

商务统计学



Business Statistics

(诺琳 R. 夏普) Noreen R. Sharpe
【美】(理查德 D. 德沃) Richard D. De Veaux 著
(保罗 F. 维尔曼) Paul F. Velleman
徐强 王远林 译

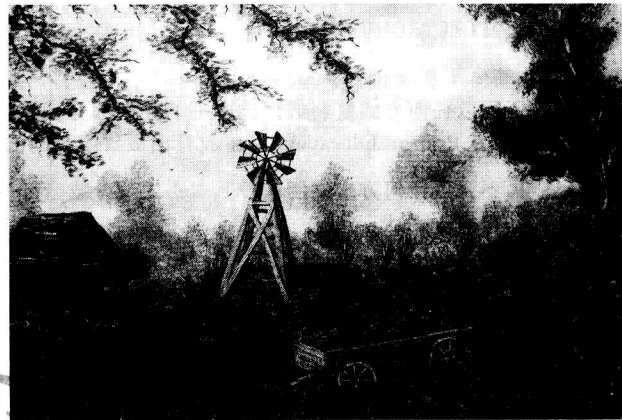


013064088

經管信
引 进 精品

F712.3
23

商务统计学



Business Statistics



(诺琳 R. 夏普) Norean R. Sharpe
 【美】(理查德 D. 德沃) Richard D. De Veaux 著
 (保罗 F. 维尔曼) Paul F. Velleman
 徐强 王远林 译

电子工业出版社
 Publishing House of Electronics Industry



F712.3
23

280380810

内 容 简 介

统计学是一门工具性学科,在众多的学科领域有着广泛的应用。本书将统计学的概念与方法应用于商务领域,从应用层面对统计学的基本方法进行了系统的讲解。全书包括四部分:探索和收集数据,理解数据和分布,探索变量间的关系,以及为决策建立模型。全书共 24 章,将方法的讲解与商务领域中的现实案例紧密结合起来,使读者掌握如何利用统计方法解决商务中的实际问题。本书还将统计软件与统计方法的应用结合起来,详细介绍各种统计方法在 Excel、Minitab、JMP、SPSS 和 Data Desk 等软件中的操作实现步骤。

本书既有坚实的理论基础,又有丰富的应用实例,是工商管理专业学生以及商务人士的必备书籍,对于其他相关专业的学生以及相关管理人员和研究人员也是一本很有裨益的参考书。

Authorized translation from the English language edition, entitled Business Statistics, 9780321426598 by Norean R. Sharpe, Richard D. De Veaux, Paul F. Velleman, published by Pearson Education, Inc., publishing as Addison-Wesley, Copyright © 2010 Pearson Education Inc.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

CHINESE SIMPLIFIED language edition published by PEARSON EDUCATION ASIA LTD., and PUBLISHING HOUSE OF ELECTRONICS INDUSTRY Copyright © 2013.

本书中文简体字翻译版由 Pearson Education(培生教育出版集团)授予电子工业出版社。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 Pearson Education(培生教育出版集团)激光防伪标签,无标签者不得销售。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

版权贸易合同登记号 图字:01-2009-3032

图书在版编目(CIP)数据

商务统计学/(美)夏普(Sharpe, N. R.), (美)德沃(De Veaux, R. D.), (美)维尔曼(Velleman, P. F.)著;徐强,王远林译. —北京:电子工业出版社,2013.7

(华信经管引进精品)

书名原文: Business Statistics

ISBN 978-7-121-20969-7

I. ①商… II. ①夏… ②德… ③维… ④徐… ⑤王… III. ①商业统计学-高等学校-教材 IV. ①F712.3

中国版本图书馆 CIP 数据核字(2013)第 158064 号

策划编辑:马 岚

责任编辑:王俊红

印 刷:三河市双峰印刷装订有限公司

装 订:三河市双峰印刷装订有限公司

出版发行:电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本:787×1092 1/16 印张:51.75 字数:1325 千字

印 次:2013 年 7 月第 1 次印刷

定 价:109.00 元

凡所购买电子工业出版社图书有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系,联系及邮购电话:(010)88254888。

质量投诉请发邮件至 zlt@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010)88258888。

译者序

近年来,我国国内出版了大量的统计学教材,其中既包括由国内学者撰写的教材,也包括引进的影印版和翻译版教材。在众多统计学教材中,由 Norean R. Sharpe、Richard D. De Veaux 和 Paul F. Velleman 撰写的《商务统计学》是一本特色突出的优秀教材。

具体而言,本教材有以下几个特色:(1) 大量引入现实中的案例。每一章的开头都有一个真实的案例,向学生们揭示了统计思维对现代商业决策为何如此重要;指导性案例详细向学生说明方法的应用过程;微型案例研究项目侧重激发学生开展研究型学习。(2) 结合统计软件介绍各种统计方法的计算过程。本书将统计软件与统计方法的应用结合起来,详细介绍了各种统计方法在 Excel、Minitab、JMP、SPSS 和 Data Desk 等软件中的实现步骤。(3) 明确说明统计方法应用中可能会出现错误和统计方法应用实践中的伦理问题。在实践中经常会出现统计方法和统计数据的种种误用,各章中“可能的错误”部分详细说明如何避免在应用中出现错误;“实践中的伦理”部分则说明统计方法应用中的伦理问题,并提出伦理解决方案。(4) 使用真实数据的习题和测验题。本书除第 1 章和第 24 章以外,各章均提供了大量使用真实数据的习题和测验题,奇数标号的习题的答案统一在书末提供,测验题的答案则在各章末尾提供。

本书的作者团队非常强大。Norean R. Sharpe 曾在鲍登学院(Bowdoin College)和巴布森学院(Babson College)讲授统计学,还曾在耶鲁大学从事研究工作,目前是乔治城大学(Georgetown University)教授。Richard D. De Veaux 是国际知名的教育家,曾在沃顿商学院和普林斯顿大学讲授过统计学,为美国运通、惠普、美国铝公司、杜邦、Pillsbury、通用电气、化学银行等“财富 500 强”企业做过咨询,目前是威廉姆斯学院(Williams College)统计学教授。Paul F. Velleman 在创新性统计教育方面具有国际声望,任教于康奈尔大学,是 Data Description 公司的创始人和 CEO,设计了 Data Desk 软件包。

本书由东北财经大学徐强博士和王远林博士翻译,其中徐强翻译第 1 章至第 20 章(包括对应的习题答案)、前言及附录,王远林翻译了第 21 章至第 24 章(包括对应的习题答案),并由徐强对全书内容进行了审定。为便于阅读,书中将国外报刊等直接用中文名表示。

我们指导的研究生也参加了初译和初校工作,其中,胡媛媛、李凯丽、方玉金、李启华、张洋、丁红、郭立革、郝胜兵、王静、沈慧和杜琨参加了初译工作,胡雪梅、李凯丽、李启华、方玉金、李文博、胡媛媛、陈丽洁、张朋飞和路鸣参加了初校工作,特此致谢。

本书涉及的内容非常广泛,由于译者水平有限,不妥或错误之处在所难免,请读者不吝指正。

前言

本书是为商科学生而写的,它将回答一个简单的问题:“怎样才能做出更好的决策?”作为企业家和顾问,应该知道,为了在当今这样的竞争环境下生存和发展,统计学是至关重要的。作为教育工作者,我们看到了向商科学生讲授统计学的方式与商业决策制定中统计学的使用方式之间的脱节。本书试图通过介绍统计方法来缩短理论与实践之间的距离。所以对学生来说,统计方法既重要又有趣。

根据数据做出商业决策的过程就像在讲一个故事,统计学在其中所扮演的角色是帮助听清楚这个故事。像其他教材一样,本书将讲授如何计算一个特定的统计量或检验,并且强调定义和公式。但是,与其他教材不同的是,本书也将讲解“为什么”,并坚持在商业决策的背景下给出结果。学生将会了解,为了做出更好的商业决策,应该如何进行统计思考、如何有效地表达分析结果并将决策告知他人。

在写作本书时,我们知道当今时代的统计学是用技术来实践的。这种见解的结果是:从对方程形式(与计算形式相比更喜欢直觉形式)的选择中得到的一切东西,都运用了对真实数据的广泛使用。但更重要的是,对技术价值的理解,使本书将重点集中于讲授统计思维而不是计算上。书中几百个例子关注的不是“怎么找出答案”,而是“如何思考答案及答案如何有助于制定出一个更好的决策”。

对统计思维的关注将书中的各章联系起来。初级商务统计学课程包含大量的新概念和方法,但是它们有一个核心部分:通过理解数据从而更加了解这个世界,做出更好的决策。从这个角度来看,学生能够知道,从数据得出推断的许多方式都是相同的核心概念的一些应用。

本书目标

如果学生不去阅读,世界上最好的教材也是没有任何价值的。下面是使本书更易学习的一些方法。

- 易读性。你会马上发现本书读起来并不像其他统计学教材。本书在寻求一种会话的、易接近的风格,并且介绍趣闻以保持学生的兴趣。在课堂测验中,老师惊讶地发现学生的自愿阅读量会超过老师布置的阅读量。学生给我们写信说,(使他们惊讶的是)他们真的很喜欢本书。

- 关注假设和条件。与其他教材不同,本书强调需要证明假设以使用统计程序,在案例和练习中也突出了这一点。本书将尽力为学生提供检查这些假设和条件的实践模板,而不是匆匆计算一个实际生活问题。

- 重视绘图和探索数据。对展示数据重要性的一贯重视是很明显的,从前面章节中的理解数据到后面章节中的复杂模型的建立都是如此。例子总是包括数据展示及通常用图形说明数据的价值,并且练习题中也强化了这一点。当将数据制图时,能够看到结构或模式。这些模式通常提出新的问题,引导统计分析和案例分析过程。当为最复杂的分析将数据制图以说明概念时,贯穿全书对绘图的重视能帮助学生看到寻找答案的简单结构。

- 一致性。书中将努力避免“做我们所说的,而不是做我们所做的”这样的陷阱。本书已经讲授了将数据制图和检验假设及条件的重要性,书中已将这种行为模型化(查看有关多元回归或时间序列章节中的习题,你会发现仍然需要前面章节中介绍的图和检验),这种一致性有助于强

化这些基本原则。

• 阅读的需要。只打算浏览本书的学生可能会发现,重要的概念、定义和抽样方案并不总是放在方框中的。本书需要学生去读,所以会尽量使阅读体验更愉快。

覆盖范围

商务统计学课程中包含的主题,通常与学生在学习和今后工作中的需要是一致的,但是这些主题的顺序和每个主题的相对重要性并不确定。在本书中,一些主题可能比你预期的更早或更晚出现。本书已经分类地写出了许多章节,所以它们能按不同的顺序讲授,但是我们极力推荐采用本书的顺序。

本书已经通过一个基本原则来安排主题的顺序,这个基本原则是:它应该是一个连贯的过程,其中的概念和方法集中在一起,让学生理解运用数据推理如何揭示新的重要真相。例如,介绍推理概念时,先介绍比例再介绍均值。学生在问卷调查和广告中见过比例,会对比例有更丰富的经验。先介绍比例,能够通过正态模型讲授推断,然后通过学生的 t 分布介绍均值的推断。

本书中很早就介绍了关联、相关和回归的概念。课堂经验表明,较早向学生介绍这些基本思想能够在课程开始时就激励他们。在学期的后期,当讨论推断时,学生回想他们所学过的知识会很自然也相对容易地通过这些方法探索数据,并以他们所见过的基本概念为基础进行推断。

在基础课程中,相对重点放在需要由老师做出计划的主题上。基础商务统计课程通常课时有限,所以很难有足够的时间去完全讲解像多元回归和模型构建这样重要的主题。本书将关于风险和概率的讨论放在主题序列的后面,以使學生能够更快进入实践中,有时间对这些重要的技能给予足够的重视。我们已经通过 GAISE(统计学教育中的评估和教学指导方针)报告^①引导选择需要强调的东西。GAISE 报告是在对学生怎样才能最好地学习统计学进行广泛研究的基础上得到的研究报告。现在已被美国统计学会官方采用和推荐的那些建议(及其他详细的建议),强烈希望统计学教育应该做到如下几点:

1. 强调统计知识和开发统计思维;
2. 使用真实数据;
3. 强调概念的理解而不仅仅是获取知识的过程;
4. 培养主动学习;
5. 在理解概念和分析数据时使用技术;
6. 使评价成为学习过程的一部分;

从这个意义上讲,本书是很现代的。

但是为了更有效率,一门课程必须很好地适应老师的偏好。使用本书有几个同样有效的途径,这取决于某个特定的教师想要强调什么。后面列出了一些其他的授课顺序,它们也能很好地处理本书中的内容。

灵活的教学纲要。本书选择了遵从 GAISE 指导方针和统计学教育中最具创新性的教育工作者的建议。这些专家一致认为,最好在课程中早一些且经常让学生面对真实的数据,以及自始至终强调对现实世界的解释。统计学教师面临着在有限的时间内(通常在一个学期中)要讲授重要内容的挑战,有两个地方背离了传统统计学的讲授顺序。

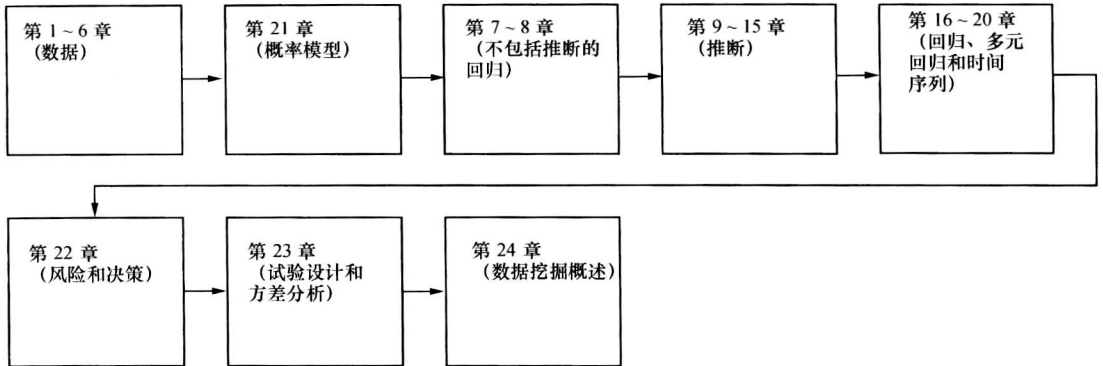
1. 本书在较前面的部分就介绍了回归(第 7 章和第 8 章)。
2. 对概率和决策制定的详细讨论(第 21 章和第 22 章)放在了教材的后面部分。这种顺序

^① www.amstat.org/education/gaise

可能并不适合每个人的需求,所以本书尽可能地将这些章节模块化——教师可以将它们挑出来,在课程的不同时点讲授。

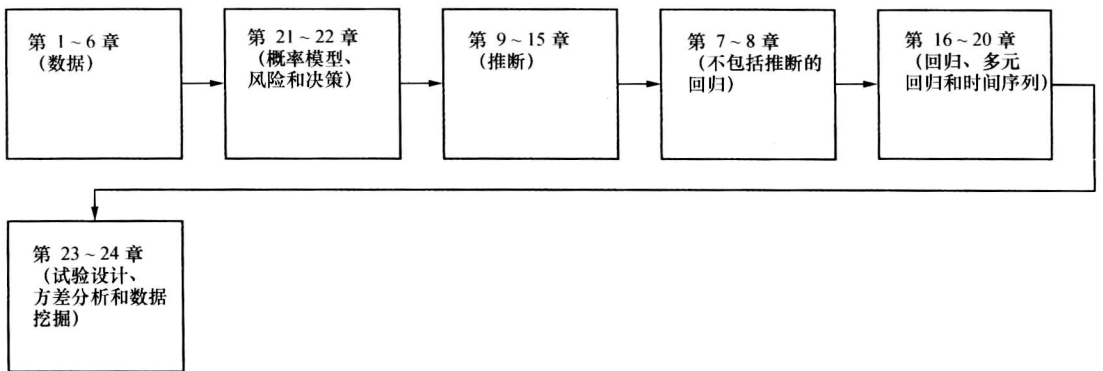
下面是主题顺序的一些选择。

关注数据和更早介绍回归(更加强调整概率)。当解释样本比例的抽样分布时,先讲授第9章,再讲授第21章,可以使教师参考二项式模型。记住,这将在理论上引出正态模型。使用正态分布表进行计算将在第9章中介绍。



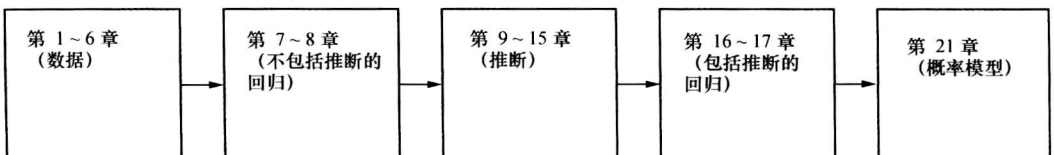
对于一个学期课程中的步骤,推荐第1~6章、第21章(概率模型)、第7~8章(不包括推断的回归)、第9~15章和第16~17章(包括推断的回归)。第21章也可以在第8章之后讲授,刚好位于有关推断的章节之前而不是直接在第6章之后。

关注概率,在后面介绍回归。为了关注概率和在课程后面介绍回归,第21章和第22章可以在第1~6章之后介绍。一种可能的顺序是:



对于一个学期课程中的步骤,推荐讲解第1~6章、第21章(概率模型)、第9~15章、第7~8章(不包括推断的回归)和第16~17章(包括推断的回归)。

关注数据和在不强调概率的情况下更早介绍回归,这是本书的标准讲授顺序。为了在一个学期课程中完成这一任务,推荐下面的顺序。



本书特色

一本教材并不仅仅是页面上的文字,而是许多特色集中起来形成的一幅大的图画。本书的特色为概念提供了一个真实世界的背景,帮助学生应用这些概念,促进问题的解决和整合技术——所有这些都帮助学生理解和领会商务统计学这幅大的图画。

引导案例

每章都以一个引导案例开头,通常来自作者的咨询经历。这些公司,如 Amazon.com、Zillow.com、Keen Inc. 和全食超市(Whole Foods Market),强调和说明了每章中的故事,向学生揭示了统计思维为什么对现代商业决策的制定如此重要。整章都会分析从这些公司得到的数据。

指导案例

明确表达统计结果的能力,对于帮助将统计学运用于商业决策的制定是非常重要的。为此,每章中的一些例子以指导案例的形式出现。整体分析遵从创新的“计划(Plan)—实施(Do)—报告(Report)”的模式。该模式以提出一个关于决策的明确问题开始分析,以回答该问题的报告结束。如果数据能够支持的话,为了强调每个问题的决策方面,将提供报告这一步作为商业备忘录,总结案例背景下的结果并提出建议。另外,尽可能地在备忘录中包括分析或模型的限制性。

微型案例研究项目

每章包括一个或两个微型案例研究项目,它们使用真实的数据并要求学生调查问题或做出决策。学生定义目标、设计过程、完成分析、报告结果。微型案例研究项目中的数据包含在本书资源包里^①,并提供各种软件所需要的格式。

可能的错误

每章都包含称为“可能的错误”的部分,它强调了最常见的统计误差和人们对统计学的误解,新的统计方法使用者所犯的最常见的错误,包括误用一种方法而不是错误地计算一个统计量。讨论的大多数错误都是作者在商务背景下而不是在课堂中经历的。目标之一是为学生提供发现统计误差的工具及提供揭示统计量误用的实践,不管是在国际上还是在国内。在这种意义上,一些练习题探究的是对这种误差的理解。

手工计算

即使建议使用技术去计算统计量,我们仍然意识到手工计算在教育层面的好处。该方框将一些更简单公式的计算拆分,通过计算工作案例来帮助学生。

可行性检查

我们经常提醒学生,统计学是关于如何理解世界和用数据进行决策的。不管多么仔细地思考所做的计算,无意义的结果很可能是错误的。错误很容易通过一些思考发现,所以要求学生在解释结果前停一下以进行可行性检查。

^① 登录华信教育资源网(www.hxedu.com.cn)可注册下载本书配套资源。

表示法

在全书中,强调明确表达的重要性。适当的符号是统计学专业词汇的一部分,但是它可能会令人生畏。学生知道数学符号 n 能够代表任意变量,但他们可能惊讶地发现在统计学中, n 总是并且仅代表样本容量。统计学家给许多字母和符号赋予特定的含义(b, e, n, p, q, r, s, t 和 z , 许多希腊字母也都有特定的含义)。当学生清楚地知道统计学家使用的字母和符号时,学习起来就会更有效率。

测验题

为了帮助学生检验他们对刚刚学习的内容的理解,书中给出了一些全章中涉及各个要点的问题。这些问题是一个快速的测验。大多数测验题只需要很少的计算。答案在每章习题的最后,所以学生能很容易检验自己以确认对关键内容的理解。这些问题还可用于激发课堂讨论。

数学方框

许多章节都提供统计方法和概念的数学基础。不同的学生以不同的方式学习,甚至同一个学生能够以多种途径理解同一内容。通过将证明、公式推导和理由与正文区分开,允许学生遵从正在学习主题的逻辑发展,同时也使数学基础能够向更深的层次发展。

小结

每章结尾的小结强调新的概念,定义该章中介绍的新概念,同时列出学生应该掌握的技能。学生可以将这些看成学习指导。如果他们理解小结中的概念,掌握了技能,就可能已经准备好去考试了。

实践中的伦理

学生经常惊讶地发现,统计学并不仅仅是将数据套进公式。大多数统计分析都需要大量的判断,这些判断最好的指导是以真诚和合乎道德的方式发现真相。做不到这一点将会导致糟糕的和危险的决策。“实践中的伦理”穿插在每章中,说明在统计分析中需要的一些判断,识别可能的误差,将这些问题与美国统计学会的道德指南联系起来,然后提出道德上和统计上更好的备选方法。

习题

书中尽量确保习题包含相关的、现代的和真实世界的问题。许多来自新闻故事,一些来自近期的研究文章。数据在本书教辅资源包中(以各种不同的格式),所以学生能够更深地挖掘它们。习题上标有①的,表示提供了数据。有时,因为数据集大小的关系,数据只能是电子版的。全书都将习题进行了配对,使每个奇数习题(书后面有答案)和后面紧接着的偶数习题是同一个统计主题。每章中的习题粗略地根据主题和难易程度进行了排序。

数据来源

在例子和习题中使用的大部分数据都来自真实世界。无论什么时候只要能够做到,都包括了使用的因特网数据来源,通常以网址的形式给出。作为网络用户,学生很了解,随着网站的变化,链接可能无效。为了将这种变化的影响最小化,书中尽量给出的是地址树的更高层。另外,

随着更多新数值的获得,数据本身通常也是变化的。所用数据可从 www.aw.com/sharpe 上查到。

技术帮助

在商务活动中,统计学是用计算机实践的,而不是使用单一的软件平台。在每章的结尾,不是强调一个特定的统计程序,而是总结学生在最常用的统计软件包中能够发现的东西,通常是有注释的输出结果。接着,会为这些最常用的统计软件包(Excel 2007 和 2003、Minitab、SPSS、JMP 和 Data Desk)提供明确的指导,以帮助学生使用他们所选的软件。

目录

第 1 篇 数据探索和数据收集		
第 1 章 统计学与变异	2	
1.1 什么是统计学	2	
1.2 本书有何用处	3	
第 2 章 数据	6	
2.1 什么是数据	7	
2.2 变量类型	9	
2.3 Where、How 和 When	12	
2.4 小结	15	
2.5 技术帮助	16	
2.6 微型案例研究项目	17	
2.7 习题	17	
第 3 章 调查与抽样	22	
3.1 抽样调查三原则	22	
3.2 是否需要普查	25	
3.3 总体和参数	26	
3.4 简单随机样本(SRS)	27	
3.5 其他抽样设计	28	
3.6 定义总体	31	
3.7 有效调查	31	
3.8 小结	36	
3.9 技术帮助:随机抽样	38	
3.10 微型案例研究项目	38	
3.11 习题	39	
第 4 章 显示和描述分类数据	45	
4.1 数据分析的 3 条规则	45	
4.2 频数表	46	
4.3 图	47	
4.4 列联表	50	
4.5 小结	60	
4.6 技术帮助:在计算机上显示 分类数据	61	
4.7 微型案例研究项目	63	
4.8 习题	63	
第 5 章 随机性和概率	76	
5.1 随机现象和概率	76	
5.2 不存在的平均数定律	78	
5.3 不同类型的概率	79	
5.4 概率法则	81	
5.5 联合概率和列联表	85	
5.6 条件概率	86	
5.7 构建列联表	88	
5.8 小结	89	
5.9 微型案例研究项目	91	
5.10 习题	92	
第 6 章 显示和描述定量数据	102	
6.1 显示分布	102	
6.2 形状	104	
6.3 中心	106	
6.4 分布的离散度	108	
6.5 形状、中心和离散度的概括	109	
6.6 五数概括和箱线图	110	
6.7 组间比较	113	
6.8 确认异常值	114	
6.9 标准化	116	
6.10 时间序列图	118	
6.11 变换有偏的数据	120	
6.12 小结	124	
6.13 技术帮助:显示和概括定量 变量	126	
6.14 微型案例研究项目	129	
6.15 习题	130	
第 2 篇 理解数据和分布		
第 7 章 散点图、关联和相关	148	
7.1 观察散点图	149	
7.2 在散点图中指定变量的 角色	151	
7.3 理解相关关系	152	
7.4 直线型散点图	159	

7.5	潜在变量和因果关系	160	*10.5	小样本的置信区间	251
7.6	小结	163	10.6	选择样本容量	251
7.7	技术帮助:散点图和相关系数	164	10.7	小结	255
7.8	微型案例研究项目	167	10.8	技术帮助:比例的置信区间	256
7.9	习题	168	10.9	微型案例研究项目	257
第8章	线性回归	182	10.10	习题	258
8.1	线性模型	183	第11章	比例的假设检验	266
8.2	相关性和直线	184	11.1	假设	267
8.3	向均值的回归	187	11.2	作为假设检验的审判	268
8.4	模型检验	189	11.3	P 值	269
8.5	从残差中学习更多知识	189	11.4	假设检验的原理	270
8.6	模型的变异和 R^2	191	11.5	备择假设	272
8.7	真实性检验:回归是否合理	193	11.6	α 水平与显著性	274
8.8	小结	196	11.7	临界值	276
8.9	技术帮助:回归	197	11.8	置信区间与假设检验	277
8.10	微型案例研究项目	199	11.9	两类错误	280
8.11	习题	200	*11.10	检验的效力	281
第9章	抽样分布和正态模型	214	11.11	小结	285
9.1	样本比例分布的建模	215	11.12	技术帮助	286
9.2	模拟	215	11.13	微型案例研究项目	288
9.3	正态分布	216	11.14	习题	289
9.4	练习正态分布的计算	217	第12章	均值的置信区间和假设检验	297
9.5	比例的抽样分布	222	12.1	均值的抽样分布	297
9.6	假设和条件	224	12.2	均值的置信区间	299
9.7	中心极限定理 ——统计学中的基本定理	226	12.3	假设与条件	300
9.8	均值的抽样分布	228	12.4	解释置信区间时需要注意的地方	305
9.9	样本容量的收益递减性	230	12.5	单样本 t 检验	306
9.10	抽样分布模型的原理	230	12.6	样本容量	309
9.11	小结	233	*12.7	自由度为什么是 $n-1$	310
9.12	微型案例研究项目	234	12.8	小结	312
9.13	习题	235	12.9	技术帮助:均值推断	313
第10章	比例的置信区间	243	12.10	微型案例研究项目	315
10.1	置信区间	243	12.11	习题	316
10.2	误差幅度:确定性与精确性	246	第13章	比较两个均值	328
10.3	临界值	247	13.1	检验两个均值的差异	329
10.4	假设与条件	248	13.2	两样本 t 检验	331
			13.3	假设和条件	331

13.4	两均值差的置信区间	335	16.11	习题	447
13.5	合并的 t 检验	337	第 17 章	对残差的理解	463
*13.6	图基快速检验	342	17.1	检验各组的残差	463
13.7	小结	343	17.2	外推和预测	466
13.8	技术帮助:两样本方法	345	17.3	不寻常和异常的观测值	468
13.9	微型案例研究项目	347	17.4	处理汇总值	471
13.10	习题	347	17.5	自相关	472
第 14 章	配对样本与区组划分	361	17.6	线性性	474
14.1	配对数据	362	17.7	转换(变换)数据	475
14.2	假设和条件	362	17.8	幂变换阶梯	477
14.3	配对 t 检验	363	17.9	小结	480
14.4	配对 t 检验的原理	366	17.10	技术帮助	481
14.5	小结	369	17.11	微型案例研究项目	482
14.6	技术帮助:配对 t 方法	370	17.12	习题	483
14.7	微型案例研究项目	372	第 18 章	多元回归	499
14.8	习题	372	18.1	多元回归模型	500
第 15 章	计数的推断:卡方检验	390	18.2	解释多元回归的系数	502
15.1	拟合优度检验	391	18.3	多元回归模型的假设和 条件	503
15.2	解释卡方值	395	18.4	检验多元回归模型	510
15.3	检测残差	396	18.5	调整后的 R^2 和 F 统计量	511
15.4	齐性的卡方检验	397	*18.6	Logistic 回归模型	513
15.5	比较两个比例	401	18.7	小结	520
15.6	独立性的卡方检验	403	18.8	技术帮助:回归分析	521
15.7	小结	408	18.9	微型案例研究项目	523
15.8	技术帮助:卡方	409	18.10	习题	523
15.9	微型案例研究项目	411	第 19 章	建立多元回归模型	538
15.10	习题	411	19.1	指示(或虚拟)变量	539
第 3 篇	研究变量之间的关系		19.2	不同斜率的调整——交互 效应项	541
第 16 章	回归的推断	426	19.3	多元回归诊断	544
16.1	总体和样本	427	19.4	建立回归模型	548
16.2	假设和条件	427	19.5	共线性	555
16.3	斜率的标准误	430	19.6	二次项	558
16.4	对回归斜率的检验	432	19.7	小结	561
16.5	相关性的假设检验	435	19.8	技术帮助:计算机回归 分析	563
16.6	预测值的标准误	436	19.9	微型案例研究项目	564
16.7	使用置信区间和预测区间	440	19.10	习题	565
16.8	小结	443			
16.9	技术帮助:回归分析	444			
16.10	微型案例研究项目	446			

第 20 章 时间序列分析	582	22.8 灵敏度	674
20.1 什么是时间序列	583	22.9 模拟	674
20.2 时间序列成分	583	22.10 概率树	676
20.3 平滑方法	586	22.11 条件的反转:贝叶斯准则	678
20.4 简单移动平均法	587	22.12 更加复杂的决策	679
20.5 加权移动平均	589	22.13 小结	680
20.6 指数平滑法	589	22.14 微型案例研究项目	682
20.7 概括预测误差	590	22.15 习题	683
20.8 自回归模型	592	第 23 章 实验的设计、分析及观察性	
20.9 随机游走	595	研究	689
20.10 基于多元回归的模型	599	23.1 观察性研究	690
20.11 加法和乘法模型	601	23.2 随机化的比较实验	690
20.12 循环和不规则成分	603	23.3 实验设计的四大原则	691
20.13 基于回归模型的预测	604	23.4 实验设计	693
20.14 时间序列预测方法的选择	608	23.5 设盲和安慰剂	696
20.15 解释时间序列模型:再次考察全食超市的数据	608	23.6 混杂变量和潜在变量	698
20.16 小结	610	23.7 单因素设计分析——单因素方差分析法	698
20.17 技术帮助	611	23.8 方差分析的假设和条件	701
20.18 微型案例研究项目	612	23.9 多重比较	704
20.19 习题	613	23.10 观测数据的方差分析	706
第 4 篇 决策建模		23.11 多因子设计分析	707
第 21 章 随机变量和概率模型	632	23.12 小结	716
21.1 随机变量的期望值	632	23.13 微型案例研究项目	722
21.2 随机变量的标准差	634	23.14 习题	722
21.3 期望值和方差的性质	636	第 24 章 数据挖掘概述	737
21.4 离散概率模型	639	24.1 直接营销	737
21.5 连续型随机变量	646	24.2 数据	738
21.6 小结	656	24.3 数据挖掘的目标	739
21.7 微型案例研究项目	658	24.4 数据挖掘的误区	740
21.8 习题	659	24.5 成功的数据挖掘	741
第 22 章 决策与风险	667	24.6 数据挖掘问题	743
22.1 行动、自然状态和结果	667	24.7 数据挖掘的算法	743
22.2 收益表和决策树	668	24.8 数据挖掘过程	747
22.3 最小化损失和最大化收益	669	24.9 总结	748
22.4 行动的期望值	669	24.10 小结	749
22.5 具有完全信息的期望值	671	习题答案	751
22.6 样本信息下的决策	671	附录 A 部分公式和表	810
22.7 估计变异性	673		

第 1 篇

数据探索和数据收集

- | | |
|-------|-----------|
| 第 1 章 | 统计学与变异 |
| 第 2 章 | 数据 |
| 第 3 章 | 调查与抽样 |
| 第 4 章 | 显示和描述分类数据 |
| 第 5 章 | 随机性和概率 |
| 第 6 章 | 显示和描述定量数据 |

第1章

统计学与变异

只须看一下 *Financial Times* (《金融时报》) 网站的一个网页, 就会发现该网页充满了“统计数字”, 很明显, 其作者认为所有这些信息都很重要, 但这就是统计学的全部吗? 嗯, 也许是, 也许不是。这个网页可能包含了许多事实, 但是正如我们将要看到的, 这一主题比电子表格和图表更加有趣, 内容也更加丰富。

“我为什么要学习统计学?” 你也许会问。“我毕竟不打算从事这种工作。事实上, 我将雇人做这些工作。”这很好, 但可能你基于数据所做的决策太重要了, 所以不能委托他人去做。你希望能够解释周围的数据并得到自己的结论。你会发现, 学习统计学比你所设想的更为重要, 也更充满乐趣。

1.1 什么是统计学

从在杂货店的每次购买, 到上网冲浪时每次点击鼠标, 似乎在每次转身时, 总有人收集关于我们的数据。美国联合包裹速递服务公司 (The United Parcel Service, UPS) 在世界范围内对每件包裹从一个地方运到另一个地方的情况进行跟踪记录, 并将这些记录保存在一个巨型数据库中。如果你发送或收到一个 UPS 包裹, 你就能访问到该数据库的一部分。该数据库约 1 700 万兆字节, 相当于一个包含美国国会图书馆所有书籍的数据库的大小 (但是, 我们怀疑, 它可能不是很有趣)。可能每个人都想知道, 如何利用这些数据呢?

对于弄清我们所处的复杂的世界, 统计学发挥了重要作用。统计学家评估美国食品和药物管理局 (Food and Drug Administration, FDA) 正在审核的转基因食品或新药的风险。统计学家预测一个国家各个地区艾滋病新病例的数量, 或预测可能会对超市的减价促销做出反应的顾客的人数。

Q: 什么是统计学?

A: 统计学是推理的一种方式, 是工具和方法的集合, 用来帮助我们理解世界。

Q: 什么是统计量?

A: 统计量是从数据中计算出来的量。

Q: 那么什么是数据? (So what is data?)

A: 你指的是“什么是数据?” (what are data?) data 是复数形式。单数形式是 datum。

Q: 那么什么是数据? (So What are data?)

A: 数据是具有一定背景的数值。

统计学家帮助科学家、社会学家和商界领袖了解失业与外界环境控制是如何相关的, 判断学童早期接受丰富的教育是否对其以后的表现产生影响, 以及判断维生素 C 是否真的可以预防疾病。无论何时, 只要你拥有数据并想了解世界, 你就需要统计学。

如果要分析学生对商业道德的看法 (这一问题我们在后续的章节中还会涉及), 是否应该对美国的每个大学生进行调查, 或就此问题对全世界的每个大学生进行调查? 这样做并不具有可行性或成本效益。那么应该怎么做呢? 要放弃这项调查吗? 也许应该尝试从一个更小、更具有代表性的学生组中取得调查结果。从手头的数据到广泛地了解世界, 统计学可以帮助我们实现这一飞跃。第 3 章会讨论抽样问题的细节, 并且“从特殊到一般”的归纳方式将会贯穿本书的始终。我们希望本书能够使

你从数据中得到结论,并对诸如下面的问题做出有效的商业决策:

- 世界上不同地方的大学生对商业道德的看法有差别吗?
- 广告对销售额有何影响?
- 积极型的“高增长”投资基金真的比保守型投资基金具有更高的收益吗?
- 你所在企业的收入和利润是否具有季节性?
- 货架的位置与麦片粥的销售额之间是什么关系?
- 你所在企业季度预测的可靠性如何?
- 你的顾客是否具有共同的特征?他们为什么选择你的产品?以及更重要的,在那些不是你的顾客的人身上是否也存在这些特征?

回答诸如此类问题并从数据中得到结论在很大程度上依赖于理解变异的能力。“变异”并不是能在一句话结尾找到的术语,但它是统计学的精髓所在。从数据中学习的关键是理解周围始终存在的变异。

数据是变化的。人之间也是有差异的。同样道理,不同月份的经济形势也是有差异的。我们无法理解每件事,更不用说完全测量它了。甚至我们进行测量时,测量也是不完美的。因此,观察之后并据之做出决策的数据充其量只是提供了世界的不完美图景。变异在统计学中处于核心位置。如何理解变异是统计学的一个主要挑战。

1.2 本书有何用处

这是一个合理的问题。本书很可能并不会变成你所期望的那样。本书并不强调计算和公式,而是强调图形和理解。不是学习如何将数字代入公式中,而是学习建立模型的过程,并且理解所分析的数据和所使用方法的局限性。每章都使用真实的数据和真实的商务场景,这样你会明白如何使用数据做出决策。

1.2.1 图形

请闭上眼睛并随机翻开本书,你打开的这一页是不是有一个统计图或统计表?请再做一次,或不妨做上十次。你很可能会看到数据有许多不同的显示方式,甚至在本书习题中也是如此。统计图和统计表有助于理解数据要表达的内容。因此,每个故事和数据集,以及每个新的统计技术都伴有统计图,这有助于理解方法和数据。

1.2.2 过程

为了帮助你使用统计学进行商务决策,我们将通过对问题思考的全过程,发现、展示你的结论并告诉别人你的发现这样的形式来引导决策。从事商务统计的3个步骤就是:计划(Plan)、实施(Do)和报告(Report)。

计划:计划居于第一位。你需要知道你的目标朝向,以及其中的原因。清晰地界定和理解目标将会节省许多工作量。

实施:大多数学生认为统计学就是有关计算的学科。计算统计数字并将之用统计图表示出来的技巧是重要的,但是,计算通常是整个过程中最不重要的部分。事实上,我们通常将计算过程交给技术性设备,而我们则着重理解结果所暗含的意义。

报告:报告所学会的东西。除非别人能理解你对结果所做的解释,否则任务并没有完成。