

Mc
Graw
Hill

Education



世界优秀教材中国版

COMPLETE BUSINESS STATISTICS

[美] AMIR D.ACZEL JAYAVEL SOUNDERPANDIAN 著

商务统计

■ 谭英平 李小虎 等译

谭英平 审校



高等教育出版社
Higher Education Press

F712.3
AKS



世界优秀教材中国版

COMPLETE BUSINESS STATISTICS

[美] AMIR D.ACZEL JAYAVEL SOUNDERPANDIAN 著

商务统计

谭英平 李小虎 等译

谭英平 审校



高等教育出版社
Higher Education Press

图字:01-2006-3072号

Aczel

COMPLETE BUSINESS STATISTICS, 6e

ISBN:0-07-286882-1

Copyright © 2006 by McGraw-Hill Companies, Inc.

Original language published by The McGraw-Hill Companies, Inc. All Rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Simplified Chinese translation/adaptation edition is published and distributed exclusively by Higher Education Press under the authorization by McGraw-Hill Education (Asia) Co., within the territory of 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. Violation of this Law is subject to Civil and Criminal Penalties.

本书中文改编版由高等教育出版社和美国麦格劳-希尔教育出版(亚洲)公司合作出版。此版本仅限在中华人民共和国境内(但不允许在中国香港、澳门特别行政区和中国台湾地区)销售。未经许可之出口,视为违反著作权法,将受法律之制裁。

未经出版者预先书面许可,不得以任何方式复制或抄袭本书的任何部分。

本书封面贴有 McGraw-Hill 公司防伪标签,无标签者不得销售。

图书在版编目(CIP)数据

商务统计/(美)阿克泽尔(Aczel, A. D.), (美)

桑德班迪安(Sunderpandian, J.)著;谭英平,李小虎译.

—北京:高等教育出版社, 2008. 6

书名原文:Complete Business Statistics

ISBN 978-7-04-022887-8

I. 商… II. ①阿…②桑…③谭…④李…

III. 商务统计 IV. F712.3

中国版本图书馆 CIP 数据核字(2008)第 067287 号

出版发行	高等教育出版社	购书热线	010-58581118
社 址	北京市西城区德外大街 4 号	免费咨询	800-810-0598
邮政编码	100120	网 址	http://www.hep.edu.cn
总 机	010-58581000		http://www.hep.com.cn
经 销	蓝色畅想图书发行有限公司	网上订购	http://www.landracom.com
印 刷	国防工业出版社印刷厂		http://www.landracom.com.cn
开 本	787×1092 1/16	畅想教育	http://www.widedu.com
印 张	44	版 次	2008 年 6 月第 1 版
字 数	1 130 000	印 次	2008 年 6 月第 1 次印刷
		定 价	63.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 22887-00

献给我的父亲 E. L. Aczel 上尉

和我的母亲 Miriam Aczel

Amir D. Aczel

献给我的父亲 V. M. Sounderpandian

Jayavel Sounderpandian

关于作者

Amir D. Aczel 就职于波士顿大学的哲学与历史科学中心,也是美国 CNN 的兼职评论员. 他在加利福尼亚大学伯克利分校获得数学学士学位和运筹学硕士学位,并在俄勒冈大学获得统计学博士学位. 在担任现在的职位之前,Aczel 曾先后在加利福尼亚、阿拉斯加、欧洲、中东以及远东等地区居住并任教.

Aczel 著有数学、商务和统计学等方面 11 本书籍. 最知名的一本名为《*Fermat's Last Theorem*》,探讨了迄今未得到证明的一个数学定理的证明方法. 最近出版的一本名为《*The Riddle of the Compass*》的书探讨了指南针的发明.

Aczel 撰写的《商务统计学》这本书中所表现出的广泛影响力,反映了作者在国际上活动的丰富经历,在教育领域的权威以及在其他著作中所表现出来的实践爱好. 目前,Aczel 与妻子和女儿居住在大波士顿区.

Jayavel Sounderpandian 是威斯康星大学定量方法的教授,在该校教了 21 年的商务统计与运筹管理. 他在印度马德拉斯科技学院获得机械工程学士学位,并在 Kent 州立大学获得商业硕士和博士学位. 在转入学术界之前,Sounderpandian 在印度一家航空制造公司担任了七年的工程师.

除本书外,Sounderpandian 还撰写了有关电子数据表在商务统计、市场调查和运筹管理中的应用等四本书籍. 他在电子数据模板应用方面有多年的课堂讲授经验,并在本书中使用了 100 多个模板. Sounderpandian 的研究兴趣在于决策和风险分析,他在一系列权威期刊上发表了大量文章,包括 *Operations Research*, *Abacus*, *Journal of Risk and Uncertainty*, *International Journal of Production Economics* 和 *Interfaces*.

在业余时间,Sounderpandian 写作并解答一些数学疑问. 本书大量的个案研究体现了他的数学创造力和从业经验. Sounderpandian 与他的妻子居住在威斯康星州拉辛,他们有三个女儿.

前言

在《商务统计学》第六版中,我们重点关注书中的许多改进之处,其中大部分是受热心的读者和其他商务统计学的讲授者们的建议所推动的。他们建议通过加强 Excel 特色、增加新的教学内容和更新原始资料等方法来进一步改进本书。

其中一个改进之处是,Excel 被完整地融合到了第六版中。虽然在本书以前的版本中就已经引进了 Excel 模板,但在这一版本中还是首次将 Excel 有规则的编排到书籍的结构和教学中。因此,本书在适用的部分将更加完整的介绍 Excel,这也使其成为本书一个更为突出的内容。

做出这一改变的动机是由于目前 Excel 和办公软件在工作室中的广泛使用,以及许多商业学院越来越需要将 Excel 的使用纳入到商务统计学课程的教授目标中。当然,统计软件包不断地被应用于商务统计学的课程中,也很少有人会认为这些软件包在统计应用中比 Excel 低级。因此,在本书的第六版中尽管更加强调 Excel,但统计软件包的使用仍然会作为一个可行的方法供偏好它的读者们选择。然而,目前绝大部分管理者和其他从业者们都直接选择 Excel。我们在《商务统计学》第六版中进一步强调 Excel 是对当前的商务环境、世界计算环境以及我们所教授的即将加入到这一社会环境中的学生们的一个回应。

第六版的其他改进之处包括对 Excel 模板的修订和增加。尽管这些模板并不是没有缺陷——最严重的是学生们可能会没有理解模板背后的概念就直接使用它们——但我们相信应用这些模板的益处还是多于它的不足。还需要注意的是,尽管我们相信这些模板对于解决问题是有用和有效的,我们仍然保留了手工计算的说明,这样学生们可以对模板中的任何结果进行手工计算。

除上述变动之外,第六版还包括以下一些改进:

1. 每章的学习目标。
2. 许多新的问题和例题。
3. 部分章节中增加了新的案例,包括 Excel 和 Excel 模板的使用。
4. 第 2 章增加了新的有关联合概率表的资料。
5. 第 3 章增加了新的有关随机变量以及线性组合的资料,增加了新的有关风险价值的章节,并对新的模板应用结果进行了更新。
6. 第 5 章增加了有关自由度的新内容。
7. 第 7 章增加了有关 β 计算的新的章节。
8. 第 10 章标题为“随机变量的线性组合”的小节中增加了有关协方差的新的资料和模板。这一资料在金融投资组合中有直接应用。
9. 许多章节都增加了新的带有屏幕点击图示的章节附录,用于介绍 Excel 和 Excel 命令的应用。
10. 本书的第五版曾首次在它的学生版 CD-ROM 中提供了“可视统计”,在第六版中将通过增加书中边缘上的“VS”图标来进一步完善该功能。这个图标会指明书中的相关内容,学生可以通过利用学生版 CD 上的可视统计进一步学习和实践来获得有关这些内容更多的收益。

至此,我们已重点关注了第六版中所作的一些修订。不过,第六版中所保留的其他内容与所

增加的新的内容同样重要。与以前的版本一样,本书涵盖了商务统计中所有重要的主题,只是在第六版中对这些内容的重要性进行了重新编排以反映商务实践的演变过程。那些在商业学院不被普遍教授的主题将被精简,而在当前的商业社会中不具有普适性的一些统计方法将被删除。第16章的部分内容被合并到第5章,但第16章的全部内容仍可在附于书末的学生版CD-ROM中找到。第17章也同样包含在其中。所有这些章节都包含了学习目标和更新资料。

与以前的版本一样,本书的第六版仍将保持它在全球的重要性,保持它在世界商务出版物中的前沿地位。世界各国的经济正变得越来越交织缠绕。亚洲和中东发生的事件会对华尔街产生直接影响,俄罗斯经济的资本主义化也会直接影响欧洲和美国。出版业的所有企业组成了一个国际大联盟;金融业的证券交易昼夜不停地在世界各地的市场中进行;零售业在世界各地进行生产来为消费者提供产品——所有这些都证实了世界经济普遍存在的全球化现象。因此,本书新的版本中大部分的例题都是有关国际问题的。我们希望这一方式会受到讲师们的欢迎,因为它将越来越反映几乎所有的商务问题。

许多人士为本书第六版的修订作出了巨大贡献,在此我们向所有人表示感谢。他们分别是:

因 Fattaneh Cauley, Pepperdine 大学

Wen Chiang, Tulsa 大学

Christiaan Heij, 荷兰 Erasmus 大学

Zhimin Huang, Adelphi 大学

Wolfgang Jank, Maryland 大学

Felix Kamuche, Morehouse 大学

John F. Kottas, William & Mary 大学

Supriya Lahiri, 洛厄尔 Massachusetts 大学

Thomas J. Page, Michigan 州立大学

David Pentico, Duquesne 大学

Rakesh Sarin, 洛杉矶 California 大学

Milan Velebit, 芝加哥 Illinois 大学

Donald Wise, Rider 大学

感谢负责撰写本书附录的作者们。Keller 管理研究所的 Lou Patille 修订了讲师手册和学生答疑手册。Texas-Arlington 大学的 Alan Cannon 更新了测试库, Morehead 州立大学的 Lloyd Jaisingh 建立了数据文件,并更新了幻灯片演示软件。Illinois 州立大学的 Don Robinson 对校样进行了审阅。还要特别感谢 Oakland 大学的 David Doane, Ronald Tracy 和 Kieran Mathieson, 在他们的许可下,统计软件包“可视统计”被用于本书的 CD-ROM。

感谢 McGraw-Hill/Irwin 出版社热心的工作人员们。感谢行政总编 Scott Isenberg 在修订第六版过程中的战略指导。感谢高级发展编辑 Wanda Zeman 对出版计划良好的管理,从而保证了计划进度的按时开展并将成本控制在预算以内。我们向 McGraw-Hill/Irwin 出版社工作团队高效的编辑、排字和印刷表示感谢。

Amir D. Aczel Jayavel Sounderpandian

波士顿大学 威斯康星 Parkside 大学

译者前言

有人说,举出一个统计没有用途的领域名称比举出一个统计作为其组成部分的领域名称要困难得多.还有人说,对于追求效率的公民而言,统计思维总有一天会和读写能力一样必要.在科学技术高度发达的今天,作为一种探寻真理的工具,统计的重要性无疑是不言而喻的.也正是出于对统计恰似哲学的本质特征的无尽热爱和兴趣,我们受高等教育出版社的委托,翻译了波士顿大学的 Amir D. Aczel 和威斯康星大学的 Jayavel Sounderpandian 共同编写的这本《商务统计学》.本书是一本优秀统计学教材,现在为大家奉上的是其经过多次修订后的第六版.

随着我国对统计学学科和统计学教育的日益重视,先后翻译出版的统计学相关教材和著作十分繁多,如果不能以更更新的角度和颇具特色的风格为读者展现丰富而迷人的统计学知识,继续这样的翻译工作将显得毫无意义.但当几经周折完成本书的翻译并成功付梓时,我们在冗长而辛苦的译稿过程中体味到了欣慰和企盼.本书以每章开篇的“统计应用”作为引领读者进入各种统计方法和思想的“小星星”,让读者在充满趣味与感叹的氛围下逐步探索和理解统计知识的“大宇宙”,每章结束部分的案例更能让读者在掌握了相应知识的基础上再次回到现实商务活动的应用.我们相信,这样的结构设计 with 内容编排能够恰如其分的向读者展现统计知识的魅力,而通俗易懂的措辞更能让不同领域的读者准确轻松的理解各种统计方法的内涵.此外,原书的两位作者受热心的读者和其他商务统计学讲授者们的建议所推动,在这一版的书稿中着重加强了 Excel 软件的应用,这无疑是帮助读者更好的适应当前高效的商务环境、计算环境一个十分有益的举措.

除两位主译人员之外,中国人民大学统计学院的孙春阳、余巍、林俊、姚睿、林春升、陶茜等硕士研究生也参与了本书的部分翻译工作,而贾俊平老师更是用他多年从事统计教学的丰富经验和对统计学的深刻理解给我们的翻译工作提供了极为有益的指导,高等教育出版社的曾飞华编辑也为本书的顺利出版付出了大量的劳动,在此对他们一并表示衷心的感谢.

翻译永远是一门遗憾的艺术,尽管我们倾注全力希望把原著自身的精彩毫无遗漏的展现在读者面前,但受译者的理论水平和文字功底所限,译作中肯定存在诸多的不足甚至错误,敬请每位热爱统计的读者积极指正.也衷心企盼当您开启这本译作的扉页时,能够踏上一趟美妙的统计知识之旅!

谭英平

2008年4月

目录摘要

第 1 章	导论与描述统计	1
第 2 章	概率	42
第 3 章	随机变量	75
第 4 章	正态分布	128
第 5 章	抽样与抽样分布	155
第 6 章	置信区间	185
第 7 章	假设检验	215
第 8 章	两总体比较	252
第 9 章	方差分析	291
第 10 章	简单线性回归与相关	342
第 11 章	多元回归	401
第 12 章	时间序列、预测和指数	478
第 13 章	质量控制和改进	511
第 14 章	非参数方法和卡方检验	532
第 15 章	贝叶斯统计与决策分析	587
附录		631
附录 A	参考书目	631
附录 B	大部分奇数题答案	635
附录 C	统计表	646

目 录

第1章 导论与描述统计	1
1-1 统计应用	1
1-2 百分位数与四分位数	6
1-3 集中趋势的测度	8
1-4 离散程度的测度	12
1-5 分组数据与直方图	18
1-6 偏度与峰度	20
1-7 均值与标准差的关系	21
1-8 展示数据的方法	22
1-9 探索性数据分析	28
1-10 小结与术语回顾	34
案例1:纳斯达克的波动	41
第2章 概率	42
2-1 统计应用	42
2-2 基本定义:事件、样本空间和概率	44
2-3 概率的性质	48
2-4 条件概率	51
2-5 独立事件	56
2-6 组合的概念	60
2-7 全概率定理和贝叶斯定理	62
2-8 联合概率分布表	67
2-9 计算机应用	68
2-10 小结与术语回顾	70
案例2:申请工作	74
第3章 随机变量	75
3-1 统计应用	75
3-2 离散型随机变量的期望	86
3-3 随机变量的和与线性组合	90
3-4 伯努利随机变量	95
3-5 二项随机变量	96
3-6 负二项分布	101
3-7 几何分布	102
3-8 超几何分布	104
3-9 泊松分布	106
3-10 连续型随机变量	108

3-11 均匀分布	111
3-12 指数分布	112
3-13 小结与术语回顾	115
案例3:微型集成电路芯片合同	125
案例4:谷物食品公司促销	126
本章附录:一些标准分布的 Excel 公式	127
第4章 正态分布	128
4-1 统计应用	128
4-2 正态分布的性质	130
4-3 模板	131
4-4 标准正态分布	134
4-5 正态随机变量的转换	139
4-6 逆变换	144
4-7 二项分布的正态逼近	148
4-8 小结和术语回顾	150
案例5:可接受的钉子	153
案例6:多种货币决策	153
本章附录:正态分布的 Excel 函数	154
第5章 抽样与抽样分布	155
5-1 统计应用	155
5-2 估计总体参数的样本统计量	157
5-3 抽样分布	162
5-4 估计量及其性质	172
5-5 自由度	175
5-6 模板	180
5-7 小结与术语回顾	181
案例7:钉子的抽样	184
第6章 置信区间	185
6-1 统计应用	185
6-2 总体标准差已知时总体均值的 置信区间	186
6-3 σ 未知时 μ 的置信区间—— t 分布	193
6-4 总体比例 p 的大样本置信区间	199
6-5 总体方差的置信区间	202

6-6 样本量的确定	205	10-2 简单线性回归模型	345
6-7 模板	207	10-3 估计:最小二乘法	347
6-8 小结与术语回顾	210	10-4 误差方差和回归估计的标准误	357
案例 8:总统民意测验	213	10-5 相关	363
案例 9:隐私问题	214	10-6 回归关系的假设检验	366
第 7 章 假设检验	215	10-7 回归效果如何?	370
7-1 统计应用	215	10-8 方差分析表与回归模型的 F 检验	374
7-2 假设检验的概念	218	10-9 残差分析和对模型不足的检查	375
7-3 计算 p -值	223	10-10 应用回归模型进行预测	384
7-4 假设检验	229	10-11 回归的 Solver 方法	386
7-5 先验决策	243	10-12 非独立随机变量的线性组合	390
7-6 小结与术语回顾	250	10-13 小结与术语回顾	394
案例 10:累人的轮胎 I	251	案例 14:卫生保健的资产回报	396
第 8 章 两总体比较	252	案例 15:风险与回报	397
8-1 统计应用	252	本章附录:LINEST 函数	399
8-2 配对数据比较	253	第 11 章 多元回归	401
8-3 独立随机样本下两总体均值 之差的检验	259	11-1 统计应用	401
8-4 两总体比例之差的大样本检验	271	11-2 k 个变量多元回归模型	403
8-5 F 分布与两总体方差是否 相等的检验	275	11-3 多元回归模型的 F 检验	407
8-6 小结与术语回顾	283	11-4 回归效果如何?	410
案例 11:累人的轮胎 II	286	11-5 单个回归参数的显著性检验	413
本章附录:使用 Excel 检验两总体 均值之差	288	11-6 检验回归模型的有效性	424
第 9 章 方差分析	291	11-7 应用多元回归模型进行预测	430
9-1 统计应用	291	11-8 定性自变量	433
9-2 方差分析的假设检验	292	11-9 多项式回归	441
9-3 方差分析的理论及计算	296	11-10 非线性模型及转换	447
9-4 方差分析表及例子	304	11-11 多重共线性	456
9-5 进一步分析	310	11-12 残差自相关与 Durbin-Watson 检验	462
9-6 模型、因子及设计	316	11-13 偏 F 检验及变量选择方法	465
9-7 双因素方差分析	318	11-14 运用 Solver 实现多元回归	469
9-8 区组化设计	330	11-15 小结与术语回顾	471
9-9 小结与术语回顾	334	案例 16:四个不同部门的资产回报	473
案例 12:统一的制服	336	本章附录:多元回归的 LINEST 函数	475
案例 13:检验收银台	337	第 12 章 时间序列、预测和指数	478
本章附录:使用 Excel 进行方差分析	338	12-1 统计应用	478
第 10 章 简单线性回归与相关	342	12-2 趋势分析	480
10-1 统计应用	342	12-3 季节性和循环波动	485
		12-4 比率移动平均法	488

12-5	指数平滑法	496	14-8	Spearman 秩相关系数	562
12-6	指数	501	14-9	拟合优度的卡方检验	566
12-7	小结与术语回顾	508	14-10	列联表分析——独立性的 卡方检验	572
案例 17: 汽车零部件销售预测		509	14-11	比例相等的卡方检验	578
第 13 章 质量控制和改进		511	14-12	小结与术语回顾	581
13-1	统计应用	511	案例 19: 北美的九个国家		584
13-2	W. Edwards Deming 的指导	512	第 15 章 贝叶斯统计与决策分析		587
13-3	统计学和质量	513	15-1	统计应用	587
13-4	$\bar{x}$ 图	519	15-2	贝叶斯定理和离散概率模型	588
13-5	R 图和 s 图	522	15-3	贝叶斯定理和连续概率分布	595
13-6	p 图	525	15-4	主观概率的评估	599
13-7	c 图	527	15-5	决策分析: 总论	600
13-8	x 图	529	15-6	决策树	603
13-9	小结与术语回顾	529	15-7	应用贝叶斯定理处理附加信息	611
案例 18: Nashua 公司的质量控制 和改进		530	15-8	效用	619
第 14 章 非参数方法和卡方检验		532	15-9	信息的价值	622
14-1	统计应用	532	15-10	计算机应用	625
14-2	符号检验	533	15-11	小结与术语回顾	626
14-3	游程检验——随机性的检验	537	案例 20: Pizzas 'R' Us		628
14-4	Mann-Whitney U 检验	542	案例 21: 新药物的开发		629
14-5	Wilcoxon 符号秩检验	547	附录		631
14-6	Kruskal-Wallis 检验——单因素 ANOVA 的非参数替代方法	552	附录 A	参考书目	631
14-7	随机化区组设计的 Friedman 检验	558	附录 B	大部分奇数题答案	635
			附录 C	统计表	646

第1章 导论与描述统计

1-1 统计应用	1-7 均值与标准差的关系
1-2 百分位数与四分位数	1-8 展示数据的方法
1-3 集中趋势的测度	1-9 探索性数据分析
1-4 离散程度的测度	1-10 小结与术语回顾
1-5 分组数据与直方图	案例 1: 纳斯达克的波动
1-6 偏度与峰度	

学习目标

通过本章的学习,应掌握以下内容:

- 区分定性与定量数据.
- 描述定类、定序、定距和定比测度.
- 描述总体与样本的区别.
- 计算并解释百分位数与四分位数.
- 理解集中趋势的测度并会计算.
- 绘制用于描述离散数据的不同形式的图形.
- 使用 Excel 模板计算各种测度并制图.

1-1 统计应用

粗略的准确胜于精确的错误.

——John Maynard Keynes

你们可能都听过关于 Malcolm Forbes 的故事,他在一次著名的热气球飞行中迷了路,最后降落在一片玉米地的中央.他看到一个人向他走来,便问道:“先生,您能告诉我,我在什么地方吗?”那个人答道:“当然可以,你在一片玉米地中央的一个篮子里.”Forbes 说:“您一定是个统计学家.”那个人说:“真奇怪,你怎么知道的呢?”“很容易,”Forbes 说:“你所提供的信息简明而准确,同时又完全无用.”^①

本书的目的在于使你相信,一个好的统计分析所得到的信息永远都是简明、准确而又有用的!事实上,前文所引用的 Keynes 的那句话恰如其分地说出了统计的灵魂.本书将教会你怎样在多数时候能够至少得出大致准确的结论.统计是一门科学,它能帮助我们在商业、经济以及其他领域中做出更好的决策.统计告诉我们如何概括数据,分析数据,并得到有意义的推论,从而做出更好的决策.我们所做出的这些决策将帮助我们改进部门、公司乃至整个经济的运转.

Statistics 这个词源自意大利语 *Stato*,意为“国家、政府”,而 *Statista* 是指那些与国家、政府

① 引自 R. Gnanadesikan 在美国统计协会的演讲,再版于 *American Statistician* 44, no. 2 (1990 年 5 月), 122 页.

事务有关的人. 因此, *Statistics* 最初的含义是指收集的对 *Statista* 有用的事实. 这个含义的 *Statistics* 被 16 世纪的意大利所使用并传播到法国、荷兰与德国. 然而, 值得注意的是, 对人口和财产的调查实际上在远古时代就已经存在了.^① 现在, 统计已不再局限于有关国家、政府的信息, 而是扩展到几乎所有人力能及的领域. 我们也不再仅仅局限于收集被称作数据的数字信息. 我们还对数据进行概括, 通过有效的方法进行展示, 并进行分析. 统计分析往往要尝试从数据进行归纳推广. 统计是一门科学——信息的科学. 信息可能是定性的或定量的. 为说明这两类信息之间的差别, 我们来看一个例子.

[例 1-1] 波士顿地区的公寓销售人员向有意向的购买者们提供了表 1-1 中的信息. 表中哪些变量是定量的, 哪些是定性的呢?

表 1-1 波士顿销售数据

售价	房间数	朝向	洗衣机/干衣机?	供暖?
\$168,000	2	E	Y	Y
152,000	2	N	N	Y
187,000	3	N	Y	Y
142,500	1	W	N	N
166,800	2	W	Y	N

解: 售价是定量变量: 它传达了一个数量信息——销售的价格. 房间数也是一个定量变量. 房间的朝向是定性变量, 因为它传达的是定性的信息(东、西、北、南). 公寓中是否有洗衣机和干衣机以及公寓售价中是否包含了供暖费都是定性变量.

定量变量可以用一个数字来描述, 对这个数字进行平均之类的代数运算是有意义的. 定性(或分类)变量只简单的记录事物的性质. 如果一个数字是用来区分定性变量的不同类别的构成, 那么这个数字的取值是任意的.

统计学涉及有关测度的问题——某些是定量的, 其他一些是定性的. 测度是一个变量的实际取值. (定性变量也可以用数字来描述, 尽管这样的描述可能是随意的; 例如, $N=1, E=2, S=3, W=4, Y=1, N=0$.)

下面由弱到强分别给出四类常用的测量尺度.

定类尺度. 对于定类测量尺度来说, 数字只是用于区分类别的标志. 如果我们的数据集由蓝色、绿色和红色等项目组成, 那么我们可以指定蓝色为 1, 绿色为 2, 红色为 3. 在这个例子中, 数字 1, 2, 3 只代表数据点隶属的类别. 定类测量尺度适用于定性数据: 蓝色、绿色、红色; 男、女; 职业类别; 地区分类等等.

定序尺度. 在定序测量尺度中, 数据可能需要按照它们的相对尺寸或质量进行排序. 消费者对四种产品的排序可能分别是 1, 2, 3, 4, 其中 4 表示最好, 1 表示最差. 在这一测量尺度中, 我们只知道一种产品比另一种产品好, 但不知道好多少.

定距尺度. 在定距测量尺度中, 零的取值是任意确定的, 因此我们不能对两个测度计算比值.

① 参见 Anders Hald, *A History of Probability and Statistics and Their Applications before 1750* (New York: Wiley, 1990), 81 至 82 页.

但我们可以**计算间距的比值**。我们对一天内时间的测度就是定距尺度一个很好的例子。我们不能说上午十点是早上五点两倍长的时间,但我们可以说,早上零点(午夜)至上午十点之间相隔的10个小时是早上零点(午夜)至五点之间相隔的5个小时的两倍。这是因为早上零点并不表示时间的不存在。另一个例子是温度。当我们说0华氏度并不表示没有热量,100华氏度并不等于50华氏度的两倍热。但将一个物体从0华氏度加热到100华氏度所需要的热量的确是从0华氏度加热到50华氏度所需热量的两倍。

定比尺度。如果两个测度是定比尺度,那么我们就可以计算它们的比值。在该尺度中存在绝对零点,例如,钱可以用定比尺度进行测度,100美元是50美元的两倍,0美元则表示没有钱,也就是绝对零点。在上文中我们已经看到,时间间隔的测度(而不是一天中的时间测度)就是定比尺度。一般的,两个定距尺度测度之间的差值是定比尺度。其他一些定比尺度的例子包括质量、容量、面积、长度等的测度。

样本与总体

在统计中,我们要对两个概念加以区别:总体和样本。总体由调查者感兴趣的所有测度的集合组成。总体也称为**样本空间**。

样本则是从总体中抽取的部分测度所组成的子集。样本的抽取通常是随机的,以保证任意 n 个元素所组成的样本被抽中的概率是相等的。通过这种方式抽取的样本称为**简单随机样本**或简称为**随机样本**。随机样本可以计算元素抽取的概率。

例如,农场主Jane有1264头羊。这些羊构成了整个**总体**。如果有15只被挑选出来修剪羊毛,那么这15只羊就代表了从Jane的羊的总体中抽取的一个**样本**。进一步的,如果这15只羊是从1264只羊中**随机**挑选出来的,那么它们就构成了羊群的一个**随机样本**。

样本和总体的定义与我们要考虑的具体问题有关。如果Jane的羊群是我们关心的全部对象,那么它们就构成一个**总体**。但是,如果我们关心的是郡里所有的羊,那么Jane的1264只羊则是这个更大的总体中的一个**样本**(虽然这个样本将不是随机的)。

在统计学中,样本和总体的区别是十分重要的。

数据和数据收集

获得的关于某个变量的一组观测值称为**数据集**。例如,10个病人的心率观测值可以构成一个数据集。我们感兴趣的变量是心率,测量的尺度是定比尺度。(心脏每分钟跳动80次的速度是每分钟跳动40次的两倍。)我们观测到的病人的心率,即数据集可能为60,70,64,55,70,80,70,74,51,80。

数据可以通过不同方法进行收集。有时候数据集就构成了我们所感兴趣的整个**总体**。如果我们有五场足球赛的比赛得分,而我们感兴趣的也只有这五场足球赛,那么这5个观测构成的数据集就是整个**总体**。(在这个案例中,我们的数据是定比尺度的。为什么呢?假设这五场比赛的数据集只告诉我们是主队还是客队胜出了,那么在这个案例中的测量尺度又是什么呢?)

在其他情况下,数据可能构成了来自某个**总体**的一个**样本**。如果这些数据被用于对它们所来自的那个总体做出某些推断,那么我们就必须十分小心的收集这些数据。基于总体的样本信息对总体做出某个结论称为**统计推断**。统计推断是本书的一个重要内容。为了确保统计推断的准确度,数据必须从人们感兴趣的总体中随机抽取出来,必须保证总体的每一个组成部分都能在样本

中得到充分的、成比例的表现。

统计推断所依赖的数据可能是在调查或实验中收集得到的,这就需要对调查或实验小心构造。例如,我们希望从人们那里获得信息,就可能要使用邮寄问卷或电话访问等比较方便的工具。然而,在这样的调查中,我们希望尽可能降低无响应误差。当我们忽略了某些人可能对调查不会做出回应这一事实时,就会产生无响应误差。这一误差将会歪曲结论,因为那些没有做出回应的人可能会倾向于属于总体中的某一特定部分。在社会调查中,有些问题可能是比较敏感的——例如,“你曾经被拘禁过吗?”这个问题很容易导致无响应误差,因为曾经被拘禁过的人们往往不太可能会回答这个问题(除非他们完全确信能够保证匿名)。由通俗杂志展开的调查也经常受到无响应误差的困扰,尤其当他们的的问题具有煽动性的时候。那些使得杂志有很好的可阅性的事物往往会导致很差的统计结论。《纽约时报》上的一篇文章报道了有关居住在美国的犹太人生活的一项调查。这项调查是通过每周六给在家的犹太人打电话来实现的——因此其结果被严重误报了,因为传统的犹太人星期六是不会接电话的。^①

假设我们想测量一辆汽车的速度或汽油消耗定额,我们就需要从实验中收集数据。在这个案例中,我们需要确保所有可能的公路状况、天气状况和其他因素都被反映出来。药物测试也是从实验中收集数据的一个例子。通常将服用药物与服用安慰剂以及不服用任何东西进行对比研究。当设计一个实验用于测试某种安眠药的效果时,人们感兴趣的变量可能是服用了安眠药之后到进入睡眠所需要的时间间隔。

与调查一样,如果希望做出正确的推断,随机化对于实验也是非常重要的。如果我们要对所有总体做出推断,那么在实验中就要随机选择个体作为研究对象。将选择的个体分配到三个组别中时也要采用随机化:服用药物,不服用药物,服用安慰剂。这样的设计能使结果的潜在偏差降至最低。

在其他情况下,数据可能来自公开发表的资料,例如不同类型的或政府出版物的统计摘要。公布的一段时期内的失业率就是其中一个例子。在这种情况下,我们直接得到数据,而对其是如何获得的无法控制。我们再次需要谨慎行事。某一特定时期内的失业率并不是任何未来可能发生的失业率的随机样本,因此在这种情况下进行统计推断可能是复杂和困难的。然而,如果我们只对这些数据所对应的那段时期感兴趣,那么它们的确就构成所要描述的整个总体。不管怎样,在任何情况下,我们都必须十分谨慎的注意那些缺失数据或不完整的观察。

本章我们将重点介绍数据的处理、概括和展示——数据分析的第一步。下一章我们则要探讨概率论,研究随机样本和总体之间的关系。第三章将基于概率的概念构建一个系统,帮助我们依据样本对总体做出合理的、一致的推断。

为什么我们要关注推断和总体?为什么不仅仅着眼于数据,对它们做出解释?如果你只对你所拥有的那些特定的观察感兴趣,那么单纯的研究数据就足够了。然而,如果你希望在那些有限的的数据上得出暗含的有意义的结论,那么统计推断就是实现它的途径。

在市场调查中,我们通常对广告与销售额之间的关系感兴趣。从某个公司随机抽取的销售额和广告费用的数据集自身可能就有一定的意义,但是如果从数据集中的信息能够推断出该公司的广告费用水平与销售额之间的潜在关系,那么这些信息就更有用了。掌握广告费用与销售额之间的真实关系将有利于我们预测任何广告费用水平下所能带来的销售额,从而可以确定能够实

^① Laurie Goodstein, "Survey Finds Slight Rise in Jews Intermarrying," *The New York Times*, September 11, 2003, p. A13.

现利润最大化的广告费用水平.

食品药物管理部门可能要求某个制药商证实其将要生产的某种新药不会导致严重的副作用.那么,基于一个随机样本对药物所做的测试结果可能会用于对所有可能使用该药物的人们进行统计推断.

某家银行可能希望对自动柜员机的普及程度进行评估.因此,银行可能从它的客户中随机抽取一个群体进行测试.而研究的结果则可以通过统计推断推广到所有的银行客户.

某工厂的电脑磁盘驱动器的质量控制师需要确保生产出来的驱动器次品率不能超过 3%.因此,这个工程师可能定期的随机抽取驱动器的样本,并对它们的质量进行检查.基于这些随机样本,工程师就可以对所有驱动器中次品所占的比例进行推断.

以上仅仅是用于介绍商业领域中应用统计推断的若干例子.在本章后续内容中,我们将介绍用于做出基本统计分析的描述统计量.下一章则将介绍从样本到总体进行推断的基本原理.

习题

1-1 某电力公司所作的一项调查包含以下问题:

- 1. 户主的年龄
- 2. 户主的性别
- 3. 家庭成员人数
- 4. 是否使用电暖器
- 5. 日常使用几件大型电器
- 6. 冬天所使用的自动调温设备
- 7. 每日暖气开放的平均时间
- 8. 暖气开放的平均天数
- 9. 家庭收入
- 10. 平均每月的用电账单
- 11. 该电力公司与先前的两家供电商相比的排名

描述这 11 个问题中暗含的定量或定性变量,并描述它们的测量尺度.

1-2 讨论本节内容所描述的不同数据收集方法.

1-3 讨论并比较不同的测量尺度.

1-4 描述以下的定量或定性变量.

基金	类型	美国/国外	10 年年均回报率	费用率
Washington Mutual	高市值股票	美国	13%	0.67%
Euro-Pacific Growth	高市值股票	国外	8.6	0.9
Bond Fund America	中级股票	美国	6.9	0.67

引自 Adrienne Carter, "America's Hottest Fund Company," Money, May 2004, p. 61.

1-5 将五种冰激凌味道按照受欢迎程度进行排序,那么其测量尺度是什么?

1-6 定性变量与定量变量之间的区别是什么?

1-7 某个城镇有 15 户人家.如果你访问了某户特定人家的每个成员,那么你所访问的是这个城镇的总体还是样本?如果是样本,那么这个样本是随机的吗?如果你有镇上每个人的名单,即所谓