

以 Excel 为决策工具的 商务与经济统计

Business and Economic Statistics
Using Microsoft Excel

肯·布莱克 (Ken Black)

休斯顿大学明湖分校

(美)

戴维 L. 埃尔德雷奇 (David L. Eldredge)

默里州立大学

著

张久琴 张玉梅 杨琳 译



附光盘



机械工业出版社
China Machine Press

经济教材译丛

以 Excel 为决策工具的 商务与经济统计

Business and Economic Statistics Using Microsoft Excel

肯·布莱克 (Ken Black)

休斯顿大学明湖分校

(美)

戴维 L. 埃尔德雷奇 (David L. Eldredge)

默里州立大学

著

张久琴 张玉梅 杨琳 译

本书主要介绍了关于应用商务统计的基础知识。书中的概念都是通过真实商业、经济数据和案例进行阐述的。本书的主要目的是对应用统计方法给出一个简洁明确的概念,另一个贯穿全书的目的是要把 Microsoft Excel 如何用于分析商务与经济数据的问题解释清楚。应用性强是本书的最大特点。书中将 Excel 的讲解与统计内容进行了完美结合,使 Excel 更易学易懂。

全书语言浅显易懂,叙述严谨,既可以用做本科生、研究生和 MBA 的入门教材,也可作为从事工商管理与经济研究的各类人员的参考书。

Ken Black, David L. Eldredge. Business and Economic Statistics Using Microsoft Excel.

EISBN: 0-324-01726-X

Copyright©2002 by South-Western, a division of Thomson Learning.

Original language published by Thomson Learning.

All Rights reserved.

CMP is authorized by Thomson Learning to publish and distribute exclusively this simplified Chinese edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan). Unauthorized export of this edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

本书原版由汤姆森学习出版集团出版。本书中文简体字翻译版由汤姆森学习出版集团授权机械工业出版社独家出版发行。此版本仅限在中华人民共和国境内(不包括中国香港、澳门特别行政区及中国台湾地区)销售。未经授权的本书出口将被视为违反版权法的行为。未经出版者预先书面许可,不得以任何方式复制或发行本书的任何部分。

版权所有,侵权必究。

981-243-993-5

本书版权登记号: 图字: 01-2002-1365

图书在版编目(CIP)数据

以 Excel 为决策工具的商务与经济统计/(美)布莱克(Black, K.), (美)埃尔德雷奇(Eldredge, D. L.)
著;张久琴等译. - 北京:机械工业出版社, 2003.9

(经济教材译丛)

书名原文: Business and Economic Statistics Using Microsoft Excel

ISBN 7-111-12732-3

I. 以… II. ①布…②埃…③张… III. 电子表格系统, Excel - 应用 - 经济统计 IV. F222

中国版本图书馆 CIP 数据核字 (2003) 第 070069 号

机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码 100037)

责任编辑: 魏 杰 版式设计: 刘永青

北京瑞德印刷有限公司印刷·新华书店北京发行所发行

2003 年 9 月第 1 版第 1 次印刷

889mm×1194mm 1/16·36.25 印张

定价: 75.00 元

凡购本书, 如有倒页、脱页、缺页, 由本社发行部调换

本社购书热线电话: (010) 68326294

统计学是一门方法论科学,研究随机现象的数量规律性。它是近代数学的重要组成部分,也是近代经济理论应用与研究的重要数学工具,在许多学科领域都有着重要应用。将统计学的概念与方法用来处理商务与经济中的各种问题,就产生了商务与经济统计学,它是解决商务与经济中各种问题的有力工具之一。

正是在这种思想的指导下,译者开始对国外各著名大学所使用的相关教材进行分析比较,最终选择了这本得到多位专家推荐的、由休斯顿大学明湖分校的肯·布莱克和默里州立大学的戴维 L. 埃尔德雷奇两位教授合著的《以 Excel 为决策工具的商务与经济统计》。我们认为它是一本很好的教材,于是把它翻译成中文介绍给我国读者。

本书内容全面,不仅包括图形统计、描述统计、概率、离散和连续分布、样本分布、假设检验和估计等一般内容,还单列出几章专门讲解了多元回归分析和时间序列预测,因为商务领域的决策者经常使用回归和预测工具。最后一章是统计质量的控制。

本书语言浅显易懂,叙述严谨,对商务统计学中的各种基本概念及关键论题进行了深入浅出的分析与介绍,既可以用作大学本科生和 MBA 的入门教材,也可作为从事工商管理与经济研究的各类人员的参考文献。

应用性强是本书的最大特点。书中将 Excel 的讲解与统计内容进行了完美结合,使 Excel 更易学易懂,同时加入 FAST \ STAT 附加软件,以对 Excel 的统计功能进行充实和提高。

本书另一大特点是与实际生活联系紧密。本书共有 7 个数据库,如制造业数据库、金融数据库、农业数据库等,它们共提供了 8 350 个数据、56 个变量。这些数据都有可靠来源,如美国劳动统计局、纽约证券交易所等。同时,本书共有 600 多道习题,其中大多数来自于真实的公司,是真实的商务与经济状况的反映。另外,每章结尾都会有一个真实案例,利用这些实际生活中的例子与资料,便于学生运用所学的统计知识去分析预测,以得出相关结论。

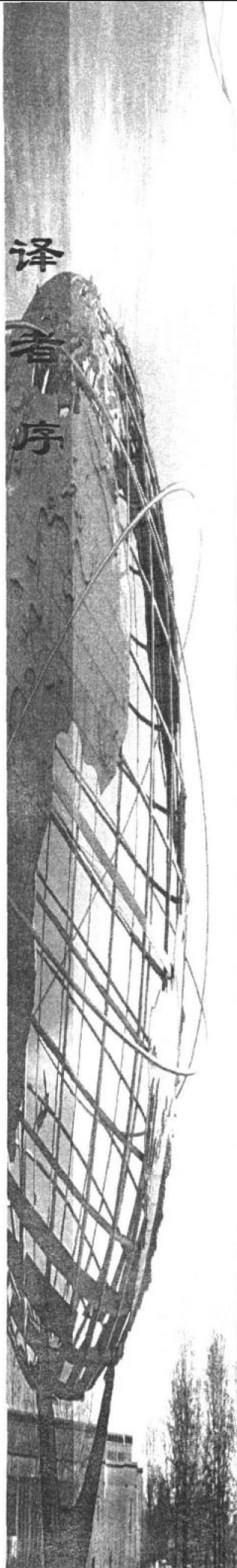
在翻译过程中,我们得到外经贸部国际贸易经济合作研究院的杨鹏和王昊宇两位同志的大力协助,在此表示衷心感谢。

由于时间和水平有限,疏漏与差错在所难免,敬请各位读者批评指正。

张久琴 张玉梅 杨琳

2003 年 4 月

0359/03



肯·布莱克(Ken Black)

出生于马萨诸塞州的坎布里奇,在密苏里长大。目前任休斯顿大学明湖分校商务与公共管理学院的决策科学教授。他在格瑞斯兰大学获数学学士学位,在埃尔帕索城的得克萨斯大学获得数学教育硕士学位,在北方得克萨斯大学获得管理科学的商务管理和教育研究博士学位。

1979年任教以来,他一直从事各种统计课、预测、管理科学、市场研究和生产/运营管理的教学工作。他已发表15篇期刊作品、20多篇专业论文和两本教科书(《商务统计:入门课程》、《商务统计:当代决策制定》)。他给许多不同的公司做顾问,有美国艾德纳人寿保险公司、休斯顿城市公司、NYLCare、美国电报电话公司、约翰逊空间中心公司、西南信息资源公司、联结公司和雄鹰工程公司。

肯·布莱克与妻子卡罗琳共有两个女儿——凯斯和温迪。他爱好吉他、阅读,担任女子垒球教练,并参加教职工田径比赛中的跳远和三级跳项目。

戴维 L. 埃尔德雷奇(David L. Eldredge)

默里州立大学商务与公共政策学院的计算机科学教授。在衣阿华州立大学获航空工程技术的工学学士学位,在俄亥俄州立大学获运作研究的硕士及博士学位。

教授目前负责默里州立大学 MBA 课程的管理科学及本科生商务管理课程中商务统计的教学工作;在商务学院担任了主任一职8年;已出版了15本书并发表了不少论文,内容主要关于激励和博弈问题、使用 Microsoft Excel 进行数据分析的问题。他已在4家世界500强企业中积累了10年的工作经验。

戴维·埃尔德雷奇和妻子朱迪有三个女儿——柯瑞斯迪、林恩和特蕾西,有7个外孙和外孙女。埃尔德雷奇一家喜欢野营度假和旅游,遍历美国50个州,并游览过许多国家。

怀着欣喜与自豪之情,我们奉献出自己的这本新书。本书可以作为本科生和 MBA 学习统计学的入门教材,共包括 15 章,主要关于应用商务统计的基础知识,而实际上所有重要的商务统计内容都包含在这本对象明确、易于阅读的教材中。我们相信这本将实际商务统计与 Microsoft Excel 完美结合的教材,正是老师们一直为自己和他们的学生所寻找的书。

本书面向那些要成为决策者而非统计员的商务及经济类学生,所以书中的概念都是通过真实的商务、经济数据和案例进行阐述的。本书的主要目的是给出应用统计方法的一个简洁明确的概念;另一个目的就是要将 Microsoft Excel 如何用于分析商务及经济数据的问题解释清楚。因此,读者会发现,本书在统计学概念/方法和理解 Excel 计算得出统计结果之间所做的平衡安排。虽然我们认为读者应该学过大学数学,但本书中并不涉及动手计算。

本书是肯·布莱克和戴维 L. 埃尔德雷奇共同努力合作的结晶。肯·布莱克出版过两本非常成功的商务统计学教材,里面有着精彩的案例与问题。戴维 L. 埃尔德雷奇是 Excel 方面的一个专家,因《*Microsoft Excel Companion to Business Statistics*》一书的出版而成名。在那本书中,他对如何利用 Excel 进行统计数据分析做了解释与说明。

本书在论述商务统计方法时融入了 Excel。第 1 章先对 Excel 的统计和图示性质给出一个大体认识,之后在本书的相关章节中还会再次提到 Excel 统计方法。换句话说,Excel 将会在需要的地方出现,而不是出现在每章的末尾或附录中。在对每种 Excel 方法进行解释时,都包括 Excel 输入对话框和某个例子的输出结果,以便让学生准确地知道如何输入数据,如何对结果进行确认和解释。实际上,因为所有问题都有原始数据,所以学生们都能够利用 Excel 进行数据分析。问题中用到的所有数据都含在本书附带的 CD-ROM 中。

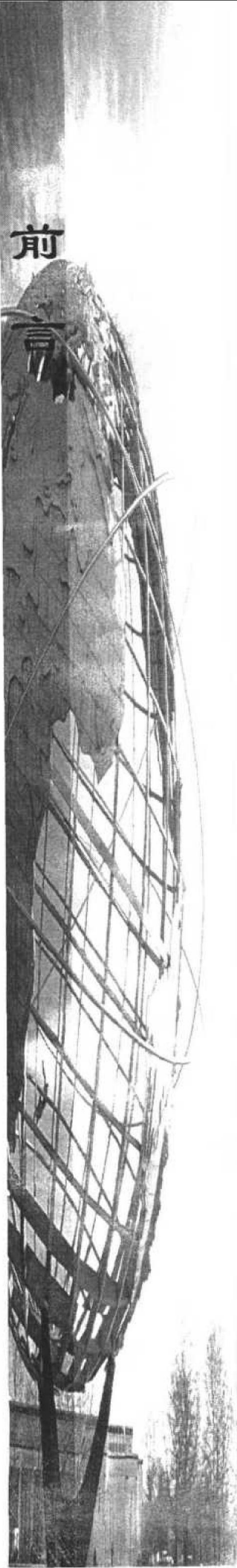
有几个统计问题用 Microsoft Excel 方法并不能得到解决。为满足这方面的要求,我们使用了 FAST \ STAT。它是一种附加软件包,形式与功能和 Excel 类似,它可用于 Excel 所不能解决的统计分析问题中。使用 Excel 和 FAST \ STAT,本书中的多数统计值都能计算出来。

FAST \ STAT

FAST \ STAT 附加软件包是本书的一大特色,这一部分由戴维 L. 埃尔德雷奇撰写,共包括 15 个宏程序,它充实和扩展了 Excel 的统计功能。FAST \ STAT 很容易安装,本书附带的光盘提供了这个软件,同时还有相关文件和安装信息。在 Windows 98 和 Windows 95 的操作系统下,安装 Excel 2000 或 Excel 97 后,就可以运行 FAST \ STAT 软件了。

FAST \ STAT 和 Excel 配合完美。在 Excel 中安装了 FAST \ STAT 后,FAST \ STAT 就变成了 Excel 菜单栏中的选择项。它有一个下拉菜单,以方便你选择想使用的宏。这些宏进行了专门设计,其输入对话框与输出样式和 Excel 相似。但 Excel 中的数据分析是静态的,FAST \ STAT 中宏的输出结果则是动态的。意思就是说,如果宏的使用者返回到原始输入数据,对任意一个输入数值进行更改,FAST \ STAT 将会重新进行运算,根据新数据更新输出结果。

FAST \ STAT 的 15 个宏包括:盒形须线图(第 3 章),概率(第 4 章),均匀分布(第 6 章),使用 Z 分布的单均值(第 8 章、第 9 章),使用 t 分布的单均值(第 8 章、第 9 章),单均值的样本



容量(第8章),单比率的C.I.和H.T.(第8章、第9章),单比率的样本容量(第8章),双比率的假设检验(第10章), χ^2 独立检验(第11章),回归C.I.和P.I.(第12章),杜宾-沃森检验(第14章), $\bar{X}$ 图和 R 图(第15章), P 图(第15章)以及 c 图(第15章)。

本书特点

特有的 Microsoft Excel 内容

现在的学生在家里、学校或工作场所都可以使用 Excel,因此他们想知道如何充分利用这个软件。Excel 是本书特有的内容,全书每章都有有关 Excel 的讲解,几乎每节都有 Excel 的内容。我们做了极大的努力,以使对 Excel 的讲解与本书的统计内容配合完美,这样 Excel 就更容易学会,也更方便使用和理解,但我们仍然以讲解统计知识为主线。本书包括了 60 个计算机输入对话框,200 多个计算机统计输出结果。前面也提到了,Excel 缺少某些统计功能,为了扩充它的用途,我们编写了 FAST $\backslash$ STAT。

内容全面

传统统计书籍的内容和 Excel 所包含的分析工具一致。本书编写的内容适合于一个学期的学习,它不仅包括了图形统计、描述统计、概率、离散和连续分布、样本分布、假设检验和估计等一般内容,而且还有一章专门讲解了单边 ANOVA、随机模块设计和双边 ANOVA。除此之外,还有拟合优度的 χ^2 检验和独立性的 χ^2 检验。因为商业领域的决策者经常使用回归和预测工具,所以本书单列出几章,专门讲解了多元回归分析和时间序列预测。最后一章是统计质量控制。

例题

实际上每一章的每一节都有例题。除了正文中举的例子外,我们还附加了例题,它有详细的解答过程,并用于扩展讨论过的统计分析内容。这个教学工具对讲解的内容进行了补充,并加深了学生对教材内容的印象。

数据库

本书包含了 7 个数据库,他们都以 Excel 文件存在光盘中。这 7 个数据库是:制造业数据库、金融数据库、股票市场数据库、国际人员雇用量数据库、能源数据库、健康护理数据库和农业数据库。这些数据库共提供了 8 350 个数值和 56 个变量。数据来源可靠,并注明了具体提供单位,包括美国劳动统计局、纽约证券交易所、美国农业部、穆迪公司普通股票指南、美国医院联合会和美国人口统计局。7 个数据库中有 4 个包含时间序列数据,其中一个包含有 168 个月份的时间序列数据,非常适合预测技术的应用与分析。

习题

本书有 600 多道习题,其中大多数都来自真实的公司,并且是真实的商业和经济状况。大多数习题都包括原始数据,可以用 Excel 或 FAST $\backslash$ STAT 进行分析。

数据库分析

数据库分析在每章的结尾部分,一般包括几个问题。它利用数据库中的变量对本章讲到的方法进行实际应用,解决这些问题有利于提高应用计算机解题的技能。

案例

每章都以案例作为结尾,这些案例都是真实的,可以使学生有机会利用本章讲到的统计概念和技术解决商业难题。有些案例是世界著名的公司,比如壳牌、可口可乐和高露洁-棕榄公司;也有些案例是关于小公司的,比如 Ther-

matrix、Robotron、Fletcher-Terry 等,这些小公司克服了很多困难和障碍才得以生存和发展。大多数案例包含原始数据,鼓励学生用本章讲到的方法进行分析,解决问题。在多数情况下,学生必须对软件输出结果进行分析,才能得出结论,做出决策。

光盘

本书附有光盘,上面有宝贵的资料,可以帮助学生学习,协助老师教学。光盘的内容包括:

- FAST \ STAT 软件和操作手册
- 本书习题和案例的 Excel 数据文件
- 以 Excel 文件存在的 7 个数据库
- 为没有使用过 Microsoft Windows 和 Excel 的学生提供的使用指导,包括一些有关工作表的练习,以使工作表得到更为有效的应用
- Excel 的数据分析工具使用指导
- Excel 的统计发现工具使用指导,包括它与数据分析工具的差别
- 附加信息和数据,以及一些为时间序列预测设置的特殊模版(第 14 章)

辅助教学和学习资料

我们为学生提供了两本学习资料,可以直接通过 <http://www.swcollege.com> 网站购买。

- WebTutor 是一种新型的电子辅助读物,它由本书的作者肯·布莱克和戴维 L. 埃尔德雷奇提供,主要有两个版本。

WebTutor 可以由教学者在上课时使用,它为学生提供了网上学习资料(包括复习资料、学习中的问题和解答、奇数习题的答案);此外还有强大的沟通和课程管理工具,包括课程安排,讨论和发给老师的 Email。

WebTutor 在 WebCT 和 Blackboard 上。要得到更多信息,请登陆 <http://webtutor.thomsonlearning.com>。

个人 WebTutor 为自学的学生提供了学习和复习资料。要得到有关这个产品的更多信息,请登陆网站 <http://pwt.swcollege.com>。

- 墨里州立大学的戴维 L. 埃尔德雷奇编写的《*Microsoft Excel Companion for Business Statistics*》一书,详细讲解了商业和经济统计学基础课程中 Excel 的应用。

目 录

译者序
作者简介
前言

第1章 使用 Excel 软件进行商务统计介绍	1	第4章 概 率	73
1.1 商务统计应用	2	4.1 概率概述	74
1.2 基本统计概念	2	4.2 分配概率的方法	74
1.3 数据	4	4.3 概率的结构	76
1.4 用 Excel 进行统计分析	8	4.4 边际概率、合并概率、联合概率和条件概率	80
1.5 数据分析工具	9	4.5 加法法则	81
1.6 统计功能	9	4.6 乘法法则	88
1.7 图表向导功能	10	4.7 条件概率法则	92
1.8 趋势线功能	13	4.8 概率的修正:贝叶斯公式	99
1.9 FAST \ STAT 附加宏	15	小结	102
小结	17	案例 4A 高露洁-棕榈公司发动“全效”冲击	106
案例 1A Digiomo 比萨店:引进冷冻比萨与 外卖比萨竞争	20	第5章 离散型概率分布	108
第2章 描述性图表	22	5.1 离散型分布和连续型分布	109
2.1 频数分布	23	5.2 二项分布	110
2.2 用图形描述数据	27	5.3 泊松分布	119
小结	33	5.4 超几何分布	126
案例 2A 肥皂公司间的竞争	36	小结	130
第3章 描述统计学	38	案例 5A 富士胶片引入 APS	134
3.1 集中趋势中的测度	39	第6章 连续型概率分布	135
3.2 描述离散程度的测度值	47	6.1 均匀概率分布	136
3.3 形状测度值	61	6.2 正态分布	141
3.4 Excel 中的概括统计量	66	6.3 运用正态分布曲线解二项分布问题	151
小结	66	6.4 指数分布	157
案例 3A 可口可乐在俄罗斯的包装越变越小	71	小结	161
		案例 6A 奔驰追求年轻客户	165

第7章 抽样和抽样分布	166	第12章 简单回归及相关性分析	341
7.1 抽样	167	12.1 简单回归分析简介	342
7.2 $\bar{X}$ 的抽样分布	175	12.2 确定回归直线方程	343
7.3 $\hat{p}$ 的抽样分布	185	12.3 分析回归直线的拟合情况:残差、 r^2 和 预测的标准差	350
小结	187	12.4 回归模型斜率的假设检验以及对模型的 总体检验	359
案例 7A 壳牌石油公司试图重返霸主地位	190	12.5 预测	364
第8章 单样本估计	191	12.6 相关检验	370
8.1 估计大样本空间的总体均值	192	小结	375
8.2 估计总体均值:小样本容量, σ 未知	199	案例 12A 三角洲线缆的强大武器:培训	378
8.3 估计总体比例	205		
8.4 估计样本容量	209	第13章 多元回归分析	381
小结	215	13.1 多元回归模型	382
案例 8A Thermatrix	217	13.2 对多元回归模型的评估	388
第9章 单样本假设检验	219	13.3 指示虚拟变量	398
9.1 假设检验简介	220	13.4 较为复杂的回归模型	403
9.2 大样本的单均值假设检验	224	13.5 模型的建立:搜寻过程	412
9.3 小样本的单均值假设检验: σ 未知	231	13.6 多重共线性	419
9.4 有关比率的假设检验	236	小结	421
小结	241	案例 13A 弗吉尼亚半导体	426
案例 9A 菲多利公司瞄准西班牙市场	243		
第10章 双样本的假设检验	245	第14章 预测	428
10.1 两个均值之差的假设检验:独立大样本	246	14.1 预测概述	429
10.2 两个总体均值之差的假设检验: 独立小样本	254	14.2 时间序列的成分	431
10.3 两个相关总体均值之差的假设检验	260	14.3 时间序列的预测程序	435
10.4 两个总体比率之差的假设检验	267	14.4 静态预测法	449
10.5 两个总体方差的假设检验	273	14.5 趋势预测法	456
小结	281	14.6 季节预测模型	478
案例 10A 塞兹公司:生产合格的驱动性齿轮 和线性机械产品	285	小结	489
第11章 方差分析及 χ^2 方应用	287	案例 14A DeBourgh 制造公司	493
11.1 实验设计介绍	288		
11.2 完全随机设计 (单向 ANOVA)	289	第15章 统计质量控制	496
11.3 随机分组设计	298	15.1 质量控制概述	497
11.4 析因设计(双向 ANOVA)	308	15.2 流程分析	502
11.5 χ^2 拟合优度检验	319	15.3 控制图	507
11.6 列联表分析: χ^2 独立性检验	327	小结	524
小结	334	案例 15A ROBOTRON	529
案例 11A 博润德:由坎坷到光明	338	附录 A	532
		附录 B 部分序号为奇数的习题 (定量型习题)答案	562

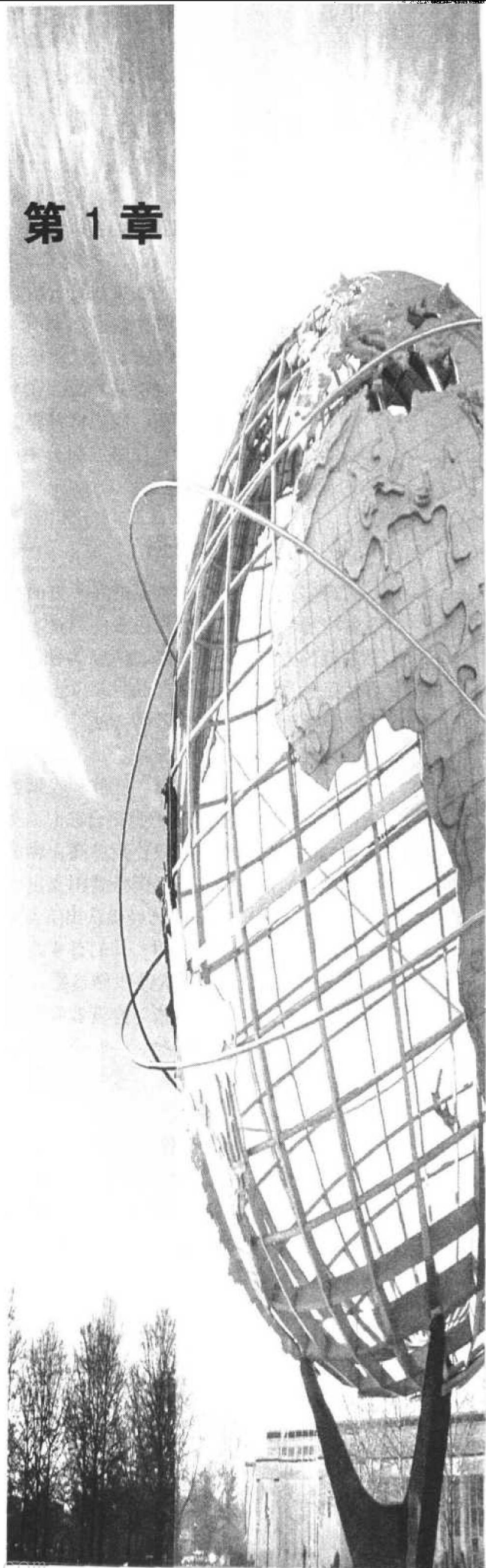
第 1 章

使用 Excel 软件进行商务统计介绍

学习目标

第 1 章的主要学习目标是向读者介绍统计学,使读者能够:

1. 定义统计学。
2. 了解统计学在商业中的广泛应用。
3. 区别描述统计学和推断统计学。
4. 根据数据的等级和种类进行分类,并理解这样做的重要性。
5. 了解 Excel 的统计分析能力。



在工作日里，每分钟都有大量与各行各业相关的决策在世界各地被制定，正是这些决策决定了公司是盈利增长还是亏损下滑。大部分决策都是依据相关的信息而得出的，这些信息包括市场信息，经济和财务信息，劳动力和竞争信息等等。这些信息常常以数据为形式或包含数据。统计就是获取和分析这些信息，并用它帮助决策的过程。所以，在 21 世纪，在活跃的商务领域中，商务统计对决策有重要的作用。

商务统计需要对大量数据进行分析，所以通常需要使用计算机。数据常以计算机表格的形式体现，因为表格工具是商务管理软件中最常用和有效的。目前，运用最广的表格程序是 Microsoft Excel。很长一段时间来，商业和工业管理者和分析员在组织他们计算、绘图和数据管理时都要用到 Excel。除了这些，Excel 目前的版本提供了多种统计分析的功能。在本书中，我们大部分的统计都是用 Excel 做的。

本章下一部分的开头，我们将首先引用一个现实中的商务统计案例。在接下来的部分我们将讨论一些关于商务统计的基本概念。在第三部分中讨论的是统计中的基本元素——数据。在最后的 6 节里将介绍使用 Excel 进行统计分析的特点。

1.1 商务统计应用

事实上，在商务中涉及的每一方面都要使用统计进行决策制定。Pitney Bowes 对美国大中型公司的 302 名负责市场销售的经理及副总裁的调查表明，差不多有 35% 的人认为直接使用邮件或产品目录是与顾客接触最划算的方法。另有 11% 的人认为使用互联网是最有成本效益的。该研究还表明，提高品牌知名度的最好办法是通过直接邮件或产品目录。该研究中收集总结的这些和其他一些统计资料可以帮助决策制定者为他们产品找到最有成本效益的方法。

如果决策制定者想要减少员工的医疗费用支出，他们可以从健康保健研究协会对 4.6 万名雇员的研究中发现，压力往往比高血糖、肥胖和吸烟更容易导致较高的医疗费用。研究表明，受压抑的员工的医疗支出要比不压抑的员工高 70%，觉得自己总是处于压力下的员工的医疗支出要比没有压力的员工高 46%，而患高血糖的员工的医疗支出只比没患高血糖的员工高 11%。该信息与此研究提供的其他统计数据表明决策者可以找到有效的途径来降低医疗费用支出。

Yankelovich Partners 已经确认出 X 代的一个分支团体，叫做“朋友化的 X 代”。这个团体是指什么呢？关于此团体的商务统计表明，他们有 4 万美元的中等家庭收入和 1.9 万美元的中等个人收入。20% 的人有至少一个大学学历，63% 的人从没结过婚，55% 的人在 25 岁以下。这些统计数据可以帮助决策者分析这个团体中的人的市场潜力，并指导经营者和产品设计者如何满足该类人群的需要。

决策者追踪经济情况。一些商务统计包括新房销售数量、消费者信心指数、国内生产总值增长的百分比、初期失业数和失业率。这些统计数据都是经济走势和财政状况的指数，可以被用来预测未来的经济形势。

图 1-1 显示出美国使用电子邮件的消费者大量增加。根据这个由 Forrester 研究中心提供的数据，美国 1999 年使用电子邮件的消费者人数是 1992 年的 20 倍，并将在未来的 5 年中再增长一倍。管理者可以用这个统计数据找到和员工更有效地沟通的方法，通信公司也可以借此信息规划新的市场机遇。

在本文中，我们将学习描述数据的几种方法和如何对这些数据重组，以便决策者可以更方便地利用它们。我们将学习从总体中提取样本的方法，因为这样可以更及时有效地研究商务世界。我们还将学习如何预测的各种方法。本文还包括假设检验和估计总体值的统计工具。许许多多的统计知识和统计技巧将在这条通往商务统计的路上等着我们，让我们开始吧。

1.2 基本统计概念

商务统计和其他学科一样，需要理解一些基本概念才能开始理解和交流。我们从对统计这个词的理解开

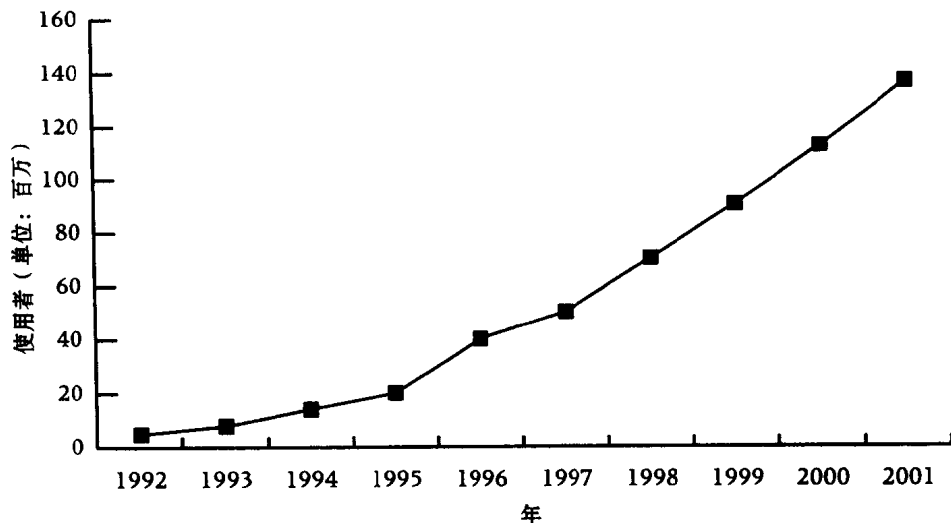


图 1-1 用 Excel 制作的图: 美国使用电子邮件的消费者统计

始讲。它有许多含义。在韦伯斯特国际辞典第3版中,给出了**统计(statistics)**的全面的定义:统计是一门收集、分析、解释和提供数据的科学。从这个角度讲,统计包括本文中所提到的所有问题。统计还是数学的一个分支,大部分统计原理都是基于数学原理或是其发展演变。包括商务在内的任何学术领域,都结合自己的特点来讲统计学;但统计学仍然还是独立的一门课。

人们常用数据统计这个词指代一组数据,如某人从商务运作中得到一组统计数据,其实指的是事实和数字。媒体常用“统计”指代死亡,成为统计这个词的这种含义可真不是件令人期待的事。

统计这个词至少还有其他两种用法。第一,统计是从样本中计算出的描述性方法,用来推测总体的值。这种用法我们以后将讨论。第二,统计是在数据分析中的分布研究。如“研究者用 t 分布来研究数据”这句话是指“研究者用 t 统计方法来研究数据”。

以下是“统计”一词的常用含义:

- (1) 获取、分析、研究、解释和提供数据
- (2) 数学的分支
- (3) 一门课程
- (4) 事实和数字
- (5) 死亡
- (6) 对样本的研究方法
- (7) 用来研究数据的一种分布

统计有多种分类方法。主要的一种是把统计分为两个分支:描述统计学和推断统计学。为了了解这两者的区别,先要知道样本和总体的定义。韦伯斯特国际辞典第3版把**总体(population)**定义为:人、物和想要研究的项目的集合。总体可以是广义的,如“所有的汽车”,也可以是狭义的,如“所有1997~1999年生产的福特马自达汽车”。总体可以是一群人,如“微软现有的所有雇员”;也可以是物,如“通用电气公司在2000年2月3日在路易斯维尔厂生产的所有洗碗机”。研究者把总体定义为任何他想要研究的东西。当研究者为了研究某方面从总体中获取数据时,称为**普查(census)**。大部分人都熟悉美国人口普查,它是每隔10年对美国人口的调查。如果研究者希望得到亚利桑那州大学所有学生的学术态度测试(SAT)分数,可以对该校现有学生进行调查。

样本(sample)是总体的一部分,是总体的代表。由于各种原因(将在第7章中解释),研究者是对样本而非总体进行研究。如,对灯泡的寿命质量进行检验时,生产者会从生产线上随机抽取75个。由于时间和经

费的限制,人力资源部经理会抽取 40 名员工进行公司道德测试,而不是对公司所有人进行。

如果经济分析员用一个组中获得的数据对同一个组进行描述或得出结论,则称其为**描述性统计 (descriptive statistics)**。如老师对班级考试成绩统计,并得出只针对该班的结论,比如班级平均成绩、成绩分布的范围或该班此次考试的其他数据。

如果商业研究者正在使用从一组中所收集到的数据描述该组或者得出有关该组的结论,这种统计也被称为描述性统计。大部分商务统计是描述性统计。比如 6 月份休假的员工的人数,丹佛办事处的平均工资,2001 年的销售收入,员工对管理层的满意程度,1998~2001 年 Lofton 公司的平均投资回报率。

另一种统计称为**推断性统计 (inferential statistics)**。如果分析员得到样本数据,并以此样本数据得出关于产生此样本的总体结论,这个统计为推断性统计,即用获得的数据来对更大的组进行推断。推断性统计也被称为归纳性统计。推断性统计的运用和重要性正在不断增加。

推断性统计可以被应用于药物研究。某些新药品价格昂贵,只能针对少数患者的服用效果进行统计。运用推断统计学,研究人员可以根据对少数样本患者进行实验得出的结论,推断总体的情况。

市场研究人员用推断统计学研究广告对不同的市场区域的影响。假设某软饮料制造公司设计了一个新广告,意思是一个会讲话的售货机同买者讲话。该公司想调查此广告对不同年龄段的消费者的影响。研究人员可以把消费者按年龄分段,从每段中随机抽取样本进行研究,再用推断统计学推出广告对不同年龄段消费者的影响。推断统计学的好处是可以高效地研究一种广泛存在的现象而不用对总体进行普查。本文中讨论的大部分问题和推断统计学有关。

对总体的描述方法称为**参数 (parameter)**。参数常用希腊字母来表示。如总体平均值(μ),总体方差(σ^2),总体标准差(σ)。对样本的描述性指标称为**统计量 (statistic)**。统计量常用拉丁字母来表示,如样本均值($\bar{X}$),样本方差(S^2),样本标准差(S)。

区别参数和统计量这两个术语的重要性主要体现在推断统计学中。统计人员常要估计或研究参数,然而由于时间和资金的限制,对参数值的直接测算是不切合实际的。所以,统计人员会从总体中抽取样本,计算样本统计量,然后推断参数值。推断统计学的基础,就是可以在不对总体做普查的情况下,对参数进行估计。

例如,洗衣机制造商要测定新洗衣机在需要维修前每次洗衣的平均重量,参数为洗衣重量的总体均值。统计员从总体中抽取样本,计算每台机器需要维修前的洗衣重量和所有样本机器的均值,并利用统计量(在本例中是样本均值)估计出总体的参数。图 1-2 表示了该过程。

对参数的估计是不确定的。除非参数是直接从事体计算出的,否则统计人员就不能确定从样本中推断出的参数值是否精确。为描述这种推断的可信程度,商务统计中要用到概率。所以本书还将涉及概率(第 4 章)。

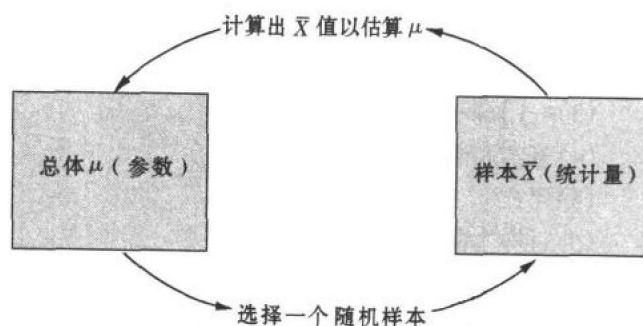


图 1-2 用推断统计学计算总体均值的过程

1.3 数据

商业中每天都要收集上百万的数据,来表示难以数计的各种研究对象。如表示生产某种产品要花费多少美元、零售点的地理位置分布、船舶载重和每年对下属员工的评估。不能对所有的数据都用同一种统计方法进行计算,因为不同的数据代表的意义是不同的。所以,分析人员要能够确认数据类型,以收集要研究的数据。

数据的不同作用可以用 40 和 80 这个例子表示。它们可以指两个在运货物的重量，两个不同的产品在消费测试中得到的比例，或是后卫与边接球手运动衫的号码。虽然 80 磅^①是 40 磅的两倍，但是边接球手的块头却不一定有后卫的两倍大。对两个重量取均值看起来是合理的，但是对足球运动衫的号码取均值却毫无意义。对数据分析的适当性取决于所收集数据的度量标准。数据所代表的现象决定了数据度量的标准。常用的 4 种度量数据的类别如下所示：

- (1) 名义
- (2) 序数
- (3) 区间
- (4) 比率

1.3.1 名义类别数据

度量数据的最低标准是使用名义值。用来代表**名义类别数据(nominal level data)**(类别常被省略)仅用来对数据进行分类。名义数据的一个例子是员工工作证号。该号码仅用来区别员工，而不是对员工价值的鉴定。调查中由许多统计问题得出的数据是名义数据，因为所提的问题仅是用来分类的。举下述这样一个例子：

下列哪一个选项最好地描述了你所在的行业？

- | | | |
|----------|---------|---------|
| A. 教育工作者 | B. 建筑工人 | C. 制造工人 |
| D. 律师 | E. 医生 | F. 其他 |

假设为了计算的目的，为教育工作者赋值为 1，建筑工人赋值为 2，制造工人赋值为 3，依此类推。这些数字仅用来为调查对象分类，数字 1 并不代表最高的级别，仅是用来区别教育工作者(1)与律师(4)。

其他经常使用名义数据来表示的变量有：性别、宗教信仰、种族划分、地理区域及出生地等。社会保险号码、电话号码、雇员 ID 号码和邮编也可用名义数据来表示。适合用于分析名义数据的统计方法是有限的，但一些使用极广的统计量，例如 χ^2 分布统计量，可以使用名义数据，常可给出有用的信息。

1.3.2 序数类别数据

序数类别数据(ordinal level data)的度量比名义数据高一个级别。除了具有名义数据的功能外，序数类别数据还可为对象排序。例如使用序数数据，一位主管可为 3 名职员的生产效率从 1~3 排序。使用序数数据，该主管可以把其中的一位雇员视为最有效率的，一个最没效率的，另一个居中。然而，该主管不能使用序数数据建立排序为 1 和 2 的职员之间与排序为 2 和 3 的职员之间的相等区间。也就是说，她不能肯定三个职员不同排序间的生产效率的差一定是相同的。使用序数数据，由连续数据所代表的距离不总是相等的。

有些研究者在问卷调查中使用的 Likert-type 标度就是序数形数据。下面就是这样的例子：

该计算机课程_____

没有用	有点用	有一定帮助	非常有用	极其有用
1	2	3	4	5

当该问题被输入到计算机时，输入的只是 1~5 这 5 个数字，而不是上面所陈述的是否有用的文字表述。在该标度中，实际上每个人都会同意 5 大于 4，为该问题的回答排序是可能的。然而，多数受访对象不会认为上述 5 个选项间的差是相等的。

有时候对作为投资工具的互助基金按照风险的类别无风险——货币风险——利率风险进行等级划分。把这三种类别按高风险、中等风险、低风险进行排序。假设对高风险赋值为 3，为中等风险赋值 2，低风险赋值 1。如果一个基金被划为 3 而不是 2，这说明它的风险更大。然而，1、2、3 类所代表的风险的差却不一定相等。因此，对这些风险的量度仅是序数的。在商业分析中另一个使用序数数据的例子是在财富杂志中为

① 1 磅 = 0.453 592 37 公斤。

50 位最值得尊敬的公司排序, 为公司排序的序号仅是序数性的。有些统计方法专门用于序数数据, 但是更多的方法不适合序数数据。

1.3.3 区间类别数据

下一个更高类别的数据是**区间类别数据(interval level data)**。区间数据除了具有序数数据的所有特征外, 连续数据间的距离是有意义的, 并且数据总是数值型数据。连续数字间的差所代表的距离是相等的, 即区间数据具有相等的间距。区间度量的一个例子是华氏温度。使用华氏温度, 可对温度进行标定, 连续数字如 20°, 21°, 22°之间的热量差数是一样的。

另外, 使用区间数据, 零点的选定是为了方便或出于惯例, 而不是通常所指的自然数或固定的零点, 即零只是坐标上的一个点, 并不表示该现象不存在。还有一些区间数据的例子, 如就业百分比的变化, 股票回报率、股票价格的美元变化。

使用区间数据, 在不同的度量单位间转换, 只需乘以某个系数 a , 再加另一个因素 b 即可, 即 $y = b + ax$ 。举个例子, 把摄氏温度转换为华氏温度, 关系式为:

$$\text{华氏度} = 32 + \frac{9}{5} \text{摄氏度}$$

1.3.4 比率类别数据

比率类别数据(ratio level data)是最高类别的数据。比率数据与区间数据具有相同的特征, 此外, 比率数据有绝对零点, 两个数字间的比值是有意义的。所谓绝对零点是指零是固定的, 代表不具有所研究特征的数据的值。零点不能任意指定, 因为它代表一个固定的数值。该定义可以使商业分析者建立数据间的比例关系。

比率数据可以是高度、重量、时间、体积及开氏温标。使用比率数据, 研究者可以说 180 磅的重量是 90 磅的两倍, 或者将二者相比 180:90。在工业中许多由机器收集的数据都是比率数据。

在商业中其他使用比率数据的例子还有生产周期、工作时间、旅行的英里数、所出售的卡车数量、每 1 万次飞行的投诉率及员工数量。使用比率数据, 在不同的度量单位间转换, 公式中没有因素 b , 即 $y = ax$ 。例如, 把高度从码转换为英尺[⊖]为: 1 英尺 = 3 码[⊖]。

1.3.5 4 种数据类别比较

图 1-3 给出了 4 种不同的数据类别使用范围关系图。同心的正方形表示高级别的数据可以使用低级别的统计方法进行分析; 此外, 还可以使用其他的统计方法, 即用于分析其他 3 类数据的统计方法可以用来分析比率数据, 同时还可以使用其他的统计方法分析比率数据。

按所用的统计分析类型, 名义数据是最受限制的一类数据。商业分析者除了可以使用分析名义数据的方法分析序数数据外, 还可以使用其他一些方法。使用比率数据, 商业分析者可进行比率间的比较, 还可以适当运用于名义、序数或区间数据的分析方法。有些统计方法只适用于比率数据, 而不能用来分析其他 3 类数据。

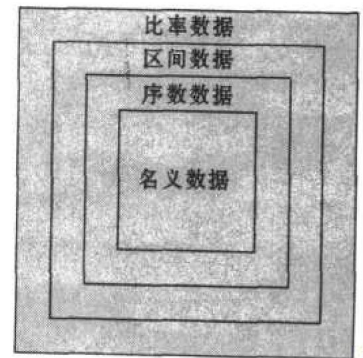


图 1-3 不同数据类别的使用范围

1.3.6 定性数据与数量数据

使用上面讨论的数据类别收集数据。这些数据类别可被分为两类数据。数据类别决定了合适的统计分析方法。

由于名义数据和序数数据来自不精确的度量, 比如人口统计问题、人或物体的分类、项目的排序等, 因

⊖ 1 英尺 = 0.304 8 米。

⊖ 1 码 = 0.914 4 米。

此名义数据和序数数据被称为非度量数据。这些数据被称为种类或**定性数据 (qualitative data)**。定性数据一般包括标记或种类，他们既可以像名义数据那样是无序的，也可以像序数数据那样是排序的。他们既可以为数值型也可以为非数值型。例如，你既可以用非数值型种类，如用一年级新生、二年级学生、三年级学生和四年级学生为大学生分类，也可以使用数值型数字 1、2、3、4 进行分类。

另一方面，区间数据与比率数据通常由精确的工具测量得出，如在生产工程过程，在国家标准测试中，在标准账户程序中产生的数据，这些数据都是数值型数据。把这些数据称为**数值型数据或数量数据 (quantitative data)**。数量数据自然包括那些数字值，因此数量数据总是数值型的。

有些统计分析方法适用于定性数据，有些适用于数量数据。定性数据通常通过计量每一类中的项目个数或每一中项目的比例而得到。第 2 章中所介绍的制表绘图的方法是适合定性数据的描述方法。定性数据的推论程序在第 7~10 章及第 15 章中介绍，另外还包括第 11 章利用 χ^2 统计量。

通过计算给定的区间内的数值或每一区间内项目的比例，可以得到数量数据。因此，第 2 章中介绍的制表及绘图的方法是对数量数据合适的描述方法。然而，你仍可以使用通常的分析方法，如计算均值，测量数据的变动性。因此，第 3 章中所介绍的数值分析方法也是对数量数据的合适的描述方法。数量数据的推导程序在第 7~10 章和第 15 章的部分节中以及第 4~6 章，第 11~第 14 章中介绍。

在结束本节前，我们对数据进一步分类。数值型数据既可以是离散的又可以是连续的，这取决于可能得到的值。如果数值仅来自于特定的可能值的清单，那么就是**离散型数据 (discrete data)**。例如，统计班级学生的数量就是离散型数据。离散型数据是来自于计数过程的数值。

另一方面，如果你不能列出数据项目所有可能的取值，该数据就是**连续型数据 (continuous data)**。例如，一个学生的重量可以说为 135 磅。然而实际的重量可能是 135.089、134.97 或者是一个可能的无限的数。因此，学生的重量是连续型数据。连续型数据是来自测量过程的数值。

数值型定性数据总是离散型数据。然而，数量数据既可以是离散型数据，也可以是连续型数据。图 1-4 是我们在本节中所讨论的所有数据类型的小结。离散型数据与连续型数据的区别对我们后面介绍的概率分布的可能取值的描述是很重要的。我们将在第 5 章讨论离散型分布，在第 6 章讨论连续型分布。

例题 1.1

保健行业发生了许多变化。由于市场竞争激烈，为了制定提高服务水平的方法，医院管理者有时会在病人出院后向他们发出服务满意度的调查表。下面列出了这种调查常问的问题。从这些问题中会得出哪种类型的数据？

1. 你出院多久了？

2. 你住院的大部分时间，主要呆在哪个病房？

☐ 冠状动脉护理中心	☐ 内科病房
☐ 特护病房	☐ 儿科病房
☐ 妇产科病房	☐ 外科病房

3. 在选择医院时，医院的地理位置重要吗？（仅选择一个）

非常重要 有些重要 不是很重要 根本不重要

4. 你第一次住院时，病情很严重吗？

☐ 危险 ☐ 严重 ☐ 一般 ☐ 轻微

5. 为你的医生的医术打分。

☐ 医术高明 ☐ 很好 ☐ 好 ☐ 一般 ☐ 差

6. 在下面从 1~7 的坐标上，标出护士的护理水平：

差 1 2 3 4 5 6 7 优