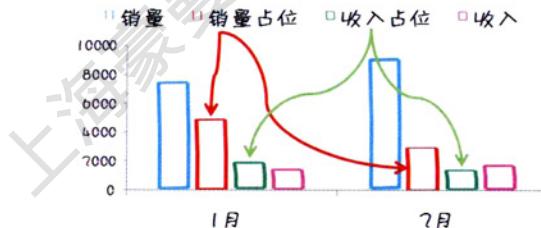


图6-23 双柱图示例数据及说明

STEP 01 选择数据制作柱形图。

STEP 02 将“收入占位”、“收入”两个系列柱图的“系列绘制”设置为次坐标轴，这时又会出现柱子两两互相重叠的情况。

STEP 03 把占位数据全部改为0，柱子重叠的情况将消失，剩下就是我们要的双柱图，如图6-24所示。

小白惊奇地说道: Mr.林，您太强了，佩服佩服！

Mr.林: 还要在图中作一些修改，如加上左右轴注释、删除占位数据图例、添加数据单位注释等，双柱图最终效果如图6-25所示。

>> 谁说菜鸟不会数据分析（入门篇）

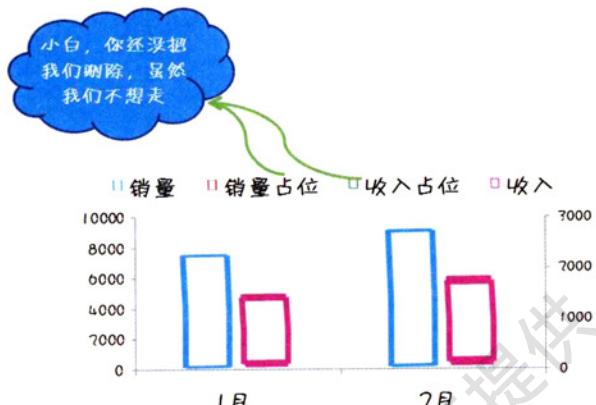


图6-24 双柱图(效果1)

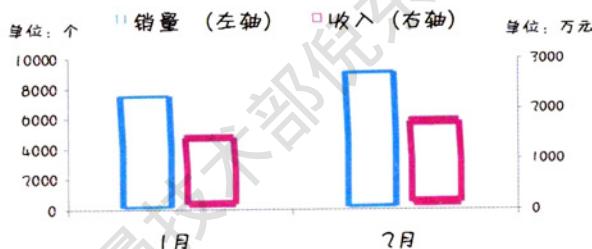


图6-25 双柱图(效果2)

6.3.3 竖形折线图

Mr.林：小白，双坐标轴图形会用了吧？

小白连忙点头：会了，会了。

Mr.林：接下来我们学习下一个图表，竖形折线图。

竖形折线图，从名字就可以知道，它是把折线图立起来，如图6-26所示，因为图形中的折线看起来像蛇一样的，所以也称之为蛇形图。竖形折线图主要在市场研究、咨询等公司使用较多，用它来展示产品功能、品牌形象等在消费者心中的评价，可用于分析与比较多个不同产品、项目在每个指标方面的表现，得出它们在每个指标属性偏向。一般采用消费者打分的形式获取数据，评分范围一般为0~5或0~10，越同意靠右的观点，选择的数字越高，越同意靠左的观点，选的数字越低。

竖形折线图就是采用“带平滑线和数据标记的散点图”绘制的，另外通过两个辅助数据来增加对应的标签。现在我们就用图6-27中的示例数据来作竖形折线图。

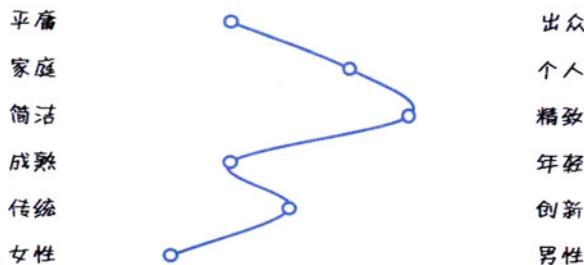


图6-26 某品牌在消费者心中的形象 (竖形折线图)

	A 平均评分 (横坐标)	B 辅助列1 (横坐标)	C 辅助列2 (横坐标)	D 纵坐标	E 辅助列1 标签	F 辅助列2 标签
1	1.5	0	5	1	女性	男性
2	2.5	0	5	2	传统	创新
3	2.0	0	5	3	成熟	年轻
4	3.5	0	5	4	简洁	精致
5	3.0	0	5	5	家庭	个人
6	2.0	0	5	6	平庸	出众

图6-27 某品牌在消费者心中形象数据示例

- STEP 01** 选择A2:A7, D2:D7两个范围的数据, 单击“插入”选项卡, 在图表“散点图”中选择“带平滑线和数据标记的散点图”。
- STEP 02** 在图表上单击右键, 选择“选择数据”项, 弹出“选择数据源”对话框, 单击“添加”按钮, 弹出“编辑数据系列”对话框, 分别添加系列名称(辅助系列1)、X轴值(B2:B7)、Y轴值(D2:D7), 单击“确定”按钮。
- STEP 03** 重复步骤2, 分别添加系列名称(辅助系列2)、X轴值(C2:C7)、Y轴值(D2:D7), 单击“确定”按钮, 操作步骤如图6-28所示, 效果图形如图6-29所示。
- STEP 04** 选择平均评分的折线图, 单击鼠标右键, 选择“添加数据标签”。
- STEP 05** 利用JWalk Chart Tools工具(使用方法详见第7章具体介绍), 添加辅助列1、辅助列2的标签。
- STEP 06** 分别通过“设置数据系列格式”功能, 把数据标记选项、线条颜色都设置为无, 这样既可以把两个辅助列隐藏起来, 又可以把辅助列的标签显示出来。
- STEP 07** 删除图例、网格线。
- STEP 08** 分别把横、纵坐标的标签、刻度线、线条颜色都设置为无, 即把它们也隐藏起来。
- Mr.林:** 竖形折线图制作完毕, 效果如图6-26所示, 小白都记住了吗? 回去要多多练习, 否则不久就会忘记的。

小白情不自禁地说: Mr.林, 我越来越佩服您了, 您是我的偶像啊!

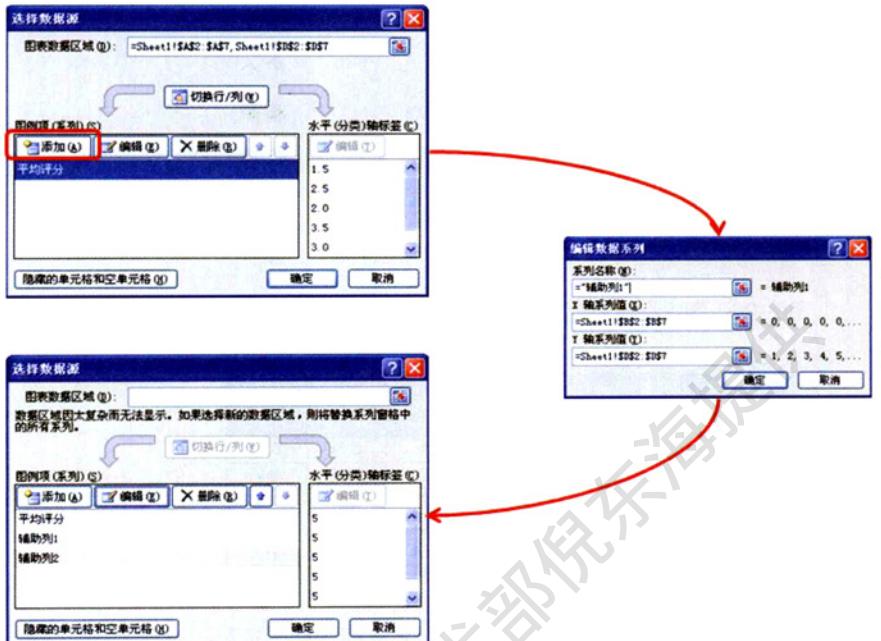


图6-28 竖形折线图添加数据系列示例图

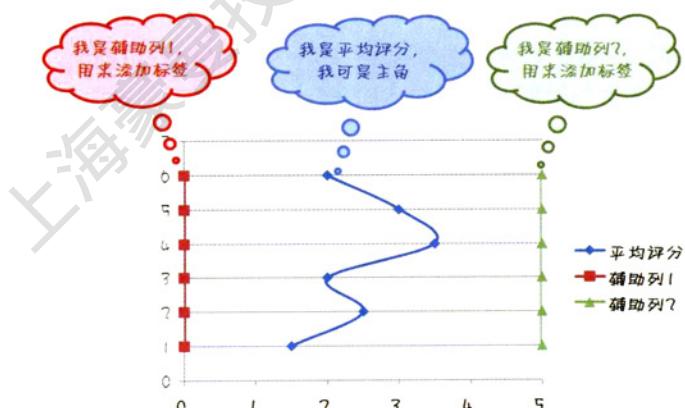


图6-29 竖形折线图过程示例图

Mr.林: 竖形折线图还有其他方式的作法，就是先作普通折线图，然后利用Excel里的照相机功能，拍照后将图形旋转90°，不过这只能显示左边的属性标签，无法显示右边的标签，需要使用辅助列添加右边的标签。这样的话，两个标签都使用辅助列的方法添加相对简单些。

小白，如果对照相机法有兴趣，可以自己回去好好研究研究。至于使用哪种方法方便、简单，就看你个人的感觉了。

6.3.4 瀑布图

Mr.林: 小白, 还记得之前说过的表示成分的图形都有哪些吗?

小白: 饼图、柱形图、条形图, 好像还有个叫瀑布图的, 我还没见过, 印象不深。

Mr.林: 说得非常正确。瀑布图, 顾名思义看起来就如同瀑布那样, 具有自上而下的流畅效果, 也有的人把它叫做步行图、阶梯图, 在企业的经营分析、财务分析中使用较多, 用于表示企业成本的构成、变化等情况。

如图6-30所示, 第1根柱子表示该集团的总成本, 后面的4根柱子为总成本的分解, 分别表示不同成本的情况。

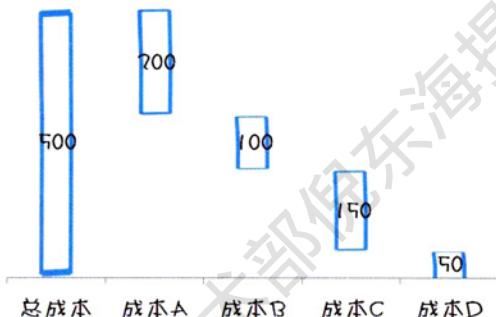


图6-30 2010年某集团成本构成示例

现在我们就一起来学习在Excel中如何制作瀑布图。首先需要思考, 如何把后面几个柱子悬空挂起来呢? 小白, 通过前面的学习, 你大概想到该怎么做了吗?

小白: 用占位的方法作图?

Mr.林: 小白, 看来你已经把所学的知识融会贯通了, 但不可以骄傲自满喔! 没错, 就是要用占位的方法作图。我们现在就以图6-31中2010年某集团成本数据为例, 学习如何制作瀑布图。

	B3	f4 = \$C\$2-SUM(\$C\$3:C3)	
1	项目	占位数	成本
2	总成本	0	500
3	成本A	300	200
4	成本B	200	100
5	成本C	50	150
6	成本D	0	50

图6-31 2010年某集团成本数据

STEP 01 准备占位数据, B2单元格中的数据为0, 需要从B3单元格开始计算占位数据, 通过观察图6-30的瀑布可以得知:

第N个占位数据的大小= (总成本 - (成本1+成本2+成本3+⋯⋯+成本N))

按以上公式我们可以得出各个占位数据，如图6-31所示。

STEP 02 选择表中A1:C6数据，单击“插入”选项卡，在图表“柱形图”中选择“堆积柱形图”，如图6-32所示。

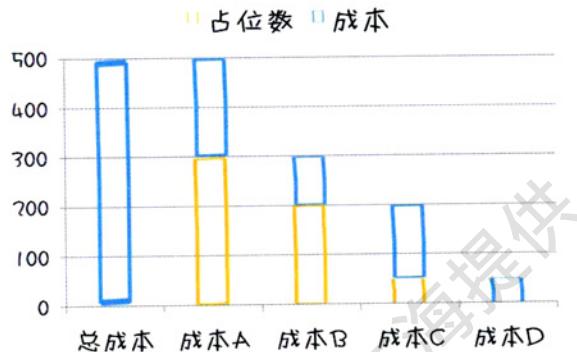


图6-32 瀑布图制作过程示例

STEP 03 选择占位数据所在的任一柱子，单击鼠标右键，选择“设置数据系列格式”功能，在“填充”“边框颜色”两个选项中分别设置“无填充”“无线条”，这样即可把占位数据柱子隐藏起来，单击“关闭”按钮。

STEP 04 删除图例和网格线。

STEP 05 把纵坐标的标签、刻度线，以及线条颜色都设置为无，即把它们也隐藏起来。

STEP 06 用鼠标选择剩下的各成本中的任一柱子，单击右键，选择“添加数据标签”。这样瀑布图就作好了，效果如图6-30所示，这个比蛇形图的制作相对简单点吧。

小白：Mr.林，好棒啊！

6.3.5 帕累托图

Mr.林：小白，考考你，你知道二八法则吗？

小白：二八法则？没有听过，请Mr.林赐教。

Mr.林：二八法则，是19世纪末20世纪初意大利经济学家巴莱多发现的。他认为，任何一组事物中，最重要的只占其中约20%，其余的80%虽然是多数，但却是次要的。比如，80%的财富掌握在20%的人手中，而剩下80%的人，只拥有那20%的财富。再比如，人生中20%的时间，决定了80%的成就，而另外80%的时间，浪费在了20%的事情上。

这启示我们在工作中要善于抓主要矛盾，善于从纷繁复杂的工作中理出头绪，把资源用在最重要、最紧迫的事情上。

接下来要学习的帕累托图就是二八法则的最佳实践工具。

帕累托图又叫排列图、主次图，是按照发生频率的高低顺序绘制的直方图（无间距的柱形图），如图6-33所示，表示有多少结果是由已确认的原因所造成。它是将出现的质量问题和质量改进项目按照重要程度依次排列而得到的一种图表，可以用来分析质量问题，用以分析寻找影响质量问题的主要因素。

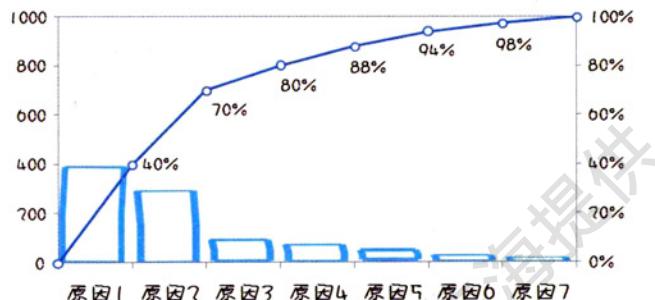


图6-33 帕累托图示例

由于质量问题的影响因素也服从二八法则，即影响质量问题的因素虽然很多，但是只有个别因素起决定性作用，而绝大多数因素的影响都是可以忽略的，所以，帕累托图是质量管理工作常用的一种统计工具，是找出影响产品质量主要因素的一种有效方法。

现在我们就来看看在Excel中如何制作帕累托图。首先要了解帕累托图的特点，如图6-34所示。

- ★ 它是一个特殊的线柱图，柱形图的数据按数值的降序排列，折线图上的数据有累积百分比数据，并在次坐标轴显示。
- ★ 折线图的起点数值为0%，并且位于柱形图第一个柱子的最左下角。
- ★ 折线图第二个点位于柱形图第一个柱子的最右上角。
- ★ 折线图最后一个点数值为100%，位于整张图形的最右上角。

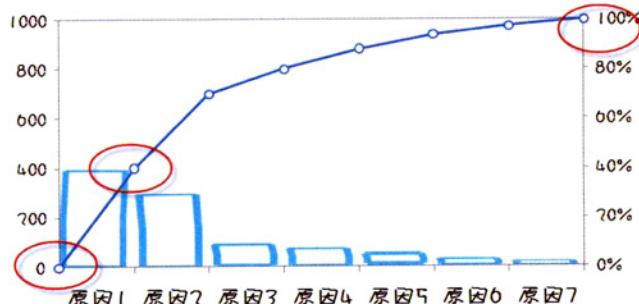


图6-34 帕累托图特点示例

>> 谁说菜鸟不会数据分析（入门篇）

STEP 01 准备数据, 如图6-35所示, 计算好累计件数、累计百分比, 并在C列第2行增加折线图起始点数据“0%”。

A	B	C	D	E	F	G
原因	件数	累计百分比	累计件数			
		0%	0			
原因1	400	40%	400			
原因2	300	70%	700			
原因3	100	80%	800			
原因4	80	88%	880			
原因5	60	94%	940			
原因6	35	98%	975			
原因7	25	100%	1000			

图6-35 帕累托图示例数据

STEP 02 选择表中A3:C9数据, 单击“插入”选项卡, 在图表“柱形图”中选择“簇状柱形图”, 如图6-36所示。

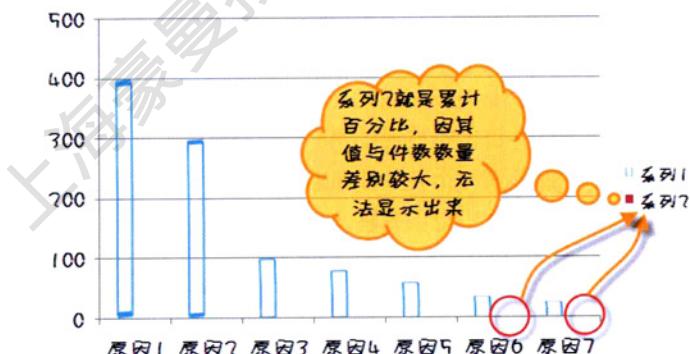


图6-36 帕累托图制作过程 (示例1)

STEP 03 根据前面介绍的双柱图制作方式, 把系列2调至次坐标轴, 并把图形更改为“带数据标记的折线图”。图6-37所示的图与最终的帕累托图相比还有4处不同。

STEP 04 单击折线图上任一点, 再单击鼠标右键, 选择“选择数据”项, 更改系列2的系列值范围为C2:C9, 如图6-38所示。这时你会发现折线图的起点已与横坐标轴相交, 需要继续进行相关参数的设置。

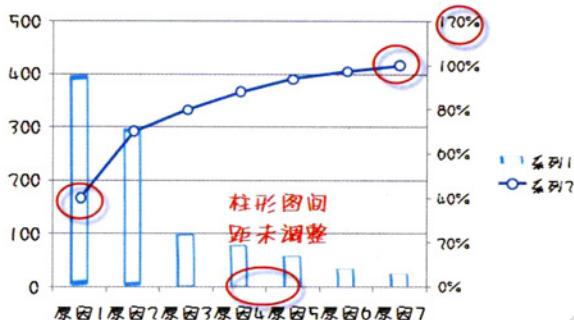


图6-37 帕累托图制作过程(示例2)



图6-38 ‘编辑数据系列’对话框

STEP 05 选中整个图表，单击“布局”选项卡→“坐标轴”→“次要横坐标轴”→“显示从左向右坐标轴”，操作步骤如图6-39所示。



图6-39 设置次要横坐标轴的显示方式

STEP 06 选中刚显示的次要横坐标轴(在图表的上方)，单击鼠标右键，选择“设置坐标轴格式”，在“设置坐标轴格式”对话框中，将“位置坐标轴”选择为“在刻度线上”，并且把“主要刻度线类型”“坐标轴标签”都设置为“无”，如图6-40所示。这时折线图都已就位，帕累托图即将成型喽！

>> 谁说菜鸟不会数据分析（入门篇）

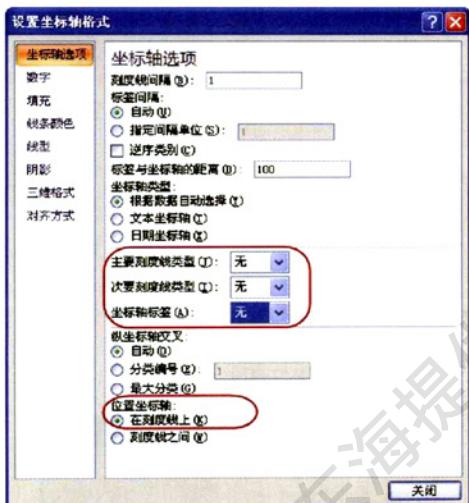


图6-40 设置次要横坐标轴格式

STEP 07 设置次纵坐标轴最大值为100%；设置柱形图各分类间隔为无间距（0%），如图6-41所示。



图6-41 设置柱形图分类间隔

STEP 08 同理，设置主要纵坐标轴最大值为累计值1000，以确保折线图第二个点位于柱形图第一个柱子的右上角。

STEP 09 用鼠标选择折线图中的任一点，单击鼠标右键，选择“添加数据标签”。

STEP 10 删除图例和网格线。

Mr.林：帕累托图制作完毕，如图6-33所示。小白，操作步骤都看明白了吗？

小白边点头边说：我在笔记本上都记下来呢，谢谢Mr.林。

6.3.6 旋风图

Mr.林：小白，记得之前给你举的生姜价格上涨与销量变化的例子吗？

小白：当然记得啦，您教的知识怎么敢忘了！我记得当时是用散点图来说明生姜价格与销量变化的。Mr.林，难道你又要教我另一种方法？

Mr.林：小白，够机灵啊！下面我要介绍一个图表给你，它有一个非常酷的名字，叫做“旋风图”，因为它看起来就像舞动着的旋风一样，如图6-42所示。它的学名叫做“成对条形图”或“对称条形图”。

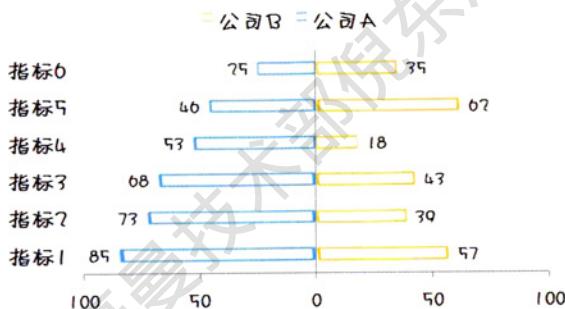


图6-42 旋风图示例

旋风图主要用在以下情形：

- ★ 同一事物在某个活动、行为影响前后不同指标的变化，如某企业促销活动开展前后，产品收入、销量等不同指标的变化。
- ★ 同一事物在某个条件变化下（指标A的变化），指标B受影响也随之变化，具有因果关系，如生姜价格与销量的关系。
- ★ 两个类别之间不同指标的比较，如某班男生和女生各项学科成绩对比，或部门A与部门B各业绩指标对比。

当然用途不限于以上三种，这里介绍的是主要作用，旋风图还可以用在其他行业的研究上，需要靠自己研究挖掘使用。

① 旋风图作法一

观察图6-42的特点，我们可知旋风图是由条形图制作而成。下面就具体介绍如何制作旋风图，我们采用图6-43示例的数据制图。

	A	B	C
1	指标	公司A	公司B
2	指标1	85	57
3	指标2	73	39
4	指标3	68	43
5	指标4	53	18
6	指标5	46	62
7	指标6	25	35

图6-43 制作旋风图的示例数据

STEP 01 选择A1:C7范围的数据，单击“插入”选项卡，在图表“条形图”中选择“簇状条形图”，效果如图6-44所示。

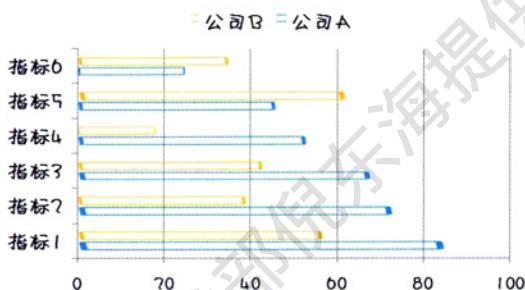


图6-44 旋风图制作过程1

STEP 02 选中蓝色条形图中任一条形，单击鼠标右键，选择“设置数据系列格式”功能，选择将系列绘制在“次坐标轴”。注意观察图表变化，这时出现两个横坐标轴，上方的横坐标轴为蓝色条形图的，下方横坐标轴为黄色条形图的，效果如图6-45所示。

STEP 03 在“设置坐标轴格式”对话框中设置两个横坐标轴的最大值都为100、最小值都为-100，单击“关闭”按钮。

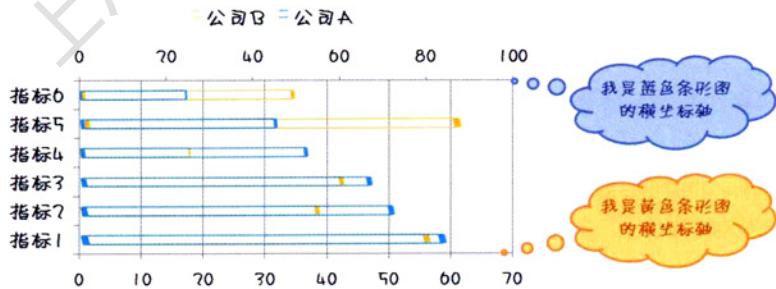


图6-45 旋风图制作过程2

STEP 04 在“设置坐标轴格式”对话框的“数字”页面，将“类别”选择为“自定义”，在“格式代码”中填写“0;0;0”，并添加至“类型”框处，选中刚添加的“0;0;0”类型，单击“关闭”按钮。操作步骤如图6-46所示，调整后的图形如图6-47所示。

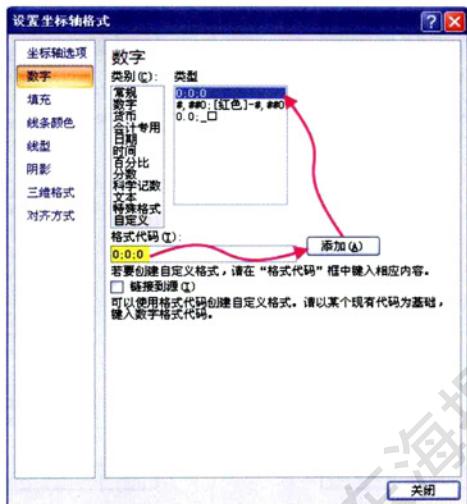


图6-46 旋风图制作过程3

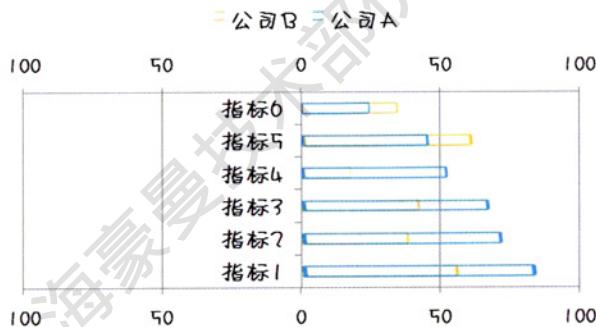


图6-47 旋风图制作过程4

此步骤的操作目的是为了让负轴不显示负数，直接显示正数。如果数值比较小，可设置为带一位小数“0.0;0.0;0.0”或带百分比“0%;0%;0%”，以此类推，设置为所需的格式。

STEP 05 选择上方的横坐标轴（蓝色条形的横坐标轴），单击鼠标右键，选择“设置坐标轴格式”，在“坐标轴选项”中，勾选“逆序刻度值”，并且把“主要刻度线类型”“坐标轴标签”都设置为“无”，操作步骤如图6-48所示。另外在“线条颜色”选项上选择“无线条”，此操作目的是要把蓝色条形图翻转至左边，并且把上方横坐标轴隐藏起来。

STEP 06 选择纵坐标轴，单击鼠标右键，选择“设置坐标轴格式”，在“坐标轴选项”中，把“主要刻度线类型”设置为“无”，“坐标轴标签”设置为“低”。此步骤目的是为了把坐标轴标签（图中的指标名称）移至图的最左边，操作步骤如图6-49所示。



图6-48 旋风图制作过程5



图6-49 旋风图制作过程6

STEP 07 分别选中条形图，单击鼠标右键，选择“添加数据标签”，将两个条形图的数据标记显示出来。

STEP 08 删除网格线。

至此旋风图制作完毕，如图6-42所示，不过为了效果更好，还可以对图表的一些格式进行调整，比如把条形间距调小，以拉宽条形。

◎ 旋风图作法2

小白：好麻烦啊！Mr. 林有没有简单一点的方法。

Mr.林：那好吧，我现在教一个简便的方法。

STEP 01 分别根据图6-43的示例数据作两个条形图，如图6-50所示。

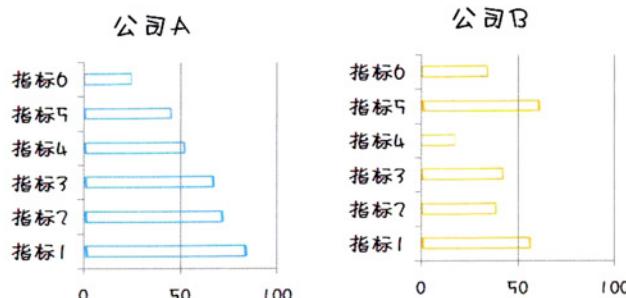


图6-50 旋风图制作过程7

STEP 02 选择公司A条形图的横坐标轴，单击鼠标右键，选择“设置坐标轴格式”，在“坐标轴选项”中勾选“逆序刻度值”。

STEP 03 选择公司A条形图的纵坐标轴，单击鼠标右键，选择“设置坐标轴格式”，在“坐标轴选项”中，将“主要刻度线类型”设置为“无”。

STEP 04 选择公司A条形图的纵坐标轴，把坐标轴标签（图中的指标名称）字色设置为白色。

小白：Mr.林，这里为何要把字色设置为白色，而不直接设置为“无”呢？

Mr.林：这个问题问得好，把纵坐标轴标签字色设置为白色，目的就是为了在不改变原有图形大小的情况下，隐藏不相关的元素或信息，使图的大小比例与公司B的条形图一致，其作用与占位数据的作用是一致的，前提是两个图形都没有调整更改默认图形的大小。小白，你可以把“坐标轴标签”设置为“无”做一下试验，看看图的大小是否发生变化。

STEP 05 分别删除两个条形图的网格线，添加数据标签。

STEP 06 把两个条形图排列在一起，旋风图就制作完毕了，效果如图6-51所示。

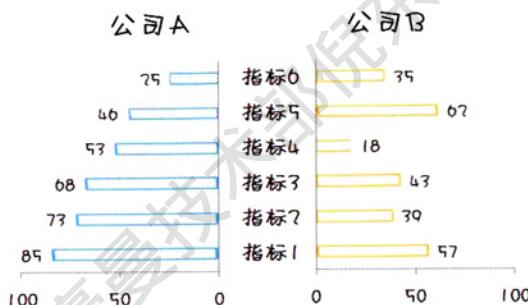


图6-51 方法二制作出的旋风图

Mr.林：小白，你看这样制作是否简便一些？

小白：是啊，这个方法简便一些。不过我觉得各种方法都有各自的好处，就看自己实际需求是什么再选择什么样的作法。

Mr.林：看来你已经把数据分析的核心精神都领会了！

小白红着小脸：谢谢Mr.林夸奖，我都不好意思了。

6.3.7 人口金字塔图

Mr.林：小白，接下来要学习的图形跟刚才介绍的旋风图很类似，可以说它就是一种特殊的旋风图，因为它专门用于反映人口的过去、现在、未来的发展情况，这种图就叫做人口金字塔图。

人口金字塔是按人口年龄和性别表示人口分布的特种塔状条形图，如图6-52所示，它形象地表示某地区人口的年龄和性别构成分布状况。水平条代表每一年龄组男性和女性的数字或比

>> 谁说菜鸟不会数据分析（入门篇）

例，金字塔中各个年龄性别组相加构成了总人口。

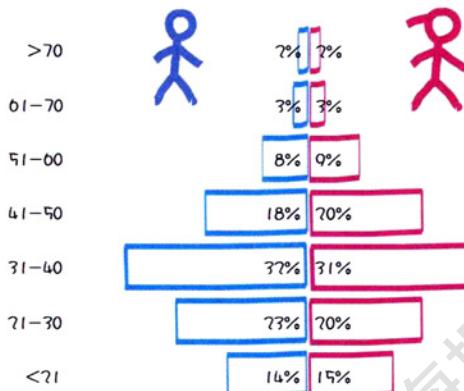


图6-52 人口金字塔图

人口金字塔图以年龄为纵轴，以人口数为横轴，按左侧为男、右侧为女绘制图形，其形状如金字塔。金字塔底部代表低年龄组人口，金字塔上部代表高年龄组人口。人口金字塔图反映了过去人口的情况，目前人口的结构，以及今后人口可能出现的趋势。

人口金字塔可分为三种类型：年轻型、成年型和年老型，它们的形状各不相同，如图6-53所示。

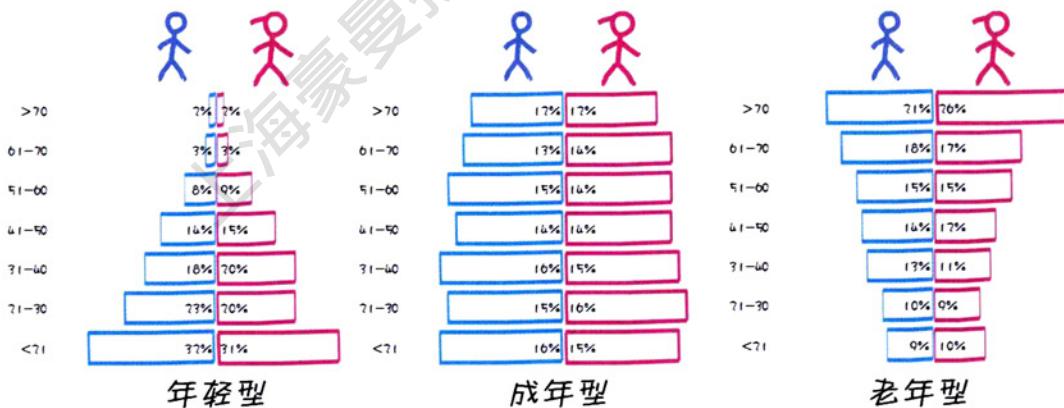


图6-53 各类型人口金字塔图

- ★ 年轻型：塔顶尖、塔底宽。
- ★ 成年型：塔顶、塔底宽度基本一致，在塔尖处才逐渐收缩。
- ★ 年老型：塔顶宽，塔底窄。

缩减型，分别与年轻型、成年型和年老型相对应。

例如要分析某地区是否具有市场潜力，就可以考虑分析当地的人口结构情况，看看当地人口是以什么样的年龄段构成的，这时人口金字塔就派得上大用场。

因人口金字塔的作法基本与旋风图作法一致，我就不进行具体的介绍了，小白，如果你有兴趣可以自己尝试下。

小白：Mr.林，好的，如果有不明白的地方再来请教您。

6.3.8 漏斗图

Mr.林：小白，在数据分析方法中我们已经介绍了什么是漏斗图，如图6-54所示，你说说看它究竟是怎么制作的，是由什么图变化而来的呢？

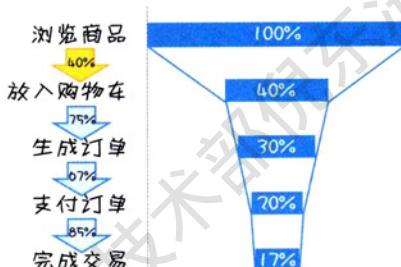


图6-54 网站转化率（漏斗图）

小白仔细观察后，兴奋地说：Mr.林，漏斗图是用堆积条形图作的，并且需要用占位数据把实际的条形图“挤”到中间去。

Mr.林：小白，非常不错，之前教给你的都记住了，并且还能融会贯通。下面就来看看漏斗图具体的作法，我就用图6-55中的数据进行介绍吧。

图6-55展示了制作漏斗图所需的数据示例。这是一个名为“漏斗图.xlsx”的工作簿中的C3单元格区域。该表格包含以下列标题：A（环节）、B（人数）、C（占位数据）、D（每环节转化率）和E（总体转化率）。表格中的数据如下：

A	B	C	D	E
环节	人数	占位数据	每环节转化率	总体转化率
1 浏览商品	1000	0	100%	100%
2 放入购物车	400	300	40%	40%
3 生成订单	300	350	75%	30%
4 支付订单	200	400	67%	20%
5 完成交易	170	415	85%	17%

图6-55 漏斗图示例数据

$$\text{第N环节占位数据} = (\text{第1环节进入人数} - \text{第N环节进入人数}) / 2$$

$$\text{第N环节转化率} = \text{第N环节进入人数} / \text{第(N-1)环节进入人数}$$

$$\text{第N环节总体转化率} = \text{第N环节进入人数} / \text{第1环节进入人数}$$

>> 谁说菜鸟不会数据分析（入门篇）

STEP 01 选择A1:C6范围的数据，单击“插入”选项卡，在图表“条形图”中选择“堆积条形图”，效果如图6-56所示。

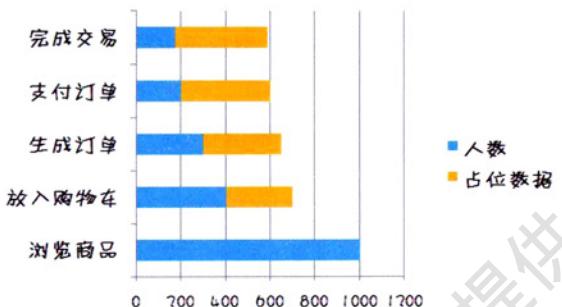


图6-56 漏斗图绘制过程示例1

由图6-56所示，这样的图形离漏斗图的样子还差很远，观察图形我们可以得知：①纵轴的顺序反了，这个好办，可以用“逆序类别”功能把它拨正。②人数的条形图应该在中间才对，这也好办，把人数跟占位数据的顺序调下即可，还要把占位数据设置隐藏起来。

- STEP 02** 选中纵坐标轴，单击鼠标右键，选择“设置坐标轴格式”，在“坐标轴选项”中勾选“逆序类别”。
- STEP 03** 在图表上单击右键，选择“选择数据”，在弹出的对话框中选择“人数”这个数据系列，单击“下移”按钮，如图6-57所示，然后单击“确定”按钮。

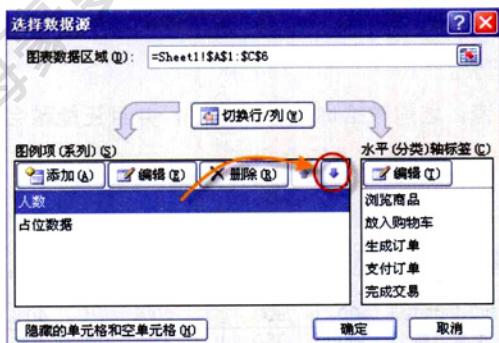


图6-57 漏斗图绘制过程示例2

- STEP 04** 选中占位数据所在的黄色条形，单击鼠标右键，选择“设置数据系列格式”，把填充及边框颜色都设置为“无”，即把它们也隐藏起来。
- STEP 05** 删除图例和网格线，此时漏斗图已成型，如图6-58所示。

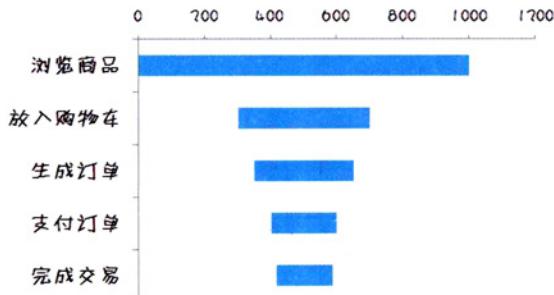


图6-58 漏斗图绘制过程示例3

通过与图6-54比较，可以发现图6-58与之还存在细小差别，主要差别在于前者有转化率数据标签，而且前一个漏斗还有外框。

总体转化率的数据标签可利用JWalk Chart Tools工具进行添加，而每个环节的转化率及其箭头则需要手工进行添加。

漏斗外框可通过下面的步骤添加：先用鼠标选中整个图表，单击“布局”主选项卡→“分析”→“折线”→“系列线”，并且设置系列线颜色与条形的颜色一致，最终效果如图6-54所示。

图6-54所示的漏斗图第一环节转化率为40%，为各环节最低，因此用黄色箭头标明，以便让读者更快注意到。颜色的使用方法我会在下次为你做具体介绍。

小白：好的，按你的计划来啊。

6.3.9 矩阵图（散点图）

Mr.林：小白，接下来要教你矩阵图的制作，还记得在数据分析方法中提到的矩阵关联分析方法吗？

小白：当然记得啦，您才教我的，如果这么快忘了，岂不是对不起您的劳动。

Mr.林：还算你有良心呦！我们现在来看看在Excel中如何绘制图6-59所示的矩阵。小白，你觉得应该用什么图来绘制呢？

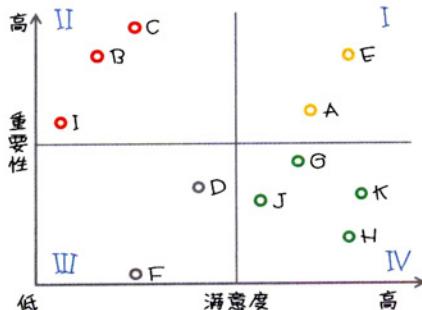


图6-59 2010年某公司用户满意度优先改进矩阵图

小白：我看只有散点图跟这个比较像，不过就不知道中间的十字线条是怎么画上去的，是手工添加上去的吗？

Mr.林：没错，我们以图6-60所示的数据介绍绘制操作。

	A	B	C
1	指标	满意度	重要性
2	A	3.2	3.1
3	B	1.5	3.6
4	C	1.8	3.9
5	D	2.3	2.4
6	E	3.5	3.6
7	F	1.8	1.6
8	G	3.1	2.6
9	H	3.5	1.9
10	I	1.2	3.0
11	J	2.8	2.3
12	K	3.6	2.3
13	平均值	2.6	2.8

图6-60 2010年某公司用户满意度数据示例

STEP 01 选择B2:C12范围的数据，单击“插入”选项卡，在图表“散点图”中选择“仅带数据标记的散点图”。作散点图时只需要选择横坐标与纵坐标对应的值即可，无须把指标名称和字段名称也选入作图数据范围。

STEP 02 删除图例和网格线，效果如图6-61所示。

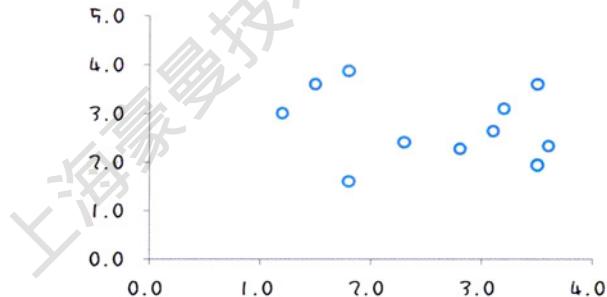


图6-61 矩阵图绘制过程示例1

Mr.林：这是常见的散点图，如何把它变成矩阵形式呢？你仔细观察图可发现，横、纵坐标是多余的，但我们能否废物利用呢？能不能移动坐标轴，比如把横坐标轴往上移，纵坐标轴往右移？

小白动手操作了一下：哎！它还真的可以移动。

Mr.林：你再按照下面的步骤操作一下。

STEP 03 选中横坐标轴，单击鼠标右键，选择“设置坐标轴格式”，在“坐标轴选项”最下方“纵坐标轴交叉”的“坐标轴值”框中填入已计算好的横坐标轴各指标满意度的平均值2.6，另外顺便把“主要刻度线类型”“坐标轴标签”都设置为“无”，操作设置如图6-62所示。



图6-62 矩阵绘制过程示例2

STEP 04 同理，选中纵坐标轴，单击鼠标右键，选择“设置坐标轴格式”，在“坐标轴选项”最下方“横坐标轴交叉”的“坐标轴值”框中填入已计算好的纵坐标轴各指标重要性的平均值2.8，另外顺便把“主要刻度线类型”“坐标轴标签”都设置为“无”，操作设置参考图6-62，矩阵图效果如图6-63所示。

Mr.林：这时矩阵图已基本成形，但与图6-59所示的还有差别，如标签、矩阵的边框箭头等要素未添加。另外，矩阵上的坐标空白较多，不仅使得各点较为集中，而且极其浪费页面空间。我们可通过设置坐标轴的最大值、最小值来让矩阵图完美适应页面空间，该项设置与PowerPoint中“使幻灯片适应当前窗口”功能类似。

STEP 05 利用JWalk Chart Tools工具添加各指标的标签。

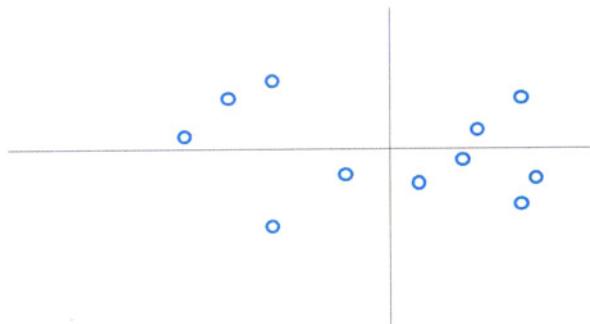


图6-63 矩阵图绘制效果

STEP 06 分别对横、纵坐标轴的最大值、最小值进行设置，可设置为比各指标满意度、重要性的最大值略大、最小值略小的值。在该例中，横坐标轴最小值、最大值可分别设置为1和4，纵坐标轴最小值、最大值可分别设置为1.5和4，以便让矩阵图完美适应页面空间。

STEP 07 单击“插入”选项卡，选择“形状”→“箭头”，绘制与矩阵相当宽度的横向箭头，然后复制绘制好的箭头并粘贴，将其调整为竖向，高度与矩阵高度相当，把两箭头的尾部相连，构成坐标系。

STEP 08 添加横向、竖向箭头的标签（满意度、重要性）及说明（高、低）。

STEP 09 对每个象限的点进行颜色的设置，比如优先改进区可以用红色来突显，维持优势区用绿色表示，高度关注区用黄色来显示，无关紧要区用灰色来显示，以提高矩阵的可读性。

Mr.林：矩阵图绘制完毕，效果如图6-59所示。你看，矩阵图的绘制也没有想象中那么难，只要了解绘制思路，就可以用Excel帮我们实现。常言道：“只有想不到，没有做不到，关键就在想法与创意”。Excel还是有很多潜力等待我们去发掘的。

小白在职场日记中记下了这样一句话：只有想不到，没有做不到。

6.3.10 发展矩阵图

Mr.林：小白，你还记得发展矩阵吗？如图6-64所示。

小白：当然，我还记得当时您问我如何表现某公司客户满意度在2009—2011年三年来的变化情况呢。

Mr.林：记性不错啊！我们现在就要学习发展矩阵的制作方法。

小白：我们已经学了如何制作矩阵图，发展矩阵与普通矩阵之间有何相似之处呢？

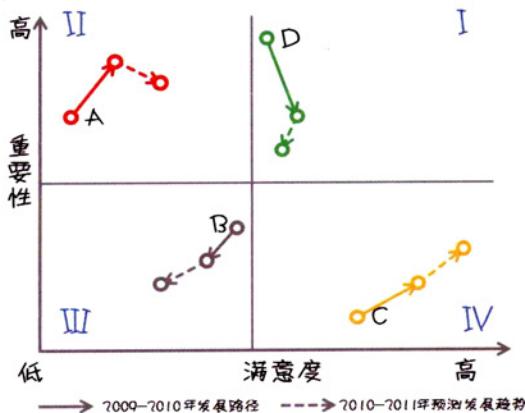


图6-64 发展矩阵示例

Mr.林: 从图的特点来对比发展矩阵与普通矩阵：发展矩阵多了箭头表示趋势。发展矩阵前面的作法都与普通矩阵相同，所以我往下接着介绍如何将普通矩阵转变为发展矩阵，即学习如何添加箭头，下面以图6-65中的示例数据作为数据源介绍绘制操作。

	A	B	C	D
1	指标	年份	满意度	重要性
2	A	2009	1.2	3.1
3		2010	1.5	3.6
4		2011	1.8	3.4
5	B	2009	2.3	2.1
6		2010	2.1	1.8
7		2011	1.8	1.6
8	C	2009	3.1	1.3
9		2010	3.5	1.6
10		2011	3.8	1.9
11	D	2009	2.5	3.8
12		2010	2.7	3.1
13		2011	2.6	2.8
14	平均值		2.4	2.5

图6-65 2009—2011年某公司用户满意度发展及预测数据示例

- STEP 01** 先以C2:D13范围的数据，绘制散点图，并按绘制普通矩阵的方式绘制好矩阵。
- STEP 02** 添加线条。选中任一数据点，单击鼠标右键，选择“设置数据系列格式”功能，在弹出的对话框中设置“线条颜色”为“实线”，并可先设置颜色为与某个区域数据点相同的颜色，如第一象限的绿色。
- STEP 03** 去掉A与B间的线条。选中B指标的第一个数据点（2009年这个点），单击鼠标右键，选择“设置数据系列格式”功能，弹出对话框。在“线条颜色”中选择“无线条”。
- STEP 04** 同理去掉B与C、C与D间的线条。
- STEP 05** 添加A指标内发展趋势箭头。第一个点与第二点之间的箭头应在第二个点上进行设置，例如想要设置2009年至2010年的箭头，那么应在2010年点上进行设置。选择要设置的数据点，单击鼠标右键，选择“设置数据系列格式”菜单，弹出对话框，设置“箭头设置”为“后端类型”，并选择一种箭头，这里选择“开放型箭头”，如图6-66所示。在设置箭头类型后，顺便也把A指标发展箭头的线条颜色设置为与所在象限（第二象限）同样的红色。
- STEP 06** 同理，添加其他各指标线条箭头，并设置箭头的线条颜色。
- STEP 07** 添加箭头图例注释。因有虚线代表预测，所以需要添加图例进行注释，可通过手工插入箭头及文本框的方式实现。

Mr.林: 小白，经过上述的操作步骤设置后，发展矩阵就基本绘制完毕，效果如图6-64所示。你可能认为发展矩阵的难点在于箭头的方向设置，我建议多动手尝试设置，就会发现其中的规律。



图6-66 发展矩阵绘制过程示例

解决问题不能光说不练，做数据分析、绘制图表也是如此，需要自己通过实践操作，才能够真正掌握。

小白边点头边在职场日记中记下了这样一句话：不能光说不练，只有亲自动手去实践才能解决问题。

6.3.11 改进难易矩阵（气泡图）

Mr.林：小白，刚才介绍的矩阵图作法，你都掌握了吗？

小白：基本了解了，不过还需要回去好好消化消化。

Mr.林：对，回去要多复习和动手实践操作。你还记得改进难易矩阵吗？如图6-67所示，现在我们学习如何制作它，它是在气泡图的基础上进行调整设置的，就如矩阵图在散点图上一样调整设置。

气泡图是一种特殊类型的散点图，它是散点图的扩展，相当于在散点图的基础上增加了第三个变量，即气泡的面积，其指标对应的数值越大，则气泡越大；相反数值越小，则气泡越小，所以气泡图可以应用于分析更加复杂的数据关系。

在我们这个例子中，气泡的大小代表着改进的难易程度，气泡越大，代表着指标的改进越难；相反，气泡越小，代表着指标的改进越容易，如图6-67所示。

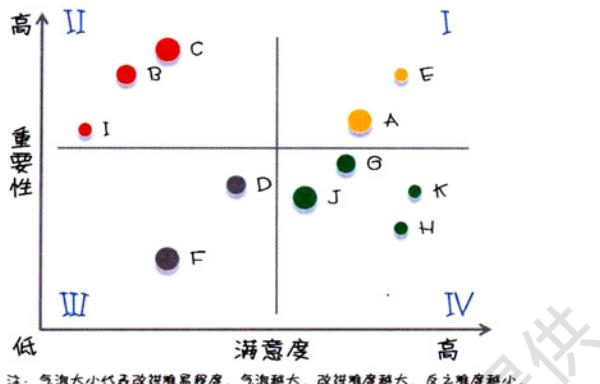


图6-67 改进难易矩阵示例

现以图6-68所示数据为例进行绘制操作的介绍。

	A	B	C	D
1	指标	满意度	重要性	改进难易程度
2	A	3.2	3.1	3.0
3	B	1.5	3.6	2.0
4	C	1.8	3.9	3.0
5	D	2.3	2.4	2.0
6	E	3.5	3.6	1.0
7	F	1.8	1.6	3.0
8	G	3.1	2.6	2.0
9	H	3.5	1.9	1.0
10	I	1.2	3.0	1.0
11	J	2.8	2.3	3.0
12	K	3.6	2.3	1.0
13	平均值	2.6	2.8	2.1

图6-68 改进难易矩阵数据示例

STEP 01 选中B2:D12范围的数据，单击“插入”选项卡，在图表“其他图表”中选择“气泡图”，效果如图6-69所示。

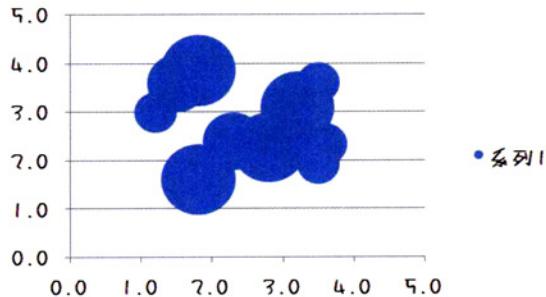


图6-69 改进难易矩阵绘制过程示例1

此时各个气泡比较大，并且重叠在一起，不便读图，我们可继续进行下述操作。

>> 谁说菜鸟不会数据分析（入门篇）

STEP 02

缩小气泡。选中任一气泡点，单击鼠标右键，选择“设置数据系列格式”菜单，弹出对话框，在“系列选项”中“大小表示”区域，选择“气泡面积”，将气泡大小缩放设置为“30”，这里的大小可根据个人习惯及实际情况进行设置，如图6-70所示。

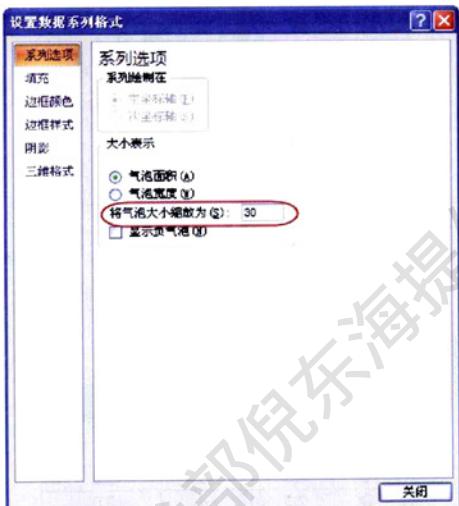


图6-70 改进难易矩阵绘制过程示例2

此时各个气泡已不再重叠，剩下的操作步骤与绘制图6-59的步骤基本一致，可以自己尝试着来绘制。

6.4 本章小结

Mr.林：小白，关于数据展现的内容今天就先介绍这些。还是老规矩，简要回顾一下。

今天介绍的数据展现方面的知识主要有：

- ★ 图表在数据分析中的作用，饼图、柱形图、折线图、条形图等经济适用图表的作用与特点，通过所要表达的关系选择恰当的图表，以及制作图表的五步法。
- ★ 表格中突出显示单元格、项目选取、数据条、图标集、迷你图等几种展现数据的表格技巧。
- ★ 平均线图、双坐标图、竖形折线图、瀑布图、帕累托图、旋风图、人口金字塔图、漏斗图、矩阵图等图表分析技巧。

小白：嗯，都是数据展现实用的技巧，听君一席话，胜读十年书呀。

Mr.林：哈哈，对你有帮助就好，我辛苦点无所谓，记得回去好好复习并在工作中灵活运用。

小白：这个是肯定的，牛董的报告还等着我完成，先回去了。

第 7 章

专业化生存——图表可以更美的

别让图表犯错

浓妆淡抹总相宜——图表美化

如虎添翼的招儿



经过Mr.林几天的“秘密培训”，小白明显感觉自己工作效率和工作质量提高了许多，为了感谢Mr.林，同时本着精益求精的态度，小白特意找了个周末的时间约Mr.林出来喝早茶。正宗的粤式早茶，环境优雅，很适合边吃边聊。

小白端起茶壶，给Mr.林倒上一杯普洱，说道：多亏您的指导！同事们都夸我最近进步很多。我说多亏我遇上个好师傅，呵呵。不过我拿自己做的图表和您做的一对比，觉得简直相差太大了，爱美之心人皆有之，有没有办法教我怎样美化一下图表呢？

Mr.林笑道：你约我喝茶，我就知道这茶不是白喝的，所以我就把笔记本电脑也一起带来了，现在就为你介绍如何美化图表。

Mr.林再添了几口爽滑清鲜的虾饺，然后一本正经开始问道：你专业吗？

Mr.林的思维还真是跳转得快，小白愕然。

Mr.林继续道：专业这个词已经逐渐成了评价一个人工作的流行标准。员工在问，老板在问，客户在问。连电影《疯狂的石头》也告诉我们，在21世纪，即便是想当个小毛贼，也要够专业！所以，专业二字，价值千金。

小白忙点头。

Mr.林仿佛对此感触很深，又补充道：就像是这家高档正宗的粤式酒家，这种专业的态度是不是让你印象深刻，下次还想再来？同样，图表要做得专业才具备说服力。专业的图表也同时传递着我们专注、敬业的形象，使我们更容易取得客户的信任和老板的赏识。

小白：专业化图表如此重要，那么它的评价标准是什么呢？

Mr.林：简单地说可以概括为3个词：严谨、简约和美观。

首先，图表是为了证明一个观点及事实而存在的。小白，你做过证明题吧，证明过程是一环套一环的，无论结论还是过程中的每一个论据和逻辑都要非常严谨，经得住推敲。专业也就意味着严谨，意味着不允许一点细微的错误，追求细节的完美。

第二是简约，即“图简意赅”，正中要点。我们不是画家，不需要像毕加索、梵高那样研究高深的绘画艺术。正如我一直强调的，图表只是为了说明观点，我们要做的是透过现象看本质，一针见血。

第三是美观。一幅精致美观、令人赏心悦目的图表才让人有看它的欲望，让人印象深刻。试问，粗糙拙劣的图表又如何让人相信其背后数据的准确性呢？

小白点点头表示赞同。

Mr.林：专业精神，博大精深。这么简单地解释其含义，你可能印象不深。下面我将通过案例详细解说如何绘制正确、严谨的图表，如何美化图表，以及一些让你事半功倍的技巧。

7.1 别让图表犯错

Mr.林：小白，你应该看过像麦肯锡、罗兰贝格等这种顶级咨询公司所做的专业图表，现在你能告诉我它们为什么专业吗？

小白紧张地摇摇头：整体感觉很专业，可是具体哪里专业我也说不上来。估计要是让我来依葫芦画瓢地学他们作图，也怕落得个“邯郸学步”或“东施效颦”的下场。

Mr.林：所以，我先给你看一些不完美不专业的图表，你来找找这些图表哪里不专业，这样就能更清楚一幅专业的图表是如何生成的。

小白表示赞同：错误和失败更能让人长记性！呵呵。

Mr.林：刚刚我们讲到专业的图表首先是正确的、严谨的。错误的图表让观众不明所以，甚至与自己要表达的信息南辕北辙。如果说正确的数据居然能产生错误的图表，那么究竟会在什么地方出现错误呢？

下面，跟我一起来玩玩“大家来找茬”的游戏，即挑出图表的毛病。小白别走神，跟着我来喔。

小白：好嘞！看我的火眼金睛！

7.1.1 让图表“五脏俱全”

Mr.林：如图7-1所示，这张图作者想要表达什么信息？

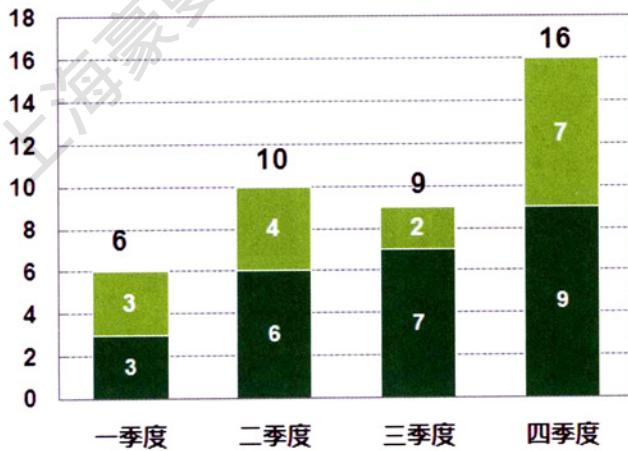


图7-1 信息不完整的图表

小白皱了皱眉：这显然是一幅残缺不全的图嘛，一没标题、二没图例，我怎么知道它要表达什么信息？

Mr.林一拍桌子：对，一张图必须包含完整的元素，才能让观众一目了然。可是新手们往往一不小心就犯了这样的错误。怎样才算包含完整的元素呢？我们通过完善图7-1来找答案。

首先，我们加上图表标题“2010年某产品的销量情况”。

小白忙插一句：嗯，现在我一看就知道了，这张图表是介绍产品销量的！

Mr.林：对！然后加上图例，帮助读者知道图中深绿色和浅绿色柱子所代表的含义：浅绿色代表品牌A的某产品销量情况，深绿色代表品牌B的某产品销量情况；再加上纵坐标数值的单位。

小白：原来纵坐标单位是百万，我还猜是万呢，这可差远了！

Mr.林：小白，你再观察一下图形的数据值，发现有什么异常没有？

小白一边仔细观察，一边思考：第三季度怎么一下子降了那么多呢，而且深绿色的部分还是处于正常上升趋势的，看来问题出在浅绿色的柱子那里了。

Mr.林：对，第三季度A产品的包装出现过问题，这是产生这样的数据背后重要的原因，是必须要加在图表上的。在图表里，我们称其为脚注。

最后，还有一个元素——资料来源，这个不是一定得有，但是专业图表最好添加这个内容，这样能增加数据的可信度。修改后的图表如图7-2所示，现在是不是觉得很完整了呢？是不是马上了解到图表所要表达的主题呢？

小白：现在的图看起来专业多了。



图7-2 2010年某产品的销量情况

Mr.林：标题、图例、单位、脚注、资料来源这些图表元素就好比是图表的五脏六腑，有了它们图表才有生命与意义。所以，做专业的图表，首先得让图表“五脏俱全”！

7.1.2 要注意的条条框框

Mr.林：刚刚讲完怎样做一幅完整的图表，到现在为止，你已经学会了如何选择合适的图表，如何制作图表，需要添加哪些图表元素，是不是早就心里痒痒了，想自己设计了？

小白急切的心情溢于言表：是啊是啊，我恨不得马上将这些天学到的十八般武艺统统用上！

Mr.林：我非常理解你这种心情，但是我相信你现在做出来的图必定漏洞百出。因为作图表还有很多的条条框框你不清楚，指不定就犯着哪一条了。

记住，不只是作图，在职场、在生活中也是如此。做事情都有规矩，懂规矩才不会犯错，将规矩看透了才能有所突破和创新。就好比武侠小说里的侠客练剑得有个心法，否则呀，很容易走火入魔喔！

小白：嗯，您说得很有道理，我记住了。

Mr.林：首先，避免生出无意义的图表。

有些搞数据的哥们想得最多的不是女朋友，而是作图！除了作图还是作图！所以难免会有这样的毛病——看见表格就不顺眼，非要整出个图出来。我想说的是，让图表出世的不是数据而是你！不要像母鸡下蛋一样随地生产，要生就要生得有意义，如果生出来的图看不出任何有价值的信息和结论，还不如省省力气回家歇着。

就像之前我们讲过，有时候表格比图能更有效地传递信息，这时候就没必要画蛇添足绘制图表了。图表贵精不贵多，决定作不作图表的唯一标准是——能否帮助你有效表达信息。用一句不恰当的古语形容“物以稀为贵，多者为贱”，图表轰炸多了，读者就麻木了，反而找不到重点。所以，作图表也讲究天时、地利、人和。

第二点，不要把图表撑破。

不要在一张图表里塞太多信息，很多新手容易犯这个错误。

小白提出疑问了：哇塞，我果然还是个菜鸟级别，头两点就全中招了。用一张图表囊括多个图表的信息，一个能抵俩，这样不好吗？

Mr.林：不行，最好一张图表反映一个观点。这样子才能突出重点，让读者迅速捕捉到核心思想。

有人做过这样的试验，用两种方法：第1种解释一张图的N条信息；第2种作N张图，每张图反映一条信息，然后一张张地给观众解释。两种方法的演示速度是一样的，但是第2种方法给观众感觉条理更清晰、印象更深刻。所以，最好一个图表阐明一个主题。

小白恍然大悟：哦，原来还是计划生育，只生一个好！

Mr.林：对，凡事都要有重点，满篇重点的结果就是看不到重点。

第三点，只选对的，不选复杂的。

有人沉迷于设计各种各样花哨的、所谓“高级”的图表。仿佛不如此无以显示自己技艺的精湛，不如此无以让别人顶礼膜拜。他们认为“你不会的我会啊，所以我很有成就感啊”。我要

说，这种敢于挑战困难突破陈规的勇气值得嘉奖，但是违背了专业精神的基本原则——简约。好的图表是能省掉用来解释的1000句话，而不是需要用1000句话来解释。

第四点，一句话标题。

Mr.林：小白，你读一读这几个图表标题：公司销售情况发展趋势；各地区的产量；薪酬与利润之间的关系……

这些是不是你见过的最泛滥的图表标题？你明白它要讲什么吗？看了这些千篇一律的标题，你还会有兴趣继续看里面的内容吗？要是你订的报纸新闻全是这种标题，你是不是恨得牙齿痒痒，想破口大骂？

小白：确实，要是这样的话，就根本没有往下看的欲望了。

Mr.林：其实这些标题就两个缺点：①没有切中图的大意；②没有吸引力。

我们的牛董每天都要看几百份图表，要是每一份图表都要看完图才知道要表达什么，不是他自己疯掉就是先让你疯掉。所以，作图时首先要把想指出的内容要点融入到标题中去。

上面的这些标题修改起来很简单，只有一个原则：就是将短语变为句子。现在上面的三个标题都是短语吧？看看我怎样把它变成句子：①公司销售额翻了一番；②C区产量居第四；③薪酬与利润之间没有关系。小白，感觉一下前后有什么区别？

小白：感觉明了很多，不需要看图，我也能很快领悟到这张图的要点。

Mr.林：另外，标题有吸引力会给图表加分，我们可以多向标题党学习。

Mr.林停顿了一会儿，接着说：刚才介绍的几点是对所有图表而言的。对于Excel里的图表原型——饼图、柱形图、条形图、折线图还有针对性的注意事项，下面我一个个给你总结出来。

① 饼图

Mr.林：小白，好记性不如烂笔头。饼图有7个注意事项，你要记好咯！

★ 对于饼图要按照时钟表盘的刻度，把数据从12点钟的位置开始排列，最重要的成分紧靠12点钟的位置。

小白不解：为什么要从12点钟位置开始排列呢？

Mr.林：因为人的眼睛都习惯从左至右、从上到下的顺序观察事物。要使读者首先抓住最重要的信息，就要把它们放在最显眼的位置，也就是12点钟的位置。

★ 数据项不要太多，保持在5项以内。

小白插了一句：为什么不能超过5项呢？

Mr.林：你若看过“定位理论”就理解了，人脑比较容易记住前5位，数据太多会分散注意力。

小白：哦，我想起来了，例如做调查问卷时，一般规定的最多评分级设置是“1分，3分，5分，7分，9分”，在这几个数值中进行打分，再多的话大家就会懵了。也是这个道理吧！

Mr.林：对的，我们继续第3点。

★ 不要使用爆炸式的“饼图分离”。

Mr.林：“饼图分离”是指将扇区都分离开来，感觉上好像摔碎的西瓜一样，如图7-3所示，这种做法不但不美观，而且也不方便阅读。不过可以将某一片扇区分离出来，前提是希望强调这一片扇区。

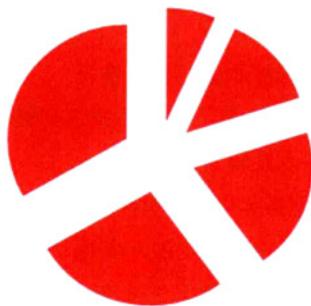


图7-3 分离后的图表

★ 饼图不要使用图例。

小白：不用图例，别人怎么知道那一部分代表什么数据呢？

Mr.林：饼图使用图例的方式阅读起来很不方便，可将标签直接标在扇区内或旁边。

★ 尽量不使用标签连线，如果要用则切忌凌乱。

★ 尽量不使用3D效果，如果要用厚度要尽量薄一些。

★ 当扇区使颜色填充时，推荐使用白色的边框线，具有较好的切割感。

Mr.林：这样说，估计你也记不得几个。所以我们继续“大家来找茬”的游戏，或许你的印象会深一些。看一看图7-4，结合我刚刚讲的几点，找找它有哪些错误。

小白一边看图，一边一条条地对着笔记说道：这张饼状图存在的问题有：

★ 没有从12点钟的位置开始排列。

★ 使用了“饼图分离”效果。

★ 使用了图例，确实不便于阅读。

★ 使用了3D效果，如果图中没有标出数字的话，我就不知道哪个大、哪个小了。

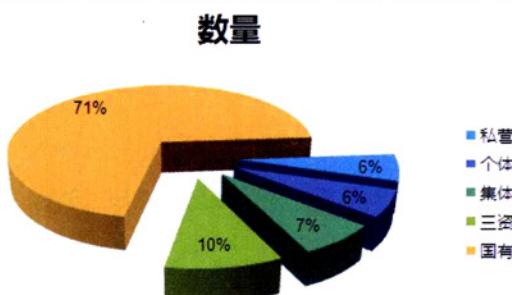


图7-4 修改前的饼图

>> 谁说菜鸟不会数据分析（入门篇）

Mr.林：是的。使用3D效果使读者无法清晰区分面积接近的扇区，容易误导读者。使用图例进行标注，这样增加了读者读图的难度。但你还漏了一点，在开始的时候我们提到的：

- ★ 标题没有反映出图表主题。

Mr.林：下面就讲一讲怎么修改。

- STEP 01** 修改为一句话标题“某市各性质企业构成情况”。
- STEP 02** 选中图表，单击鼠标右键，选择“设置数据系列格式”，在弹出的对话框中将“第一扇区旋转角度”调整为“无旋转”。
- STEP 03** 继续在上一步的“设置数据系列格式”对话框中，调整“饼图分离程度”为“不分离”。
- STEP 04** 使用“更改系列图表类型”将图表更改为普通饼图，也就是无3D效果。
- STEP 05** 删掉图例，将文本标签直接标在扇区上或旁边，尽量不使用标签连线。在Excel中添加标签的操作如下：在图表上单击鼠标右键，选择“设置数据标签格式”，在弹出的对话框的“标签选项”→“标签包含”项中勾选“类别名称”。
- STEP 06** 可以把每个扇区边框颜色设置为白色，既美观，而且打印时可以清晰显示。

得到的效果图如图7-5所示。

某市各性质企业构成情况

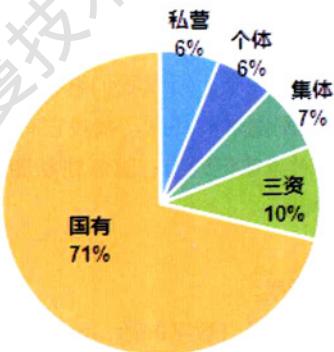


图7-5 修改后的饼图

小白拍手：哇，修改后的饼图果然清晰明了很多。

◎ 复合饼图

停顿了一会儿，小白又问了：您刚刚说饼图的成分不能超过5块，那如果数据项超过5项怎么办呢？

Mr.林立马答道：这个问题好解决！我们可以将5项以外的数据放入到第二个饼图中，构建复合饼图或复合条形图，如图7-6所示。

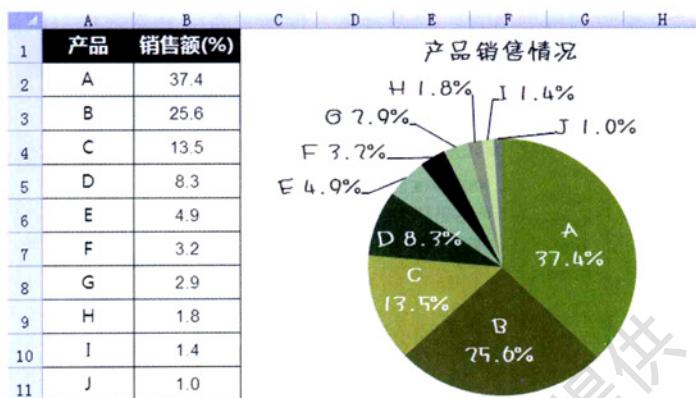


图7-6 产品销售情况图表

应用复合饼图或复合条饼图可以将分类数据中所占比例较小的分类合并成“其他”项，使图表反映的数据更能突出重点，操作步骤如下。

STEP 01 在如图7-6所示的工作表中选取A2:B11单元格区域，在“插入”主选项卡中，单击“饼图”下拉菜单，选择“二维饼图”里面的复合饼图，如图7-7所示。

得到如图7-8所示的复合饼图，这个复合饼图还不符合我们的要求，我们要将销售额占百分比小于6%的都归为第二绘图区，这样就要做一些设置了。



图7-7 “饼图”下拉菜单

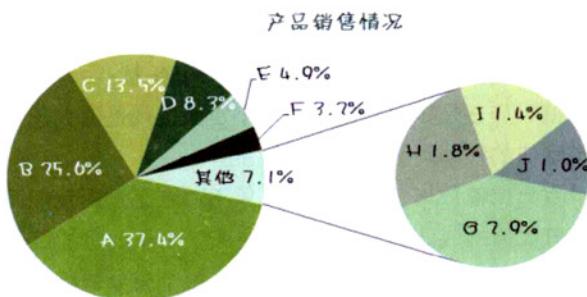


图7-8 复合饼图（初始图）

STEP 02 在Excel里，将第一个饼图称为“第一绘图区”，第二个饼图称为“第二绘图区”。如何按照我们的要求调整每一个产品的绘图区呢？首先选取图表中的饼图，单击鼠标右键，打开“设置数据系列格式”对话框，如图7-9所示。

例如，本例要求第一个饼图包含5个成分（其中一个为“其他”项），第二绘图区则包含剩

下的6个成分。在“设置数据系列格式”的设置是：保持“系列分割依据”默认选项“位置”，再将“第二绘图区包含最后一个”右边的数值选项调为“6”，如图7-9所示。

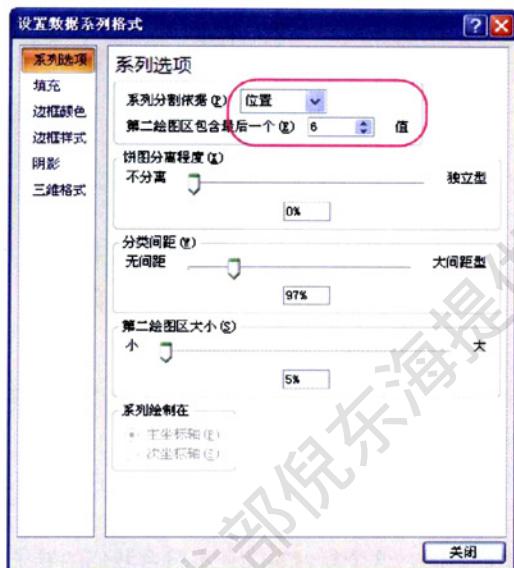


图7-9 数据系列格式

小白边听边跟着操作：等等！“系列分割依据”除了默认的“位置”项，它的下拉菜单还有“值”“百分比值”“自定义”的选项。什么时候用到这些选项？

Mr.林露出欣慰的笑容：嗯，好奇心是数据分析师必备的品质之一。但是，光会发现问题还不够，还得学会自己寻找答案。你好好想一想，“大胆猜测，小心求证”！

小白一边操作，一边用试探的语气问Mr.林：那就麻烦您看看我猜得对不对？

“值”应该是指由我们设定一个定值，凡是值大于或等于该定值的成分则放入第一个饼图，小于该定值的移动到第二个饼图。

“百分比值”与“值”类似，只不过这里是由我们设定一个百分比值，即占比大于或等于规定的百分比值的成分则放入第一个饼图，小于规定的百分比值的则移动到第二个饼图。

“自定义”就是可以任由自己调整每个成分的绘图区，但是怎么调整呢？

Mr.林：很好！“值”和“百分比值”你都说对了，我补充一下，“百分比值”这个选项用得比较广泛，例如在研究产品市场份额时，市场份额大于或等于20%的产品是我们特别关注的对象，我们想将其放在第一个饼图，小于20%的是次要关注对象，故将其放到第二个饼图，这时候用“百分比值”选项，设置“第二绘图区包含所有小于该值的值”为“20%”即可，非常方便。

下面你再选择“自定义”选项，然后用鼠标在原图上选中需要更改绘图区的成分，然后看看“点属于”右边的下拉菜单选项中有“第一绘图区”与“第二绘图区”，选择你要放置的绘图

区，即可完成对某个成分的绘图区设定，如图7-10所示。

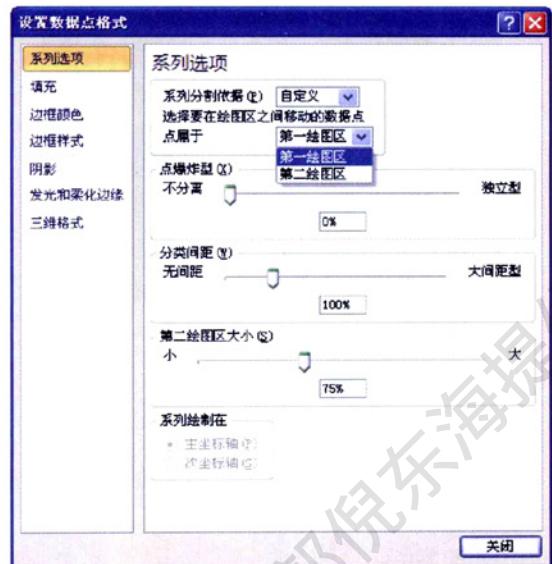


图7-10 自定义绘图区

另外，还可以在“分类间距”区域调整两个饼图之间的距离，在“第二绘图区大小”区域调整两个饼图的大小比例，然后单击“关闭”按钮完成复合饼图的制作，得到如图7-11所示的效果图。

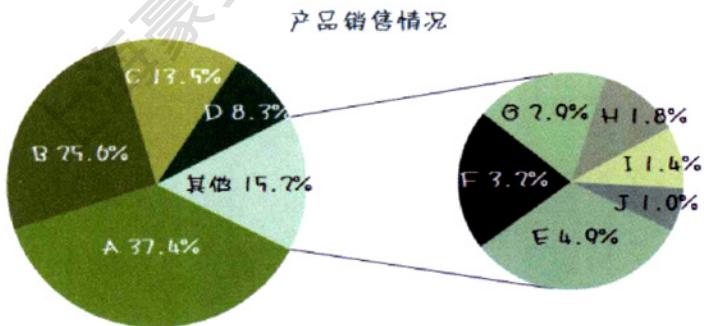


图7-11 复合饼图

Mr.林：我在前面提到，除了可以使用复合饼图表示多个成分，还可以使用复合条形图，效果是一样的。例如我们将上面的复合饼图转化为复合条形图，只需要一步操作。

STEP 03 选取图表中的饼图，单击鼠标右键，选择“更改图表类型”，选择“饼图”区域中，条形图的图标 \square 。然后单击“确定”按钮，将图表修改为复合条饼图，如图7-12所示。

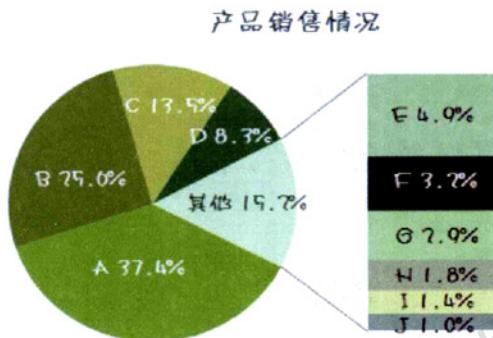


图7-12 复合条形图

◎ 柱形图

Mr.林：柱形图在日常工作中也是经常用到的，比如在数据对比分析时就会经常遇到。现在我们来看看柱形图需要注意的地方。

★ 同一数据序列使用相同的颜色。

★ 不要使用倾斜的标签，别让读者歪着脑袋看。

小白听到第2条，不由得嘿嘿一笑：我作图的时候还真没有想过这么多呢。

Mr.林：当项目名称，也就是横坐标轴的标签文字过长时，可以采用条形图来代替柱形图。

★ 纵坐标轴一般刻度从0开始。

同样，我给个例子一起来找茬，如图7-13所示。

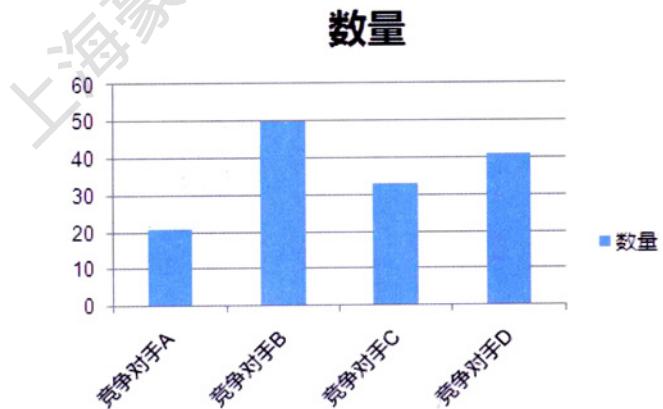


图7-13 修改前的柱形图

小白：这个错误很明显啊！

★ 图表横坐标轴的标签倾斜显示。

- ★ 没有使用一句话标题。

Mr.林补充道：嗯，不错，我还有两点要补充。

- ★ 一般来说，柱形图最好添加数据标签，这样让读者一眼就能看到具体数值。
- ★ 如果柱形图已经有了数据标签，纵坐标刻度线和网格线则显得多余了，最好删除。

Mr.林：你再看我是如何修改的。

- STEP 01 调整横坐标轴标签的对齐方式为 0° ，或通过缩小标签字体大小实现横向显示。这里要提醒一下的是，如果以上两种方式都无法改变倾斜显示，可以使用条形图作图。
- STEP 02 在数据条图中添加数据标签，并设置为Arial字体，这种字体在英文印刷品中经常用到。
- STEP 03 去掉网格线。
- STEP 04 去掉纵坐标轴。

修改后的柱形图如图7-14所示。



图7-14 修改后的柱形图

小白兴奋地说道：果然，经Mr.林一加工，图表的效果不同凡响。

Mr.林也开玩笑道：少拍马屁了啊，我可不是牛董，呵呵。

◎ 条形图

Mr.林：接下来我们看看条形图，条形图与柱形图的注意事项是类似的。

- ★ 同一数据序列使用相同的颜色。
- ★ 尽量让数据由大到小排序，方便阅读。
- ★ 不要使用倾斜的标签。
- ★ 最好添加数据标签。

Mr.林：同样，我们来找一找图7-15存在的问题，因为与前面的柱形图类似，所以我直接说了。小白，你听一听，想想还有什么要补充的。

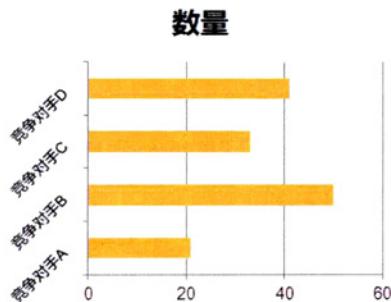


图7-15 修改前的条形图

这张条形图的错误有：

- ★ 没有使用一句话标题。
- ★ 图表纵坐标轴标签倾斜显示。
- ★ 条形图中各数据条未添加数据标签，给读者读图带来了不便。
- ★ 因为考虑添加数据标签，所以网格线和横坐标轴是多余的。

在没有数据标签的时候可以加上网格线，让图中的数据条有标准数据刻度可参照。可当有数据标签时，网格线纯属多余。

小白抢着补充：哈哈，还有一点，刚刚讲到的，让数据按由大到小的顺序排列，方便阅读。

Mr.林赞许地点头：嗯。对于这张图表，修改的思路与步骤是：

- STEP 01** 将标题改为“各竞争对手数量”。
- STEP 02** 调整纵坐标轴标签的对齐方式为 0° 。
- STEP 03** 在数据条图中添加数据标签，并设置为Arial字体。
- STEP 04** 去掉网格线。
- STEP 05** 去掉横坐标轴。
- STEP 06** 按各数据条的大小进行排序，使得数据条从上至下，由长到短排列。

小白：怎样将数据条从上至下，由长到短排列？

Mr.林：对源数据先进行升序排列，即可得到如图7-16所示的条形图。

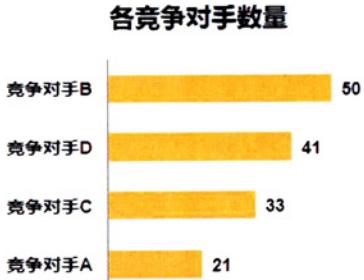


图7-16 修改后的条形图

小白：酷！您已经介绍完饼图、柱形图、条形图，还剩下一个折线图是吧？

◎ 折线图

Mr.林：对！最后一个要介绍的就是折线图。折线图有下面几点需要注意。

- ★ 折线选用的线型要相对粗些，最好比网格线、坐标轴等更突出。
- ★ 线条一般不超过5条，否则像意大利面条，非常杂乱。如果线条太多的话可以分开做图表。
- ★ 不要使用倾斜的标签。
- ★ 纵坐标轴一般刻度从0开始。

Mr.林：小白，这次由你来找找图7-17中的错误。

产品A价格走势

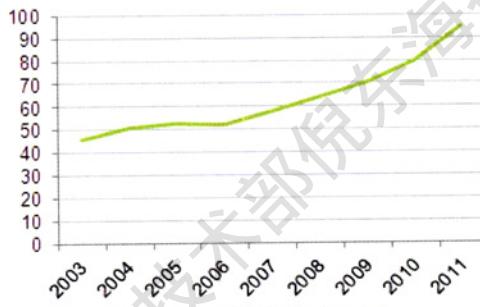


图7-17 修改前的折线图

小白胸有成竹地说：好的，您看我找的对不对，有没有遗漏：

- ★ 图表横坐标轴的标签倾斜显示。
- ★ 网格线多余。
- ★ 图中折线未添加数据标签，给读者读图带来不便。

然后呢，图表修改思路可以是这样的：

STEP 01 调整横坐标轴标签的对齐方式为 0° ，或通过缩小标签字体大小实现横向显示。

STEP 02 去掉网格线。

STEP 03 在折线图上添加数据标签，并设置为Arial字体。

Mr.林：呵呵，别得意过头了，还有两点我没教你，我来给你补充一下。

STEP 04 这里需要将2011年数据点的线条线型更改为虚线，因其属于预测值，并非实际值，故用虚线表示。

STEP 05 因为这些都是离散的数据点，而且数量不多，最好通过“设置数据系列格式”对话框设置数据标记。这里设置圆圈标记样式，圆圈标记内填充白色，以折线图的线条颜色为圆圈标记线颜色，并加粗，得到如图7-18所示的效果图。

产品A价格走势

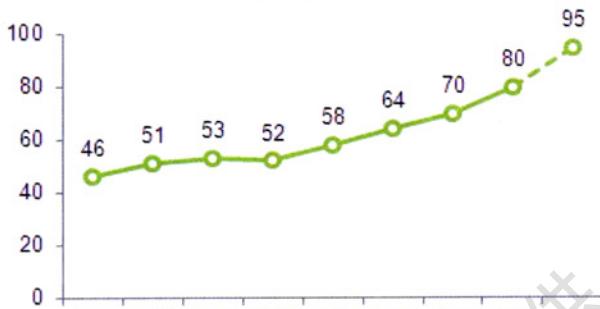


图7-18 修改后的折线图

小白：姜还是老的辣，原来还留了一手呀！

7.1.3 图表会说谎

Mr.林：小白，你见过图表说谎吗？

小白惊讶地重复了一遍：图表会说谎？

Mr.林娓娓道来：是的。刚刚介绍的条条框框，说的都是新手容易犯的错误，而有时候，具有一定经验的专业人士可以利用自己的技能夸大或掩盖事实，也就是明知故犯。

大多数人面对海量的数字都有一种天生的畏惧感，但是对图表却是另眼相待。特别是那种生动的乃至夸张的图表很容易吸引观众的眼球，所以有些数据分析师正是利用这一点，对图表做了个易容术就轻轻松松地欺瞒了观众的眼睛。据我所观察到的现象，这种被掩盖或夸张了的图表在很多地方被滥用。

小白：又没有篡改数据，怎么会欺瞒观众呢？

Mr.林：呵呵，给你看看一位网友写的一首打油诗“一个富翁上千万，邻居都是穷光蛋，平均数字一核算，人人都是富百万”。这是利用平均数掩盖真实情况的最佳说明和解释。数据尚且能够蒙蔽人的眼睛，何况是花样繁多的图表呢？下面我给你一一列举些例子。

◎ 虚张声势的增长

Mr.林：人们最喜欢研究一条线的发展趋势了，例如股市的发展趋势，CPI的发展趋势，房价的发展趋势，销售额的增长趋势……。有时候有些报刊杂志为了吸引读者故意夸大变化趋势，如图7-19所示，这是某产品的价格增幅，我们从原始图看到的增长是缓慢的。

小白：是啊，价格走势较为平稳。

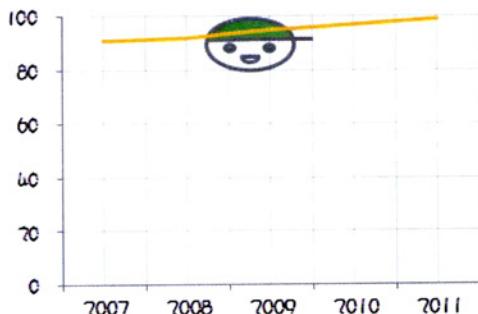


图7-19 正确的趋势图

但是，你再看看图7-20这张图表，是不是感觉这价格在飞速增长？到底哪张图表才是正确的呢？

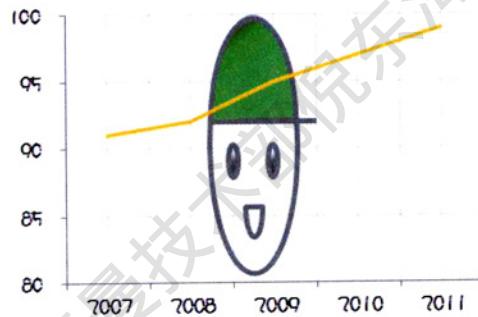


图7-20 拉长了的趋势图

小白：哈哈，这个你教过我，我说来吧。其实制图者只是调整了纵坐标轴的刻度，本来应该是从0开始的，结果他偷偷地换成从80开始了。原来纵坐标总长度是100，现在变为20了，相当于纵向拉长了5倍。当然得到的图形也大相径庭，读者一不小心就会被表象欺骗了。

Mr.林满意地说：不错。你看我还特意在这两幅图后面加了个小人，这小人是不是也感觉被人扯长了很多呀！

小白：就像照哈哈镜一样，呵呵。

◎ 3D效果的伪装

Mr.林：前面在讲图表注意事项时，提到过我们要去掉3D效果的图形。

为什么呢？很多人认为3D的图表别人都没用过，够炫，能激起读者的兴趣。可是，根据“简约”原则，3D效果实在是无关紧要的装饰，反而使图表显得拖沓繁冗。而且，3D效果的图表让人理解起来困难重重，读者可能因为3D角度的关系而无法看清相关坐标上的数据。如果非要用，建议用得薄一些，不宜显得太厚。

Mr.林：看看下面两柱形图，一个有3D效果（如图7-21所示），另一个无3D效果（如图7-22所示），图7-21看上去是不是A→B→C→D→E依次递增？其实再看图7-22，D是高于E的！

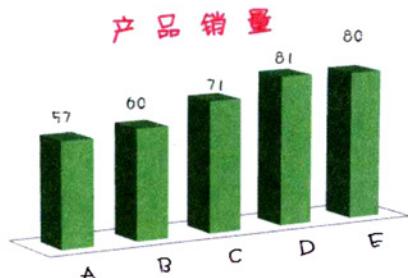


图7-21 有3D效果的柱形图

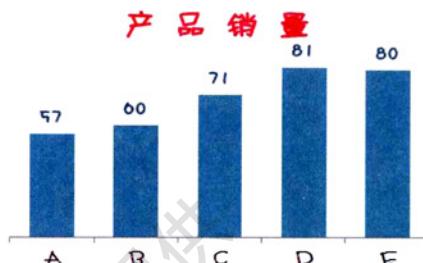


图7-22 无3D效果的柱形图

小白：怎么会这样？确实有一种被欺骗的感觉。

◎ 逆序排列的误导

Mr.林：再拿一个柱形图作为例子，如图7-23所示，给你的第一印象是不是数据是依次递减的？

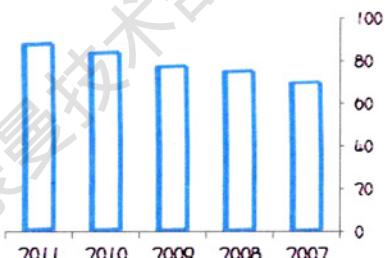


图7-23 逆序排列的柱形图

小白：是啊！不对，你看它的横坐标是从2011年到2007年排列的，干嘛将它逆序排列呀！

Mr.林：嗯，被你发现了，实际的图表为图7-24，数据是依次递增的，这样读者就不会被误导了。

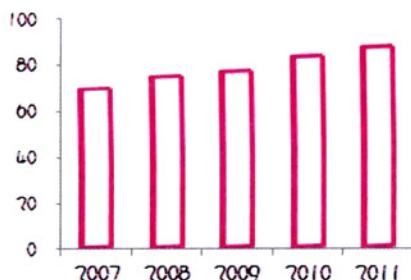


图7-24 实际的柱形图

小白点头：嗯！差点就被它给骗了。

◎ 一维图形的障眼法

Mr.林：有时候报刊杂志为了增强视觉效果，吸引读者，就将简单的柱状图、条形图、饼状图用图片填充，这样给人感觉更形象。小白，你看看图7-26的原型图7-25，它要传递的信息是今年的粮食产量是去年的两倍。



图7-25 粮食产量图的原型

图7-26 粮食产量图

小白：我咋觉得这今年的米袋跟去年的米袋比较起来，哪是大两倍，感觉上大好几倍呢！这又是视觉效果造成的吗？

Mr.林：呵呵，是制图者又耍了一个小伎俩，本来应该将米袋图片增高2倍，他却同时也将长宽增大了两倍，所以体积却变成了8倍。理解吗？

小白：似乎有点理解了，难道不能用剪切画代替原图了吗？

Mr.林：呵呵，可以的。那你再看下面两幅奶牛图。同样，图表的信息是2011年的奶牛养殖量是2006年的2倍。图7-27是采用刚才那个制图者的做法，将奶牛图的长和宽同时扩大了2倍。其实正确的做法是将其采用图7-28的形式，这才是正确的视觉效果。

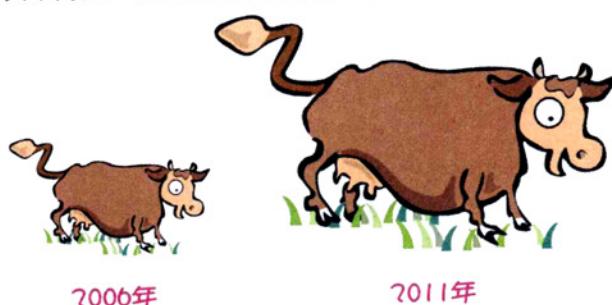


图7-27 奶牛养殖量对比图1（错误）

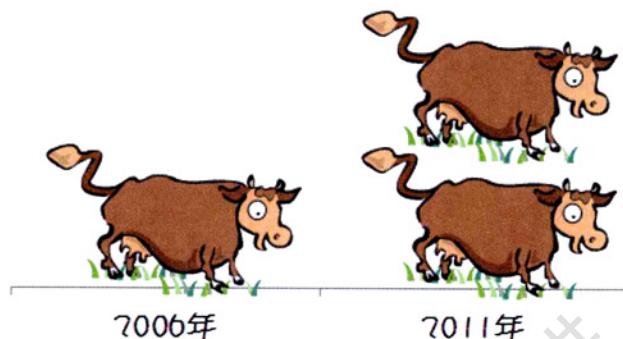


图7-28 奶牛养殖量对比图2(正确)

小白：这下才感觉对了！原来图表还可以这样子撒谎的，神不知鬼不觉就把读者给糊弄了。今天是长了见识，本姑娘以后可没那么容易上当了，呵呵。

7.2 浓妆淡抹总相宜——图表美化

Mr.林：小白，见证了这么多错误的图表，现在有什么感想？

只见小白眉头紧锁：我现在知道标准的图表应该如何绘制了。但是我有一个问题，到现在为止您都是在教我怎么一步步地按程序操作，难道数据分析师在图表方面不能有一点创新吗？

Mr.林释然，笑道：不是你想的这样。的确，数据分析需要大量的数据计算、逻辑分析、依程序画图，但是数据分析绝不仅仅是个力气活，它是一门艺术。数据分析师不仅要懂业务、懂管理、懂分析、懂工具，即开动你的“左脑思维”；还要懂设计、懂创意、懂艺术，即调用你的“右脑思维”。所以说，做数据分析得双管齐下，没两把刷子是做不了数据分析师的！

小白：愿闻其详。

7.2.1 图表美化的三原则

Mr.林突然停顿了一下，走到窗户旁，示意小白望向窗外：小白，你看看大街上行色匆匆的路人，他们的着装都有些什么风格？

不知道Mr.林葫芦里卖的什么药，小白不以为然地回答道：有鞋尖发亮圆肚直挺的老板，有穿豹纹皮袄高跟靴的性感女郎，有西装笔挺肩背电脑包的IT白领，有一身蓝色中山装的古怪老人，还有衣衫褴褛邋遢不堪的乞丐……

Mr.林：那在这形形色色的人群中，哪种风格给你的感觉最舒服？

“我呀，我这人头脑比较简单”，小白边说边手指着一个方向：您看，那位在等公交车的女

士就是我感觉最舒服的，简约、干净、得体、自然。

Mr.林回归正题了：我之所以这样问你，是因为图表美化也和着装一样，不仅要穿暖，还要穿得有讲究，给别人一种舒适、得体的感觉才是最好的。图表美化要遵循什么样的原则呢？

◎ 简约

漫画大师Scott McCloud曾说过“如果艺术家能以简单的手法直接表露事物的本质，那么他的创作意义将被大幅提升”。

简明扼要、清晰明了，才是一名图表设计师所追求的境界。但是我们所处的灯红酒绿的纷繁世界里，很多人都不愿意凡事从简，把简单定义为没有想法、没有深度、才疏学浅，而追求“多总比少好”的思想，往往弄得自己和他人不堪重负，劳人伤己。所以，追求事物的本质才是现在我们所应推崇的处事原则。

不过，需要澄清的是“简约”不是偷工减料、规避复杂，也不是推崇浅陋、毫无内涵的表达方式，我们定义的“简约”是指清晰明了，让人一看就明白，而不是给人遮遮掩掩，晦涩的感觉。我举个例子，小白，你逛过宜家家居吗？

说到宜家，小白兴趣就来了：前几天还去过！很喜欢那里，宜家的家具给人感觉色彩鲜明、简单大方且独具匠心。

Mr.林：嗯，宜家的简单设计并没有给你感觉很低俗吧！另外你还可以学习学习《经济学人》《华尔街日报》《纽约时报》这些世界顶级商业杂志上的图表，还有像罗兰贝格、麦肯锡等这些世界顶尖的咨询公司，它们都有专门的图表部门或图表顾问团队负责设计和制作图表。下面就是从国际知名咨询公司的报告里截取出来的图表，如图7-29与图7-30所示。

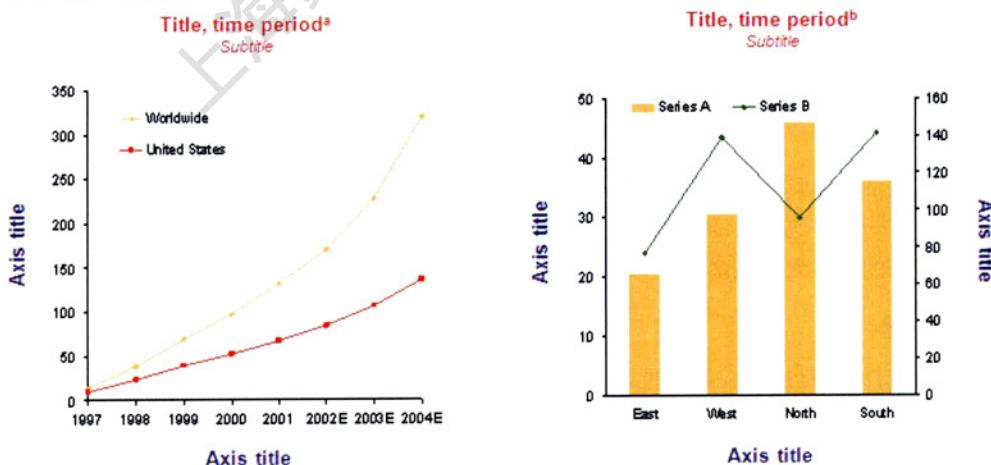


图7-29 某知名咨询公司图表模板

>> 谁说菜鸟不会数据分析（入门篇）

Mr.林: 有没有发现，越是这种专业的大公司，绘制的图表越是简约。所以，我们作图也要追求“图简意赅”！

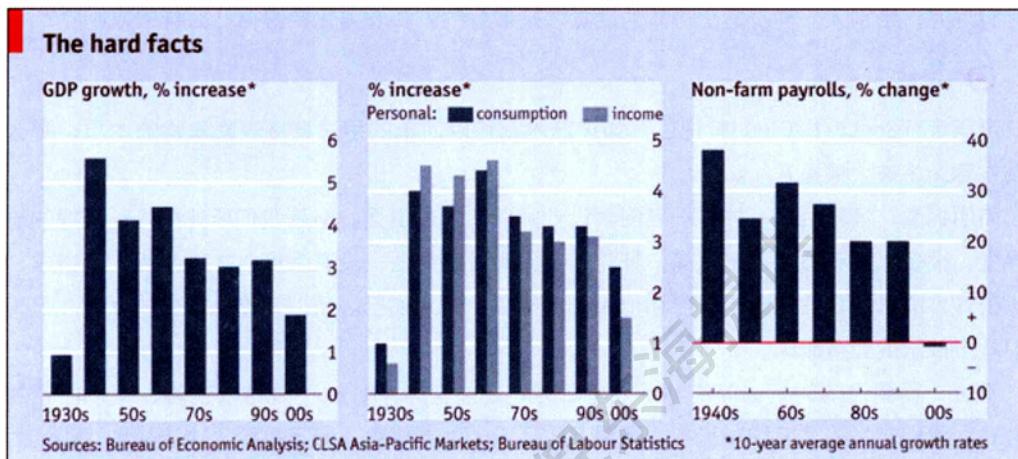


图7-30 《经济学人》图表模板

◎ 整洁

Mr.林: 刚刚你讲的那位等车女郎，她给你感觉很舒服的原因还有一个：整洁。仅仅是简单还不够，乞丐也穿着简单啊，为什么你不喜欢呢？所以说，整洁也是美化原则中很重要的一点。

整洁体现在图表中的特点是：①整整齐齐；②干干净净；③和谐自然。

在我们的图表中有很多的元素，这些元素不能排列得杂乱无章。还记得我们讲过做条形图时要让数据由大到小排序吗？这就是为了让图表看起来整齐。否则会给人不用心、不精致、不专业的感觉。如果要把图表当成一种设计，一种艺术，并乐于使之更完美，就要让图表中每一个元素之间要像被无形的线条贯穿在一起一样，排列整齐，整洁自然，这样能使读者觉得更清晰，理解起来更快。

另外，整洁也意味着自然。在图表美化中怎样体现出自然呢？有一个亲近法则，也就是把相关的内容放在一起。这样，读者会自然而然地假设那些距离较近的内容是相关的，同样也会认为那些距离较远的内容之间联系没有那么紧密，整个图面看起来结构更清晰，不会像一盘散沙。

◎ 对比

Mr.林: 对比，是指突出某些重要元素，帮助读者迅速抓住信息。小白你想象一下，在水墨画中的万绿丛中一点红，是不是格外娇艳？情节跌宕起伏的电影总会勾起你看下去的欲望吧？相反，读一本平淡无奇的小说会让你感觉味同嚼蜡，而且半天摸不透作者的意图；你也见过许多那种千篇一律的列表式的简历，让人索然无趣，看完直接想扔进垃圾箱。

小白：我想起，牛董每次提到“我们今年的销售目标是……”声调就突出高了八度，就像图7-31这样，也是通过对比强调其重要性吧，呵呵。



图7-31 提到重点时要特别强调

Mr.林：所以，我们要掌握对比的艺术，将它也应用到图表设计中。对比能更深刻地让读者体会到图表所要突出强调的信息，同时调动他们的兴趣。

对比原则运用在图表设计中，主要体现在字体（大小、粗细）、颜色（明暗、深浅）或者构图（分散、前后）等方面。在后面的美化技巧中，我会用详细的例子阐述如何巧妙地应用对比原则。

7.2.2 略施粉黛——美化图表的技巧

Mr.林：仅仅知道美化原则只是懂了图像美化的一半，就像是女人的化妆术一样，只会说“我想画得杏面桃腮，柳眉如烟，明眸善睐，再添朱樱小嘴一点点”，却不懂该怎么化妆，都是枉然。

小白机灵地接道：所以我们在了解美化原则的基础上，得学会美化图表的技巧。

Mr.林忙点头：对头！刚刚教会你要做成什么样，现在教你具体怎样去做。

◎ 最大化数据墨水比

Mr.林：首先，我要跟你介绍下最大化数据墨水比的概念，即指图表中的每一滴墨水都要有存在的理由。小白，你想象一下，若是将制作完成的这张图表打印出来，会有多少碳粉是必不可少的？有多少碳粉是被浪费掉的？由此可见，数据墨水比也就是简约原则的衍生概念，对我们的图表应该更多地关注“减”而非“加”，即尽量减少和弱化非数据元素，增强和突出数据元素。

小白：什么是数据元素？什么是非数据元素？

Mr.林：例如在图表中，曲线、条形、扇形等代表的是数据信息，故称为数据元素；而网格线、坐标轴、填充色等跟原始数据无关的就叫做非数据元素。

小白：哦，跟数据源有瓜葛的就是数据元素，跟数据源非亲非故的就是非数据元素，是吧？

>> 谁说菜鸟不会数据分析（入门篇）

Mr.林无奈地笑道：呵呵，你这样理解也对！小白，是不是想知道怎样最大化数据墨水比？我们现在换一个游戏，不玩“大家来找茬”了，玩个“擦擦看”的游戏。如图7-32所示，你想象一下手中拿着橡皮擦，仔细查看图中的每一小块、每一点内容，感觉一下用橡皮擦擦掉它会不会影响数据信息的传递，如果不受影响，那就告诉我，我来把它擦掉，直到没有地方可以再被擦掉为止。你说，我来操作，怎么样？

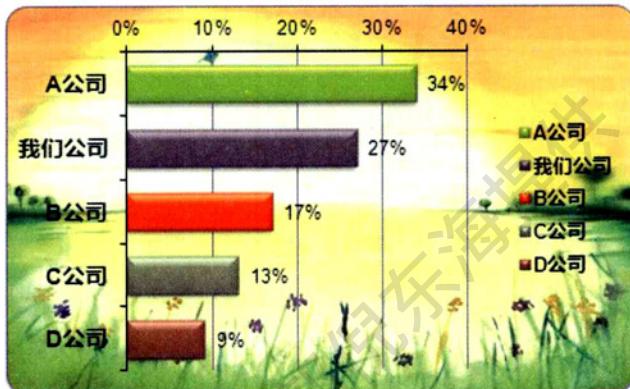


图7-32 修改前的条形图

小白：好的。哇塞！这张图我都看不下去了，真恐怖，让我想到浓妆艳抹的潘金莲的脸！那我开始“动刀”了。

首先，删除背景不会影响数据结果，可以全部擦掉；竖条的网格线也是一样，可以擦掉；还有，纵坐标有公司名称了，所以图例可以不要。哦，对了，图表边框也可以擦掉！

Mr.林：好的，根据你的指示，去除背景图片、网格线、图例、边框线条，现在条形图变成图7-33这个样子了，还需要做补充吗？

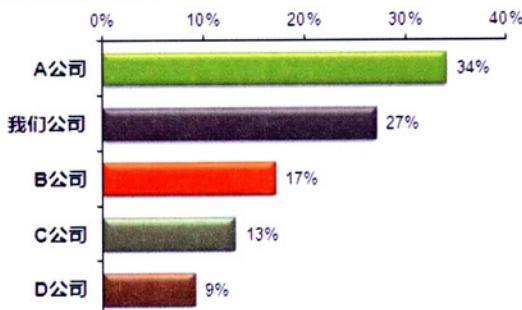


图7-33 小白修改后的条形图

小白：呃……横坐标轴貌似也不用，数据标签不是都已经标出确切值了么？

Mr.林：对，我们再把横坐标轴去除，纵坐标的线条也可以去除。还有吗？看看数据条？

小白：数据条不是反映数据内容的吗，是不能擦掉的吧？

Mr.林：的确，但是需要用到3D效果吗？之前讲过，3D效果在这里也是多此一举的。

另外再补充一点，我们最关注的信息其实就是“我们公司的市场份额排名第二”，是吧？可是在这张图里重点信息没有合理地凸显出来，花花绿绿的条状图形让读者一下子眼花缭乱，不知道要重点看什么。所以，不需要每个条状图形都用特别的颜色，只需要将“我们公司”的颜色突出一下就可以了。

最后，加上一句话标题。这就是最大化数据墨水比后的效果图，如图7-34所示。

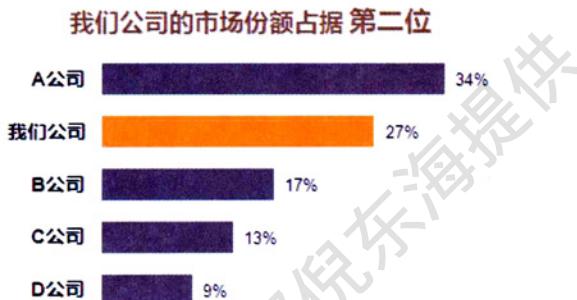


图7-34 最化数据墨水比的条形图

小白：这图改过之后果然顺眼多了。

Mr.林：那当然。再回顾一下刚刚的过程，看我们是怎样一步步做到数据墨水比最大化的。一般情况下，数据墨水比最大化的步骤如图7-35所示。

- ① 去掉不必要的背景填充色
- ② 去掉无意义的颜色分类
- ③ 去掉装饰性的渐变色
- ④ 去掉网格线、边框
- ⑤ 删掉不必要的图例
- ⑥ 去掉不必要的坐标轴
- ⑦ 去掉装饰性图片
- ⑧ 以上不能去掉的元素就尽量淡化
- ⑨ 对需强调的数据元素进行突出标识

图7-35 数据墨水比最大化的步骤

◎ 找出隐形的线

Mr.林：刚刚讲到整洁原则的时候，提到一个很重要的评价标准：要让图表中每一个元素像

被无形的线条贯穿在一起一样。最好的方法是找一条明确的线，并用它来对齐，使元素与元素之间存在着某种视觉纽带。我举个例子，小白，你看看如图7-36所示的员工对薪酬现状与期望评分图，分别是员工满意度调查中关于薪酬制度的三道题的评分，你觉得有什么问题？

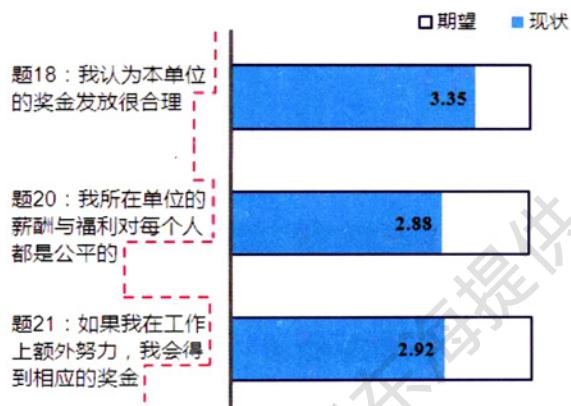


图7-36 员工对薪酬现状与期望评分 (修改前)

小白无奈地摇摇头：不知道。

Mr.林引导小白：你看，这里的题目是左对齐的，也就是在左边有一条明确的线。另外，沿着条形图的左边也有一条明确的线（坐标轴）。不过，在题目和条形图之间“留”了一些空白，看我用红色虚线标记的部分，这部分形状也不规则，很难看，而且由于留下的这部分空白，使得题目和条形图隔开了。

另外，还有个小地方，看它的图例，期望用蓝色外框表示，现状用浅蓝色方块表示，期望数据都是在现状数据的外面的，所以我们的图例也要调转一下两者的顺序，图例中的期望数据靠外侧，而且不能超出条形图。

调整后的效果，如图7-37所示，是不是显得整洁很多？

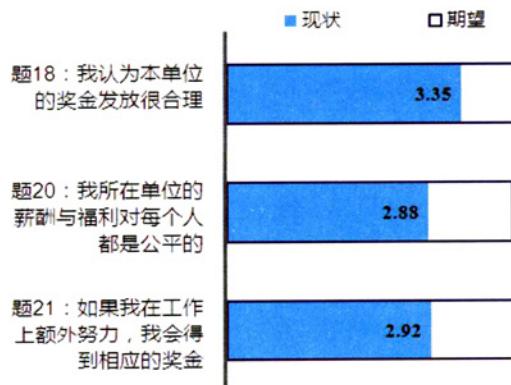


图7-37 员工对薪酬现状与期望评分 (修改后)

Mr.林: 小白，我用这个例子是想告诉你，在描绘图表的每一个元素时，都要特别注意它的位置，应当找出能够与之对齐的元素，尽管这两者之间的地理位置可能很远，但是在图中还是要做到就像是有人在中间牵了一条隐形的线将两者关联起来一样。切记不要一会儿居中一会儿右对齐，在图表上逮着个空白地方就添东西。

◎ 图表喜欢的数字格式

Mr.林: 刚刚我们讲找出隐形的线是为了使图表元素显得井然有序，这里再教你一个小技巧以增加图表的整洁感。当我们添加的数据标签很多，或者坐标轴刻度很密集时，图表上就会显示很多数字，排版时稍不注意就会显得非常拥挤和凌乱。此时，我们可以将所有数字格式都调整成Arial字体，数字将会显得“规矩”很多，效果如图7-38所示。如果图表中有英文，那么字体同样建议调整为Arial字体。

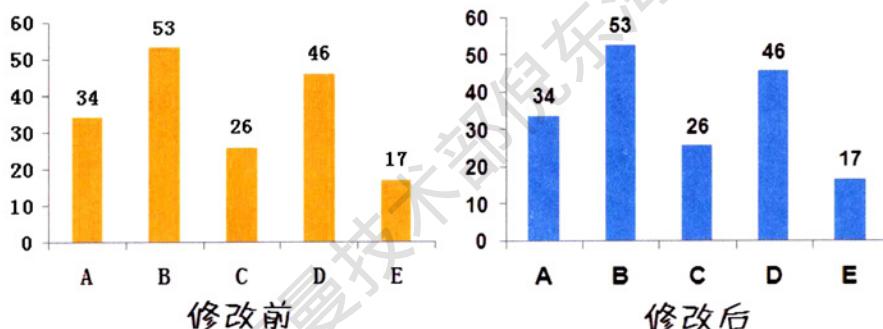


图7-38 Arial数字字体的图表

小白: 是喔，这样一对比，宋体的数字真的很丑！呵呵。

◎ 如何突出对比

Mr.林: 前面介绍完了整洁原则的美化技巧，下面再介绍对比原则的技巧。如果你要运用对比原则，突出与众不同的元素，就不要缩头缩脚，要大胆地让它变得非比寻常。

小白: 就像图7-34一样，使用醒目的颜色将元素突出吗？

Mr.林: 你说对了，最方便快捷的方式，就是改变颜色！利用对比色更能增强突出效果。

除此之外，还可使用直线、箭头或者阴影等手法。还记得我教你画的平均线图吗？有时候，我们只是简简单单地加一条线，就能进行很好的对比。如图7-39所示，产品的需求量是随着月份而变化的，而产品的每月生产量是由机器固定生产出来的，在图7-39中我们只是简单地添加了一条表示产量为恒值的直线，就能很清晰地对比出需求量与产量的关系，从而分析产量是否应该做出调整。

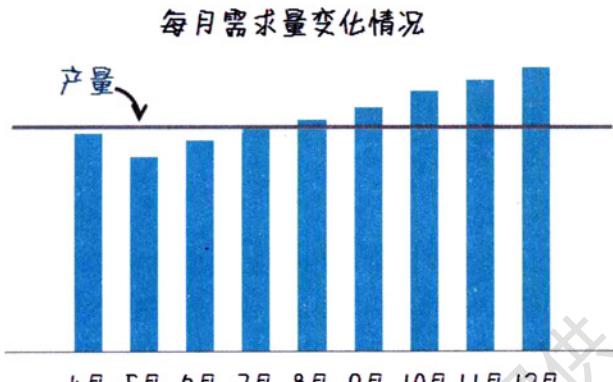


图7-39 对比手法1——直线

采用箭头的方式也有很不错的强调效果。如图7-40所示，图中描绘的是2005年至2010年某公司销售额的变化情况，这里使用箭头的目的有两个：一是将人们的注意力集中在2008年这个特殊点，由于受金融危机的影响销售额出现反常的下降；二是箭头的长度反映出下降的幅度。

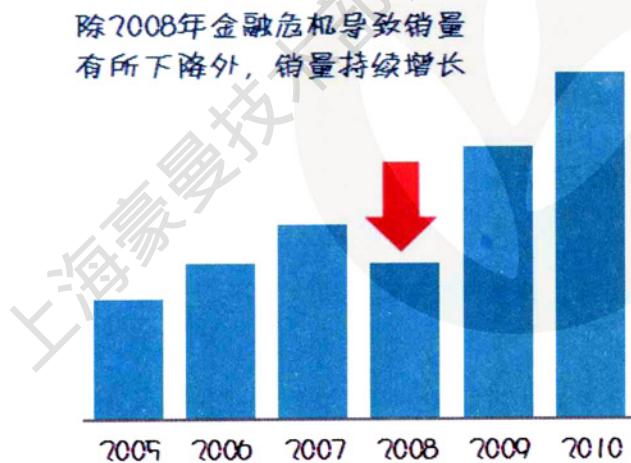


图7-40 对比手法2——箭头

其实呀，对比的方式有很多，关键在于要动脑筋想，用善于发现的眼睛去挖掘灵感。灵感藏在哪里？动动脑筋，俯首皆是。

如图7-41所示，它采用的就是简单且很有意思的对比方式：因为我们已经习惯看完整的圆饼，这里故意留出一个缺口，产生意想不到的效果，更能引人注意，而且缺口部分还起到暗示的作用，提醒你应该将整体补全，而图中的箭头更进一步提醒你这点。

经理进行自我学习的时间相对较少

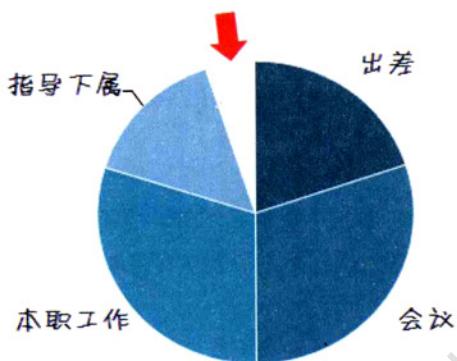


图7-41 利用饼图缺口进行强调

Mr.林：不过，小白，切记对比的目的只有两个：首先是让读者快速地领悟重要信息；其次才是吸引读者的眼球，调动他们的兴趣。不能本末倒置，为了视觉效果，把图表弄得式样繁杂，弄得人不明所以。

所以，记住简约在三个原则中永远稳坐第一把交椅，任何原则都务必以它为前提。

小白：好的，我记住了。

7.2.3 图表也好“色”

Mr.林：小白，衣服的颜色搭配你应该留意过吧？

小白：这回可算您遇上行家了！您教了我这么多，我也教教您我多年珍藏的秘笈，如图7-42所示。

主色	配色
红色	白色、黑色、蓝灰色、米色、灰色
咖啡色	米色、鹅黄、砖红、蓝绿色、黑色
黄色	紫色、蓝色、白色、咖啡色、黑色
绿色	白色、米色、黑色、暗紫色、灰棕色
蓝色	白色、粉蓝色、酱红色、金色、银色、橄榄绿、橙色、黄色

图7-42 衣服颜色搭配

Mr.林：看样子，你对颜色还比较有研究。色彩学可是一门大学问，但我们并不需要精通，懂点三脚猫功夫，在职场里无论是美化图表还是制作PPT都会加分不少。

其实色彩搭配在生活中无处不在，例如刚刚说的服装搭配，还有家居装饰、广告宣传等等，平常我们都可以多留意，毕竟我们的生活就是多姿多“彩”的嘛！

◎ 色彩的C大调

Mr.林：颜色这东西说来话长，我们不是搞美术的，也不是搞广告设计的，所以了解基本原理就够了。就像乐律，俺只懂个C大调，不过唱歌从不跑调。

小白：嘿嘿，您是从不着调吧？

Mr.林：不开玩笑了，要了解色彩的基础知识首先从学习色环开始。

1. 色环

小白：哇塞，这图7-43所示的色环，让我看得眼花缭乱了！

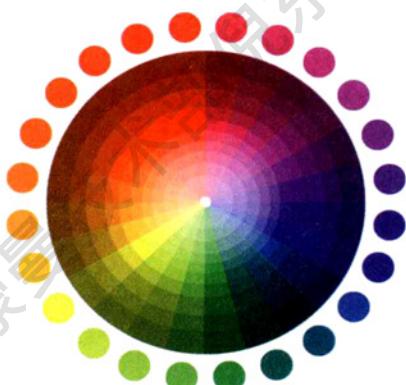


图7-43 色环

Mr.林：不慌，其实色环里只有最基本的三种颜色——红、黄、蓝，它们可是色环中所有颜色的“本源”，其他颜色都是由这三种颜色变换出来的。

小白：嗯，貌似在小学美术课本里学过。

Mr.林：然后，三原色两两混合，组成了二次色， $\text{红}+\text{黄}=\text{橙}$ ， $\text{黄}+\text{蓝}=\text{绿}$ ， $\text{红}+\text{蓝}=\text{紫}$ 。二次色所处的位置是两种原色一半的地方；二次色再两两搭配，组成了三次色，三次色即由相邻的两个二次色调合而成，如图7-44所示。

Mr.林：色彩还可以继续两两调和下去，现在你懂了吧，重要的是每一种颜色都拥有部分相邻的颜色，如此循环成一个色环。

小白：原来是这样，我有点明白了。

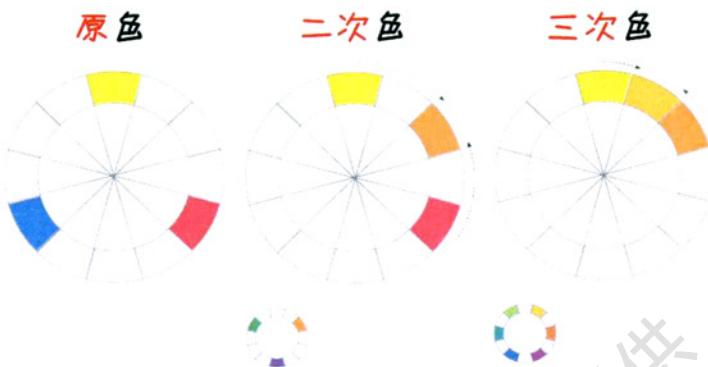


图7-44 色调原理

2. 相似色

相似色是由一种色调及其相应的多种亮色和暗色组成，如图7-45的左图所示。相似色组合在一起给人的感觉很素雅、正式，但是也要注意可能会导致的两个问题：①画面较平淡；②对象间区分度不够，有时候会让人忽视各个对象之间的差别。

3. 邻近色

邻近色，顾名思义就是色环上相邻的颜色，如图7-45中间的图所示，橙黄色、橙色和橙红色，它们都有相同的基础色，所以邻近色的色调统一和谐，能够带给观众一种明显情感特征。

4. 对比色

色环上相对的颜色是对比色，如图7-45的右图所示。由于它们相互对立，所以在表示强调和对比时可利用对比色，例如在表示产品的盈亏情况时，盈利可以用蓝色，亏损可以用红色表示。最常用的对比色是：深色与浅色、亮色与暗色、冷色与暖色，其中深色与浅色的经典用色是黑色与白色。



图7-45 相似色·邻近色·对比色

5. 冷暖色

小白: 咱是冷色与暖色啊?

Mr.林: 除了相似色、邻近色、对比色的区分, 颜色还有冷暖之分。冷色调给人的感觉是安静的、冷酷的、稳重的, 暖色调给人的感觉是热情的、奔放的、温暖的。

小白: 那红色肯定是暖色调, 而蓝色是冷色调咯? 还有哪些颜色是暖色调呢?

Mr.林: 其实不需要刻意记忆哪些是冷色调哪些是暖色调, 这主要取决于你想带给别人的感觉, 在Excel的标准调色板里也可以看到左上方是冷色调, 右下方是暖色调, 如图7-46所示。

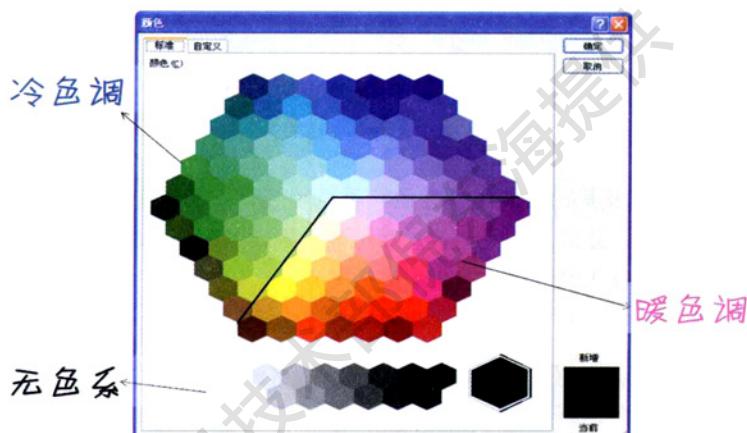


图7-46 冷暖色

在这里, 我要告诉你的就是, 我们可以通过增加一些红色或黄色将某些颜色“加热”, 也可以增加不同程度的蓝色对其“降温”。另外, 冷色总趋于后退, 而暖色是趋于前进的, 所以, 暖色不费吹灰之力就会抢走你的视线。

讲了这么多色彩知识, 最后我将概要做个总结, 如图7-47所示。

项目	说明
色系	同色系(红、黄、蓝、绿等)、无色系(黑、白、灰)
色调	暖色调(代表色:红、黄, 火热感觉), 一般冬、春季使用 冷色调(代表色:绿、蓝, 冰凉感觉), 一般夏、秋季使用
对比色	对比色指两种可以明显区分的色彩, 如深色和浅色, 冷色和暖色, 亮色和暗色, 可用于突出主题、内容或表现不同类别等
相似色	相似色指一个颜色在明度上的深浅变化(也就是同色系的颜色), 如蓝色有深蓝、浅蓝等颜色, 可用于表现类似、过渡的事物
邻近色	邻近色指如偏绿的蓝色、偏紫的蓝色, 都是蓝色邻近绿色和紫色之间的颜色, 可用于表现类似、过渡的事物

◎ 慎用的颜色

Mr.林：在所有颜色中，有三个颜色在使用的时候要注意。小白，你知道是哪三个吗？

小白：不会是交通灯那三种颜色吧？

Mr.林：没错，正是红色、黄色和绿色这三种颜色，因为它们具有特殊的含义，如图7-48所示。

红绿灯	交通	企业经营分析
红	危险，禁止通行	显示指标存在重大问题
黄	提醒即将有危险，起警示作用	显示指标存在潜在问题
绿	表示安全，准许通行	显示指标发展良好

图7-48 红绿灯

红色代表禁止或者危险，黄色代表警告和提醒，绿色代表安全、正常，所以在我们的图表里也要注意按照它们的含义来使用，而且尽量避免使用红色。

还记得我们在讲“给数据量体裁衣”的“图标集”时讲到的例子吗？如图7-49所示。

项目	A企业	B企业	C企业
第一季度收入目标（亿元）	1	1	1
完成值（亿元）	1.1	0.95	0.73
完成率	✓ 110%	⚠ 95%	✗ 73%

图7-49 2010年第一季度某集团下属三企业的收入目标及完成情况

我们当时设置的规则为：完成率大于或等于100%的企业，用绿色带钩圆圈表示，完成率大于或等于90%且小于100%的用黄色带感叹号圆圈表示，小于90%的用红色带叉圆圈表示。其实这里特别注意了颜色的内在含义：用绿色标记完成业绩目标的企业，红色标记未完成业绩的企业，黄色标记快完成业绩目标的企业。

7.3 如虎添翼的招儿

Mr.林：最后，根据我多年积累的实践经验，教你一些实用的招儿。有了这些招儿，如同武林高手配备了独门利器，就能如虎添翼，提高工作效率，使你的图表制作变得更得心应手。

Mr.林的形容果然奏效，小白竖起了耳朵往下听。

7.3.1 我的图表模板

Mr.林: 首先, 创建自己的图表模板。有时候你好不容易制作了一些自我感觉良好的图表想把它保存下来, 或者做了一个图表模板, 在以后的工作报表中想经常调用。小白, 你会怎么做呢?

小白: 每次打开这个文档, 再复制→粘贴→更改数据源?

Mr.林: 哎, 你要这样做可真是“小白”了。在Excel里有内置存放模板的操作, 例如我们用鼠标选中如图7-50所示的饼状图, 选择“图表工具”主选项卡中的“另存为模板”, 在弹出的对话框中定义自己模板的名称, 这里定义名称为“绿色饼状图”, 单击“保存”按钮就行了, 如图7-50所示。

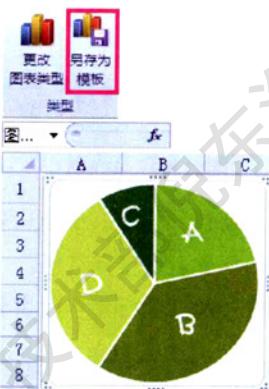


图7-50 保存绿色饼状图模板

如果下次你还想调用这张图表模板, 可以在“更改图表类型”对话框中, 选择模板“绿色饼状图”, 甚至还可以将它设置为默认图表类型, 如图7-51所示。



图7-51 “更改图表类型”对话框

7.3.2 快速制图

Mr.林: 教会你怎样保存自己的图表模板，接着再教你一招儿快速制图的方法，以图7-52数据表制图为例。

区域	第一季度	第二季度	第三季度	第四季度
东部	20.4	27.4	90.0	30.4
西部	30.6	38.6	34.6	41.6
北部	45.9	46.9	45.0	53.9
南部	32.3	37.6	56.5	41.9

图7-52 某产品2010年销售情况——数据表

选中如图7-52所示的数据区域，然后按F11键，则Excel会即刻新建一个图表工作表“Chart1”并创建柱形图，如图7-53所示。

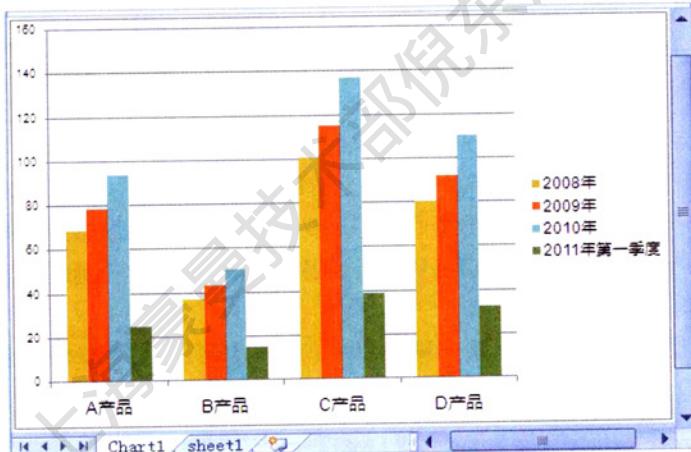


图7-53 快捷键方式

使用快捷键方式创建的图表为Excel默认的图表类型，并且这个图表会绘制在新建的工作表中（如图7-53中的Chart1）。这就有个弊端，当默认图表类型不是我们想要的图表时，就需要重新更改图表类型。至于更改图表类型的方法，前面介绍过：单击图标→右击鼠标选择“更改图表类型”。

小白顺溜地接过Mr.林的话茬：所以，当我们采用快速制图方法时，得事先把默认图表类型设置成自己想要的图表格式。嘿嘿，我发现这个方法非常适用于批量生产相同格式的图表。

Mr.林：呵呵，是的，其实采用普通的插入图表的方式作图也不麻烦。在前面提到过：单击“插入”主选项卡，在图表功能栏里选择你要插入的图表类型，例如这里插入“堆积柱形图”，效果如图7-54所示。

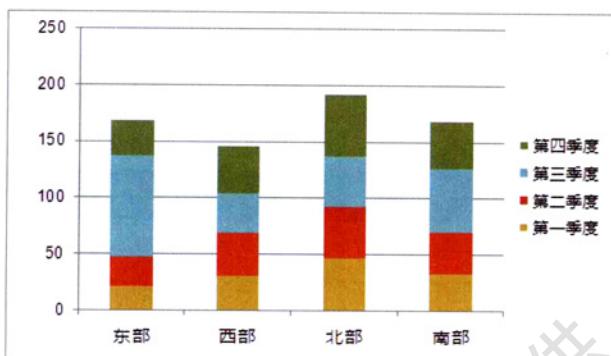


图7-54 用插入图表的方式快速作图

7.3.3 添加标签小工具

Mr.林：还记得我们介绍绘制散点图注意事项时的例子吗？如图7-55所示，十个品牌知名度和满意度散点图看上去并不是很直观，是不是？

小白：是啊，我想知道每一个点分别对应的是哪个品牌。

Mr.林：我们经常会添加“数据标签”，以分辨每个数据点对应的名称，如果采用在“图表工具”主选项卡选择“布局”里的“数据标签”，要么添加的是X轴或Y轴的数据，要么添加的是“系列一”，总之不是我们想要的品牌名称，所以很多人就只好一个个地手动添加标签。这个问题是Excel的一大缺陷，困扰了很多人。

其实，只要我们借用一个小工具——JWalk Chart Tools加载宏，轻轻松松就能搞定标签问题！

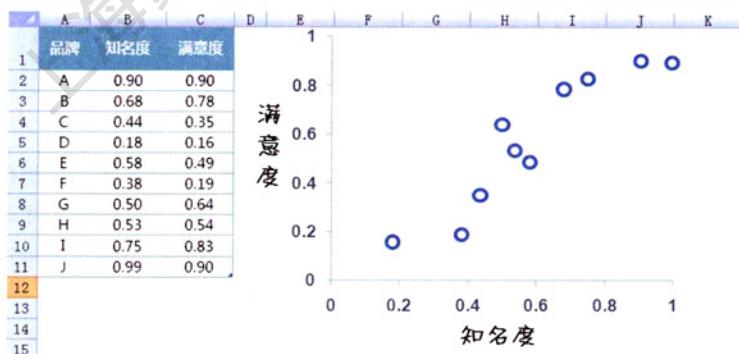


图7-55 十个品牌知名度和满意度散点图

下面我告诉你具体怎么操作。

STEP 01 加载JWalk Chart Tools宏。单击Excel的“文件”主选项卡→“选项”→“加载项”，

注意最底端的“管理”项右侧有个下拉菜单，选择“Excel加载项”，再单击“转到”按钮，如图7-56所示。



图7-56 转到Excel加载项

随即弹出如图7-57所示的“加载宏”对话框，此时，“可用加载宏”下面的文本框是没有“JWalk Charting Tools”选项的。单击“浏览”按钮，打开存放JWalk Chart Tools工具的路径，因为之前我将JWalk Chart Tools工具放置到桌面，所以单击“桌面”→双击“JWalk Chart Tools”文件夹→单击“chart tools”→单击“确定”按钮，Excel就返回到之前的“加载宏”对话框。



图7-57 加载宏

如图7-57所示，“可用加载宏”下面的文本框内多了一个“JWalk Charting Tools”的选项，勾选此项，再单击“确定”按钮。此时，就完成了加载宏JWalk Chart Tools的操作，如图7-58所示，我们的Excel里多了一个“加载项”的主选项卡，JWalk Chart Tools工具就藏在这里了。

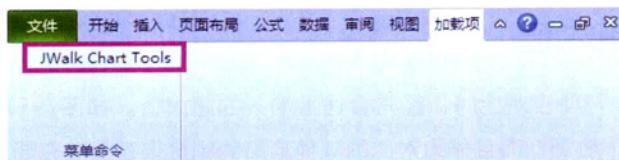


图7-58 加载后的宏

- STEP 02** 单击选中图7-55的所有数据点，在“加载项”里选择刚添加的工具JWalk Chart Tools，即弹出如图7-59所示的对话框。

>> 谁说菜鸟不会数据分析（入门篇）

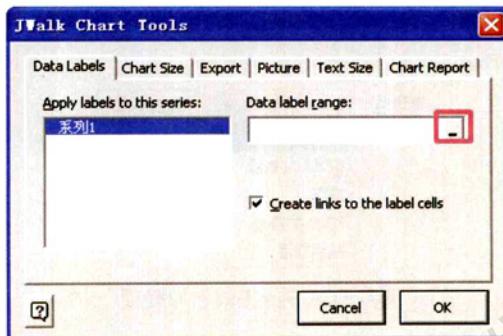


图7-59 添加数据标签对话框

STEP 03 单击“Data label range”下拉列表框，就是图7-59所示的用红色方框标记的部分，然后选择我们要添加的品牌名称，即A2:A11区域。

STEP 04 单击“OK”按钮，则完成了标签的添加，效果如图7-60所示。

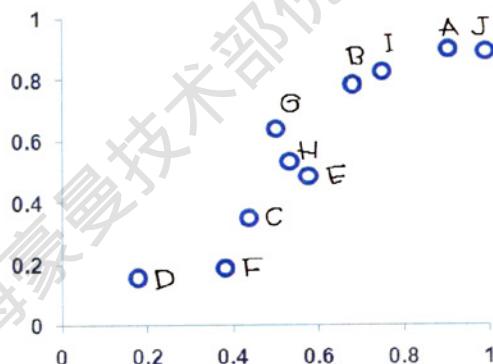


图7-60 添加文本标签后的散点图

小白：这都能搞定？哈哈，我又学到了一个别人不会的招儿！

7.3.4 修剪超大值

Mr.林：我们在进行项目对比时，经常会遇上恼人的超大值。如图7-61所示，产品E的销量一枝独秀，是不是很不协调？而且作为对比的其他产品的销量很容易被忽略。

小白：哇，产品E的確是“鹤立鸡群”。

Mr.林：而且，我们的目的并不是为了突出E产品销量有多么的高，这早在我们的预期之内了，我们想看看每个产品的销量如何。这时候，可以适当为产品E剪剪枝。

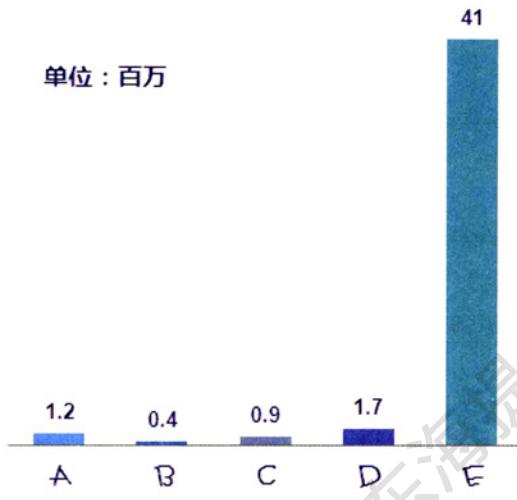


图7-61 各产品年销量

STEP 01 更改产品E的数据，让E的柱子显示为适当的高度。这里我们将41改成5，刚好合适。

小白：你这样更改数据，不是违规了吗？而且数据标签也变为“5”了。

Mr.林：别急啊！接着按下面的步骤操作。

STEP 02 单击两下E的数据标签，更改为“41”。这里说的“单击两下”不同于“双击”，中间要有一点时间间隔。

STEP 03 做个截断标记表示中间省略的一截数据。这个截断标记任由你设计，不过要注意美观、专业。这里我仿照麦肯锡的图表绘制了一个类似平行四边形“”的图形，并填充为白色，覆盖在E数据条的上方，代表中间省略了一段数据，如图7-62所示。

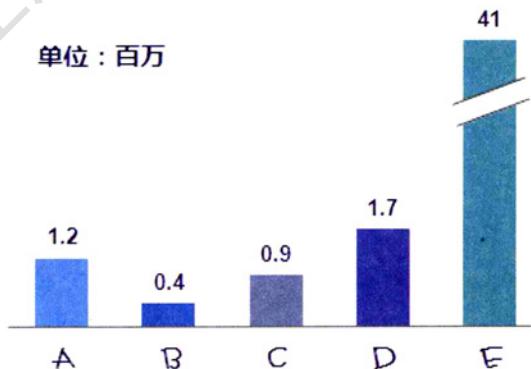


图7-62 截断后的各产品销量图

小白：哈哈，这么简单就解决了超大值问题，下次我也试试！

>> 谁说菜鸟不会数据分析（入门篇）

7.4 本章小结

Mr.林：到现在为止，基本上所有的图表技巧我都倾囊相授了。现在我们来简要回顾复习下今天学过的知识。

小白：好的。

Mr.林：今天主要讲了制作专业图表的三部分内容。

- ★ 图表的基本组成要素，制作图表时的注意事项，并且举出了一些在报纸、杂志、网站上故意夸大事实或者隐瞒事实的图表，让你看到图表背后的真相，并引以为鉴不犯同样的错误。
- ★ 图表最大化数据墨水比原则、图表配色等图表美化技巧。
- ★ 专业图表的制作技巧，如保存图表模板、快速制图、用JWalk Chart Tools添加数据标签等技巧。

小白：嗯，都是非常实用的技巧，以前都不知道图表还有这么多讲究呢。

Mr.林：呵呵。当然我们还讲到要有专业主义的精神，要常怀一颗“童心”，多问几个“为什么？”小白，现在你觉得自己的专业吗？

看小白迟疑没有回答，Mr.林继续说道：其实，我并不是要你马上回答。专业是一种长期的修为，不是一时一刻就能达到的。所以，永远不要满足于现状。“路漫漫其修远兮”，要不断努力，怀揣永不厌倦的好奇心和进取心，才能走向卓越。

第8章

专业的报告——体现你的职场价值

初识数据分析报告

数据分析报告的结构

撰写报告时的注意事项

报告范例

Mr. 林，
这是分析报告。



小白，干得不错！



>> 谁说菜鸟不会数据分析（入门篇）

在得到Mr.林关于数据分析方法及图表制作的真传后，经过两天的数据分析与图表制作，小白终于将数据结果整理出来，并用PowerPoint整理成一份所谓的员工满意度调查报告，不过她这次留了个心眼，先请Mr.林帮忙指导，待检查过关后，再递交给牛董，免得又被牛董K。

小白带着她完成的报告来到Mr.林办公桌前： Mr.林，真不好意思又打扰您。

Mr.林： 说吧，又遇到什么难分析的数据了？

小白连忙解释道： 不是不是，经过您的提点，数据已经分析完了，并且我也写了报告，想先请您指导指导。

Mr.林： 不错嘛。来，让我看看你的报告。

Mr.林接过小白手中的报告，翻了翻，眉头逐渐锁紧。小白也观察到Mr.林的表情变化，不免担心起来。

看完报告Mr.林终于开口说话了： 小白，我建议你先别交给牛董，恐怕你还要加一加班改报告。

小白： 好的，听您的，请问我应该怎么改呢？

Mr.林： 别急，我先跟你说说如何写报告，讲完后你自然就知道要怎么改了。不过这个可不是我说一说你就能完全学会的。要做到真正的应用自如，还需要一段时间的体会和积累。

小白笑嘻嘻地说道： 好好，那我们就快点开始吧。

8.1 初识数据分析报告

Mr.林： 小白，首先我们来了解一下什么是数据分析报告。

8.1.1 数据分析报告是什么

数据分析报告是根据数据分析原理和方法，运用数据来反映、研究和分析某项事物的现状、问题、原因、本质和规律，并得出结论，提出解决办法的一种分析应用文体。

这种文体是决策者认识事物、了解事物、掌握信息、搜集相关信息的主要工具之一，数据分析报告通过对事物数据全方位的科学分析来评估其环境及发展情况，为决策者提供科学、严谨的依据，降低风险。

8.1.2 数据分析报告的写作原则

Mr.林： 接下来，我们来看看撰写数据分析报告需要注意哪些原则，如图8-1所示。



图8-1 撰写数据分析报告应注意的原则

◎ 规范性

数据分析报告中所使用的名词术语一定要规范，标准统一，前后一致，要与业内公认的术语一致。

◎ 重要性

数据分析报告一定要体现数据分析的重点，在各项数据分析中，应该重点选取关键指标，科学专业地进行分析。此外，针对同一类问题，其分析结果也应当按照问题重要性的高低来分级阐述。

◎ 谨慎性

数据分析报告的编制过程一定要谨慎，基础数据必须真实、完整，分析过程必须科学、合理、全面，分析结果要可靠，内容要实事求是。

◎ 创新性

当今科学技术的发展可谓日新月异，许多科学家也都提出各种新的研究模型或者分析方法。数据分析报告需要适时地引入这些内容，一方面可以用实际结果来验证或改进它们，另一方面也可以让更多的人了解到全新的科研成果，使其发扬光大。

总之，一份完整的数据分析报告，应当围绕目标确定范围，遵循一定的前提和原则，系统地反映存在的问题及原因，从而进一步找出解决问题的方法。

8.1.3 数据分析报告的作用

Mr.林：小白，接下来为你介绍数据分析报告的作用，不要写了半天数据分析报告，到头来都不知道自己为什么要写这个报告。

小白：是啊，之前我在写这份报告时也一直在思考，到底为什么要去写它？

Mr.林：作为数据分析的结果，无论其数据搜集过程有多么的科学，数据分析方法多么的高深，数据处理多么的先进，如果不能将它们有效地组织展示出来，并与决策者进行沟通与交流，就无法向决策者提供一个满意的解决方案。

因此，分析报告实质上是一种沟通与交流的形式，主要目的在于将分析结果、可行性建议以及其他有价值的信息传递给管理人员。它需要对数据进行适当的包装，让阅读者能对结果做出正确的理解与判断，并可以根据其做出有针对性、操作性、战略性的决策。

数据分析报告主要有三个方面的作用，即展示分析结果、验证分析质量，以及为决策者提供参考依据，如图8-2所示。



图8-2 数据分析报告的三大作用

◎ 展示分析结果

报告以某一种特定的形式将数据分析结果清晰地展示给决策者，使得他们能够迅速理解、分析、研究问题的基本情况、结论与建议等内容。

◎ 验证分析质量

从某种角度上来讲，分析报告也是对整个数据分析项目的一个总结。通过报告中对数据分析方法的描述、对数据结果的处理与分析等几个方面来检验数据分析的质量，并且让决策者能够感

受到整个数据分析过程是科学并且严谨的。

◎ 提供决策参考

大部分的数据分析报告都是具有时效性的，因此所得到的结论与建议可以作为决策者在决策方面的一个重要参考依据。虽然，大部分决策者（尤其是高层管理人员）没有时间去通篇阅读分析报告，但是在其决策过程中，报告的结论与建议或其他相关章节将会被重点阅读，并根据结果辅助其最终决策。所以，分析报告是决策者二手数据的重要来源之一。

8.1.4 数据分析报告的种类

Mr.林：由于数据分析报告的对象、内容、时间、方法等情况的不同，因而存在着不同形式的报告类型。我们常用的几种数据分析报告有专题分析报告、综合分析报告、日常数据通报等，如图8-3所示。



图8-3 数据分析报告种类

◎ 专题分析报告

专题分析报告是对社会经济现象的某一方面或某一个问题进行专门研究的一种数据分析报告，它的主要作用是为决策者制定某项政策、解决某个问题提供决策参考和依据。

专题分析报告具有以下两个特点。

1. 单一性

专题分析报告不要求反映事物的全貌，主要针对某一方面或某一问题进行分析，如用户流失分析、提升用户消费分析、提升企业利润率分析等。

2. 深入性

由于专题分析报告内容单一，重点突出，因此便于集中精力抓住主要问题进行深入分析。它不仅要对问题进行具体描述，还要对引起问题的原因进行分析，并且提出切实可行的解决办法。这就要求对公司业务的认识要有一定的深度，由感性上升至理性，切忌蜻蜓点水，泛泛而谈。

◎ 综合分析报告

综合分析报告是全面评价一个地区、单位、部门业务或其他方面发展情况的一种数据分析报告。例如世界人口发展报告、全国经济发展报告、某某企业运营分析报告等等。

综合分析报告具有以下两个特点。

1. 全面性

综合分析报告反映的对象，无论一个地区、一个部门还是一个单位，都必须以这个地区、这个部门、这个单位为分析总体，站在全局的高度，反映总体特征，做出总体评价，得出总体认识。在分析总体现象时，必须全面、综合地反映对象各个方面的情况。例如在分析方法论时提到的4P分析法，就是从产品、价格、渠道、促销四个角度进行企业运营分析的。

2. 联系性

综合分析报告要把互相关联的一些现象、问题综合起来进行全面系统的分析。这种综合分析不是对全面资料的简单罗列，而是在系统地分析指标体系的基础上，考察现象之间的内部联系和外部联系。这种联系的重点是比例关系和平衡关系，分析研究它们的发展是否协调，是否适应。因此，从宏观角度反映指标之间关系的数据分析报告一般属于综合分析报告。

◎ 日常数据通报

日常数据通报是以定期数据分析报表为依据，反映计划执行情况，并分析其影响和形成原因的一种数据分析报告。这种数据分析报告一般是按日、周、月、季、年等时间阶段定期进行，所以也叫定期分析报告。

日常数据通报可以是专题性的，也可以是综合性的。这种分析报告的应用十分广泛，各个企业、部门都在使用。

日常数据通报具有以下三个特点。

1. 进度性

由于日常数据通报主要反映计划的执行情况，因此必须把计划执行的进度与时间的进展结合起来分析，观察比较两者是否一致，从而判断计划完成的好坏。为此，需要进行一些必要的计算，通过一些绝对数和相对数指标来突出进度。

2. 规范性

日常数据通报基本上成了数据分析部门的例行报告，定时向决策者提供。所以这种分析报告就形成了比较规范的结构形式。一般包括以下几个基本部分：

- ★ 反映计划执行的基本情况。
- ★ 分析完成或未完成的原因。
- ★ 总结计划执行中的成绩和经验，找出存在的问题。
- ★ 提出措施和建议。

这种分析报告的标题也比较规范，一般变化不大，有时为了保持连续性，标题只变动一下时间，如《XX月XX日业务发展通报》。

3. 时效性

由日常数据通报的性质和任务决定，它是时效性最强的一种分析报告。只有及时提供业务发展过程中的各种信息，才能帮助决策者掌握企业经营的主动权，否则将会丧失良机，贻误工作。

对于大多数公司而言，这些报告主要通过微软Office中的Word、Excel和PowerPoint系列软件来表现。这三种软件各有优劣势，具体内容如图8-4所示。

项目	Word	Excel	PowerPoint
优势	<ul style="list-style-type: none">● 易于排版● 可打印装订成册	<ul style="list-style-type: none">● 可含有动态图表● 结果可实时更新● 交互性更强	<ul style="list-style-type: none">● 可加入丰富的元素● 适合演示汇报● 增强展示效果
劣势	<ul style="list-style-type: none">● 缺乏交互性● 不适合演示汇报	<ul style="list-style-type: none">● 不适合演示汇报	<ul style="list-style-type: none">● 不适合大篇文字
适用范围	<ul style="list-style-type: none">● 综合分析报告● 专题分析报告● 日常数据通报	<ul style="list-style-type: none">● 日常数据通报	<ul style="list-style-type: none">● 综合分析报告● 专题分析报告

图8-4 Office各软件制作报告的优劣势对比

8.2 数据分析报告的结构

小白：Mr.林，现在我已经明白了数据分析报告的三大作用和四项基本原则，以后我在写报告时，就会更有目的地来组织行文。刚才看到你在阅读我的报告时，似乎不太满意。是不是我写得不好呢？

Mr.林：从内容上来说，并不是不好，但是整体来看，你的报告缺少结构，导致内容太散，不利于阅读。

小白：也就是说在撰写数据分析报告时，需要遵从某种特定的结构？

Mr.林：数据分析报告确实有特定的结构，但是这种结构并非一成不变，不同的分析师、不同的老板、不同的客户、不同性质的数据分析，其最后的报告可能会有不同的结构。最经典的报告结构还是“总一分一总”结构，它主要包括：开篇、正文和结尾三大部分，如图8-5所示。

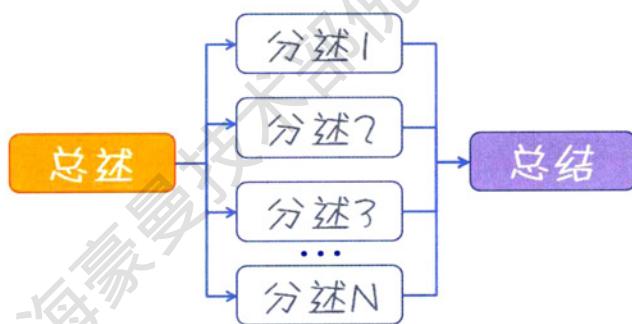


图8-5 “总一分一总”报告结构

在数据分析报告结构中，“总一分一总”结构的开篇部分包括标题页、目录和前言（主要包括分析背景、目的与思路）；正文部分主要包括具体分析过程与结果；结尾部分包括结论、建议及附录。下面将对这几部分进行具体介绍。

8.2.1 标题页

Mr.林：顾名思义，标题页需要写明报告的题目，题目要精简干练，根据版面的要求在一两行内完成。标题是一种语言艺术，好的标题不仅可以表现数据分析的主题，而且能够激发读者的阅读兴趣，因此需要重视标题的制作，以增强其艺术性的表现力。图8-6列出了常用的标题类型。

◎ 标题常用的类型

1. 解释基本观点

这类标题往往用观点句来表示，点明数据分析报告的基本观点，如《不可忽视高价值客户的保有》《语音业务是公司发展的重要支柱》等。

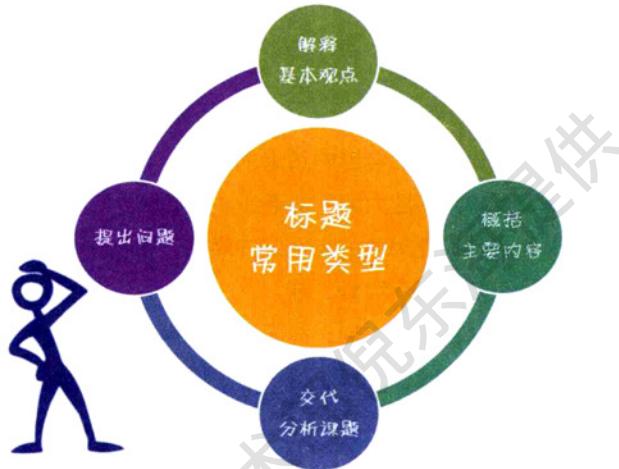


图8-6 标题的常用类型

2. 概括主要内容

这类标题重在叙述数据反映的基本事实，概括分析报告的主要内容，让读者能抓住全文的中心，如《我公司销售额比去年增长30%》《2010年公司业务运营情况良好》等。

3. 交代分析主题

这类标题反映分析的对象、范围、时间、内容等情况，并不点明分析师的看法和主张，如《发展公司业务的途径》《2010年运营分析》《2010年部门业务对比分析》等。

4. 提出问题

这类标题以设问的方式提出报告所要分析的问题，引起读者的注意和思考，如《客户流失到哪里去了》《公司收入下降的关键何在》《1500万元的利润是怎样获得的》。

◎ 标题的制作要求

1. 直接

数据分析报告是一种应用性较强的文体，它直接用来为决策者的决策和管理服务，所以标题必须用毫不含糊的语言，直截了当、开门见山地表达基本观点，让读者一看标题就能明白数据分析报告的基本精神，加快对报告内容的理解。

2. 确切

标题的撰写要做到文题相符，宽窄适度，恰如其分地表现分析报告的内容和对象的特点。

3. 简洁

标题要直接反映出数据分析报告的主要内容和基本精神，就必须具有高度的概括性，用较少的文字集中、准确、简洁地进行表述。

◎ 标题的艺术性

数据分析报告的标题大多容易雷同，如《关于XXX的调查分析》《对XXX的分析》等，这类模式化的标题使用太泛，千篇一律，必然会影响读者的阅读兴趣。因此，标题的撰写，除了要符合直接、确切、简洁三点基本要求，还应力求新鲜活泼、独具特色，增强艺术性。

要使标题具有艺术性，就要抓住对象的特征展开联想，适当运用修辞手法给予突出和强调，如《我的市场我做主》《我和客户有个约会》等。

有时，报告的作者也要在题目下方出现，或者在报告中要给出所在部门的名称。为了将来方便参考，完成报告的日期也应当注明，这样能够体现出报告的时效性，如图8-7所示。

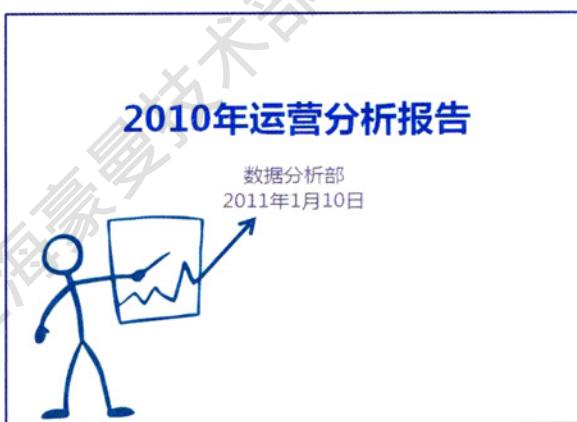


图8-7 标题页示例

8.2.2 目录

Mr.林：目录可以帮助读者快捷方便地找到所需的内容，因此，要在目录中列出报告主要章节的名称。如果是在Word中撰写报告，在章节名称后面还要加上对应的页码，对于比较重要的二级目录，也可以将其列出来，如图8-8所示。所以，从另外一个角度说，目录也就相当于数据分析大纲，它可以体现出报告的分析思路。但是目录也不要太过详细，因为这样阅读起来让人觉得

冗长并且耗时。

此外，通常公司或企业的高层管理人员没有时间阅读完整的报告，他们仅对其中一些以图表展示的分析结论会有兴趣，因此，当书面报告中有大量图表时，可以考虑将各章图表单独制作成目录，以便日后更有效地使用。



图8-8 目录示例

8.2.3 前言

Mr.林: 前言是分析报告的一个重要组成部分，主要包括分析背景、目的及思路三方面，即：

- ★ 为什么要开展此次分析？有何意义？
- ★ 通过此次分析要解决什么问题？达到何种目的？
- ★ 如何开展此次分析？主要通过哪几方面开展？

所以，前言的写作一定要经过深思熟虑，前言内容是否正确，对最终报告是否能解决业务问题、能否给决策者决策提供有效依据起决定性作用。

◎ 分析背景

对数据分析背景进行说明主要是为了让报告阅读者对整个分析研究的背景有所了解，主要阐述此项分析的主要原因、分析的意义，以及其他相关信息，如行业发展趋势等内容，如图8-9所示。

◎ 分析目的

数据分析报告中陈述分析目的主要是为了让报告的阅读者了解开展此次分析能带来何种效果，可以解决什么问题。有时将研究背景和目的意义合二为一，如图8-9所示。

>> 谁说菜鸟不会数据分析（入门篇）

图8-9中所述的分析目的就很清楚：通过分析企业市场环境的变化，及时回答市场拓展工作中需要研究解决的各种问题，把握市场机会，借以指导并推动市场拓展工作。所以分析报告需要有一个明确的分析目的，目的越明确，针对性就越强，越能及时解决问题，就越有指导意义；反之，数据分析报告就没有生命力。

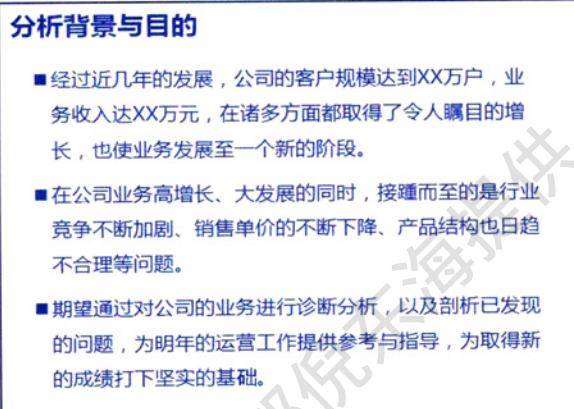


图8-9 分析背景与目的示例

① 分析思路

分析思路用来指导数据分析师如何进行一个完整的数据分析，即确定需要分析的内容或指标。这是分析方法论中的重点，也是很多人常常感到困扰的问题。只有在营销、管理理论的指导下，才能确保数据分析维度的完整性，分析结果的有效性及正确性，如图8-10所示。

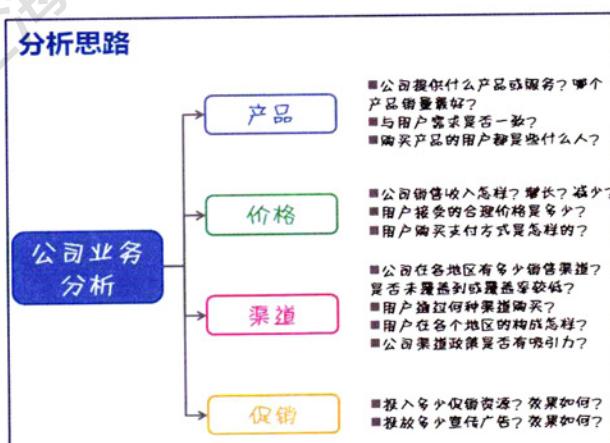


图8-10 报告分析思路示例

在报告的分析思路中，有时会使用到高级的数据分析方法，如回归、聚类等，此时就需要在分析思路中对使用到的高级分析方法略加说明，不需要涉及太过专业的描述，只需把分析原理进行言简意赅的阐述，让报告阅读者对此有所了解。

8.2.4 正文

Mr.林：正文是数据分析报告的核心部分，它将系统全面地表述数据分析的过程与结果。

撰写报告正文时，根据之前分析思路中确定的每项分析内容，利用各种数据分析方法，一步步地展开分析，通过图表及文字相结合的方式，形成报告正文，方便阅读者理解，如图8-11所示。



图8-11 报告正文示例

正文通过展开论题，对论点进行分析论证，表达报告撰写者的见解和研究成果的核心部分，因此正文占分析报告的绝大部分篇幅。一篇报告只有想法和主张是不行的，必须经过科学严密的论证，才能确认观点的合理性和真实性，才能使别人信服。因此，报告主题部分的论证是极为重要的。

报告正文具有以下几个特点：

- ★ 是报告最长的主体部分。
- ★ 包含所有数据分析事实和观点。
- ★ 通过数据图表和相关的文字结合分析。
- ★ 正文各部分具有逻辑关系。

8.2.5 结论与建议

Mr.林： 报告的结尾是对整个报告的综合与总结、深化与提高，是得出结论、提出建议、解决矛盾的关键所在，它起着画龙点睛的作用，是整篇分析报告的总结。好的结尾可以帮助读者加深认识，明确题旨，引起思考，如图8-12所示。

结论是以数据分析结果为依据得出的分析结果，通常以综述性文字来说明。它不是分析结果的简单重复，而是结合公司实际业务，经过综合分析、逻辑推理形成的总体论点。结论是去粗取精、由表及里而抽象出的共同的、本质的规律，它与正文紧密衔接，与前言相呼应，使分析报告首尾呼应。结论应该措词严谨、准确、鲜明。

建议是根据数据分析结论对企业或业务等所面临的问题而提出的改进方法，建议主要关注在保持优势及改进劣势等方面。因为分析人员所给出的建议主要是基于数据分析结果而得到的，会存在局限性，因此必须结合公司的具体业务才能得出切实可行的建议。

结论与建议

- 各产品销量结构相对合理，销售收入发展良好，呈上升趋势
- 华南与华东两地区未来市场拓展潜力大，市场拓展空间较大
- XXX促销方式更能促进其销量增长，但XXX广告投放方式效果不甚理想
-

建议公司对以下发展策略进行商榷：

- 主推XXX产品，并将XXX产品价格进行调整
- 集中公司资源，大力拓展华南、华东两地区市场
- 重新考量XXX的广告投放方式
-

图8-12 结论与建议示例

8.2.6 附录

小白：以上应该就是数据分析报告的基本结构了吧？

Mr.林：没错，但是还有一个部分不容忽视，那就是报告的附录。它是数据分析报告的一个重要组成部分。一般来说，附录提供正文中涉及而未予阐述的有关资料，有时也含有正文中提

及的资料，从而向读者提供一条深入数据分析报告的途径。它主要包括报告中涉及的专业名词解释、计算方法、重要原始数据、地图等内容。每个内容都需要编号，以备查询，如图8-13所示。

当然并不要求每篇报告都有附录，附录是数据分析报告的补充，并不是必需的，应该根据各自的情况再决定是否需要在报告结尾处添加附录。

附录A—数据透视表术语	
术语	内容
轴	数据透视表中的一个维度，例如行、列或页
数据源	创建数据透视表的数据表、数据库等
字段	数据信息的种类，相当于数据表中的列
字段标题	描述字段内容的标志，可通过拖动字段标题对数据透视表进行透视分析
透视	通过改变一个或多个字段的位置来重新安排数据透视表
汇总函数	Excel用来计算表格中数据的值的函数。数值和文本的默认汇总函数分别是求和与计数
刷新	重新计算数据透视表，以反映目前数据源状态

图8-13 报告附录示例

8.3 撰写报告时的注意事项

小白：这么说，在撰写报告之前，需要先构建出清晰合理的报告结构，之后再详尽地进行分析？

Mr.林：不，数据分析报告的内容最好详略得当，一份报告的价值，并不取决于其篇幅的长短，而在于其内容是否丰富，结构是否清晰，是否有效反映业务真相，提出的建议是否可行。因此，在撰写报告时，有以下几个问题需要特别注意。

◎ 结构合理，逻辑清晰

一份合格而优秀的报告，应该有非常明确、清晰的构架，呈现简洁、清晰的数据分析结果。如果报告的分析过程逻辑混乱、各章节界限不清晰、没有按照业务逻辑或内在联系有条理地论证，等等，那么报告阅读者就无法从中得出有用的决策依据。因此，报告的结构是否合理、逻辑条理是否清晰是决定此份报告成败的关键因素。

◎ 实事求是，反映真相

数据分析报告最重要的就是必须具备真实性。真实性的含义不仅包括基于分析得到的结论是事实，而且包括数据在内，不允许有虚假和伪造的现象存在。此外，对事实的分析和说明也必须遵从科学、实事求是的做法，符合客观事物的本来面目。一定要保持中立的态度，不要加入自己的主观意见。

◎ 用词准确，避免含糊

报告中的用词必须准确，即如实、恰如其分地反映客观情况，在分析报告中最好的做法就是尽量用数据说话，避免使用“大约”“估计”“更多（或更少）”“超过50%”等模糊的字眼。报告必须明确告知阅读者：什么情况合理（或好），什么情况不合理（或坏）。

◎ 篇幅适宜，简洁有效

报告的价值主要在于提供给决策者所需要的信息，并且这些信息能够解决他们的问题。换句话说，就是报告要满足决策者的需求。如果一份关于消费者满意度的分析研究报告中没有回答满意度的驱动因素，没有关于满意度指标的评估等有价值的内容，报告写得再长也没有很大的参考意义。

◎ 结合业务，分析合理

一份优秀的报告不能仅基于数据而分析问题，或简单地看图说话，必须紧密结合公司的具体业务才能得出可实行、可操作的建议，否则将是纸上谈兵，脱离实际。因此，分析结果需要与分析目的紧密结合起来，切忌远离目标的结论和不现实的建议。当然，这也要求数据分析人员对业务有一定的了解，如果对业务不了解或不熟悉，可请业务部门的同事一起参与讨论分析，以得出正确的结论及提出合理的建议。

8.4 报告范例

Mr.林：刚才说了那么多，估计你已经听得头晕了吧！

小白茫然地看着Mr.林：是有一些晕了，不过我已经记好笔记了，会好好复习的。

Mr.林：呵呵，没关系，我来给你看一份数据分析报告的范例，这样会比较有感觉。

小白立刻兴奋了起来，连说：好啊好啊。

Mr.林拿出一份报告，边翻边给小白讲解起来：这个是报告的标题页，如图8-14所示，你应该不陌生。

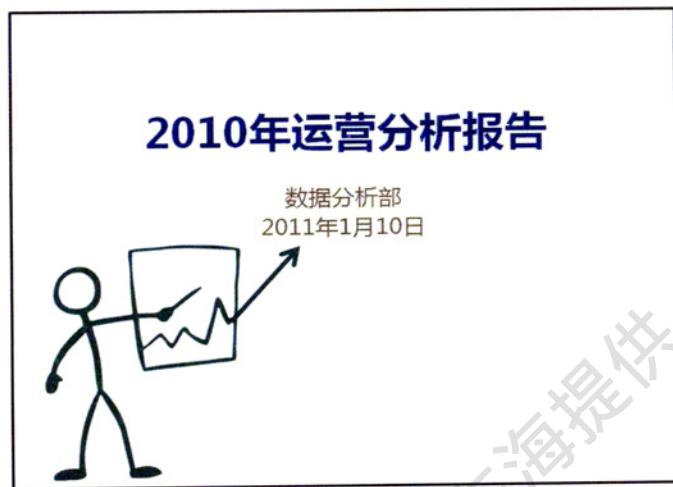


图8-14 标题页

Mr.林：接下来是报告的目录，如图8-15所示，里面列出了主要章节的名称，这样就能对整份报告的分析思路一目了然。



图8-15 目录

Mr.林：下一页是报告的前言，如图8-16所示，包括了分析背景与目的。

分析背景与目的

- 经过近几年的发展，公司在客户规模达XX万户、业务收入达XX万元等诸多方面都取得了令人瞩目的增长，也是业务发展来到了一个新的阶段。
- 在公司业务高增长、大发展的同时，接踵而至的是行业竞争不断加剧、销售单价的不断下降、产品结构也日趋不合理等问题。
- 期望通过对公司业务进行诊断分析，以及存在的问题进行剖析，为明年的运营工作提供参考与指导，为取得新的成绩打下坚实的基础。

图8-16 前言

Mr.林：之后就是重头戏了——报告的分析思路，如图8-17所示。这里清晰地展现了数据报告的分析结构，当然也方便读者就自己所关心的部分重点阅读。

分析思路

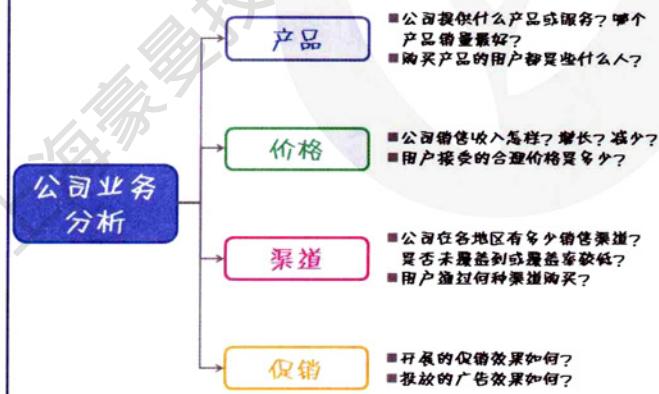


图8-17 分析思路

Mr.林：下一页就开始报告的正文了。首先是产品分析，如图8-18所示，这一页主要介绍公司产品的销量构成情况。

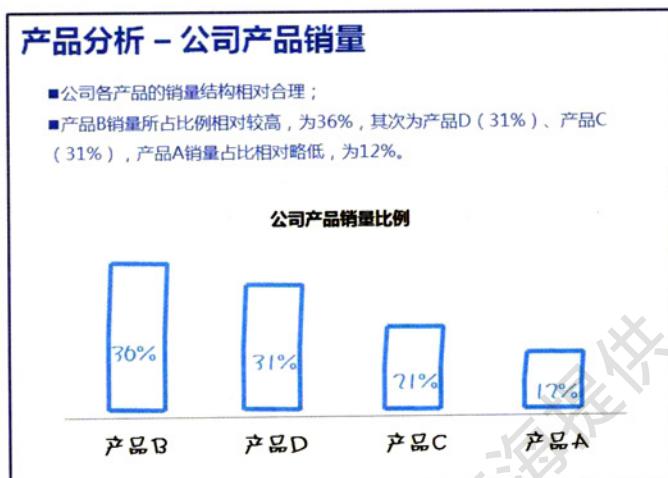


图8-18 正文——产品分析(1)

Mr.林：下面的这一页通过百分比堆积柱形图回答了产品分析的第二个问题，如图8-19所示，清晰地表现出用户在每种产品上所占的比例。

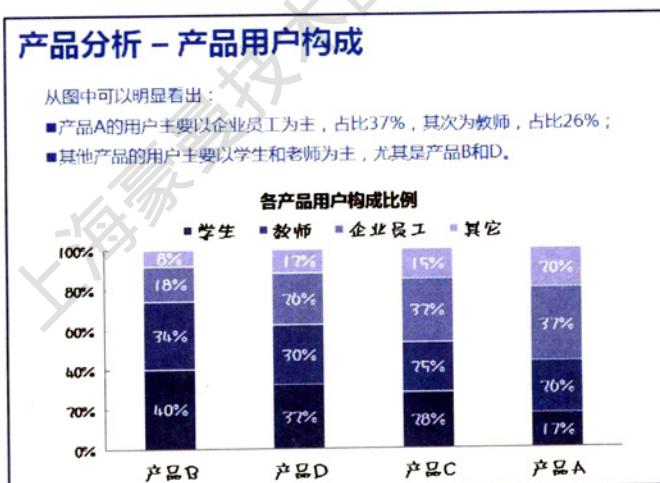


图8-19 正文——产品分析(2)

Mr.林：接下来是第二个分析主题——价格分析。首先通过堆积柱形图显示出公司产品的月度平均销售收入走势，如图8-20所示，通过这个图，读者能够很快了解产品在2010年的销售情况。

价格分析 – 产品销售收入

- 2010年公司产品平均每月销售收入发展良好，呈增长趋势，使得2010年整体运营收入达XX万元，环比增长XX%。
- 公司中各产品收入相对均衡，运营状况良好。

2010年公司各产品平均每月销售收入

单位：万元 ■ 产品B ■ 产品D ■ 产品C ■ 产品A

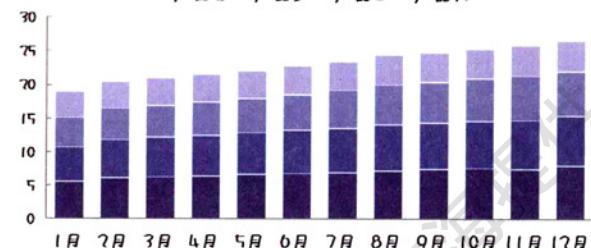


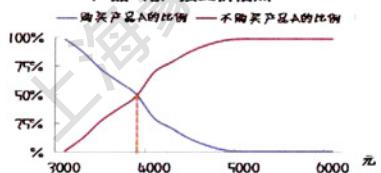
图8-20 正文——价格分析(1)

Mr.林：价格分析的第二点是产品最优价格点分析。这一份报告中采用折线图展现，如图8-21所示，能够明显看出两条曲线的交叉点即产品的最优价格点。

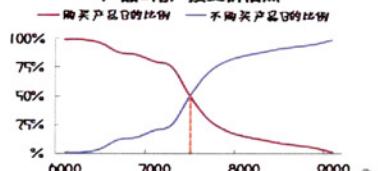
价格分析 – 价格接受点

- 每个产品都有各自的用户接受价格点，两条线的交叉点即为价格最优点，意味着.....
- 四个产品的用户接受价格点分别为XXX, XXX, XXX, XXX。

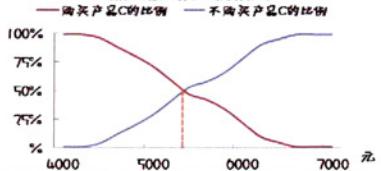
产品A用户接受价格点



产品B用户接受价格点



产品C用户接受价格点



产品D用户接受价格点

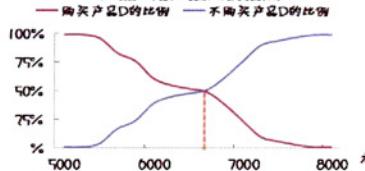


图8-21 正文——价格分析(2)

Mr.林：价格分析完毕后，就进入第三个主题——渠道分析。首先用百分比堆积柱形图表现各销售渠道在各地区的分布情况，如图8-22所示。这里能够直观地看出产品在特定地区的主要

销售渠道是什么。

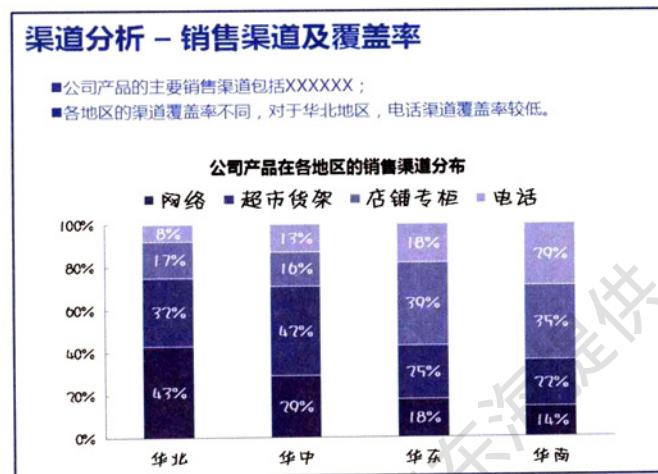


图8-22 正文——渠道分析(1)

Mr.林：对于产品市场覆盖率的分析，这里用条形图展现，如图8-23所示。其中蓝色的部分表示市场覆盖率，红色的部分则表示用户分布。红蓝条的此消彼长能够说明产品在市场上的覆盖情况，以及未来的市场发展潜力。

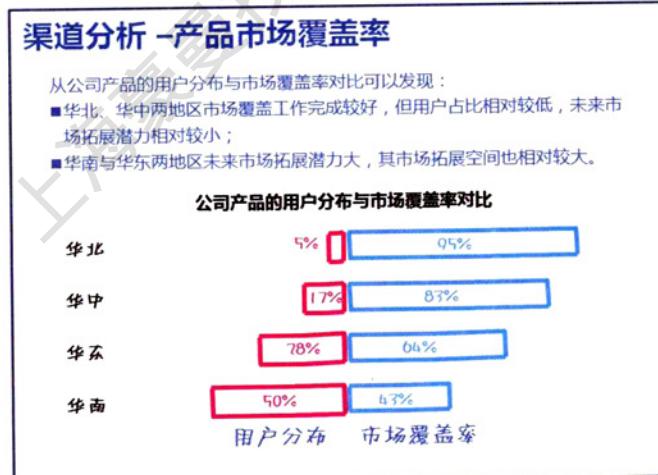


图8-23 正文——渠道分析(2)

Mr.林：接下来，再用百分比堆积柱形图显示公司产品在每种渠道下的销售比例，如图8-24所示。建议类似这样的分析可以先按某一产品的比例高低进行排列，做出的图就能够让读者一目

了然。

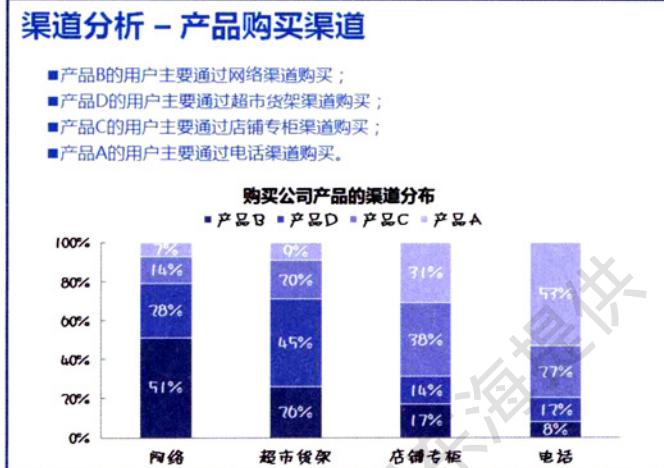


图8-24 正文——渠道分析(3)

Mr.林：报告中的最后一个分析是促销分析。首先对产品的促销活动进行分析，如图8-25所示。通过柱形图，直观形象地描述出产品A在2010年通过促销活动使销售收入有所增长，以此反映该促销活动对产品销售起到的积极作用。

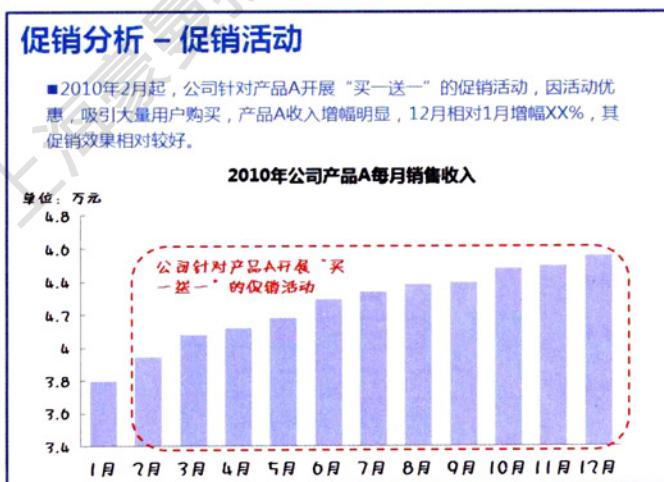


图8-25 正文——促销分析(1)

Mr.林：接着是广告投放分析，如图8-26所示。用柱形图显示出产品C在投放广告后销售收入的涨幅。但是由结果可以发现该广告并未明显促进产品C的销售。

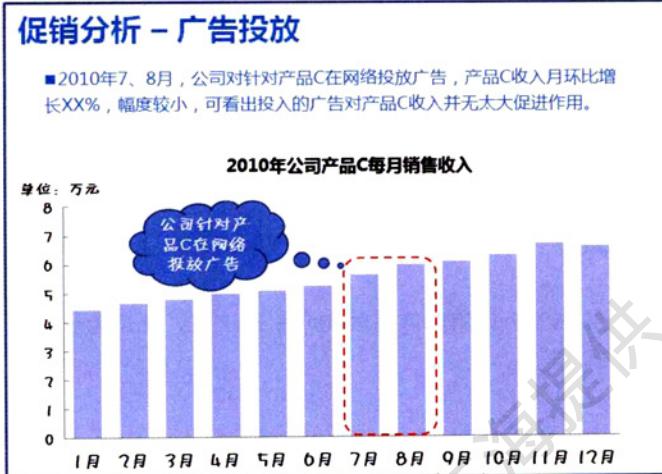


图8-26 报告正文分析——促销分析(2)

Mr.林：正文到此告一段落，最后是根据前面的分析得出的结论和建议，如图8-27所示。

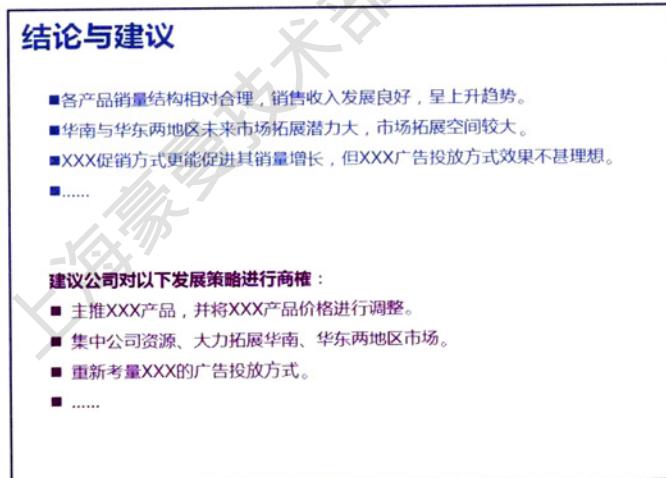


图8-27 结论和建议

Mr.林：好啦，看到这里，我想你对数据分析报告的整个框架和结构已经了然于心了吧。

小白底气十足地点点头道：嗯，是啊，现在看到具体的报告，已经基本能够理解前面所讲的内容啦。

Mr.林：要记住，写数据分析报告的能力不是一蹴而就的，需要平时慢慢地积累。这里看的报告范例只是例子，并不一定是适合每个分析。实际工作时还是需要具体问题具体分析，根据自

己所在公司的实际业务情况进行调整，灵活运用，切忌生搬硬套。只有深刻理解公司业务，才能较好地完成基于业务的数据分析，否则就是纸上谈兵了。

另外，对数据的敏锐洞察力，以及对报告内容的组织能力也是需要不断提高的。相信你在不久的将来就会成为一名优秀的数据分析师了。

8.5 本章小结

Mr.林：小白，关于数据分析报告的内容就先介绍到这里，还是老规矩，我们一起简要回顾一下数据分析报告中的知识要点。

小白：好的。

Mr.林：今天介绍的数据分析报告方面的知识主要有：

- ★ 数据分析报告的三大作用，即展示分析结果、验证分析质量，以及为决策者提供参考依据，还有规范性、重要性、谨慎性以及创新性四项基本原则。
- ★ 数据分析报告的三种常用报告类型，即专题分析报告、综合分析报告、日常数据通报，以及各自的特点。
- ★ 数据分析报告的标题、目录、前言、正文、结论与建议、附录六大组成部分及特点。
- ★ 撰写数据分析报告的五大注意事项：第一，合理的结构以及清晰的逻辑；第二，实事求是，反映真相；第三，用词准确，避免含糊；第四，篇幅适宜，简洁有效；第五，结合具体业务进行合理的分析。

小白补充道：我们还欣赏了一份完整的数据分析报告范例。

Mr.林：哈哈，小白，现在你已经掌握了数据分析报告的精华，相信你很快就可以出师了。现在只需要将这些原则性的内容应用到工作中，并在实际操作中不断积累经验，你的数据分析能力就会与日俱增，要不了多久，就可以挂上优秀数据分析师的头衔了。走，为了庆贺你即将出师，上次你请我喝早茶，一会儿我请你吃饭。

小白：真的呀！太好了，Mr.林，您好帅喔！

Mr.林：哈哈，走吧！

写在后面的Q/A

Q1：本书为何不继续介绍一些高级的分析方法？

A：这是一本写给普通大众的数据分析书，根据二八原则，日常学习、工作中80%以上的数据分析甚至100%数据分析只用到数据分析知识中20%的内容，所以本书就专门介绍这20%的数据分析知识，并且这部分知识是人们容易忽视的，比如对比分析，有多少分析师能明确并使用呢？

在相应的章节，我们也提到了高级数据分析方法，如果读者有需要，可参考其他专业书籍。

Q2：本书为何不多介绍一些Excel在数据分析方面的知识？

A：Excel在数据分析方面的知识很多，数据分析这门技术、知识是写不完的，数据分析方面的技术写上10本都写不完。有兴趣的读者可以专门阅读Excel书籍。本书旨在起抛砖引玉，带领各位读者入门数据分析。俗话说：师傅领进门，修行在个人，所以，剩下的事就要靠读者自己多看、多想、多练，从而提高自己的数据分析水平。

Q3：本书是关于数据分析的书，但为何没有介绍统计函数？

A：前言中已经介绍了本书的定位，本书基于大多数人日常工作使用的办公软件Excel 2010版介绍数据分析方法，并且尽量避免涉及较为复杂的方法（如统计函数、宏代码编写等），尽可能使用简单明了的方式解决日常遇见的问题。而采用数据透视表已经基本能满足大部分读者日常的数据分析工作了，所以本书重点介绍数据透视表。掌握了数据透视表，相信会令你的工作效率大大提高。

Q4：本书主人公为何是小白？

A：本书通过小白跟师傅学习数据分析的方式，介绍数据分析的各个步骤及方法，并通过把数据分析步骤与生活中的烹饪、制衣等流程结合起来，加深读者对知识点的理解，避免了通常教科书、专业书籍易出现的枯燥晦涩等问题，增加了学习的趣味性。

另外，小白问的问题也是数据分析初学者常遇到并且需要深刻理解的问题。

Q5：本书为何有两种类型的图表风格？

A：本书有一种规范专业的图表风格，直接采用的是Excel标准图表。除此之外，还有一种俏皮化的图表风格，采用的是漫画字体和手绘画的表达方式。前者意在介绍图表规范时，让读者有专业图表的概念，可以以此为模板作图；后者是为了增强漫画效果，让图文生动有趣，使读者更容易理解和记忆。

附录A 网络学习资源

资源类型	名称	网络地址	微博地址
	199IT-互联网数据中心	http://www.199it.com/	http://weibo.com/199it
	DataV	http://datavlab.org/	http://weibo.com/tbdatav
	Excel技巧网	http://www.exceltip.net/	http://weibo.com/exceltipnet
	Matlab技术论坛	http://www.matlabsky.com/	http://weibo.com/hgsz2003
	Matlab中文论坛	http://www.ilovematlab.cn/	
	SAS爱好者	http://sasor.feoh.net/	
	SAS中文论坛	http://www.mysas.net/forum/	http://weibo.com/mysas
	SAS资源资讯	http://saslist.net/	http://weibo.com/saslist
	SPSS数据分析论坛	http://www.spsschina.cn/	
数据分析	互联网分析沙龙	http://www.techxue.com/	http://weibo.com/tianxuefeng
网站	蓝鲸的网站分析笔记	http://bluewhale.cc/	http://weibo.com/askcliff
	人大经济论坛	http://bbs.pinggu.org/	http://weibo.com/pingguorg
	社会网络与数据挖掘	http://www.snsdatamining.com/	http://weibo.com/dmonsns
	数据仓库之路	http://www.dwway.com/	http://weibo.com/dwway
	数据科学 - 数据极客	http://www.itisbi.net/	http://weibo.com/itisbi
	数据堂	http://www.datatang.com/	http://weibo.com/datatang
	数据挖掘论坛	http://www.businessanalysis.cn/	http://weibo.com/1650042843
	数据挖掘研究	http://www.chinakdd.com/	http://weibo.com/chinakdd
	数据熊猫论坛	http://www.datapanda.net/	http://weibo.com/lande
	统计之都	http://cos.name/	http://weibo.com/cosname

续表

资源类型	名称	网络地址	微博地址
	网站分析社区	http://www.wasns.com/	http://weibo.com/wasns
	网站分析星期三	http://www.webanalyticswednesday.cn/	http://weibo.com/beijingwaw
	网站分析与电子商务	http://www.wachina.net/	http://weibo.com/davezheng
	网站分析在中国	http://www.chinawebanalytics.cn/	http://weibo.com/webanalytics
	网站数据分析	http://webdataanalysis.net/	http://weibo.com/joeghwu
	演绎生动Excel图表	http://visualvivid.com/	http://weibo.com/vividchart
	中国商业智能网	http://www.chinabi.net/	http://weibo.com/1593239290
	中国统计网	http://www.itongji.cn/	http://weibo.com/itongji
	Excel大全	http://blog.sina.com.cn/excelchrisfang	http://weibo.com/excelbible
	SAS_Miner	http://blog.sina.com.cn/racoon	http://weibo.com/racoon2002
	车品觉	http://blog.sina.com.cn/herbertchia	http://weibo.com/herbertchia
	刘思喆	http://www.bjt.name/	http://weibo.com/sunbjt
	刘万祥ExcelPro	http://excelpro.blog.sohu.com/	http://weibo.com/excelpro
	庖丁的小刀	http://blog.sina.com.cn/u/2125736794	http://weibo.com/u/2125736794
	沈浩老师	http://shenhaolaoshi.blog.sohu.com/	http://weibo.com/shenhaolaoshi
	数据分析与研究	http://blog.sina.com.cn/spssjiaoliu	
	数据分析那些事		http://weibo.com/datafenxi
数据分析	数据海洋	http://blog.sina.com.cn/wyx11s	http://weibo.com/wyx11s725
博客	数据化管理	http://chemyhuang.blog.163.com/	http://weibo.com/chemyhuang
	数据挖掘与数据分析	http://spss-market.r.blog.163.com/	http://weibo.com/betterdata
	数据小兵	http://hi.baidu.com/new/datasoldier	http://weibo.com/shujuxiaobing
	数据小宇军	http://blog.sina.com.cn/u/2687675833	http://weibo.com/u/2687675833
	数据元素	http://blog.sina.com.cn/greatdoor	http://weibo.com/luckydata
	天岸网站分析	http://we.univcite.com/scyan/	
	网站分析臭皮匠	http://qing.weibo.com/wabook	http://weibo.com/wabook
	小蚊子乐园	http://blog.sina.com.cn/xiaowenzi22	http://weibo.com/xiaowenzi22
	张磊IDMER	http://idmer.blog.sohu.com/	http://weibo.com/idmer
	张文彤老师	http://blog.sina.com.cn/medstatstar	http://weibo.com/wintone

>> 谁说菜鸟不会数据分析（入门篇）

续表

资源类型	名称	网络地址	微博地址
	郑来轶	http://blog.sina.com.cn/zhenglaiyi	http://weibo.com/analyst
	NordripPT	http://www.nordridesign.cn/	http://weibo.com/nordri
	PPT精选	http://www.80ppt.com/	http://weibo.com/pptlove
PPT网站	扑奔PPT	http://www.pooban.com/	http://weibo.com/pooban
	我爱PPT	http://iloveppt.org/	http://weibo.com/iloveppt
	无忧PPT	http://www.51ppt.com.cn/	http://weibo.com/51ppt
	Lonely Fish	http://lonelyfish1920.blog.163.com/	http://weibo.com/lonelyfish1920
	PPTworld	http://hi.baidu.com/pptworld/	
	PPT马建强	http://blog.sina.com.cn/majq	http://weibo.com/majq
	般若黑洞	http://hi.baidu.com/paratop/	http://weibo.com/paralife
	大乘起信	http://pptdesign.blogbus.com	http://weibo.com/edd lou
PPT博客	枫桥PPT创意坊	http://blog.sina.com.cn/pptcreate	http://weibo.com/pptcreate
	李旸旸	http://blog.sina.com.cn/liyangyang1126	http://weibo.com/u/1708082727
	且行资源	http://www.qiexing.com	http://weibo.com/qiexing
	秋叶	http://www.70man.com/	http://weibo.com/qiuyept
	孙小小	http://xiaoxiaosun1978.blog.sohu.com/	http://weibo.com/sunxiaoxiao1loveppt
	小秦策略PPT	http://blog.sina.com.cn/xiaoqinstrategy	http://weibo.com/2014252331
	蝇子	http://quick-learning.blogbus.com/	http://weibo.com/1680857295