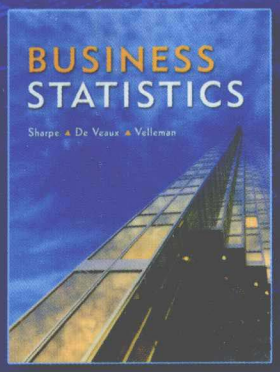
英文版

PEARSON

商务统计学

Business Statistics



Norean Radke Sharpe
[美] Richard D. De Veaux 著
Paul F. Velleman

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

商务统计学

（英文版）

Business Statistics

Norean Radke Sharpe

［美］ Richard D. De Veaux 著

Paul F. Velleman

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

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四部分内容，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，让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DataDesk等软件中的操作实现步骤。

本书可作为大学本科生和研究生的教材，也可供从事工商管理和经济分析的人士参考。

版权贸易合同登记号　图字：01-2010-1368

图书在版编目（CIP）数据

商务统计学=Business Statistics：英文/（美）夏普（Sharpe, N. R.），（美）德维克斯（Veaux, R. D. D.），（美）维尔曼（Velleman, P. F.）著. -北京：电子工业出版社，2010.6

ISBN 978-7-121-10632-3

Ⅰ.①商…　Ⅱ.①夏…　②德…　③维…　Ⅲ.①商业统计学-英文　Ⅳ.①F712.3

中国版本图书馆CIP数据核字（2010）第056102号

策划编辑：马　岚

责任编辑：许菊芳

印　　刷：北京市天竺颖华印刷厂

装　　订：三河市鑫金马印装有限公司

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：54.5　　字数：1395千字

印　　次：2010年6月第1次印刷

定　　价：99.00元（含光盘1张）

凡所购买电子工业出版社的图书有缺损问题，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zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

服务热线：（010）88258888。

序

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，应该知道为了在今天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。本书将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。

根据数据做出一个商业决策有一个故事要讲，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。学生们将会了解到，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（比计算形式更喜欢直觉形式）的选择中得到的一切东西，都运用到了对真实数据的广泛使用中。但是更重要的是，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"如何思考答案以及它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术语、概念和方法，但是它们有一个核心部分：通过理解数据告诉如何更加了解这个世界，怎样做出更好的决策。从这个角度来看，学生们能够知道从数据中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

如果学生们不去阅读，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。**你会马上发现这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们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）他们真的很喜欢这本书。
- **关注假设和条件。**同其他教材不同，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，从前面的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以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，贯穿全书中对绘图的重视能帮助学生们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查假设及条件的重要性，书中已将这种行为模型化（检查有关多元回归或时间序列章节中的练习题，你会发现仍然需要和展示在前面章节中介绍的绘图和检查），这种一致性有助于强化这些基本的原理。
- **阅读的需要。**只打算浏览这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。这本书需要阅读，所以会尽量使这种阅读体验更愉快。

覆盖范围

商务统计学课程中包含的主题，通常与学生在他们学习和今后的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，但是极力推荐采用我们所选择的顺序。

我们已经通过一个基本原理来引导主题的顺序，这个基本原理是：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能揭示新的、重要的真相。例如，介绍推理概念时，先介绍比例再介绍均值。学生们在问卷调查和广告中见过比例，会对比例有更丰富的经验。通过先介绍比例，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。课堂中的经验表明，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，学生们回想他们所学过的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，相对重点是放在需要由老师做出计划的主题上。基础商务统计学课程通常课时有限，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。本书将对风险和概率的讨论放在主题序列的后面，这样学生们能够更快进入到实践应用中，有时间对这些重要的技能给以足够的重视。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（以及其他详细的建议），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，这本书是很现代的。

但是为了更有效率，一门课程必须很好地适合老师的偏好。使用这些材料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，它们也能够很好地处理这些材料。

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

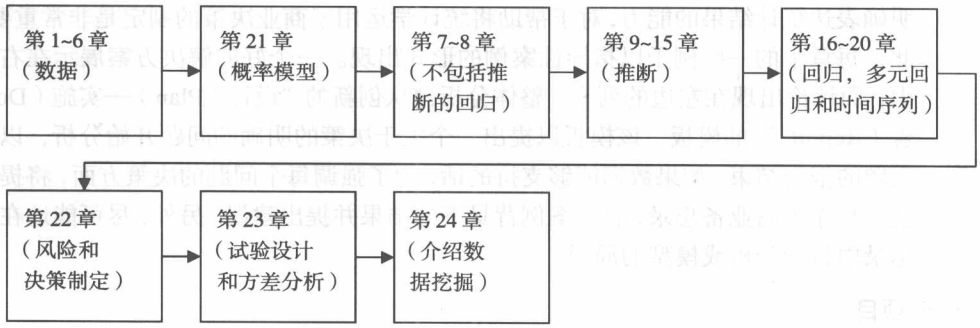
① www.amstat.org/education/gaise。

对现实世界的解释。我们深知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. 第 7 章和第 8 章较早地在导言部分介绍了回归（初步地，而不是基于推断）。
2. 将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。认识到这种顺序可能并不适合每个人的需求，所以尽可能地将这些章节模块化——可以将它们挑出来，在课程中的不同时点讲授。

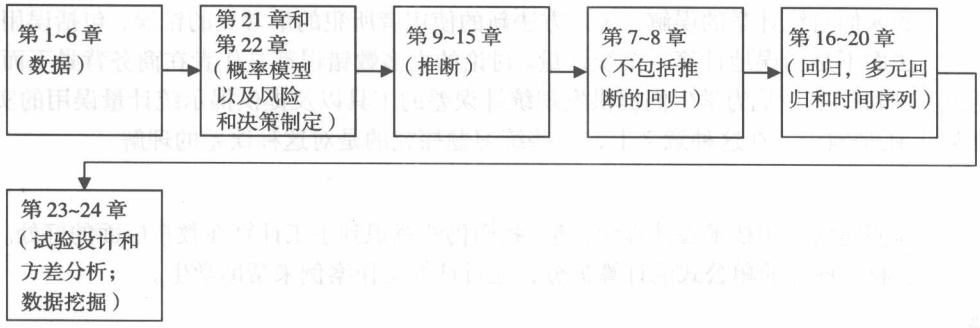
下面是主题顺序的一些选择。

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。



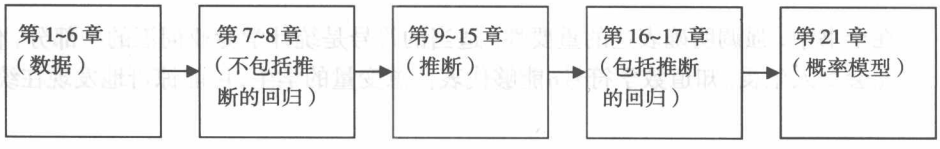
对于在一个学期课程中的这个步骤，推荐第 1~6 章、第 21 章（概率模型）、第 7~8 章（不包括推断的回归）、第 9~15 章和第 16~17 章（包括推断的回归）。第 21 章也可以在第 8 章之后讲授，刚好位于有关推断的章节之前而不是直接在第 6 章之后。

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：



对于在一个学期课程中的这个步骤，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副大的图画。本书的特色为概念提供了一个真实世界的背景，帮助学生们应用这些概念，促进问题的解决和整合技术——所有这些都帮助学生们理解和领会商务统计学这幅大的图画。

激发案例

每一章都以一个激发案例开头，通常来自于作者的咨询经历。这些公司——例如Amazon.com、Zillow.com、Keen Inc.和Whole Foods Market——强调和说明了每一章中的故事，向学生们揭示了统计思维如何和为什么对现代商业决策的制定如此重要。全章都会分析从这些公司取得的数据。

逐步的指导性案例

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性案例的形式出现。一个好的解决方案展示在右侧的列中，而评论出现在左边的列中。整体分析遵从创新的“设计（Plan）—实施（Do）—报告（Report）”的模板。该模板以提出一个关于决策的明确的问题开始分析，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，尽可能地在总结备忘录中包括分析或模型的局限。

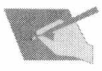
微型案例研究项目

每一章中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们去调查一个问题或做出一个决策。学生们定义目标，设计过程，完成分析，报告结果。微型案例研究项目中的数据可以从本书教辅资源包和网站上获得，并提供各种软件所需要的格式。

什么可能出错

每一章中都包含一个称为“什么可能出错？”的创新部分，它强调了最常见的统计误差和人们对统计学的误解。统计方法新的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。目标之一是为学生们提供发现统计误差的工具以及提供揭示统计量误用的实践，不管是在国际上还是国内。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



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。手工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。

现实检查

我们经常提醒学生们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现实检查。

注意符号

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。知道数学符号n能够代表任意变量的学生，可能惊讶地发现在统计学中，

n 总是并且仅代表样本容量。统计学家们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以及许多希腊字母都有特定的含义）。当学生们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阅读的材料的理解，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。答案在每一章习题集的最后，所以学生能很容易检验自己以确认他们对关键思想的理解。这些问题还可用于激发课堂讨论。

数学方框

MATH BOX

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材料。通过将证明、公式推导和理由与正文区分开，允许学生们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。

学到了什么

章节结尾的总结强调新的概念，定义该章中介绍的新术语，同时列出学生们应该掌握的技能。学生们可以将这些看成是学习指导。如果他们理解总结中的概念，知道术语和掌握技能，他们可能已经准备好去考试了。

行为中的道德规范

ETHICS IN ACTION

学生们经常惊讶地发现，统计学并不仅仅是将数据塞进公式里。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努力去发现真相。做不到这一点将会导致糟糕的和危险的决策。行为中的道德规范穿插在每一章中，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。

练习题

书中尽量确保练习题包含相关的、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。在可能的情况下，数据都保存在本书教辅资源包和网站上（总是以各种不同的格式），所以学生们能够更深地挖掘它们。练习题上标有 T 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。全书都将练习题进行了配对，使每个奇数练习题（书后面有答案）和后面紧接着的偶数练习题是同一个统计主题。每章中的练习题粗略地根据主题和难易程度进行了排序。

数据来源

在例子和练习题中使用的大部分数据都来自真实世界。在这个版本中列出了许多来源。无论什么时候只要能够做到，都包括了使用的互联网数据来源的参考，通常以网址的形式给出。作为互联网的使用者（也就是学生）很了解，随着网站的变化网址可能成为无效的链接。为了将这种变化的影响最小化，书中指出在地址树上越高处越可行。另外，随着更多更近数值的获得，数据本身通常也是变化的。使用的数据通常是放在本书 CD-ROM 中，也放在合作网站 www.aw.com/sharpe 上。

DVD 上带可选字幕的视频

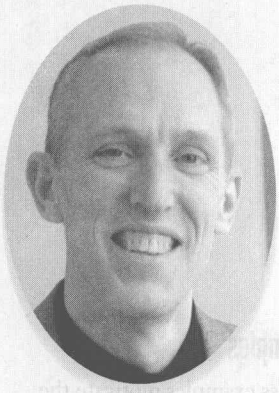
体现本书作者特点的这些视频，能够帮助学生回顾每章中的主要内容，而概念性视频（商务洞察能力视频）关注的是统计概念，因为它们适合于真实世界。与教材一样，这些视频也是学生友好型的，也强调批判性思维。DVD 格式能够更加简单和方便地在家里或学校的电脑里观看这些视频。视频中包含字幕。也可以从在线 MyStatLab 课程中观看这些视频。

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，而是总结学生们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，以帮助学生们开始使用他们所选的软件。

Meet the Authors

As a researcher of statistical problems in business and a professor at a business school, **Norean Radke Sharpe** (Ph.D. University of Virginia) understands the challenges and specific needs of the business student. She is currently Professor of Statistics and Operations Research at Babson College, where she is also Chair of the Division of Mathematics and Science. Prior to joining Babson, she taught statistics and applied mathematics courses for several years at Bowdoin College and conducted research at Yale University. Norean is coauthor of the recent text, *A Casebook for Business Statistics: Laboratories for Decision Making*, and has authored over 30 articles—primarily in the areas of statistics education and women in science. Norean currently serves as Associate Editor for CAUSE (Consortium for the Advancement of Undergraduate Statistics Education) and Associate Editor for the journal *Cases in Business, Industry, and Government Statistics*. Her research focuses on business forecasting and statistics education. She is also co-founder of DOME, Inc., a nonprofit foundation that works to increase Diversity and Outreach in Mathematics and Engineering for the greater Boston area. She has been active in increasing the participation of women and underrepresented students in science disciplines for several years and has two children of her own.



Richard D. De Veaux (Ph.D. Stanford University) is an internationally known educator, consultant, and lecturer. Dick has taught statistics at a business school (Wharton), an engineering school (Princeton), and a liberal arts college (Williams). While at Princeton, he won a Lifetime Award for Dedication and Excellence in Teaching. Since 1994, he has been a Professor of Statistics at Williams College. Dick holds degrees from Princeton University in Civil Engineering and Mathematics and from Stanford University in Dance Education and Statistics, where he studied with Persi Diaconis. His research focuses on the analysis of large data sets and data mining in science and industry. Dick has won both the Wilcoxon and Shewell awards from the American Society for Quality and is a Fellow of the American Statistical Association. Dick is well known in industry, having consulted for such Fortune 500 companies as American Express, Hewlett-Packard, Alcoa, DuPont, Pillsbury, General Electric, and Chemical Bank. He was named the "Statistician of the Year" for 2008 by the Boston Chapter of the American Statistical Association for his contributions to teaching, research, and consulting. In his spare time he is an avid cyclist and swimmer. He also is the founder and bass for the doo-wop group "Diminished Faculty" and is a frequent soloist with various local choirs and orchestras. Dick is the father of four children.

Paul F. Velleman (Ph.D. Princeton University) has an international reputation for innovative statistics education. He designed the Data Desk® software package and is also the author and designer of the award-winning ActivStats® multimedia software, for which he received the EDUCOM Medal for innovative uses of computers in teaching statistics and the ICTCM Award for Innovation in Using Technology in College Mathematics. He is the founder and CEO of Data Description, Inc. (www.datadesk.com), which supports both of these programs. He also developed the Internet site, *Data and Story Library* (DASL) (www.dasl.datadesk.com), which provides data sets for teaching Statistics. Paul coauthored (with David Hoaglin) the book *ABCs of Exploratory Data Analysis*. Paul has taught Statistics at Cornell University on the faculty of the School of Industrial and Labor Relations since 1975. His research often focuses on statistical graphics and data analysis methods. Paul is a Fellow of the American Statistical Association and of the American Association for the Advancement of Science and baritone of the barbershop quartet *Alchemy*. Paul's experience as a professor, entrepreneur, and business leader brings a unique perspective to the book.

Dick De Veaux and Paul Velleman have authored successful books in the introductory college and AP High School market with Dave Bock, including *Intro Stats*, Third Edition (Pearson, 2009), *Stats: Modeling the World*, Third Edition (Pearson, 2010), and *Stats: Data and Models*, Second Edition (Pearson, 2008).

From the Classroom...

Providing Real Business Context

Chapter Openers

Each chapter opens with an interesting business example. The stories of companies such as Amazon.com, The Home Depot, and KEEN Inc. enhance and illustrate the message of each chapter, showing students how and why statistical thinking is vital to modern business decision-making. We analyze data from these examples throughout the chapter.

Table 4.2 A relative frequency table for the same data.

100.01%?
If you are careful to add the percentages in Table 4.2, you will notice the total is 100.01%. Of course the real total has to be 100.00%. The discrepancy is due to individual percentages being rounded. You'll often see this in tables of percents, sometimes with explanatory footnotes.

In-Text Examples

Real business examples motivate the discussions, often returning to the chapter-opening company.

4.3 Charts

The Area Principle

Now that we have a frequency table, we're ready to follow the three rules of data analysis and make a picture of the data. But we can't make just any picture; a bad picture can distort our understanding rather than help it. For example, here's a graph of the frequencies of Table 4.1. What impression do you get of the relative frequencies of visits from each source?

While it's true that the majority of people came to KEEN's website from Google, in Figure 4.2 it looks like nearly all did. That doesn't seem right. What's wrong? The lengths of the sandals *do* match the totals in the table. But our eyes tend to be more impressed by the *area* (or perhaps even the *volume*) than by other aspects of each

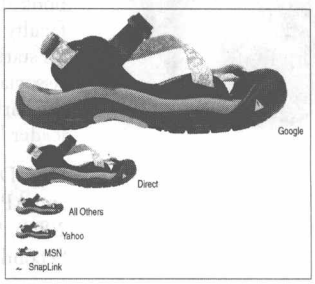


Figure 4.2 Although the length of each sandal corresponds to the correct number, the impression we get is all wrong because we perceive the entire area of the sandal. In fact, only a little more than 50% of all visitors used Google to get to the website.

Applying the Concepts

The Sharpe Edge: Plan, Do, Report

There are three simple steps to doing Statistics right: ***Plan***, ***Do***, and ***Report***.

We lead students through the process of making business decisions with data. The first step is planning how to tackle a problem, the second is doing the calculations, and the third is reporting the results and conclusions. In each chapter, we apply the new concepts learned in *Guided Examples*. Examples are structured to reflect the way statisticians approach and solve problems. These step-by-step examples show students how to produce the kind of solutions and reports that clients expect to see.

PLAN first. Know where you're headed and why. Clearly defining and understanding your objective will save you a lot of work. What do you know? What do you hope to learn? Are the assumptions and conditions satisfied?

DO is the mechanics of calculating statistics. This is what most people think Statistics is about. But the computations don't tell the whole story.

REPORT what you've learned. Until you've explained your results in the context of the business question in your Plan, the job isn't done. We present the report step as a memo to emphasize the decision aspect of each example.

GUIDED EXAMPLE | *New York Stock Exchange Trading Volume*

Are some months on the NYSE busier than others? Boxplots of the number of shares traded by month are a good way to see such patterns. We're interested not only in the centers, but also in the spreads. Are volumes equally variable from month to month, or are they more spread out in some months?

PLAN

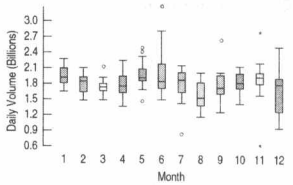
Setup Identify the variables, report the time frame of the data, and state the objective.

We want to compare the daily volume of shares traded from month to month on the NYSE during 2006.

The daily volume is quantitative and measured in number of shares. We can partition the values by month and use side-by-side boxplots to compare the volume across months.

DO

Mechanics Plot the side-by-side boxplots of the data.



REPORT

Conclusion Report what you've learned about the data and any recommended action or analysis.

MEMO:

Re: Research on trading volume of the NYSE

We have examined the daily sales volume on the NYSE (number of shares traded) for each month of 2006. As the attached display shows, sales volume follows a seasonal pattern with lower volume in March and August. The highest median trading activity is found in May and November. The variability of trading volume also shows a pattern. June and December have higher variability than the rest, and March has noticeably less variability. There were several unusually high-volume days that bear investigation and one extremely low-volume day in November.

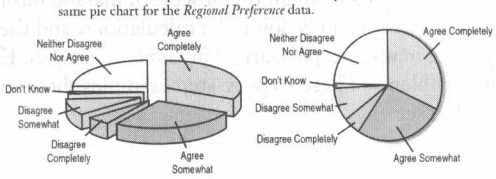
Promoting Understanding

What Can Go Wrong?

The most common mistakes for those new to statistical analysis usually involve misusing a method, not miscalculating a statistic. We acknowledge these mistakes with *What Can Go Wrong?*, found at the end of each chapter. Our goal is to arm students with the tools to detect statistical errors and offer practice in debunking misuses of statistics.

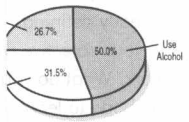
WHAT CAN GO WRONG?

- **Don't violate the area principle.** This is probably the most common mistake in a graphical display. Violations of the area principle are often made for the sake of artistic presentation. Here, for example, are two versions of the same pie chart for the *Regional Preference* data.

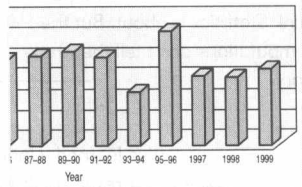


The one on the left looks interesting, doesn't it? But showing the pie three dimensionally on a slant violates the area principle and makes it much more difficult to compare fractions of the whole made up of each category of the response—the principal feature that a pie chart ought to show.

- **Keep it honest.** Here's a pie chart that displays data on the percentage of high e in specified dangerous behaviors as reported by ntrol. What's wrong with this plot?



iges. Or look at the 50% slice. Does it look right? e percentages of? Is there a "whole" that has been e proportions shown by each slice of the pie must i individual must fall into only one category. Of a slant makes it even harder to detect the error. le. This bar chart shows the number of airline rity screening.



Math Box

The mathematical underpinnings of statistical methods and concepts are set apart to avoid interrupting the explanation of the topic at hand. We use these derivations to increase students' understanding of the underlying mathematics, but they can be skipped by less mathematically inclined students.

MATH BOX

Standardizing the variables first gives us an easy to understand expression for the correlation.

$$r = \frac{\sum z_x z_y}{n-1}$$

But sometimes you'll see other formulas. Remembering how standardizing works gets us from one formula to the other.

Since $z_x = \frac{x - \bar{x}}{s_x}$

and $z_y = \frac{y - \bar{y}}{s_y}$,

we can substitute these and get

$$r = \left(\frac{1}{n-1}\right)\sum z_x z_y = \left(\frac{1}{n-1}\right)\sum \frac{(x-\bar{x})}{s_x}\frac{(y-\bar{y})}{s_y} = \sum \frac{(x-\bar{x})(y-\bar{y})}{(n-1)s_x s_y}.$$

That's one version. And since we know the formula for standard deviation,

$$s_y = \sqrt{\frac{\sum (y-\bar{y})^2}{n-1}},$$

we could use substitution to write:

$$r = \left(\frac{1}{n-1}\right)\sum \frac{(x-\bar{x})}{s_x}\frac{(y-\bar{y})}{s_y}$$

$$= \left(\frac{1}{n-1}\right)\frac{\sum (x-x)(y-y)}{\sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}\sqrt{\frac{\sum (y-\bar{y})^2}{}}}$$

$$= \left(\frac{1}{n-1}\right)\frac{\sum (x-\bar{x})}{\left(\frac{1}{n-1}\right)\sqrt{\sum (x-}}$$

$$= \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}.$$

That's the other common version. If you ever h tion by hand, it's easier to start with one of thes ber how correlation works, stick with the first fo

Finding the correlation coefficient by hand

To find the correlation coefficient by hand, we'll use a formula in original units, rather than z-scores. This will save us the work of having to standardize each individual data value first. Start with the summary statistics for both variables: $\bar{x}$, $\bar{y}$, s_x, and s_y. Then find the deviations as we did for the standard deviation, but now in both x and y: $(x - \bar{x})$ and $(y - \bar{y})$. For each data pair, multiply these deviations together: $(x - \bar{x}) \times (y - \bar{y})$. Add the products up for all data pairs. Finally, divide the sum by the product of $(n - 1) \times s_x \times s_y$ to get the correlation coefficient.

Here we go.

Suppose the data pairs are:

x	6	10	14	19	21
y	5	3	7	8	12

Then $\bar{x} = 14$, $\bar{y} = 7$, $s_x = 6.20$, and $s_y = 3.39$

Deviations in x	Deviations in y	Product
6 − 14 = −8	5 − 7 = −2	−8 × −2 = 16
10 − 14 = −4	3 − 7 = −4	16
14 − 14 = 0	7 − 7 = 0	0
19 − 14 = 5	8 − 1 = 1	5
21 − 14 = 7	12 − 7 = 5	35

Add up the products: 16 + 16 + 0 + 5 + 35 = 72
Finally, we divide by $(n - 1) \times s_x \times s_y = (5 - 1) \times 6.20 \times 3.39 = 84.07$
The ratio is the correlation coefficient:

$$r = 72/84.07 = 0.856$$

entury—
nt single
es it even

(continued)

By Hand

Although we encourage using technology to perform statistical calculations, we recognize the benefits of knowing how to compute by hand. *By Hand* boxes explain formulas and help students through the calculation of a worked example.

Checking Understanding

Just Checking

Once or twice per chapter, *Just Checking* asks students to stop and think about what they've read. These questions are designed to check student understanding and involve little calculation. Answers are provided at the end of the chapter so students can easily check their work.

JUST CHECKING

So that they can balance their inventory, an optometry shop collects the following data for customers in the shop.

		Eye Condition			
		Near Sighted	**Far Sighted**	**Need Bifocals**	Total
Sex	**Males**	6	20	6	**32**
	Females	4	16	12	**32**
	Total	**10**	**36**	**18**	**64**

1. What percent of females are far-sighted?
2. What percent of near-sighted customers are female?
3. What percent of all customers are far-sighted females?
4. What's the distribution of *Eye Condition*?
5. What's the conditional distribution of *Eye Condition* for males?
6. Compare the percent who are female among near-sighted customers to the percent of all customers who are female.
7. Does it seem that *Eye Condition* and *Sex* might be dependent? Explain.

What Have We Learned?

These chapter-ending summaries highlight the concepts introduced in the chapter, define new terms, and list the skills presented in the chapter. If students understand all these parts, they're probably ready for the exam.

What have we learned?

We've learned that we can summarize categorical data by counting the number of cases in each category, sometimes expressing the resulting distribution as percents. We can display the distribution in a bar chart or a pie chart. When we want to see how two categorical variables are related, we put the counts (and/or percentages) in a two-way table called a contingency table.

- We look at the marginal distribution of each variable (found in the margins of the table).
- We also look at the conditional distribution of a variable within each category of the other variable.
- We can display these conditional and marginal distributions using bar charts or pie charts.
- If the conditional distributions of one variable are (roughly) the same for every category of the other, the variables are independent.

Terms

Area principle A principle that helps to interpret that in a statistical display, each da

Bar chart (relative frequency bar chart) A chart that represents the count variable as a bar, allowing easy vi

Column percent The proportion of each column

Conditional distribution The distribution of a variable res of individuals.

Contingency table A contingency table displays cou falling into named categories on individuals on all variables at onc may be contingent on the catego

Distribution The distribution of a variable is

- all the possible values of the v
- the relative frequency of each

Frequency table (relative frequency table) A table that lists the categories in a categorical variable and gives the number (the percentage) of observations for each category. The row percent is the proportion of each row contained in the cell of a frequency table, while the column percent is the proportion of each column contained in the cell of a frequency table.

Independent variables Variables for which the conditional distribution of one variable is the same for each category of the other.

Marginal distribution In a contingency table, the distribution of either variable alone. The counts or percentages are the totals found in the margins (usually the right-most column or bottom row) of the table.

Skills

PLAN

- Recognize when a variable is categorical and choose an appropriate display for it.
- Understand how to examine the association between categorical variables by comparing conditional and marginal percentages.

DO

- Summarize the distribution of a categorical variable with a frequency table.
- Display the distribution of a categorical variable with a bar chart or pie chart.
- Construct and examine a contingency table.
- Construct and examine displays of the conditional distributions of one variable for two or more groups.

REPORT

- Describe the distribution of a categorical variable in terms of its possible values and relative frequencies.
- Describe any anomalies or extraordinary features revealed by the display of a variable.
- Describe and discuss patterns found in a contingency table and associated displays of conditional distributions.

Integrating Technology

Technology Help

In business, Statistics is practiced with computers. We offer specific guidance for several of the most common Statistics software (Excel® 2007 and 2003, Minitab®, SPSS®, JMP®, and DataDesk®), often with an annotated example to help students get started with the technology of their choice.

Technology Help: Confidence Intervals for Proportions

Confidence intervals for proportions are so easy and natural that many statistics packages don't offer special commands for them. Most statistics programs want the "raw data" for computations. For proportions, the raw data are the "success" and "failure" status for each case. Usually, these are given as 1 or 0, but they might be category names like "yes" and "no." Often we just know the proportion of successes, $\hat{p}$, and the total count, n. Computer packages don't usually deal with summary data like this easily, but the statistics routines found on many graphing calculators allow you to create confidence intervals from summaries of the data—usually all you need to enter are the number of successes and the sample size.

In some programs you can reconstruct variables of 0's and 1's with the given proportions. But even when you have (or can reconstruct) the raw data values, you may not get *exactly* the same margin of error from a computer package as you would find working by hand. The reason is that some packages make approximations or use other methods. The result is very close but not exactly the same. Fortunately, Statistics means never having to say you're certain, so the approximate result is good enough.

EXCEL

Inference methods for proportions are not part of the standard Excel tool set.

Comments

For summarized data, type the calculation into any cell and evaluate it. Confidence intervals for a proportion are available in the DDXL add-in. Select the range of data holding the variable. Then choose **Confidence Intervals** from the **DDXL** menu. Choose **1 Var Prop Interval** from the menu, indicate the variable, and click **OK.**

MINITAB

Choose **Basic Statistics** from the **Stat** menu.

- Choose **1Proportion** from the Basic Statistics submenu.
- If the data are category names in a variable, assign the variable from the variable list box to the **Samples in columns** box. If you have summarized data, click the **Summarized Data** button and fill in the number of trials and the number of successes.
- Click the **Options** button and specify the remaining details.
- If you have a large sample, check **Use test and interval based on normal distribution.**
  Click the **OK** button.

Comments

When working from a variable that names categories, MINITAB treats the last category as the "success" category. You can specify how the categories should be ordered.

SPSS

SPSS does not find confidence intervals for proportions.

JMP

For a **categorical** variable that holds category labels, the **Distribution** platform includes tests and intervals for proportions. For summarized data, put the category names in one variable and the frequencies in an adjacent variable. Designate the frequency column to have the **role** of **frequency**. Then use the **Distribution** platform.

Comments

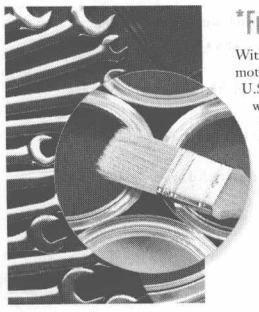
JMP uses slightly different methods for proportion inferences than those discussed in this text. Your answers are likely to be slightly different, especially for small samples.

DATA DESK

nts

narized data, open a Scratchpad to compute the
deviation and margin of error by typing the calcula-
n use **z-interval for individual μs.**

Mini Case Study Projects



*Fuel Efficiency

With the ever increasing price of gasoline, both drivers and auto companies are motivated to raise the fuel efficiency of cars. Recent information posted by the U.S. government proposes some simple ways to increase fuel efficiency (see www.fueleconomy.gov): avoid rapid acceleration, avoid driving over 60 mph, reduce idling, and reduce the vehicle's weight. An extra 100 pounds can reduce fuel efficiency (mpg) by up to 2%. A marketing executive is studying the relationship between the fuel efficiency of cars (as measured in miles per gallon) and their weight to design a new compact car campaign. In the data set **ch07_MCSP_Fuel_Efficiency** you'll find data on the variables below.

- Model of Car
- Engine Size (L)
- Cylinders
- MSRP (Manufacturer's Suggested Retail Price in $)
- City (mpg)
- Highway (mpg)
- Weight (pounds)
- Type and Country of manufacturer

Describe the relationship of Weight, MSRP, and Engine Size with fuel efficiency (both City and Highway) in a written report. Be sure to transform the variables if necessary.

Mini Case Study Projects

Each chapter includes one or two *Mini Case Study Projects* that use real data and ask students to investigate a question or make a decision. Students define the objective, plan the process, complete the analysis, and report their conclusion. Data for the *Mini Case Study Projects* are available on the DVD and website, formatted for various technologies.

Tackling Problems

Exercises

We have worked hard to ensure that exercises contain relevant and modern questions with real data. The exercises generally start with a straightforward application of the chapter ideas, then tackle larger problems. Many break a problem into several parts to help guide the student through the logic of a complete analysis. Finally, there are exercises that ask the student to synthesize and incorporate their ideas with less guidance. Data for exercises marked T are available on the DVD and website, formatted for various technologies.

EXERCISES

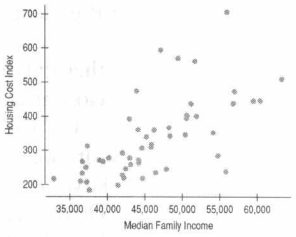
40. Selling fuel economy 2007. In 2006, a study by *Consumer Reports* found that 37% of nationwide respondents reported that they were considering replacing their current car with one with greater fuel economy. Here are advertised horsepower ratings and expected gas mileage for several 2007 vehicles (www.kbb.com/KBB/ReviewsAndRating).

Vehicle	Horsepower	Highway Gas Mileage (mpg)
Audi A4	200	32
BMW 328	230	30
Buick LaCrosse	200	30
Chevy Cobalt	148	32
Chevy TrailBlazer	291	22
Ford Expedition	300	20
GMC Yukon	295	21
Honda Civic	140	40
Honda Accord	166	34
Hyundai Elantra	138	36
Lexus IS 350	306	28
Lincoln Navigator	300	18
Mazda Tribute	212	25
Toyota Camry	158	34
Volkswagen Beetle	150	30

a) Make a scatterplot for these data.
b) Describe the direction, form, and strength of the plot.
c) Find the correlation between horsepower and miles per gallon.
d) Write a few sentences telling what the plot says about fuel economy.

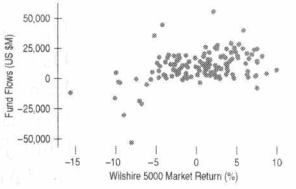
T **41.** Pizza sales. Here is a scatterplot for the weekly sales (in pounds) for every fourth week of a brand of frozen pizza versus the unit price of the pizza for a sample of stores in the Dallas area.

T **42.** Housing costs. Concern over the possibility of a "home cost bubble" has lead many economists to examine housing costs. The Office of Federal Housing Enterprise Oversight (www.ofheo.gov) collects data on various aspects of housing costs around the United States. Here is a scatterplot of the *Housing Cost Index* versus the *Median Family Income* for each of the 50 states. The correlation is 0.65.



a) Describe the relationship between the *Housing Cost Index* and the *Median Family Income* by state.
b) If we standardized both variables, what would the correlation coefficient between the standardized variables be?
c) If we had measured *Median Family Income* in thousands of dollars instead of dollars, how would the correlation change?
d) Washington, DC, has a *Housing Cost Index* of 548 and a median income of about $45,000. If we were to include DC in the data set, how would that affect the correlation coefficient?
e) Do these data provide proof that by raising the median income in a state, the *Housing Cost Index* will rise as a result? Explain.

T **43.** Mutual funds. Here is a scatterplot showing the association between money flowing into mutual funds (fund flows in $M) and a specific type of market return (Wilshire Index) for each month from 1990 to 2002.



a) Is it appropriate to calculate a correlation? Explain.
b) Identify the largest outlier in the scatterplot. Find and discuss

ETHICS IN ACTION

Beth Tully owns Zenna's Café, an independent coffee shop located in a small mid-western city. Since opening Zenna's in 2002, she has been steadily growing her business and now distributes her custom coffee blends to a number of regional restaurants and markets. She operates a microroaster that offers specialty grade Arabica coffees recognized as some as the best in the area. In addition to providing the highest quality coffees, Beth also wants her business to be socially responsible. Toward that end, she pays fair prices to coffee farmers and donates funds to help charitable causes in Panama, Costa Rica, and Guatemala. In addition, she encourages her employees to get involved in the local community. Recently, one of the well-known multinational coffeehouse chains announced plans to locate shops in her area. This chain is one of the few to offer Certified Free Trade coffee products and work toward social justice in the global community. Consequently, Beth thought it might be a good idea for her to begin communicating Zenna's socially responsible efforts to the public, but with an emphasis on their commitment to the local community. Three months ago she began collecting data on the number of volunteer hours donated by her employees per week. She has a total of 12 employees, of whom 10 are full time. Most employees volunteered less than 2 hours per week, but Beth noticed that one part-time employee volunteered more than 20 hours per week. She discovered that her employees collectively volunteered an average of 15 hours per month (with a median of 8 hours). She planned to report the average number and believed most people would be impressed with Zenna's level of commitment to the local community.

ETHICAL ISSUE *The outlier in the data affects the average in a direction that benefits Beth Tully and Zenna's Café (related to Item C, ASA Ethical Guidelines).*

ETHICAL SOLUTION *Beth's data are highly skewed. There is an outlier value (for a part-time employee) that pulls the average number of volunteer hours up. Reporting the average is misleading. In addition, there may be justification to eliminate the value since it belongs to a part-time employee (10 of the 12 employees are full time). It would be more ethical for Beth to: (1) report the average but discuss the outlier value; (2) report the average for only full-time employees, or (3) report the median instead of the average.*

rop, de-

r box is
a boxes

Ethics in Action

Our ethics vignettes in each chapter illustrate some of the judgments needed in statistical analysis, identify possible errors, link the issues to the ASA's Ethical Guidelines, and then propose ethically and statistically sound alternatives.

Acknowledgements

Many people have contributed to this book from the first day of its conception to its publication. This book would not have been possible without the many contributions of David Bock, our co-author on several of our other textbooks. Many of the explanations and exercises in this book benefit from Dave's pedagogical flair and expertise. We are honored to have him as a colleague and friend. *Business Statistics* would have never seen the light of day without the assistance of the incredible team at Pearson. Our Editor in Chief, Deirdre Lynch, was central to the support, development, and realization of the book from day one. Sara Oliver Gordus, Associate Editor, and Chere Bemelmans, Senior Project Editor, kept us on task as much as humanly possible. Peggy McMahon, Senior Production Supervisor, and Laura Hakala, Senior Project Manager at Pre-Press PMG, worked miracles to get the book out the door. We are indebted to them. Christina Lepre, Assistant Editor; Dana Jones, Editorial Assistant; Alex Gay, Marketing Manager; Kathleen DeChavez, Marketing Associate; and Dona Kenly, Senior Market Development Manager, were essential in managing all of the behind-the-scenes work that needed to be done. Christine Stavrou, Media Producer, put together a top-notch media package for this book. Barbara Atkinson, Senior Designer, and Geri Davis are responsible for the wonderful way the book looks. Evelyn Beaton, Manufacturing Manager, and Ginny Michaud, Senior Manufacturing Buyer, worked miracles to get this book and DVD in your hands, and Greg Tobin, President, was supportive and good-humored throughout all aspects of the project.

Special thanks go out to Pre-PressPMG, the compositor, for the wonderful work they did on this book and in particular to Laura Hakala, the project manager, for her close attention to detail.

We'd also like to thank our accuracy checkers whose monumental task was to make sure we said what we thought we were saying: Eugene Allevato, Woodbury University; Dave Bregenzer, Utah State University; Ann Cannon, Cornell College; Joan Donohue, University of South Carolina; David Doorn, University of Minnesota, Duluth; David Hudgins, University of Oklahoma, Norman; John Lawrence, California State University, Fullerton; Joe Kupresanin, Cecil College; Monnie McGee, Southern Methodist University; Jackie Miller, The Ohio State University; Doug Morris, University of New Hampshire; Michael Polomsky, Cleveland State University; Gary Smith, Florida State University; Joe Sullivan, Mississippi State University; Dirk Tempelaar, Maastricht University; William Warde, Oklahoma State University; and Jim Zimmer, Chattanooga State University.

We wish to thank the following individuals who joined us for a weekend to discuss business statistics education, emerging trends, technology, and business ethics. These individuals made invaluable contributions to *Business Statistics*:

Dr. Taiwo Amoo, City University of New York, Brooklyn
Dave Bregenzer, Utah State University
Joan Donohue, University of South Carolina
Soheila Fardanesh, Towson University
Chun Jin, Central Connecticut State University
Brad McDonald, Northern Illinois University
Amy Luginbuhl Phelps, Duquesne University
Michael Polomsky, Cleveland State University
Robert Potter, University of Central Florida
Rose Sebastianelli, University of Scranton
Debra Stiver, University of Nevada, Reno
Minghe Sun, University of Texas—San Antonio
Mary Whiteside, University of Texas—Arlington