

工商管理经典译丛

BUSINESS STATISTICS
商务统计学

(第5版)

[美] 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)
黄耀锋 王小勇 等 译
胡大源 审校

PEARSON



中国人民大学出版社

BUSINESS ADMINISTRATION CLASSICS 工商管理经典译丛



工商管理
经典译丛

BUSINESS STATISTICS

商务统计学

(第5版)

戴维·M·莱文 (David M. Levine)

[美] 蒂莫西·C·克雷比尔 (Timothy C. Krehbiel) 著

马克·L·贝伦森 (Mark L. Berenson)

黄耀锋 王小勇 等 译

胡大源 审校



中国人民大学出版社

·北京·

图书在版编目 (CIP) 数据

商务统计学. 第5版/ (美) 莱文, (美) 克雷比尔, (美) 贝伦森著; 黄耀锋等译.
北京: 中国人民大学出版社, 2010
(工商管理经典译丛)
ISBN 978-7-300-12492-6

- I. 商…
- II. ①莱… ②克… ③贝… ④黄…
- III. 商业统计学
- IV. F712.3

中国版本图书馆 CIP 数据核字 (2010) 第 139380 号

工商管理经典译丛

商务统计学 (第5版)

戴维·M·莱文

[美] 蒂莫西·C·克雷比尔 著

马克·L·贝伦森

黄耀锋 王小勇 等 译

胡大源 审校

Shangwu Tongjixue

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)

010-82501766 (邮购部)

010-62515195 (发行公司)

网 址 <http://www.crup.com.cn>

<http://www.ttrnet.com> (人大教研网)

经 销 新华书店

印 刷 北京宏伟双华印刷有限公司

规 格 185 mm×260 mm 16 开本

印 张 34.25 插页 2

字 数 791 000

邮政编码 100080

010-62511398 (质管部)

010-62514148 (门市部)

010-62515275 (盗版举报)

版 次 2010 年 8 月第 1 版

印 次 2010 年 8 月第 1 次印刷

定 价 65.00 元

版权所有 侵权必究

印装差错 负责调换

出版说明

随着中国改革开放的深入发展，中国经济高速增长，为中国企业带来了勃勃生机，也为中国管理人才提供了成长和一显身手的广阔天地。时代呼唤能够在国际市场上搏击的中国企业家，时代呼唤谙熟国际市场规则的职业经理人。中国的工商管理教育事业也迎来了快速发展的良机。中国人民大学出版社正是为了适应这样一种时代的需要，从1997年开始就组织策划《工商管理经典译丛》，这是国内第一套与国际管理教育全面接轨的引进版工商管理类丛书，该套丛书凝聚着100多位管理学专家学者的心血，一经推出，立即受到了国内管理学界和企业界读者们的一致好评和普遍欢迎，并持续畅销数年。全国人大常委会副委员长、国家自然科学基金会管理科学部主任成思危先生，以及全国MBA教育指导委员会的专家们，都对这套丛书给予了很高的评价，认为这套译丛为中国工商管理教育事业做了开创性的工作，为国内管理专业教学首次系统地引进了优秀的范本，并为广大管理专业教师提高教材甄选和编写水平发挥了很大的作用。据统计，本丛书现已成为国内管理院校和企业培训中采用率最高、影响最大的引进版教材。其中加里·德斯勒的《人力资源管理》获第十二届“中国图书奖”；斯蒂芬·罗宾斯的《管理学》获全国优秀畅销书奖。

进入21世纪后，随着经济全球化和信息化的发展，国际MBA教育在课程体系上进行了重大的改革，从20世纪80年代以行为科学为基础，注重营销管理、运营管理、财务管理到战略管理等方面的研究，到开始重视沟通、创业、公共关系和商业伦理等人文类内容；并且增加了基于网络的电子商务、技术管理、业务流程重组和统计学等技术类内容。另外，管理教育的国际化趋势也越来越明显，主要表现在师资的国际化、生源的国际化 and 教材的国际化方面。近年来，随着我国MBA和工商管理教育事业的快速发展，国内管理类引进版教材的品种越来越多，出版和更新的周期也在明显加快。为此，我们这套《工商管理经典译丛》也适时更新版本，增加新的内容，同时还将陆续推出新的系列和配套的案例教材、教学参考书，以顺应国际管理教育发展的大趋势。

本译丛入选图书，都是世界著名的出版机构，如培生教育出版集团、美国麦格劳-希尔教育出版公司、圣智学习出版公司、约翰威立出版公司等畅销全球的工商管理教材，被世界各国（地区）的著名大学商学院和管理学院所普遍选用，是国际工商管理教育界最具影响力的教科书。本丛书的作者，皆为欧美管理学界享有盛誉的著名教授，他们的这些教材，经过了美国和世界各地数千所大学和管理学院教学实践的检验，被证明是论述精辟、视野开阔、资料丰富、通俗易懂，又具有生动性、启发性和可操作性的经典之作。本译丛的译者，大多是各大学的优秀中青年学术骨干，他们大都曾留学欧美，在长期的教学研究和实践中积累了丰富的经验，具有较高的翻译水平。

本丛书的引进和运作过程,从市场调研与选题策划、每本书的推荐与论证、对译者翻译水平的考察与甄选、翻译规程与交稿要求的制定、对译稿质量的严格把关和控制,到版式、封面和插图的设计等各方面,都坚持高水平和高标准的原则,力图奉献给读者一套译文准确、文字流畅、从内容到形式都保持原著风格的工商管理精品图书。

本丛书参考了国际上通行的MBA和工商管理专业核心课程的设置,并充分兼顾了我国管理各专业现行通开课与专业课程设置,以及企业管理培训的要求,故适应面较广,既可用于管理各专业不同层次的教学,又可供各类管理人员培训和个人自学使用。

为了本丛书的出版,我们成立了由中国人民大学、北京大学、中国社会科学院等单位专家学者组成的编辑委员会,德高望重的袁宝华同志、黄达教授和中国人民大学校长纪宝成教授,都给了我们强有力的支持,使本丛书得以在管理学界和企业界产生较大的影响。许多我国留美学者和国内管理学界著名专家教授,参与了原著的推荐、论证和翻译工作,原我社编辑闻洁女士在这套书的总体策划中付出了很多心血。在此,谨向他们致以崇高的敬意并表示衷心的感谢。

愿这套丛书为我国MBA和工商管理教育事业的发展,为中国企业管理水平的不断提升继续做出应有的贡献。

中国人民大学出版社

校译者序——怎样学习商务统计学

商务统计学可以定义为在准确调查研究基础上，对可靠数据进行分析与概括，整理出一套有助于经营管理者在不确定条件下作出正确决策的原理和方法。这一定义的关键在于准确的观察、科学的分析和正确的决策。

多年前，我曾同一群来访的美国商学院的 MBA 学生座谈，当问到哪门课他们最不愿意学时，回答正是商务统计。前不久，听美国商学院的一位教授讲，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，商务统计恐怕是唯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”），讲的是一位数学天分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。尽管这篇短文的故事情节有些夸张，结果颇具戏剧性，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，加之多花上些工夫，数学基础薄弱的学生仍有很大的可能取得好成绩。

在过去的十余年间，我们曾为数以千计的北京大学中国经济研究中心本科双学位学生和已有多年工作经验的北大国际（BiMBA）硕士研究生讲授商务与经济统计课程，之所以选用莱文、克雷比尔和贝伦森的这本《商务统计学》作教材，最主要原因就是这本书的实用性强，并且有大量的不断更新的应用案例和练习题。

理解商务统计的关键在于结合现实问题进行思考和分析。要学好商务统计，熟悉并透彻理解一些典型案例不失为一条事半功倍的学习途径。案例越生动，越令人信服，就越容易给人留下深刻的印象，日后遇到类似的问题，对这个案例的分析便自动浮现。对本教材感兴趣的读者，学习商务统计大多不在于对纯理论的兴趣，而是为了系统地了解在不确定条件下通过调查研究进行判断与决策的分析思路与方法，以便更好地解决现实世界中遇到的实际问题。与领军作战相似，常常有这样的情形：面对突然发生的不确定事件，无论怎样处置都难

免会有失败的风险。此时,指挥官必须当机立断,迅速作出判断以便下定决心。在这种情况下,总结类似环境下的案例便大有益处。善于触类旁通的决策者可以根据掌握的各种信息与丰富的个人经历迅速进行类比,分析异同,作出判断,制定对策。这正是百年哈佛商学院案例教学法的精妙之处。本书中的许多应用实例都十分生动有趣。现在许多人的计算机都装有垃圾邮件过滤软件,但这个软件是怎样设计的?在过滤时为什么会犯错误?知道的人并不多。在第4章的“想一想”栏目中讨论了垃圾邮件过滤器的工作原理,介绍了计算机软件如何应用贝叶斯定理来学习识别垃圾邮件的过程。

与国内众多统计学教材相比,这本商务统计教材中的数学推导很少。从理论角度来看,也许算不得系统严谨。但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,因此本书更适合那些注重学以致用读者的需求。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,一方面可使读者免去复杂数学计算之苦,另一方面也帮助读者在学习商务统计知识的同时提高计算机软件的使用能力。

初次看到本书的读者可能会觉得篇幅太多。对于在教师指导下的学生,这也许算不上什么缺点,任课老师自会选择要讲授的章节,其他部分可作为参考,供有兴趣的读者翻看。这种编写教材的方式在国内外的大学和研究生教材中已司空见惯。因此,对于自学为主的读者,大可不必被这本书的篇幅吓倒。事实上,作者由浅入深、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更好地了解国际商务运营环境。

这次翻译的这本《商务统计学》第5版是著名国际教育出版商 Pearson 公司 2010 年的最新版本,由黄耀锋、王小勇、陆奔和朱锐翻译。2010 年新版的改进之处包括:结合一些章节增加了“想一想”短文,以便读者深入思考所学的知识,并且更好地了解统计知识在现实应用中的重要性;根据《华尔街日报》和《今日美国》等媒体报道编写增补了很多新的应用实例和练习题;在许多章节中引入学生调查;扩充和改进了 Excel 等统计软件的应用演示;在各章的结尾增加了“统计应用”,以便更好地掌握该章的内容及其应用。

胡大源

目 录

第 1 章 概述与数据收集	(1)
1.1 为什么学习统计学	(1)
1.2 商业世界中的统计学	(2)
1.3 统计学的基本术语	(3)
1.4 本书的组织方式	(4)
1.5 数据的来源	(5)
1.6 数据的类型	(7)
1.7 Microsoft Excel 和 Minitab	(8)
1.8 学习使用统计程序	(9)
第 2 章 用图表演示数据	(15)
2.1 属性数据的图表演示.....	(16)
2.2 组织数值数据.....	(23)
2.3 数值数据的图表演示.....	(26)
2.4 交叉分类表.....	(40)
2.5 散点图和时间序列图.....	(43)
2.6 图表的误用和道德问题.....	(47)
第 3 章 数值描述度量	(60)
3.1 集中趋势的度量.....	(61)
3.2 变异程度和分布形状.....	(65)
3.3 总体的数值型描述性度量.....	(77)
3.4 四分位数和盒须图.....	(82)
3.5 协方差和相关系数.....	(89)
3.6 描述性数值度量：缺陷和道德问题.....	(95)
第 4 章 概率论基础	(103)
4.1 概率论基础	(104)
4.2 条件概率	(113)
4.3 贝叶斯定理	(121)
4.4 道德问题和概率	(126)



第 5 章 离散概率分布	(132)
5.1 离散随机变量的概率分布	(133)
5.2 二项分布	(137)
5.3 泊松分布	(144)
第 6 章 正态分布	(154)
6.1 连续概率分布	(155)
6.2 正态分布	(155)
6.3 估计正态性	(170)
第 7 章 抽样和抽样分布	(180)
7.1 抽样分布	(180)
7.2 估计调查的价值	(186)
7.3 抽样分布	(190)
7.4 均值的抽样分布	(190)
7.5 比例的抽样分布	(200)
第 8 章 置信区间估计	(208)
8.1 总体均值的置信区间估计 (已知 σ 时)	(209)
8.2 总体均值的置信区间估计 (未知 σ 时)	(215)
8.3 比例的置信区间估计	(224)
8.4 样本容量的确定	(227)
8.5 置信区间估计和道德问题	(233)
第 9 章 假设检验基础: 单侧检验	(242)
9.1 基本假设检验方法	(243)
9.2 均值的 t 假设检验 (σ 未知)	(256)
9.3 单侧检验	(264)
9.4 比例的 Z 假设检验	(269)
9.5 假设检验的潜在缺陷和职业道德问题	(273)
第 10 章 两个样本数值数据假设检验和单向方差分析	(280)
10.1 两个独立总体的均值比较	(281)
10.2 两个相关样本的均值比较	(290)
10.3 两个独立总体的比例差异的比较	(299)
10.4 两个总体方差差异的 F 检验	(305)
10.5 单向方差分析	(311)
第 11 章 χ^2 检验	(335)
11.1 两个比例差异的 χ^2 检验	(335)
11.2 多个比例差异的 χ^2 检验	(343)

11.3 独立性的 χ^2 检验	(348)
第 12 章 一元线性回归	(364)
12.1 回归模型的类型	(365)
12.2 建立一元线性回归方程	(367)
12.3 离差的度量	(376)
12.4 假设条件	(381)
12.5 残差分析	(382)
12.6 有关斜率和相关系数的统计推断	(386)
12.7 均值的估计和单个数值的预测	(394)
12.8 回归的缺陷	(399)
第 13 章 多元回归	(412)
13.1 建立多元回归模型	(412)
13.2 r^2 , 调整后的 r^2 和整体的 F 检验	(419)
13.3 多元回归模型的残差分析	(423)
13.4 总体回归系数的统计推断	(425)
13.5 虚拟变量和交互项在回归模型中的使用	(429)
第 14 章 质量管理中的统计应用	(443)
14.1 控制图论	(444)
14.2 比例的控制图: P 图	(446)
14.3 红珠子实验——理解过程波动	(452)
14.4 波动幅度和均值的控制图	(454)
14.5 全面质量管理	(460)
14.6 六西格玛	(462)
附 录	(474)
自我测验答案和部分偶数习题答案	(498)

第1章

概述与数据收集

学习目标

在本章中你将学到：

- 商业中是如何使用统计学的
- 数据的来源
- 数据的类型
- Microsoft Excel 的基本使用
- Minitab 的基本使用

统计应用

Good Tunes

Good Tunes 是一家拥有 4 家店面、成长中的家庭娱乐系统零售商，它试图在未来的 3 年中使其拥有的店面数翻倍。为了支持扩张所需要的资金，经理们已经决定同本地银行接触。他们需要准备电子幻灯片和正式的内容说明书来说明 Good Tunes 是一家蓬勃发展的企业和扩张的良好候选人。

你被要求来协助准备幻灯片和内容说明。你如何完成这份工作？学习更多的统计学将是一个好的开始。

1.1 为什么学习统计学

人们每天都用数字来描述或者分析我们生活的这个世界。例如，考虑这些最近的大字标题：

●“股票价格可能会下降，但高层管理者的薪酬却未必”（G. Farrell and B. Hansen, usatoday.com, April 10, 2008）——在 2007 年，S&P 500 中最大的 50 家公司高管的中位数薪酬为 1 570 万美元。



● “美国人消费更多的瓶装水” (USA Today, February 28, 2007, p. 1D) ——瓶装水的人均消费量从 2001 年的 18.8 加仑增长到 2006 年的 28.3 加仑。

● “用信用卡支付得更多” (USA Today, March 6, 2007, p. 1D) ——消费者用信用卡付款从 1995 年的 18% 增长到 2005 年的 25%，同期使用现金付款从 21% 下降到 14%。

● “实际最有价值的选手” (R. Adams, *The Wall Street Journal*, April 14, 2007, pp. 1, 4) ——经济学家已经开发了模型来预测棒球选手表现的真正价值，且发现其与选手的薪酬相关。

如果你懂统计学，你对这些大字标题背后的故事里的数字会有更好的感觉。统计学 (statistics) 是将数字转化为对决策者有用信息的数学的分支。变化是任何决策制定过程的一部分，统计学提供一种理解和进一步减少但不是消除变化的方式，并且能够告诉你作出一个决定所伴随的风险。

统计学通过提供一组分析数字的方法而实现。这些方法帮助你发现数字的模式，并且使你可以确定数字的不同是不是由于偶然性。你学习这些方法的同时，也会学习应用这些方法的适用条件。为了实际收益，有如此多的统计方法必须用计算机处理，同时学习统计学时，你也需要学习帮助在商业世界里运用统计学的程序。

1.2 商业世界中的统计学

在商业世界中，统计学有 4 种重要的应用：

- 总结商业数据
- 根据数据得出结论
- 作出商业行动的可靠预测
- 改进运营过程

统计学领域包括两个分支：统计描述和统计推断。

统计描述 (descriptive statistics) 注重数据的收集、总结、演示和分析。你可能熟悉诸如表格和图表之类的演示以及以前在学校学过的平均数和中位数。

统计推断 (inferential statistics) 利用从一个小组收集的数据而得出有关更大组的结论。这些方法被用来作出关于什么投资会带来更高的收益和什么市场策略可能带来销售的增加的决策。

看看上面的圆点 (以圆点●开头的内容)，统计描述允许你创建不同的表格和图表来总结数据。它也提供了统计量诸如均值、中位数和标准差来描述你的数据的不同特征。

从数据中得出结论是统计推断的核心。利用这些方法使你能够基于数据而不是直觉来作决策。

作出可靠的预测包括开发用于预测的统计模型。这些模型使你能够发展对未来行为的更精确的预测。

改进运营过程包括利用诸如六西格玛之类的关注提高质量的管理方法。这些方法是数据驱动的且使用统计方法作为质量提高方法的部分。

为了帮助你发展作出更好决策的技能，本书的每一章都有统计应用场景部分。这些场景描述了真实情形，你被要求作出决策来将数据转换为统计信息。例如，在某一章中你将会被要求决定超市中一种可乐饮料的货架位置，该位置能够最好地增加这种可乐饮料销售；在另一章中，你将被要求预测一家服装店的销售。

在本章开始的场景中，你肯定会提出下列问题。你应该使用哪些数据来让银行家信服，从而提高 Good Tunes 需要的信用额度？你如何展示那些数据？

但是，呈现给银行家数以千计的交易可能淹没他们，并且不是很有用。你需要以某些有用的方式总结每个交易的细节，从而将交易数据转换成信息，使得银行家能够揭示随着时间变化的销售的令人喜欢的模式。

银行家可能想要看到的一条信息是每年销售总额。记录和加总销售额是将数据转换成信息的一种普遍方法。当记录销售额或者其他你选择使用的关于 Good Tunes 相关数据的时候，你根据正常商业实践并且根据商业时期记录，例如月、季度或者年。当这样做的时候，你得到了多重值：今年的销售额、去年的销售额、前年的销售额，等等。如何决定查阅这些多重值的最好的方法，需要学习统计学的基本术语。

1.3 统计学的基本术语

变量是项目或者个人的特征，是使用一种统计方法所分析的对象。在 Good Tunes 场景中，销售额、每年的支出和净利润都是那些银行家想要分析的变量。

变量

变量 (variable) 是项目或者个人的一个特征。

当被用于日常交谈时，变量意味着某些事情改变或者变化了。你可能期望销售额、支出和净利润每年的值各不相同。这些不同的值就是与变量相关联的数据 (data)，更简单地说，被分析的数据。

变量可能因为时间之外的原因而不同。例如，你如果想分析一个大型讲座课的组成成分，可能需要包括下列变量：班级级别、性别和主修领域。这些变量也可能变化，因为班级中的每个学生都是不同的。一个学生可能是大学二年级、男性、会计专业，而另一个可能是大学一年级、女性、金融专业。

你需要记住值是无意义的，除非它们的变量有**通用的定义** (operational definitions)。这些定义对所有与该分析相关的人而言是被普遍接受的含义。即使每年的销售额的实际定义似乎是清楚的，如果一个人指整个连锁店的每年销售额而另一个指每家店每年的销售额，错误的传达就会发生。即使变量的个体值有时也是需要定义的——例如，对班级级别变量，大学新生和大学二年级的确切意思是什么？(可能含糊定义最著名的例子是 2000 年美国总统选举期间佛罗里达州的有效选票的定义。实际定义的含糊最终要求美国最高法院裁决。)

4 个基本术语是：总体、样本、参数和统计量。

总体

总体(population)就是所考虑的全部元素。

样本

样本(sample)就是从总体中挑选出来用于分析的一部分。

参数

参数(parameter)就是描述总体特征的概括性度量。

统计量

统计量(statistic)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。

特定年度的所有 Good Tunes 的销售交易、所有本周在 Good Tunes 购物的顾客、一个学院招收的所有全日制学生和俄亥俄州所有登记的选民,都是总体的例子。上述 4 个总体的样本是审计师研究所随机选择的 200 次 Good Tunes 销售交易、被要求完成顾客满意度调查的 30 个 Good Tunes 的顾客、为市场研究而选择的 50 个全日制学生和为民意测验而电话联系的 500 个俄亥俄州登记选民。在每个样本中,样本中的交易或者人代表了组成总体的项目或个人的一部分。

本周在 Good Tunes 购物的顾客的平均消费是参数的一个例子,因为整个总体的消费量是需要的。相对的,完成顾客满意度调查的 30 个顾客的平均消费量是统计量的一个例子,因为仅要求样本的 30 人的消费量。

1.4 本书的组织方式

如果你喜欢快速浏览一本书从而了解一本书的组织方式,你应当知道我们是根据 1.2 节中列出的 4 个商业目标组织本书章节的。我们用那些目标——总结商业数据,根据数据得出结论,作出商业行动的可靠预测,改进运营过程——作为组织章节的一种方式。

第 1~3 章讨论了演示和描述信息的方法。第 4~11 章讨论了利用样本信息得出关于总体的结论。第 12 和 13 章提供了作出可靠预测的方法。第 14 章介绍了如何应用统计方法来改进运营过程。

如果你的讲师没有覆盖所有的章节,请不要担心——介绍性商业统计课程在范围、长度和学分数方面各不相同。传统上,讲师从描述统计开始,因为这些方法中的一些你可能从日常行为中已经熟悉了,这也是我们将它们放在前三章的原因。第 4 章讨论了概率论的概念,这些概率是第 5~7 章所需要的。第 8~11 章包含了统计推断。第 8 章介绍了置信区间的方法,第 9 章讲述了假设检验的基本概念,并在第 10 和 11 章进一步扩展。理解这些概念将使你能够回答诸如“你如何基于一些(样本)而作出关于许多(总体)的决定”之类的问题,这些问题是未受过统计教育的商业人士在他们挑战统计应用时常问的。第 12 章和第 13 章包含了用

于作出可靠预测的回归模型。第12章讲述了简单回归模型，而第13章将这一简单模型扩展到更复杂的情形。第14章提供了一些质量管理和六西格玛质量提高方法的背景以及用于质量改进的统计方法介绍。

记住本书的目标不是让你成为一个统计学家。本书的目标是帮助你理解和正确应用你最感兴趣的商业领域常用的统计方法，这些商业领域有：会计、经济、金融、信息系统、市场或者管理。

顺便说一句，如果你喜欢先读一本书的最后一页，请不要这样阅读本书。商务统计介绍是需要累积的。如果你刚开始学习统计学，你跳到本书的后面会觉得迷惑不解。因此如果你碰巧发现诸如奇怪的表格之类的东西，不要着急，到你在本书中需要用到这些表格的时候，你会知道如何有效使用它们。

1.5 数据的来源

Good Tunes 的经理们相信，如果他们向银行家展示 Good Tunes 的顾客对其所接受的服务很满意，那么他们将有更强的扩展论据。经理们如何证明优质服务是在 Good Tunes 的典型的顾客体验？

在上述 Good Tunes 场景中，经理们现在面临两个挑战：首先识别顾客满意度研究的相关变量，然后设计一种**数据收集**（data collection）的方法，也就是，收集这些变量的值。

在很多种情况下我们都需要数据：

- 市场调研者需要了解一个新电视广告的效果。
- 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。
- 某公司的运营经理要改进一个生产或服务过程。
- 某个审计人员要了解公司的财务活动，确认这些活动是否符合普遍接受的会计准则。

在这些例子中，同 Good Tunes 的经理们一样，从总体中的每一个项目或者个人收集数据太难或者太费时间。因为这是典型情况，数据收集几乎一直包括从样本收集数据。（第7章讨论了样本选择的方法。）

数据来源分为**原始来源**（primary sources）和**二手来源**（secondary sources）。当数据收集者就是使用数据进行分析的人时，来源是原始的。当进行统计分析的人不是数据收集者时，来源是二手的。收集和发布数据的组织和个人作为原始来源自己使用数据，而让其他人当二手来源使用。例如，美国联邦政府为了公众和私有目的用这种方式收集和发布数据。劳工统计局收集失业率的数据且发布月度消费者价格指数。人口普查局监测多种正在进行的关于人口、住房和制造业的调查，且承担诸如犯罪、旅行和卫生保健之类主题的研究。

有四个主要的数据来源：

- 组织或者个人已经发布的数据。
- 可以设计一个实验来获得必要的数据。
- 可以进行调查。
- 可以通过观察研究的方式。

市场研究公司和同业公会发布有关特定行业或市场的数据。投资服务商例如 Mergent (参看 www.mergent.com) 提供各个公司的金融数据。而诸如 A. C. 尼尔森 (A. C. Nielsen) 这样的综合服务机构则向其委托人提供信息, 以比较委托人产品和委托人竞争对手的产品。每天的报纸上也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。

第二种数据来源是设计的实验。这些结果是实验的结果, 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。如何正确设计实验过程通常是更高级的教材所涉及的内容, 因为这需要更复杂的统计程序。但是, 本书的第 10 章讲述了一些设计实验的基本的概念。

第三种数据来源是进行调查。询问接受调查的人一些关于他们的看法、态度、行为和其他方面的问题。例如, 人们可能被要求回答他们关于哪种品牌的洗涤剂对于特定的污渍去污效果最好的观点。(这可能导致与被问同样问题的设计实验不同的结果。)

第四种获得数据的重要方法是通过观察研究。研究者直接观察行为, 通常在自然或中性的环境中。观察研究是商业领域收集数据的常用工具。举一个例子, 市场研究者利用焦点小组访谈来启发对于开放式问题的各种各样的回答。由一个组织者来引导讨论, 所有的参与者都回答提出的问题。观察研究技术也被用来增强团队合作或者提供产品和服务的质量。

确定最合适的来源是关键任务, 因为如果有偏见的、含糊的或者其他类型的错误使得收集的数据有缺陷, 那么即使是最有经验的统计方法也不能产生有用信息。在 Good Tunes 的例子中, 有关顾客体验的变量可以采取有关顾客体验多方面的调查问题的形式。例子可见图 1—1。为了进一步的分析, 调查也可提出将顾客分组的问题。

1. 您从订货到收货所花费的天数是多少? ____
2. 您是否购买 Good Tunes 销售宣传单上的特色商品? 是 ___ 否 ___
3. 这是您在 Good Tunes 首次购买吗? 是 ___ 否 ___
4. 您是否会在将来的 12 个月里从 Good Tunes 购买额外的商品? 是 ___ 否 ___
5. 您在将来的 12 个月里打算花多少钱 (美元计) 在立体声系统和消费电子设备上? ____
6. 您如何评价最近购物中 Good Tunes 所提供的服务?
完美 很好 一般 不好
7. 您如何评价与其他的家庭娱乐系统销售商相比 Good Tunes 所提供的产品选择?
完美 很好 一般 不好
8. 您如何评价最近从 Good Tunes 购买产品的质量?
完美 很好 一般 不好

图 1—1 Good Tunes 顾客体验的问题

对 Good Tunes 而言, 一种避免数据收集缺陷的好方法是将问卷分配到顾客的一个随机样本 (第 7 章解释了如何收集一个随机样本)。差一点的方法是依赖商业评级网站, 该网站允许在线访问者对店主评级。这些网站不能保证那些评级的人是顾客总体的代表甚至是顾客。

1.6 数据的类型

统计学家将变量分为属性变量和数值变量，并进一步将数值变量分为离散或连续。图 1—2 显示了这些关系并提供了每种类型变量的例子。

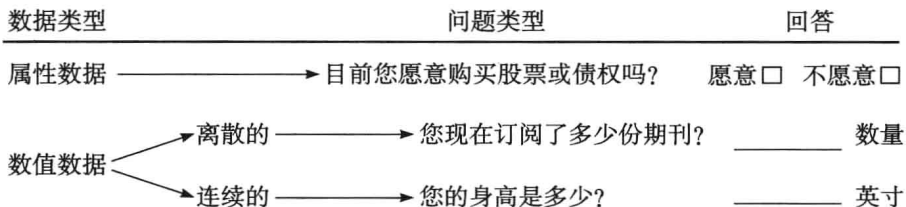


图 1—2 数据的类型

属性变量 (categorical variables) (也称为**定性变量** (qualitative variables)) 给出定性的回答，比如是或者不是。举一个例子，问题“你是否经常持有股票或有价证券?”和图 1—1 中问题 2~4 就是属性变量的例子，因为它们的答案只局限于是或者不是。属性变量可能会有多于两个的可能回答。例如，你可能问顾客他们购物是星期几。在图 1—1 中，问题 6~8 有 4 个可能的回答。

数值变量 (numerical variables) (也称为**定量变量** (quantitative variables)) 给出定量的回答。例如，在图 1—1 中的问题 1~5 就是数值变量。数值变量进一步分为离散的和连续的。

离散变量 (discrete variables) 是由计数过程得出的数值回答。“你订阅了多少种杂志”就是一个离散数值变量的例子，因为对于此问题的回答是有限整数中的一个。你只能订阅零种、一种、两种或更多种杂志。一个顾客购买的项目数也是一个离散数值变量，因为你可以计算购买的项目的数目。

连续变量 (continuous variables) 是由测度过程得出的数值回答。你在银行等候出纳员的时间就是一个连续变量的例子，因为对于此问题的回答是在一个连续统或者区间内取得的一个数值，数值的精确程度取决于测度工具的精确程度。例如，你等候的时间可能是 1 分钟、1.1 分钟、1.11 分钟或者 1.113 分钟，这取决于你所用测量工具的精确度。

从理论上说，如果使用的测度工具足够精确，没有两个连续值是一样的。但是在实际中，大多数测度工具都不够精确从而发现极其细微的差别。因此，即使变量实际上是连续的，在实验或者调查的数据中，我们也会经常发现相同的观测值。

1.6 习题

掌握基础知识

1.1 某家快餐店中出售三种不同的饮料——软饮料、茶和咖啡。解释为什么出售的饮料种类是属性变量。

1.2 某家快餐店中出售三种不同量的饮料——小杯、中杯和大杯。解释为什么出售的饮料量是属性变量。

1.3 假设你正在测量从网上下载