

(原书第15版)

商务与经济统计方法

Statistical Techniques in Business & Economics (15th Edition)



附光盘

道格拉斯 A. 林德 (Douglas A. Lind)

卡罗来纳海岸大学和多伦多大学

威廉 G. 马歇尔 (William G. Marchal)

多伦多大学

塞缪尔 A. 沃森 (Samuel A. Wathen)

卡罗来纳海岸大学

【美】

著

聂巧平 叶光 译



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

商务与经济统计方法 (原书第 15 版)/(美) 林德 (Lind, D. A.), (美) 马歇尔 (Marchal, W. G.), (美) 沃森 (Wathen, S. A.) 著; 聂巧平, 叶光译. —北京: 机械工业出版社, 2015.9 (经济教材译丛)

书名原文: Statistical Techniques in Business & Economics

ISBN 978-7-111-51734-4

I. 商… II. ①林… ②马… ③沃… ④聂… ⑤叶… III. ①商业统计-统计方法-高等学校-教材 ②经济统计-统计方法-高等学校-教材 IV. ①F712.3 ②F222.1

中国版本图书馆 CIP 数据核字 (2015) 第 239294 号

本书版权登记号: 图字: 01-2015-2385

Douglas A. Lind, William G. Marchal, Samuel A. Statistical Techniques in Business & Economics, 15th Edition.

ISBN 0-07-340180-3

Copyright © 2012 by McGraw-Hill Education.

All Rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including without limitation photocopying, recording, taping, or any database, information or retrieval system, without the prior written permission of the publisher.

This authorized Chinese translation edition is jointly published by McGraw-Hill Education and China Machine Press. This edition is authorized for sale in the People's Republic of China only, excluding Hong Kong, Macao SAR and Taiwan.

Copyright © 2015 by McGraw-Hill Education and China Machine Press.

版权所有。未经出版人事先书面许可, 对本出版物的任何部分不得以任何方式或途径复制或传播, 包括但不限于复印、录制、录音, 或通过任何数据库、信息或可检索的系统。

本授权中文简体字翻译版由麦格劳-希尔(亚洲)教育出版公司和机械工业出版社合作出版。此版本经授权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。

版权 © 2015 由麦格劳-希尔(亚洲)教育出版公司与机械工业出版社所有。

本书封面贴有 McGraw-Hill Education 公司防伪标签, 无标签者不得销售。

本书是统计学最优秀的教材之一, 为读者深入领会统计概念的含义提供了最有效的工具。本书编排新颖, 讲解清晰简洁, 结合大量的案例, 使统计技术不再枯燥, 变得生动有趣, 方便读者理解和学习。

本书不仅适用于全国各财经院校统计专业的广大师生, 而且为主修经济、财政、金融、市场营销、会计和管理专业的师生提供了一本应用统计技术的指导手册。

出版发行: 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: 王金强

责任校对: 董纪丽

印刷: 三河市宏图印务有限公司

版次: 2015 年 11 月第 1 版第 1 次印刷

开本: 185mm × 260mm 1/16

印张: 30

书号: ISBN 978-7-111-51734-4

定价: 79.00 元 (附光盘)

ISBN 978-7-89405-830-0 (光盘)

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88379210 88361066

投稿热线: (010) 88379007

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjg@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光/邹晓东



译者序

统计学是对数据进行收集、整理、展示、分析和解释，以帮助人们更为有效地做出正确决策的一门方法论科学。统计学在众多的专业、学科领域中，特别是在商务和经济活动中，都有着重要的应用。只要我们研究的问题具有不确定性，例如政府对国家经济形势的分析和对未来趋势的预测，保险公司对人寿保险费率的确定，一种预防心脏病的新药是否有效及其市场前景等商务和经济问题，均需要运用统计学的方法和技术。统计学是解决商务与经济中各种问题的重要工具。

美国卡罗来纳海岸大学道格拉斯 A. 林德等三位教授合著的《商务与经济统计方法》是统计学科最优秀的教材之一，历经 14 次修订仍畅销不衰，被美国及世界各地众多高等学校广泛采用，是经济、管理类专业的本科生、研究生学习统计学的经典教材或必读参考书。2005 年本书第 11 版中文版在中国出版，得到了广大读者的认可与积极的反响。2008 年我们翻译了本书的第 13 版，2014 年本书第 15 版面市了，我们以最快的速度将新版的中文译本奉献给读者。

本书是为主修经济、财政、金融、会计、市场营销、工商管理以及其他经济、管理类专业的学生提供的一本介绍描述统计和推断统计诸多应用的教材。在介绍各种统计方法在商务与经济中的实际应用时，本版延续了以前版本的风格和特色，尽量避免烦琐的数学推导，采用深入浅出、循序渐进的叙述方法，更多地将重点放在数据和解释的结果上。自始至终与读者保持互动友好的界面，使得统计方法技术不再是枯燥无味的书本知识。书中穿插了大量生动翔实的案例，配有大量的例题和练习，结构严谨、讲解清晰、图表活泼，易于读者理解与掌握。

本书的突出特色是配有大量的例题、自测题、习题以及应用测试，并且它们的背景和数据大都源自诸如美国劳工统计局、美国国土安全局等美国官方网站，以及《今日美国》《商业周刊》等杂志，此外，还包括家得宝、UPS、苹果公司等世界 500 强的实际问题，这些题目具有极强的应用性和可读性。

本书可以作为经济、管理类专业本科生、研究生学习中级统计学的最佳教材，也可以作为从事工商管理和经济分析的各类人员应用统计方法与技术的参考书。读者可以根据时间和需要，有选择地学习有关内容。

本书翻译工作安排如下：叶光翻译第 1~2 章，聂巧平、王梦颖翻译第 3~7 章，聂巧平、刘东红翻译第 8~12 章，刘东红、叶光翻译第 13~19 章。康凯、李波、

周修飞、王晓军也参加了本书的部分翻译工作。黄河、汪小雯、张坤、程晶蓉、冀思轩、鲁大成、刘峻岭、李宾等也对本书翻译定稿提供了帮助。还要特别感谢南开大学冯燕奇教授的指导和支持。

全书译稿由聂巧平和叶光进行整理、总纂和通校。感谢本书的责任编辑王金强，他严谨的工作作风、尽职尽责的工作态度使本书能在最短的时间内与读者见面。

鉴于译者水平，疏漏和差错之处在所难免，恳请读者批评指正。

译者

2015年6月2日于天津商业大学



前言

多年来，本书得到大家的一致赞扬，并深受学生喜爱，这是对我们最大的认可，我们将会继续努力。

《商务与经济统计方法》旨在为主修管理、市场、金融、会计、经济以及工商管理等其他领域的学生，初步介绍描述统计和推断统计的诸多应用。我们针对的主要是商务应用，但同时我们也使用了许多与当前大学生息息相关的练习和例子。你只需具备大学一年级的数学水平，即使以前从没接触过统计学有关的课程，也可以顺利完成本书的学习。

在本书中，我们展示了一个初学者成功学习基础统计学课程所需的每一个步骤。这种循序渐进的方法，有助于提高学生的学习效果，加速学生的接受过程，明显增强学生的学习热情。本书着重于让学生通过基本概念的理解、大量案例的阅读和足够的习题训练，了解统计方法在商务和经济领域的应用。

本书第1版于1967年出版，那时收集相关的商业数据是一件非常困难的工作。现在这种情况已经发生了变化。今天，数据的收集已不再是一个难题。你在食品店购买的食物数量在收银台会被自动地记录下来。电话公司能记录下通话的次数、通话的时间和通话人的电话号码。信用卡公司能够记录卡号、时间和日期以及购买金额信息。医疗设施能自动地监控患者的心率、血压和体温。大量的商业信息会被及时地记录和公布出来。例如，在CNN、《今日美国》和MSNBC的网站上，你都能浏览到延迟不超过20分钟的股票价格。

今天，我们需要的是能够处理大量数据信息的技术。首先，我们会成为信息的主要消费者，从其他人那里获取信息；其次，我们必须能够把大量的信息简化为简明的和有意义的形式，以便能够进行有效的解读、判断和决策。现在，所有的学生都有计算器，大部分学生要么有个人计算机，要么能在校园实验室里使用个人计算机，Excel和MINITAB等统计软件都可以在计算机上应用。在每个章节中，我们也使用了很多屏幕截图，以帮助学生熟悉软件的输出结果。

正是由于计算机和软件的便利性，我们再也不必纠缠于烦琐的数据计算。我们使用许多阐释性例题来代替计算性例题，以帮助学生更好地理解 and 解读统计结果。此外，我们现在还更加重视统计问题的概念内涵。在做出上述改进的同时，我们依然尽自己最大的努力，继续结合有趣的相关案例，来介绍统计学的基本概念。

本版新增的内容

本版的变化，相信对读者非常有用。

- 我们修改和增加了学习目标，使它们变得更加具体，同时我们还将学习目标与章节内容直接联系起来。
- 我们替换了第 1~4 章的主要案例，新的案例包含更多的变量和更多的观测值，与真实商务情境更为贴近。
- 我们在一些章节增加或者改写了部分新内容：
 - 第 7 章新增了关于指数分布的讨论。
 - 重新编排第 9 章，使其内容更加流畅，更利于教学。
 - 重新编排第 13 章，同时增加对回归斜率的假设检验。
 - 第 17 章增加了正态性的图形检验和卡方检验。
- 新的习题和例题使用 Excel 2007 和最新版 MINITAB 的屏幕截图，并扩大了截图尺寸，以保证清晰度。
- 我们仔细复查了每章中间、结尾和复习部分的练习题，同时还增加和修改了很多练习题。你可以从中选择你喜欢的例题，当然你也可以引入新的实例。
- 新添加部分章节号码，使每节的主题更明确，引用更方便。
- 对包含数据文件的练习题，用易于识别的图标特别标出。
- 修正每章结束的数据集习题。棒球数据已经更新到 2009 年最近结束的一季，同时还引入一个新的商业应用，有关 Buena School District 校车的使用与维护。
- 每章引言部分给出大量新的观点。

致谢

《商务与经济统计方法》是众人共同努力的结果：学生、同事、审阅人和麦格劳-希尔公司的工作人员，我们感谢他们。我们想要向调查和讨论团队的参与者，以及所有的审阅人表示诚挚的感谢。

正是他们的建议，以及对上一版和本版原稿的详尽审阅，使得这本教材更加完善。

特别需要感谢的有：Nevada-Reno 大学的 Debra K. Stiver，他审阅了原稿和排好版的校样，并检查了习题以确保准确无误；South-west Tennessee Community 大学的 Joan McGrory 检查了试题库的准确性；南卡罗来纳大学的 Kathleen Whitcomb 教授准备了学习指南；Coastal Carolina 大学的 Samuel Wathen 博士准备了测试题和试题库；南俄勒冈大学的 Rene Ordonez 教授准备了 PowerPoint 演示文档、在线系统 Connect 中的引例，以及 screencam 教程中的大部分内容；Denise Heban 女士和作者共同准备了教师手册。

我们还想要向麦格劳-希尔公司的工作人员表示感谢，包括执行编辑 Steve Schuetz、高级营销编辑 Wanda Zeman、高级项目经理 Diane Nowaczyk，以及那些我们虽然不知道姓名，但为本书做出宝贵贡献的人。



目 录

译者序 前 言

第1章 什么是统计学 1

学习目标	1
引言	1
1.1 为什么学习统计学	2
1.2 什么是统计学	3
1.3 统计学的类型	4
1.4 变量的类型	7
1.5 测量尺度	8
1.6 伦理与统计	12
1.7 计算机应用	12
本章小结	13
习题	13

第2章 描述数据：频数表、频数 分布和图形表示 15

学习目标	15
引言	15
2.1 构造频数表	16
2.2 构造频数分布：定量数据	20
2.3 一个软件例子	23
2.4 相对频数分布	23
2.5 频数分布的图形表示	25
本章小结	30
习题	31

第3章 描述数据：数值度量 33

学习目标	33
------------	----

引言	33
3.1 总体均值	34
3.2 样本均值	35
3.3 算术平均数的性质	36
3.4 加权平均数	37
3.5 中位数	38
3.6 众数	40
3.7 软件解法	42
3.8 均值、中位数和众数的相对位置	42
3.9 几何平均数	44
3.10 为什么要研究离散程度	45
3.11 离散程度的度量	46
3.12 软件解法	52
3.13 标准差的解释与应用	53
3.14 分组数据的均值和标准差	55
3.15 伦理和结果报告	57
本章小结	58
符号含义	59
习题	59

第4章 描述数据：数据展示和 数据开发 61

学习目标	61
引言	61
4.1 点状图	61
4.2 茎叶图	63
4.3 位置度量	66
4.4 偏度	72
4.5 描述两个变量之间的关系	75
本章小结	78

符号含义	79	本章小结	154
习题	79	习题	155
第1~4章复习	81	第5~7章复习	157
第5章 概率论概述	87	第8章 抽样方法和中心极限	
学习目标	87	定理	163
引言	87	学习目标	163
5.1 概率	88	引言	163
5.2 分配概率方法	90	8.1 抽样方法	164
5.3 几种计算概率的法则	93	8.2 抽样误差	170
5.4 列联表	99	8.3 样本均值的抽样分布	171
5.5 树形图	101	8.4 中心极限定理	174
5.6 贝叶斯定理	103	8.5 样本均值抽样分布的应用	179
5.7 计数原理	106	本章小结	181
本章小结	110	符号含义	182
符号含义	111	习题	182
习题	112	第9章 估计和置信区间	185
第6章 离散型概率分布	114	学习目标	185
学习目标	114	引言	185
引言	114	9.1 均值的点估计	186
6.1 概率分布	115	9.2 总体均值的置信区间	187
6.2 随机变量	116	9.3 比例的置信区间	198
6.3 概率分布的均值、方差和标准差	117	9.4 选择适当的样本容量	200
6.4 二项概率分布	120	9.5 有限总体修正因子	203
6.5 超几何分布	126	本章小结	205
6.6 泊松概率分布	129	习题	205
本章小结	132	第8~9章复习	207
习题	133	第10章 单样本假设检验	210
第7章 连续型概率分布	135	学习目标	210
学习目标	135	引言	210
引言	135	10.1 假设	211
7.1 均匀概率分布族	136	10.2 假设检验	211
7.2 正态概率分布族	138	10.3 假设检验五步法	212
7.3 标准正态概率分布	139	10.4 单侧与双侧显著性检验	215
7.4 二项分布的正态近似	148	10.5 总体均值的检验：已知总体	
7.5 指数分布族	152	标准差	216

10.6	假设检验中的 p -值	219	引言	289	
10.7	总体均值的检验: 总体标准差未知	221	13.1	相关分析	290
10.8	关于比例的检验	227	13.2	相关系数	291
10.9	第 II 类错误	229	13.3	相关系数的显著性检验	297
	本章小结	232	13.4	回归分析	299
	符号含义	233	13.5	斜率的显著性检验	304
	习题	233	13.6	评估回归方程的预测能力	306
第 11 章	两样本假设检验	235	13.7	预测区间估计	309
	学习目标	235	13.8	变换数据	313
	引言	235		本章小结	316
11.1	两样本假设检验: 独立样本	236		符号含义	317
11.2	比例的两样本检验	240		习题	317
11.3	总体均值的比较: 总体标准差未知 (合并 t 检验)	243	第 14 章	多元回归与相关分析	321
11.4	两样本的假设检验: 相依样本	250		学习目标	321
11.5	相依样本与独立样本的比较	252		引言	321
	本章小结	255	14.1	多元回归分析	322
	符号含义	255	14.2	评价多元回归方程	325
	习题	255	14.3	多元线性回归的推断	329
第 12 章	方差分析	259	14.4	评价多元回归的假定	334
	学习目标	259	14.5	定性自变量	339
	引言	259	14.6	具有交互作用的回归模型	341
12.1	F 分布	259	14.7	逐步回归	343
12.2	比较两总体的方差	260	14.8	多元回归回顾	346
12.3	ANOVA 的假定	263		本章小结	350
12.4	ANOVA 检验	264		符号含义	351
12.5	关于成对处理均值的推断	270		习题	351
12.6	双因素方差分析	272		第 13~14 章复习	355
12.7	具有交互作用的双因素方差分析	276	第 15 章	指数	359
	本章小结	280		学习目标	359
	符号含义	281		引言	359
	习题	281	15.1	简单指数	360
	第 10~12 章复习	284	15.2	为什么把数据转换成指数	361
第 13 章	线性回归与相关	289	15.3	指数的构造	362
	学习目标	289	15.4	未加权指数	363
			15.5	加权指数	364
			15.6	价值指数	368
			15.7	特殊用途指数	369

15.8 消费价格指数	373	引言	425
15.9 调整基期	376	18.1 符号检验	426
本章小结	378	18.2 相依样本的威尔科克森符号秩检验	431
习题	378	18.3 独立样本的威尔科克森秩和检验	435
第16章 时间序列与预测	380	18.4 克鲁斯卡尔-沃利斯检验：秩方差分析	437
学习目标	380	18.5 秩相关	441
引言	380	本章小结	444
16.1 时间序列的构成	381	符号含义	445
16.2 移动平均	383	习题	445
16.3 加权移动平均	385	第17~18章复习	446
16.4 线性趋势	387	第19章 统计过程控制与质量管理	449
16.5 最小二乘法	388	学习目标	449
16.6 非线性趋势	390	引言	449
16.7 季节波动	392	19.1 质量控制简史	449
16.8 消除数据的季节影响	396	19.2 差异产生的原因	452
16.9 杜宾-瓦特森统计量	399	19.3 诊断图	453
本章小结	403	19.4 质量控制图的用途和类型	455
习题	403	19.5 受控和失控情形	459
第15~16章复习	406	19.6 属性控制图	461
第17章 非参数方法：检验	408	19.7 验收抽样	466
学习目标	408	本章小结	468
引言	408	符号含义	469
17.1 拟合优度检验：期望频数相等	408	习题	469
17.2 拟合优度检验：期望频数不等	412	附录A 数据集[⊖]	475
17.3 卡方的局限性	413	附录B 表格	475
17.4 检验一组数据是否来自正态总体	415	附录C 奇数题号的各章习题答案	475
17.5 正态检验的图形和统计方法	417	附录D 奇数题号的复习题答案	475
17.6 列联表分析	419	附录E 应用测试答案	475
本章小结	423		
符号含义	423		
习题	423		
第18章 非参数方法：分级数据分析	425		
学习目标	425		

⊖ 本书附录见光盘。为方便读者学习，本书光盘含有大量辅助资料，同时，我们将全书各章偶数号习题一并放入。

第1章

什么是统计学



学习目标

1. 列举统计学的用途。
2. 了解描述统计学和推断统计学的差异。
3. 理解样本和总体的区别。
4. 区分定性变量和定量变量。
5. 区分离散变量和连续变量。
6. 分辨数据的测量尺度。

巴诺书店最近开始销售 Nook。利用这个设备，你可以下载逾 1 500 本电子书在显示屏上阅读，而不必购买实体书籍。假如你现在掌握有位于加利福尼亚州河滨市 Market Commons Mall 的巴诺书店上个月 Nook 的日销售额数据，哪种情形下可以将这些信息视为一个样本？哪种情形下同样的数据却应该被视为一个总体？

引言

100 多年前，英国作家、历史学家 H. G. 威尔斯 (H. G. Wells) 曾经说过，将来有一天，定量分析对于有效地行使公民的权利和义务而言，将如同阅读能力一样必不可少。他没有提到商业，因为当时工业革命才刚刚开始。威尔斯先生无疑是正确的。随着“商业经验”、对未来的“精心预测”和“直觉”成为成功管理者的三大关键特质，单就管理过程中的量化决策而言，今天的商业问题已经变得十分复杂。

统计学是进行决策的工具之一，不仅商业人士如此，日常生活中我们也会用到一些统计概念。例如，新的一天开始了，你打开热水器让水先流一会儿，然后用手试试水温，决定是否需要加些热水或冷水，如果水温刚好合适，你就开始淋浴。又比如，假如你在好市多超市里，准备买一张冻比萨饼，每个比萨饼商家都有自己的摊位，每个摊位都会提供小块的比萨饼以供品尝，经过品尝你才能决定是否要在这个摊位购买。上面的两个例子中，决定的做出和行为的选择不建立在抽取样本的基础之上。

商业活动中也经常会遇到类似的问题。家乐氏公司必须保证 25.5g 盒装的葡萄干小麦片的平均重量与标签相符。为此，他们设定一个比标签说明稍高的“目标”重量，对每个装满食品的包装盒进行称重，称重机器每小时都会报告这一时间段内包装盒中食品重量的分布，以及由于重量达不到标签说明而被“开除”的数目。质检部门也会随机从生产线抽取样本，并检查产品质量和盒子重量，如

果平均产品重量与目标重量显著不符, 或者被清除的包装盒过多, 就需要对整个过程进行调整。

作为一名商学或经济学专业的学生, 你需要具备组织、分析和处理数据并提供信息的基本知识和技能。本书将告诉你一些基本的统计技术和方法, 使你进行个人和商业决策的能力得到更好的开发。

1.1 为什么学习统计学

浏览你所在大学的课程表, 你会发现在许多学院统计学都是必修课。为什么会是这样? 工学院、文学院的心理学和社会学系以及商学院开设的统计学课程又有哪些不同? 事实上, 课程内容基本上是相同的, 最大的区别在于所使用的案例。商学院的学生对利润、工作时间和工资等问题感兴趣, 心理学系的学生对测试分数感兴趣, 而工学院的学生感兴趣的则是特定机器可以生产多少件产品。不过, 这三个学院的学生对数据的典型值和离散程度却有着共同的兴趣。此外, 对数学基础的要求也有所不同, 工学院的统计学课程通常需要计算, 商学院和教育学院则更多地强调应用。如果你已经完成了高中的代数课程, 就完全具备了本书所需要的数学基础。

究竟为什么这么多专业都需要学习统计学呢? 其首要原因在于数字化信息随处可见。翻开报纸 (USA Today)、新闻杂志 (Time、Newsweek、U. S. News and World Report)、商业杂志 (Business Week、Forbes)、大众杂志 (People)、女性杂志 (Ladies Home Journal or Elle)、体育杂志 (Sports Illustrated、ESPN The Magazine), 你将完全处于数字化信息的狂轰滥炸之中。

下面就是几个例子:

- 以 1982 ~ 1984 年美元计, 从 2009 年 1 月到 2010 年 1 月周薪的平均增幅是 8.32 美元。
- 2010 年 1 月, 每个家庭信用卡的平均欠款额为 7394 美元, 相对于 2009 年 7 月的 7801 美元有所下降。一项联邦储备委员会于 2010 年的调查显示, 75% 的美国家庭至少持有一张信用卡。
- 下表汇总了波音公司于 2006 ~ 2009 年生产的商用飞机数量。

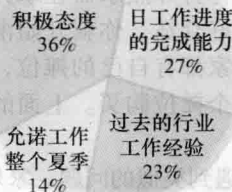
波音飞机销售额						
飞机型号						
年份	737	747	767	777	787	总计
2006	733	72	8	77	160	1 050
2007	850	25	36	143	369	1 423
2008	488	4	29	54	94	669
2009	197	5	7	30	24	263

- 《今日美国》(www.usatoday.com) 的“Snapshots” (实录) 栏目会刊登不同的研究组织、基金和联邦政府的调查结果。下图汇总了招聘方在雇用季节性员工时所看重的因素。

《今日美国》快照

招聘季节性员工时最看重“积极态度”

招聘季节性员工时最看重什么?



第二个学习统计学的原因在于，对我们日常生活产生影响的各种决策经常需要利用统计技术，也就是说，统计技术已经影响到个人生活的质量。下面有几个例子：

- 保险公司通过统计分析来确定家庭、汽车、人寿和健康保险的费率。利用网页 www.ssa.gov/OACT/STATS/table4cb.html（注意字母的大小写）上的一些表格，可以估算出20岁女性的平均剩余寿命为60.25年，80岁女性的平均剩余寿命为4.56年，50岁男性的平均剩余寿命则为27.87年。在这些关于未来寿命预期的基础上，人寿保险的保费就可以计算出来。
- 环境保护组织（Environmental Protection Agency）非常关注Erie湖的水质问题。他们定期采集湖水样本，确定湖水污染程度，以将水质控制在合格水平。
- 医学研究者对使用不同药物和不同治疗方法的病人的治愈率进行研究。例如，对膝盖受损的病人采用手术治疗或物理治疗的效果如何？如果你每天服用一片阿司匹林，患心脏病的风险是否会因此而降低？

第三个学习统计学的原因在于，统计方法能够帮助你了解决策是怎样做出的，并使你更加清楚这些决策会对你产生什么样的影响。

不论选择了什么样的工作，你都会发现自己面临着许许多多的决策问题，而掌握一种数据分析技术对此将大有裨益。为了做出明智的决策，你需要：

- (1) 确定现有信息是否充分，是否需要额外的信息。
- (2) 如果需要收集额外信息，确保不要被这些信息误导。
- (3) 将信息以有效且直观的形式加以汇总。
- (4) 对可得信息进行分析。
- (5) 得出结论并进行推断，同时对错误结果的风险进行评价。

本书所讲述的各种统计方法将为你提供进行决策的框架。

总而言之，至少有3个原因让我们学习统计学：①数据随处可见；②对我们日常生活产生影响的各种决策经常需要利用统计技术；③无论从事什么样的工作，你都不可避免地要与数据打交道并做出很多专业的决策。对统计方法的了解将会使你的决策更加有效。

1.2 什么是统计学

日常生活中，我们经常遇到统计（statistics）这个词。我们应该怎样定义它呢？它具有两层含义。在比较普遍的用法中，统计指的是数字化信息。例如，大学毕业生的平均起始工资；上一年度因为酒精中毒而死亡的人数；从昨天到今天，道琼斯工业平均指数的变化，以及2010赛季芝加哥队本垒打的次数。上述例子中，统计是一些数值或者百分比。其他的例子还有：

- 在美国，一般汽车每年行驶11 099英里，公共汽车每年行驶9 353英里，卡车每年行驶13 942英里。在加拿大，相应的数据为汽车10 371英里，公共汽车19 823英里，卡车7 001英里。
- 等待技术支持的时间平均为17分钟。
- 自1945年以来，商业周期的平均长度为61个月。

这些都是统计的实例。统计就是一组数字化信息的集合。

实践中的统计 1-1

我们请你注意一个具有特色的标题——实践中的统计。仔细阅读每一篇文章，你可以充分了解统计学在管理、经济、护理、法律实务、体育及其他领域中的广泛应用。

- 2009年,《福布斯》公布的美国富豪排名榜中,微软公司的创始人威廉·盖茨名列第一。据估计,他的净资产约为590亿美元(www.forbes.com)。
- 2009年,按总收益排名,美国最大的四家公司依次为沃尔玛公司、埃克森-美孚公司、雪佛龙公司和通用汽车(www.forbes.com)。
- 在美国,一般的高中生毕业生一生的总收入为120万美元;获得学士学位的大学毕业生一生的总收入为210万美元;获得硕士学位的研究生一生的总收入为250万美元(usgovinfo.about.com/library/weekly/aa072602a.htm)。

统计信息还经常可以表示为图形的形式,而在吸引读者注意力和描述大量数据信息方面,图形是非常有用的。例如,图1-1给出在美国超市的点和薯片类食品中,Frito-Lay公司的销售量及所占的市场份额。你只需快速一瞥,就可以看出薯片的销售总量为8亿磅,而Frito-Lay公司的市场份额为64%。从图中还可以看出,Frito-Lay公司的玉米饼市场份额为82%。

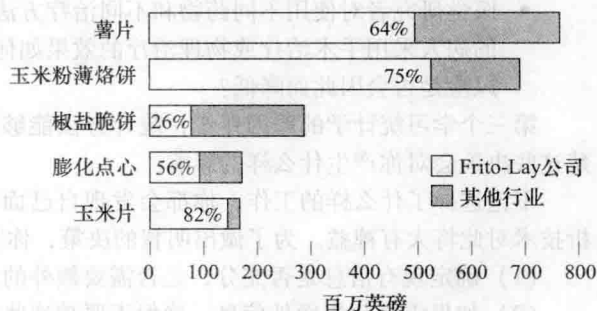


图1-1 美国超市的点心类食品中,Frito-Lay公司的销售量和市场份额

正如我们将在书中探讨的那样,统计学所涉及的范畴要比单纯地收集和公布数字化信息广泛得多。我们将统计学定义为:对数据进行收集、整理、展示、分析和解释,以帮助人们更有效地进行决策的科学。

如定义所述,收集数据是研究问题的第一步。所收集的数据必须按照一定的方式加以整理,或者像图1-1那样用图的形式加以展示。只有把数据整理好以后,我们才能对它们进行分析和解释。这里有一些需要收集数据的例子。

- Merrill Lynch公司的研究人员在做出“买入”或“卖出”某只股票的建议之前,要对这只股票的许多方面进行评价。他们收集公司过去的销售数据,利用这些数据对公司未来的收益进行估计。此外,他们在提出建议之前,还要考虑其他各种问题,如世界范围内对公司产品的需求情况、竞争激烈程度,以及新的联合管理合同的影响等。
- Colgate-Palmolive公司是一家肥皂生产厂商,该公司的市场部要负责对新推出的一系列具有葡萄、橘子和菠萝等不同水果香味的洁面香皂的潜在盈利性进行评估。在得出最后决策之前,市场部主管将选择几个市场进行试验,可能会在堪萨斯州的Topeka和佛罗里达州的Tampa做广告并试销。根据这两个地区的试销效果,Colgate-Palmolive公司决定是否要将这个产品系列推广到全国范围内。
- 管理者需要对产品和服务的质量做出判断。例如,顾客如果遇到了不能解决的软件问题,会打电话给软件公司,寻求技术建议。此时可以利用技术顾问回电之前顾客所需等待的时间来评价客户服务的质量。软件公司可能将一分钟确定为反应时间的目标,而后通过收集和分析反应时间的相关数据,以决定一周中每天或者一天中每个时段的平均反应时间是否存在差异。如果反应时间增加了,管理者就要在特定的时间段安排更多的技术顾问。

1.3 统计学的类型

关于统计学的研究通常分为两类:描述统计学和推断统计学。

1.3.1 描述统计学

前文给出的统计学定义中，“整理、展示、分析……数据”方面的内容通常被称为**描述统计学** (descriptive statistics)。

例如，美国政府报告美国人口总数 1960 年为 179 323 000 人，1970 年为 203 302 000 人，1980 年为 226 542 000 人，1990 年为 248 709 000 人，2000 年为 265 000 000 人，2010 年为 308 400 000 人。上述信息都属于描述统计。如果我们计算从一个十年到下一个十年人口的百分比增长，这依然属于描述统计。但如果我们利用这些数据预测美国 2010 年的人口总数或者从 2000 ~ 2010 年美国人口的百分比增长，这就不再属于描述统计了。为什么？因为这些统计量不是对过去人口数据的描述和概括，而是对未来人口的预测。下面列出的是其他一些关于描述统计的例子：

描述统计学

以某种便于提供信息的方式对数据进行整理、概括和展示的方法。

- 美国洲际高速公路的总长度为 46 837 英里。州际高速公路系统仅占美国全部道路的 1%，却承担了 20% 以上的交通量。最长的一条州际公路是 1-90 号公路，从波士顿一直延伸到西雅图，全长 3 081 英里；最短的一条是纽约市的 1-878 号公路，全长仅为 0.70 英里。阿拉斯加州没有一条州际公路，得克萨斯州则拥有 3 232 英里的州际公路，总长度是全美最长的；纽约拥有 28 条高速公路，线路是全美最多的。
- 2010 年，人均购买传统的情人节商品的支出是 103.00 美元，比 2009 年增长了 0.50 美元。前些年，男性在节日的花费几乎是女性花费的 1 倍，男性平均支出 135.35 美元购买礼物，而女性只花费 72.28 美元。家庭宠物也能感受到情人节的爱意，人们在宠物朋友上的平均花费为 3.27 美元，比去年增加了 2.17 美元。

对于大量未经整理的数据，如人口普查数据、成千上万的计算机程序员的周工资数据，以及 2 000 位注册选民对美国总统选举的个人选择数据，其价值十分有限。但可以使用统计技术对这些数据进行整理，使它们具有一定的实际价值。某些数据可以表示为**频数分布** (frequency distribution) 的形式 (具体步骤参见第 2 章)。各种各样的**图表** (charts) 也可以用来描述数据，第 4 章将对一些基本的图表形式进行介绍。

关于数据中心位置的度量 (如均值) 是对一组数据中心值的描述。还有一些统计度量是用来描述数据在其平均值周围紧密程度的。这些关于数据集中趋势和离散程度的度量将在第 3 章进行讨论。

1.3.2 推断统计学

第二类统计学是**推断统计学** (inferential statistics)，或称为**统计推断** (statistical inference)。推断统计学主要关注的是如何从总体中抽取的样本去发现总体的某些特征。例如，最近的一次调查显示，中学高年级学生中只有 46% 的学生能够解决涉及分数、小数和百分比的问题，只有 77% 的学生能够正确计算餐馆菜单上色拉、汉堡、油炸食品和可乐的总成本。由于上述结果是在样本数据基础上对总体 (所有中学的高年级学生) 的推断，我们称之为推断统计学。你可以将推断统计学看作在样本信息基础上对总体水平的“最优猜测”。

推断统计学

在样本基础上对总体的某些性质进行估计的方法。

总体

所研究的个人、事物或者来自个人、事物度量值的全部集合。

注意，在推断统计学的定义中所出现的**总体** (population) 和**样本** (sample) 这两个词。我们经常提到美国 3.088 亿的常住人口或者中国的 13.361 亿人口。虽然人口和总体的英文单词相同，都

样本

所研究总体的一部分。

是 *population*，但在统计学中，总体有着更广泛的含义。总体可以由个人 (individuals) 组成，例如，犹他州立大学的注册学生、学习初级会计学的所有学生，或者《财富》500 强企业的 CEO。总体也可能由事物 (objects) 组成，例如，Cooper Tire and Rubber 公司在俄亥俄州 Findlay 的工厂中生产的所有 Cooper G/T 型轮胎、10 月底 Lorrance 塑胶公司的应收账款，以及 2010 年第一季度 State Farm Insurance 公司东北地区办事处接受的汽车险申请。人们感兴趣的度量值可能是第一次初级会计学考试所有学生的分数、Cooper Tire 轮胎的胎面磨损、Lorrance 塑胶公司的应收账款总额，或者是 State Farm Insurance 公司的汽车险申请数量。因此，在统计学意义上“总体”这个词并不一定指人。

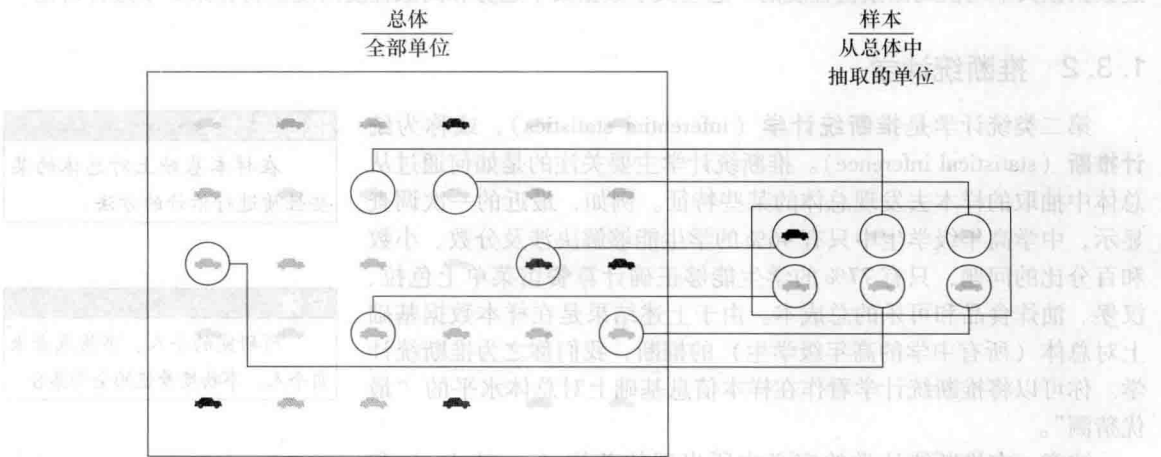
要推断总体的某些特征，我们通常需要从总体中抽取一个样本。

为什么要抽取样本，而不是对总体的每个单位都进行研究呢？由于选举前联系数百万选民的成本高得无法承受，从中抽取一些选民作为样本就显得十分必要；检测小麦的含水量会毁坏小麦，因此抽取样本进行检测就势在必行；如果品酒师尝遍了全部的葡萄酒，就没有产品可供销售了；对少数的海洋生物学家而言，捕捉海洋中全部海豹并进行标记，实际上是根本不可能实现的（这些和其他一些需要进行抽样的原因将在第 8 章讨论）。

正如下面的例子表明的，利用样本对总体特征的研究在商业、农业、行政和政府管理领域有着广泛的应用。

- 电视网请尼尔森 (Nielsen) 和其他组织对电视观众的偏好进行抽样调查，以便于监控各电视台节目的受欢迎程度。例如，在由 800 名黄金时段的电视观众组成的一个样本中，320 名或 40% 的观众表示上周他们收看的是 FOX 电视台播出的《美国偶像》。调查显示的节目受欢迎程度等级将用于制定广告费率或者作为取消某些节目的依据。
- Gamous and Associates 是一家公立会计公司，该公司正在对 Pronto Printing 公司进行审计。会计公司首先随机抽取 100 张发票组成一个样本，并检查每张发票是否准确，结果发现其中 5 张发票至少出现了一个错误。因此会计公司估计所有发票中将有 5% 的发票至少出现一个错误。
- 从 4 年制营销专业的毕业生中随机抽取 1 260 人组成一个样本，调查结果表明他们的平均起始工资为 42 694 美元。因此，我们可以估计所有 4 年制营销专业毕业生的平均起始工资为 42 694 美元。

样本和总体之间的关系可以用下图表示。例如，我们想估计 SUV 每加仑汽油的平均行驶里程。从总体中抽取 6 辆 SUV。这 6 辆汽车每加仑汽油行驶里程的平均值就是对总体平均行驶里程的估计。



下面是一道自测题。每一章中都会穿插这样的自测题，完成这些题目可以帮助你检查自己是否理解前面讲述的内容。答案和解答方法在每章结尾给出。建议读者认真完成每一道题目后再核对答案。