

# BUSINESS STATISTICS

A FIRST COURSE, THIRD EDITION



# 商务统计学

## ——初级教程 (第3版)

戴维·M·莱文 (David M. Levine)  
蒂莫西·C·克雷比尔 (Timothy C. Krehbiel) 著  
马克·L·贝伦森 (Mark L. Berenson)  
李鹏宇 许红燕 等译 胡大源 李冰 审校



中国人民大学出版社

# BUSINESS STATISTICS

A FIRST COURSE, THIRD EDITION

# 商务统计学

## ——初级教程（第3版）

戴维·M·莱文 (David M. Levine) 著  
蒂莫西·C·克雷比尔 (Timothy C. McHugh) 著  
马克·L·贝伦森 (Mark L. Berenson) 著  
李鹏宇 许红燕 等译 李泳 审校

## 图书在版编目(CIP)数据

商务统计学:初级教程:第3版/莱文等著;李鹏宇等译.

北京:中国人民大学出版社,2004

ISBN 7-300-05727-6/F·1802

I. 商…

II. ①莱… ②李…

III. 商务统计学—教材

IV. F712.3

中国版本图书馆 CIP 数据核字(2004)第 067220 号

## 商务统计学——初级教程(第3版)

戴维·M·莱文 蒂莫西·C·克雷比尔 马克·L·贝伦森 著

李鹏宇 许红燕 等译 胡大源 李冰 审校

---

出版发行 中国人民大学出版社

社 址 北京中关村大街 31 号

电 话 010-62511242(总编室) 010-62511239(出版部) 010-82501766(邮购部)  
010-62514148(门市部) 010-62515195(发行公司) 010-62515275(盗版举报)

邮政编码 100080

网 址 <http://www.crup.com.cn>  
<http://www.ttrnet.com>(人大教研网)

经 销 新华书店

印 刷 中煤涿州制图印刷厂

开 本 787 毫米×1092 毫米 1/16

版 次 2004 年 10 月第 1 版

印 张 54.75 插页 1

印 次 2004 年 10 月第 1 次印刷

字 数 1 329 000

定 价 69.00 元

---

版权所有 侵权必究 印装差错 负责调换

## 译者序

什么是商务统计学？一种常见的定义为：商务统计学是对一组准确的观察数据进行概括、整理得出结论用以支持企业经营管理决策的原理和方法。这一定义的关键在于准确的观察、科学的整理和正确的决策。

几年前，我曾同一群来访的美国商学院的 MBA 学生座谈，问到哪门课他们最不愿意学，回答正是商务统计。前不久，听一位美国商学院的教授讲，商务统计已被美国最有影响力的商学院资质认定机构 AACSB 定为 MBA 学生的必修课程，由此可见商务统计课有两个特点：一是抽象难懂；二是在经营管理决策过程中必不可少。

商务统计之所以被许多 MBA 学生认为抽象难懂，不外乎是由于下列两个原因。首先，MBA 是一个实战性很强的职业学位，在 MBA 的主要课程中，涉及数量分析方法的课程并不多，除了运营管理中有时会讲些运筹学的基本方法以外，商务统计恐怕是惟一大量涉及数学知识的课程。更何况近十年来以优化数学方法为主要内容的运筹学在美国商学院有越讲越少的趋势，商务统计对于数学基础较为薄弱的那部分 MBA 学生而言，就越发“令人讨厌”。

如果说第一个原因主要是由于部分学生的教育背景薄弱造成的，那么第二个原因则主要在于教师。对于一个善于总结经验的教师来说，为 MBA 学生讲授商务统计课程或许会成为一件趣事。2001 年《青年文摘》上曾经刊载了一篇杰温·汤姆的文章，题为“第二种思维方式”，讲的是一位数学天份不足的学生在善于揣摩学生心理的教授的帮助下，如何学好统计学的故事。尽管这篇短文的情节似乎有些夸张，结果颇为戏剧性，但是细细品味之下，也不无道理。统计学的数学基础主要是概率论。而概率论是研究不确定性的学问，与传统的数学学科，如代数、几何、微积分等，就其基本思路而言并不相同。只要学习方法得当，加之多花上些功夫，数学基础薄弱的学生仍有很大的可能取得好成绩。

在过去的 9 年间，我先后为数以千计的北大中国经济研究中心双学位学生和北大国际 MBA 学生讲授商务统计课程。选用 Berenson, Levine 和 Krehbiel 的这本《商务统计基础》作为教材，最主要的考虑就是这本书的实用性强，有大量的应用案例和不断更新的练习题。

理解商务统计的最佳途径莫过于应用。要学好商务统计，透彻地理解和记住一些典型应用案例是一个行之有效的学习方法。实例越生动，越令人信服，就越容易给人留下深刻的印象，从而使学生免受死记硬背之苦，更重要的是，对 MBA 学生来说，学习商务统计的目的并非为了理论研究，而是为了解决企业经营管理中的实际问题。与领军作战相似，常常有这样的情形：面对突然发生的事件，无论怎样处置都难免会有失败的风险。此时，并没有足够的时间进行理论推导或模型演算，决策者必须当机立断，迅速做出决定。在这种情况下，总结

类似环境下的案例便大有用处。善于触类旁通的决策者可以根据掌握的信息与经历的实例进行类比，分析异同，加以判断，做出决策。这正是哈佛商学院案例教学法的精妙之处。

与国内众多统计学教材相比，这本商务统计教材并没有多少数学推导。从传统的学术观念来看，算不上系统严谨。在系统严谨与实用易懂不可兼得的权衡之下，作者选择了后者。因此，这本书应该更适用于那些有一定实践经验、注重学以致用用的学生。

这本书在过去十年中数次再版的一个主要原因就是为了跟上计算机软件和互联网技术发展的时代步伐。书中不断更新对于各种常用计算机统计软件的示例，一方面可使学生免去复杂数学计算之苦，另一方面也帮助学生在学商务统计知识的同时提高计算机软件的使用能力。

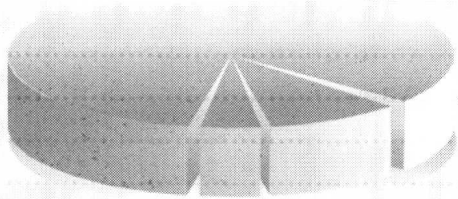
对于每位拿到这本书的读者来说，这本书有一个显而易见的缺点——篇幅太多，对于在读的 MBA 学生，这也许算不上什么大问题，任课老师自会选择需要讲授的章节，其他部分则只是作为参考，以供有特殊需要的学生阅读。这种情形在美国的大学和研究生教材中是很常见的。

对于自学为主的读者，大可不必被这本书的篇幅所吓倒。事实上，这本书由浅入深、循序渐进的表述方法是很适合自学的。此外，书中大量应用案例不但增加了这本教材的可读性，而且帮助读者从不同的角度了解美国的商务环境。书中在每一章后面还在对该章总结概括的基础上做出了总结图表，以便读者提纲挈领地把握该章的内容。本书仅选取了原书的附录 A、E、F，删去了附录 B、C、D，附录顺序仍依照原书。

该书的翻译工作历时较长，我的学生李冰、许红燕、崔成儿、赵鹏飞、李鹏宇、林莉尔、张彤等参与了部分翻译工作，此外微软中国研究中心的软件工程师马一帆对书中涉及的计算机术语进行了校对，在此我一并向他们表示感谢。

胡大源

2004年9月



# 目 录

<b>第 1 章 概述与数据收集</b> .....	(1)
1.1 经理人为什么需要了解统计学 .....	(2)
1.2 现代统计学的发展 .....	(4)
1.3 本书的读者需要了解哪些有关微软 Excel 和 PHS tat, Minitab 或者 SPSS <sup>®</sup> 的知识 .....	(5)
1.4 为什么需要数据 .....	(6)
1.5 数据的来源 .....	(7)
1.6 数据的类型 .....	(12)
1.7 调查研究的设计 .....	(15)
1.8 不同类型的抽样方法 .....	(18)
1.9 评价调查的价值 .....	(25)
案例研究——校友会调查 .....	(36)
附录 1.1 Windows 系统基本操作 .....	(38)
附录 1.2 Microsoft Excel 简介 .....	(40)
附录 1.3 Minitab 简介 .....	(53)
附录 1.4 SPSS <sup>®</sup> 简介 .....	(55)
<b>第 2 章 用图表演示数据</b> .....	(58)
2.1 数值数据整理 .....	(59)
2.2 数值数据的图表演示 .....	(64)
2.3 两变量数值数据的图形演示 .....	(74)
2.4 属性数据的图表演示 .....	(79)
2.5 两变量属性数据的图表演示 .....	(87)
2.6 追求卓越的图表演示 .....	(92)

案例研究——Springville Herald .....	(112)
附录 2.1 使用 Microsoft Excel 制作图表 .....	(116)
附录 2.2 使用 Minitab 制作图表 .....	(126)
附录 2.3 使用 SPSS®制作图表 .....	(131)
<b>第 3 章 数值描述度量 .....</b>	<b>(138)</b>
3.1 数值数据及其特征 .....	(139)
3.2 集中趋势、变异程度和分布形状的度量 .....	(140)
3.3 探索性数据分析 .....	(160)
3.4 计算总体的描述性概括度量 .....	(167)
3.5 相关系数 .....	(171)
3.6 描述性数值度量的缺陷和道德问题 .....	(177)
案例研究——Springville Herald .....	(189)
附录 3.1 应用 Microsoft Excel 计算描述性统计量 .....	(189)
附录 3.2 应用 Minitab 计算描述性统计量 .....	(195)
附录 3.3 应用 SPSS®计算描述性统计量 .....	(197)
<b>第 4 章 概率论基础 .....</b>	<b>(201)</b>
4.1 概率论基础 .....	(202)
4.2 条件概率 .....	(212)
4.3 道德问题和概率 .....	(219)
附录 4.1 使用 Microsoft Excel 计算基础概率 .....	(224)
<b>第 5 章 概率分布 .....</b>	<b>(226)</b>
5.1 离散随机变量的概率分布 .....	(227)
5.2 协方差及其在金融方面的应用 .....	(231)
5.3 二项分布 .....	(237)
5.4 正态分布 .....	(245)
5.5 检验正态分布的假设 .....	(261)
案例研究——Springville Herald .....	(277)
附录 5.1 Microsoft Excel 在协方差和概率分布中的应用 .....	(279)
附录 5.2 利用 Minitab 计算概率分布 .....	(283)
<b>第 6 章 抽样分布和置信区间估计 .....</b>	<b>(287)</b>
6.1 抽样分布 .....	(288)
6.2 置信区间估计绪论 .....	(301)
6.3 总体均值的置信区间估计 ( $\sigma$ 已知) .....	(302)

6.4	总体均值的置信区间估计 ( $\sigma$ 未知)	(306)
6.5	比率的置信区间估计	(314)
6.6	样本容量的确定	(317)
6.7	置信区间估计和道德问题	(325)
	案例研究——Springville Herald	(334)
附录 6.1	应用 Microsoft Excel 获取抽样分布、置信区间, 确定样本容量	(338)
附录 6.2	应用 Minitab 计算抽样分布和置信区间估计	(342)
附录 6.3	应用 SPSS <sup>®</sup> 计算置信区间估计	(346)
<b>第 7 章</b>	<b>假设检验基础: 单样本检验</b>	<b>(347)</b>
7.1	假设检验方法论	(348)
7.2	关于均值的假设的 Z 检验 ( $\sigma$ 已知)	(354)
7.3	单侧检验	(361)
7.4	均值的 $t$ 假设检验 ( $\sigma$ 未知)	(365)
7.5	比率的 Z 假设检验	(372)
7.6	假设检验的潜在缺陷和道德问题	(377)
附录 7.1	使用 Microsoft Excel 进行单样本假设检验	(385)
附录 7.2	使用 Minitab 进行单样本假设检验	(388)
附录 7.3	使用 SPSS <sup>®</sup> 进行均值的单样本假设检验 ( $\sigma$ 未知)	(389)
<b>第 8 章</b>	<b>两个或多个数值数据样本的假设检验</b>	<b>(391)</b>
8.1	两个独立样本的比较: 两个均值差异的检验	(392)
8.2	两个方差差异的 $F$ 检验	(403)
8.3	两个相关样本的比较: 均值差异检验	(411)
8.4	完全随机化的设计: 单向方差分析	(421)
	案例研究——检验圆珠笔的营销与促销	(450)
	案例研究——Springville Herald	(452)
附录 8.1	使用 Microsoft Excel 进行双样本或多样本数值数据假设检验	(455)
附录 8.2	使用 Minitab 进行双样本或多样本数值数据假设检验	(462)
附录 8.3	使用 SPSS <sup>®</sup> 进行单个或多个类型数据样本的假设检验	(465)
<b>第 9 章</b>	<b>对两个或多个属性数据样本的假设检验</b>	<b>(468)</b>
9.1	两个比率差异的 Z 检验	(469)
9.2	两个比率差异的 $\chi^2$ 检验	(475)
9.3	多个比率的 $\chi^2$ 检验	(483)
9.4	独立性的 $\chi^2$ 检验	(490)
	案例研究——Springville Herald	(505)



附录 9.1 使用 Microsoft Excel 对来自两个或多个样本的属性数据进行假设检验 .....	(507)
附录 9.2 使用 Minitab 对来自两个或多个样本的属性数据进行假设检验 .....	(510)
附录 9.3 使用 SPSS®对来自两个或多个样本的属性数据进行假设检验 .....	(511)
<b>第 10 章 一元线性回归 .....</b>	<b>(514)</b>
10.1 回归模型的类型 .....	(515)
10.2 建立一元线性回归方程 .....	(518)
10.3 离差的度量 .....	(528)
10.4 假设条件 .....	(533)
10.5 残差分析 .....	(534)
10.6 自相关的度量：Durbin-Watson 统计量 .....	(539)
10.7 有关斜率和相关系数的统计推断 .....	(548)
10.8 均值的估计和单个数值的预测 .....	(556)
10.9 回归的缺陷以及道德问题 .....	(560)
10.10 一元线性回归的计算 .....	(564)
案例研究——东西部搬运公司 .....	(583)
案例研究——Springville Herald .....	(584)
附录 10.1 Microsoft Excel 在一元线性回归中的应用 .....	(586)
附录 10.2 Minitab 在一元线性回归中的应用 .....	(589)
附录 10.3 SPSS®在一元线性回归中的应用 .....	(591)
<b>第 11 章 多元回归 .....</b>	<b>(593)</b>
11.1 建立多元回归模型 .....	(594)
11.2 多元回归模型的残差分析 .....	(605)
11.3 多元回归模型显著性的检验 .....	(607)
11.4 总体回归系数的统计推断 .....	(610)
11.5 多元回归模型的偏检验 .....	(613)
11.6 二次型回归模型 .....	(621)
11.7 虚拟变量模型 .....	(633)
11.8 数学变换在回归模型中的应用 .....	(647)
11.9 多重共线性 .....	(651)
11.10 构建模型 .....	(653)
11.11 回归的缺陷以及道德话题 .....	(666)
案例研究——山脉马铃薯公司 .....	(676)
附录 11.1 Microsoft Excel 在多元回归中的应用 .....	(677)
附录 11.2 Minitab 在多元回归中的应用 .....	(682)

附录 11.3 SPSS <sup>®</sup> 在多元回归中的应用 .....	(684)
<b>第 12 章 时间序列预测</b> .....	(687)
12.1 商务预测的重要性 .....	(688)
12.2 古典多元时间序列模型的构成要素 .....	(689)
12.3 平滑年度时间序列 .....	(691)
12.4 最小二乘趋势拟合与预测 .....	(699)
12.5 趋势拟合和预测的自回归模型 .....	(717)
12.6 选择适当的预测模型 .....	(728)
12.7 月份数据和季度数据的时间序列预测 .....	(734)
12.8 时间序列预测的缺陷 .....	(744)
案例研究——外币交易 .....	(752)
案例研究——Springville Herald .....	(752)
附录 12.1 使用 Microsoft Excel 进行时间序列预测 .....	(754)
附录 12.2 使用 Minitab 进行时间序列预测 .....	(756)
附录 12.3 使用 SPSS <sup>®</sup> 进行时间序列预测 .....	(758)
<b>第 13 章 质量管理和生产作业管理中的统计应用</b> .....	(761)
13.1 质量管理和生产作业管理理论发展历史 .....	(762)
13.2 戴明 14 条：管理理论 .....	(763)
13.3 控制图概论 .....	(766)
13.4 不合格率控制图—— $p$ 图 .....	(768)
13.5 红珠子实验——理解过程波动 .....	(777)
13.6 波动幅度 ( $R$ ) 和均值 ( $\bar{X}$ ) 的控制图 .....	(779)
案例研究——锁线装订机公司 .....	(797)
案例研究——Springville Herald .....	(801)
附录 13.1 应用 Microsoft Excel 生成控制图 .....	(804)
附录 13.2 应用 Minitab 生成控制图 .....	(806)
附录 13.3 应用 SPSS <sup>®</sup> 生成控制图 .....	(808)
<b>课后习题答案</b> .....	(810)
<b>附录 A 回顾算术、代数和对数</b> .....	(837)
<b>附录 E 附表</b> .....	(840)
<b>附录 F 关于 PHStat2</b> .....	(859)

# 第1章

## 概述与数据收集

### 学习目标

- ◎对统计学的学科内容及其应用做一个全面概述；
- ◎区分统计描述和统计推断；
- ◎讨论数据的来源；
- ◎讨论数据的类型；
- ◎介绍不同的抽样方法；
- ◎学习如何评价调查的价值。

**统计应用引例**

**Good Tunes 顾客满意度调查**

Good Tunes 是一家通过互联网做电子商务的厂商，主要销售高品质的音响设备，它试图向顾客提供更加优质的服务和更高质量的商品。为了了解顾客的看法，Good Tunes 请它的顾客填写并迅速寄回一张顾客满意度调查问卷。这张问卷包含了以下问题：

- 从你订购商品之日起到你收到商品共需多少天？ \_\_\_\_\_
- 在接下来的 12 个月里你打算花多少钱（以美元表示）用于购买音响设备？ \_\_\_\_\_
- 就你近期内购买 Good Tunes 的商品的经验而言，你如何评价 Good Tunes 的总体服务质量？

比预期的要好得多       比预期的要差   
比预期的要好       比预期的要差得多   
跟预期的差不多

- 你怎样评价您最近购买的 Good Tunes 音响设备的产品质量？

比预期的要好得多       比预期的要差   
比预期的要好       比预期的要差得多   
跟预期的差不多

- 在接下来的 12 个月里你是否打算再向 Good Tunes 购买其他产品？  
是 \_\_\_\_\_ 否 \_\_\_\_\_

请您对以上调查问卷做出评论：

- 填写完整的问卷将会提供何种类型的信息？
- Good Tunes 应如何利用这些信息来改进它们的服务和产品的质量？
- 你认为这张调查问卷还应包含哪些其他的问题？

## 1.1 经理人为什么需要了解统计学

21 世纪的经理人会接触到大量的信息。一个重要的问题是如何利用可获得的信息来做出更好的决策。

本书就是从如何做出明智决策的角度来撰写的。出于以下四个关键的原因，经理人对统计学有一定了解：

- (1) 为了知道如何正确地演示和描述信息；
- (2) 为了知道如何依据样本信息得出有关总体的结论；

- (3) 为了知道如何改进运营过程；  
 (4) 为了知道如何获得可以信赖的预测。

在图 1—1 中，我们可以看到本书的一幅结构图，它就是从学习统计学的这四个原因的角度来构造的。从这张结构图中，我们可以看到，前 3 章包括了有关收集、演示以及描述信息的各种方法。第 4 章和第 5 章是概率论的基本概念以及二项分布和正态分布的知识，这样在第 6~9 章中，你就可以学习如何依赖样本信息得出有关总体的结论。第 10~12 章集中讲述一元回归、多元回归问题和时间序列分析，这些章节将教给你预测的方法。第 13 章包含统计学在质量和生产管理上的应用问题，这对于改进运营过程是至关重要的。

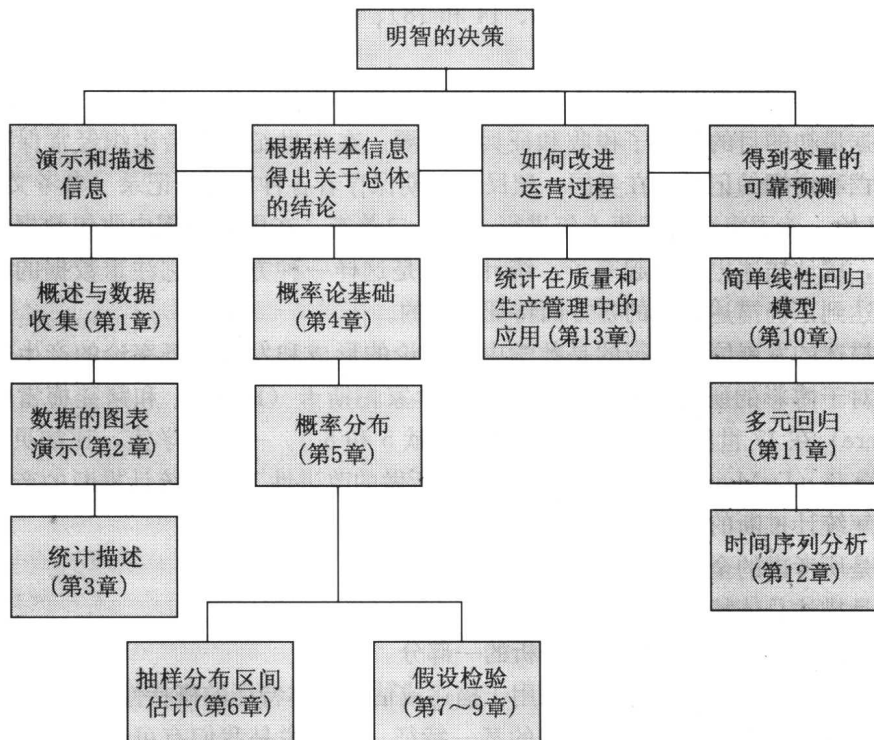


图 1—1 本书的结构图

统计学方法在商业活动的职能部门应用得十分普遍：会计、财务、管理和营销都离不开统计学。会计部门使用统计学方法选择样本进行审计，并在成本会计中找出成本动因。财务部门运用统计学方法在不同的投资组合中做出选择，并且追踪财务指标随时间的变化趋势。管理部门运用统计学方法来改进产品质量或服务。营销部门运用统计学方法来估计偏爱某种产品的顾客的比例，以及他们为什么喜欢这种产品，并且由此做出决策：采取何种广告战略能够最有效地扩大产品销量。

在整本书中，“统计应用引例”将贯穿始终，这些案例让你以一个职业经理人的身份置身于真实商业活动中，使用统计学方法来做出决策。在“统计应用引例”中提出的问题将引导你阅读本书后面的章节，并利用其中的统计学技巧来得到答案。通过阅读本书和研究案例，

你将会知道 21 世纪的职业经理人是如何运用统计学做出决策、解决问题以及改进他们的产品和服务质量的。

## 1.2 现代统计学的发展

从历史上看，现代统计学的产生和发展可以追溯到 3 个相互独立的现象：政府收集公民的有关数据的需要（参考文献 5、6、13、14 和 18）、数学中概率论的发展和电子计算技术的不断进步。

在所有有史料记载的时期内，人们都在做着数据收集的工作。在埃及、希腊和罗马文明中，收集数据最初的目的是为了税收和征兵的需要。在中世纪，教会组织经常保存着有关人口出生、死亡和婚姻的记录。在美国，殖民地时期保存着各种各样的记录（参考文献 19），而从 1790 年开始，美国宪法要求每十年进行一次人口普查。在国家范围内收集数据的需要不断加强的同时，统计描述也发展起来了，统计描述是这样一种方法：它注重数据的收集、演示和刻画，以达到正确描述数据的各方面特征的目的。

与统计描述的发展同时进行的是数学中概率论的形成和发展。概率论的产生部分基于文艺复兴时期对于博彩的研究，它可以追溯到数学家帕斯卡（Pascal）和赌徒梅雷骑士（Chevalier de Mere）在 17 世纪中叶的通信（参考文献 5 和 6）。一些数学家，比如贝努里（Bernoulli）、列莫弗（DeMoivre）和高斯（Gauss）所做的改进性工作为统计推断的产生奠定了基础。为了理解统计推断的含义，可以参见下边的定义。

总体就是所考虑的全部元素。

参数就是描述总体特征的概括性度量。

样本就是从总体中挑选出来用于分析的一部分。

统计量就是根据样本数据计算出来用于描述或估计总体特征的概括性度量。

统计推断使得我们有可能估计总体的某一特征，或者说是我们有可能仅依据样本的结果做出有关总体的决策。在 20 世纪，像皮尔逊（Pearson）、费雪（Fisher）、格西特（Gosset）、纽曼（Neyman）、韦尔德（Wald）和塔基（Tukey）这些统计学家都是统计推断方法发展中的先驱者，今天，他们发展的这些方法广泛地应用在诸多领域中。今天使用这些方法还需要应用现代的抽样技术。随着总体不断变大，要想从总体中获取信息不仅要耗费大量金钱、时间，而且非常麻烦。因此，我们只能通过样本信息做出有关总体特征的决策。概率论帮助我们确定从样本得到的结论与总体的结论在多大程度上是相同的。

这些发展在很大程度上受益于电子计算技术的进步，现在，电子计算技术的进步使得个人用户拥有了早期统计学家无法想象的计算能力。从 20 世纪六七十年代开始，“统计软件包”出现了，它可以帮助用户自动完成统计分析中的计算部分。20 世纪 80 年代，诸如 SAS、SPSS 和 Minitab（参考文献 9、16 和 17）这些软件的销售商开始为不断成长起来的个人计算机市场生产各种版本的软件。在这些软件投放到市场的同时，另一类电子表格软件，比如

Microsoft Excel, 逐渐成熟起来并具备了统计功能。用户可以使用这些程序创建能够交互操作数据的模型, 只要把原来写在纸质工作表格上的数据录入到计算机上, 即在计算机上形成包含规则的列和行的表格 (使用这种表格的方法很久以来一直用于财务数据的管理和分析)。

软件工具广泛的可获得性导致统计学方法在商业决策上的应用实现了前所未有的增长。现在, 企业经理人面临着在信息量越来越大的基础上做出越来越多决策的双重压力, 而全球经济也正变得日益电子化和网络化, 在这种情况下, 统计学在商业中的应用只会增加, 不会减少。

### 1.3 本书的读者需要了解哪些有关微软 Excel 和 PHStat, Minitab 或者 SPSS®的知识

本书中所有的运算结果都是用一种电子工作表 (spreadsheet) 程序——Microsoft Excel 以及统计软件——Minitab 和 SPSS®表示出来的。此外, 每一章后的附录都包含了如何使用 PHStat, Minitab 和 SPSS®解决该章所讨论问题的详细步骤说明。

为了更加有效地使用本书, 一个方法是使用 Microsoft Excel 计算出所有的统计学问题, 这种电子表格程序既可以独立销售, 也是一种叫做 Microsoft Office 的一套程序的一部分。尽管并非毫无缺陷 (参见下面的评注部分), 对于商学初学者来说, Microsoft Excel 仍不失为介绍和演示统计学方法的良好工具, 此外它还有一个优点是容易获得。

每一章附录中的 Microsoft Excel 使用指南要么讲述如何运用常见的 Excel 解决问题, 要么讲述如何使用 PHStat 解决问题。对于一些特定的统计分析, PHStat 可以对数据进行预处理, 而这些预处理工作如果依靠人工进行的话, 会非常复杂且容易出错。PHStat 还可以以一种新的方式把 Microsoft Excel 的基本功能结合起来, 进行其他分析。

使用 PHStat 只需要使用 Microsoft Excel 的最基本的能力。但是, 读者若要在计算机系统上安装和运行 PHStat, 就要按照附录 F 中的详细说明一步步进行。

#### 评注

本书中使用 Microsoft Excel, 并不意味着它可以取代当前应用的各种统计软件。正如其他人指出的那样 (参考文献 7), Microsoft Excel 的一些例行统计程序含有缺陷, 有时可能得出不可靠的结果, 尤其是在使用非常大的数据集或者数据集包含异常统计特征时。有些缺陷总是会造成不可靠的结果, 在那些情况下, 我们会指出缺陷并使用其他方法计算结果。在其他情况下, 我们会选择要分析的数据以避免这些缺陷的影响, 并且让读者从 Microsoft Excel 给出的结果中得出正确的结论。

另一个办法是使用诸如 Minitab 和 SPSS®这些统计软件解决本书中的统计学问题。

尽管统计软件和空白表格程序方便了非常复杂的分析, 你也应该认识到, 如果统计学经

验不够丰富的学生不能正确理解即将讲到的统计过程中的假设条件和局限性，那他们就会被从计算机软件上得到的结果误导，问题也就出现了。因此，出于教学上的原因，本书中讲述的所有统计方法的应用，我们都通过给出运算结果的例子来演示，这也是非常重要的。

## 1.4 为什么需要数据

获得有用的信息对于商业活动来说是至关重要的。数据就包含了在一个特定情况下做出更明智决策所需要的信息。在很多种情况下我们都需要数据：

- 某个市场调研者需要了解产品的特征以便区分一种产品与另一种产品。
- 某个研制药物的人员想要知道某种新药是否比当前正在使用的药物更加有效。
- 某公司的生产经理要规范地监测生产装配过程，以确定正在生产的产品质量是否符合公司的标准。
- 某个审计人员要了解公司的财务活动，确定这些活动是否符合大家普遍接受的会计准则。
- 某个潜在的投资者想知道在经济复苏的时候哪些行业里的哪些企业最有可能加速发展。
- 某个学生为了满足其好奇心，想得到有关他的同学最喜爱的摇滚乐队的的数据。

正如要点 1.1 显示的那样，收集数据主要有 6 个原因。

### 要点 1.1

收集数据的原因：

- (1) 需要用数据为调查提供必要的信息；
- (2) 需要用数据为研究提供必要的信息；
- (3) 需要用数据来衡量某项正在进行的服务或生产过程的表现；
- (4) 需要用数据来评价某项事物是否符合标准；
- (5) 需要用数据来帮助人们在决策过程中形成可供选择的行动方案；
- (6) 需要用数据来满足人们的好奇心。

“统计应用引例”中的 Good Tunes 顾客满意程度调查体现了以上第 (1)、(3)、(4) 和 (5) 条原因。例如，Good Tunes 在收集调查问卷后整理数据，然后分析这些数据，利用这些数据衡量他们的工作业绩，评价他们的产品和服务质量，并且如果需要的话，他们还可以策划其他的行动方案。

在开始进行任何统计分析时，找出最合适的数据来源是极为重要的。如果数据出现偏差，有模棱两可的情况或者出现其他类型的错误，那么即使使用最精妙最先进的统计学方法也可能无法弥补这些缺陷。



## 1.5 数据的来源

正如要点 1.2 显示的那样，有 4 个主要的数据来源。

### 要点 1.2

主要的数据来源：

- (1) 政府、行业或者个人已经发布或出版了各种数据；
- (2) 可以设计一个实验来获得必要的数据；
- (3) 可以进行调查；
- (4) 可以通过观察研究的方式。

正如要点 1.2 显示的那样，第一种获得数据的方法是通过政府、行业或者个人的数据来源渠道。美国政府为了公共的和私人的需要设置专门机构收集和整理数据。劳动统计局 (the Bureau of Labor Statistics) 负责收集有关就业方面的数据，并且负责核算每个月的消费物价指数。人口普查局 (the Bureau of the Census) 除了依照宪法的要求每十年进行一次人口普查外，还负责进行许多有关人口、住房和制造业的调查。此外，它还承担着一些特殊的研究课题，比如犯罪、旅行和医疗卫生。

除了美国政府，各种贸易出版物也会提供涉及各个行业组织的数据。像穆迪 (Moody) 这样的投资咨询机构会提供某个企业的财务数据。而像 A. C. 尼尔森 (A. C. Nielsen) 这样的综合服务机构则向它的委托人提供信息，以比较委托人与其竞争对手的产品。每天的报纸上也都充满了数不胜数的关于证券价格、天气情况以及体育方面的信息。

第 2 种数据来源是实验。做实验时，要对实验过程施加严格控制。例如，在一项测试洗涤剂的去污能力的研究中，研究者想知道哪种品牌的洗涤剂对于脏衣物的去污效果最好，那他们就要亲自用洗涤剂来洗净衣物，而不是询问消费者他们认为哪种品牌的洗涤剂最有效。如何正确设计实验过程通常是更高级的教材会涉及的内容，因为这需要更复杂的统计程序。但是，为了让大家对测试和实验有一个认识，本书的第 7~9 章讲述了设计实验的最基本的原则。

第 3 种数据来源是进行调查。调查时，不需要对接受调查的人的行为有任何限制，只是问他们一些关于他们的看法、态度、行为和其他方面的问题，然后再对他们的回答进行整理、编码、列表，用以分析。

第 4 种获得数据的方法是通过观察研究。研究者直接观察行为，通常是观察自然界中的行为。大部分有关动物的知识都是通过这种方法获得的，其他一些自然科学领域，比如天文学和地理学也可以通过这种方法获得知识。在这些领域中，要进行实验和调查，即使是可能