

“十三五”国家重点出版物出版规划项目

CENGAGE
Learning®

经济科学译丛

统计学

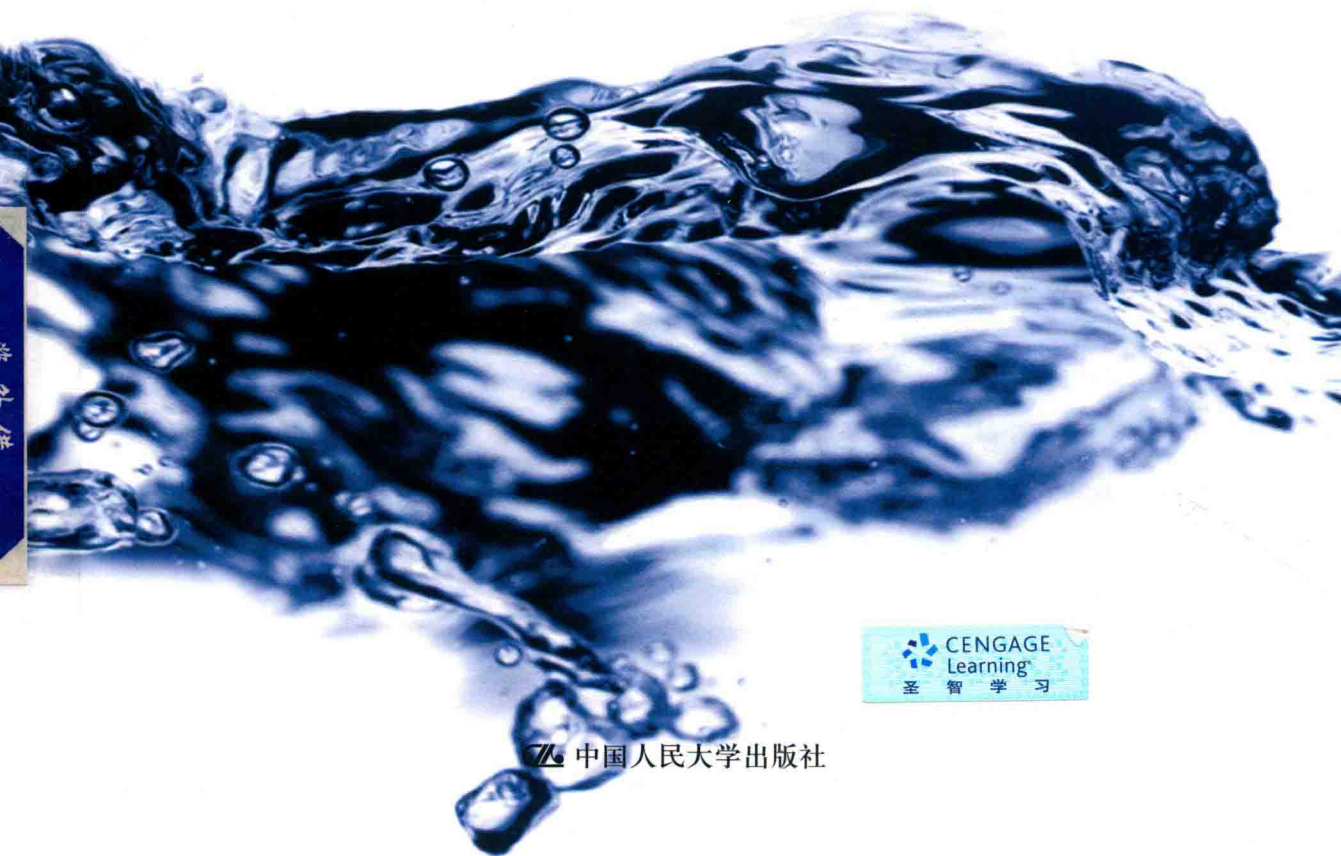
在经济和管理中的应用 (第10版)

上册

杰拉德·凯勒 (Gerald Keller) 著

Statistics for Management
and Economics

(Tenth Edition)



CENGAGE
Learning®
圣智学习

中国人民大学出版社

“十三五”国家重点出版物出版规划项目

经济科学译丛

统计学

在经济和管理中的应用（第10版）

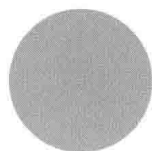
上册

杰拉德·凯勒（Gerald Keller）著

夏利宇 韩松涛 李君 张伟等译

Statistics for Management
and Economics

(Tenth Edition)



中国人民大学出版社
·北京·

《经济科学译丛》

编辑委员会

学术顾问

高鸿业 王传纶 胡代光 范家骧 朱绍文 吴易风

主 编

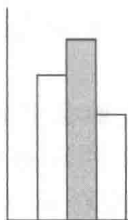
陈岱孙

副主编

梁 晶 海 闻

编 委 (按姓氏笔画排序)

王一江 王利民 王逸舟 贝多广 平新乔 白重恩
朱 玲 刘 伟 许成钢 李 扬 李晓西 李稻葵
杨小凯 汪丁丁 张宇燕 张维迎 林毅夫 易 纲
金 碚 姚开建 钱颖一 徐 宽 高培勇 盛 洪
梁小民 樊 纲



《经济科学译丛》总序

中国是一个文明古国，有着几千年的辉煌历史。近百年来，中国由盛而衰，一度成为世界上最贫穷、落后的国家之一。1949年中国共产党领导的革命，把中国从饥饿、贫困、被欺侮、被奴役的境地中解放出来。1978年以来的改革开放，使中国真正走上了通向繁荣富强的道路。

中国改革开放的目标是建立一个有效的社会主义市场经济体制，加速发展经济，提高人民生活水平。但是，要完成这一历史使命绝非易事，我们不仅需要从自己的实践中总结教训，也要从别人的实践中获取经验，还要用理论来指导我们的改革。市场经济虽然对我们这个共和国来说是全新的，但市场经济的运行在发达国家已有几百年的历史，市场经济的理论亦在不断发展完善，并形成了一个现代经济学理论体系。虽然许多经济学名著出自西方学者之手，研究的是西方国家的经济问题，但他们归纳出来的许多经济学理论反映的是人类社会的普遍行为，这些理论是全人类的共同财富。要想迅速稳定地改革和发展我国的经济，我们必须学习和借鉴世界各国包括西方国家在内的先进经济学的理论与知识。

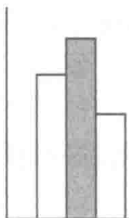
本着这一目的，我们组织翻译了这套经济学教科书系列。这套译丛的特点是：第一，全面系统。除了经济学、宏观经济学、微观经济学等基本原理之外，这套译丛还包括了产业组织理论、国际经济学、发展经济学、货币金融学、公共财政学、劳动经济学、计量经济学等重要领域。第二，简明通俗。与经济学的经典名著不同，这套丛书都是国外大学通用的经济学教科书，大部分都已发行了几版或十几版。作者尽可能地用简明通俗的语言来阐述深奥的经济学原理，并附有案例与习题，对于初学者来说，更容易理解与掌握。

经济学是一门社会科学，许多基本原理的应用受各种不同的社会、政治

或经济体制的影响，许多经济学理论是建立在一定的假设条件上的，假设条件不同，结论也就不一定成立。因此，正确理解并掌握经济分析的方法而不是生搬硬套某些不同条件下产生的结论，才是我们学习当代经济学的正确方法。

本套译丛于 1995 年春由中国人民大学出版社发起筹备并成立了由许多经济学专家学者组织的编辑委员会。中国留美经济学会的许多学者参与了原著的推荐工作。中国人民大学出版社向所有原著的出版社购买了翻译版权。北京大学、中国人民大学、复旦大学以及中国社会科学院的许多专家教授参与了翻译工作。前任策划编辑梁晶女士为本套译丛的出版做出了重要贡献，在此表示衷心的感谢。在中国经济体制转轨的历史时期，我们把这套译丛献给读者，希望为中国经济的深入改革与发展做出贡献。

《经济科学译丛》编辑委员会



序 言

越来越多的企业使用统计方法从数据中提取信息。对于准备进入商业领域的学生来说，仅仅掌握统计和计算方法是远远不够的。统计课程和教科书必须能够全面介绍统计概念以及统计方法在实际生活中的应用。《统计学：在经济和管理中的应用》一书的写作目的就是证明，对于今天的管理者和经济学家来说，统计方法是非常有效的工具。

为了完成这个目标，本书具有如下特征：首先，使用了大量包含数据的例子、习题和案例来说明统计方法能够被市场部经理、金融分析师、会计师、经济学家、运营经理和其他工作人员使用，很多例子和习题中使用的数据都是来自现实世界的较大数据集；其次，加强了本书的应用特色，即教会学生如何选择正确的统计方法；最后，介绍了一些统计概念以更好地解释统计结果。

本书的目的

商业世界非常复杂，因此只有有效的管理才能取得成功。管理复杂的事务需要很多技巧。竞争者越来越多，销售产品的渠道越来越多，招聘员工的公司也越来越多，总之，竞争变得越来越激烈。因此，有效的决策比以前更加重要。另外，经理们更容易获得规模较大和详细的数据，这些数据都是可能的信息来源。然而，要获得这些可能的信息就要求经理们能够从数据中提取信息。仅仅知道如何计算统计量是不够的。但是，这恰恰是大多数统计学教科书的内容——使用手工计算来说明一堆不相关的统计方法。这种模式已经延续了很多年。我们需要的是对应用统计方法的全面介绍。

1971年，作者开始讲授统计学时，教学中重点强调如何计算统计量，有时还包括如何推导公式。当时的观念认为，只有通过手工计算才能让学生理解各种方法和概念。

1988年作者初次出版本书的一个重要目标就是指导学生识别恰当的方法。接下来的八版又做了改进，同时强调解释与决策。作者用自己的方法把各种统计问题的解决都分成三步走，并在所有举例中都这样做：(1) 识别方法，(2) 计算统计量，(3) 解释结果。计

算过程可以通过手工（使用计算器辅助）、Excel、Minitab 三种方式中的任何一种或全部三种来完成。对于那些强调计算机使用的课程，手工计算可以简化乃至完全剔除。而对于强调手工计算的课程，亦可把计算机的使用当作选讲内容或完全忽略。具体情况可由教师根据需要做出选择。

作者认为，其思路有以下几项优点：

- 对识别和解释的强调使学生能够掌握可运用于其所面对的问题的可行技术，无论其所学课程偏重手工计算还是计算机的运用。

- 使学生深入体会到统计学是将数据转化为信息的一种工具。967 个数据文件以及相应的需要学生解释统计结果的习题，为学生运用数据进行分析与决策提供了广泛的机会。

- 计算机的选用使得学生可以面对规模更大以及现实性更强的习题和案例。

给出问题的计算过程使得教师能够关注决策问题更加重要的方面，例如，如何解释统计结果。为了恰当地解释统计结果，我们需要了解统计方法背后的概率论和统计概念以及问题的背景。本书将指导学生创建 Excel 工作簿供其进行假设分析，借此向他们介绍这些概念。学生可以很直观地看出改变统计方法的一部分所带来的影响，如增加样本量的影响。

统计学在教学方案中的位置使得教授统计课程变得非常困难。在大多数本科教学计划中，统计学都出现在第一年或者第二年。在研究生教学计划中，统计学也往往在三学期项目的第一学期或者两年项目的第一年开设。但是，会计学、经济学、金融学、人力资源管理、营销学和运营管理等课程则往往在统计学之后开设。因此，很多学生不能全面地理解统计方法的应用。本书通过加入专门介绍统计应用的章节和应用专栏来解决这个问题。对于那些学生不熟悉的应用，在介绍统计应用之前，本书会给出相关的背景知识。

- 例如，在介绍图示方法时，作者使用了一个例子来比较两种投资收益率的直方图。为了解释金融分析师想从直方图中得出什么信息，需要让学生知道收益率的方差可以用来度量风险。因此，在介绍这个例子之前，作者先给出了“在金融中的应用”应用专栏，并介绍了如何计算和使用投资收益率。

- 再如，在讲解正态分布时，作者加入了另一个“在金融中的应用”应用专栏来介绍为什么收益的标准差可以用来衡量风险。

- 本书共有 36 个应用专栏。

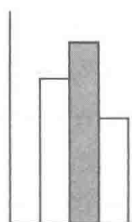
某些应用的规模太大，因此作者使用整个一节或者一小节来介绍这个主题。例如，在介绍总体比例的置信区间估计量的那一章中，作者用专门的一节介绍了市场细分的知识。在那一节，作者讲解了由总体比例的置信区间估计量如何得到细分市场规模的估计量。在其他章中，作者通过展示市场部经理如何应用统计方法来判断各细分市场的差别介绍了大量的统计方法。本书共有多个诸如此类的章节。那些名为“在……中的应用”的章节会向学生详细介绍“如何应用相应的统计方法”。

本书的创新之处

8 个大型数据集为我们提供了 369 道新习题，学生将会有更多机会分析真实数据，教师能够利用这些数据集寻找额外的例子和习题。

作者更新了第9版中很多真实数据的例子、习题和案例，包括棒球、篮球、足球和曲棍球球员的获胜次数、工资和上座率数据，纽约证券交易所、纳斯达克、多伦多证券交易所的股票收益率数据，以及全球变暖数据。

本书编写了很多新的例子和习题。第10版中共有145个附答案的例子、2148道习题、27个案例、967个数据集和35个附录（包含37个附有答案的例子、98道习题、25个数据集），即本书中共包含182个附有答案的例子、2246道习题、27个案例和992个数据集。

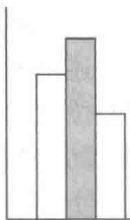


目 录

| | |
|--|-----|
| 第 1 章 什么是统计 | 1 |
| 1.1 主要统计概念 | 4 |
| 1.2 商业中的统计应用 | 5 |
| 1.3 大型的真实数据集 | 6 |
| 1.4 统计与计算机 | 7 |
| 附录 凯勒网站的应用说明 | 10 |
| 第 2 章 图表描述方法 I | 11 |
| 2.1 数据和信息的类型 | 12 |
| 2.2 对一组定类数据的描述 | 18 |
| 2.3 对两定类变量间关系的描述及两组或两组以上定类数据间的比较 | 37 |
| 第 3 章 图表描述方法 II | 50 |
| 3.1 对一组定距数据的描述 | 51 |
| 3.2 对时间序列数据的描述 | 71 |
| 3.3 两个定距变量间关系的描述 | 84 |
| 3.4 图表描述的艺术与科学 | 92 |
| 第 4 章 数字描述方法 | 106 |
| 4.1 中心位置的测度 | 107 |
| 4.2 变异性的测度 | 116 |
| 4.3 相对位置的测度和箱线图 | 125 |
| 4.4 线性相关性的测度 | 134 |

| | | |
|-------------|--------------------------|------------|
| 4.5 | 在职业运动中的应用：棒球（可选） | 151 |
| 4.6 | 在金融中的应用：市场模型（可选） | 154 |
| 4.7 | 图表方法和数字方法的比较 | 160 |
| 4.8 | 对数据描述的总结 | 163 |
| 附录 | 描述性方法的回顾 | 170 |
| 第5章 | 数据的收集与抽样 | 172 |
| 5.1 | 收集数据的方法 | 173 |
| 5.2 | 抽样 | 175 |
| 5.3 | 抽样方案 | 177 |
| 5.4 | 抽样与非抽样误差 | 183 |
| 第6章 | 概率 | 185 |
| 6.1 | 事件概率的分配 | 186 |
| 6.2 | 联合概率、边缘概率和条件概率 | 191 |
| 6.3 | 概率法则和概率树 | 204 |
| 6.4 | 贝叶斯法则 | 211 |
| 6.5 | 识别正确的方法 | 222 |
| 第7章 | 随机变量和离散型概率分布 | 231 |
| 7.1 | 随机变量和概率分布 | 232 |
| 7.2 | 二元（维）分布 | 245 |
| 7.3 | 在金融中的应用：投资组合多样化和资产分配（可选） | 253 |
| 7.4 | 二项分布 | 262 |
| 7.5 | 泊松分布 | 270 |
| 第8章 | 连续型概率分布 | 281 |
| 8.1 | 概率密度函数 | 282 |
| 8.2 | 正态分布 | 288 |
| 8.3 | 指数分布（可选） | 305 |
| 8.4 | 其他连续分布 | 308 |
| 第9章 | 抽样分布 | 322 |
| 9.1 | 均值的抽样分布 | 323 |
| 9.2 | 比例的抽样分布 | 333 |
| 9.3 | 两均值之差的抽样分布 | 339 |
| 9.4 | 由此推断 | 342 |
| 第10章 | 估计 | 345 |
| 10.1 | 估计的概念 | 346 |

| | | |
|---------------|--------------------|------------|
| 10.2 | 总体标准差已知时对总体均值的估计 | 349 |
| 10.3 | 样本容量的选择 | 362 |
| 第 11 章 | 假设检验入门 | 368 |
| 11.1 | 假设检验的概念 | 369 |
| 11.2 | 总体标准差已知时检验总体均值 | 372 |
| 11.3 | 犯第二类错误的概率 | 393 |
| 11.4 | 应用前瞻 | 402 |
| 第 12 章 | 总体的推断 | 406 |
| 12.1 | 标准差未知时对总体均值的推断 | 407 |
| 12.2 | 总体方差的推断 | 423 |
| 12.3 | 总体比例的推断 | 431 |
| 12.4 | 在市场营销中的应用：市场细分（可选） | 447 |
| 第 13 章 | 两总体对比的推断 | 462 |
| 13.1 | 两总体均值差异的推断：独立样本 | 463 |
| 13.2 | 观测数据和试验数据 | 491 |
| 13.3 | 两总体均值差的推断：配对试验 | 494 |
| 13.4 | 两总体方差比的推断 | 509 |
| 13.5 | 两总体比例差异的推断 | 516 |
| 附录 | 回顾第 12 章和第 13 章 | 544 |



什么是统计

引 言

统计是从数据中获取信息的一种方法，这就是统计的定义。本书的大部分内容讲述了管理者和统计工作者如何、何时以及为什么要引导和控制统计过程。读者可能会问：“如果那就是统计的全部，为什么这本书（包括大部分其他统计学教材）的内容如此之多呢？”这是因为应用统计学的学生需要面对形形色色的信息和数据。后面我们将会用一个案例和两个实例来说明这个问题。

例 3.3 商务统计学课程成绩（见第 3 章）

一名报名学习商学的学生正在上统计学的第一堂课。他有一点担心，因为他认为这门课程非常难。为了消除自己的紧张感，他向教授询问了去年的课程成绩。与其他统计学教授一样，这位教授非常友好且乐于助人，因此，他向这个学生提供了去年期末成绩表，其中包含平时成绩和期末考试成绩。这个学生可以从这些数据中得到什么信息呢？

这是一个典型的统计问题。这名学生拥有数据（成绩）且需要应用统计方法来获取他想要的信息。这就是描述统计学的功能。

□ 描述统计学

描述统计学（descriptive statistics）是以便利化和信息化的方式对数据进行整理、汇总、显示的一种方法。描述统计学的一种形式是图示方法，这种方法可使统计工作者以读者比较容易获取有用信息的方式呈现数据。在第 2 章和第 3 章中我们将会介绍多种作图方法。

描述统计学的另一种形式是利用数值方法对数据进行汇总。我们常用的平均数或均值就是这样一种方法。与计算一个公司员工平均年龄的方法相同，我们可以算出去年的统计学课程的平均成绩。第4章为大家介绍了几种描述数据不同特征的数值统计测度。

实际应用什么统计方法取决于我们想要得到什么信息。在上述例子中，我们可以看到至少3种重要信息。首先是“典型”成绩。我们称之为对中心位置的度量。平均值就是一种度量中心位置的方法。在第4章中，我们将会介绍另一个用来度量中心位置的量——中位数。如果这名学生被告知去年的平均成绩至少为67分，这条信息是否足够消除他的忧虑？他很可能做出否定的回答。因为，他想要知道，大多数人的成绩是都接近67分，还是很零散地分布在低于或高于均值的位置。也就是说，他需要对离散程度的度量。度量离散程度的最简单的量是全距，即最大值减去最小值。假设最高成绩为96分，最低成绩为24分。但是，仅仅依靠两个值，全距提供的信息还是太少了。在第4章中，我们将介绍其他度量离散程度的量。此外，该学生还想了解关于成绩的更多信息，尤其需要知道成绩在24分至96分之间是如何分布的。实现这一目的的最好方法是图示法，即在第3章中要介绍的直方图。

案例 12.1

百事可乐公司和某大学的专营协议

数据：C12-01

在过去的几年中，不少大学和各式各样的公司签订了专营协议，协议约定该学校在校园里只能销售该公司的产品。许多食品公司和饮料公司都纷纷和各所大学签订了协议。

一所共有50 000名学生的大学向百事可乐公司提出了签订专营协议的要约，该协议将授予百事可乐公司下一年对这所大学的所有校园便利店的专营权，以及以后各年优先签订协议的选择权。作为回报，校方将收取百事可乐公司35%的校园销售收入，此外，每年还一次性收取200 000美元的费用。百事可乐公司方面有两周的考虑时间。

百事可乐公司的管理层迅速回顾了他们所掌握的信息。软饮料市场是用12盎司的罐为单位来测量的。百事可乐公司目前平均每周的销售量为22 000罐饮料（每年可以在校园内销售40个星期），每罐饮料平均卖1美元，成本中包括每罐0.30美元的劳动力成本。百事可乐公司并不确定它的市场占有率，但是它估计大大低于50%。一份市场分析显示，假如百事可乐公司现有的市场占有率达到了25%，这项专营权就能使百事可乐公司每周售出88 000罐饮料，这样，每年的销售量将达到3 520 000罐（用每周销售量88 000乘以40周）。收入总额可以用以下方法计算^①：

$$\text{收入总额} = 3\,520\,000 \text{ 罐} \times 1.00 \text{ 美元/罐} = 3\,520\,000 \text{ 美元}$$

由于学校要抽取总收入的35%，所以这个数字必须乘以65%：

$$65\% \times 3\,520\,000 \text{ 美元} = 2\,288\,000 \text{ 美元}$$

^① 我们已经构造了对案例中的有关数据进行了计算的 Excel 工作表。参见凯勒网站的附录1，了解如何下载该工作表以及成百上千的数据集。

扣除每罐的劳动力成本 0.30 美元（即 1 056 000 美元）和一次性支付给学校的 200 000 美元，百事可乐公司得到的净利润为

$$2\,288\,000 \text{ 美元} - 1\,056\,000 \text{ 美元} - 200\,000 \text{ 美元} = 1\,032\,000 \text{ 美元}$$

百事可乐公司目前的年利润为

$$40 \text{ 周} \times 22\,000 \text{ 罐/周} \times 0.70 \text{ 美元/罐} = 616\,000 \text{ 美元}$$

如果目前百事可乐公司的市场占有率为 25%，那么该项协议的潜在收益为

$$1\,032\,000 \text{ 美元} - 616\,000 \text{ 美元} = 416\,000 \text{ 美元}$$

上述分析存在的唯一问题是，百事可乐公司不知道软饮料每周在学校里的销量。而且，可口可乐公司也不可能告知百事可乐公司它的销售情况。百事可乐公司和可口可乐公司的产品加起来几乎占据了整个软饮料市场。

于是百事可乐公司安排了一个刚毕业的大学生来为它调查未知信息。这个学生在校园里随机选取了 500 名学生，让他们记录下在未来的七天中一共在学校里买了多少罐软饮料。

该调查得到的信息储存在凯勒网站附录的文件 C12-01 里，读者可到该网站下载。

□ 推断统计学

在案例 12.1 中，我们想要估计由独家经营协议带来的年利润。数据就是样本中的 500 名学生接下来的 7 天内软饮料的购买数量。我们可以使用描述统计方法来分析这些数据。

当然，在案例 12.1 中，我们对这 500 名学生的回答并不是十分感兴趣，因为我们想要了解的是学校全部 50 000 名学生的平均软饮料消费量。为了实现这一目标，我们需要用到统计学的另一个分支——推断统计学。

推断统计学是利用样本数据信息对总体特征做出推断的一系列方法。在案例 12.1 中，总体是指该大学的 50 000 名学生，我们感兴趣的特征是这一总体的软饮料消费量，而对每一位学生进行调查是极其费时费力的。统计方法的存在使得我们不必逐人调查。相反，我们可以从中抽取相对来说很少的部分学生作为样本（该样本容量为 500），利用样本信息，就可以推知所有 50 000 名学生软饮料的消费量了。据此就可对百事可乐公司的年利润做出估计。

例 12.5 票站调查（见第 12 章）

在政府选举期间，电视台通常会取消常规电视节目而代之以有关选举的新闻报道。计票完毕，结果才公布于众。但是，对一些很重要的职位，比如进行总统或大州参议员的选举时，各电视台就会积极竞争看谁能首先预测出获胜者，而这正是通过票站调查完成的。所谓票站调查（exit polls），就是从投完票离开票站的选民中随机抽取一部分，询问他们选的是谁。根据调查数据可以计算出支持某个候选人的选民的比例。应用统计方法就可以判断暂时领先的候选人是否能获得足够多的选票从而赢得选举。假设在 2000

年美国大选中佛罗里达州的票站调查结果被记录下来。虽然在大选中有若干候选人角逐总统，但是民意测验专家只记录了两个有希望赢得大选的人的得票情况，那就是民主党候选人艾伯特·戈尔（代码=1）和共和党候选人乔治·W. 布什（代码=2）。投票结果（共有765人投了布什或者戈尔的票）储存在文件 Xm12-05 中。这样，电视台的分析家就可以判断出乔治·W. 布什是否将在佛罗里达州大选中胜出。

例 12.5 介绍了推断统计学的一个简单应用。电视台要做出推断的总体是投票给布什或戈尔的约 500 万位佛罗里达选民，样本由投票调查公司随机选择的投票给两位主要候选人的 765 位选民组成。我们感兴趣的总体特征是投票给布什的选民比例，尤其想了解是否有超过 50% 的选民选择了布什（只计投票给共和党 and 民主党候选人的选民）。必须清楚的一点是，由于我们不能确切得知 500 万位选民每个人的投票情况，因此预测的结果不会 100% 地准确。这是统计工作者甚至学过统计学的学生都应该知道的常识。样本作为总体的一小部分只能以一定的概率对总体做出正确推断。统计工作者可以控制样本容量，而且一般设置正确率为 90%~99%。

出乎意料的是，2000 年 11 月的美国大选之夜，电视台犯了一个大错。利用票站调查和前期选举的结果，晚上 8 点，四大电视台得出结论：戈尔将在佛罗里达州胜出。刚过晚上 10 点，大部分计票工作已经完成，电视台又调过头来宣称乔治·W. 布什将会在该选举中胜出。到凌晨 2 点，又一个预测结果宣布出来：得票数很相近以至于难以判断谁将获胜。自那以后，统计学教师在讲授过程中很可能把它作为怎样使用统计的一个例子。

与你之前认为不同的是，数据可能并不一定是数值型的。当然，例 3.3 中的成绩和案例 12.1 中每个学生每周购买的软饮料数量都是数值型的；但是，例 12.5 中的选票却不是数值型的。在第 2 章中，我们将会详细讨论你在统计应用中可能会遇到的数据类型以及相应的处理方法。

1.1 主要统计概念

统计推断问题有三个主要概念：总体、样本和统计推断。下面我们将分别予以详细讨论。

□ 1-1a 总体

所谓**总体** (population)，就是统计工作者研究对象的全体。一般而言，总体容量是很大的，而且可能无限大。在统计范畴中，总体并不一定指的是一组人，也可以是事物。比如，某大工厂生产的全部滚珠轴承就是一个总体。在案例 12.1 中，总体是该学校的 50 000 名学生。而在例 12.5 中，投票给布什或戈尔的佛罗里达州选民则构成了总体。

对总体的描述性测度称为**参数** (parameter)。在案例 12.1 中，我们感兴趣的参数是该大学所有学生软饮料的平均消费量。在例 12.5 中，参数则是 500 万位佛罗里达州选民中投票给布什的选民所占的比例。在大多数推断统计学的实际应用中，参数代表我们感兴趣的信息。

□ 1-1b 样本

样本 (sample) 是从所研究的总体中抽取的若干数据的集合。对样本的描述性测度是**统计量** (statistic), 我们就是用统计量对参数进行推断的。在案例 12.1 中, 计算的统计量是样本中的 500 名学生一周内软饮料的平均消费量。然后, 我们利用样本均值对总体均值进行估计; 而总体均值正是该问题中我们想要了解的参数。在例 12.5 中, 样本由 765 位佛罗里达州选民组成, 我们计算两名主要候选人中每一个的得票比例, 然后用样本统计量对 500 万位选民所构成的总体比例进行推断。这样, 在计票结果出来之前, 我们就可以对大选结果进行预测。

□ 1-1c 统计推断

统计推断 (statistical inference) 是运用样本数据对总体进行估计、预测和决策的过程。由于总体一般相对较大, 因此, 对总体中的每一个体进行调查是不切实际的, 并且花费很高。而从总体中抽取一个样本就很简单, 花费也低得多, 并且运用样本信息很容易对总体进行估计并做出判断。当然, 这些估计和结论并不一定都正确, 故我们在统计推断中加入可靠性测度。可靠性测度共有两种: **置信水平** (confidence level) 和**显著性水平** (significance level)。置信水平是一个估计过程正确的概率。比如在案例 12.1 中, 对 50 000 名学生软饮料的平均消费量就可建立一个置信水平为 95% 的区间估计, 这意味着, 以这种统计推断形式为基础进行的估计有 95% 的概率是正确的。当统计推断的目的是对总体做出判断时, 显著性水平测度了错误结论发生的频率。例如, 考虑对例 12.5 的结果进行分析, 我们确定超过 50% 的选民将票投给了乔治·W. 布什, 这样, 他将在佛罗里达州大选中获胜。5% 的显著性水平意味着由该样本推断布什将在大选中获胜的结论的错误率是 5%。

1.2 商业中的统计应用

商业和经济学项目中统计学课程的一个重要功能是, 证明统计分析实际上在商业和经济学的各个方面都起着重要作用。我们希望通过例子、习题和案例来说明这一功能。我们假设第一次接触统计学课程的这些学生还没有学过管理学项目中的其他课程, 然而, 为了更好地理解统计学在商业、经济学以及其他课程中的应用, 对这些课程有一些基本了解是非常有必要的。为了给统计学的应用提供充分的背景知识, 我们会介绍统计学在会计学、经济学、金融学、人力资源管理、市场营销和运营管理方面的应用。我们将会通过以下两种方式向读者介绍这些应用的背景知识。

□ 1-2a 应用章节和分节

我们通过五个章节来描述统计学在商业领域职能部门的应用。例如, 在 7.3 节, 我们介绍了统计学在金融学中的应用, 描述了一个金融分析人员如何应用概率论和数理统计的知识构造能够降低风险的投资组合。

我们用一整节和一个小节的篇幅来说明概率论和数理统计在某些行业的应用。4.5

节介绍了统计学在职业棒球中的应用。6.4 节中的一个小节介绍了统计学在医学测试（医学测试在医药行业是非常重要的）中的应用。

□ 1-2b 应用专栏

对于其他不需要很详细描述的问题，我们以应用专栏的方式给出一个相对简短的背景知识描述，然后介绍例子或者练习。这些应用专栏将会贯穿整本书。例如，在第 3 章中，我们将会讨论一个营销经理为了确定产品的合理价格所需做的工作。为了更好地理解这个例子，我们需要对营销经理这个角色做一个简单的描述，然后介绍统计学的应用。

1.3 大型的真实数据集

我们认为，学生可以在亲自解决统计问题的过程中学习统计。当学生离开学校后，我们希望他们具备整合大量真实数据、获取制定决策所需信息的能力。为了提供练习这项重要技能的数据，我们创建了 8 个大型的真实数据集，读者可以到凯勒网站下载。这些数据来自综合社会调查（GSS）和美国国家选举调查（ANES）。

□ 1-3a 综合社会调查

自 1972 年起，GSS 已经跟踪调查了美国民众对诸多问题的态度。除了美国人口普查外，GSS 是最常被用到的美国社会信息来源。当前，该调查每两年进行一次，涉及数百个变量和数千个观测样本。我们引用最近 6 次（2002 年、2004 年、2006 年、2008 年、2010 年和 2012 年）的调查数据，保存文件分别为 GSS2002，GSS2004，GSS2006，GSS2008，GSS2010 和 GSS2012，样本容量分别为 2 765，2 812，4 510，2 023，2 044 和 1 974。我们将变量的数量减少到 60 个左右，删除了含有缺失数据（不知道、拒绝回答等）的样本。

数据集中包括一些人口学变量，例如年龄、性别、种族、收入和受教育程度。其他变量是关于公众的政治态度、对政府活动和工作的支持率。每次调查的全部变量列表保存在配套网站的附录 GSS2002，GSS2004，GSS2006，GSS2008，GSS2010 和 GSS2012 中。

本书中的很多例子和习题源于这些数据集。

□ 1-3b 美国国家选举调查

美国国家选举调查在总统选举时期进行，目的是提供分析美国公民选举行为的数据。我们引用了 2004 年和 2008 年的调查数据。像 GSS 一样，ANES 包括一些人口学变量。同时，它还包括总统候选人的支持意向、政治主张和政治背景变量。配套网站的附录 ANES2004 和 ANES2008 中有这些变量的名称和定义。

2008 年的调查数据中重复抽选了非洲裔和西班牙裔选民，我们通过在这两个种族群体中随机删除样本来“调整”数据。

就像处理 GSS 数据集一样，我们删除了缺失数据。