

妙趣横生的统计学入门书 从零开始 文科生也能学好统计学

心理统计

日常生活中的统计推理

(原书第3版)

杰弗里 O. 贝内特 (Jeffrey O. Bennett)

科罗拉多大学波尔得分校

(美) 威廉 L. 布里格斯 (William L. Briggs)

科罗拉多大学丹佛分校

著

马里奥 F. 特里奥拉 (Mario F. Triola)

达奇斯社区学院

封文波 译

Statistical Reasoning for Everyday Life 3rd Edition



机械工业出版社
China Machine Press

总有一天，统计学的思维方式就像读写能力一样，会成为你的基本素质。



02030113

本书是一本极具特色的统计学入门书，利用丰富有趣的真实案例，帮助学生（尤其是需要发展一定统计素养的人文社科学生）从零开始学好统计，同时也适合那些想要利用统计学提高批判性思维能力、做出更优决策的人。

本书特色

- **以生活中的问题为切入点，简单易懂**

统计数字的含义和奇妙之处在于与我们的生活有着千丝万缕的联系。本书的内容不但包含人文社科（心理学、社会学、经济学等）学生所必需的统计学基础知识，更加强调以生活中的问题为切入点，使统计学问题从遥不可及、高深晦涩变为平易近人、简单易懂，有效提升学习效率。

- **写作深入浅出，帮助你逐步建立统计学思维**

本书的三位作者均为教学经验丰富、著作颇丰的统计学教师，所以本书的写作逻辑严谨周密却又轻松易读，思考深入浅出，结合生活实际，以理性的思维方式来面对感性的学习者，帮助学习者从零开始、循序渐进地建立统计学思维方式。

- **丰富具体的实例，增加学习过程的乐趣**

本书的最大特色就是拥有极其丰富的具体实例，使内容充实、言之有物，大大提高了学生学习的兴趣和对内容的理解，让人一旦拿起，便不忍释卷。书中每一章后穿插的生活中的完整案例，也开拓了学生的眼界，使学生更加全面地以统计学思维方式来解读生活和我们的世界。

PEARSON

PEARSON

www.pearson.com

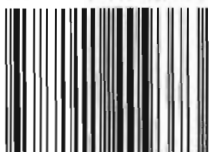


客服热线: (010) 88379210 88361066
购书热线: (010) 68326294 88379649 68995259
投稿热线: (010) 88379007

读者信箱: hzjg@hzbook.com
华章网站: www.hzbook.com
网上购书: www.china-pub.com

上架指导: 心理学

ISBN 978-7-111-40937-3



9 787111 409373 >

定价: 65.00元

心理统计

日常生活中的统计推理

(原书第3版)

杰弗里 O. 贝内特 (Jeffrey O. Bennett)

科罗拉多大学波尔得分校

(美) 威廉 L. 布里格斯 (William L. Briggs)

科罗拉多大学丹佛分校

著

马里奥 F. 特里奥拉 (Mario F. Triola)

达奇斯社区学院

封文波 译

Statistical Reasoning for Everyday Life 3rd Edition



机械工业出版社
China Machine Press

图书在版编目(CIP)数据

心理统计: 日常生活中的统计推理(原书第3版)/(美)贝内特(Bennett, J. O.), (美)布里格斯(Briggs, W. L.), (美)特里奥拉(Triola, M. F.)著; 封文波译. —北京: 机械工业出版社, 2013. 1
(心理学精品教材译丛)

书名原文: Statistical Reasoning for Everyday Life

ISBN 978-7-111-40937-3

I. 心… II. ①贝… ②布… ③特… ④封… III. 心理统计—教材 IV. B841.2

中国版本图书馆CIP数据核字(2012)第304651号

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问 北京市展达律师事务所

本书版权登记号: 图字: 01-2009-7318

本书的特色是注重统计学在日常生活中的应用, 内容设计旨在帮助读者理解统计学的核心理念, 具备基本的统计素养, 形成批判性思维。因为日常生活中无论是开展商业计划还是安排财务预算, 甚至只是观看电视新闻, 统计也无处不在。作为现代社会的一员, 理解统计学背后的核心理念, 对于成功至关重要。

本书适合社会学、经济学、教育学、心理学等本专科学生使用。

Jeffrey O. Bennett, William L. Briggs, Mario F. Triola, Statistical Reasoning for Everyday Life, 3rd Edition.

ISBN 978-0-321-28672-7

Copyright © 2009 by Pearson Education, Inc., Publishing as Addison-Wesley.

Simplified Chinese Edition Copyright © 2013 by China Machine Press.

Published by arrangement with the original publisher, Pearson Education, Inc. This edition is authorized for sale and distribution in the People's Republic of China exclusively (except Taiwan, Hong Kong SAR and Macau SAR).

All rights reserved.

本书中文简体字版由 Pearson Education (培生教育出版集团) 授权机械工业出版社在中华人民共和国境内(不包括中国台湾地区和香港、澳门特别行政区)独家出版发行。未经出版者书面许可, 不得以任何方式抄袭、复制或节录本书中的任何部分。

本书封底贴有 Pearson Education (培生教育出版集团) 激光防伪标签, 无标签者不得销售。

机械工业出版社(北京市西城区百万庄大街22号) 邮政编码 100037

责任编辑: 赵艳君 版式设计: 刘永青

北京诚信伟业印刷有限公司印刷

2013年2月第1版第1次印刷

185mm×260mm·23.75印张

标准书号: ISBN 978-7-111-40937-3

定 价: 65.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88379210 88361066

投稿热线: (010) 88379007

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjg@hzbook.com

FOREWORD

译者序

与数字打交道的统计学往往是枯燥乏味的代名词，作为一个讲授统计学课程的大学教师，每当在课堂上看到学生以一脸茫然和焦灼的神情看着我时，一种挫败感和自责就会油然而生，如何让学生愉快地学习这门课程、轻松领悟其思维方式、快速掌握其方法和原理，就成了我，我想也应该是所有统计学任课教师的一个特别想实现的理想。

学好一门课程，最重要的在于积极主动的思考，对于统计学课程来说，这种主动性和积极性尤为重要。怎样才能做到这点呢？一是个人具有独立思考的习惯，二是所学习的事物让我们容易产生思考。思考的习惯有赖于终身学习的塑造，一个人的中小学时代是形成这种习惯的关键时期。是否形成思考的良好习惯，我们的学校与老师对此负有不可推卸的责任，当然，学生自己才是最终的责任者，在这一点上，与那些有着良好思考习惯的同学一起来学习这门感觉上很枯燥的课程，作为教师才能真正体会到什么是教学相长。当然，这取决于机遇的安排，你不可能期望所有同学都对数字有着敏感的洞察力和强烈的兴趣，因此在统计课堂上尽量关注和引导学生积极思考，也是统计课程教师努力要达成的目标之一。除此之外，如果将要学习的内容与我们的生活息息相关，就更容易引起学习者的思考，无论思考与否，生活都会继续，但思考有助于生活则是毋庸置疑的，从某种角度上讲，思考甚至是生活的一个重要目标。在这里我很想引用赫伯特·乔治·威尔斯的一句话作为我的期望，“和读写能力一样，终有一天统计思维会成为每个合格市民的必备素质”。

学好一门课程，另一个重要方面就是有个好“老师”，这个好老师可以是一位教师，当然也可以是一本好书。只不过和书这个“老师”打交道的话就更有赖于我们前面所说的第一点，你要喜欢思考，并且善于思考，当然好老师会尽力引导和启发你思考，但归根结底，再好的老师要发挥作用，也在于个体自身的主动性和积极性。如果有一本非常好的参考书，这就是不会说话但却能够与你良好交流的老师。

一个好老师大体要有以下几方面的要求：一是能够引导和启发你思考；二是能在你思考时给你助力；三是能与你顺畅交流。从这几点来说，这本书可以说算是一个无声的好老师了。要做到第一点，主要在于所选择的主题，本书所选的一些统计学主题，力图让学习者明白，统计学离我们很近，统计数字的含义和奇妙之处，在于它与我们的生活有着千丝万缕的联系，而不是象牙塔中一门高深晦涩的学科，本书的内容编排在符合基本统计学理论架构的前提下，更加强调以生活中的问题为切入点，而不是呆板地关注统计学学科体系，在这一点上，我们有些参考书真的应该好好借鉴。要做到第二点，主要在于作者的思维方式和思考逻辑。如果作者的逻辑严谨周密而不

晦涩, 思考深入浅出, 结合生活实际, 帮助我们理解统计在生活中的意义, 通篇融合统计学的思维来解读生活, 以理性的思维方式来面对感性的学习者, 那么当你读完本书, 你可能会发现你思考问题的角度和方式已经在不知不觉中有改变了。要做到第三点, 在于引用的例证和语言表述的方式。案例的编写不是为了哗众取宠, 也不是说教, 而是运用统计学思维方式和你娓娓而谈, 说一些深刻而简单的道理, 更难能可贵的是本书丰富具体的实例, 使内容不空洞而言之有物, 又使得学习过程不枯燥乏味, 大大提高学习的兴趣和对内容的理解。章后穿插的生活中的完整案例, 也开拓了学生的眼界, 使其学会以更加全面的统计学思维方式解读生活和世界。

生活是有序的, 也是无序的。学习是有规律的, 也是有路径的; 是循序渐进的, 也不总是与“刻板”和“艰辛”相联系的。我个人更欣赏“书山有路‘思’为径, 学海无涯‘乐’作舟”的学习理念。学习是生活的一部分, 但不是全部。如果太在意知识体系以致忽略学习的规律, 则得不偿失; 如果硬把学习置于全部生活之上, 则有本末倒置之嫌。学习方式也应该是多元化的、兼容并包的, 选择权最终应该给学习者, 我们只是择其善者而荐之。这本书能够做到内容选择的精致而合理, 思维方式流畅简洁且逻辑严谨周密, 语言表达准确恰当。与一位好老师打交道时, 学生应该是自然放松的状态, 并且受益匪浅, 从这个角度来讲, 这本书与这个目标离得应该很近了。

我在翻译这本书的时候, 体验到了与作者交流的乐趣, 感受到了作者严谨而生动的思维方式, 因此就更想与大家分享我学习的快乐和经验心得。但是我个人的能力有限, 这倒并不是自谦, 如果做不到上述这些, 不能归咎于原书作者的水平, 而只能埋怨我这个译者力有不及了。

兴趣是最好的老师, 引导而不是逼迫, 这样引发的学习动力是惊人的。当然, 本书还有很多内容有待读者发现, 这也算是一个思考的过程吧。总之, 我认为, 这是一本很好的统计学入门参考书。

本书的翻译需要许多人的支持和帮助, 首先感谢机械工业出版社给予我这样的机会, 从而能与这样的美国统计学同行进行交流, 在翻译这本书的过程中, 收获颇多, 也促进了自己的思考。此外, 也非常感谢我的研究生邓苗会、赵玲玲、孟曦、柳晓燕、葛向楠、李楠、武泰、张飞翔, 在我翻译本书的过程中他们给予了许多帮助和支持。

PREFACE

前言

为什么学习统计学

科幻小说作家威尔斯 (H. G. Wells) 曾经说过：“和读写能力一样，终有一天统计思维会成为每个合格市民的必备素质。”威尔斯设想的未来是这样的，统计学将成为日常生活重要的一部分，无论你是要开展一项新的商业计划，还是要安排财务预算，甚至只是观看电视新闻，统计将无处不在。从民意测验到经济报道，再到最近的癌症预防研究，都在应用统计学。作为现代社会的一员，理解统计学背后的核心理念，对于成功而言是至关重要的。

在本书中你会学到什么样的统计学

统计学是一个非常丰富的研究领域，丰富到即使一个人穷其一生去学习和研究它，仍然会感到有很多内容需要深入探索。尽管如此，在大学阶段经过一个学期或者一学年的学习，你依然可以掌握统计学的核心理念。本书的设计旨在帮助你理解这些核心理念。在本书中你将学到的统计知识是日常生活中所需要的，也是你在一些课程学习中必备的统计学基础知识。尤其需要强调的是，本书的设计旨在实现以下三个目标：

- (1) 为你提供在**大学**课程学习中所必需的统计学基础知识，尤其是在诸如经济学、心理学、社会学和政治学等社会科学领域。
- (2) 帮你提高利用统计信息进行推理的能力，这是现代社会里任何**职业**都不可或缺的能力。
- (3) 促使你提高自己的评估、判断和决策能力，关于如何评估**日常生活**中与统计研究有关的新闻报道；如何形成关于这些报道内容的看法；如何看待新闻报道的结论；以及是否因此受其影响或改变你自己的生活方式。

本书适合哪些读者

我们希望本书对每个人都将有益处，但是它首先针对的是那些不打算继续深造高级统计课程

的学生。尤其是，对于大多数人类和社会科学范畴的学生而言，因其主修专业属于一个广泛的领域并且需要一定的统计素养，本书将提供适宜的统计学入门知识。此外，这本书的文字水平适用于任何一位已经完成两年高中数学课程的学习者。

学习方法

本书提供了一个设计方案以帮助你理解重要的量化统计观念，以及在对这些观念进行分类时如何使用定性技术。下面列出了一些指导本书写作的关键性教学策略以供参考。

- **从统计学的重点和框架开始。**对于统计学，大多数开始学习这门课程的人多少都会有一些了解，当我们学习一些具体概念和方法时，了解统计学的整体目标是非常重要的，因此我们从第1章的统计概述开始这本书。在此章中，我们会解释样本和总体的关系，讨论抽样方法和统计研究的各种类型，呈现大量的案例来帮助你确定是否可以相信一项统计研究的结果。这种方式，为读者更深入地学习统计学思想提供了可靠的基础。
- **逐步建立统计思想。**统计学课程的所有目标都是帮助学生真正地理解统计学问题。当然，为了理解更复杂的统计研究，从简单案例开始学习通常会更容易，在每一章我们都应用这种策略，由简入繁，逐步引向实证和特定个案研究。
- **通过计算增进理解。**本书的基本目标在于帮助学生理解统计学的概念和思想，我们坚定地认为要达到这个目标至少也要做一些必要的计算练习。本书之所以包含较多练习题也是因为它们可以增进我们对统计思想的深层理解。
- **将概率和统计学紧密联系。**自从人们意识到概率在科学统计学中起着必不可少的作用，大多数统计学教材会包括基础概率知识，但是对于学生来说，概率的概念通常不能与其他统计学的主题相联系，这让人非常惭愧。本书从第1章就开始讨论这一点，结合统计研究的基本结构，然后依次贯穿全书，特别在第6章我们会介绍更多概率的思想。对于一些不大强调概率知识专业的学生而言，第6章可以作为可选章节学习。
- **强调目标：将统计推理应用到日常生活。**统计学是一门涉及广泛领域的学科，它到底包括哪些主题，是比较难以确定的。当写作本书时，我们总要返回头来，看看本书所反映的主题，看看读者在日常生活中会需要哪些应用统计学推论。如果我们感到一个主题在日常生活中不是经常遇到的，我们就不予考虑。按照同样的原则，我们比其他统计学教材增加了一些这样的主题，比如第2章的百分比、第3章的制图法。这些是一般统计学教材中不经常包括的内容，但却是日常生活常用到的统计学的主要部分。

模块结构

虽然本书按一定结构编排，以期读者能够从头到尾依序阅读，但我们认识到，还有些教师希

望能以不同的顺序,或者根据不同的课时长度及针对不同水平的学生只选取书中的一部分材料进行讲解。因此本书以模块结构方式进行组织,以便教师能够根据课程的具体安排自行选取材料。依据不同模块,全书分为10章,每章包括若干相对独立的小节,每小节都有一个核心的统计主题或应用。大多数情况下,教师可以根据需要按任意顺序来教授各个章节,也可以跳过某些章节。注意每一章都是依照以下特定结构设计的。

- **学习目标。**每章开始都有学习目标,帮助学生指导自己的学习。
- **相关知识和思考题。**各章都有很多相关知识介绍,主要是社会常识以方便学生理解。为了促进学生更积极主动地思考,各章有许多思考题,这也可以帮助教师发起精彩的课堂讨论。
- **练习题。**每节最后都有相关日常应用的练习题,主要包括以下方面:统计素养和批判性思维,概念与应用,网络和远程项目,新闻中的统计学等。大部分奇数题号的练习题答案都列于本书末尾。
- **复习题和小测验。**每章结尾处都有简短的复习题和小测验,这些题便于学生自学,所有答案均列于本书末尾。
- **聚集主题。**每章都包括两篇题为“聚焦……”的文章,以深入讨论当下的重要统计主题,这些主题被选来阐释统计学发挥作用的各个领域,包括历史、环境、农业以及经济等。“聚焦……”栏目包括一系列要完成或者讨论的问题。

TO THE STUDENTS

致学生：怎样学好统计学这门课程

如果你正在阅读本书，你可能已经选修某一种类型的统计学课程。在你的课程中成功的关键在于，首先要怀有一颗开放和乐观的心去接近学习材料，其次要密切关注生活中有用和有趣的统计现象，最后要学会有效地学习。下面的内容可能对于学习统计学是非常有用的提示。

如何使用本书

- 在你形成通常的学习习惯之前，下面的一些指导会帮助你更有效地使用本书。
- 学习各章内容之前，请认真阅读给定的材料两遍：第一遍，快速阅读以了解给定材料和概念的框架结构；第二遍，更深入地阅读，以更仔细地掌握知识内容。
 - 在第二遍阅读中，注意那些对你后续学习有帮助的内容，你还可以用钢笔或铅笔在需要特别注意的地方作标记，这样会帮助你在这些重要的地方备加留心。
 - 要想学得更好，必须要多做练习。在你完成阅读以后，务必要完成大量的练习题和复习题。除了老师布置的一些习题之外，尽量完成那些本书后面有答案的练习题。
 - 学会利用网络上更多可资利用的研究资源。

合理分配你的学习时间

与课堂学习相结合，你可以自行安排学习时间。当然，如果你是一边学习一边工作，那么你需要更合理地安排你的时间。下面是关于如何安排你学习时间的大概指导。

你的课程 (学分)	阅读规定内容 (每周, 小时)	家庭作业 (每周, 小时)	复习和备考 (平均每周, 小时)	总学时 (每周, 小时)
3	1~2	3~5	2	6~9
4	2~3	3~6	3	8~12
5	2~4	4~7	4	10~15

如果发现你的学习时间比表中所建议的要少，你可能需要更多的学习才能提高你的成绩。如果你的学习时间比表中建议的要多，你可能学习效率不高，在这种情况下，你可以去咨询你的指导老师怎样学习才能更有效。

学习的一般策略

(1) 不要缺课。听老师讲和参与讨论比起你阅读其他人的笔记益处更多，积极地参与可以帮助你长久地保持你所学习的内容。

(2) 合理安排你的学习时间。每天花一两个小时学习，比起在考试前熬夜学习，会更有效的，也更少痛苦。

(3) 如果遇到一个你不太理解的新概念，你要学会查阅相关资料，尽量自己先去解答问题。如果你还是有问题，那就寻求帮助，你会发现朋友、同学和老师都很乐意帮助你。

(4) 和朋友一起学习是非常有效的，尤其是在帮助你解决难题的时候。无论如何，你要和同学一起学习，但不要依赖他们。

考前准备

(1) 复习你做过的练习题，尝试做一些附加练习确定你理解了这些概念。翻看这一学期先前你所完成的作业、小测验或者是试题。

(2) 复习你的课堂和讨论笔记，特别注意那些老师重点讲过的内容。

(3) 再次阅读教材的有关章节，特别注意你在书旁的标记。

(4) 在和同学组成一个学习小组之前先独自学习，只有当每个人都准备好积极参与时，学习小组才是有效的。

(5) 不要总是在考试之前才临阵磨枪，人不可能一口吃成个大胖子（想一想任何知识的学习都需要一个过程）。

(6) 在考试之前和考试过程中都要保持放松的心态。如果你认真学习过，你就有能力考好，保持放松的心情会让你更加清楚地思考。

CONTENTS

目 录

译者序

前言

致学生：怎样学好统计学这门课程

第1章 统计学绪论 / 1

1.1 统计学是什么 / 1

1.2 样本 / 9

1.3 统计研究的类型 / 16

1.4 统计研究可信吗 / 25

聚焦社会学 日托培养的方式增加攻击性了吗 / 36

聚焦公众健康 你的生活方式健康吗 / 37

第2章 统计测量 / 39

2.1 数据类型和测量水平 / 39

2.2 误差控制 / 43

2.3 百分数在统计学中的应用 / 50

2.4 指数 / 55

聚焦政治 谁会从减税中受益 / 61

聚焦经济 我们的生活水平提高了吗 / 62

第3章 数据的直观呈现 / 65

3.1 次数分布表 / 65

3.2 绘制数据分布图 / 73

3.3 媒体中的图表 / 82

3.4 有关图表的几个注意事项 / 92

技术应用 / 101

聚焦历史 战争可以用图表描述吗 / 102

聚焦环境 我们怎样才能直观呈现全球变暖 / 104

第4章 数据的描述 / 107

4.1 什么是集中量数 / 107

4.2 分布形态 / 116

4.3 差异量数 / 122

4.4 统计悖论 / 132

技术应用 / 140

聚焦股票市场 什么是道琼斯平均工业指数 / 141

聚焦经济 富人变得更富了吗 / 143

第5章 正态的世界 / 146

5.1 什么是正态 / 146

5.2 正态分布的比例 / 152

5.3 中心极限定理 / 159

技术应用 / 167

聚焦教育 我们能从 SAT 考试的发展趋势中了解到些什么 / 168

聚焦心理学 我们比父母更聪明吗 / 170

第6章 统计中的概率 / 172

6.1 概率在统计学中的作用: 统计显著性 / 172

6.2 概率基础知识 / 176

6.3 概率的大数定律 / 185

6.4 风险观念和预期寿命 / 192

6.5 组合事件的概率(补充部分) / 198

聚焦社会科学 彩票公平吗 / 207

聚焦法律 DNA 指纹识别可靠吗 / 209

第7章 相关和因果 / 211

7.1 寻求相关 / 211

7.2 对相关的解释 / 221

7.3 最佳拟合线和回归 / 228

7.4 寻找因果关系 / 234

技术应用 / 240

聚焦教育 是什么帮助儿童学习阅读 / 241

聚焦环境 是什么原因导致全球变暖 / 242

第8章 从样本到总体 / 246

8.1 抽样分布 / 246

8.2 总体平均数的估计 / 256

8.3 总体比率的估计 / 262

技术应用 / 268

聚焦历史 统计学的起源 / 269

聚焦文化 莎士比亚总共认识多少单词 / 271

第9章 假设检验 / 273

9.1 假设检验的理论基础 / 273

9.2 总体平均数的假设检验 / 281

9.3 总体比率的假设检验 / 292

技术应用 / 297

聚焦健康和教育 你所受的教育是否有利于你长寿 / 298

聚焦农业 转基因食品是否安全 / 300

第10章 t 检验、双向表和方差分析 / 303

10.1 平均数的 t 检验 / 303

10.2 双向表的假设检验 / 309

10.3 方差分析（一元方差分析） / 317

技术应用 / 324

聚焦犯罪学 当你看到一个事物时能否辨别真假 / 326

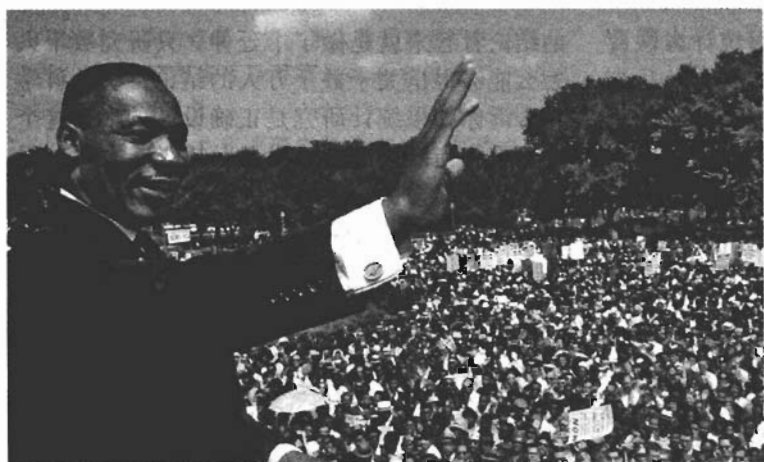
聚焦教育 四年级的学生能用统计学做什么 / 329

结语：统计学展望 / 331

附录A 标准分数表 / 332

附录B 随机数字表 / 335

参考答案 / 336



一个基本的法则：谁掌握了最有用的信息，谁就能成为成功人士。

——本杰明·迪斯雷利

第 1 章 统计学绪论

学习目标

■ 1.1 统计学是什么

理解术语“统计”的两个意思以及一些统计研究背后的基本设想。包括总体、样本、样本统计量和总体参数等概念以及它们的相互关系。

■ 1.2 样本

理解选择一个代表性样本的重要性，熟悉几种常用的抽样方法。

■ 1.3 统计研究的类型

理解观察研究和实验研究的差异，掌握实验研究的关键词，包括实验组和控制组的选择、安慰剂效应、盲试。

■ 1.4 统计研究可信吗

评估你在媒体中听到的统计研究，判断其结果是否有意义。

你的饮用水安全吗？大多数人赞同总统的税收计划吗？我们是否掌握了花钱少而有效的保健方法？上述问题只有通过统计学进行研究。在第 1 章，我们将讨论统计研究的基本概念，为后续章节深入的统计研究打下基础。以同样的方式，我们将讨论大量巧妙设计的统计研究个案，以便为社会政策和个体决策提供指导，同时也讨论一些容易被人误解的统计案例。

1.1 统计学是什么

如果你像大多数使用这本书的同学一样，是一位统计学的初学者，你也许不太确定为什么要学习统计学，也可能对未来的学习会充满焦虑。但是我们希望当你开始阅读本书时，能够充满期待和欣喜，可以把其他杂念放在一边。

统计学经常被归类于方法或技术，但是它与现代社会的方方面面都有关联。统计可以告诉我们一种治疗癌症的新药是否有效；它可以

帮助农产品检验人员判断我们的食品是否安全；它还是所有民意测验的关键所在；商业的市场研究和广告设计会应用统计学；甚至体育运动中也会应用到统计学，比如经常作为运动队和运动员的一种排序方法。的确，在某些重要领域中如果你能想到的理念与统计学无关，那是非常不容易的。

■ 相关知识

统计学起源于针对各州都要有公平的普查和税收数据的收集，这是为什么统计学的词根来自州这个词的原因。

本书的首要目标是帮助你掌握统计方法背后的统计思想。这些基本观念不难理解，尽管掌握它们背后的细节和方法可能需要几年的学习。统计学的一个重要作用就是即使书中一个很小的理论也会给你理解日常生活提供强大的支持，当你试图理解在新闻报道中所看到的、你的学习和工作中所遇到的，以及你日常生活中所经历的统计学问题时，这些理论可以让你的理解变得容易许多。

还有一个地方值得注意，那就是“统计”(statistics)这个术语本身，它在英语中用作单数或者用作复数时，意思是不一样的。当它在英语中是单数时，翻译为汉语就是统计学的意思，它帮助我们理解搜集、整理和解释有关某些观念的数据信息；当它在英语中用复数时，翻译为汉语就是统计的意思，它指的是描述某些事物属性的实际数据。例如，如果你的班里有30名学生，他们的年龄从17~64岁，数字“30个学生”、“17岁”、“64岁”都是以某种方式描述你所在班级某些属性的统计数字。

还需要注意的是，在英语中术语“data”(数据)这个词经常用来作“information”(信息)的同义词，但严格意义上它是复数，一个信息点叫做“datum”，两个以上的信息点才叫做“data”。

1.1.1 如何应用统计学

你观看美式橄榄球超级碗大赛吗？广告客户非常想知道这个答案。因为在大赛期间一个30秒的广告时间价格大约为300万美元。如果有足够多的人观看这个比赛，那这个广告价格就值。例如，新闻报道说有9320万美国人观看了印第安纳波利斯小马队赢得第41届超级碗的那场比赛，但你可能困惑的是：谁去数出这些人数呢？

答案是没有人在做这样的计数。声称9320万美国人观看了赢得印第安纳波利斯小马队第41届超级碗的那场比赛，是来自一个叫做尼尔森媒体研究的调查公司所实施的统计调查，这个公司发布的研究结果就是著名的尼尔森媒体收视率。值得注意的是，尼尔森媒体中心仅仅通过监控5000个住户内人们的收视习惯就编制出了这个收视率。

如果你对统计研究而言是个新手，尼尔森

的结论看起来就是做了个延伸。只研究数千人怎么能够得出关于数千万人的结论？统计科学将告诉你如果统计研究是正确设计的话，这个结论是相当准确的。让我们以超级碗大赛的尼尔森收视率为例，通过几个关键问题来阐明通常情况下统计研究如何进行。

1. 统计研究的目标是什么

尼尔森媒体研究的目标是确定观看超级碗比赛的美国人数目。在统计学语言中，我们称尼尔森感兴趣的总体(population)是全体美国人。尼尔森希望确定的数据是美国观看超级碗的人数，是总体的某个特征。在统计学语言中，总体的特征叫做总体参数(population parameters)。

尽管我们通常以为总体是一个人群，但是一个统计学上的总体可以是一个人群，也可以是一群动物或者一组其他事物。例如，在一个机动车安全的研究中，总体可以是所有在路上行驶的汽车，同理，术语“总体参数”可以指的是总体的任何一个特征。在这个机动车安全的案例中，总体参数可以是某一段时间内正在路上行驶的所有车的数量、在路上行驶的汽车的事故率或者行驶在路上的汽车重量的范围。

相关知识

尼尔森创建他的公司并且发明市场调查是在1923年。他在1942年推出了尼尔森收音机指数来计算广播节目的收听率，并在1960年将这种方法扩展到电视节目上。



例 1-1 总体和总体参数

对于下面每一个情境，请描述其研究的总体，确定感兴趣的总体参数。

(1) 你为农夫保险公司工作，公司要求你确定没

有侧安全气囊的汽车事故受害人接受赔付的平均数。

(2) 你被麦当劳公司雇用来确定每周用于法式炸薯条的土豆用量。

(3) 你是一个商业研究报告者,你在调查基因泰克公司的一种新疗法对于治疗儿童白血病是否有效。

解答

- (1) 总体包括所有的无侧安全气囊的汽车发生事故而接受保险赔付的人,相关的总体参数是赔付人们金额的平均数(参考第4章所讨论的平均数和其他集中量数更多的细节)。
- (2) 总体包括所有的用于法式炸薯条的每周土豆用量,相关的总体参数包括土豆的平均重量和重量的变异性(例如,大多数每周用量是接近还是远高于平均值?)。
- (3) 总体包括所有患有白血病的儿童,重要的总体参数是没有接受新疗法患儿的百分比和接受新疗法患儿的百分比。

2. 实际研究的是什么

如果尼尔森媒体研究是无所不能的,他们可以通过询问每一位个体来确定观看超级碗的人数,但是没有人能做到这些,因此可以通过研究一小部分人,从而估计全部美国人观看超级碗比赛的人数。换句话说,尼尔森媒体研究试图通过监控美国人的一个较小样本(sample)的收视习惯,从而了解全部美国人这个总体。需要特别注意的是,尼尔森在大约5 000个家庭中安装了与电视相连的监控设备(叫做“收视记录仪”),生活在这些家庭中的大约13 000人,就是尼尔森所研究的美国人的样本。(为了与人们日益变化的包括有线电视和互联网在内的收视习惯保持一致,尼尔森也增加了安置收视记录仪的人数,但是这5 000个家庭仍然能够反映通常的收视率采集过程。)

尼尔森媒体研究对所选择的5 000个家庭中的人们进行的个体测量结果构成了原始数据(raw data)。尼尔森搜集了很多原始数据,例如,什么时间人们在家庭里收看电视,收看多长时间,在收看什么节目,是谁在家庭里收看电视等。尼尔森将这些原始数据合并成关于样本特征的一组数字,诸如收看每一个单独电视节目的收视人数百分比,或者样本中收看超级碗比赛的全部人数,这些数字就叫做样本统计量(sample statistics)。

定义

- ◆ **样本(sample)**是指总体的一个子集,数据是通过它们进行实际测量而获得的。
- ◆ **对样本进行实际测量或者观察所搜集的结果组成了原始数据(raw data)。**
- ◆ **样本统计量(sample statistics)**是建立在合并或者概括原始数据基础之上的样本的某些特征。

例1-2 失业调查

美国劳工部将公民劳动力定义为那些已经就业或者正在积极寻找工作的人。每个月,美国劳工部都会报告失业率,失业率是指那些正在寻找工作的人占全部公民劳动力的百分比。为了确定失业率,美国劳工部调查了60 000个家庭,在失业率报告中,分别会描述:

- a. 总体
- b. 样本
- c. 原始数据
- d. 样本统计量
- e. 总体参数

解答

- a. **总体**是美国劳工部想掌握的劳动力人口,也就是指构成公民劳动力的全部人口。
- b. **样本**是指60 000个家庭调查中所包括的全部人口。
- c. **原始数据**是指调查中采集的全部信息。
- d. **样本统计量**是由样本的原始数据计算而来的。在本例中,有关的样本统计量是指正在寻找工作的人口所占的百分比。(美国劳工部也计算某些小群体在总体中的百分比,这样类似的样本统计量,诸如青少年、男性、女性、退伍军人失业的百分比。)
- e. **总体参数**是指与样本统计量所对应的全部总体人口特征,在本例中,有关的总体参数是指真正的失业率。(注意,美国劳工部并没有真正地测量总体参数,数据仅仅是从样本中采集而来的,并且以此来估计总体参数。)

相关知识

根据美国劳工部的定义,一些没有参加工作的人并不是失业。例如,回家照顾孩子的爸爸和妈妈并不计入失业的人群之中,除非他们再次寻找工作。那些试图寻找工作但是因为挫折而放弃的人也不计算在失业人群之中。

3. 样本统计量如何与总体参数相关联

假设尼尔森调查发现样本中的5 000个家庭

中 31% 的人观看美式橄榄球超级碗大赛。这个“31%”就是样本统计量，因为它是样本的特征量数。但是尼尔森真正想知道的是相应的总体参数，全美国人观看美式橄榄球超级碗大赛的百分比是多少。

对于尼尔森的调查者来说，想知道确切的总体参数值，是根本不可能的，因为他们的研究只是基于一个样本。当然，尼尔森的调查者希望他们所做的工作保证样本统计量是总体参数的一个很好的估计值。换句话说，样本中美式橄榄球超级碗大赛 31% 的收视率，估计总体参数也大约是 31%。统计学的一个主要目标就是帮助研究者评估这一类研究结论的有效性。

思考题

假设尼尔森调查的结论是美式橄榄球超级碗大赛的收视率是 30%，这表示有多少美国人观看美式橄榄球超级碗大赛（美国的全部人口接近 3 亿）？

统计科学提供了一种方法，使得研究者可以评价以一个样本统计量来估计总体参数的合理性。举例来说，调查结论或者民意测验结果经常涉及误差幅度（margin of error）的概念，样本统计量加减误差幅度后，会得到一个值域，叫做置信区间（confidence interval），其中包含总体参数的可能性很大。在大多数案例中，误差幅度被定义为：这个范围内包含总体参数的可信度为 95%。我们将在第 8 章进一步讨论并明确“可能性”或者“95% 的可信度”的意思，现在你可以通过《纽约时报》（见图 1-1）得到一个大概的解释。

民意测验是如何进行的

上一次的《纽约时报》/CBS 新闻的民意测验是于 10 月 23 ~ 28 日进行的，调查了 1 315 名纽约州的成人。其中 1 026 人说他们登记要投票。测验在说英语和说西班牙语的人中进行。理论上，基于这些样本的测验结果，会比基于全州的调查结果，有 19/20 的概率不超过 3% 的正确率。如果样本再小一点，抽样误差会相应地增大。

图 1-1 调查结果或者民意测验的误差幅度，经常被描述为总体参数所在的一个可能的范围（95% 的可信度，意味着 20 次中有 19 次是可信的），引自《纽约时报》关于 3% 误差幅度的解释

统计学最重要的一个标志性发现是，从较小

的样本中得到有意义的结论是可能的。不管怎样，大样本自然是很好的（如果可行的话），因为通常情况下，样本越大，误差幅度就越小。例如，一个设计良好的民意测验在 95% 的可信度条件下，样本人数为 400 时，误差幅度约是 5%，样本人数是 1 000 的话，误差幅度就成了 3%，如果人数范围超过 10 000 的话，误差幅度就只有 1% 了（参考第 8 章，了解误差幅度如何计算）。

定义

◆ 统计研究中的误差幅度（margin of error）经常表示为可能包含总体参数的值域或者置信区间，以研究得来的样本统计量加减误差幅度后作为置信区间（confidence interval）。也就是说，可能包含总体参数的值域。

从（样本统计量 - 误差幅度）

到（样本统计量 + 误差幅度）

误差幅度经常被定义给定的 95% 置信区间，意思是指，以研究中的样本大小计算出样本统计量，然后据此估计总体参数所在值域范围，其包含真正的总体参数的可能性是 95%（有 5% 的可能性不包括）。

例 1-3 人类登录火星

皮尤人口和压力研究中心以“关于未来的态度”作为问题，调查了 1 546 名美国成年人，问“人类在未来 50 年内是否会登上火星”，1 546 名美国人中的 76% 回答说会或可能会，民意测验的误差是 3%，解释样本统计量 76% 的意思，关于“人们在未来 50 年内是否会登录火星”这个问题，我们怎样才能得到答“是”的人数所占的百分比呢？

解答

总体是全部的美国成年人，样本是所调查的 1 546 名美国成年人，样本统计量 76%，是样本中确实认为人们在未来 50 年内肯定会或可能会登录火星的人数所占百分比，样本统计量 76% 和误差幅度 3% 告诉我们，包含总体参数的值域：

从 $76\% - 3\% = 73\%$

到 $76\% + 3\% = 79\%$

这个值域成立有 95% 的可能性。这就是在此案例中认为人们在未来 50 年内肯定会或可能会登录火星的美国成年人数的百分比。

例 1-4 克隆人？

皮尤研究中心的另外一个调查问题是“人们是否相信人类将在未来的 50 年会被克隆”，51% 的人回答

会或者可能会，同样，误差幅度也是3个百分点，我们是否确信大多数美国成年人认为人类将在未来的50年会被克隆？

解答

不是的。

为了确定全部美国成年人中认为克隆人类将会肯定或者可能发生的人数所占百分比的值域，我们将样本统计量51%分别加上和减去3%的误差幅度。得到的这个值域是从48%~54%，因为这个值域在50%上下浮动，因此我们不能确定大多数美国成年人（指的是超过50%）认为人类将在未来50年内会被克隆。

相关知识

克隆是一种对亲本基因的完全复制。例如，你自身的克隆在遗传上与你完全一样，只不过以一个婴儿出生并且与你相比有着完全不同的经验。1997年成功克隆了第一个成年的哺乳动物。那是伊恩·威尔穆特和他的同事在苏格兰克隆了一只绵羊，名字叫多莉，它没有父亲，所有基因均来自母亲。现在科学家已经成功地克隆了多种哺乳动物。



思考题

在本周新闻中寻找一份关于民意测验的报告。该报告给出了一个误差幅度吗？这个数字意味着什么？

1.1.2 资料汇总：统计研究过程

尼尔森媒体研究的研究过程同样也被应用于许多统计研究中，图1-2以及下段内容概括总结了统计研究的基本步骤。注意这些步骤是有些理想化的，实际操作中每个研究都是会有差别的。当然，这些步骤背后的细节更为重要。例如，在步骤2中一个错误的

样本选取将致使整个研究毫无意义，从较小的样本结果中推论有关总体的结论更是需要加倍谨慎。

统计研究的基本步骤

步骤1：明确表述你的研究目标。确定你想研究的总体，以及明确你了解有关这个总体的什么内容；

步骤2：从总体中选择样本（使用恰当的抽样方法，在下一节中会讨论此问题）；

步骤3：从样本中搜集原始数据，通过计算感兴趣的样本统计量来概括数据；

步骤4：使用样本统计量对总体进行推论；

步骤5：做出统计结论：确定你所了解的是否达到你的目标。

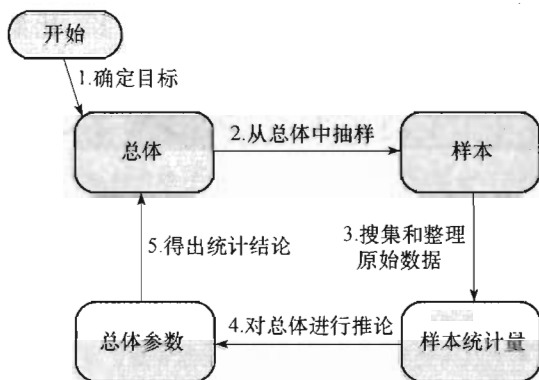


图1-2 统计研究的流程

相关知识

统计学家经常将他们的课程分成两大部分，描述统计是通过图表和样本统计量来描述原始数据，推断统计是指从样本数据中推断（或者估计）总体参数。在本书中，第2~5章主要覆盖了描述统计，第6~10章则关注推断统计。

例1-5 明确步骤

结合在例1-3和例1-4中描述的皮尤研究中心的调查，明确研究者如何在一个统计研究中应用这5个步骤。

解答 步骤应用如下。

1. 研究者的目标在于了解美国人对未来趋势的特定态度。选择美国成年人作为总体而有意不考虑儿童。
2. 研究者选择了1546名美国成年人作为研究样本，虽然并未告知我们样本如何选择，但我们可以假

定通过抽样，可以让这 1 546 名美国成年人代表全部美国成年人这个总体。

3. 通过要求样本中的人们回答一些选择题来搜集原始数据，原始数据是每个人对问题的反应，通过计算诸如每个问题中回答是或否的人数百分比等样本统计量来概括数据。
4. 统计技术使得研究者可以推论总体特征，在这个案例中，推论的方式包括估计特定的总体参数以及计算误差幅度。
5. 确定研究是恰当的，说明了总体参数的估计值，在此基础上研究者得到关于全体美国人对未来趋势的态度的结论。

1.1.3 统计：在不确定中做出决策

尼尔森媒体收视率和我们目前所讨论的大多数案例都采用了问卷调查或民意测验的形式，当然，统计学的研究主题还包括更多，比如，检验一种新的医学治疗方法是否有效的实验设

计；分析全球变暖的危害性；甚至对大学教育质量的评价。总而言之，统计研究的根本目标是，当我们面对很多可能的选择时，统计学帮助我们做出更恰当的决策。

统计学的目标

统计学有很多用途，但它最重要的目标是帮助我们在面对不确定的问题时做出决策。

这个目标在大多数统计研究和本书所讨论的案例中都是清晰的，但有时我们也必须要讨论一点开始看起来很抽象的理论。如果你能谨记统计的总体目标，最后回头来看，你会明白这些理论是怎样帮助我们理解整个世界的。下面的一个案例研究可以给我们一些启示，它包含了几个最重要的理论观点，正是这些观点带来了 20 世纪公众健康最伟大的成就。

案例研究

索尔克脊髓灰质炎疫苗

如果你是一个 20 世纪四五十年代的患者，你最害怕患上的一种疾病就是小儿麻痹症。在小儿麻痹症长期肆虐的每一年里，数以千计的儿童因为这种疾病而导致身体麻痹。1954 年，实施了一个大型实验以检验一种新疫苗的有效性，该疫苗是乔纳斯·索尔克博士（1914—1995）发明的。样本包括 40 万名儿童样本，其中一半儿童注射疫苗，另外一半不注射疫苗只注射生理盐水（生理盐水是安慰剂，见 1.3 节），注射疫苗的儿童仅有 33 个患上小儿麻痹症，而没有注射疫苗的一组有 115 名儿童患上了小儿麻痹症。随后应用统计科学技术进行研究，研究者得出结论说，疫苗对于预防小儿麻痹症是有效的。因此他们决定进一步研究改进疫苗，并将疫苗应用于所有儿童。结果发现，美国和其他发达国家开始将注射疫苗作为一种常规防治手段，这样一来，令人惊恐的小儿麻痹症大规模爆发就已成为历史。

相关知识

在索尔克脊髓灰质炎疫苗成功推广之后，脊髓灰质炎在美国就变得很少。但是在一些不发达国家还是很常见的。对全球的儿童注射疫苗以预防脊髓灰质炎的努力从 1998 年开始，现在虽然还没有达成完全根除脊髓灰质炎的目标，但已经是非常成功了。



1.1 练习题

统计素养和批判性思维

1. 总体和样本 总体和样本分别是什么？它们之间有什么不同？
2. 统计和统计学 假定一个人进行棒球统计而另一个人统计使用一种特定药物进行治疗是否有效，两人使用的术语“统计”是否是同一个意思？如果不是的话，区别是什么？
3. 统计量和参数 样本统计量和总体参数分别是什么？它们之间有什么不同？
4. 误差幅度 统计研究中的误差幅度是什么？为什么它很重要？

判断正误：判断练习题 5~10 中的陈述是

有意义的(表述正确)还是无意义的(表述错误),并加以解释。并非所有的陈述都有明确的答案,所以你的解释比判断更重要。

5. 统计量和参数 教授设计了一个不能够计算任何样本统计量的统计研究,但是他成功地确定了总体参数,并且误差幅度很小。
6. 可怜的民意测验 选举两周前进行的民意测验显示史密斯将获得70%的选票,误差幅度为3%,结果他却输掉了选举。
7. 民意测验的确定性 毫无疑问约翰逊将赢得选举,民意测验显示他将获得54%的选票,而误差幅度仅为3%。
8. 对抗尼尔森 一个新开业的公司试图与尼尔森媒体研究竞争,同样的价格他们将提供具有更大误差幅度的数据。
9. 压力样本 研究目标是了解遭受了家庭灾难的人们的压力。所以研究者计划把过去一个月里生过病的人作为总体,从其中选择一个样本。
10. 新产品 市场研究部门调查了1000个顾客关于我们新产品的态度,由于样本中的人们非常热切地希望购买我们的新产品,因此我们决定在全国展开广告宣传活动。

概念与应用

样本、总体、统计量和参数: 练习题11~14的每道题都描述了一种统计研究,请确定每个案例中的总体、样本、样本统计量和总体参数。

11. 干细胞研究 普林斯顿国际调查研究协会设计的发布在《每周新闻》上的民意测验中,调查了1002名成年美国人支持还是反对“联邦税收资助取自人类胚胎的干细胞医学研究项目”,结果显示48%支持。
12. 黑色星期五 盖洛普公司对1236名成年人进行民意测验,9%的人说他们的坏运气产生于黑色星期五。
13. 银河系的距离 天文学家通过测量银河其中几个星球的距离并取它们的平均数来推断银河系的距离。
14. 抗过敏药物 内舒拿(Nasonex)是用于治疗过敏症的药物,在内舒拿药物实验中,374名3~11岁的儿童服用100微克内舒拿,其中17%会出现头疼的副作用。

确定值域: 练习题15~18中,用给出的样本统计量和误差幅度确定最可能包含总体参数的值域(置信区间)。

15. 干细胞研究 普林斯顿国际调查研究协会设计的发布在《每周新闻》上的民意测验中,调查了1002名成年美国人支持还是反对“联邦税收资助取自人类胚胎的干细胞医学研究项目”,基于调查结果,估计48%的成年人持支持的态度,误差幅度是3个百分点。
16. 北极国家野生动物保护区的油田 在哥伦比亚广播公司新闻部和《纽约时报》的民意测验中,调查1241名成年人是否赞成开采北极国家野生动物保护区的油田。基于调查结果,估计45%的人不赞成,误差幅度是3个百分点。
17. 体质指数 随机选择40岁的美国成年人测量他们的体质指数(BMI)。基于样本的结果,估计BMI平均数为26.0,误差幅度为3.4。
18. 体温 研究者随机选择106名成年人并且测量了他们的体温,估计平均体温为98.2°F,误差幅度为0.1°F。
19. 解释民意测验结果 州参议员只有两人参与竞选,在选举前进行了一个民意测验。结果显示58%的选举者支持共和党的候选人,误差幅度是3个百分点。共和党会获胜吗,为什么?
20. 解释味觉测验结果 在一个联邦实况转播的味觉测验中,抽取了一个经常饮用百威公司啤酒的人群样本,采用盲试方法给予两种品牌 Michelob 和 Schlitz 的啤酒。基于调查结果,我们估计52%的被试喜欢 Michelob,误差幅度是10%。我们可否得出结论说,饮用百威公司啤酒的人,在两种品牌 Michelob 和 Schlitz 的啤酒中更喜欢 Michelob,为什么?
21. 人们投票会说谎吗? 在某个1002人的调查中,701个(或70%)调查者说他们将参与投票,调查的误差幅度是3个百分点。然而,实际的投票结果是只有61%的合格选民实施了投票。这是否意味着他们在调查中说谎了呢?请加以解释。
22. 为什么有差异? 伊格顿研究所的一个民意测验,调查男性是否同意下面的陈述“堕胎是私人事务,应该让妇女自己决定,政府不应该干预”。那些由女性调查者进行访问所调查的男性,77%同意上面表述,那些由男性调查者进行访问所调查的男性,70%同意上面陈述。假定这个差异是显著

的，这个差异如何解释？

解释下述真实研究：对于练习题 23 ~ 26，完成如下任务。

- a. 基于给定信息，表述你所想到的研究目标。
确定一个可能的总体以及感兴趣的总体参数。
 - b. 简略描述研究中的样本、原始数据、样本统计量。
 - c. 基于样本统计量和误差幅度，确定你所感兴趣的总体参数可能的所在值域。
23. 改行 光辉国际咨询顾问公司调查了 1 733 名行政主管，发现 51% 的人说如果能够再次择业的话，他们将会改行。误差幅度为 3 个百分点。
 24. 超能力 在一项对 1 018 个成年人的调查中，结果发现 11% 的人宁愿选择能够隐形这样的超能力，误差幅度为 3.2 个百分点。
 25. 失业率 基于最近的一个关于 6 万个家庭的成年人调查，美国劳工部报告失业率为 4.6%，误差幅度为 0.2 个百分点。
 26. 未雨绸缪 洛佩尔调查公司调查了 2 000 个美国成人，64% 的被调查者采用定期账户存钱，误差幅度是 2.0 个百分点。

统计研究的五步骤：在练习题 27 ~ 30 中，陈述你在一项统计研究中是怎么应用五个基本环节的。

27. 手机和开车 你想确定，在过去一周里，至少有一次在开车时打手机的人数在有驾照的人们中所占的比例。
28. 信用评分 消费者信用评分是用来量化消费者信用水平的常规指标，你想知道所有美国成年人的信用评分的平均水平。
29. 乘客体重 认识到民航超重会影响飞行安全，你想知道民航乘客的平均体重。
30. 心脏起搏器电池 因为电池用于心脏起搏器，非常之重要，你想知道在故障发生前电池的平均使用寿命。

网络和远程项目

请在网址 www.aw.com/bbt 上打开第 1 章的相关链接“Links for Internet Projects”。

31. 近期的尼尔森收视率 过去一周的尼尔森收视率，三个最流行的电视节目是什么？解释每个电视节目的收视率和收视份额的意思。
32. 尼尔森方法 尼尔森媒体研究连续修订了数据选择方法的细节，浏览中心的网站并且

分析其最近电视节目评级方法，采用标准化的项目符号列表形式总结这些方法。

33. 航线比较 美国交通部例行公事，发布了不同航空公司的航班正点率，行李遗失率以及其他统计指标。寻找近期的一个统计指标的例子，基于你的发现，说某一个特定的航空公司比起其他航空公司是出奇的好或者特别不好是公平的吗？
34. 劳动统计 浏览劳动统计局的网页寻找过去 12 个月的每月失业率，如果你假定每月的调查误差幅度大约为 0.2 个百分点，在过去的一年里失业率有什么引人注目的变化吗？请说明。
35. 统计学和安全 选择一个已经完成的（或者将要完成的）提高司机和乘客安全的研究。简要说明统计对这个研究的重要性。
36. 皮尤研究中心 皮尤人口和压力研究中心研究了压力、政治和公众政策问题。浏览它的网页寻找最新的态度调查。选定某一个近期的调查，撰写一个关于这个调查的总结，主要包括研究的是什么，如何实施的调查，结果发现什么等内容。

新闻中的统计学

37. 新闻中的统计 选择过去 1 周中涉及统计学的 3 个故事，在每个例子中，简要写出一段陈述，说明统计在这个故事中的作用。
38. 统计学在你的主修专业中的应用 写一个简要的介绍，说明在你的主修专业领域中如何应用统计学（如果你还没有选定一个主修专业，从你想选择的专业角度来回答这个问题）。
39. 统计和娱乐 尼尔森媒体收视率对于电视收看的影响是众所周知的，选择另一个经常用于娱乐产业的统计方法，简要地说明统计在这个应用中的作用。
40. 运动中的统计 选择一个运动项目，描述至少三种不同的通过参赛者或观众进行追踪的常用统计。在每一个例子中，简要地说明统计在其中的重要性。
41. 经济统计 政府会定期发布许多不同的经济统计，诸如失业率、通货膨胀率、联邦财政的盈余或赤字。研究近期的报纸，选择 5 个重要的经济统计，简要地解释这 5 种统计中每一个的目标。



1.2 样本

知道总体参数真值的唯一方法就是观测总体中的每个成员。例如，想知道你所在学校全体学生的平均体重，你需要测量每个学生的体重。搜集总体每个成员的数据叫做普查（census），遗憾的是，建立一个普查经常是不可操作的。在某些情况下，总体往往实在太以致搜集每个成员的数据是非常费时费力的；在另外一些情况下，普查是不可行的，主要是由于它与研究的目标发生冲突。例如，设计检验销售前的棒棒糖质量，这个试验是不能普查的，因为那样的话意味着需要检验每一个棒棒糖，那将没有完整的棒棒糖可以出售了。

大多数统计研究可以不必费力地建立一个普查，作为搜集总体中每个成员数据的替代，我们可以从总体中选择一个样本并且使用样本统计量来推断总体参数，当然，至少就研究的特点而言，只有样本能够很好地代表总体时，这个推断才是合理的。也就是说，我们要寻找总体的代表性样本（representative sample）。

定义

◆ **代表性样本（representative sample）**是指样本组成成员的相关特征与总体特征是基本是一致的，这样的样本称为代表性样本。

例 1-6 身高的代表性样本

假定你想知道你所在学校全部学生的平均身高，下面哪一个更符合你研究的代表性样本的条件呢？是男子篮球队队员还是你所上统计课程的班级同学？

解答 男子篮球队队员不是你所研究群体的身高的代表性样本，因为篮球队只包括男生并且篮球队员的身高比一般水平更高些。你所在统计课程的班更接近所有学生的平均水平，所以统计课程班级里的学生比篮球队员更符合代表性样本的要求。 ■

1.2.1 偏差

设想一下，电视收视率样本中的 5 000 个家庭，如果尼尔森仅仅选择常常上夜班的工薪家庭，因为夜班工人不在家，看不了午夜段的电

视，尼尔森将会发现午夜时段的电视节目在样本里的家庭中是不受欢迎的。很明显，这个样本不是所有美国家庭的代表性样本，如果据此得出结论说午夜时段的电视节目不受欢迎是错误的，我们说这样的一个样本是有偏差的，因为样本中的家庭不同于“典型”的美国家庭。实际上，尼尔森会小心谨慎地避免在样本选择中出现这样的偏差。通常，术语**偏差（bias）**是指设计或建构一个统计研究时由于倾向某一特定结论而出现的任何问题。我们不能信任一个有偏差的研究结论。

偏差产生的途径有很多种，例如：

(1) **样本选取的偏差**，即样本中的成员在某个方面与总体中的一般成员有所不同。这种情况下，研究的结论将更多地反映样本的不寻常的特征而非总体的真实特征。

(2) **研究者偏差**。即一个研究者个人如果倾向于某一定特定结果就会产生偏差。在这种情况下，研究者会有意无意地歪曲研究数据的意义。

(3) **数据组自身也会产生偏差**，如果在数据收集时有意无意地选择了并不代表总体的那些数据。

(4) 即使研究很成功，它也可能在报告时存在偏差，例如，一个应用图表表述的数据可能仅仅告诉我们事实的一部分或者歪曲的数据（见 3.4 节）。

相关知识

许多医学研究采用实验设计，以检验一种新药是否有效。根据发表在《美国医学会杂志》上的文章来看，作者发现研究是呈阳性的结果（药物是有效的）要比研究是呈阴性的结果（药物是无效的）更容易发表。总体来讲，这种“出版物偏差”会使得一种新药看起来会比实际上更有效。



避免偏差是统计研究中的一个巨大挑战，寻找偏差是评估一项统计研究或媒体报道中的统计研究最重要的环节。

例 1-7 为什么采用尼尔森的结果

尼尔森媒体研究通过向电视台和网站出售它的服务获得收入，例如，NBC 支付尼尔森费用以获得电视节目的收视率。为什么 NBC 自己不调查它的收视率而是交给尼尔森这样的公司？

解答 电视节目的广告费用高低依赖于它的收视率，收视率越高，广告的入网费就越高。这意味着 NBC 如果对自己的收视率进行调查将会有个明显的偏差。广告客户也不会信任 NBC 自己进行调查的结果。通过雇用像尼尔森这样的独立调查机构，NBC 可以提供让广告客户更可能信任的收视率信息。 ■

思考题

NBC 付费给尼尔森的服务，这个事实可能看起来给了尼尔森一个财务激励，使得尼尔森的调查结果会让 NBC 的收视率看起来更高些。尼尔森也提供其他网络公司的收视率，这个事实怎样有助于减少 NBC 收视率的偏差？

1.2.2 抽样方法

一个好的统计研究必须具有一个代表性样本，否则样本是有偏差的，结论也是不可信的，让我们来学习以下几种抽样方法，至少理论上可以抽出代表性样本。

1. 简单随机抽样

大多数情况下，获得代表性样本的方法是总体中随机抽样。随机抽样（random sample）是指总体中的每一个成员都有被选入样本的相同机会。例如，你可以让总体中的每一个成员掷骰子，然后选择掷 6 点的人，通过这样的方式获得一个随机样本。反过来，如果你选择每个高于 6 英尺^①的人进入样本就不是随机的，因为每个人被选择的机会不是相同的。

统计学上，我们经常预先决定需要的样本容量。在简单随机抽样（simple random sampling）中，具有特定样本容量的每个可能样本都有同等的被选中的机会。例如，从你所在学

校的全部学生中选择一个包括 100 个学生的随机样本，你应该给学校的所有学生全部编上号，然后抽签从中抽出 100 人。每个学生的编号是确定和唯一的，由 100 个学生组成的各种可能的样本被选中的机会是相同的，为了使抽签效率更高，可以借助电脑或者随机数码发生器。

思考题

在计算器上寻找随机数码键（带统计功能的计算器上有此键）。你按下它会怎么样？你如何用随机数码键抽取 100 个学生样本？

因为随机抽样对给定样本量的每一个样本被选中的机会完全相同，如果样本量足够大的话，它更有可能成为总体的一个代表性样本。简单随机抽样的样本容量越大就越是总体很好的代表性样本。

例 1-8 电话号码本抽样

你想以某个城镇的所有居民为总体进行一个民意调查，你能否通过从当地电话号码本上选取名字的方式抽取一个随机样本。

解答 从电话号码本选取的一个样本不是小镇全体居民的随机样本，因为电话本常常会遗漏好多人，那些遗漏的人是没有机会被选中的。例如，当两个或两个以上的人共用一部电话时有些人名就不列在电话本上，而有些根本没有固定电话而使用手机，或者有些人（如无家可归者）根本就没有电话。 ■

2. 系统抽样

简单随机抽样效果良好，但在许多情况下我们可以通过更简单的方法得到同样的结果。假定检验英特尔公司生产的微机芯片，芯片从流水线上生产出来后，你可以决定检验每个第 50 块芯片，这应该是一个代表性样本，因为没有理由说明每个第 50 块芯片与其他芯片有什么特殊的不同之处，这种形式的抽样，诸如从总体中抽取每一位第 50 个成员，这称为系统抽样（systematic sampling）。

例 1-9 博物馆测量

自然航空航天博物馆想检验举办一个新的太阳系展览的创意是否可行，博物馆工作人员采用系统抽样的方法调查一个参观者的样本。他每隔 15 分钟就访

① 1 英尺 = 0.3048 米。

问1个参观者,选择对象是在那个时刻刚好进入正在举办的太阳系展览入口的参观者。相较于简单随机抽样,他更愿意选择系统抽样,你认为这是为什么呢?在本案例中,系统抽样会产生一个代表性样本吗?

解答 如果采用简单随机取样的话,可能偶尔会发生所抽取的两个参观者先后时间间隔很短,以至于工作人员没有时间去和他们交流,系统抽样每间隔15分钟选取1个参观者作为调查对象避免了上述问题的发生。因为没有理由说,某个特定时间进入展览入口的人们会比早些或晚些时有什么差异,所以这个时间间隔下的抽样是可以选为一个参观者总体的代表性样本的。 ■

相关知识

例1-9中所描述的抽样是太阳系模型建设时的前期调研,一个永久的太阳系模型建成在美国国家广场,从航空航天博物馆一直延伸到史密森学会。下面的照片显示了展览的设计者(也是本书的一位作者)抱着她儿子触摸太阳系模型。



例1-10 什么情况下系统抽样会失败

你准备建立一个男女同校的住校学生调查,在这所学校里,男生被安排在编号单数的宿舍,女生被安排在编号为双数的宿舍。你能否通过选择每个第10间宿舍获得1个代表性样本。

解答 不可以。如果你从一个编号为单数的宿舍开始,每隔10个宿舍依然是单数(比如选中房间号为3,13,23,...)。同样地,如果你从1个编号双数的宿舍开始,每隔10个宿舍依然是双数。这样你不可能获得一个男生女生都包括的样本,它也不是男女同校学生的代表性样本。 ■

3. 便利抽样

系统抽样比简单随机取样容易些,但是

在很多情况下是不可操作的。例如,假定你想知道你所在学校左利手学生的比率。你可能要花很大力气去进行一个简单随机取样或者系统抽样。因为这两种抽样都需要知道全校的所有学生。相反,如果你所在统计班的学生作为样本将是很容易的,你只需要让左利手的同学举手就可以了。这种形式的抽样方法称为便利抽样(convenience samples),因为它抽样非常方便,并不需要一个复杂的过程。为了确定左利手同学所占的比率,你统计课程班里的便利样本是不错的,也没有理由证明你统计班里左利手人的比率与其他人群相比有什么差异。但是如果你想了解的是各个专业的学生比率,样本将有可能是偏差的,因为有些专业需要学习统计课程,而有些专业是不需要学习统计课程的。通常,便利样本比起其他形式的抽样来,可能会存在出现偏差的更多可能性。

例1-11 沙拉口味测试

一家超市想看看是否应该引进一种新品牌的沙拉,所以它在一家店面营业闲暇时提供免费品尝,然后询问人们怎么评价这款新沙拉。采取哪一种抽样方式较好?该样本有可能是所有顾客这个总体的代表性样本吗?

解答 停下来免费品尝沙拉这样的顾客样本是个便利样本,因为这些顾客是刚好来到这家店面也希望尝试一种新产品的人群。(这种类型的便利样本,是他们自己选择成为或者不成为样本的一部分,所以也叫做自我选择样本,我们将在1.4节中进一步介绍自我选择样本。)这个样本不太可能是所有顾客的代表性样本,因为不同的顾客会选择在不同的时间进行购物(例如,居家的父母比那些正在工作的父母更可能在中午购物),并且仅仅是喜欢沙拉的人们会参与品尝。当然,这个数据仍然是有用的,因为喜欢沙拉的顾客的态度在这个案例中可能是最重要的。 ■

4. 分组抽样

分组抽样(cluster sampling)是从所有成员的随机选择转变成随机选择组或者群,假定你在农业部工作并且想了解使用有机农业技术的农场主的百分比。如果采用简单随机抽样或者系统抽样将会很困难且花费巨大,因为需要参观大量相距很远的个体农场。某个县农场主的便利样本会有偏差因为不同地区的农业生产方式会有很大差异。你可以决定从全美随机选

择12个县并且调查这些县里的所有农场主。我们知道每个县都会包括一群农场主，选择的样本包括随机抽取的这些县里的每一个农场主。

例 1-12 汽油价格

你想知道距离飞机场的出租车站附近1英里内的加油站汽油的平均价格。说明你在本案例中可能使用的分组抽样。

解答 你可以随机抽取全国范围内的几个机场，在抽中的几个机场，你可以调查距离飞机场出租车站附近1英里内的所有加油站。

5. 分层抽样

假定你建立一个民意测验来预测下届美国总统选举。研究的总体是所有可能的投票者，你可以从总体中随机选择一个样本。美国总统由各个州选举人团选举产生，只有了解每个州的选民偏好才能更好地预测总统选举结果。你最终的样本应该包括50个州中每个州各自的随机样本。在统计技术中，50个州代表了子群体，或者叫层(strata)。因为你的最终样本包括了每层中随机选择的成员，这就叫做分层抽样(stratified sampling)。

相关知识

由美国国会授权，实际上投票选举美国总统的是叫做“总统选举人”的一小部分人。每个州可以选择一些总统选举人作为它的国会成员(包括参议员和众议员)。当你投票选举总统时，实际上是给所在州的总统选举人投票，这些总统选举人对选民承诺投票给某一特定的总统候选人，这些总统选举人在普选之后数周之内进行他们的投票。



例 1-13 失业率数据

美国劳工部每月调查60 000个家庭完成失业率报告(见1.1节中的例1-2)。为了选择这些家庭，美国劳工部首先将所有的县和城市分成2 000个地区，然后在这些不同地区随机选择家庭进行调查。这是分层抽样吗？层是什么？在这个案例中为什么分层抽样很重要？

解答 失业率调查是分层抽样的一个例子，因为它首先将总体分成了层。层，指的是2 000个地区的人们，在本例中分层抽样是重要的，因为不同地区的失业率是有差异的。例如，堪萨斯州农村与硅谷的失业率是非常不一样的，通过采用分层抽样，美国劳工部可以保证样本能公平地代表所有地区。

1.2.3 抽样方法总结

下述文字和图1-3概括了我们讨论的5种抽样方法。没有一种单独的方法是“最好的”，每种方法都有自己的用途(有些研究甚至采用两种或者多种抽样方法)，但是不管采用哪种抽样方法，以下三个关键点都应该牢记在心。

- 只有当样本对总体具有好的代表性时，一个研究才可能成功。
- 一个有偏差的样本不太可能是总体的代表性样本。
- 甚至一个精选的样本也可能被证明是不具有代表性的，可能只是因为抽样时的糟糕运气。

通常的抽样方法

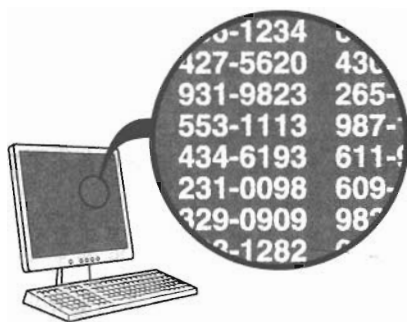
(1) 简单随机抽样：我们以这样的方式来抽样，那就是每个具有相同样本量的可能样本被选取的机会是相等的。

(2) 系统抽样：我们采用一个简单的系统去选取样本，比如总体中每第10个或者每第50个成员均被抽选出来。

(3) 便利抽样：我们选取一个样本因为这个样本刚好很方便获得。

(4) 分组抽样：我们先将总体分成不同的组，再随机地选择其中的一些组，被选中的组的所有成员构成样本。

(5) 分层抽样：我们使用这种方法是因为总体中的不同层或者群是有差异的，我们首先定义层，然后在每个层中随机抽取样本，总的样本包括所有从每个单独的层中抽取出来的样本。



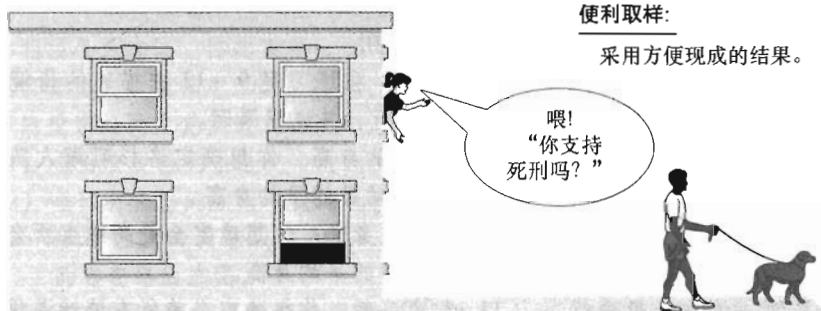
简单随机抽样:

每个同样容量的样本都有相同的被抽取的机会,经常采用计算机生成随机数字的方式来进行随机抽样。



系统抽样:

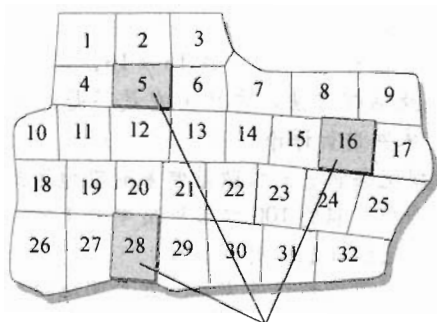
依次选取每一位第K个成员。



便利取样:

采用方便现成的结果。

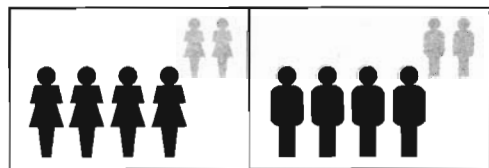
在选择卡尔森市的行政区



在选中的地区选取所有投票

分组抽样:

将总体分成若干组,然后随机选取其中的一些组,被选中的组里的所有成员构成了样本。



分层抽样:

将总体区分为至少两层,然后从每一层中再进行抽样。

图 1-3 常用抽样方法

例 1-14 抽样方法

确定下面每个案例的抽样方法。

a. 一个果园苹果收获时共采摘了 1 200 篮。1 位农业检验员随机选取了 25 篮并检查其中的每个苹果是否有虫害。

b. 一个教育研究者想知道在某个大学，班级里的男生还是女生倾向于问更多的问题。在这所大学的 10 000 名学生中，她随机选取 50 名男生和 50 名女生进行了访谈。

c. 为了了解行星系是不是常见的，天文学家建构一个调查，在 100 颗距离最近的星球中寻找行星。

d. 为了确定谁将获得球星亲笔签名的足球，在

观众爆满的体育场里，一个计算机程序随机地抽取了 11 位观众的门票号码。

解答

a. 农业检验员用的是分组抽样，因为他随机地选取了几组（篮子）并且检验选取的每组中的所有苹果。

b. 男生和女生分别代表了研究中两个不同的层，这是一个分层抽样的例子。

c. 天文学家关注离地球较近的星球因为它们容易研究，这是一个便利抽样的例子。

d. 因为电脑随机地选择了 11 个门票号码，每个门票号码都有相同的被选取的机会，这是一个简单随机取样的例子。

1.2 练习题

统计应用和批判性思维

1. 普查和抽样 什么是普查，什么是抽样，它们之间有什么不同？
2. 偏差样本 大选前的民意测验，从一个民主党注册的名册下随机抽取一个样本。这种抽样方法有什么错误？
3. 分组和分层抽样 分组和分层抽样都是从总体的子群体中选取被试，这两种抽样的不同在哪里？
4. 学生样本 一次，本书的一位作者想进行一个研究统计他班里的右利手学生的比率。目标是得到大学生中右利手比率的结论。采用的是哪一种抽样方法？样本可能会产生偏差吗，为什么？

判断正误：判断练习题 5~10 中的陈述是有意义的（表述正确）还是无意义的（表述错误），并解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. 毕业年龄 为了完成一个统计课的项目，我实施了一个普查以调查大学生获得学士学位的平均年龄。
6. 便利样本 为了完成一个统计课的项目，我选取了一个便利样本，但结果仍然是有意义的。
7. 偏差样本 某项研究一定会有偏差，因为它的样本包含了 75% 超过 6 英尺高的美国人。
8. 死囚牢房 现在有 3 366 名犯人被关押在死囚牢房（基于司法援助局 2007 年统计数据），我们将他们的名单进行编号，然后通过电脑从 1~3 366 这些数字中随机抽取 20 个

号码，将与号码相对应的犯人抽取出来组成一个简单随机样本。

概念与应用

普查：在练习题 9~12 中描述的情境下普查是否可行，并说明原因。

9. 湖人队的身高 你想确定洛杉矶湖人队所有篮球运动员的平均身高。
10. 高中生身高 你想确定全美高中生所有篮球队队员的平均身高。
11. 智商分数 你想确定全美所有统计课程教师的平均智商分数。
12. 教师年龄 你想确定科罗拉多州立大学所有统计教师的平均年龄。

代表性样本？练习题 13~16，识别样本、总体以及抽样方法。你认为它是不是总体的代表性样本并进行评论。

13. 参议院条款 一位政治学家随机地选择了当前国会任职的 100 位参议员中的 4 位，计算他们已经任职的时间长度。
14. 美国橄榄球超级碗大赛 在美国橄榄球超级碗大赛期间，尼尔森媒体研究中心随机抽取了 5 108 个家庭进行了一项调查，结果发现 44% 的家庭收看橄榄球超级碗大赛。
15. 克隆 盖洛普公司随机调查了 1 012 名美国成年人，89% 的人说不应该允许克隆人类。
16. 邮件调查 一位纽波特大学的大学生实施了一个关于美国成人研究项目，她给所认识的 500 个人发送了调查邮件，她请他们回复邮件回答问题“你更喜欢使用电子邮件还是传统邮件（美国邮政服务）”，她收到 65 件

回复,其中42件说更喜欢传统邮件。

评估样本选择: 练习题17和18均描述了研究目标,并给你提供了可能的样本。在每个案例中,确定哪个样本最可能是代表性样本,并解释为什么其他选择不是研究的代表性样本。

17. 信用卡债务 你想确定佛罗里达州成年消费者信用卡债务的平均数值。

(1) 佛罗里达州自己拥有路虎车并注册过的驾驶者。

(2) 在劳德代尔堡市电话号码本上排列在前1000名的佛罗里达州居民。

(3) 在一个很全的佛罗里达州电话号码本上排列在前1000名的佛罗里达州居民。

(4) 那些回复《迈阿密先驱报》上发布的一个调查问卷的佛罗里达州居民。

18. 加利福尼亚选民 你想实施一个调查,以确定有资格的加利福尼亚选民在下次选举中可能投票给民主党候选人的比率。

(1) 全部有资格的圣迭戈市的选民。

(2) 全部有资格的索诺玛县的选民。

(3) 所有回复美国在线(AOL)网络调查的有资格的选民。

(4) 在全部的加利福尼亚选民名单上每隔1000名抽取一名。

偏差: 练习题19~22描述的情境中存在偏差吗?说明原因。

19. 电影评论家 一位为ABC新闻撰稿的电影评论家发表了她关于迪斯尼最近一部电影的观点,迪斯尼公司刚好控股ABC公司。

20. 汽车评论 《消费者报道》杂志,不接受任何免费产品和广告,发布了新款汽车评论。

21. 转基因大豆 孟山都公司雇用了独立的大学科学家来确定它的新转基因大豆对环境造成的任何危害。

22. 药物研究基金 《美国医学会杂志》发表了一篇文章评估某种药物,发表该文章的一些生理学家得到了生产此种药物的医药公司的基金支持。

抽样方法: 练习题23~38,识别每一个抽样分别采用简单随机抽样、系统抽样、便利抽样、分组抽样和分层抽样的哪一种,在每一个案例中,你认为该抽样方法得到是一个代表性

样本还是有偏差样本?请说明原因。

23. 临床试验 在对一种用以提高红细胞数量的新药进行试验的第二阶段,研究者拿出一堆写有每个实验被试的名字和住址的信封。她想由12个被试组成的子样本中增加剂量,因此她在一个箱子里彻底打乱了所有信封,然后抽出12个信封以确定哪个被试将会给予增加剂量。

24. 酒后驾驶检查人员 一位作家观察到,一位检查酒后驾驶的警察总是每隔5个驾驶者就让他停车受检。

25. 投票后民调 总统选举投票当天,新媒体随机选择某些投票站组织了一次投票后民调,在选中的投票站所有投票人离开这一区域时都被调查。

26. 教育和运动 一位斯伯丁运动装备公司的研究人员研究受教育水平和积极参加运动的关系。他随机选择了40个高尔夫球手、40位网球运动员和40位游泳选手进行了一个调查。

27. 人类工程学 一位工程学的学生测量了他的家庭成员经常用来按键的手指长度。

28. 逃税 一位美国国内收入署的研究者,通过对20个随机选择的饭店的所有男女服务员进行问卷调查,调查了税收申报时偷漏税的情况。

29. MTV调查 一位关于MTV的市场专家计划对500人进行问卷调查,从10~19岁、20~29岁等依次类推的这些年龄组中随机抽取。

30. 信用卡数据 本书的一位作者以他所有学生为样本,调查了他们所持有的信用卡张数。

31. 筹集资金 纽波特大学的资金筹集人,按字母排列顺序的校友名单每间隔100个抽取1个进行电话营销活动。

32. 电话民意测验 盖洛普公司对1059名成年人通过电话进行民意测验,被访问者是通过电脑随机抽取其电话号码而选中的。

33. 市场调查研究 一位市场调研员将加利福尼亚居民划分为失业者、全职工作者和兼职工作者三类,她从每一类中选取了50名进行调查。

34. 学生饮酒者 受一个酗酒而死亡学生的影响,纽波特大学组织了一项关于学生饮酒情

况的研究，随机选取 10 个不同的班级并且调查每个班级所有学生饮酒的情况。

35. 杂志问卷调查 《人物》杂志以“穿着最佳名人”为题发布了一个问卷调查，并收集了那些回复了邮件的读者的反馈。
36. 心脏移植 一位约翰霍普金斯大学的医学研究者得到了一个列有所有等待心脏移植手术的患者名单，用电脑随机生成 50 个数字，然后选择了与这 50 个数字相对应的患者。
37. 质量控制 通过电脑随机生成给每一张 CD 分配一个 1~1 000 之间的数字，然后选取得到 1 000 这个数字的 CD 作为产品样本。
38. 安全带 每间隔 500 个就抽取 1 条安全带通过强力拉拽进行破坏性检验。

选择一个抽样方法：练习题 39~42，提供一种更可能得到代表性样本的抽样方法，解释你为什么采用这种方法而不是用其他方法。

39. 学生会主席选举 你想预测即将来临的学生会主席选举的获胜者。
40. 血型 你想知道全美每种主要血型（A，B，AB，O）的人口比例。
41. 心脏病死亡 你想确定每年死于心脏病的人数比率。
42. 金枪鱼体内汞含量 你想确定美国居民消费的金枪鱼内的平均汞含量。

网络和远程项目

请在网址 www.aw.com/bbt 上打开第 1 章的相关链接“Links for Internet Projects”。

43. 公众态度民意测验 利用调查机构可利用的网络资源，比如盖洛普、哈里斯、皮尤、扬科洛维奇，回答下述问题。

(1) 怎么选择样本被试是恰当的？

(2) 基于你已经学习的知识，你认为民意测验的结果是可信的吗，为什么？

44. 失业样本 利用劳动统计局的网页确定在每月报告中它是如何选取家庭样本的细节的？撰写一个关于这个过程简短的摘要，它为什么更可能产生的是一个代表性样本？
45. 选择性投票 美国奥斯卡电影金像奖、海斯曼奖和《纽约时报》畅销书榜正好代表了获奖者是通过特定的被选择的个体投票的结果而定的。挑选其中一个选举过程，说明由谁来投票以及这些人是怎么被选中的。分析这个过程偏差来源。

新闻中的统计学

46. 新闻报道中的抽样 寻找一则你感兴趣的有关统计研究的近期新闻报道，撰写一篇简短的关于研究样本是怎样被选取的摘要，简要地讨论在这个研究中，样本是否是总体的代表性样本。
47. 民意测验样本 寻找一则近期新闻机构（比如盖洛普、哈里斯、《今日美国》、《纽约时报》或者 CNN）发布的民意测验报道，简要描述该样本及其选取过程，样本选取的方式有可能出现偏差吗？说明原因。
48. 政治组织民意测验 寻找一则近期政治组织（比如共和党或民主党、试图影响国会某些特定决策的组织）实施的民意测验结果，主要讨论样本是什么样的以及如何被选取的，样本选取的方式有可能出现偏差吗？政治组织是否比新闻机构实施的民意测验更有可能出现偏差？说明原因。

免影响或改变我们所要观察的属性。尼尔森收视率就是一个很好的观察研究的案例，因为尼尔森使用收视记录仪仅仅观察被试正在观看的电视节目，并不干预他们看什么。

注意观察研究可能包括的活动超越了我们一般意义上所定义的观察，测量人们的体重需要和人们互动，比如要求他们站在一个体量测量仪所对应那块位置，但是在统计研究中，这些测量是一种观察，因为这种互动并不改变人们的体重，同样，研究者与被试深入交流的民意调查也被看做一种观察研究，因为研究者仅仅是想了解人们的态度，而非改变。

1.3 统计研究的类型

统计研究有多种不同的类型，在所有案例中，人、动物（或者其他生物）或者被选中的其他物体叫做研究被试（subjects），如果被试是人，通常将他们看做研究的参与者（participants）。

统计研究有两种基本类型：观察研究和实验研究。在观察研究（observational study）中，我们观察或测量某种特殊属性，但是会尽力避

相反,比如用一个医学研究设计检验大剂量的维生素 C 是否能预防感冒。实施这个研究,研究者会要求样本中的一部分人每天服下大剂量的维生素 C。这种类型的统计研究叫做**实验(experiment)**,一个实验的目标就是研究某些**实验处理(treatment)**的效果,在这个案例中,实验处理是指服用大剂量的维生素 C。

两种基本类型的统计研究

(1) 在一个**观察研究**中,研究者观察或测量被试的某种属性或特征,但是并不去影响或改变这种属性或特征。

(2) 在一个**实验**中,研究者对被试实施一些实验处理然后观察这些处理是否产生某些效果。

例 1-15 研究类型

识别下面的研究是观察研究还是实验研究。

(1) 索尔克脊髓灰质炎疫苗的研究(见 1.1 节的研究案例)。

(2) 一个民意测验调查人们下一次选举中计划投票给谁。

解答

(1) 索尔克脊髓灰质炎疫苗研究是一个实验研究,因为研究者实施了实验处理。在这个案例中,处理是指注射疫苗是否降低了脊髓灰质炎的发病率。

(2) 民意测验是一个观察研究,因为它只是想了解人们的投票倾向而不是要影响人们的投票。 ■

1.3.1 变量识别

无论是观察研究还是实验研究,通常我们将试图测量的东西称为**关注变量(variables of interest)**。术语变量是指一个取值是变化的项目或者属性。关注变量是我们想要了解的变量。例如,尼尔森电视收视习惯研究的关注变量包括被收看的节目和观众的人数。变量“收看的节目”可以取不同的值比如“超级碗”、“60 分钟”、“迷失”,变量“观众人数”是某一特定节目的收视人数。本质上讲,任何统计研究的原始数据都是关注变量的不同取值。

当考虑到事件发生的原因和结果时,我们有时将关注变量分为两类。例如,有关维生素 C 和感冒关系研究中的每一个人,在某一时期内服用不同剂量的维生素 C 会导致发生感冒的次数也不一样。因为我们试图了解服用维生素

C 是否会降低感冒的次数。我们说每日维生素 C 服用剂量是一个**解释变量(explanatory variable)**,它可以用来解释或者是引起感冒发生次数改变的一个原因,我们称感冒发生次数为**反应变量(response variable)**,因为它随着解释变量的变化而反应。

例 1-16 变量识别

确定下面每个研究的关注变量

(1) 索尔克脊髓灰质炎疫苗研究。

(2) 一个民意测验调查人们下一次选举中计划投票给谁。

解答

(1) 索尔克脊髓灰质炎疫苗研究所关注的两个变量分别是,疫苗和脊髓灰质炎,它们之所以是变量是由于它们可以有不同的取值:一个儿童可以注射疫苗,也可以不注射疫苗,可能得脊髓灰质炎,也可能不得脊髓灰质炎。在这个案例中,研究目的在于确定疫苗是否可以预防脊髓灰质炎,我们说疫苗是解释变量(它可以解释脊髓灰质炎发病率的变化),脊髓灰质炎是反应变量(假定它受疫苗的影响而发生反应)。

(2) 选举前的民意测验的关注变量是候选人姓名,因为人们一定会选择候选人之一并计划投票给他,支持率意思是指民意调查中计划投票给某一候选人的人数比例(在这个研究中没有原因和效应,所以我们不需要区分解释变量或反应变量)。 ■

1.3.2 观察研究

我们讨论的观察研究比如尼尔森收视率、民意调查还有确定学生的平均体重,存在这样的共性,通常研究的数据都是同时搜集的。有时,观察研究也会着眼于过去的的数据或者计划关注未来一段时间内的数据。

一个**回顾性研究(retrospective study)**,有时也称为个案对照研究,是使用过去数据的观察性研究,比如官方数据或者做过的访谈,来了解相关问题。回顾性研究在那些不可操作的或者违背道德的实验操作的案例中特别有价值。例如,假定我们想了解怀孕期间饮酒对新生儿的影响,因为早就知道怀孕期间饮酒可能会对新生儿有伤害,如果要求怀孕的妈妈进行饮酒的实验处理是不道德的。尽管如此,由于很多妈妈在过去的怀孕期间饮过酒(不知道会有危险或者忽视危险)。我们可以实施一个回顾性研

究，比较那些妈妈在怀孕期间饮酒和没有饮酒而出生的儿童的情况。

有时，当我们需要清楚的结论时，过去的记录是无效的，在这些案例中，研究者可以实施**前瞻性研究**（prospective study），也可称作纵向研究，指从那些具有共同因素的群体中搜集未来观察数据的研究设计。一个经典的前瞻性研究是哈佛护士健康研究，这个研究开始于1976年，旨在研究不同生活方式对妇女健康的影响（见本章末尾“聚集公众健康”栏目）。这个研究一直持续到今天，跟踪研究了数千名护士30多年，搜集了关于她们的生活方式和健康的大量数据。

观察研究的变式

最常见的观察研究是那些数据被一次性搜集的（或者时间距离非常近），观察研究的两种变式也是较常见的。

回顾性研究，或者称做个案比较研究，采用过去的记录，比如官方数据或者过去的访谈。

前瞻性研究，或者称做纵向研究，研究计划从那些具有共同因素的群体中搜集未来的数据。

例 1-17 观察研究

你想知道早产儿和足月儿在小学是否表现一样，你可以进行哪一种模式的研究？

解答 在这个案例中一个回顾性的观察研究是唯一可行的。你可以搜集他们的出生时的数据，然后比较早产儿和足月儿在小学的表现。

1.3.3 实验研究

因为实验研究需要相互作用，诸如提供实验处理，我们必须特别小心以保证实验设计可以提供给我们所要掌握的信息。让我们来分析一下实验设计中可能发生的问题。

1. 实验需要控制

分析一个实验，给予研究参与者服用维生素C以确定它预防感冒的效果，假定在3个月内服用维生素C平均会发生1.5次感冒，研究者怎样才能知道不服用维生素C会发生更多的感冒？为了回答这种类型的问题，研究者必须通过两组被试实施他的实验：一组被试每天服用大剂量的维生素C，另一组被试则不服用。简要地强调一下，大多数实验中研究参与

者被随机地分成两组是很重要的。

随机安排服用维生素C的一组被试叫做**实验处理组**（treatment group），因为它的成员接受了打算验证的实验处理（服用维生素C）；另一组不服用维生素C的人群叫做**控制组**（control group）。如果实验组的被试患感冒的次数显著地少于控制组的被试，研究者就可以确定服用维生素C是有效的。控制组的名称来自它能够帮助我们控制对实验结果的解释。

例 1-18 实验组和控制组

再看看索尔克脊髓灰质炎疫苗研究案例（见1.1节的案例研究），实验处理是什么？哪一组儿童构成了实验组，哪一组构成了控制组？

解答 实验处理是索尔克疫苗，实验组就是注射了索尔克疫苗的一组儿童，对照组包括没有注射索尔克疫苗而是注射了生理盐水的一组儿童。

例 1-19 莫扎特实验

一个研究将大学生分成两组。一组在完成一个设计好的特定任务之前聆听莫扎特或者其他经典音乐；另一组只是被分配完成任务。研究者发现那些聆听经典音乐者完成任务时稍胜一筹，但只是在刚开始听音乐的几分钟表现出来（两组在稍后完成的任务是一样的）。确认实验处理以及实验组 and 对照组。

解答 实验处理是经典音乐。实验组是那些聆听音乐的大学生；对照组是那些没有听音乐的大学生。

相关知识

所谓的莫扎特效应指的是聆听莫扎特音乐可以使婴儿变得更聪明。这个效应使得针对儿童的莫扎特产品工业化。佐治亚州甚至开始向新妈妈分发莫扎特音乐CD碟片。然而，更多关于莫扎特效应的最近研究发现并不能证实所声称的效应。



2. 混淆变量

使用控制组可以帮助我们说明那些影响研究结果的已知变量。尽管如此,研究者并不能完全清楚地或者不能够说明或解释其他重要的变量。分析下面这样的一项实验,统计学教师想验证合作学习(在学习中与其他人合作)的学生是否比独立学习的学生能够获得更高的分数。老师选取了5位学生采用合作学习的方式(实验组),5位学生独立学习(控制组)。为了保证这些学生具有相同的学习能力并且能够合作学习,老师只是选取了高分数的学生,在学期末,老师发现采用合作学习的学生比独立学习的学生获得了更高的分数。

研究的关注变量是合作学习(是否采用合

作学习方式)和期末考试成绩。但是对于老师来说不知道的在于,合作学习的学生都住在同一间宿舍,晚上宿舍会准时闭灯以保证他们有足够睡眠时间,这个事实可能引进一个新的变量,我们称之为适量的睡眠,这可以用来解释结果,换句话说,实验结果看起来支持合作学习是有益于成绩的,但因为老师没有考虑学生睡眠的情况,导致这个结论不公正。

在统计技术中,这个研究出现了混淆变量(confounding variables),学生的高分可能由于关注变量的影响(合作学习),也可能是由于适量的睡眠或者是两者的结合。因为老师没有考虑到学生睡眠时间上的差异,睡眠时间就叫做这个研究的混淆变量,你也可以发现其他像这样的潜在的影响研究的混淆变量。

案例研究

混淆药物结果

辉瑞公司研制了一种用于预防住院患者真菌感染的新药(叫做氟康唑)。一些研究发现新药比老药(叫做两性霉素B)更有效。然而,一些其他研究者后来的分析发现老药在被建议注射时患者却经常口服。这样研究就产生了一个混淆变量:原始的研究者发现新药比老药效果好,但是他们没有考虑到老药是被口服的这个混淆变量,在此基础上,新研究者考虑到了这个因素(只比较两种药在都被正确口服情况下的案例),结果发现新药的效果并不比老药更好。

3. 实验组和控制组的分配

如合作学习的实验所阐明的,如果实验组和控制组在某些重要方面有所不同(除了接受与不接受实验处理之外的其他因素),实验结果必定受到混淆变量的干扰。研究者通常采用两种策略来避免这种差别,以保证实验组和控制组在比较时是公平的。首先,可以随机分配实验参与者进入实验组和控制组,也就是说采用的实验设计方法要保证每一个实验参与者分配进每一组的机会是完全相同的。实验被试是随机分配的,这样实验组和控制组在影响实验结果的某些方面就不可能有差异了。其次,研究者努力保证实验组和对照组成员是足够多的。例如,合作学习实验,每一组包括50个学生比5个学生生活在同一间宿舍产生误差的可能性要小多了。

实验组和控制组的选取策略

(1) 随机分组:要能确保实验被试是随机分配到实验组和控制组中,这也意味着每一个被试被分配到任何一组中的机会是等同的。

(2) 选取充分的大样本:要能确保实验组和控制组都是充分的大样本,在某些明显的方面不可能是有差异的(除了一组采取实验处理,另一组没有非实验处理之外)。

例 1-20 索尔克研究的分组

简要说明索尔克脊髓灰质炎疫苗研究中选取实验组和控制组时是怎样运用两种策略的。

解答 总共大约40万名儿童参与了这项研究,其中一半注射了索尔克脊髓灰质炎疫苗(实验组),另一半注射了生理盐水(控制组)。第一种研究策略的实施是对所有的儿童随机分配到其中一组,第二种研究策略是选取了大量的研究参与者(每一组各20万名儿童),所以两组儿童基本不可能有差别。

4. 安慰剂效应

当参与一项实验时,人们可能仅仅只是知道他们参与了实验就会产生一些效果。例如,假定你要检验一种新的抗抑郁新药是否有效,你找到500个陷入抑郁中的人,并把他们随机

分配到服用新药的实验组和不服用新药的控制组，几个星期后，和这些患者进行访谈后结果显示，实验组的被试比控制组的被试感觉会更好些。你能说这是新药发生作用了吗？

不幸的是，服用新药的人心情有所改善，很有可能只是他们愿意接受某种治疗，并不能确定新药真正有效。只是因为他们相信自己接受了一种有用的治疗，情况就有所改善，这种形式的效应，叫做安慰剂效应（术语“安慰剂”来源于拉丁语“to please”）。

为了区分效果是由安慰剂引起的，还是真的由实验处理引起的，研究者努力保证实验的参与者并不知道他们是实验组还是控制组的成员，为了达到这个目标，研究者会给控制组的人也服用安慰剂：就是看起来或者感觉上像是和被检验的药物一样的东西，但是缺少它的活性成分。例如，一种丸药的实验中，安慰剂也被做成一样大小形状的药丸，只不过用糖代替了真正的药物成分。一种疫苗注射剂的实验中，安慰剂也被制成注射用生理盐水用来取代真正的疫苗成分。近期的检验中医针刺疗法效果的一项实验中，安慰剂包括真正针刺疗法所使用的针，只不过这种针刺的部位并不是针灸师所宣称的重要特定身体部位。

只要实验参与者并不知道他们接受的治疗是真正的治疗还是安慰剂，安慰剂效应对实验组和控制组的人影响就是等同的，如果两组的实验结果差异是显著的，这个差异就可以归因于实验处理。例如，在一项抗抑郁药物的研究中，只有当控制组也接受了安慰剂，但是实验组成员比起控制组成员效果更好时，我们才可以认定这种药物是有效的。为了达到更好的实验控制，一些实验采用三组：一个实验组、一个安慰剂组和一个控制组。安慰剂组是给予安慰剂而控制组则是什么也不给予。

安慰剂（placebo）是指在一项医疗方式检验研究中缺乏活性成分但看起来或者感觉上像是治疗药物或方法，以致实验参与者不能区分他们是接受了安慰剂还是接受了真正的治疗。安慰剂效应（placebo effect）是指由于患者仅仅相信自己接受了一种有用的治疗，病情就有所改善的情况。

注意：虽然所有实验参与者并不应该知道他们属于实验组还是控制组，但是出于伦理的

原因，实验参与者们还是应该被告知他们中的一部分给予的是安慰剂而不是真正的治疗。

思考题

过去的十年，研究者经常告诉一项研究的所有参与者他们接受了真正的治疗，但实际上有一半是接受的安慰剂，这为什么被看做不符合伦理道德的？讨论以下几种可能的原因。研究者可以被允许使用过去但是违背今天的伦理道德标准的研究结果吗？就你的观点进行辩论。

相关知识

安慰剂效应有着非凡的作用。在某些研究中，超过75%的参与者实际上的提高是受到了安慰剂效应的影响。安慰剂效应是如此之强，使得被试在被告知给予的是安慰剂而不是真正的治疗后，还乐于继续治疗，然而，有些研究者关于安慰剂效应的强度持有异议，甚至有一些人质疑这个效应的真实性。

例 1-21 疫苗安慰剂

索尔克脊髓灰质炎疫苗研究的安慰剂是什么？研究者为什么要使用安慰剂？

安慰剂是给予控制组儿童注射生理盐水。为了理解为什么要对控制组使用安慰剂，假定没有使用安慰剂，观察到实验组有效时，那将不可能知道是归因于疫苗的作用还是安慰剂效应。为了克服这种干扰，所有的实验参与者必须相信他们是给予了同样的治疗，这保证了安慰剂效应会公平地在两组都会发生，这种情况下如果两组还有差别就可以归因于疫苗了。

相关知识

一个相关的效应，就是著名的霍桑效应，实验被试以某种方式做出不同的反应，只是因为他们是实验的一部分，而不管他们是以什么样的方式进行实验。霍桑效应得名是由于在对厂矿工人的一个研究中的第一次观察是来自西方电气霍桑工厂。

5. 实验者效应

即使研究中被试并不知道他们接受的是真的治疗还是安慰剂，实验者仍然会产生影响。检验一种抗抑郁的新药，例如，实验者会与患

者进行访谈,看看他们是否感觉好些。但如果实验者知道谁是接受的真的治疗,谁是接受安慰剂的话,他会不自觉地对实验组的人们抱以更多微笑。这些微笑可能会改善参与者的心理,这会使得好像是治疗产生效果了而事实上是由实验者的原因引起的。实验者对于实验结果产生影响,这种类型的混淆,叫做**实验者效应**(experimenter effect),也叫做**罗森塔尔效应**。避免实验者效应的唯一途径就是保证实验者也不知道那一位被试在哪一组。

例 1-22 虐待儿童了吗

在一个著名的案例中,来自加利福尼亚贝克斯菲尔德的两对夫妇,被控告在他们的日托中心虐待 12 个学前儿童,虐待的证明最初来自和儿童的访谈。尽管如此,一个男子在监狱服刑 12 年后控告被推翻了,法官重新翻阅了调查记录,得出结论说,儿童给出了调查者想要听到的答案。如果我们以访谈者为实验者,这就是一个典型的实验者效应,因为调查者通过言语和某种方式的问话影响了儿童的回答。

6. 盲试

在统计方法中,关于谁在实验组谁在控制组,人们并不知情状况下进行操作叫做**盲试**(blinding)。**单盲实验**(single-blind)是指实验的参与者不知道自己属于实验的哪一组,但是实验者知道。如果参与者自己和实验者都不知道他属于哪一组,这种研究叫做**双盲实验**(double-blind)。当然,会有人追踪分组的情况以便于评估实验的最后结果。在双盲实验中,研究者建立一项研究并雇用实验者与参与者保持必要的接触,研究者自己避免和参与者有所接触,确保不以任何方式影响参与者,索尔克脊髓灰质炎疫苗研究就是一项双盲实验,因为无论是参与者(儿童)还是实验者(实施注射的护士和诊断灰质炎的医生)都不知道到底谁注射了真的疫苗,谁注射了安慰剂。

相关知识

和设想的在日托中心和幼儿园存在普遍虐待一样,许多类似的案例被再次审核以确定是否存在导致错误指责的实验者效应(那些面谈过儿童的实验者)。有关实验者效应的声明也包括被压抑记忆的案例,在这个案

例中咨询被设想为可以帮助人们恢复创伤性体验的记忆。



例 1-23 实验在哪出错了

下面所描述的每一个实验,澄清问题所在并说明怎样避免发生这样的问题。

(1) 一种用来治疗注意力缺陷障碍(ADD)的新药可以促使儿童更有耐心,随机选择一些患有 ADD 的儿童然后分为实验组和控制组,实验是单盲实验,实验者通过一对一的访谈评估儿童的耐心水平。

(2) 教育研究者想了解学习时聆听古典音乐是否能提高学习效果。他给两组学生同样的 2 小时的课程,然后学习一段时间进行一个小测验,一组包括 50 个喜欢并且在他们学习时会听古典音乐的学生。另一组,包括 50 个不喜欢听古典音乐且在学习时保持安静的学生。结果显示听古典音乐的一组学生测验成绩会更好一些。

(3) 研究者希望知道一种罕见的退行性疾病能否通过训练延缓其发展的进程,他们用 6 位确诊的患者做被试,并随机将其中 3 位分到实验组每天训练,另外 3 位分到控制组不进行训练,6 个月以后,比较两组的成员各自退行的程度。

(4) 脊椎按摩师对 25 位后背疼痛的患者进行了调整,之后,其中 18 位患者说感觉好多了,因此他得出结论这种调整是有效的。

解答

(1) 实验者通过访谈调查了儿童的耐心,但是由于他知道哪一个儿童服用了真正的药物,在访谈中实验者可能在无意之间对他们的语气不同,由于他们知道哪一个被试服用了真正的药物,所以在解释儿童行为时也会不一样。这是能混淆实验结果的实验者效应,实验应该是双盲实验。

(2) 这个问题是研究者并不是随机地分成两组,将喜欢古典音乐的人分在一组,而将不喜欢古典音乐的学生分在另一组,研究者的设计使这两组学生并不具备相同的一般特征。

(3) 研究的结果将难以解释,因为样本量不够充分,当然,对于这样一种退行性疾病,发现适合的研

究被试是非常困难的。

- (4) 25 个患者接受了调整代表了一个实验组，但是研究缺少一个控制组，患者感觉好一些可能只是安慰剂效应而非调整所起到的作用，脊椎按摩师可以在控制组雇用一名演员假装按摩改进他的研究（一个看起来一样，但是并不真正按照按摩方法和原则做）。这样，就可以比较两组的结果看是否伴有安慰剂效应。

例 1-24 确定研究类型

对于下面每一个问题，哪一种研究方式最可能得到答案？尽可能详细地说明。

- (1) 股票经纪人的平均收入是多少？
- (2) 安全带可以挽救生命吗？
- (3) 举重可以提高运动员百米跑的成绩吗？
- (4) 皮肤接触一种特殊的胶是否会引起皮疹？
- (5) 草药疗法能够减少重感冒的发生吗？
- (6) 使用白藜芦醇（一种红葡萄提取物）能够延长生命吗？

解答

- (1) 采用观察研究就可以告诉我们股票经纪人的平均收入。我们只需要对股票经纪人进行调查，调查并不能改变他们的收入。
- (2) 如果告诉某些人系安全带，告诉另一些人系安全带，这样的试验是违背伦理道德的。所以一个研究确定是否系安全带会挽救生命，必须是观察研究。由于一些人选择系安全带，另一些人选择不系，比较系与不系安全带的人发生事故后的死亡率，我们就可以知道系安全带是否会挽救生命。
- (3) 我们需要一个实验确定举重是否可以提高运动员百米跑的成绩，我们从一群跑步运动员中随机选取一些人分为实验组和控制组，实验组的运动员实施一个举重训练的计划，而控制组的运动员则不用举重。我们必须保证这些运动员除举重之外的其他训练是完全等同的。这样我们就可以知道举重训练对于实验组比起控制组是否提高了他们跑步的成绩。我们不可能采用盲试，因为我们无法不让实验的参与者知道他们是否要举重。
- (4) 可以进行一项实验，以确定皮肤接触某种胶是否会引起皮疹。在这个案例中，最好采用单盲实验，一组提供真正的胶，而另一组则提供的是外表看起来一样但是缺少活性成分的安慰剂。这不需要双盲实验，因为实验者似乎并不可能影响一个人是否产生皮疹（尽管如此，如果被试是否产生皮疹还需要诊断，实验者知道谁真正属于实验组就会影响到他的诊断）。

- (5) 我们应该采用双盲实验来检验草药疗法是否能够减少重感冒的发生，一些实验参与者接受真正的草药治疗，另外一些接受安慰剂。我们需要双盲实验，因为被试的心情和其他一些因素会受到实验者无心的态度的影响，这会进一步影响被试的重感冒，在双盲实验中，实验者并不知道被试属于哪一组，这样他对于两组的态度就不会不一样了。
- (6) 发现白藜芦醇并且提取成可利用的产品是近期的事，我们需要很多年的数据才能确定它是否能够延年益寿，我们可以通过前瞻性研究设计持续数年对实验参与者监控。实验参与者保留他们是否服用和服用多少白藜芦醇的书面记录，最后研究者分析此数据讨论白藜芦醇是否能够延长寿命。

相关知识

实验显示，服用大剂量白藜芦醇（源自红葡萄表皮的一种提取物）的小白鼠寿命会更长并且耐力更久，然而，服用大剂量的白藜芦醇对于人类是否有同样效果尚未可知，并且也不知道服用它是否安全。而且，小白鼠服用的白藜芦醇剂量很大，由于我们人类体型的原因，在人身上使用同样的剂量是不可行的。



1.3.4 元分析

所有个体的统计研究不是观察研究就是实验研究，最近几年来，统计学家发现将以往研究作为“我的”一个研究整体是有用的，从而发现我们能否从中得到一些不能从个体研究中得到的。例如，数以百计的研究都在讨论维生素 C 对于感冒的影响，研究者可以将已往研究作为一个整体来回顾这些研究中的数据。研究者将过去的许多研究作为一个整体回顾，这种形式的研究叫做元分析（meta-analysis）。

案例研究

抗抑郁药物：一个元分析

政府研究者估计每5个美国人就有一个人在他们一生中有时会陷入抑郁，每年因此失去劳动能力而花费的费用至少要400亿美元。直到20世纪90年代，唯一有效的抗抑郁药是经典的三环抗抑郁药。但是在过去的10年，几种新药也使用得很平常了。最有名的是百忧解，百忧解处方开的如此之多以致好多人笑称我们生活在“普罗萨克王国”。但是百忧解和其他新药有效吗？为了回答这个问题，美国卫生保健政策研究机构的研究者进行了一项元分析。

研究者通过搜索研究抑郁的医学文献，他们发现在过去的9年里发表了超过8000篇的研究报告。通过寻找符合某一标准的研究，将如观察患者至少超过6周的，并且对新药和老药进行比较的，这样缩小范围至300项研究，另外也分析了关于药物副作用的600项研究，元分析将所有这些研究作为一个整体探讨其结果，去发现在个体研究中不明显的趋势。

他们发现，新药对于严重的抑郁比安慰剂更有效，大约50%的人对新药有反应，只有32%的人对安慰剂有反应。尽管如此，老药三环抗抑郁药是同样有效的，而且价格更低些。新药和老药的副作用是不一样的，但是副作用的严重水平大致相当。这对于新药的生产者和使用者来说既是坏消息又是好消息，元分析可能对这些说不清楚的研究会有一些作用。

思考题

再看看抗抑郁药的元分析中，新药反应比例（50%），安慰剂的反应比例（32%）。与安慰剂相关的大反应是指如果真药有潜在副作用的话是否应该先给予安慰剂，为什么？如果你是一个心理学家，对于那些对真药没有反应的50%的人，你将给出什么样的建议？

1.3 练习题

统计素养和批判性思维

1. 安慰剂 安慰剂是什么？为什么安慰剂在检验一种药物是否有效的实验中是非常重要的？
2. 盲试 什么是盲试？为什么在检验一种药物是否有效的实验中盲试实验是非常重要的？
3. 混淆变量 在检验一种新疫苗是否有效的试验中，假定研究者使用男性作为实验组，使用女性作为安慰剂组，混淆变量是什么？它是如何影响这个试验结果的？
4. 不给予医疗措施 在一个著名的研究中，399名非洲裔美国籍梅毒患者并不给予任何治疗他们的医疗措施。这个目的在于了解非洲裔美国籍梅毒患者的病情发展。这是一个试验研究还是一个观察研究？为什么这个研究在道德和法律上都是错误的？
5. 衣服颜色 一个研究者计划对人们下面这个信念进行调查，“在夏天的烈日下穿浅颜色的衣服会比穿深颜色的衣服更舒适些”。在这种情况下应用双盲实验是有意义的吗？请说明原因。
6. 草坪实验 一位研究者计划检验一种促进草坪生长的肥料效果。在这种情况下应用双盲

实验有意义吗？

7. 疼痛测量 许多研究都涉及治疗后的疼痛测量方法。描述一种实验效应如何影响结果的不良反应，并且描述这种反应怎么可以避免。
8. 秃头的治疗 在一个秃发的实际研究中，结果发现安慰剂组比使用药物的实验组头发长得更好。这样的结果可能吗？这样的结果对于治疗来说意味着什么？

概论和应用

研究类型：对于练习题9~20，说明研究的类型并确定关注变量。请尽量具体一些。

9. 接触疗法 9岁大的埃米莉·罗莎在做了专业触摸治疗家检测以后，在《美国医学联合会杂志》上发表了一篇文章（参考第10章的“聚集教育”栏目）。使用一个纸板进行隔断，她将手放在治疗家的手上，治疗家被要求来确定埃米莉选择的是哪一只手。
10. 质量控制 联邦药物管理机构随机选择了拜耳公司的阿司匹林药片样本，每片阿司匹林的质量都被精确测量。
11. 磁性手镯 豪华游轮上给其中一些乘客准备了一副磁性手镯，要求乘客带上它以消除或

- 者减少晕船的效应。另外的乘客也会给予一副手鐲但是没有磁性。
12. 性别选择 在遗传和试管受精研究所的一项关于 YSOST 性别选择方法的研究中, 152 对夫妇生下了 127 个男孩和 25 个女孩。
 13. 双生子 在一项关于数百对瑞典双生子的研究中, 结果发现同卵双生子(来源于同一个卵子的双生子)比起异卵双生子(来源于两个独立的卵子的双生子)的心理技能水平更一致(《科学》)。
 14. 心脏疾病 欧洲一项关于具有异常水平的高半胱氨酸的 1 500 人的研究发现, 这些个体比一般人患心脏病的风险高出两倍。然而, 那些在研究中服用维生素 B 补充剂的人患心脏病的风险会显著降低(*Journal of the American Medical Association*)。
 15. 前列腺癌 关于 11 个个体的一个研究分析, 试图确定在输尿管切除术和前列腺癌发生率之间有没有确定性的联系(*Chance*)。
 16. 美国在线的民意调查 美国在线对 38 410 人进行了调查, 问题是“你在收听到长期的天气预报之后会相应地做多少食物储备?”在这些人的回答中 47% 的人说很少或没有。
 17. 转基因玉米 纽约大学的一项研究发现当在转基因玉米花粉上喷洒杀虫剂时, 这种农药可通过其根系进入土壤(*Nature*)。
 18. 骨髓干细胞研究 普林斯顿调查研究联合会进行的一项民意调查发表在《新闻周刊》上, 针对 1 002 名美国成人的研究, 其中 48% 的人说他们同意使用联邦税收资金去进行从胚胎上提取的骨髓干细胞的医学研究。
 19. 磁疗 在一项关于背部疼痛的磁疗效果的研究中, 一些被试使用磁疗工具进行治疗, 另外一些被试则给予看起来一样的但没有磁疗效果的工具治疗, 结果发现在治疗背部疼痛中, 磁疗并没有产生效果(*Journal of the American Medical Association*)。
 20. 高压线和癌症 根据已有的数百项关于高压线和居住在它附近的人们的癌症增加率是否有关的科学和统计研究的一项总结性结论说, 在高压线和癌症之间并没有明显的关联(*Journal of the American Medical Association*)。
- 下面的实验有什么错误之处: 在练习题 21~28 中, 在所描述的每项试验中找出可能产生混淆的一些问题, 并且解释怎么样可以避免这些问题, 并讨论任何其他可能影响结果的问题。
21. 白杨树的生长 有一个实验是用来评估灌溉和肥料对于白杨树生长的效果的: 在一个潮湿多雨的地区对一组白杨树使用肥料, 在一个干旱的地区对一组白杨树进行灌溉。
 22. 网络购物 通过在《个人电脑杂志》中的一则广告, 招聘了 200 名志愿者, 让他们来参与某项研究以比较通过网络进行购物的花费和不通过网络进行购物的花费, 要求每一个人选择是分在网络购物组还是非网络购物组, 一个月之后比较两组被试在购物中所花的总钱数。
 23. 辛烷值 在一项比较不同汽油辛烷值实验中, 24 辆大篷车使用 87 号辛烷值汽油, 28 辆越野车使用 91 号辛烷值汽油, 在每一辆交通车行驶 250 公里之后, 测量所消耗的汽油量。
 24. 阿司匹林实验 在临床试验的第一阶段, 要检验阿司匹林在预防心肌梗死中的效果, 给 3 个人服用阿司匹林, 给另外 7 个人服用安慰剂。
 25. 高血压药物实验 在一个临床实验中, 检验一种药物对于治疗高血压的效果, 被试被告知他们服用的是真正的药还是安慰剂。
 26. 香港脚 在一个临床实验中, 一种洗液被用来治疗足癣以检验它的效果, 评估实验结果的生理学家知道哪些被试给予了治疗, 哪些被试给予的是安慰剂。
 27. 举重 在一项关于举重对血压影响的实验中, 一组被试要经历包括举重程序的实验处理, 另外一组被试只是举起网球。
 28. 绘画颜料的混合 在两种不同的绘画颜料的耐用性实验中, 评估实验结果的研究者知道哪一个样本是哪两种颜料的混合。
- 分析实验:** 练习题 29~32 中, 每题呈现了一些可能出现在实验中的问题, 如果你是实验的设计者, 你将怎么选择实验组和控制组。实验应该是单盲、双盲还是都不用? 解释你的原因。
29. 贝多芬和智力 聆听贝多芬的音乐会使婴儿变得更聪明吗?
 30. 阿托伐他汀和胆固醇水平 服用药物阿托伐他汀会降低胆固醇水平吗?

31. 乙醇和公里数 将乙醇加进汽油中会降低公里数吗?
32. 外墙壁 铝制外墙壁要比木质外墙壁更耐用吗?

网络和远程项目

请在网址 www.aw.com/bbt 上打开第1章的有用链接“Links for Internet Projects”。

33. 在压抑记忆中的实验者效应 在网络中寻找一篇关于如何看待压抑记忆的文章或者信息。基于你所浏览的网页内容,简要概括一两个这样有意义的案例,你认为被修复的记忆是否会被实验者效应所影响。
34. 实验中的伦理道德 每个时代的伦理道德标准都有所变化,一个臭名昭著的不符合伦理道德的案例是1932~1972年之间发生在亚拉巴马州塔斯基吉市的一项梅毒研究(见本节练习题4),在这项研究中,非洲裔的美国籍男性被告知他们正在进行梅毒的治疗,但实际上他们并没有给予任何治疗,实验者隐瞒了他们的研究目标,是为了观察这种疾病的长期发展结果。使用网络了解塔斯基吉梅毒研究的历史,举行一个班级讨论。针对包括这个案例的道德问题,或者写一篇小论文。总结这个案例,以及它的道德教训。
35. 争论:我们应该使用非道德实验中的数据吗

过去的研究经常不符合今天的伦理道德标准,在一些极端的案例中,比如纳粹德国一些医生所实施的研究,这些研究者有时会杀死他们的实验被试,这些过去的非道德研究明显地侵犯了实验被试的人权,但在这些案例中它又有可能帮助我们今天人类的一些发现,使用这些非道德研究的一些结果是符合道德的吗?

36. 一些终止的研究 有时一些研究在完成前就被终止了,应用网络寻找这样一个研究的例子。这个研究为什么会终止?它应该被终止吗?或者说完成这个研究会更好吗?

新闻中的统计学

37. 观察研究 从前几周的报纸中,寻找一个应用统计的观察研究的例子,简要地描述这个研究,并概括它的结论。
38. 实验研究 从前几周的报纸中,寻找一个包含实验内容的统计研究,简要地描述它的研究,并概括它的结论。
39. 回顾性研究 从过去几周的报纸中,寻找一个观察性的回顾性研究的例子,简要地描述它的研究,并概括它的结论。
40. 元分析研究 从过去几周的报纸中,寻找一个元分析研究的例子,简要地描述它的研究,并概括它的结论。

1.4 统计研究可信吗

这本书余下的大部分,主要为了帮助你深入理解到目前为止我们学习的这些概念和定义。但是你早已充分了解本书要达到如下一个目标:学会回答这个问题,“你应该相信一项统计研究吗?”

大多数研究者诚实而完整地实施他们的研究,大多数统计研究也是严谨而仔细的。尽管如此,统计研究是复杂异常的,偏差可能以多种不同的方式发生。仔细地研究统计分析报告是非常重要的,“我应该相信一个统计研究吗”,这个问题没有统一明确的答案。无论如何,下面的八条指导原则是非常有帮助的。沿着这个思路,我们也介绍一些以备后面讨论之用的定义和概念。

批判性评估统计研究的八条指导原则

- (1) 明确统计研究的目标、总体以及研究类型;
- (2) 分析偏差来源,尤其是研究者是否存在偏差;
- (3) 检查抽样方法确定是否能够得到一个代表性样本;
- (4) 寻找界定或测量关注变量中存在的会使解释报告结果变得困难的问题;
- (5) 找出会使研究结论无效的混淆变量;
- (6) 考虑调查或民意测验中的文本措词,寻找可能导致不正确或不诚实作答的任何影响因素;
- (7) 公正地检查用图表表达和文字陈述的研究结果,因为研究者和媒体经常会绘制误导性的图表,得出研究结果并不支持的结论;
- (8) 回顾并仔细思考结论,研究是否实现了它的目标,结论是否有意义,研究是否有实际应用价值。

新闻报道并不总是提供足够的信息，以便于你应用上述指导原则。但是你可以利用网络以外的信息。寻找“NASA 报道”和《新英格兰医学杂志》上的线索，帮助你追索原始资料和其他相关信息。

1.4.1 指导一：明确统计研究的目标、总体和研究类型

评估一个统计研究的第一步就是理解研究目标，基于你所听到和所读到的研究，试着回答以下问题：

(1) 研究设计的目的是什么？

(2) 研究的总体是什么？总体是否明确和恰当地定义？

(3) 研究是观察性研究、实验研究还是元分析？如果是观察性研究，是不是回溯性研究？如果是实验性研究，是单盲实验还是双盲实验？实验组和控制组是否恰当地随机分配的？研究类型与研究目的是否对应恰当？

例 1-25 研究类型是否恰当

想象一下如下新闻报道（假设的）：“研究者给出 100 个实验参与者他们每个人的星座分析结果，问他们星座分析是否准确，85% 的会说他们的星座分析结果是精确的，研究者因此得出结论，星座说大多数情况下是有效的。”按照第一条指导原则分析这个研究。

解答 研究的目标是确定星座分析结果是否有效。基于这个新闻报道，它呈现的研究是观察研究，研究者只是简单地询问实验参与者星座分析结果是否准确。无论怎样，星座说是主观的，研究应该是一个有控制的实验，一些人给予真实的星座分析结果，另一些人给予伪造的星座分析结果，研究者能够通过两组找出差异，尽管如此，研究者提问研究参与者的方式也容易影响结果，所以应该采用双盲实验。概括地说，研究类型对于研究目的而言是不恰当的，研究结论也是无意义的。

思考题

你自己尝试对星座说进行检验。寻找 12 星座昨天的星相图，将 12 个星座各自的结果分析放在一个单独的纸张上，不做任何标记，随机混合这些纸片，然后问一些人哪一个纸片上的结果分析是与自己相符合的？有多少人会选择与自己星座对应的答案？讨论这个结果。

相关知识

调查发现接近半数的美国人相信占星术。然而，在一项控制实验中，占星术所预测的结果成为现实的并不比随机猜测的机会更高。



例 1-26 阿司匹林能预防心脏病吗

《新英格兰医学杂志》上的一篇研究报告（第 318 卷第 4 期）试图确定阿司匹林是否能预防心脏病。它包括 22 000 名医生诊断为存在心脏病风险的男性。这些男性被分为服用阿司匹林的实验组和不服用阿司匹林的控制组，研究结果令人信服地支持阿司匹林的作用，但是实验在完成前因为伦理道德原因被中止了，被试被告知了结果。许多报纸在以这样的标题报道说“阿司匹林能用于预防心脏病。”应用指导原则一来分析这个报道的标题。

解答 这是个不恰当的实验研究，它的结果看起来是让人信服的。然而，这个研究的样本只有男性被试，这意味着研究结果只能应用于男性总体。因为在男性身上检验的实验结果并不一定适合用于女性，报道的标题错误陈述了结果，因为它并没有明确区分总体。

相关知识

许多最近的研究显示出男性和女性对同样医学治疗的反应有着本质差异。例如，阿司匹林在稀释血液方面的作用在男性身上更明显（稀释血液被认为可以帮助某些人们预防心脏病）；吗啡止痛对于女性更有效；而布洛芬对男性更有效；女性比男性的心脏移植排斥反应更强。这些差异源于荷尔蒙的交互作用，或者是男性女性在代谢不同药物方面的差异。

1.4.2 指导二：分析偏差来源

统计研究被看做客观的，但是实施研究的和资助研究的人们可能会导致偏差。因此，考虑研究资料，评估那些可能使研究结果无效的潜在偏差，这些是非常重要的。

为市场促销或者其他商业目的而设计的统计研究，其中的偏差可能是显而易见的，例如，一个牙膏广告宣称“4/5的牙医推荐我们的牙膏品牌”，这里面就存在统计偏差，但是我们没有获得这个调查结论是怎么得出的详细过程。因为广告商显然只想说他们的牙膏品牌好，但是如果关于这个结果如何获得的更详细的信息，做出一个严格意义上的统计结论是很困难的。

另一些案例中的偏差可能是更微妙的。假定一个精心设计的实验，用来检验一种治愈癌症的新药。表面上，研究可能看起来非常可信，但是如果研究的基金来源于一家医药公司，假如新药被证明是有效的话，这家公司可以通过出售这种新药获得上亿美元的收益。研究者也许可以公正无私地设计他们的实验而不考虑基金的来源，但是它仍然需要有一些额外的调查来确认。

重要的统计研究经常需要中立的专家进行评估。例如，一个科学家检查其他人的研究过程叫做同行评议（peer review，因为做评估工作的科学家是实施研究者的同行），声誉良好的学术期刊在研究报告发布给公众之前需要对其进行同行评议。同行评议并不能保证研究是有效的，但是会使得它是可信的，因为这表明其他专家认为这个研究的设计是合理恰当的。

例 1-27 吸烟有害健康吗

到1963年，研究清楚地表明吸烟对健康的危害，美国卫生部长公开声明吸烟有害健康。从那时起实施的很多研究进一步支持了这个结论。然而，当大量的主要研究显示吸烟有害健康，但是也有很少数的研究表明吸烟无害甚至有益健康。这些研究通常是由烟草公司基金支持的烟草研究机构实施的。按照指导原则二分析这些研究。

解答 即使在这样的案例中，仍然难以确定谁更可信。尽管如此，显示吸烟有害健康的研究主要来自经过同行评议的研究，相反，烟草研究机构设计的研究存在一个清晰的可能性偏差。偏差的可能性并不意味着研究一定是有偏差的，但它几乎与所有主要的研究都相矛盾的，这个事实应该是与此有关的。

相关知识

在争论了10年之后，1999年10月，世界上最大的烟草制造商菲利普·莫里斯公司公开承认吸烟可以导致肺癌、心脏病、肺气肿和其他严重疾病。但随后不久，菲利普·莫里斯公司就更为奥驰亚集团。

例 1-28 科学新闻发布会

假定夜间电视新闻中展示了一个新闻发布会，科学家宣布他们新发现了一种能够终止老化过程的一种化学物质。这个研究还没有进行同行评议，按照指导原则二分析这个研究。

解答 科学家经常在新闻发布会上公布他们的研究结果，以便公众可以知道他们在研究什么。然而，大量的实验需要评估研究中可能的偏差或其他误差，这是一个同行评议的过程，直到经过同行评议并且发表在一个声誉良好的学术期刊上，这些发现才被初步认可，特别是能够终止老化这样令人惊奇的研究结果。

1.4.3 指导三：检查抽样方法

如果样本不能很好地代表总体的话，一项统计研究将是无效的，不恰当的抽样方法总是会产生一个有偏差的样本，这样会使研究结果无效。

很多方式都会产生有偏差的样本，但是两个密切相关的问题是最常见的。第一个问题是，选择性偏差（selection bias，或者叫选择效应），当研究者选择样本时带有某种倾向使得样本并不代表总体，这种偏差就会产生。例如，一个选举前的民意调查仅仅针对的是注册的共和党人，选择性偏差就会产生，因为结果似乎并不反映所有政党的（包括没有党派）投票人的观点。

第二个问题是被试偏差（participation bias），当某些人自己决定成为研究的一部分时也会发生偏差，也就是说，这些实验被试是志愿者。大多数被试偏差发生于在自选择调查（self-selected surveys，或者是自愿响应调查）中，调查或者问卷让人们自己选择是否成为研究的参与者。在这些案例中，对某一问题具有强烈情绪的人更愿意成为实验的参与者，他们的观点可能并不能代表大多数人的观点。

案例研究

1936 年《文学文摘》调查

《文学文摘》，是 20 世纪 30 年代美国的一本流行杂志，它进行大型的民意调查几次成功地预测了总统选举的结果。1936 年，《文学文摘》的编辑在总统选举前设计了一个特别庞大的民意调查，他们从各种不同的名单上随机选取 1 000 万人的样本，名单包括电话号码本和乡村俱乐部的名单。他们邮寄了明信片选票给这 1 000 万人，大约 240 万人回复了明信片，《文学文摘》的编辑预测阿尔夫·兰登将赢得 57% 的选票，对比富兰克林·罗斯福将只获得 43% 的选票，事实上，罗斯福获得全部投票人 62% 的选票。为什么这么庞大的民意调查会错得这么离谱？

这个样本包含了选择性偏差和被试偏差，选择性偏差存在是因为《文学文摘》选择了 1 000 万人的方式偏向于富人。例如，从电话号码本上选择名字意味着选择的人仅仅包括在 1936 年能付得起电话费用的人。同样，乡村俱乐部成员通常是非常富有的，这个选择性偏差偏向于共和党的兰登，因为 20 世纪 30 年代的富人倾向于投票给共和党的候选人。

被试偏差产生是因为回复明信片选票是自愿的，只有对选举怀有强烈情感的人更愿意回复明信片，这个偏差也有利于兰登，因为他是挑战者，那些不喜欢总统罗斯福的人们更愿意回复明信片，这样，即使这是一个样本巨大的调查，两种形式的偏差使得抽样结果变得无意义。

相关知识

一位年轻的名叫乔治·盖洛普的调查者在 1936 年大选之前建立了他自己的调查。他只是向随机选取的 3 000 人发放了明信片进行了调查，他就不仅正确地预测了大选的结果，也预测了《文学文摘》调查的结果，误差在 1% 之内，随后盖洛普建立了非常成功的调查机构。

相关知识

当被要求做一个调查时，超过 1/3 的美国人会粗暴地关上房门或者挂掉电话，从而使合法的调查者不得不进行自选择问题。人们挂电话的一个原因是伪装成市场研究的推销的泛滥，一名电话推销员假装向你调查问题而目的在于推销产品。

例 1-29 自选择测验

电视节目《晚间报道》进行了一项调查，电视观众被询问联合国总部是否应该设在美国，观众可以通过支付 50 美分拨打电话号码“900”，对此问题做出回答来表达自己的观点，这个调查接到了 18.6 万个电话回复，其中 67% 希望联合国总部离开美国，就同一问题进行的一个调查，采用简单随机取样选取了 500 人，其中的 72% 的人希望联合国总部继续设在美国，哪一个调查更像是能代表美国人的观念呢？

解答 《晚间报道》的样本存在严重的偏差，

它的选择性偏差由于它的样本是仅仅是这个节目的电视观众中选取的，而不是来自全部美国人。调查本身也是一个自选择调查，电视观众不仅要选择是否回答，还要支付 50 美分。这个费用使的那些具有改变的强烈需要的人才会做出反应。尽管回复的人数非常多，《晚间报道》调查并不能给出一个有意义的结果。相反，随机抽取的 500 个人的样本更具有代表性，因此这个小型调查的结果更能代表全部美国人的真实观点。

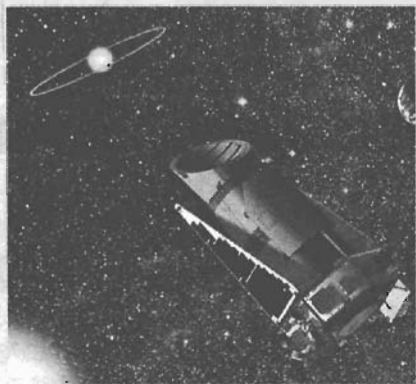
例 1-30 环绕其他恒星的行星

直到 20 世纪 90 年代中期，关于我们自己所在的太阳系以外的行星，天文学家还从来没有发现决定性的事实。但是随着科技的进步，发现这样的行星开始具有可能性了。到 2007 年，大约有 200 多个这样的行星被发现了。现有的技术使得发现大行星比小行星容易，发现其运转轨道离其恒星近的比远的更容易，按照太阳系行星导出的规律，大行星的运转轨道距离它的恒星非常近是非常罕见的，但这种现象却在发现的 200 多颗行星中非常普遍。这是不是意味着按照太阳系行星现象推出的理论错了呢？

解答 虽然理论上推定大行星的运转轨道距离它的恒星非常近是极少见的，但是科技让这种少见的例子非常容易被发现。许多这样的行星发现可能反映了选择效应，使之更倾向于包括以往比较罕见的行星类型，从而产生样本（已经发现的行星）偏差。实际上，在这个案例中几乎是确定的，因为许多“正常”的行星被发现，只要对现存的理论稍微调整一下，天文学家们就可以解释为什么早期发现中这种不常见的行星类型反而更多。

相关知识

美国国家航空航天局的开普勒计划，向太空发射一个轨道望远镜，主要用于观测围绕其他恒星运行的和地球一样大小的行星。它通过每一次行星轨道通过开普勒望远镜前面时，恒星稍微变暗的光来发现行星，这意味着我们能够探测的行星只是其行星系统刚好在我们视线之内的很小一部分星体。



1.4.4 指导四：定义或测量关注变量

如果研究中变量难以定义和测量，那么统计研究的结果也就难以解释了。例如，设想构建一个运动量影响你静息心率的研究。关注变量是运动量和静息心率。这两个变量都是难以定义和测量的。以运动量为例，这个定义包含什么是很不清楚的，比如它包括步行到学校去吗？即便我们详细说明这个定义，我们怎样测量那些各种形式的运动量？

思考题

你怎么测量你的静息心率？描述定义和测量静息心率时会有哪些困难？

例1-31 钱能买来爱情吗

一项发布在《今日美国》上的盖洛普民意测验，调查对象包含最富有的1%的美国人，调查发现人们愿意支付平均487 000美元换取真爱，407 000美元换取伟大的智慧，285 000美元换取才能，259 000美元换取永葆青春。按照指导四分析这个结果。

解答 研究变量是非常难以定义的。比如，你怎么定义真爱，它意味着爱一天、一生还是一段时间？同样地，在鼻子上能使一个勺子平衡的能力算是才能

吗？由于变量如此难以定义，仁者见仁，智者见智，不同的人会有不一样的解释，因此研究结果也就变得难以解释了。

例1-32 非法药品供应

一个经常被引用的统计数据是执法机关成功地制止了10%~20%非法药物进入美国。你相信这个统计结果吗？

解答 这个有关非法药物拦截的研究中存在两个可能的变量：被拦截的非法药物数量和未被拦截的非法药物数量。测量政府执法机关拦截的非法药物的数量是相对容易的。然而，由于这些药物是非法的，不管是谁报告未被拦截的非法药物数量看起来都是不可能的，那又是怎么能知道拦截的非法药物是总量的10%~20%？据《纽约时报》的分析，一个警官引述说，他的同事管这种类型的统计数据叫做“无中生有”。

1.4.5 指导五：找出混淆变量

通常，不属于研究想要的变量会使得结果难以正确解释。如同1.3节所讨论的，这些混淆变量并不总是很容易发现的。有时，在完成数年后它们才被发现，有时，它们就根本发现不了，这样，研究结论可能会被接受，尽管它不正确。幸运的是，混淆变量有时相当明显，只要认真想一想影响研究结果的因素，它们就能被发现了。

例1-33 氡气和肺癌

氡气是一种自然形成的（铀的衰退）具有放射性的气体，这种气体通过地基可能浸入建筑物中，并且如果门窗是密闭的话还会累积到较高浓度，设想（猜想）一个研究，要确定氡气是否导致肺癌，比较氡气水平一般的科罗拉多和氡气水平偏低的香港的肺癌发生率。假定研究发现两地的肺癌发生率是一样的，然后推论说氡气不是显著导致肺癌的原因，这样推论是合理的吗？

解答 关注变量是氡气的数量（本研究中的解释变量）和肺癌发生率（反应变量）。氡气不是导致肺癌唯一可能的原因。例如，吸烟也会导致肺癌，因此，吸烟率可能会是本研究的混淆变量，特别是香港的吸烟率要比科罗拉多州高很多。这样的话，如果不考虑吸烟率（也可能存在其他变量），我们不能得出任何氡气和肺癌关系的结论。事实上，严格的研究总是显示氡气会导致肺癌，美国环境保护局（EPA）建议人们采取措施预防室内氡气。

相关知识

许多五金商店销售一种用来检测你房间内氧气累积量的设备，它是通过安装一种合适的氧气消除系统来解决这个问题。该装置通常包括一个在氧气进入房屋之前就将氧气抽至户外的风扇。

1.4.6 指导六：考虑调查的设置和措词

即使调查选取了合适的样本，明确定义概念及界定问题，你可能还要当心语言和措词引发的问題，这可能会导致不正确的或者不诚实的反应。当调查涉及敏感问题，比如个人习惯或收入等时，尤其可能会引起不诚实的反应。例如，当问到“你会偷逃个人收入所得税吗”，那些真的偷逃个人收入税的人也不可能诚实回答，除非他们确认完全保密（即使真的那样也未必诚实作答）。

在其他情况下，如果问题的措词招致偏差，即使诚实作答也不一定准确。有时，问题中的措词顺序也会影响结果。一个在德国实施的调查问到下面两个问题：

A：你认为交通运输引起的污染比工业生产大还是小？

B：你认为工业生产引起的污染比交通运输大还是小？

这两个问题的唯一区别是交通运输和工业生产的前后顺序，但是这个差别戏剧性地改变了调查结果：第一个问题中，45%回答是交通运输引起的污染大，32%认为工业生产引起的污染大；第2个问题中，只有24%回答是交通运输引起的污染大，却有57%认为工业生产引起的污染大。

例 1-34 你想减税吗

在美国政府运行年度预算有所盈余的时期，共和党的国会议员提出了一个减税方案，并且共和党全国委员会委托实施了一项民意调查，以了解美国人民是否支持这项提案。当问题为“你赞同实施一项减税方案吗”，67%回答是。我们可以断定美国人赞同这个提案吗？

解答 当问题为“你赞同实施一项减税方案吗”，这是有偏差的，因为它没有给出其他选项。实际上，同时进行的一个独立民意调查，则给出应用剩余收入的各种反应选项。结果发现31%希望这些钱用于社会安全，26%希望用于减少国家债务，只

有18%愿意用于减税（余下25%反应则是选择了各种其他选项）

相关知识

在一个调查中人们更可能选择排在前面的选项。这是心理学家所说的有效性误差，基于心中会想到什么是有效的而产生的一种判断倾向。专业调查机构特别注意避免这种问题，例如，他们所设计的包括两个选项的问题，会给一半被调查者以一种选项顺序呈现，而对另一半被调查者则呈现相反的选项顺序。

例 1-35 敏感性调查

在波士顿地区进行的对天主教徒的一项调查，询问他们避孕药是否可用于未婚女性，第一个调查是个人访谈，46%的人回答说是，第2个调查是通过邮件和电话实施的，结果75%的人回答是。哪一个调查可能会更准确？

解答 避孕药是一个敏感主题，尤其在天主教徒中（因为天主教反对使用避孕药），个人访谈的调查可能会鼓励不诚实的回答，通过邮件和电话实施的调查，看起来更可能反映被试的真实观点。

1.4.7 指导七：检查用图表表示的结果及推出的结论是否公正

即使一个统计研究实施得很理想，当它用图表表达结果和推导出结论时也很可能是错误的。研究者偶尔会误解他们的研究结果，或者得出研究结果并不支持的结论，特别是当他们自己存在偏见时。新闻报告者更可能曲解调查或者得出一个没有根据的结论，这样做是为了让故事看起来更引人注目。曲解的图表特别常见（我们将在第3章详细探讨这个主题）。你应该努力寻找研究的解释（图表中和文本中）和它的实际数据之间不一致的地方。

例 1-36 学校董事会需要统计课程吗

在科罗拉多大学波尔得分校的学校董事会上，当宣布波尔得分校28%的学生阅读水平低于其年级平均水平时，引起了一阵喧哗，由此推断下来，阅读的教学方法需要改变。这个宣布是基于阅读测验而来的，28%的波尔得分校学生阅读成绩低于他们所在年级的州平均水平。这些数据支持董事会的结论吗？

解答 实际上，28%的波尔得分校学生阅读成绩低于他们所在年级的州平均水平，72%的学生高于平均水平。这样，学校董事会关于学生阅读低于年级平

均水平的陈述是无意义的，除非“年级水平”是指某一特定年级的州平均成绩。“年级水平”的解释是令人惊奇的，因为它意味着全州的学生总是有一半人在这个水平之下，而无论分数是多少。本案例中，可能在教学方法方面确实需要提高，但是所提供的数据并不支持这个结论。 ■

1.4.8 指导八：退一步再思考一下结论

最后，即使一项研究按照前面所有的指导来分析都是合理的，你仍然需要退一步来思考以下结论，问你自己如下的问题：

- (1) 这个研究实现了目标吗？
- (2) 这个结论有意义吗？
- (3) 你能排除基于研究结果的其他解释吗？
- (4) 如果结论是有意义的，它们有什么实践价值吗？

例 1-37 超乎寻常的公告

假定（设想）一个研究得出结论说，佩戴上一条金链可以提高你在一场交通事故中生存机会的

10%，这个声称是基于一项关于生存率和穿着关系的统计分析而来的。仔细分析这个研究，它的设计是合理而严格的。当驾驶汽车时你会佩戴一条金链吗？

解答 尽管这个研究非常严谨，但是宣称一条金链可以在交通事故中提高生存机会，这仍然是让人难以置信的。毕竟，一条细链子怎样在交通碰撞中帮助你？很可能是存在未知效应的，才能使金链子与碰撞结果发生关联，这样的结果可能更像是侥幸或者不确定的混淆变量发生作用（例如，可能那些佩戴金链的人更富有，并且他们开的车是具备更多安全设施的新车，从而降低了他们的死亡率）。 ■

例 1-38 实践意义

一个有关减肥的实验设计，用以比较“快餐添加剂”与采用其他方法减肥的效果，8周以后，结果显示，实验组减肥效果比控制组平均多0.5磅^①。假定这种减肥方法没有不良副作用，研究是否意味着“快餐添加剂”是人们减肥的一种好方法。

解答 与人们的平均体重相比，0.5磅几乎算不上什么，因此，从实验结果上看是有效的，但看起来实践意义却不强。 ■

1.4 练习题

统计素养和批判性思维

1. 同行评议 什么是同行评议？它的用处在哪里？
2. 选择偏差和参与者偏差 描述并比较样本的选择偏差和参与者偏差。
3. 自选择调查 为什么说自选择调查总是会产生参与者偏差？
4. 混淆变量 什么是混淆变量？它能导致什么结果？

判断正误：判断练习题5~8中的陈述是有意义的（表述正确）还是无意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断会更重要。

5. 调查效度 电视调查访问了超过100万个电话用户，因此它比由专业调查者设计的调查更有效。专业调查面试只包括了几百个人的面谈。
6. 调查所在地区 调查所在地区会包括一个选择偏差，因为问卷主要在天主教堂发放。
7. 维生素C和感冒 我的实验证明了维生素C能降低感冒的严重性，因为我非常精细地设计

了实验，并尽可能地排除了每一个混淆变量。

8. 加速慢跑 参加慢跑练习的每个人，应该尝试一种新的训练方案，因为非常精细的实验研究表明能提高你的速度1%。

概念与应用

应用指导：在练习题9~16中，确定与哪一条评估统计研究的指导原则最相关，解释你的理由。

9. 质量控制 公共关系部实施了一项研究，关于公司生产的断路器的残次品率。
10. 性别选择 在一项检验性别选择方法是否可以提高女孩出生率的研究后发现，出生的婴儿23个是女孩，21个是男孩，所以公司发言人声称，这种方法在大多数时候是有效的。
11. 伦理道德 在对250名律师的研究中，每个人都询问他是否具有良好的职业道德。
12. 农业 研究者得出结论说，在加利福尼亚州的一种灌溉系统对西红柿的生长要比在亚利桑那州的另一个类似系统更有效。
13. 监狱囚犯的调查 本书其中一位作者所在大

① 1磅=0.4536千克。

学通过信件调查的方式，根据是否累犯对监狱的囚犯进行分类，结果发现，这些犯人中的10%是累犯。

14. 选举调查 在下述标题下：特纳将以压倒性优势获得选举胜利。这条新闻报道了在一个选前的民意调查中，特纳获得了55%的支持，要高于他的竞争对手45%的支持率。
15. 核能调查 随机挑选一些成人，让他们回答以下问题：你是否同意增加可能会杀死成千上万无辜人群的核能利用。
16. 假冒伪劣产品 一家大制造商计划研究在美国生产的假冒伪劣产品2000年的价值与近年来他们所生产产品的价值对比。

偏差：在练习题17~20中确定并解释下面所描述的研究中至少一项偏差来源。然后针对如何避免偏差提出你的建议。

17. 健康巧克力 在《纽约时报》发表的一篇文章中，包括如下内容：“在过去很长的一段时间内，在食物金字塔结构中，巧克力的价值越来越被正确地认识到，尽管在红酒、水果、蔬菜和绿茶这些健康食品的高调包围中。据《营养学杂志》的报道中发现，在食用完巧克力以后，检验被试血液中的抗氧化物质的水平有所提高，巧克力中含有黄酮类抗氧化物等与降低心脏病和中风风险有关的物质。糖果公司和巧克力制造联合会赞助了这项研究。”
18. 名著 作者写下了 *Women and Love: A Cultural Revolution in Progress*，他发出了10万份问卷给各种各样的女性群体，最后收到了4500份回复，在这个基础上得到了他的结论。
19. 纳普斯特 《新闻周刊》杂志进行了一项关于纳普斯特网站下载音乐的调查。读者可以在《新闻周刊》的网页做出他们的回答。
20. 调查方法 你计划在你所在的州实施一个调查，计算能够叫出副州长名字的（他计划竞选联邦参议员）人数的百分比。你获得了一份在本州拥有自有住房的人的地址，你从这份名单中随机挑选了850个人，并给他们邮寄了问卷进行调查。
21. 都是措词不同的结果吗 普林斯顿调查研究联合会为《新闻周刊》杂志进行了一项研究，以说明在调查中措词不同的效应，就像

下面两个问题所表述的：

- (1) 你个人认为流产是错误的吗？
- (2) 不管你个人对流产的观点是怎么样，你同意还是反对一个女性在她的医生建议下选择流产？

对于第一个问题，被调查者中57%回答说是，36%回答说否。对于第二个问题，被调查者中的69%同意这个选择，只有24%反对这个选择。讨论这两个问题为什么会产生不一样的结果。如果选择其他人群来回答这个问题，结果会是怎么样的。

22. 税收或开支 一项盖普洛调查问到人们如下两个问题：

- (1) 你同意减税还是在其他政府组织中增加开支？结果75%回答是减税。
- (2) 你同意减税还是增加新的退休养老金，就和增加教育、国防、医药卫生和其他项目的一些开支一样？结果60%同意增加开支。

你想知道些什么？练习题23~26列出了可能来自统计研究的两个相关问题，简要地讨论这两个问题有什么不同，并且说明这些不同会怎样影响研究目标和研究设计。

23. 网络约会
第一个问题：由网络约会而至结婚的人数百分比是多少？
第二个问题：百分之多少的婚姻开始于网络约会？
24. 全职教育
第一个问题：校园中的基础课程有百分之多少是由全职教师来任教的？
第二个问题：全职教师中的百分之多少担任基础课程的授课？
25. 酗酒
第一个问题：大学生经常酗酒吗？
第二个问题：酗酒的人经常是大学生吗？
26. 统计课程
第一个问题：上过一门统计课程的大学毕业生的比率是多少？
第二个问题：所有统计课程中给大学生开设的比率是多少？

准确的标题：练习题27和28给出了在报纸头条发表的统计新闻故事的一个标题及其简要描述，在每一个例子中，讨论标题是否准确

地表达了故事。

27. 标题: 98% 的电影中出现了吸毒镜头。

故事概要: 一项政府研究声称, 吸毒、饮酒和抽烟在 98% 的电影内容中有所描述。

28. 标题: 性比工作更重要。

故事概要: 一项调查发现, 500 位被电话访谈的人中 82% 将满意的性生活放到了重要和非常重要的位置上, 79% 的人则将工作满意度放到了重要和非常重要的位置上。

字数统计: 政治家通常认为, 应该让他们的政治声明非常简短, 因为听众的注意力集中时间是非常短的。在统计新闻报道中也会发生同样的效应, 主要的统计研究经常会被概括为一两句话。练习题 29~32 是关于各种新闻资料中的统计报告的概括, 说明在给出的陈述中, 什么样的重要信息被遗失了, 以及什么是你想知道更多的。

29. 军事上的自信心 《今日美国》上发表了一篇哈里斯的调查报告宣称, “对军事领导人具有巨大信心”的成年人百分比达到了 54% (从 1997 年的 37% 提高到 2007 年的 54%)。

30. 高级酒店 美国有线电视新闻网络 (CNN) 报告了扎格特对美国高级酒店的调查, 结果发现, “最高评级为 30 级, 只有 9 家酒店达到了罕见的 29 级, 而这些酒店没有一家是在纽约城的。”

31. 天气预报 一则《今日美国》的报道标题说“很多公司尝试信任天气预报的结果”, 文章给出了很多公司相信长期的天气预报是可信的例子, 其中 4 个公司被特别引用。

32. 中国的人口出生 《今日美国》的一则报告标题说“中国男女婴出生的平衡被打破了”, 并且通过图表显示, 每 100 名女孩出生就会有 116.9 个男孩出生。

网络和远程项目

请在网址 www.aw.com/bbt 上打开第 1 章的有用链接 “Links for Internet Projects”。

33. 分析一个统计研究 寻找一些近期引起你关注的统计研究报告中的细节, 应用本节中所给出的八条指导原则写一个简短的报告 (在你分析的某个特定的研究中, 有一些指

导原则可能没有被应用, 在这种情况下, 解释指导原则为什么没有被应用)。

34. 哈勃指数 上网寻找哈勃指数并且检查几个近期被印证的统计数字, 如果有可能, 获取统计的原始资料, 选择 3 个你特别感兴趣的统计并且讨论, 基于本节所给出的指导原则, 你是否相信它们。

35. 双生子研究 研究者从生物学、心理学和社会学角度对双生子现象做了大量统计研究, 双生子经常被用来研究他们的某一特质是遗传自父母还是从出生后的环境获得的。同卵双生子来自母亲的同一个卵子, 有着相同的遗传物质, 异卵双生子来自母亲的两个独立的卵子, 他们的遗传物质有一半相同。寻找一个公开发表的双生子研究, 讨论同卵双生和异卵双生怎么样被分成案例组和控制组。应用指导原则 1~8 条, 研究和评论你发现的报告中的结论是否有冲突。

36. 美国的人口统计学 查阅《美国人口统计数据》的一个主题, 一本用来概括美国人生活方式的统计研究的非学术性杂志。选择某一篇文章, 应用这一节的理念, 评估和概括这个研究, 确定你所相信的印证信息是有所遗漏的, 应该提供一个更完整的信息。

新闻中的统计学

37. 应用指导原则 寻找一篇近期的你感兴趣的统计研究的新闻或电视报道。应用本节给出的八条指导原则写一个简短的报告。

38. 让人信服的结果 寻找一篇近期的新闻报道, 上面的研究结论你相信它是有意义的和重要的, 用少于一页纸的篇幅概括这个研究, 并且解释为什么你认为它是可信任的。

39. 不可信任的研究结果 寻找一篇关于统计研究的新闻报道, 你不相信它的结果是有意义的和重要的, 用少于一页纸的篇幅概括这个研究, 并且解释你为什么认为它是不可信任的。

40. 自选择调查 寻找近期的一个调查报告的例子, 所用的样本是自选择的, 描述样本是怎么选择的, 并且说明自选择是如何影响调查结果的。

第1章复习题

- 家庭中的枪支 一项盖洛普民意测验中包括对所选择的美国成人中的一个调查，他们被问到这个问题：“你家里拥有枪支吗”，在收到的1012个回复中，38%回答说是，报告中的误差幅度是3个百分点。
 - 在确定所有家庭拥有枪支比例可能取值的全距的基础上，解释误差幅度。
 - 确定总体。
 - 这是一项观察研究还是一项实验研究？说明理由，确定关注变量。
 - 报告中所提到的38%是一个总体参数还是样本统计量？为什么？
 - 如果你知道调查中的被试是在看到杂志中的文章后要求他们电话回复答案的一些读者，你认为这个调查结果是有有效的吗？为什么？
 - 描述采用随机抽样方法来选择调查被试的过程。
 - 描述采用分层抽样方法来选择类似的调查被试的过程。
 - 描述采用分组抽样方法来选择类似的调查被试的过程。
 - 描述采用系统抽样方法来选择类似的调查被试的过程。
 - 描述采用便利抽样方法来选择类似的调查被试的过程。
- 简单随机样本 这一章的一个重要元素就是简单随机样本的概念。
 - 什么是一个简单随机样本？
 - 美国劳动统计局实施了一项调查，它开始先将美国成年人口分成2007个部分，称之为基本抽样单元。假定这些基本抽样单元都包含同样数目的美国成年人。如果你随机地从每一个基本抽样单元中挑选一个人，将他们组成一个样本的话，是不是一个简单随机样本？为什么？
 - 依据第2小题中所说的基本抽样单元，描述一个结果为简单随机样本的抽样计划。
- 辛伐他汀片试验 在用于治疗高水平胆固醇的药物辛伐他汀片的一个临床实验中，观察1583位被试服用辛伐他汀片的副作用。结果发现3.5%的辛伐他汀片服用者感到头疼。
 - 基于已知信息，你能得出结论说在某些情况下，辛伐他汀片会导致头疼吗？为什么？
 - 当1583位被试给予辛伐他汀片治疗的同时，另外157名被试则给予了安慰剂，5.1%的安慰剂组被试感到了头疼。这个附加的信息可以提示说服用辛伐他汀片会出现头疼的副作用吗？
 - 在这个临床试验中，盲试是什么样的？为什么采用这种方式来检测辛伐他汀片的副作用是重要的？
 - 这个临床试验是一项观察研究还是一项实验研究？说明理由。
 - 什么是实验者效应，这个效应怎样才能最小化？
- 调查问卷的措词 在《超级调查者》中，戴维W. 穆尔描述了一个实验，在这项实验中被试被问到他们是否同意如下的表述：

第一种表述：太少的钱被用于福利事业了。

第二种表述：太少的钱被用于对穷人的帮助了。

即使一个人属于穷人也接收到了福利，当使用福利这个措词时只有19%的人表示同意。但是当使用对穷人的帮助这样的措词时会有63%的人表示同意。

 - 两个问题中的哪一个应该被用于调查中？为什么？
 - 如果你为一位保守党的国会议员候选人工作，你想强调反对使用联邦基金去扶持穷人，你将使用两个问题中的哪一个？为什么？
 - 故意在问卷中采用某些措词影响被试的回答是符合伦理道德的吗？为什么？

第1章小测验

- 对于下面每一个问题选择一个最佳答案，用一个或几个完整的句子来说明你的理由。
- 你实施了一项民意测验，从得克萨斯州随机选择了1000名注册选民，并且问他们是否

- 赞成他们现任州长的工作。这个研究的总体是：
- (1) 得克萨斯州的所有注册选民。
 - (2) 你所调查的1000人。
 - (3) 得克萨斯州的州长。
2. 在练习题1的民意测验中，如果你选择如下样本中的哪一个最可能会出现偏差。
- (1) 得克萨斯州的所有注册选民。
 - (2) 拥有得克萨斯州驾照的所有人员。
 - (3) 所有给州长竞选广告捐款的人。
3. 当我们说这个样本是总体的代表性样本时，我们的意思是指：
- (1) 样本中所发现的结果与我们在全部总体所计算的那些结果是一样的。
 - (2) 样本非常大。
 - (3) 样本是用一种可能的最佳方法挑选出来的。
4. 思考一个实验设计，观察现金激励能否促进学校出勤率。研究者选择了两组各100名高中学生。他给予其中一组学生每周全勤10美元的奖励，他告诉另一组学生他们是实验的一部分但是没有给以他们任何激励措施。那些没有接受激励的学生代表的是：
- (1) 实验组
 - (2) 控制组
 - (3) 观察组
5. 练习题4中所描述的实验属于：
- (1) 单盲实验
 - (2) 双盲实验
 - (3) 公开实验
6. 安慰剂的目标是：
- (1) 预防实验参与者知道他们属于实验组还是控制组。
 - (2) 在随机-控制研究中区分是随机因素还是控制因素。
 - (3) 决定一种疾病如果没有治疗能否治愈。
7. 如果我们在一个实验中看到了安慰剂效应，实验是用来检验一种新的疗法能否治愈疣状物：
- (1) 没有恰当地使用双盲实验。
 - (2) 实验组的人数太少了。
 - (3) 控制组的一些成员中疣状物被治愈了。
8. 一个实验属于单盲实验，如果：
- (1) 它缺少实验组。
 - (2) 它缺少控制组。
 - (3) 实验参与者并不知道他属于实验组还是控制组。
9. X民意测验指出鲍威尔的支持率是49%，同时Y民意测验也指出他的支持率是53%。两个民意调查误差幅度都是3个百分点。
- (1) 两个民意调查中的一个一定是进行得不充分。
 - (2) 两个民意调查彼此是一致的。
 - (3) 鲍威尔的支持率是51%。
10. 一项调查显示12%美国人相信埃维斯·普里斯利仍然活着。误差幅度是4个百分点。这个调查的置信区间是：
- (1) 10%~14%
 - (2) 8%~16%
 - (3) 4%~20%
11. 一项美孚石油公司实施的研究显示阿拉斯加的一个大的漏油事故没有持续的危害。这个结论：
- (1) 当然是无效的，因为这个研究是有偏差的。
 - (2) 也许是正确的，因为潜在的偏差意味着你会更认真地分析他的结论是怎么得到的。
 - (3) 如果它落在研究的置信区间内可能是正确的。
12. 电视节目《美国偶像》从任何想投票的人员中选择投票的获胜者，这意味着获胜者
- (1) 是大多数想赢的美国人中的一个人。
 - (2) 可能是也可能不是大多数想赢的美国人中的一个人，因为投票会招致参与者偏差。
 - (3) 可能是也可能不是大多数想赢的美国人中的一个人，因为投票应该是双盲的。
13. 考虑一个实验中要测量一个6岁儿童的体重，这一研究中的关注变量是：
- (1) 样本容量
 - (2) 6岁的儿童体重
 - (3) 研究中的儿童年龄
14. 设想一个调查，随机挑选那些被发现在过去一年中使用防晒油却反而被晒伤更严重的人们。下面关于这个结果的解释哪一个更适合？
- (1) 防晒油是无用的。

- (2) 研究中使用防晒油的人们，所使用的防晒油是过期的。
- (3) 使用防晒油的人们可能会花更多时间在太阳底下的。

15. 从每个可能的角度仔细设计一项统计研究，

那么：

- (1) 结果一定是正确的。
- (2) 我们可以信任这个结果，但它仍然有可能是错误的。
- (3) 我们说这个研究是绝对有偏差的。

| 聚焦社会学 |

日托培养的方式增加攻击性了吗

数十年前，大多数母亲会居家照顾他们的孩子。但是随着越来越多的女性参加工作，越来越多的孩子会在各种类型的日托机构度过他们的清醒时分。1975年，6岁以下孩子的母亲在社会上做全职或兼职工作的不足40%。现在，超过60%的母亲参加工作。日托儿童数量的急剧增长引起了心理学家和社会学家的关注，他们研究了日托的效果。美国联邦儿童健康和人类发展研究所（NICHD）在1991年开始了儿童保育的研究，他们追踪了1300多名在日托培养的儿童，追踪持续到了这些儿童15岁，这些儿童来自美国10个不同的研究地点。注意这种研究启动模式使得它成为一个前瞻性的观察研究（见1.3节）。



很多著名的学者参与了这项实验，但结果仍然备受争议。比如，依据2001年发表的这项研究报告的结果：儿童日托时间越久，他（她）长大后攻击性也就越强。媒体对日托及其可能带来的不良影响所做的广泛报道，让职业女性经常因内疚而感到痛苦，因为她们担心日托对她们的孩子产生不良影响。但我们应该相信这个结论吗？有很多方法去评估这个结论，出于可操作性原则，我们采用1.4节里提到的八条指导。

指导一：这个研究的目标看起来很明显（研究日托的效果），研究总体是在日托中的所有儿童。然而这种类型的研究存在一个潜在的问题，这是一个观察研究而非随机实验。因为控制父母是否送儿童去日托是不符合伦理原则的。然而可惜的是，观察研究只能建立变量间的联结（或相关），它不能分析事物的原因和结果（见第7章）。在这个案例中，研究建立了日托和攻击行为的关联，但是却不能说明它们之间的因果关系。问题在于，因为由父母选择儿童是在家还是进入日托，我们都不能确定样本和对照组的儿童是可以真正进行比较的。例如，有这样的可能性，必须哄孩子进入日托的母亲工作压力可能更大，这样使得日托儿童比待在家庭中的儿童承受的压力更大，或者由于儿童比较易吵闹才被送入日托机构。因此，不管日托和攻击性的关联是否有效，这种观察研究的属性都不可能支持“日托导致攻击性”的这个结论。

指导二：这项研究实施得很认真，就研究人员这方面而言，我们没有任何理由去怀疑其存在独有的偏见。尽管如此，日托是一个带有情感意味的问题，研究人员可能抱有强烈的个人看法，这就会存在潜在的偏见。事实上，不同的研究人员对研究结果的解释也是不一样的，这刚好说明偏见产生于对研究结果的解释而不是针对研究本身。

指导三：研究者采用分层抽样的方法去招募被试参与试验，因为他们想从人口统计学的角度使被试能够匹配，考虑的因素包括社会背景、种族、家庭结构这些方面。这样抽样的方法看起来更适合这个研究。但是被招募的家庭有权利决定是否参与这个研究。这种自选择会使得被试偏见产生的可能性增加了，这种偏见可能会影响研究的结果。

指导四：这个关于日托和潜在攻击性的研究包含两个变量：一个儿童在日托度过的时间总量和儿童攻击性的水平。这两个变量都很难界定。例如儿童保育研究定义：日托的意思是除了母亲以外的任何人的照料，这意味着父亲和祖父母的照料都属于日托。很多人反对这样的定义。定义攻击性也很困难。例如，分清动作性游戏和真正的攻击就很困难。

指导五：在儿童保育研究中有很多潜在的混淆变量，正如之前提到的，研究并没有提供下面的数据，孩子的母亲如果是职业女性比起居家的母亲是否压力更大？这个变量却可以干扰研究结果。可是让人担忧的是，没有一个可行的方法去测量或者控制日托的质量，因此，甚至一些参与此项研究的人员也提出说，观察到的儿童攻击性行为是较差日托的一个结果，而不能归因于一般意义上的日托本身。

指导六：研究人员采用面谈的方式确定日托时间和类型以及每个儿童的攻击性水平。但由于攻击性是很难定义的，并且每个人的看法也都不一样。没有精确的方法保证所有的访谈者以同样的方式准确地提问（不仅是用词，也包括表情和语调），也不能保证所有的被访谈者都能同样理解问题。

指导七：儿童保育研究者发表的这篇报告承认研究可能存在潜在的瑕疵，并且强调这个研究只是呈现了攻击性和日托可能的联结。日托会产生攻击性这个结论来自哪里？这种表述方式来自媒体。因为结果并不支持这样的解释，这样的新闻报道明确地反应不是偏见就是误解。例如没有学过统计学的新闻记者可能不知道两个变量间关联和因果的区别。另外一种可能，就是媒体故意耸人听闻以吸引更多的眼球。在某些情况下，很多认为女性应该居家的人们可能会故意曲解这个结果，以便为他们的观点找到科学依据。

指导八：有关这个问题的研究具有非常重要的实践意义。从个体层面上来说，很多职业女性不相信关于日托有可以供她们选择的余地，由于财务和职业的原因她们必须工作，因此她们会更可能选择日托，即使可以证明居家照顾儿童会更好。从社会层面上来说，这个发现也可能具有十分重要的现实意义，它会使得政府制定新政策来促进儿童保育，但是只要这个结果不能从因果的角度被证明，它就不能给政策制定者清晰明确的指导。

总之，我们发现，一个充满刺激性的有关日托和攻击性的新闻头条，在仔细检视下崩溃了。这个研究仍然可以是非常有价值的，尤其是研究者自己已经意识到了研究中潜在的瑕疵。这个研究会导致将来的研究带来更多具有说服力的结果，研究结果中的信息将会使很多工作中的父母充满感激。但是在那之前，我们需要详细了解应如何解释儿童保育的研究结果，更应该小心应对那些发表于媒体头条耸人听闻的研究结果。

问题讨论

1. 第三个潜在的混淆变量是儿童看电视和玩电子游戏的时间，这个变量也与攻击性行为有关。你想过这个变量会影响儿童保育研究的结果吗？假如想过，它是怎样影响的呢？还有其他混淆变量会影响研究结果的吗？请清晰地解释你的答案。
2. 假设你正在设计一个后续的研究，试图得到更明确的结果，基于你了解的儿童保育研究，你的研究与其有什么区别？
3. 就你个人的观点，你认为父母是居家照料儿童还是出去工作哪个更好？如果儿童保育研究得出确定性的结论，你的意见会受其影响吗？为什么？
4. 你认为儿童保育研究花费如此多时间和费用物有所值吗？说明理由。

建议阅读

1. 阅读更多儿童保育研究 (SECC)，请登录网页 <http://secc.rti.org.carey>。
2. Carey, Benedict. "Study Finds Rise in Behavior Problems After Significant Time in Day Care." *New York Times*, March 26, 2007.
3. Stolberg, Sheryl Gay. "Science, Studies, and Motherhood." *New York Times*, April 22, 2001.
4. Sweeney, Jennifer Foote. "The Day-Care Scare, Again." *Salon.com*, April 20, 2001.

| 聚焦公众健康 |

你的生活方式健康吗

从统计研究的角度来思考下面所陈述的研究发现：

- (1) 吸烟会增加患心脏疾病的风险。
- (2) 经常食用黄油会增加患心脏疾病的风险。
- (3) 每天一杯红酒可以预防心脏疾病，但却增加了患乳腺癌的风险。

你可能已经很熟悉这样的结论了，甚至你也很可能根据这些发现已经改变了自己的生活方式。但是这些发现是怎么得到的？值得注意的是，这些以及数百项有关公众健康的重要发现均来自同一项庞大的前瞻性研究，这个研究包含了数百项小型的统计研究。这就是著名的哈佛护士健康研究，它是已经实施的最持久的公

众健康研究了。如果它还没有改变你的生活方式，在将来一定会影响你的。

哈佛护士健康研究于1976年开始实施，当时哈佛医学院的一位教授弗兰克 E. 斯佩萨 (Frank E. Speizer) 打算研究一种口服避孕药的长期效果。他发放了将近 370 000 份问卷，回收到了多于 120 000 份。他之所以调查护士，是因为他相信护士的医学训练会使得其反应的答案比普通人群更加可信。

弗兰克 E. 斯佩萨博士和他的同事们筛查收回的问卷数据时，他们认识到这个研究不仅仅能说明避孕药的效果。今天，这个研究团队继续追踪 120 000 个初始被试中的 90%。上百个研究者使用这些数据研究公众健康。

每年的问卷仍然是研究中很重要的一部分。这些问卷让研究人员了解了很多情况，护士们吃什么食物；服用了什么药物和维生素；是否参与体育锻炼，多久运动一次；饮酒、吸烟的情况；曾经患过的疾病。一些护士还提供了他们的血液样本，这可以用来测量他们的很多指标。例如胆固醇水平、激素水平、基因还有农作物残留农药和环境对他们身体的危害水平。斯佩萨博士对护士的信任被证明是合理的，他们总能可靠地完成调查，并且几乎都能正确按照要求抽取血液样本并做好标记。

经过 30 多年的通信联络，研究者和护士彼此都感觉很亲近。很多护士期待着听到研究者的消息并且说研究帮助他们更加注意怎样生活。当研究者必须记录其中的护士的死讯时也会感到深切的悲伤。

这种悲伤的死讯会在这个研究中日益增多，因为这些护士中好多人年龄已经七八十岁了。但这种悲伤同时也提供了具有非常珍贵价值的数 据，涉及老年人健康长寿的影响因素。研究者希望这些数据可以指出一条明路，以便我们确切地理解以下问题，比如：健康的饮食由什么组成；环境污染和化学接触是怎样影响健康的；我们怎么预防退行性疾病如骨质疏松症和阿尔茨海默症等。直到这 120 000 名女性的生命全部走到终点，哈佛护士健康研究带给未来人们的礼物就是学会如何更好地生活。

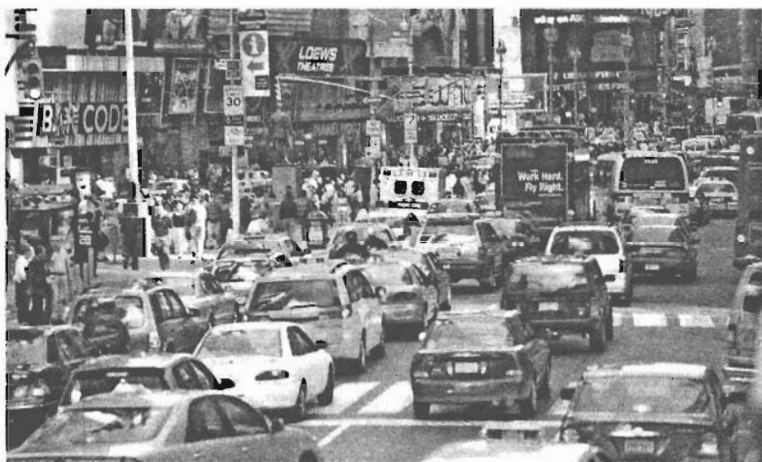


问题讨论

1. 思考一下哈佛护士健康研究未来一二十年间可能发现的结果，什么类型的结果你认为是最重要的？你认为这些发现会改变你的生活方式吗？
2. 解释为什么哈佛护士健康研究是一个观察研究？批评者指出如果这个研究包含某些实验的话会比观察更有效，讨论采用符合伦理道德规范以及可操作的实验方法收集同样的数据是否可行？
3. 从理论角度讲，哈佛护士健康研究是有被试偏差的，因为发放的 370 000 问卷只收回了 120 000 份，研究者是否应该考虑这个问题？为什么？
4. 另外一个潜在风险来自问卷中经常涉及有关个人健康的一些敏感问题，研究者没有办法保证护士的回答都是诚实的。你认为不诚实的回答会让研究得出不正确的结论吗？阐明你的观点。
5. 哈佛护士健康研究的参与者都是女性，你认为得到的结论适用于男性吗？为什么？
6. 在网络上搜索讨论哈佛护士健康研究结果的新文章，选择一个你感兴趣的近期结果，讨论它意味着什么以及它可能怎样影响公众和你自己未来的健康。

建议阅读

1. 阅读关于哈佛护士健康研究浏览网站：<http://www.nurseshealthstudy.org>.
2. Brophy, Beth. "Doing It for Science." *U. S. News & World Report*, Vol. 126, March 21, 1999, p. 67.
3. Conwell, Vikki. "Forever Friends: Ties Can Last a Lifetime and Sustain Body and Soul." *Atlanta Journal-Constitution*, January 2, 2007.
4. Parker-Pope, Tara. "Drinking Has Hidden Health Risks for Women." *Wall Street Journal*, December 26, 2006.
5. Yoon, Carol Kaesuk. "In Nurses' Lives, a Treasure Trove of Health Data." *New York Times*, September 15, 1998.



事实上，当人们谈论新闻中的数字时，没人知道自己在讲什么。因为我们总是在引用连自己都不知道自己在讲什么的人的话，如政客和股市分析师。

——莫莉·艾文斯
报纸专栏作家

第2章 统计测量

学习目标

■ 2.1 数据类型和测量水平

能够明确定性和定量数据，测量水平（称名、顺序、等距或比率）。

■ 2.2 标准控制

理解随机误差和系统误差的差别，能够采用绝对和相对的方式描述误差，掌握测量中准确性和精确度的差异。

■ 2.3 百分数在统计学中的应用

理解百分比在报告统计结果时的用途，认识到错误应用它们的情况。

■ 2.4 指数

理解指数的概念，特别是常用于通货膨胀测量的消费者价格指数（CPI）。

我们都知道怎么样测量诸如身高、体重和体温的数量，在统计研究中有许多种测量。我们必须确定这些是可定义的、可获得的和仔细报告的。在本章中，我将讨论一些与测量有关的重要概念，正如你所看到的，对于理解量化你日常生活的统计报告来说，这些概念是非常有用的。

2.1 数据类型和测量水平

统计学的一个挑战是如何完美地概括和呈现数据，不同类型的数据需要不同的概括。在本节，我们会讨论数据是怎么样分类的，这对于后面章节中概括和呈现数据是非常有帮助的。

2.1.1 数据类型

数据有两种基本类型：定性和定量。定性

数据（qualitative data）的取值为一种非数字的类别（由于此原因，定性数据有时也叫做分类数据）。例如，眼睛颜色是一个定性数据，将眼睛颜色分为蓝、棕和黑色。其他一些定性数据包括喜爱的冰激凌、雇用者的姓名、动物的性别、电影或饭店的“星级”（诸如3星或4星）。但是这些数字不是必要的并且也不能用来计算，我们同样也可以用优、良、中、差这4个非数字的类别来对电影进行分级。

定量数据（quantitative data）的取值是以计数或测量的结果来表达的。运动员完成田径比赛的用时、大学毕业生的收入、不同班级的学

生人数都是定量数据的例子。

数据类型

定性数据，或者也叫分类数据 (categorical data)，是指其取值为一种非数字的类别。

定量数据是指取值表示计数或测量的结果。

例 2-1 数据类型

区分下列数据属于定性数据还是定量数据：

- (1) 一个消费者调查涉及的鞋子品牌名。
- (2) 一个选择题测验的分数。
- (3) 一个论文设计以字母表示的分数。
- (4) 一个篮球队里用来表示某个球员的球衣上的数字。

解答

- (1) 品牌属于分类变量，表示的是定性数据。
- (2) 一个选择题测验的分数属于定量数据，因为它表示的是正确回答的题数。
- (3) 一个论文设计的以字母表示的分数属于定性数据，因为它表达了完成任务的不同类别（从失败到优秀）。
- (4) 球衣上的数字属于定性数据，因为数字表示的不是计数或测量的结果，它们只是用来表示唯一的标志。你可以说这些数字属于定性数据而非定量数据，因为你不能用这些数字进行计算。比如，对不同球员球衣上的数字进行加法或减法是毫无意义的。

2.1.2 离散数据与连续数据

定量数据可以进一步分为离散数据和连续数据。在一个给定的区间内可以任意取值的数据就是连续数据 (continuous data)，例如，一个人的体重可以从 0 到数百磅之间的一个值，所以表示体重的数据是连续的。离散数据 (discrete data) 指的是取值只能是特定的，之间的值没有意义，例如，学生的人数是离散数据，因为只能是整数，鞋的号码也是离散数据，因为只能是整数和半整数，诸如 7 ， $7\frac{1}{2}$ ， 8 和 $8\frac{1}{2}$ 。

离散数据与连续数据

连续数据在给定的区间内可以取任意值。

离散数据只能取特定的个别值，相邻的两个值之间没有其他取值。

例 2-2 离散数据还是连续数据

指出下列每种数据是离散的还是连续的：

- (1) 测量步行 1 公里所花的时间。
- (2) 日历上的年份（比如 2007，2008，2009）。
- (3) 不同牧场的奶牛数。
- (4) 一个农场的牛奶生产量。

解答

- (1) 时间可以取任何值，所以时间的测量属于连续数据。
- (2) 日历的年份是离散数据，因为它不可以有小数，例如，在 2009 年的除夕，年份将从 2009 年前进到 2010 年，我们不能说是 $2009\frac{1}{2}$ 年。
- (3) 每一个农场都有我们能计数的全部奶牛数，因此这个数据是离散的（例如，你不能说牛的数量有小数）。
- (4) 一头牛生产的牛奶数量可以是一个取值区间的任意值，所以挤的牛奶量是连续数据。

2.1.3 测量水平

另一个数据分类的方式是测量水平。最初级的测量水平应用于诸如眼睛颜色、喜爱的冰激凌、动物性别等此类变量。这些变量仅仅以名称、标志或分类来描述。我们称这样的数据是测量的称名水平 (nominal level of measurement)，术语称名指的是对分类变量命名。测量的称名水平的数据没有等级和顺序。例如，我们不能说蓝眼睛的等级比棕眼睛高，也不能说香草味的冰激凌顺序排在巧克力味的冰激凌前面。

当我们以等级或顺序的方式描述数据时，诸如电影或酒店的星级，我们采用的是测量的顺序水平 (ordinal level of measurement)，术语顺序是指对变量排序。这样的数据通常不能计算平均数等量数，例如，将电影的星级加起来是无意义的，看 3 场 1 星级的电影并不等于看 1 场 3 星级的电影。

思考题

分析这样一个调查问题“你最喜爱的冰激凌的口味是什么样的”，我们可以说冰激凌的口味表达的是称名水平的数据。为了方便起见，我们假定研究者录入研究数据时以不同的数字代表不同的口味，1 = 香草味，2 = 巧克力味，3 = 奶油曲奇，诸如此类，这样是否将冰激凌口味数据从称名水平改变成了顺序水平？为什么？

测量的顺序水平提供了一个等级系统,但是它并不能使我们确定测量之间的明确差异。例如,没有办法可以界定一个3星级电影和一个2星级电影的准确差别。相反,81°F比80°F高出的1°F,相比28°F比27°F高出的1°F是等值的,温度的数据是一个较高水平的测量,因为一个温度单位的间距(距离)总是意味着同样数量。不管怎样,华氏温度的等距(比如减法)是有意义的,但比率(比如除法)则无意义。例如,20°F比10°F热2倍,-40°F要比-20°F冷2倍这样的说法是不成立的。原因在于华氏温度值的比率是无意义的,它的零点是人为规定的,并不代表“没有温度”的含义。如果等距是有意义的而比率是无意义的,我们说这种数据是测量的等距水平(interval level of measurement)。

当等距和比率都是有意义时,我们说这样的数据是测量的比率水平(ratio level of measurement)。例如,距离的数据是测量的比率水平,因为10公里的距离确实是5公里距离的2倍。通常,测量的比率水平适用于具有真正零点的数据,真正零点的值意味着“没有”东西被测量到,在距离的例子中,零距离意味着“没有距离”。其他测量的比率水平数据还包括体重、速度和收入等。

注意称名和测量的顺序水平数据总是定性的,等距和测量的比率水平数据总是定量的(当然也可以区分为间断的和连续的),图2-1概括了可能的数据类型和测量水平。

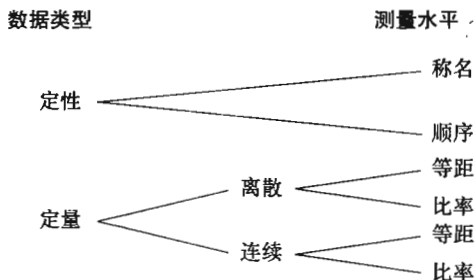


图2-1 数据类型和测量水平

测量水平

测量的称名水平是指数据的属性只包括命名、标志或者分类。数据是定性的,数据没有等级也不能排序。

测量的顺序水平应用于能够以某种规则排序的

定性数据(例如从低到高)。以这种顺序水平测量的数据计算通常是没意义的。

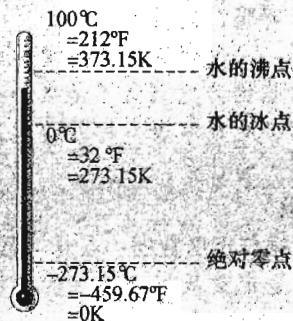
测量的等距水平应用于间距有意义但是比率无意义的定量数据,这个水平的数据具有相对零点。

测量的比率水平应用于间距和比率都是有意义的定量数据,这个水平的数据具有绝对零点。

相关知识

科学家经常采用开尔文绝对温标来测量温度,开尔文绝对温标属于测量的比率水平,因为开尔文绝对温标具有绝对零点。开尔文绝对温标零度是可能的最冷温度,也叫做绝对零度,开尔文绝对温标0度大约等于摄氏-273.15°C或者华氏-459.67°F(开尔文绝对温标通常不使用单位符号)。

温度刻度



例2-3 测量水平

判断下面每一种数据的测量水平(称名、顺序、等距、比率)。

- (1) 篮球队运动员球衣上标明的数字。
- (2) 学生给自助餐厅的食物排序优、良、中、差。
- (3) 某个历史事件的日历年份,诸如1776, 1945或者2001。
- (4) 摄氏温度。
- (5) 波士顿马拉松运动员的成绩。

解答

- (1) 如同在例2-1中所讨论的,篮球队运动员球衣上标明的数字不能计数也不能测量任何事物,它属于测量的称名水平,因为它只是标志,并不代表任何顺序。
- (2) 一种等级系列表明数据为测量的顺序水平,因为分类(优、良、中、差)包含一种特定的顺序。
- (3) 相同日历年份的间距总是相同的意思,但是年份的比率是没有意义的,因为年份0是人为规定的,并不意味着时间的开始,日历年份属于测量

的等距水平。

- (4) 像华氏温度一样，摄氏温度也属于测量的等距水平， 1°C 的间隔总是相同的，但是，零点 (0°C = 水的冰点) 是人为规定的，并不意味着

没有温度。

- (5) 马拉松时间的比率是有意义的，例如，6 小时是 3 小时的 2 倍时间长短，因为 0 小时的时间是一个绝对的零点。

2.1 练习题

统计素养和批判性思维

1. 定性数据与定量数据 定性数据和定量数据之间有什么不同？
2. 定性数据和定量数据 一个足球运动员选修了一门统计课程，这名运动员的名字和他所在球队的名称是定性数据，能否通过在他们统一的运动服上进行编号，从而使这些定性数据变成定量数据？这样正确吗？为什么？
3. 定性数据和定量数据 当一个研究者争论说所有的数据既是定性数据也是定量数据，他这样说正确吗？说明原因。
4. 邮政编码 一位研究者争辩说美国邮政编码是定量数据因为它测量了地区，小数字表示东边地区，大数字表示的西边地区。他的说法正确吗？为什么？

概念与应用

定性数据与定量数据：在练习题 5 ~ 16 中，下面描述的是定量数据还是定性数据并解释为什么。

5. 血型 血型分为 A、B、O、AB 等 4 种。
6. 体重 在一种新药的临床试验中被试的体重。
7. 身高 在一种新药的临床试验中被试的身高。
8. 电影 电影类型（戏剧、喜剧等）。
9. 电影时长 电影时间长度（以分钟计）。
10. 对调查内容的回答 当被试在调查中被问到一个问题时所做的反应（是，否）。
11. 尼尔森电视收视率 尼尔森媒体研究所调查的家庭正在收看的电视节目。
12. 尼尔森电视收视率 尼尔森媒体研究调查时正在使用电视的家庭数量。
13. 拒绝调查 在进行调查时拒绝回答问题的人数。
14. 鞋子的号码 实验被试鞋子的号码（比如 8 和 10.5）。
15. 薪资 全美政府官员的薪水。
16. 区号 调查被试电话号码所在地区的编码（比如 617）。

离散变量与连续变量：在练习题 17 ~ 28

中，描述下列的实际变量是离散变量还是连续变量并解释原因。

17. 行人交通信号灯按钮 在纽约市，在十字路口有 3 250 个行人信号灯按钮，其中 2 500 个并不能使用（数据来源于《纽约时报》的文章“*For Exercise in New York Futility, Push Button,*” by Michael Luo）。
18. 一美分硬币重量 近期铸造的一美分硬币的重量。
19. 一美分硬币数量 美国西尔斯百货公司每天所收集的硬币数量。
20. 盖洛普调查 在盖洛普公司不同时间的调查中，被试被问到他们是否拥有自己的枪支，在一段时间内记录下来的回答说是的人数。
21. 化学 一个重复的化学试验，所记录的化学反应发生所需要的时间。
22. 测验时间 学生完成一个统计测验所需要的时间。
23. 测验分数 一个统计测验的分数，测验包括 25 道选择题，分数就是正确回答题目的数量。
24. 交通流量 每小时穿过金门桥的汽车数量。
25. 汽车行驶速度 汽车穿过金门桥中心的行驶速度。
26. 电影星级 电影按照如下标准进行分级：0 星、1/2 星、1 星，依此类推。
27. 恒星数量 宇宙中每个星系的恒星数量。
28. 卡路里 足球运动员在一场比赛中所消耗的卡路里数量。

测量水平：针对练习题 29 ~ 40 中所描述的数据，说明它的测量水平是称名、顺序、等距还是比率水平的。

29. 身高 洛杉矶湖人队所有篮球运动员的身高。
30. 电影分级 电影评论的一种分级标准，将电影分为“必看”“良好”“一般”“差片”“烂片”五级。
31. 电影类型 电影类型（戏剧、喜剧等）。

32. 体温 一个统计课堂上的所有学生的体温。
33. 汽车 汽车按照吨位分为微型车、小型车、中型车和标准车四等。
34. 临床试验 临床试验结果包括“真阳性”“假阳性”“真阴性”“假阴性”四种。
35. 成绩 一门课程的最终成绩分为 A、B、C、D、E 五等。
36. 重量 美国联合包裹服务公司 (UPS) 今天处理的所有包裹重量。
37. SSN 社会保障号码。
38. 年份 民主党候选人赢得美国总统选举的年份。
39. 航空行李 美国联合航空公司 15 号线每天接受安检的行李重量。
40. 汽车安全水平 《消费者报道》将汽车安全水平分为从 0 = 不安全到 3 = 安全四等。
是否有意义的比率? 在练习题 41 ~ 48 中, 所给出的一种陈述是否表达了一种有意义的比率, 是不是属于比率水平测量的应用? 说明原因。
41. 电影分级 一部 4 星级的电影是一部 2 星级电影的两倍。
42. 马拉松比赛 10 公里的马拉松比赛路程是 5 公里的马拉松比赛路程的两倍。
43. 航空器飞行速度 一个航空器每小时飞行 450 公里是另外一个每小时飞行 150 公里的航空器速度的 3 倍。
44. 温度 8 月 6 日, 纽约的温度是华氏 80 °F, 9 月 7 日的温度是华氏 40 °F, 因此 8 月 6 日的温度是 9 月 7 日温度的两倍。
45. 鱼的重量 一条鱼发现它重量是 2 磅, 另外一条鱼的重量 4 磅, 所以较重的鱼是较轻的鱼重量的两倍。
46. 碳测定年代 通过碳测定年代, 一块木头的样本年代是另一块木头的两倍, 因为发现第

一块样本为 2 000 年前的, 而第二块样本是 1 000 年前的。

47. 薪资 一位雇员薪资为 150 000 美元, 另外一个雇员薪资为 75 000 美元, 前一位雇员是后面一位雇员薪资的两倍。
48. SAT 分数 一个人在 SAT 测验中的分数是 2 200, 是作为一个合格大学生所要求成绩 1 100 的两倍。

完成分类: 在练习题 49 ~ 56 中, 确定所描述的数据是定性数据还是定量数据, 并且说明数据的测量水平。如果数据是定量数据, 表明它是离散数据还是连续数据。并给出简要的解释。

49. 马拉松时间 纽约市马拉松比赛的完成时间。
50. 马拉松比赛名次 在马拉松比赛中运动员完成比赛的排名 (比如第 1 名、第 2 名、第 3 名等)。
51. 雇员的 ID 号码 随机生成 6 位识别码分配给 Teletronics 公司的雇员。
52. 雇员的工作年限 每一位雇员自从被雇用后在 Teletronics 公司工作的时间长短。
53. 被雇用年份 雇员被雇用的那一个年份 (比如 2003, 1998, 2007)。
54. 政治调查 在一个关于投票人偏好的调查中, 为方便记录对于政党的反应被编成号 1, 2, 3, 4, 5 (1 = 民主党, 2 = 共和党, 3 = 自由党, 4 = 保守党, 5 = 其他)。
55. 产品等级 《消费者报道》杂志将不同的计算机中的每一台进行评级, 分别评为“优秀产品”“推荐产品”“不推荐产品”3 级。
56. 质量控制 纽波特电子公司检测它所制造的每一台收音机, 并将每一台分别贴上不同的标签, 分别是“合格产品”或者“残次品”。

度对大量的误差进行讨论时, 它同样适用于预测或估计等, 诸如总体参数估计或预测一家公司的收入。



2.2 误差控制

你理解了测量的不同类型和水平, 我们开始讨论如何控制测量误差。首先, 我们考虑误差产生的各种情况, 然后我们讨论陈述结果时怎样计量可能的误差, 注意当我们从测量的角

2.2.1 误差类型: 随机误差和系统误差

大致说来, 测量误差分为两类: 随机误差和系统误差, 下面以一个例子说明这种差异。

假定你在医院小儿科工作, 要使用电子秤

测量婴儿的体重。如果你曾经和婴儿一起相处过，你就知道将他们放于一个电子秤上，他们通常会是很不高兴的，他们哭喊着试图摇晃天平，使得读数器上下波动。如图 2-2 所示，我们说这个摇晃产生了一个**随机误差**（random errors），因为特定的测量不是高就是低。

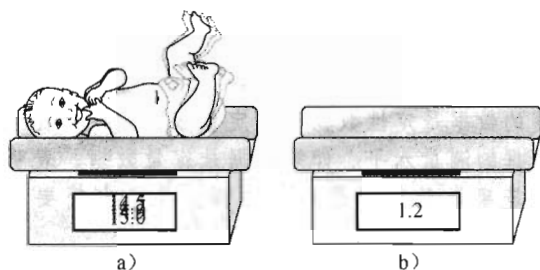


图 2-2 测量婴儿体重时的随机误差和系统误差
a) 婴儿的活动导致随机误差 b) 量具空着时读数却是 1.2 磅，这导致系统误差，使得所有测量偏高 1.2 磅。

现在假定你全天一直都在使用图 2-2 呈现的电子秤测量婴儿的体重，在一天的最后，你注意到了量具显示 1.2 磅可是上面却什么也没有，在这个案例中，你的每次测量都高出了 1.2 磅，这种类型的误差叫做**系统误差**（system error），因为它使得一个测量误差呈现系统性，一个误差持续地（系统地）影响所有测量结果。

相关知识

测量刻度与真值不一致所导致的一种系统误差也叫做**校准误差**。你可以将一个已知重量的物体放到量具上面对其进行校准，比如 0，5，10，20 磅的重量，这样就可以确保量具给出预期的读数。

两种类型的测量误差：

随机误差的发生是由于在测量过程存在随机的和内在的不可预测的事件。

系统误差的发生是指在测量存在着一个以同样方式影响所有测量的问题。

一个系统误差以同样的方式影响所有的测量，这样会使它全部偏高或者偏低。如果你发现一个系统误差，你可以返回头来调整所有受影响的测量。相反，不可预测的随机误差的性质使得它不大可能被校正。尽管如此，你可以

通过多次测量然后取平均值使随机误差的影响最小化。例如，如果你测量婴儿的体重 10 次，你的测量结果可能有时高有时低，通过将 10 次测量结果平均以后你将会得到一个更恰当的值。

思考题

打开一个能给出当前时间的网页（例如 www.time.gov），你的钟表或手表与它的时间相差多少？描述你的计时工具可能的系统误差和随机误差的来源。

相关知识

城市地区要比那些人类活动较少的地区更热这个事实叫做**城市热岛效应**。这个效应的主要原因包括机动车、家庭和工业生产大量燃烧燃料，以及人行道和高大建筑物吸收的太阳能热量。



例 2-4 全球气候变暖数据的误差

研究全球气候变暖的科学家需要掌握全球气温的平均值，或者说全球平均气温如何随时间而变化的。考虑到说明 20 世纪以前的历史气温数据存在两个困难：（1）使用简单的温度计测量气温并且手工记录；（2）大多数气温测量并做出记录的地点是在靠近城区附近的区域。在这些区域，人类活动释放的热量使得其气温要高于广阔的农村地区。讨论这两个困难中的每一个是否产生了随机误差或系统误差，思考这些误差的含义。

解答

第一个困难包含随机误差，因为人们使用温度计

时肯定会有一些偶然的误差, 比如在读温度计的刻度值时, 校正温度计时, 或者记录温度计的结果时。没有办法可以预知一个人在读温度计的刻度值时是准确的、偏高的还是偏低的。尽管如此, 在同一地区同一天多测量几次气温, 然后取它们的平均值就可以使随

机误差达到最小水平。

第二个困难包含系统误差, 因为城区过量的热量总是会使气温测量结果高于其他地区, 如果研究者可以估计到影响气温测量的系统误差是多少, 就可以校正这个问题影响的数据了。

案例研究

普查

美国宪法规定每 10 年实施一次人口普查, 美国人口普查局进行普查 (也进行其他人口统计学研究)。为了尝试人口汇总, 美国人口普查局依靠大量的调查数据, 假定对每个美国人都进行调查并且获得反馈。无论怎样, 许多随机误差发生在调查过程中。例如, 许多人填写表格不正确, 美国人口普查局的雇员对人们反应的记录不正确。

普查也受一些类型的系统误差的影响。例如, 对无家可归者发放问卷进行调查是非常困难的, 统计非外籍移民也是非常困难的, 因为暴露他们在美国的存在让他们非常不愿意。系统误差会导致人口的不完全统计。另一个多计的系统误差的例子是一些大学生的重复计数, 父母会报告他们, 所在学校也会报告。

2000 年的人口普查, 例如, 最初统计的人口数为 2.814 亿, 然而, 后续研究提出普查漏报了 760 万, 430 万人被重复报告了两次。如果这些统计结果是正确的, 普查中少计的人口数超出了 300 万。

思考题

美国人口普查局是否允许基于统计调查调整它的“官方”数据是一个非常有争议的问题。宪法需要一个人口的“真实计数”(见 2.2 节), 你相信阻止或允许在官方数据基础上采用统计调查的补充这样的做法吗? 也请讨论一下为什么在人口统计中民主党更倾向于抽样的方法而共和党却反对。

2.2.2 误差的大小: 绝对误差和相对误差

除了想了解一个误差是随机的还是系统的, 我们经常也想了解误差的大小。例如, 误差是否足够大到影响了测量结果或者误差很小根本不重要? 让我们来探讨误差大小的概念。

假定你想去商店购买你想要的 6 磅汉堡, 但是由于商店的磅秤校正不正确, 你实际上只得到了 4 磅汉堡。你会为这两磅的误差非常烦恼。现在假定你为一个集镇的肉野餐购买汉堡, 你订购了 3000 磅汉堡, 但实际上你只得到了 2998 磅。像前一个例子你同样少了两磅汉堡, 在这个案例中, 误差看起来不那么重要。这个简单的例子显示了一个误差的大小取决于你怎样看待它。

采用统计语言表述, 两个案例中少两磅的误差是**绝对误差** (absolute error), 它描述了声称或测量的数值与真值之间的差距, **相对误差** (relative error) 将绝对误差与真值进行比较。第

一个例子的相对误差相当大, 因为绝对误差两磅是真实重量 4 磅的 $1/2$, 我们说相对误差是 $2/4$, 或者 50%。相对而言, 第二个例子中的相对误差则非常小, 绝对误差 2 磅除以真实重量 2998 磅, 意味着相对误差仅仅是 $2/2998 \approx 0.00067$, 或者 0.067%。

绝对和相对误差

绝对误差描述了合计或测量的数值与真值之间的差距。

$$\text{绝对误差} = \text{声称值或测量值} - \text{真值} \quad (2-1)$$

相对误差将绝对误差与真值进行比较, 它常用百分比表示。

$$\text{相对误差} = \text{绝对误差} / \text{真值} \times 100\% \quad (2-2)$$

注意, 当声称值或测量值大于真值时绝对误差和相对误差是正的, 当声称值和测量值小于真值时绝对误差和相对误差是负的。

例 2-5 绝对误差和相对误差

计算下列例子中的绝对和相对误差。

(1) 你体重的真值是 100 磅, 用一个体重磅秤称的结果是 105 磅。

(2) 政府宣称一个项目花费 990 亿美元, 但是审计的结果是花费了 1000 亿美元。

解答

(1) 测量值是 105 磅, 真值是 100 磅

$$\begin{aligned} \text{绝对误差} &= \text{测量值} - \text{真值} \\ &= 105 \text{ 磅} - 100 \text{ 磅} \\ &= 5 \text{ 磅} \end{aligned}$$

$$\begin{aligned}\text{相对误差} &= \text{绝对误差} / \text{真值} \times 100\% \\ &= 5 \text{ 磅} / 100 \text{ 磅} \times 100\% \\ &= 5\%\end{aligned}$$

体重测量值比真值多 5 磅，或者说多了 5%。

(2) 政府声称项目花费数是 990 亿美元，真实花费数是 1 000 亿美元

$$\begin{aligned}\text{绝对误差} &= \text{声称值} - \text{真值} \\ &= 990 \text{ 亿} - 1000 \text{ 亿} \\ &= -10 \text{ 亿}\end{aligned}$$

$$\begin{aligned}\text{相对误差} &= \text{绝对误差} / \text{真值} \times 100\% \\ &= -10 \text{ 亿} / 1000 \text{ 亿} \times 100\% \\ &= -1\%\end{aligned}$$

政府声称花费比真实花费少 10 亿美元，或者说少了 1%。 ■

2.2.3 结果表述：准确性和精确度

一旦要报告一个测量结果，我们应该从潜在误差的角度评估它是否可靠。我们尤其应该考虑任何数值报告时的两个关键所在：它的准确性和精确度。在日常用语中这两个术语经常互相替代，但是它们却表示不同的科学概念。

测量的目标就是观测值尽可能与真值接近。**准确性 (accuracy)** 是指测量值与真值的接近程度，越准确的测量值越接近真值（接近通常以很小的相对误差来界定，而不是采用很小的绝对误差）。**精确度 (precision)** 是指测量值应用数字表达时的保留位数。例如，假定一次普查你所在城镇人口的总数是 72 453 人，但是人口数的真值却是 96 000 人，72 453 人的普查值精确度很高，但是却没能告诉我们真实的情况，它是非常不准确的，因为它比真实人口数 96 000 人少了将近 25%。注意准确性经常以相对误差来界定，而较少用绝对误差。例如，一家公司的某个项目销售额为 73 亿美元，真实的销售额是 73.2 亿美元，我们说这个数额是相对准确的，因为它的误差不到 1%，尽管这个 0.2 亿误差值代表的是 2 000 万美元。

相关知识

1999 年，美国国家航空和宇宙航行局失去了价值 16 亿美元的火星气候探测器，当时工程师向它发出了非常精确的英国单位（磅）计算机指令，而航天器接受这些指令

却是以国际单位（千克）来解读的。换句话说，这次失误的发生是由于非常精确的计算机指令实际上非常不准确造成的。美国国家航空和宇宙航行局接受了这次教训，此后另一个 2008 年登陆火星的计划（名叫凤凰），所发射的 4 个火星航天器都成功了。



我们通常以反映实际测量过程所报告的数字来界定精确度。如果我们说你的体重是 132 磅，意思是指你的体重测量精确到 1 磅，这样说来，更精确的测量是你的体重结果为 132.2 磅或 131.6 磅（都围绕 132 磅上下浮动）。比较起来，如果我们说你的体重为 132.0 磅，那就意味着测量精确到 0.1 磅，上面关于精确度的说明，意味着以实际上并没有应用的测量精度数字来进行报告就是虚假的。例如，你实际上测量体重最小的单位是 1 磅，如果你报告说你的体重为 132.0 磅就是虚假表述，因为它意味着你测量所应用的最小单位是 0.1 磅。

例 2-6 体重的准确性和精确度

假定你真实的体重为 102.4 磅，使用医生办公室的磅秤，它能精确到 1/4 磅，测量你的体重为 102 (1/4) 磅。还有一个健身房的磅秤，它能读出的最小刻度为 0.1 磅，它测量你的体重为 100.7 磅，哪一个体重量具更精确，哪一个更准确？

解答 健身房的磅秤更精确，因为它能测量的体重最小单位为 0.1 磅，而医生办公室的磅秤精确到 1/4 磅。然而，医生办公室的磅秤更准确，因为它测量的体重更接近真值。 ■

思考题

在例 2-6 中，我们需要掌握你体重的真值以便确定哪一个磅秤更准确。但是怎么样才知道你真实的体重？你确定你知道它吗？请解释。

案例研究

普查能否测量到真正的人口数

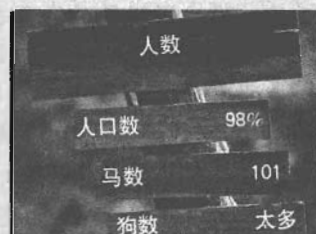
在完成了2000年的人口普查之后,美国人口普查局报告了人口总数为281 421 906人(截至2000年4月1日),这包含所有生活在美国的人口的一个完全数据。遗憾的是,这样精确的数据并不可能像它看起来意味的那样准确。

即使在理论上,要想得到一个生活在美国的完全人口数,唯一的办法是即时将每个人计数。否则,这个数字将会不存在,因为人口的变化是非常快的。平均算来,在美国每分钟有8个人出生和4个人死亡,同时每3分钟会有一个新移民进入美国。实际上,美国人口普查局搜集这些数据需要花费好几个月,此外,随机误差和系统误差也会使得几亿人口的人口普查不准确。

在现实中,美国人口普查局不可能知道特定一天完全的人口数,实际上不确定的人口至少会有数百万之多,基于这些事实,声称报告的是一个完全人口数是不诚实的。更诚实的报告应采用小一些的精确度,例如,表述时说人口数约为“2.8亿”。对美国人口普查局不太公平的是,他们的精确报告中解释了人口数的不确定性,但这些不确定性根本不被新闻机构所关注。

相关知识

一个数值中实际测量得出的数字叫做有效数字,一个数值中,除了用于准确定位小数点的0以外的所有数字都是有效数字,例如,0.001 234有4个有效数字,这些0是用于准确定位小数点的,同样的原因,数值1 234 000 000也有4个有效数字,而数值132.0有4个有效数字,在这里0是有效数字,因为这个0并不用于准确定位小数点,它的意思是指实际的测量值是0个1/10。



案例研究

不精确的预算导致万亿误差

早在2001年,政治家通报未来10年的美国联邦预算盈余为5.6万亿美元。到2007年,政府并没有获得万亿美元的盈余,而是多了3万亿美元的债。换句话说,到这个10年规划还有4年左右时间,这个计划已经产生的误差将近9万亿美元(从盈余的5.6万亿美元到负债3万亿美元)。

你可能听过或阅读过关于这个数万亿美元政府财富转换的新闻报道,不同政治派别的人们可能会对它的意义感觉相当不一样。但我们的目标在于,最引人注意的事情是关于原始计划的不恰当精确度,毕竟,5.6万亿美元的计划所引用的精确度是0.1万亿美元(这意味着1000亿美元),意味着真正的价值在5.5和5.7万亿美元之间,这个数字会让你惊奇地“哎呀”出来。

经济学家看待这个计划还是客观的,他们知道这里面有很多不确定性因素。不管怎么样,计划制定后发生了很多变化,比如“9·11”恐怖袭击、阿富汗和伊拉克战争、重要的减税,这些都使得计划与最开始时大相径庭,但这也让我们清醒地知道:下一次当你听到政治家在媒体中很精确地谈到将来的财政预算时,记住更可能的是,这些数字将证明它是多么不准确。

2.2.4 小结:误差控制

在本节中,主要涉及一些统计技术,但非常重要是理解测量和误差,让我们主要概括一下这两者之间的关系。

- 误差发生的方式很多,但是通常只分为

两种基本类型:随机误差和系统误差。

- 无论误差的来源是什么,描述误差大小的形式可以分为两种:一种是绝对误差,一种是相对误差。
- 当报告一个测量结果时,我们可能从准确性和精确度两方面对其进行评估。

2.2 练习题

统计素养和批判性思维

1. 误差类型 在一个实验中记录驾驶员的反应时，一个技术员不正确地记录了一次实际的时间。这是随机误差还是系统误差？说明理由。
2. 标准重量 精确表达 1 公斤的标准量被保存在联邦标准和技术研究院里。如果你将这 1 公斤的标准量放到一个天平上，天平的刻度显示是 1 002 克，这测量的是绝对误差还是相对误差？
3. 身高测量 一门统计课程中的学生测量了他们指导教师的身高，并且记录是 174.012 366 8 厘米。这名指导教师的真实身高接近于 176 厘米。描述所记录的身高的准确性和精确度，并做出解释。
4. 人口发展的里程碑 媒体报道说在 2006 年 10 月 17 号上午 7 点 46 分（美国东部时间），美国人口将达到 3 亿人。描述所报告时间的准确性和精确度。

判断正误：判断练习题 5~8 中的陈述是有意义的（表述正确）还是无意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. 鱼的种类 地球上共有 24 627 种鱼类。
6. 相对误差 一个微生物学家对细胞的测量的相对误差一定会小于一个天文学家对星系的测量，因为细胞要比星系小得多。
7. 扫描误差 詹金斯超市经理声称在购物环节所出现的扫描误差是随机的，大约一半的误差是对超市有利的。
8. 汽车的里程表 当购买一辆汽车时，汽车的里程表显示每加仑汽油能行驶 22.3655 英里^①。因为它是非常精确的所以我们可以信任它的结果。

概念与应用

9. 税收审计 税收审计主要为了寻找税收返还的几种问题，包括这两种：
 - (1) 在税收返还中录入或者计算等出现的错误；
 - (2) 纳税人不诚实地报告收入。
 讨论每种问题所包含的随机误差或者系统

误差。

10. 飞行安全 在起飞之前，飞行员按要求应该设置一下评估飞机高度的测高仪。从丹佛市（海拔 5 280 英尺）起飞的飞行员将他的测高仪调整至 2 500 英尺。说明测高仪的读数贯穿飞行始终会产生什么效应？这种误差是什么？
11. 技术规格 纽波特制造公司接到了一份订单，要求制造 10 000 个 25 毫米长的金属螺丝。在金属螺丝制造完成以后，一个质量控制分析员认真地作了测量，发现金属螺丝的平均长度是 25 毫米，但是大约一半金属螺丝要长于 25 毫米，另一半金属螺丝要短于 25 毫米。我们要控制哪种类型的测量误差？说明原因。
12. 自杀数据 依据搜集到的自杀数据，《纽约时报年鉴》宣称，“大多数专家认为自杀的统计要比报道中的严重得多，他们争论说大量的自杀被归因于事故或其他死亡原因。”如果在这种情况下，哪一种误差被带进了自杀数据中，它又是如何影响数值的呢？

误差的来源：在练习题 13~20 中，请确定如下表述中至少一个随机误差来源，至少一个系统误差来源。

13. 调查 电话调查 200 个人的年收入。
14. 税收返还 200 个人的年收入来自税收返还的部分。
15. 乘客体重 美国联邦航空局要求航空公司实际上测量一个乘客样本的体重，测量这些乘客体重时要求包括他们的衣服和托运行李。
16. M&M 公司 M&M 公司的筒装糖果是将一颗糖果包裹在一个纸杯中，只计算糖果的重量而不考虑纸杯的重量。
17. 雷达测速仪 一位警官使用雷达测速仪测量汽车的速度。
18. 假冒伪劣产品 纽约市的警官估计每年在这个城市出现的假冒伪劣产品的价值。
19. 香烟的销售 加利福尼亚州的健康专员基于香烟盒上的印花来计算香烟销售包数，进而估计在本州抽烟的数量。

① 1 英里 = 1.609 344 千米。

20. 长度测量 场地管理员使用1英尺长的尺子测量了足球场的长和宽。

绝对误差和相对误差：在练习题21~24中计算绝对误差和相对误差的值。

21. 信用卡账单 本书的作者收到了一张2 999美元的账单但是包括一个1 750美元的无效费用。
22. 牛排的重量 本书作者的一个学生对一块牛排进行称重,发现只有18盎司,但是菜单上说其重达20盎司。
23. 汽车的公里数 汽车上的计时器显示在高速公路上每加仑汽油可以行驶22英里,但实际的测量结果显示,在高速公路上的每加仑可以行驶24英里。
24. 12个面包 面包店的价格单上显示,一个包裹里含有12个甜甜圈,但实际上,面包师在每个包裹里通常会放13个甜甜圈。
25. 将误差最小化 25个人,包括你自己测量一个房子的长度精确到10毫米。假如每个人都使用同样的校正好的测量工具,比如卷尺。

- (1) 全部25个人的测量不可能完全一样,测量将会导致一些误差,这些误差是系统误差还是随机误差,并做出解释。
- (2) 如果在确定房子的长度时,想将随机误差的影响最小化,下面哪一种办法是最好的选择,报告你自己的测量结果,还是报告25个人测量结果的平均数,并做出解释。
- (3) 描述在房屋测量中所有可能出现的系统误差。
- (4) 将25个人的测量结果平均的过程能否降低系统误差?为什么?

26. 将误差最小化 用来测量22F型汽车蓄电池重量的测量工具是非常精确的,通过连续10次测量获得蓄电池的重量的数据。

- (1) 全部的10次测量一般不会是完全一样的,这样测量一定会导致一些误差,这些误差是系统误差还是随机误差,并做出解释。
- (2) 如果你想确定蓄电池的真实重量并将随机误差最小化,下面方法哪一种是

最好的选择。随机选择10次测量中的一次进行报告,还是报告10次测量的平均数,并做出解释。

- (3) 在10次测量中说明所有可能的系统误差来源。
 - (4) 将10次测量的结果取平均能否降低系统误差?为什么?
27. 测量一艘巡洋舰的重量的准确性和精确度 一艘新的巡洋舰重达3 273磅,制造商的测量结果精确到10磅测出的结果是3 250磅,美国交通部的测量结果精确到0.1磅,测量结果是3 298.2磅,哪一个结果更精确,哪一个结果更准确,并做出解释。
28. 身高测量的准确性和精确度 假定美国总统乔治·布什实际的身高是71英寸[⊖],假定他的身高是用一个读数精确到1/8英寸的卷尺来测量的,结果发现是71 (1/8)英寸,使用医生办公室的一个读数精确到0.05英寸的新的激光测高仪,测出他的身高是71.15英寸,哪一个结果更精确,哪一个结果更准确,并做出解释。
29. 重量测量的准确性和精确度 假定你的体重是52.55千克,一台健康称读数精确到0.5千克,测得你的体重是53千克,健身房的1台磅秤读数精确到0.01千克,测得你的体重是52.88千克,哪一个结果更精确,哪一个结果更准确,并做出解释。
30. 重量测量的准确性和精确度 假定你的体重是52.55千克,一台健康称读数精确到0.5千克,测得你的体重是52 (1/2)千克,健身房的一台磅秤读数精确到0.01千克,测得你的体重是51.48千克,哪一个结果更精确,哪一个结果更准确,并做出解释。

事实可信任吗?练习题31~38给出了来自统计测量的事实描述,对于每一段描述简要讨论测量的误差来源,然后说明所给出测量的精确度并讨论你认为这个事实是否可信?

31. 人口 1860年的美国人口是31 443 321人。
32. 机动车的交通事故死亡率 去年一年中,美国有44 758人死于机动车交通事故。
33. 中国人口 去年中国的人口1 372 557 236人。

⊖ 1英寸=0.025 4米。

34. 最高的建筑物 台北 101 大厦高达 1 671 英尺，是世界上最高的建筑物。
35. 圣路易拱桥 圣路易拱桥高达 630. 237 759 969 4 英尺。
36. 手机 据《信息世界》报道，全世界的手机用户已经达到 2.1 亿。
37. 大学生 美国当前大学生的人数是 18 000 000。
38. 将要灭绝的生物种类 按照美国政府的统计，现在有 1 879 种动植物将要灭绝或者生存受到威胁。

网络和远程项目

请在网址 www.aw.com/bbt 上打开第 2 章的有用链接“Links for internet projects”。

39. 2012 年人口普查 浏览美国人口普查局的网站了解 2012 年人口普查的计划，将在何时以及如何搜集数据？与 2011 年的人口普查相比，数据采集过程有什么显著的改变，看起来 2012 年的人口普查是否比 2011 年更准确，发表你的观点。
40. 人口普查的争论 浏览美国国会图书馆关于“Thomas”网页寻找有关人口普查数据收集的未决议案，如果你寻找到了很多未决议案选择一个进行深入研究，总结议案的目标并简要讨论支持和反对争论的焦点。
41. 手表误差 寻找一个能准确预报当地时间的网页，并调整你的手表精确度至秒，然

后比较你和你朋友的手表时间差别，记录这个误差，如果手表走得快将误差记做正值，如果手表走得慢将误差记做负值，使用这章的概念描述你所采集的样本中手表的准确性。

新闻中的统计学

42. 随机误差和系统误差 寻找一个最近的新闻报道，上面有统计测量的数量（例如有关人口、平均收入或者无家可归的人数的报道）。写下一个简短的这个数量是如何测量而来的描述，并简要说明任何可能的随机误差和系统误差来源，最后你认为报道的数量是准确的吗？为什么？
43. 绝对误差和相对误差 寻找一则近期的新闻报道，内容是描述一些测量误差的或者是有关估计以及项目数字是错误的（例如一个项目预算被证明是不正确的）使用统计学的语言描述绝对误差和相对误差的大小。
44. 准确性和精确度 寻找一则近期的新闻文章，在这篇文章中你质疑它的准确性和精确度。例如，文章报道了一个比你所认为的更精确的数字，或者是陈述了一个你知道是不正确的数字。写一页有关这个报道的总结，并说明你为什么质疑它的准确性和精确度。



2.3 百分数在统计学中的应用

统计结果经常以百分数表示，百分数是分数的一种简单表达方式。术语“百分数”意味着“除以 100”。然而，百分数经常以这样的方式应用，思考《纽约时报》中一篇文章开头部分所陈述的：

“十年级的学生吸烟率达到了 45%，上升了 18.3%；八年级的学生吸烟率达到了 44%，上升了 10.4%。”

在这个陈述中，所有的百分数的应用都是正确的，但是它们容易让人迷惑。例如，“达到了 45%，上升了 18.3%”意思是指什么？在这一节，我们将讨论应用百分数的微妙之处以及

百分数的滥用。在开始之前，你应该复习一下以下的分数和百分数互相转换的基本规则。

分数和百分数之间的相互转换

1. 百分数转换成一般的分数：

将百分号去掉后除以 100，然后作必要的简化。

例子： $25\% = 25/100 = 1/4$

2. 百分数转换成小数：

去掉百分号后将小数点后移两位（也就是除以 100）。

例子： $25\% = 0.25$

3. 小数转换成百分数：

向前移动两位小数点（也就是乘以 100）然后加上百分号。

例子： $0.43 = 43\%$

4. 分数转换成百分数:

首先将分数转换成小数, 然后再小数转换成百分数。

例子: $1/5 = 0.2 = 20\%$

思考题

比较绝对变化、相对变化和绝对误差、相对误差的公式。依据 2.2 节的内容描述你观察到的它们的相同点。

例 2-7 报纸调查

一家报纸报道说调查了 1 069 人, 其中 44% 说总统工作很称职, 问有多少人总统很称职?

解答 44% 代表了说总统很称职这样反应的一个分数, 因为“其中的”通常指示的是乘法, 我们将两者相乘得

$$44\% \times 1\,069 = 0.44 \times 1\,069 = 470.36 \approx 470$$

1 069 人中的大约 470 人说总统很称职, 注意我们说约 470 人是为了得到整数。(符号 $\approx$ 意思是约等于。)

例 2-8 世界人口增长

1950 年世界人口总数是 26 亿, 到 2000 年开始时, 已经超过 60 亿了, 描述世界人口从 1950 ~ 2000 年的绝对变化和相对变化。

解答 参考值是 1950 年 26 亿的人口总数, 新值是 2000 年 60 亿的人口总数。

$$\begin{aligned}\text{绝对变化} &= \text{新值} - \text{参考值} \\ &= 60 \text{ 亿} - 26 \text{ 亿} \\ &= 34 \text{ 亿}\end{aligned}$$

$$\begin{aligned}\text{相对变化} &= (\text{新值} - \text{参考值}) / \text{参考值} \times 100\% \\ &= (60 \text{ 亿} - 26 \text{ 亿}) / 26 \text{ 亿} \times 100\% \\ &= 130.7\%\end{aligned}$$

世界人口总数从 1950 ~ 2000 年增长了 34 亿人, 或者说增长了 130.7%。

2.3.1 用百分数描述变化

在统计学中, 百分数经常用来描述随着时间而改变的数据。在一个例子中, 假定某一个城镇的人口数 1970 年时为 10 000 人, 而到了 2000 年则为 15 000 人, 我们可以以下面两种基本方式表示这种人口的变化。

- 因为人口数增加了 5 000 人 (从 10 000 到 15 000), 我们说人口的绝对变化 (absolute change) 是 5 000 人。
- 因为增长的 5 000 人是开始人口 10 000 的 50%, 我们说人口的相对变化 (relative change) 是 50%。

通常, 计算一个绝对或相对变化总是包括两个数字: 一个起始数, 或者叫参考值 (reference value), 还有一个新值 (new value)。当我们确定了这两个数值, 我们就可以根据下述公式计算绝对变化和相对变化。注意如果新值大于参考值, 绝对变化和相对变化就是正数, 如果新值小于参考值, 绝对变化和相对变化就是负数。

绝对变化和相对变化

绝对变化描述了从参考值到新值间实际的增长或减少:

$$\text{绝对变化} = \text{新值} - \text{参考值} \quad (2-3)$$

相对变化描述了绝对变化的量与参考值的比, 可以表示成一个百分数:

$$\text{相对变化} = (\text{新值} - \text{参考值}) / \text{参考值} \times 100\% \quad (2-4)$$

2.3.2 用百分数作比较

百分数也经常用来比较两个数字, 这时, 两个数字是参考值和比较值。

- 参考值 (reference value) 是我们用来比较的对象的那个数字。
- 比较值 (compared value) 是另一个数字, 我们用它和参考值做比较。

我们可以用公式表达两个数值间绝对的和相对的差别, 与绝对变化和相对变化非常相似。注意如果比较值大于参考值, 这种绝对差别和相对差别就是正数, 如果比较值小于参考值, 这种绝对差别和相对差别就是负数。

绝对差别和相对差别

绝对差别描述了比较值和参考值间的差别:

$$\text{绝对差别} = \text{比较值} - \text{参考值} \quad (2-5)$$

相对差别描述了绝对差别的量与参考值的比, 可以表示成一个百分数:

$$\text{相对差别} = (\text{比较值} - \text{参考值}) / \text{参考值} \times 100\% \quad (2-6)$$

相关知识

没有一个人知道俄罗斯人预期寿命较低的原因，但有一个确定发生影响的因素就是酗酒，俄罗斯人比美国人的酗酒更常见。



例 2-9 俄罗斯人和美国人的预期寿命

美国男人的预期寿命约为 75 岁，俄罗斯男人的预期寿命约为 59 岁。以绝对差别和相对差别的方式，比较俄罗斯人和美国人的预期寿命（见 6.4 节关于平均预期寿命的讨论）

$$\begin{aligned}\text{绝对差别} &= \text{比较值} - \text{参考值} \\ &= 75 \text{ 岁} - 59 \text{ 岁} \\ &= 16 \text{ 岁}\end{aligned}$$

$$\begin{aligned}\text{相对差别} &= (\text{比较值} - \text{参考值}) / \text{参考值} \times 100\% \\ &= (75 \text{ 岁} - 59 \text{ 岁}) / 59 \text{ 岁} \times 100\% \\ &= 27\%\end{aligned}$$

美国男人的预期寿命比俄罗斯男人多 16 岁，或者说美国男人的预期寿命比俄罗斯男人多 27%。

2.3.3 “是多少”与“多多少”

一个有关百分数的微妙表述来自用词。人数从 200 变到 600 是 3 倍的变化。用百分数表述有以下两种等价的方式。

- 使用“多多少”：新的人数比原始人数增长 200%。在这里，我们看到的是人数的相对变化。

$$\begin{aligned}\text{相对变化} &= (\text{新值} - \text{参考值}) / \text{参考值} \times 100\% \\ &= (600 - 200) / 200 \times 100\% \\ &= 200\%\end{aligned}$$

- 使用“是多少”：新的人数是原始人数的 300%，也就是说原来是原来人数的 3 倍，在这里，我们看到的是新人数与原始人数的比值。

$$\text{新人数} / \text{原始人数} = 600 / 200 = 3.00 = 300\%$$

值得注意的是：百分数表示中的“是多少”和“多多少”的相互关系是 $300\% = 100\% + 200\%$ ，通常会有下面的关系。

“是多少”与“多多少”（或者“少多少”）

如果新值或比较值比参考值多 $P\%$ ，那么，它们是参考值的 $(100 + P)\%$ 。

如果新值或比较值比参考值少 $P\%$ ，那么，它们是参考值的 $(100 - P)\%$ 。

例如，比参考值多 40% 就是参考值的 140%，如果比参考值少 40% 就是参考值的 60%。当你听到统计学上引用百分数时，仔细分辨清楚是“多多少”（或者少多少）还是“是多少”是非常重要的，当然也希望百分数的应用者知道这个差别。

例 2-10 世界人口

在例 2-8 中，我们发现 2000 年的人口数比 1950 年增长了 131%，用“是什么”的表达方式说明这个变化。

解答 2000 年的世界人口比 1950 年的增长了 131%，由于 $(100 + 131)\% = 231\%$ ，2000 年的人口是 1950 年的 231%，这意味着 2000 年的人口是 1950 年的 2.31 倍。

例 2-11 促销

一家商店促销时“降价 25%”，与原价相比现在以什么样价格出售？

解答 “降价 25%”意思是指促销价比原价降低了 25%，也就是说，促销价是原价的 $(100 - 25)\% = 75\%$ 。举例来说，如果原价是 100 美元，现在价格是 75 美元。

思考题

一家商店的广告说“每件商品均降价 $1/3$ ”。另一家商店的广告说“所有产品促销价均是原价的 $1/3$ ”，哪一家商店的促销力度更大？请解释。

相关知识

近年来，世界人口每年增长大约为 7500 万人，这意味着大约 4 年时间增加人口就相当于全部美国人的数量。

2.3.4 百分数的百分数

当值本身就是百分数时，以百分数表示变化和差别就使人特别困惑了。假定银行将存款利率从3%提高到4%。这可能诱使你表述说，利率增长了1%，但是这个表述很容易引起歧义。利率增长了一个百分点，相对变化是33%：

$$(4\% - 3\%) / 3\% = 0.33 \times 100\% = 33\%$$

这时你可以说银行提高利率为33%，也可以说实际利率增长只有一个百分点（从3%到4%）。

百分点与百分数

当我们看到一种变化或差异以百分点来表示时，你可以推定这是一种绝对变化或绝对差别。如果它以百分数（%）来表达，它很可能是相对变化或相对差别。

例2-12 误差幅度

依据你所在学校一个学生样本的调查，你得出结论说全部学生中素食者可能在20%~30%。你可以报告你的结果“25%误差幅度5%”还是“25%，5个百分点的误差幅度”？说明理由。（见1.1节，参考误差幅度的含义。）

解答 20%~30%的范围来自25%减去或加上一个5个百分点的绝对差异，也就是：

$$20\% = (25 - 5)\% \text{ 到 } 30\% = (25 + 5)\%$$

这样说来，正确的表述是“25%和5个百分点的误差幅度”。如果你说“25%，误差幅度5%”，那意味着误差将是25%的5%，只有1.25%。

例2-13 仔细措词

假定40%的卡尔森市注册选民是共和党人，仔细阅读下面的问题给出最恰当的回答。

(1) 弗里敦的注册选民中共和党人比卡尔森市多出25%，弗里敦的注册选民是共和党人的百分比是多少？

(2) 弗里敦的注册选民中共和党人比卡尔森市多出25个百分点，弗里敦的注册选民是共和党人的百分比是多少？

解答

(1) 我们说明“25%”是相对差别，40%的25%是10%，（因为 $0.4 \times 0.25 = 0.1$ ）。弗里敦的注册选民中共和党人的百分数就是 $40\% + 10\% = 50\%$ 。

(2) 在这种情况下，我们说明“25个百分点”是绝对差别，因此我们简单地将卡尔森市注册选民中共和党人的百分数加上25%这个百分数，这样弗里敦的注册选民中共和党人的百分数是 $40\% + 25\% = 65\%$ 。

2.3 练习题

统计素养和批判性思维

1. 百分比的解释 思考这一节开始所引用的内容“在十年级学生中，抽烟人数的比率已经上升到45%，上升了18.3个百分点，八年级的学生中抽烟人数的比例已经上升到44%，上升了10.4个百分点”简要解释在这段陈述中每一个百分数的含义。
2. 百分数 《纽约时报》的一篇编辑社论列举出1张描述洗牙的图表，介绍到“洗牙降低了牙齿中斑块的300%”，如果说洗牙去除了所有的斑块，那么百分之多少的斑块比例已被消除了。降低斑块的比例达到300%可能吗？
3. 百分点 一个盖洛普调查访问了1236名成人，其中的5%的人相信祸不单行。这个百分数的误差幅度是1.2个百分点。为什么“5%的百分数具有1.2%的误差幅度”这样的陈述容易导致误解？
4. “是多少”和“多多少” 一个民意测验调查了1400个回答者，1400人的5%是多少？

比1400人多5%又是多少？

判断正误：判断练习题5~8中的陈述是有意义的（表述正确）还是无意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. 移动电话 拥有手机人数的百分比增长了1200万人。
6. 百分数 公司的CEO比CFO的年收入多出50%，所以CFO的年收入比CEO少50%。
7. 利率 杰斐逊谷银行的新车贷款利率提高了100%。
8. 利率 杰斐逊谷银行的新车贷款利率提高了100个百分点。

概念与应用

9. 比例、小数和百分数 将下列每一个数字分别用3种不同的形式表述：比例、小数和百分数。
(1) $2/5$ (2) 1.50
(3) 0.25 (4) 30%
10. 比例、小数和百分数 将下列每一个数字分

别用3种不同的形式表述：比例、小数和百分数。

(1) 225% (2) 0.375

(3) $1/20$ (4) -0.12

11. 实践中的百分数 一项关于1028名犯罪嫌疑人的辩护词的研究。在这些犯罪嫌疑人中,956名承认有罪,其中329名被判决入狱服刑。另外72名犯罪嫌疑人,他们辩护自己无罪,其中58名被判决入狱服刑(based on data from "Does It Pay to Plead Guilty?" by Brereton and Casper, *Law and Society Review*, Vol. 16, No. 1)

(1) 犯罪嫌疑人承认有罪的人数所占百分比是多少?

(2) 犯罪嫌疑人被判入狱服刑的人数的百分比是多少?

(3) 在承认有罪的犯罪嫌疑人中,被判入狱服刑的人数百分比是多少?

(4) 在辩护自己无罪的犯罪嫌疑人中,被判入狱服刑的人数百分比是多少?

12. 实验的百分数 一项研究用以确定抛掷或者旋转一枚硬币对于正面向上的百分比的影响。在49434次试验中,包括29015次抛掷硬币,其中14709次正面向上。另外20422次试验是旋转硬币,结果是9197次正面向上的(based on data from Robin Lock as given in *Chance News*)。

(1) 实验中抛掷硬币的次数所占的百分比是多少?

(2) 实验中旋转硬币的次数所占的百分比是多少?

(3) 在抛掷硬币的次数中,正面向上的次数所占的百分比是多少?

(4) 在旋转硬币的次数中,正面向上的次数所占的百分比是多少?

相对变化:在练习题13~20中,每一道题提供了两个取值,对于每一组数值,使用百分数表达相对变化或者差异。以第二个给出的数值作为参考值,用最接近的百分点表达结果。同时,用一句话来描述结果。

13. 报纸 美国现在的日报数量是1456家,1900年这个数量是2226家。

14. 家用汽车 现在美国注册的家用汽车是136651000辆,而1980年,注册的家用汽

车是121601000辆。

15. 航空器 美国现在有8206架民用航空器,而在1980年这个数量是3808架。

16. 破产 去年一共有1725300件破产案例,而在2000年破产案例的数量是1276900件。

17. 报纸 《华尔街日报》的发行量大约是175万份(美国发行量最大的日报),《纽约时报》每日的发行量大约是109万份(美国发行量第3的日报)。

18. 汽车销售 在最近一年,一共销售了13525辆克莱斯勒300C汽车,而本田混合动力汽车的销售量是45782辆。

19. 飞机场 俄亥俄州芝加哥飞机场去年接待了3400万乘客。世界上最繁忙的机场亚特兰大哈兹菲尔德飞机场去年接待了4200万乘客。

20. 旅游者 在最近一年,法国在旅游目的地中排行第1,接待了7510万国际旅游者。美国排名第3,接待了4610万国际旅游者。

调查:一些重要调查结果的分析,需要你知道回答落入某个分类的实际被试人数。在练习题21~24中,根据给出的百分比计算其实际人数。

21. 私人电话 在AT-A-Glance调查中,访问了1385名政府工作人员,其中4.8%的人在在工作时间不接打私人电话。

22. 调查错误 在一项对150名经理人的调查中,其中47%说最常见的面谈错误是对公司很少了解或者几乎不了解。

23. 电视 在对1005名成人的调查中,其中83%报告说家庭中有超过一台以上的电视。

24. 手机 在对1109名超过11岁的消费者的调查中,81%报告说在他们的家庭中至少有一部手机。

“是多少”与“多多少”:将练习题25~28中的括号部分用百分数填充,并简要解释你在每一个例子中的理由。

25. 重量 如果一部卡车的重量超过一部汽车的重量40%,那么卡车的重量是汽车重量的____%。

26. 面积 如果挪威的面积比科罗拉多州的面积多24%,那么挪威的面积是科罗拉多面积的____%。

27. 人口 如果蒙大拿州的人口要比新罕布什尔的人口少20%,那么蒙大拿州的人口是新罕布什尔人口的____%。

28. 薪水 如果杰里诺的工资大约要比戴维·莱特曼的工资少18%，那么杰里诺的薪水是戴维·莱特曼薪水的____%。
29. 误差幅度 盖洛普一项对1012名成人的调查显示89%的美国人认为克隆人类不应该被允许，误差幅度是3个百分点。如果说一家报纸所报道的内容误差幅度是3%你会介意吗？请加以解释。
30. 误差幅度 皮尤研究中心对3002名成人的调查显示，收听国家公共广播的人数百分比可能在14%~18%之间，报纸应该如何报道其误差幅度？请加以解释。

百分数的百分比：练习题31~34中描述了百分数测量的自身变化，通过两种方式表达这种变化：①用百分点来表达绝对差异；②用百分比来表达相对差异。

31. 高中高年级学生的饮酒百分比从1975年的68.2%下降到今天的52.7%。
32. 世界上生活在发达国家的人口的百分数从1970年的27.1%降低到今天的19.5%。
33. 因为各种原因所导致的癌症的5年生存率从20世纪60年代的39%上升到今天的61%。
34. 非洲裔美国人所患各种各样的癌症的5年生

存率从20世纪60年代的27%上升到今天的48%。

网络和远程项目

请在网址 www.aw.com/bbt 上打开第2章的有用链接“Links for internet projects”。

35. 世界人口 寻找美国普查局关于世界人口的估计，描述从1999年世界人口达到60亿到现在的百分比变化，也寻找一下人口中的估计是怎么完成的，并且讨论一下对世界人口估计的不确定性。
36. 药物使用的统计 浏览联邦药物成瘾和药物滥用研究中心的网页，寻找近期关于药物滥用的报告，写一页这个新研究的总结，在结论中至少要用到一些百分数。

新闻中的统计学

37. 寻找3篇近期的新闻报道，其中描述统计结果的时候用到了百分数，描述每一篇报道中百分数的意义。
38. 百分数变化 寻找近期的一篇新闻报道，在表达从一个时期到另外一个时期的统计结果的变化时用到了百分数（比如人口的增长或者儿童抽烟的人数）。描述这种变化的意义，请注意寻找关键词“是多少”或“多多少”。

到1965年的价格是1975年价格的55.0%：

$$\begin{aligned} 1965 \text{ 年价格} / 1975 \text{ 年价格} &= 31.2 \text{ ¢} / 56.7 \text{ ¢} \\ &= 0.550 = 55\% \end{aligned}$$

每一年的价格都重复这个过程，我们可以计算出每年价格是1975年价格的百分比。表2-1的第3列显示了这个结果，注意1975年的百分比是100%，因为我们将1975年的价格作为参考值。

表2-1 汽油平均价格（每加仑）

年份 (年)	价格 (美分)	是1975年价格 的百分比(%)	价格指数 (1975=100)
1955	29.1	51.3	51.3
1965	31.2	55.0	55.0
1975	56.7	100	100
1985	120.5	210.9	210.9
1995	155.0	272.5	272.5
2000	230.1	405.8	405.8
2005	230.1	405.8	405.8



2.4 指数

如果你听过晚间商业报道，你可能听过指数。比如消费者价格指数、产品价格指数或者消费者信心指数。指数在统计学中是非常常见的，因为它们提供了不同时间和不同地方测量结果之间的一个比较，在这一节，我们将探讨指数的含义及如何运用指数。关注的焦点是消费者价格指数(CPI)，让我们以汽油价格的例子开始。

表2-1显示了1955~2005年每年美国汽油的平均价格（这是那些年实际的价格，只是没有做通货膨胀后的价格调整）。假定，代替价格自身，我们想知道不同年份与1975年的价格相比较的价格是怎么样的？一种方法就是将每一年的价格表示为1975年价格的百分比。例如，将1965年的价格除以1975年的价格，我们得

⊖ ¢ 是货币单位美分的符号，1 ¢ = 0.01 \$。

现在我们看表 2-1 的最后 1 列，它与第 3 列是相同的，只是去除了百分号，这个数字的简单变化是将百分数转变成一个价格指数，这是一种指数类型。分类栏中的说明“1975 年 = 100”显示 1975 年的价格是参考值。在这种情况下，百分数和指数之间相比较，实际上并没有什么差别，只不过指数是一种选择，使用起来更方便些。然而，就像我们下面看到的，在同时考虑许多因素的情况下，使用指数而不是百分数是一种惯例。

相关知识

表 2-1 中的价格是一年内的平均数，并不能反映本年内价格的波动，例如，2005 年，在卡特里娜飓风之后汽油价格达到一个短暂的高峰为每加仑^①3.13 美元。

指数

指数 (index number) 提供的是不同时间和不同地点测量结果之间简单的比较。必须选择某一时间 (或某一地方) 的数值作为参考值。另一个时间 (或地方) 的指数就是：

$$\text{指数} = \text{数值} / \text{参考值} \times 100 \quad (2-7)$$

例 2-14 计算指数

假定今天的汽油价格为每加仑 3.2 美元，使用 1975 年的价格作为参考值，计算今天的汽油价格指数是多少？

解答 表 2-1 显示 1975 年的汽油价格为每加仑 0.567 美元，如果将 1975 年的价格作为参考值，今天的价格是每加仑 3.2 美元，今天的汽油价格指数就是：

$$\begin{aligned} \text{指数} &= \text{当前价格} / 1975 \text{ 年价格} \times 100 \\ &= 3.2 \text{ 美元} / 0.567 \text{ 美元} \times 100 = 564.4 \end{aligned}$$

当前的价格指数是 564.4，这意味着现在的汽油价格是 1975 年的 564.4%。

思考题

在附近的一家加油站查清今天汽油的实际价格，以 1975 年的价格为参考值，今天的汽油价格指数是多少？

相关知识

术语“指数”用于各种进行比较的数字中，即使这些数字不是标准指数。例如，体型指数 (BMI) 提供了一个对人们的身高和体重进行比较的方法，但是它的定义中没有任何参考值。体型指数被定义为体重 (以公斤为单位) 除以身高 (以米为单位) 的平方。

2.4.1 用指数作比较

指数的目标就是让对比变得非常容易进行。例如，假定你想知道 2000 年的汽油比 1975 年贵了多少。我们很容易从表 2-1 中找出答案，我们以 1975 年价格为参考值，表中显示 2000 年的价格指数是 273.4，这意味着 2000 年的汽油价格是 1975 年的 273.4%。同样，我们也可以说 2000 年的汽油价格是 1975 年的 2.734 倍。

我们也可以用另一个值为参考值进行比较，例如，我们想知道 1995 年的汽油比起 1965 年价格贵了多少，通过将这两年的指数进行除法运算就可以回答这个问题：

$$\begin{aligned} &1995 \text{ 年指数} / 1965 \text{ 年指数} \\ &= 212.5 / 55.0 = 3.86 \end{aligned}$$

1995 年价格是 1965 年的 3.86 倍，或者是 386%。换句话说，在 1965 年价格为 1.00 美元的汽油在 1995 年价格是 3.86 美元。

例 2-15 应用汽油价格指数

参照表 2-1 回答以下问题。

- (1) 假定在 1975 年加满你的油箱需要 \$7.00，那么在 2005 年买同样数量的汽油需要花费多少钱？
- (2) 假定在 1995 年加满你的油箱需要 \$20.00，那么在 1955 年买同样数量的汽油需要花费多少钱？

解答

- (1) 表 2-1 显示 2005 年的价格指数 (1975 = 100) 是 407.4，这意味着 2005 年的汽油价格是 1975 年价格的 407.4%，所以在 1975 年价值 7.00 美元的汽油在 1995 年的价格应该是：

$$407.4\% \times 7.00 \text{ 美元} = 4.074 \times 7.00 \text{ 美元} = 28.52 \text{ 美元}$$

- (2) 表 2-1 显示 1995 年的价格指数 (1975 = 100) 是 212.5，1955 年的价格指数是 51.3，1955 年和 1995 年的汽油价格相比较是：

$$1955 \text{ 年指数} / 1995 \text{ 年指数} = 55.3 / 212.5 = 0.2614$$

① 1 加仑 = 4.546 09 升

1995 年价格为 20.00 美元的汽油 1955 年要花费
 $0.2414 \times 20.00 = 4.83$ 美元。 ■

2.4.2 消费者价格指数

我们已经了解汽油价格随着时间持续上涨,大多数物价和薪资也在上涨,这个现象叫做通货膨胀(inflation),物价和工资偶尔也会下调,这是紧缩。如果不与通货膨胀率进行比较的话,汽油价格的变化是没什么意义的,通货膨胀率的指标一般是消费价格指数(Consumer Price Index, CPI)。

消费价格指数是由美国劳动统计局每月进行计算和报告的,它代表了一个包括食品、服务和住房等的样本平均价格。每月样本包括超过 60 000 个项目,数据搜集和指数计算的细节是相当复杂的,但 CPI 本身却是一个简单的指数。表 2-2 呈现了 30 多年间每年的平均 CPI。当前, CPI 的参考值是 1982~1984 年之间的平均价格,这也是表中“1982~1984=100”的含义。

方法注解

美国政府采用两个消费价格指数: CPI-U 是基于那些反映城市消费者购买习惯的产品, CPI-W 则基于只涵盖工薪阶层购买习惯的产品(表 2-2 呈现的是 CPI-U)。

消费者价格指数

消费者价格指数是每月进行计算和报告的,基于一个超过 60 000 个项目包括食品、服务和住房等的样本而来。

表 2-2 每年的消费者价格指数
(1982~1984=100)

年份	CPI	年份	CPI	年份	CPI
1977	60.6	1987	113.6	1997	160.5
1978	65.2	1988	118.3	1998	163.0
1979	72.6	1989	124.0	1999	166.6
1980	82.4	1990	130.7	2000	172.2
1981	90.9	1991	136.2	2001	177.1
1982	96.5	1992	140.3	2002	179.9
1983	99.6	1993	144.5	2003	184.0
1984	103.9	1994	148.2	2004	188.9
1985	107.6	1995	152.4	2005	195.3
1986	109.6	1996	156.9	2006	201.6

CPI 允许我们比较不同时期的总体价格。

例如,为了知道 2005 年比 1995 年的价格高多少。我们将这两年的 CPI 相除:

$$\begin{aligned} & 2005 \text{ 年的 CPI} / 1995 \text{ 年的 CPI} \\ &= 195.3 / 152.4 = 1.28 \end{aligned}$$

按照 CPI 来说,2005 年的价格是 1995 年的 1.28 倍。例如,一个在 1995 年价值 1 000 美元的代表项目,在 2005 年价格为 1 280 美元。当然,个体项目的价格变化不同于平均水平,例如,同样性能的计算机从 1995~2005 年价格明显下降,这意味着在 2005 年你花同样的钱或者更少一些就可以购买一台性能更好些的电脑。相反,在同一段时间内,健康保险的平均价格涨到了两倍以上。我们说健康保险的价格比总体上的通货膨胀率要上涨快得多。

例 2-16 CPI 变化

假如你在 2000 年维持某一生活水平需要 30 000 美元,那么你在 2006 年维持同样的生活水平需要多少钱?假定你购买的商品价格上涨的水平与消费者价格指数(CPI)是一样的。

解答 我们比较 2006 年与 2000 年的消费价格指数
 2006 年的 CPI/2000 年的 CPI
 $= 201.6 / 172.2 = 1.17$

这样,2006 年的代表性价格大约是 2000 年的 1.17 倍,如果你在 2000 年需要 30 000 美元,那么你在 2006 年需要 $1.17 \times 30\,000$ 美元 = 35 100 美元维持同样的生活水平。 ■

2.4.3 通货膨胀后的价格调整

2000 年中期,汽油价格突然上涨,到达一个高峰值每加仑 1.87 美元,虽然这个价格在今天看起来还是很便宜的,在当时,它可使消费者异常愤慨,新闻报道中则大肆惊呼这是“历史上最高的汽油价格”。

相关知识

职业运动员的薪资过去一直保持在较低的水平上,因为运动员不允许在市场上自由流动(“自由球员”),这一状况在棒球明星科特·弗勒德 1970 年向棒球大联盟的起诉之后产生了变化。在最高法院裁定偏向棒球大联盟后,弗勒德最终败诉了,但是这个过程(趋向“自由球员”)并没有停止。

由于实际上的价格(张贴在加油泵上),每加仑 1.87 美元深深地打破了 1981 年以来每加仑 1.41 美元的最高纪录,但这个价格真的

很高吗？如果我们客观地进行比较，就必须考虑通货膨胀的效应。这样的话，我们必须要看一下1981年和2000年的代表价格的变化，依据表2-2中的数据，我们将这两年的CPI相除：

$$\begin{aligned} & \text{2000年的CPI/1981年的CPI} \\ & = 172.2/90.9 = 1.89 \end{aligned}$$

由于CPI上涨1.89的因素，1981年汽油价格每加仑1.41美元换成是2000年的价格应该为 $1.89 \times 1.41 \text{ 美元} = 2.78 \text{ 美元}$ 。用经济学的语言来说，我们说“1981年美元”的1.41美元相当于“2000年美元”的2.78美元。因为2000年的实际价格比2.78美元低，经过通货膨胀率调整后的2000年的汽油价格并没有1981年高。图2-3呈现了50多年来汽油每年的价格数据，包括实际价格和以2006年美元为基准进行通货膨胀率调整后的价格，注意，扣除物价因素，直到2006年1981年的平均汽油价格还保持着最高的纪录。

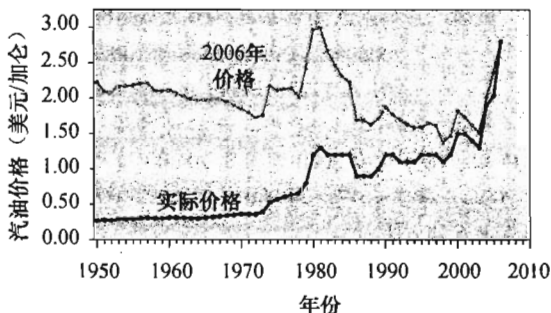


图2-3 每年汽油平均价格（1950~2006年）

注：我们使用2006年价格作为通货膨胀调整后的价格。

资料来源：American Petroleum Institute.

例2-17 棒球运动员的薪资

1987年，美国棒球联盟运动员的平均薪资是412 000美元，2006年涨到了2 867 000美元。将棒球联盟运动员平均薪资的增长与以CPI为指标的通货膨胀率进行比较。

解答 首先，我们比较2006年和1987年的消费价格指数：

$$\begin{aligned} & \text{2006年的CPI/1987年的CPI} \\ & = 201.6/113.6 = 1.77 \end{aligned}$$

然后，我们比较这两年的棒球运动员的薪水：

$$\begin{aligned} & \text{2006年棒球运动员的平均薪资/1987年棒球运动员的} \\ & \text{平均薪资} = 2\,867\,000 \text{ 美元} / 412\,000 \text{ 美元} = 6.96 \end{aligned}$$

在这两个时间段内，平均价格（以CPI为指标）上升了大约77%，而棒球运动员的平均工资则上升了几乎600%。换句话说，美国棒球联盟运动员的平均薪资相比较通货膨胀率，上升了大约 $600/77 \approx 8$ 倍。 ■

2.4.4 其他指数

消费价格指数仅仅是你在新闻报道中见到的众多指数之一。另外有一些也是价格指数，比如生产价格指数（PPI），意思是指生产者（制造商）支付他们所购买原料的价格的指数（不是消费者支付的价格）。另外一些指数试图去测量更多定性变量。例如，消费者信心指数基于对消费者态度的调查，旨在测量人们更愿意消费还是储蓄。新的指数经常一起呈现以便能做简单的比较。

相关知识

想成为喜剧演员的话，你可能要去测评你的笑点指数（确有其事），然后才能被喜剧俱乐部录取。

2.4 练习题

统计素养和批判性思维

1. 指数 一篇新闻报道说2007年的汽油价格指数是每加仑2.27美元，这个陈述有什么错误？
2. 指数 如果将2000年的计算机价格设定为100，以它作为计算机价格指数的基础，2005年的计算机价格指数是20，那么2005年的计算机与2000年的计算机在价格上相比，我们可以知道些什么？
3. CPI 如果食物、服务和住房的价格增长了，消费价格指数一定增长吗？说明原因。
4. CPI 如果消费价格指数增长了，那么工资也

一定增长吗？说明理由。

概念与应用

汽油价格指数：在练习题5~8中，应用表2-1中所列出的汽油价格指数，简要说明你在每一题中的推理过程。

5. 当前数据 假定今天的汽油价格是每加仑3.25美元，以1975年的汽油价格作为参考值，今天的汽油价格指数是多少？
6. 2006年指数 2006年汽油的平均价格是每加仑2.62美元，以1975年的汽油价格作为参考值，2006年的汽油价格指数是多少？

7. 1998 年价格 以 1975 年价格作为参考值, 1998 年的汽油价格指数是 197.5, 1998 年的每加仑汽油价格是多少?

8. 1978 年价格 以 1975 年汽油价格作为参考值, 1978 年的汽油价格指数是 114.6, 1978 年的每加仑汽油价格是多少?

9. 重建汽油价格指数 以 1965 年的汽油价格作为参考值, 重新确定表 2-1 中 7 个年份的汽油价格指数 (提示: 在表格后面重新生成一栏, 内容是本年价格相对 1965 年价格的百分比, 后面一栏则给出汽油价格指数, 参考值 1965 = 100)。

10. 重建汽油价格指数 以 2000 年的汽油价格作为参考值, 重新确定表 2-1 中 7 个年份的汽油价格指数 (提示: 在表格后面重新生成一栏, 内容是本年价格相对 2000 年价格的百分比, 后面一栏则给出汽油价格指数, 参考值 2000 = 100)。

11. 使用汽油价格指数 如果在 1985 年加满你的油箱需要花费 16.74 美元, 那么在 2000 年加满同样的油箱需要花费多少?

12. 使用汽油价格指数 如果在 2000 年加满你的油箱需要花费 23.25 美元, 那么在 2005 年加满同样的油箱需要花费多少?

消费者价格指数: 在练习题 13 ~ 16 中, 应用表 2-2 中的消费者价格指数。

13. 私立大学的费用 4 年制的私立大学每年的平均费用 (包括学费、杂费、住宿费、伙食费等) 从 1980 年的 5 900 美元上升至 2004 年的 27 516 美元。计算从 1980 ~ 2004 年增长的百分比, 并且将它与以消费者价格指数为指标的通货膨胀率作比较。

14. 公立大学的费用 4 年制的公立大学每年的平均费用 (包括学费、杂费、住宿费、伙食费等) 从 1980 年的 2 490 美元上升至 2004 年的 11 354 美元。计算从 1980 ~ 2004 年增长的百分比, 并且将它与以消费者价格指数为指标的通货膨胀率作比较。

15. 南部住房价格 美国南部一所新的独立房屋, 其价格 (中位数) 从 1990 年的 75 300 美元上涨到 2004 年的 155 500 美元。计算 1990 ~ 2004 年的住房价格增长的百分比, 并且将它与以消费者价格指数为指标的通货膨胀率作比较。

16. 西部住房价格 美国西部一所新的独立房屋, 其价格 (中位数) 从 1990 年的 129 600 美元上涨到 2004 年的 241 300 美元。计算 1990 ~ 2004 年的住房价格增长的百分比, 并且将它与以消费者价格指数为指标的通货膨胀率作比较。

住房价格指数: 房地产经纪人使用一个指数比较全国范围内的主要城市的住房价格。在下表中给出了几个城市的指数。如果你知道你所在城镇的某一处房屋的价格, 你可以用这个指数去计算另外一个城镇类似房屋的价格。

$$\text{价格 (其他城镇)} = \frac{\text{价格 (你所在城镇)}}{\text{其他城镇的住房价格指数}} \times \text{你所在城镇的住房价格指数}$$

在练习题 17 ~ 20 中应用如下表的住房价格指数

城市	指数	城市	指数
丹佛	100	波士顿	358
迈阿密	194	拉斯维加斯	101
凤凰城	86	达拉斯	81
亚特兰大	90	夏延	60
巴尔的摩	150	旧金山	362

17. 住房价格 在丹佛一所住房的价格是 30 万美元, 计算在迈阿密和夏延一所类似房屋的价格。

18. 住房价格 在波士顿一所住房的价格是 50 万美元, 计算在巴尔的摩和凤凰城一所类似房屋的价格。

19. 住房价格 在夏延一所住房的价格是 25 万美元, 计算在旧金山和波士顿一所类似房屋的价格。

20. 住房价格 在波士顿一所住房的价格是 100 万美元, 计算在旧金山和夏延一所类似房屋的价格。

网络和远程项目

请在网址 www.aw.com/bbt 上打开第 2 章的相关链接 “links for internet projects”。

21. 消费者价格指数 浏览消费价格指数主页, 寻找近期关于 CPI 的图形指数的新闻, 总结新闻中发布的内容以及 CPI 的重要趋势。

22. 产品价格指数 浏览产品价格指数主页, 阅读以往的历史数据和近期新闻发布的数据, 写一篇简短的总结描述 PPI 的用途, 以及它和 CPI 的区别。对 PPI 的重要近期发展趋势也作一个总结。

23. 消费者信心指数 使用搜索引擎寻找近期关于消费者信心指数的新闻, 在研究新闻以后, 写出一篇简短的总结, 概括消费者信心

指数测量的是什么内容，并且描述消费者信心指数近期的发展趋势。

24. 人类发展指数 美国发展计划经常发布它的人类发展报告，这个报告最重要的一项内容就是人类发展指数，主要测量一个国家发展的三个主要方面的成就：预期寿命、教育水平和调节收入。寻找这个报告的近期副本，并且研究 HDI 是如何定义和计算的。
25. 便利商店指数 去本地的一家超级市场，计算它的几种主要商品价格，比如面包、牛奶、水果和咖啡。计算这些项目的总体价格，然后去一些小的便利店，计算同样项目的价格。以超级市场的总价作为参考值，计

算便利店的价格指数。

26. 汽油价格 搜集数据，制作一个有说服力的图表，写出有说服力的证据来支持或者反驳下面关于汽油价格的表述：汽油价格在过去的 30 年没有下降与生活价格是有关联的。

新闻中的统计学

27. 消费者价格指数 寻找一则近期的新闻报道，包括消费者价格指数的相关内容。简要描述消费者价格指数多么重要。
28. 指数 寻找一则近期的新闻报道，包括一些不是消费者价格指数的其他指数，描述这个指数及其意义，并且讨论这个指数在报道中的重要性。

第2章复习题

1. 泰坦尼克号 当时在泰坦尼克号上的人总共是 2 223 人，当泰坦尼克号在 1912 年 4 月 15 日星期一沉没时，其中 31.76% 的人生还。
- (1) 计算当泰坦尼克号沉没时实际生还的人数。
- (2) 生还人数是一个离散数据还是一个连续数据？请加以解释。
- (3) 当时在泰坦尼克号上面的人们，其中 531 人是妇女和儿童，泰坦尼克号乘客中妇女和儿童的百分比是多少？
- (4) 当时在泰坦尼克号上面共有 45 个女孩，男孩子比女孩子多 42%，在泰坦尼克号上一共有多少个男孩？
- (5) 如果我们搜集泰坦尼克号上面乘客的年龄，这些年龄是一个什么水平的测量（称名、顺序、等距、比率）？
- (6) 如果按照男人、女人、男孩子、女孩子这种分类列出泰坦尼克号上面的全部乘客，这是一个什么水平的测量（称名、顺序、等距、比率）？
2. 美国在线（AOL）投票 美国在线 3 309 人的投票中，66% 的人说他们有时会在肯德基就餐，26% 的人说他们从来都不在肯德基就餐，8% 的人说他们经常在肯德基就餐。
- (1) 说他们从来不在肯德基就餐的人数是多少？
- (2) 如果 1 224 人回答说在消除反式脂肪方面肯德基既没有好处也没有坏处，那么做出这种陈述的人数百分比是多少？
- (3) 如果给出的可选项是总是、有时、从不，这些是一个什么水平的测量（称

名、顺序、等距、比率）？

- (4) 已知这个投票是通过网页上的问题询问美国在线的使用者所得到的，关于这个投票结果你会得到什么结论？这个结果能够反映一般大众的观点吗？
3. 卫生保健费用 在美国，卫生保健费用从 1973 年的 800 亿美元增长到 2004 年的 1.8 万亿美元。消费者价格指数 1973 年是 44.4，2004 年是 188.9 (1982 ~ 1984 = 100)。从 1973 ~ 2004 年，将卫生保健费用的变化与以消费者价格指数为指标的通货膨胀率进行比较。
4. 最低保障工资 下表列出了 60 年来的联邦每小时的最低保障工资。分别以当时的实际美元数和 1996 年的美元价值数来计算。除 2006 年之外，下表完全反映了每年最低保障工资的变化（数据来源于美国劳工部）。

年份	当时的实际美元数	以 1996 年美元价值计算
1938	0.25	2.78
1939	0.30	3.39
1945	0.40	3.49
1950	0.75	4.88
1956	1.00	5.77
1961	1.25	6.41
1967	1.40	6.58
1968	1.60	7.21
1974	2.00	6.37
1976	2.30	6.34
1978	2.65	6.38
1979	2.90	6.27
1981	3.35	5.78

(续)

年份	当时的实际美元数	以 1996 年美元价值计算
1990	3.50	4.56
1991	4.25	4.90
1996	4.75	4.75
1997	5.15	5.03
2006	5.15	4.01

- (1) 按照这个表, 1938 年的 0.25 美元相当于 1996 年的多少美元?
- (2) 按照这个表, 1956 年的 1.00 美元相当于 1996 年的多少美元?
- (3) 为什么 2006 年的最低保障工资实际上的美元数要大于换算成的 1996 年的美元数?

第2章小测验

1. 作为国家健康研究的一部分, 随机选择被试, 记录他的眼睛颜色, 这些眼睛颜色是一个什么水平的测量 (称名、顺序、等距、比率)?
2. 作为国家健康研究的一部分, 随机选择被试, 记录其头围, 这些测量值是离散的还是连续的?
3. 如题 2 中所描述的, 测量的头围数值是一个什么水平的测量 (称名、顺序、等距、比率)?
4. 一个研究者测量了一个被试的头围, 记录的数据是 45.4 厘米, 但是被试的头围实际数值是 55.4 厘米, 绝对误差是多少?
5. 一个研究者测量了一个被试的头围, 记录的数据是 45.4 厘米, 但是被试的头围实际数值是 55.4 厘米, 相对误差是多少?
6. 在盖洛普对 1 038 位成人的调查中, 540 人说二手烟是非常有害的。说二手烟非常有害的成人百分比是多少?
7. 在盖洛普对 1 038 位成人的调查中, 5% 的人说二手烟根本没有危害, 说二手烟根本没有

危害的人数是多少?

8. 两个不同的学生测量了他们指导教师的身高, 教师的实际身高是 178.44 厘米, 第一个学生测量的结果是 178 厘米, 第 2 个学生的测量结果是 179.18 厘米。哪一个学生的测量更准确? 哪一个学生的测量结果更精确?
9. Telektronics 公司已经运作 5 年了, 下表列出了这些年的年利润。使用第一年的利润作为参考值, 计算第 2 年利润的指数。

年份	利润 (美元)
1	12 335
2	15 257
3	23 444
4	31 898
5	47 296

10. 参考第 9 题中的表格, 如果第 6 年的利润要比第 5 年的利润多 12%, 第 6 年的利润是多少?

| 聚焦政治 |

谁会从减税中受益

政治家有种特别的能力, 总能找到哪怕极细微的数字来支持他们的观点。仔细分析图 2-4 显示出的两种图形, 显示的是同一个减税提议, 但似乎支持两种完全不同的结论 (这项减税内容并没有成为法律)。

图 2-4a 呈现的数据是由支持减税的共和党提供的。它显示了减税可让所有纳税的家庭从减税中获益的百分点大体是相当的, 中等收入的家庭受益要稍多于贫困家庭和富裕家庭。图 2-4b 呈现的数据是由反对减税的民主党提供的。这个图表显示受益者明显比例失调, 对于富有者更为有利。我们假定这两种说法都掩盖了事实真相, 怎么得到这两种截然不同的结论的呢? 答案在于两者选择了怎样的方式来计算减税的“获益”。

共和党计算的是每组家庭通过减税受益的平均值。例如, 图 2-4a 中最后一个柱状图显示了收入超过 200 000 美元的家庭可以获得的减税平均值为 2.9%。当然, 2.9% 的减税对于高纳税人与那些低纳税人相比较实际上意味着更多的美元。比如, 应交纳 100 000 美元税款的纳税人可以节省 2 900 美元, 而只需要交纳 1 000 美元税款的人只节省了 29 美元。

民主党计算的是总的获益分配到各组的百分比。例如, 图 2-4b 中最后一个柱状图显示了收入超过 200 000 美元的家庭从减税中得到了总获益的 28.1%。但并不意味着这些家庭可以得到 28.1% 的减税率。究其原因, 我们设想一下, 如果针对每个人一概减税同样比例, 其效果会是怎样。收入超过 200 000 美元的家庭支付了



美国政府所征收全部税收的 1/4，这些家庭将享受全面减税获益的大约 1/4。例如，如果政府征收了 1 000 000 美元税金，一个全面的 10% 的减税意味着所有纳税人将节省 100 000 美元。这样的话，由于收入超过 200 000 美元的家庭缴纳了 1 000 000 美元税金中的 1/4，他们也就享受了节省的那 100 000 美元的 1/4。因此这些家庭将从减税中得到 25% 的获益，而实际上他们仅仅享受到 10% 的减税率。哪种说法更公平？实际上，哪一种说法都不确切。共和党忽略了富有者仍然获取了减税所节省下来的大部分美元这样一个事实，而民主党则选择忽略了富有者仍然缴纳了大多数税金这个现实。遗憾的是，当以数据的方式来呈现时，尤其是跟政治扯上关系的时候，这种“选择性事实”会变得普遍。

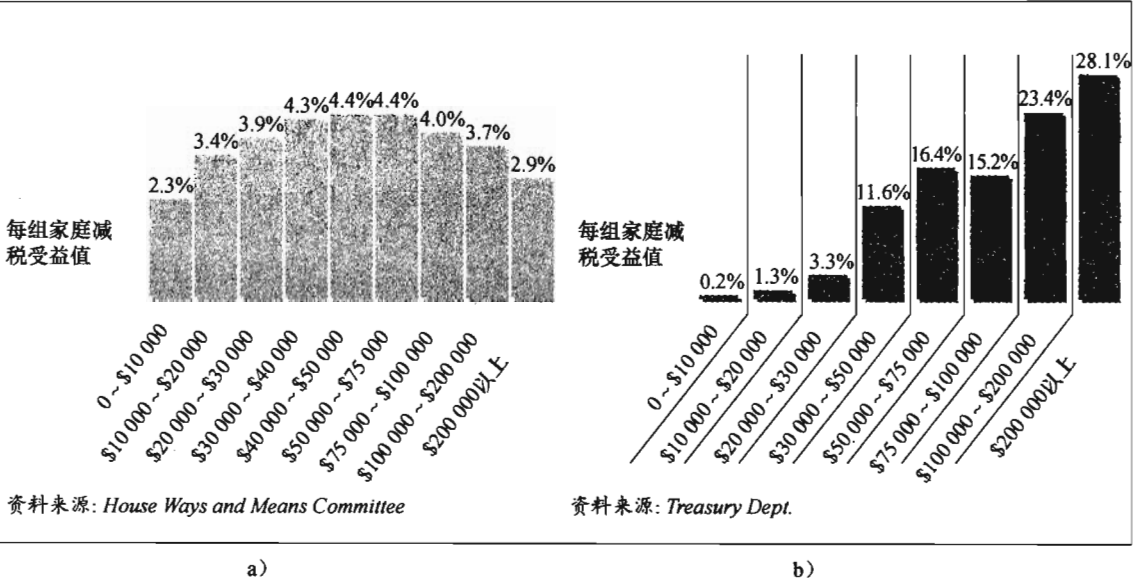


图 2-4 减税的两种看法
a) 共和党数据：全民平等减税 b) 民主党数据：少数人获益

问题讨论

- 图 2-4a 中的百分点显示了不同收入群体相对应的减税获益。对于每个群体来说，这些相对的受益跟绝对的受益又如何对应上呢？（提示：在估算绝对税收获益时，应注意任何一个家庭支付的总税额是他们收入的某个百分比，而且低收入家庭支付的税额，通常在他们的收入中占更少百分比。）
- 两个图呈现差异的第 2 个原因是两党对收入的定义不同。例如，民主党选择将全部利润分配给股票持有者作为个人收入，也就意味着民主党认为持有股票的人将会有更高的收入。这种界定收入的差异是如何影响两个图表的？
- 你认为图 2-4 中是否有一个图可以准确揭示被提议的减税的全部“公平”。如果有，是哪一个？原因是什么？如果没有，你认为怎样描述数据才可以更加公平？
- 今年美国国会或者美国总统有没有减税或增税的计划？讨论现行提案的公平性。

建议阅读

- “Two Views of a Tax Cut.” *New York Times*, April 7, 1995.
- Andrews, Edward. “Your Taxes: Cracking the Tax Code.” *New York Times*, February 11, 2006.
- Andrews, Edward. “Tax Cuts Offer Most for Very Rich, Study Says.” *New York Times*, January 8, 2007.

| 聚焦经济 |

我们的生活水平提高了吗

我们中的大多数人都希望自己的生活水平可以逐年提高，理论上通过以下例证容易说明：如果你的收入增长速度快于通货膨胀速度，你的生活水平将会提高；反之，如果你的收入增长速度赶不上通货膨胀速度，

你的生活水平将会下降。

尽管如此，事实上情况要复杂得多。例如，即使你的收入显著增加，如果你增加了新的开销项目，你的生活水平也可能下降，这些开销包括你的保险中没有覆盖的有关健康方面的风险；一个孩子或你自己的大学学习的费用；或者因抵押贷款的浮动利率上调而增加的月供。

相类似的复杂性也影响着国民生活水平问题。经济学家普遍认为美国人的平均生活水平在第二次世界大战后到20世纪70年代初之间，以及整个20世纪90年代都有实质性的较大提高。但是，在20世纪七八十年代以及21世纪的最初几年中，除非你处于财富金字塔的顶端（他们的收入持续增长，见第4章的最后“聚焦经济：富人变得更富了吗”栏目），关于一般人生活水平是否提高存在分歧和争论。

争论的关键问题在于：真正的收入，即基于通货膨胀校正后的收入，是上升还是下降了。于是问题转换成了如何测量通货膨胀。通常采用2.4节中提到的消费者价格指数（CPI）来对通货膨胀进行测量。如果采用CPI作为通货膨胀的测量指标，图2-5显示了1980年以来大多数美国人的真实薪水情况。例如，真实薪水的变化为0代表在这一年中实际薪水和CPI的变化恰好同步。当实际薪水的增长大于CPI，真实薪水的变化呈正值；而如果实际薪水的增长小于CPI，真实薪水的变化则呈负值。

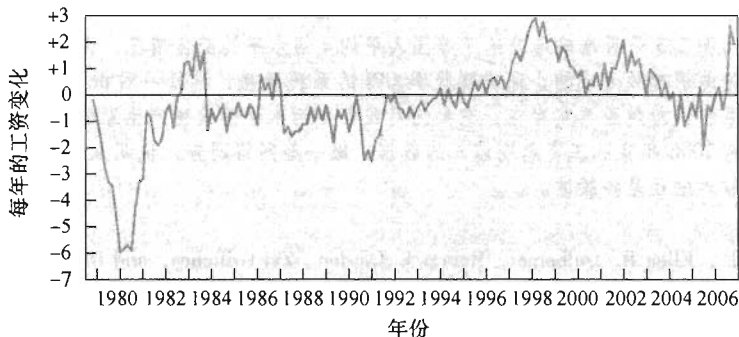


图2-5 私营企业工作的非管理职位工人经CPI调整后的工资变化。这些工人涵盖了劳动力的大约80%，不包括政府雇员，负值表示实际工资下降而正值表示实际工资提高。
《纽约时报》，2007年1月3日

资料来源：U. S. Department of Labor；University of Michigan；National Bureau of Economic Research.

值得注意的是，照此算来，自20世纪70年代后期以来，真实薪水在大多数年份都是降低的（呈负值）。的确，如果你要计算整个时期（往后再推几年）总的变化，图片显示了2006年的真实薪水要比30年前的水平更低。在2006年的选举中，民主党人在国会议员竞选时使用这些数据声明，大部分美国人的薪水已经将近30多年停滞不前了。

但我们生活水平的变化真的如图所示如此惨淡吗？抑或更加惨淡？答案取决于你问的是谁，因为关于CPI，到底是放大、缩小还是精确测量了通货膨胀，不同的经济学家有不一样的观点。

让我们从认为CPI会放大通货膨胀的观点开始。如果是这样的话，真实的通胀率应该比CPI显示的，也就是说图2-5中的基线要比实际较高。例如，如果CPI放大了每年通胀率的一个百分点，基线实际该是图2-5中的-1的高度水平，而不是0。这个变化使得图中大部分年份显示的数据将由负值变成正值，这就意味着在过去的30多年中平均生活水平实际上显著提高了，而不是停滞不前。

争论的焦点，大多数清楚地记载于1996年的“Boskin报告”（见建议阅读），CPI的计算标准里所包含的系统误差放大了生活消费水平的增长。报告指出了能够导致这种结果的两种主要的系统误差。首先，CPI是依据特定商店特定商品从这个月到下个月价格的变动而计算来的。实际上，当一个商店的商品价格上涨时，消费者常常会去另一个商店购买比较便宜的。如果一种商品的价格上调太多，消费者会选择一款类似的但价格更低的商品，这意味着不过与更换商品品牌一样简单，因此消费者发现他们的实际开销增加并不像CPI显示的那么多。其次，CPI考察的是“典型”消费者在任意给定时间购买商品的价格，但是这些商品是随时变化的，

特别是进入高科技时代，但是 CPI 并没有考虑到这些变化对人们生活水平的影响。例如，20 年前，没有一个人拥有现在的 DVD 机，也没有一台 iPod，也没有一台高清电视，也没有接入互联网的高速电脑。大多数经济学家认为这些产品的使用提高了我们的生活水平，但是在 CPI 计算中这些产品并没有“典型”消费者现在购买商品的作用大。由于这两种类型的系统误差的存在，“Boskin 报告”指出，CPI 会将每年真实的通货膨胀率放大 0.8% ~ 1.6%。

当大多数经济学家都认同了“Boskin 报告”说明的系统误差时，另外一些经济学家却指出了一些相反方向的系统误差。例如，许多中年工人现在承担着照料父母的大笔开销，这些在几十年前是不存在的。另外，从平均的角度说来，我们的收入越来越多的部分用于支付养老金和医疗保险，而这些过去是负担很少的。这些开销使得我们只能将收入的很少一部分，用于一般意义上认为的提高生活水平的商品上。的确，有些经济学家认为这些变化如此明显，CPI 实际上是缩小了通货膨胀水平，这样说来，我们的生活水平比图 2-5 显示的工资变化还要低得更多。

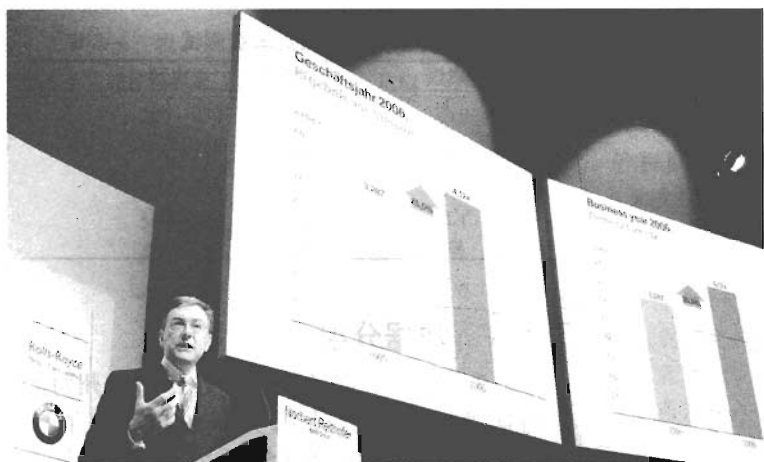
应该搞清楚这种争论对于我们的意义，CPI 的计算结果和实际工资的统计数据都是非常清楚的，但仍然会有持不同政策和经济理念的经济学家从此数据中得到截然不同的结论。那么我们的生活水平究竟是提高了还是下降了，关于这样的问题，为什么会关系到本着诚信善意原则的争论呢？这也同样说明，对于任何一种类型的统计研究的结论，我们都应该去认真思考，而不是仅仅接受具有表面意义上新奇的政治口号。

问题讨论

1. 根据图 2-5 呈现的数据，描述过去的 30 年里实际工资是怎么变化的。例如，描述哪个阶段工资提高了，哪个阶段工资降低了。
2. 总体看来，你认为图 2-5 是否准确地表达了美国人平均生活水平的变化情况，说明你的观点。
3. 思考文章中所讨论的引起放大或缩小通货膨胀率 CPI 的系统误差，选择一种做更深入的调查。结合大量的实例，分析这个系统误差是否真实存在，它如何对我们生活水平的改变产生影响，得出你自己的结论。
4. 在图 2-5 中，注意 2006 年实际工资急剧增长的数据，做一些网络调查，说明从那时候起工资发生了什么变化，这个增长是暂时的还是持续的。

建议阅读

1. Boskin, Michael J., Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale W. Jorgenson.
2. “Toward a More Accurate Measure of the Cost of Living.” Final Report to the Senate Finance Committee, from the Advisory Commission to Study the Consumer Price Index. Washington, DC, Senate Finance Committee, 1996.
3. Greenhouse, Steven. “Falling Fortunes of Wage Earners.” *New York Times*, April 12, 2005.
4. Gross, Jane. “Elder-Care Costs Deplete Savings of a Generation.” *New York Times*, December 30, 2006.
5. Leonhar, David. “It’s the Year to Keep an Eye on Paychecks.” *New York Times*, January 3, 2007.



图表最大的价值在于强迫我们注意到不希望看到的东西

——约翰·塔基

第 3 章 数据的直观呈现

学习目标

■ 3.1 次数分布表

能够编制和解释次数分布表。

■ 3.2 绘制数据分布图

能够编制和解释基本的条形图、点柱图、饼图、直方图、茎叶图、线图和时序图。

■ 3.3 媒体中的图表

掌握怎样解释多种类型的复杂图表，这些图表经常出现在新闻媒体中。

■ 3.4 有关图表的几个注意事项

学会批判性地评估图表，明确图表通常被曲解的方式。

如 果你看报纸，一个公司的年度报表或者政府研究，你总是能够看到统计图表。一些图表非常简单，另外一些比较复杂。一些图表使你很容易理解数据，另一些会让你困惑甚至误解。在这一章，我们将探讨统计图表通常以什么样的方式呈现。在今天的信息社会，用图表来表达概念是非常有效的，这一章的技术几乎在每一种职业中都是非常重要的。



3.1 次数分布表

德兰尼教授记录了如下给她的 25 个学生的一次论文测验的成绩单。

A C C B C D C C F D C C C

B B A B D B A A B F C B

这个成绩单包括了从 A 到 F 的各种成绩，但是非常不容易浏览。一个更方便的呈现这些数据的方式，就是编制一个记录各种成绩出现次数（也叫次数分布）的表，结果如表 3-1 所示，叫做次数分布表。5 种可能的成绩（A、B、C、D、F）叫做这个表的分类（或者叫组）。

表 3-1 一组论文测验成绩的次数分布表

成绩	次数
A	4
B	7
C	9
D	3
F	2
总计	25

定义

一个基本的次数分布表（frequency table）有两栏：

- ◆ 一栏列出数据的所有分类。
- ◆ 另一栏列出每一种类别的次数，或者是这种类别相对应的数据值。

例 3-1 味觉测试

洛基山饮料公司想对它的一种新产品珊瑚可乐进行产品调研，对 20 个人进行了味觉测试，每一个人均被要求在一个 5 点量表中对这种可乐的味道进行打分。

味道很差 1 2 3 4 5 味道很好

20 个人打分如下：

1 3 3 2 3
3 4 3 2 4
2 3 5 3 4
5 3 4 3 1

针对这个数据编制一个次数分布表。

解答 关注变量是味道，这个变量有 5 个可能取值：味道分类从 1 ~ 5（注意这个包括 5 种类别的数据是定性数据，是测量的顺序水平）。我们编制一个次数分布表，左边一栏是 5 个类别，右边一栏是对应的次数，如表 3-2 所示。

表 3-2 味觉等级测试

成绩	次数
1	2
2	3
3	9
4	4
5	2
总计	20

3.1.1 数据分组

表 3-3 中的数据呈现的是美国 50 个州每人年均耗费的能量，使用英制热单位（百万 BTUs）。在这组数据中的 50 个数字（不计算全美国的平均值）小到每人 2.12 亿 BTUs（罗得岛），大到每人 11.75 亿 BTUs（阿拉斯加），大部分数字仅出现一次，我们怎样有效地编制这种数据的次数分布表？

表 3-3 每人年均耗费的能量（百万 BTUs）

州	每人年均	州	每人年均	州	每人年均
亚拉巴马	447	路易斯安那	822	俄亥俄	349
阿拉斯加	1 175	缅因	366	俄克拉何马	425
亚利桑那	246	马里兰	281	俄勒冈	295
阿肯色	415	马萨诸塞	248	宾夕法尼亚	321
加利福尼亚	229	密歇根	313	罗得岛	212
科罗拉多	297	明尼苏达	355	南卡罗来纳	389
康涅狄格	255	密西西比	411	南达科他	345
特拉华	383	密苏里	322	田纳西	388
佛罗里达	252	蒙大拿	410	得克萨斯	560
佐治亚	343	内布拉斯加	372	犹他	296
夏威夷	248	内华达	292	佛蒙特	252
爱达荷	341	新罕布什尔	254	弗吉尼亚	329
伊利诺伊	310	新泽西	298	华盛顿	316
印第安纳	470	新墨西哥	353	西弗吉尼亚	433
艾奥瓦	400	纽约	220	威斯康星	335
堪萨斯	410	北卡罗来纳	314	怀俄明	919
肯塔基	456	北达科他	624	美国平均	339

注：数据包括各种能源使用，包括家庭、商业、工业和交通。
资料来源：United States Energy Information Administration（2007 tables, showing 2003 data）。

答案是按照一定的数据取值范围进行分类，例如，我们可以将所有的取值在 200 ~ 299 百万 BTUs 范围内的数据归为 1 类，取值在 300 ~ 399 百万 BTUs 范围内的数据归为 1 类，依此类推。

这样，我们就可以对每类数据进行计数，最后编制成次数分布表（见表 3-4）。这个过程叫做数据分组（binning），每个类别实际上是一个单独的组，我们可以将数据按照取值分进相应的组内。

表 3-4 表 3-3 人均耗费能量的次数分布表

每人年均耗费的能量 (百万 BTUs)	次数 (州的数量)
200 ~ 299	16
300 ~ 399	19
400 ~ 499	10
500 ~ 599	1
600 ~ 699	1
700 ~ 799	0
800 ~ 899	1
900 ~ 999	1
1 000 ~ 1 099	0
1 100 ~ 1 199	1
总计	50

相关知识

1BTU,是指将一磅水的温度提高1°F所需要的能量,它等于252卡路里或者1055焦耳。

例 3-2 道琼斯指数样本股票

关于道琼斯工业指数的 30 只样本股票,表 3-5 呈现了它们的年收入(亿美元)、年回报率和在美国大公司的《财富》500 强中的排名。编制一个关于收入的次数分布表,从正反两方面讨论分组的选择。

解答 收入数据的范围从 216 亿美元(麦当劳)到 3 511 亿美元(沃尔玛)。在这个范围内有很多可能的数据分组的方式,下面是一种较好的方式,理由如下:

- (1) 我们定义数据分组范围从 0 ~ 4 000 亿美元,这就涵盖了所有数据,这个范围包括了最低的和最高的收入数值。
- (2) 每一组的组距定义为 500 亿美元,这样我们可以将 0 ~ 4 000 亿美元分成 8 组,500 亿美元的间距是个比较方便的数字,这样会帮助我们更容易读表。
- (3) 由于给出的数据值有效数字是 1 亿美元,因此我们也这样定义每组的上下限,以使有效数字不被覆盖,这样,分组为 0 ~ 499 亿美元、500 亿 ~ 999 亿美元,依此类推。

表 3-6 显示的次数分布表的结果。

表 3-5 道琼斯工业平均指数样本股票 2006 年的数据

公司	收入 (亿美元)	回报率	排名	公司	收入 (亿美元)	回报率	排名
沃尔玛	3 511	0.1%	1	强生制药有限公司	533	12.4%	35
埃克森美孚公司	3 473	39.2%	2	辉瑞制药有限公司	524	15.3%	39
美国通用汽车公司	2 073	64.2%	3	联合科技公司	478	13.7%	42
通用电气公司	1 683	26.5%	5	微软	443	15.8%	49
花旗银行	1 468	19.4%	8	卡特彼勒公司	416	7.9%	55
AIG	1 132	6.0%	10	英特尔	354	-17.2%	62
摩根大通公司	1 000	25.5%	11	迪士尼	343	44.5%	64
弗莱森电讯	932	34.6%	13	霍尼韦尔国际公司	314	24.2%	69
惠普公司	917	45.3%	14	美国铝业公司	309	3.5%	71
IBM	914	19.8%	15	杜邦公司	290	18.7%	74
家得宝公司	908	0.9%	17	美国运通公司	271	19.1%	79
阿尔特里亚集团	703	19.9%	23	可口可乐公司	241	23.1%	94
宝洁公司	682	13.3%	25	3M 公司	229	3.0%	97
AT&T	631	52.9%	27	默克制药公司	226	42.9%	99
波音公司	615	28.4%	28	麦当劳	216	34.6%	108

表 3-6 表 3-5 数据中每年收入的次数分布表

年收入 (亿美元)	次数 (公司数)
0 ~ 49.9	13
50 ~ 99.9	10
100 ~ 149.9	3
150 ~ 199.9	1
200 ~ 249.9	1
250 ~ 299.9	0
300 ~ 349.9	1
350 ~ 399.9	1
总计	30

相关知识

《华尔街日报》编辑挑选的 30 只股票构成了道琼斯指数,股票的名单有时也会变化。例如,2004 年 4 月,AT&T 公司、柯达公司、国际纸业等被从指数中移除。取而代之的是美国铝业公司、辉瑞制药有限公司、Verizon 公司。不久,AT&T 公司重回指数,因为名单中另一家公司西南贝尔通信公司与



之合并，并且更名为 AT&T 公司。

思考题

试想表 3-6 中数据其他 3 种可能的分组方式，取值范围从 0 ~ 4 000 亿美元分成 4 组，取值范围从 0 ~ 3 750 亿美元分成 11 组，取值范围从 0 ~ 3 600 亿美元分 36 组。简要地从正反两方面讨论这几种选择。

3.1.2 相对次数分布

让我们来看看表 3-1 中的论文测验成绩单。我们不仅想知道每一种成绩的学生人数，也想知道每种成绩的学生所得的分数或所占的百分比。我们称每个类别所占的这个分数（或者是比率或者是百分比）叫做相对次数（relative frequencies）。例如，25 个学生中的 4 个成绩为 A，其相对次数就是 4/25，或者是 0.16，或者是 16%。表 3-7 再现的是表 3-1 中的数据，但是这次增加了一栏相对次数。

表 3-7 相对次数分布表

成绩	次数	相对次数
A	4	4/25 = 0.16
B	7	7/25 = 0.28
C	9	9/25 = 0.36
D	3	3/25 = 0.12
F	2	2/25 = 0.08
总计	25	1

相对次数的总和必须等于 1（或者 100%），因为每组的相对次数都是总次数的一个分数（有些时候，由于四舍五入取整的原因会和总数 1 之间有些很小的差距）。

定义

- ◆ 任何一组的相对次数都是分到本组的次数所占总次数的比例或者百分比。

相对次数 = 本组次数 / 总次数 (3-1)

3.1.3 累加次数分布

方法注解

大多数次数表从最低组开始，但是成绩表是从最高组（A）开始，分类顺序并不影响次数和相对次数，但是会影响累加次数。例如，如果表 3-8 是相反的顺序，成绩 C 对应的累加次数就变成 C 以下次数（表 3-8 是 C 以上的次数）。

再看一次表 3-1 的论文测验成绩，如果我们想知道成绩在 C 及 C 以上的人数有多少，我们可以将成绩是 A、B 和 C 的人数相加，结果是 20 个人成绩在 C 及 C 以上。像这样进行运算的就构成了累加次数，或者在某一分组所对应的数据是此前所有组的次数和。表 3-8 重现了表 3-1 的数据，但这一次新加的一栏为累加次数。注意累加次数的最后一组次数一定等于总次数。

表 3-8 累加次数分布表

成绩	次数	累加次数
A	4	4
B	7	4 + 7 = 11
C	9	4 + 7 + 9 = 20
D	3	4 + 7 + 9 + 3 = 23
F	2	4 + 7 + 9 + 3 + 2 = 25
总计	25	25

定义

- ◆ 任意一组的累加次数是本组前面所有组的次数之和。

应该牢记只有当分类是有序的时，累加次数才有意义，也就是说，只有测量数据是顺序水平、等距水平和比率水平时我们才能够应用累加次数。如果测量的是称名水平的数据时，累加次数是无意义的。

例 3-3 味觉测试再分析

使用例 3-1 中的味觉测试数据，编制一个包括相对次数和累加次数的次数分布表。给可乐打分最高的一组所占百分比是多少？给可乐打分低的 3 组合起来所占百分比是多少？

解答 我们通过将每一组次数除以总次数 20 得出相对次数，通过将每一组对应的本组及以下组的次数依次相加得出累加次数。表 3-9 呈现了这个结果。给可乐打分最高一组的相对次数为 0.1，或 10%，给可乐打分低的 3 组合起来是 14 个人，占总数 20 的 70%。

表 3-9 味道测试的相对次数和累加次数

成绩	次数	相对次数	累加次数
1	2	$2/20 = 0.10$	2
2	3	$3/20 = 0.15$	$2 + 3 = 5$
3	9	$9/20 = 0.45$	$2 + 3 + 9 = 14$
4	4	$4/20 = 0.20$	$2 + 3 + 9 + 4 = 18$
5	2	$2/20 = 0.10$	$2 + 3 + 9 + 4 + 2 = 20$
总计	20	1	20

方法注解

累加次数除以总次数叫做相对累加次数。例如,表 3-9 中 3 对应的相对累加次数就是 $14/20 = 0.7$,意思是指 70% 的数据值在 1, 2, 3 这三组中。

例 3-4 能量数据

对于表 3-4 呈现的能量数据的次数分布表,添加两栏,内容为相对次数和累加次数。讨论看起来特别有启发性的令人惊奇的趋势。

解答 我们可以用每一组的次数除以总次数得到相对次数,本例中的总数是 50,我们可以将每一组的次数及以下各组的次数依次相加得到本组对应的累加次数,表 3-10 呈现了这个结果。

表 3-10 分组的能量消耗数据

每人年均耗费的能量 (百万 BTUs)	次数	相对次数	累加次数
200 ~ 299	16	0.32	16
300 ~ 399	19	0.38	35
400 ~ 499	10	0.20	45
500 ~ 599	1	0.02	46

(续)

每人年均耗费的能量 (百万 BTUs)	次数	相对次数	累加次数
600 ~ 699	1	0.02	47
700 ~ 799	0	0.00	47
800 ~ 899	1	0.02	48
900 ~ 999	1	0.02	49
1 000 ~ 1 099	0	0.00	49
1 100 ~ 1 199	1	0.02	50
总计	50	1	50

本表中具有启发性的令人惊奇的事实是各州每人年度所消耗能量是不同的。例如,相对次数一栏告诉我们 38% (或者 0.38) 的州每人每年消耗能量在 300 ~ 399 百万 BTUs,累加次数显示 50 个州中的 35 个州分布在前两个组中,或者说每人年度消耗能量在 400 百万 BTUs 以下。3 个年度每人消耗能量最高的州,相比任何一个年度每人消耗能量低的 35 个州,都要高出两倍以上(我们将在第 4 章使用统计学术语进行讨论,我们说这 3 个州属于异常值,因为它们的值与大多数其他州相比有很大不同)。

思考题

仔细说明例 3-4 中的表 3-10,例如,在 200 ~ 299 百万 BTUs 这组的相对次数 0.32 意思是说,50 个州中有 32% 每人年消耗能量是在这个范围内。是不是这也意味着所有美国人中有 32% 每人年消耗能量在 200 ~ 299 百万 BTUs 区间之内? 为什么? (提示: 返回到表 3-3 然后考虑各州的人数)

3.1 练习题

统计素养和批判性思维

1. 次数分布表 什么是次数分布表? 说明通过分类(或者分组)和次数我们想要做什么。
2. 相对次数 一个次数分布表包括 5 组,对应的次数分别是 2, 9, 14, 12, 3。这 5 组各自的相对次数是多少?
3. 累加次数 一个次数分布表包括 5 组,对应的次数分别是 2, 9, 14, 12, 3。这 5 组各自对应的累加次数是多少?
4. 次数分布表 次数分布表显示表格中第一组对应的组限是 15.0 ~ 15.9。使用关于这个次

数表的唯一信息,能否确定 12 个原始样本数据值归入相应的分组中? 说明理由。

判断正误: 判断练习题 5 ~ 8 中的陈述是有意义的(表述正确)还是无意义的(表述错误),并加以解释。并非所有的陈述都有明确的答案,所以你的解释比判断更重要。

5. 次数分布表 一个朋友告诉你他的一个次数分布表有两栏,分别标记为所在州和收入中位数。
6. 相对次数 一个次数分布表中分类为 A 的相对次数是 1.4。
7. 累加次数 一个次数分布表中分类为 A 的累

加次数是1.4。

8. 组距 对于一个给出的数据组，当组距的宽度减小时，分组数会增多。

概念与应用

9. 次数分布表练习 迪亚兹教授记录了下列他所教课程的学生们的最终成绩。

A A A A B B B B B B B C
C C C C C C C D D D F F

针对这些成绩编制一个次数分布表，包括相对次数和累加次数，并简要地解释每列的意义。

10. 次数分布表练习 纽约市的一本指导书列出了5家5星级酒店、10家4星级酒店、20家3星级酒店、15家2星级酒店、5家1星级酒店。针对这些星级酒店编制一个次数分布表，包括相对次数和累加次数，并简要地说明每一列的意思。

11. 可口可乐的重量 根据下面所列出的36听普通可口可乐的净重（单位为磅）编制一个次数分布表。最低组的下限从0.7900开始，组距设定为0.0050磅。讨论你所编制的次数分布表。

0.819 2 0.815 0 0.816 3 0.821 1
0.818 1 0.824 7 0.806 2 0.812 8
0.817 2 0.811 0 0.825 1 0.826 4
0.790 1 0.824 4 0.807 3 0.807 9
0.804 4 0.817 0 0.816 1 0.819 4
0.818 9 0.819 4 0.817 6 0.828 4
0.816 5 0.814 3 0.822 9 0.815 0
0.815 2 0.824 4 0.820 7 0.815 2
0.812 6 0.829 5 0.816 1 0.819 2

12. 健怡可乐的重量 根据下面所列出的36听健怡可乐的净重（单位为磅）编制一个次数分布表。最低组的下限从0.7750开始，组距设定为0.0050磅。讨论你所编制的次数分布表。

0.777 3 0.775 8 0.789 6 0.786 8
0.784 4 0.786 1 0.780 6 0.783 0
0.785 2 0.787 9 0.788 1 0.782 6
0.792 3 0.785 2 0.787 2 0.781 3
0.788 5 0.776 0 0.782 2 0.787 4
0.782 2 0.783 9 0.780 2 0.789 2
0.787 4 0.790 7 0.777 1 0.787 0
0.783 3 0.782 2 0.783 7 0.791 0
0.787 9 0.792 3 0.785 9 0.781 1

13. 奥斯卡电影金像奖最佳男演员 下列的数据呈现了近年来奥斯卡电影金像奖最佳男演员获奖时的年龄。针对如下数据编制一个次数分布表，使用年龄分组，组限为20~29、30~39等依次类推。讨论你从中所发现的现象。

32 37 36 32 51 53 33 61 35
45 55 39 76 37 42 40 32 60
38 56 48 48 40 43 62 43 42
44 41 56 39 46 31 47

14. 体温 下列数据呈现了随机选择的被试的身体温度（华氏温度）。编制一个分成7组的次数分布表，分组如下：96.9~97.2，97.3~97.6，97.7~98.8，依次类推。

98.6 98.6 98.0 98.0 99.0
98.4 98.4 98.4 98.4 98.6
98.6 98.8 98.6 97.0 97.0
98.8 97.6 97.7 98.8 98.0
98.0 98.3 98.5 97.3 98.7
97.4 98.9 98.6 99.5 97.5
97.3 97.6 98.2 99.6 98.7
99.4 98.2 98.0 98.6 98.6
97.2 98.4 98.6 98.2 98.0
97.8 98.0 98.4 98.6 98.6

15. 遗漏的信息 下表呈现了英语课程中的作文成绩。表格是不完整的，根据已知信息将表格中遗漏的信息填写完整。

分类	次数	相对次数
A	?	?
B	?	18%
C	?	24%
D	11	?
E	6	?
总数	50	?

16. 遗漏的信息 下表呈现了戏剧课程中的表演成绩。表格是不完整的，应用已知信息将表格中遗漏的信息填写完整。

分类	次数	相对次数
A	?	1
B	6	?
C	7	?
D	?	23
E	?	25
总数	?	?

17. 不公平骰子 本书的一位作者在1枚骰子上面钻了一个洞,并且在里面灌满了铅,然后抛掷这枚骰子,结果如下表中的次数分布表。

结果	次数
1	27
2	31
3	42
4	40
5	28
6	32

- (1) 按照这个数据,总共抛掷这枚骰子多少次?
 (2) 结果大于2的次数是多少?
 (3) 结果为6的次数所占的百分比是多少?
 (4) 列出各种结果次数所对应的相对次数。
 (5) 列出各种结果次数所对应的累加次数。
18. 说明家庭数据 思考下面关于美国家庭孩子数量的次数分布表。

- (1) 按照这个数据,美国有多少个家庭?
 (2) 有两个以下孩子的美国家庭是多少?
 (3) 没有孩子的美国家庭数占的百分比是多少?
 (4) 有3个以上孩子的美国家庭数占的百分比是多少?

孩子数量	家庭数量(百万)
0	35.54
1	14.32
2	13.28
3	5.13
4及以上	1.97

19. 电脑键盘 传统的电脑键盘配置叫做打字机式(Qwerty)键盘,因为上面一行是字母Qwerty。这种键盘发明于1872年,打字机式键盘强迫人们打字慢一些,避免打字机卡住。德沃夏克键盘发明于1936年,德沃夏克键盘将最常用的按键放到最方便触及的中间一排(基准键位),通过这样安排提供了一个有效的设计。

《发现》杂志上的一篇文章建议通过指向率系统来测量是否方便打字:将基准键位的每一个字母计数为0,将基准键位上面1排的每一个字母计数为1,将基准键位下面1排的每一个字母计数为2。例如,单词

“statistics”在打字机式键盘上最后计数结果是7,而在德沃夏克键盘上最后计数是1,如下表所示。

	s t a t i s t i c s	
打字机式 键盘	0 1 0 1 1 0 1 1 2 0	总和=7
德沃夏克 键盘	0 0 0 0 0 0 0 0 1 0	总和=1

应用这种指向率系统对美国宪法开篇的52个单词进行计数,我们得到如下的数据值。

打字机式键盘单词指向率:

2 2 5 1 2 6 3 3 4 2 4 0 5
 7 7 5 6 6 8 10 7 2 2 10 5 8
 2 5 4 2 6 2 6 1 7 2 7 2 3
 8 1 5 2 5 2 14 2 2 6 3 1 7

德沃夏克键盘单词指向率:

2 0 3 1 0 0 0 0 2 0 4 0 3
 4 0 3 3 1 3 5 4 2 0 5 1 4
 0 3 5 0 2 0 4 1 5 0 4 0 1
 3 0 1 0 3 0 1 2 0 0 0 1 4

- (1) 针对打字机式键盘的单词指向率,编制一个次数分布表,采用组限0~2, 3~5, 6~8, 9~11, 12~14。包括一系列相对次数。
 (2) 针对德沃夏克键盘的单词指向率,编制一个次数分布表,采用组限0~2, 3~5, 6~8, 9~11, 12~14。包括一系列相对次数。
 (3) 按照从第1小题和第2小题中得到的结果,哪一种键盘的设计更有利于打字?并加以解释。
20. 双向次数分布表 统计课程中的学生对学生在高中时代使用的交通工具进行了一项调查。在这些数据中,记录了年龄和从家到学校之间所使用的交通工具。下表是所收集到的一些数据。关于年龄的数据,1=14岁、2=15岁、3=16岁、4=17岁、5=18岁。关于交通工具的数据:1=步行、2=校车、3=公交车、4=自驾车、5=其他。

学生编号	年龄	交通工具	学生编号	年龄	交通工具
1	1	1	4	3	5
2	5	1	5	4	3
3	2	2	6	1	1

(续)

学生编号	年龄	交通工具	学生编号	年龄	交通工具
7	5	2	14	5	5
8	2	1	15	5	5
9	3	4	16	4	4
10	1	3	17	2	2
11	3	5	18	3	1
12	5	5	19	3	3
13	1	2	20	1	4

- (1) 将两个变量年龄和交通工具进行归类，分别是定量变量和定性变量，说明每个变量的测量水平。
- (2) 为了便于分析和呈现数据，数据必须按照两个变量分别进行分组处理。对 $5 \times 5 = 25$ 种（年龄/交通工具）分类分别进行统计，并且将相应的次数填写进下表中相应的空格中。

		交通工具				
		1	2	3	4	5
年龄	1					
	2					
	3					
	4					
	5					

网络和远程项目

请在网站 www.aw.com/bbt 上打开第3章的有用链接“links for internet projects”。

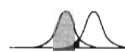
- 21. 能源表 美国能源资料协会（EIA）的网页上提供数十个与能源使用、能源价格和污染有关的表格。寻找一个你感兴趣的原始数据表并将其转换成一个合适的次数分布表。简要讨论从这个次数表中你了解到什么原始数据表中并不明显的现象或趋势。
- 22. 濒危物种 世界保护监测中心位于大不列颠岛，其网站上提供了已经灭绝的、濒临灭绝的和受到威胁的动物种类数据。浏览这些数据，并且用次数分布表概括你感兴趣的发现。
- 23. 肚脐数据 肚脐比被定义为一个人的身高与从脚下到肚脐之间高度的比值。一个古老的理论说，就平均而言，人类的肚脐比符合黄金分割律 $(1 + \sqrt{5})/2$ 。测量你所在班级的每

一个人的肚脐比，肚脐比在黄金分割律5%内的人数所占的百分比是多少？肚脐比在黄金分割律10%内的人数所占的百分比是多少？这个古老的理论看起来可信吗？

- 24. 你自己的次数分布表（不分组） 搜集有关你自己的一些不需要进行分组的数据编制次数分布表（例如，你可以要求朋友们品尝一些不同品牌的饼干进行口味测试所搜集到的数据）。通过列出一张所有原始数据的表格，陈述你所收集的数据。然后将这些数据编制成一个次数分布表，包括一列相对次数和一列恰当的累加次数（除了称名变量水平的测量之外，对其他所有测量水平的数据累加次数都是适用的）。
- 25. 你自己的次数分布表（分组） 搜集有关你自己的一些需要进行分组的数据编制次数分布表（例如，你朋友们的体重或者在最近一次测验中的分数）。陈述你怎样搜集这些数据的，并且编制一个所有原始数据的表格，然后将这些数据编制成一个次数分布表，包括一列相对次数和一列累加次数。

新闻中的统计学

- 26. 次数分布表 寻找近期的一则新闻报道文章，内容包含某些类型的次数分布表。简要描述这个表格以及它对新闻报道起到怎样的作用。你认为这张表格对于这篇文章而言是最佳的吗？如果是的话，为什么？如果不是的话，你能编制一个不同的次数分布表吗？
- 27. 相对次数 寻找近期的一则新闻报道文章，内容至少包含一些相对次数的形式。简要描述这些数据，并且讨论为什么这个表格中的相对次数是有用的。
- 28. 累加次数 寻找近期的一则新闻报道文章，内容至少包含一些累加次数的形式。简要描述这些数据，并且讨论为什么在这个表格中累加次数是有作用的。
- 29. 温度数据 在报纸上寻找天气预报部分，列出美国许多城市的预报最高气温（制作一个关于这些数据的象形图并请你的指导老师批阅）。选择一个恰当的分组，编制预报最高温度的次数分布表，包括一列相对次数和一列累加次数，简要说明为什么以及怎样选择的分组。



3.2 绘制数据分布图

次数分布表给我们展示了一个变量在给定的分类内是如何分布的，因此我们说次数分布表概括了数据的分布。由于表格特别有用，我们通常会通过制图来更深刻地了解数据分布。在这一节中，我们探讨一些最常用的数据图示方法。

变量的分布 (distribution) 是指所有可能的取值散落的形态。我们可以通过制表或绘图来概括一组数据的分布。

3.2.1 条形图、点柱图和帕累托图

条形图 (bar graph) 是描述数据分布最简单的图形，简称条图。条形图常用于定性数据。每一个条表示一种类别的次数或相对次数：次数越大，条形越高。条可以绘制在纵轴上，也可以在横轴上。

让我们以表3-1的数据为例绘制论文测验成绩的条图。需要画出5个条形，每一个条形对应一种成绩类别 (成绩A、B、C、D、F)，每一个条形的高度对应某一类别的次数。图3-1呈现了绘图的结果。请注意以下条形图的图形结构特点：

表 3-1 论文测验成绩 (重复)

成绩	次数
A	4
B	7
C	9
D	3
F	2
总计	25

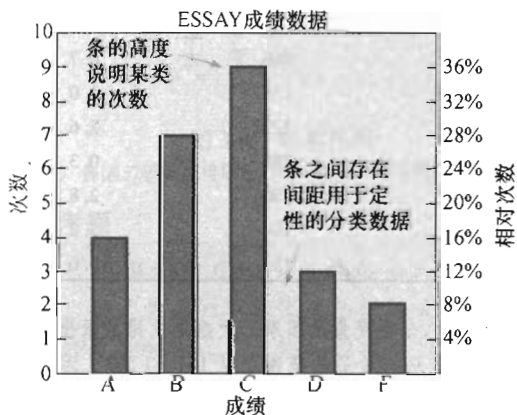


图 3-1 论文测验成绩条形图

- 因为最大的次数是9 (成绩C的人数)，我们选择让纵轴的刻度从0~10。这保证了最高的条形也不会超出图形顶部。
- 条形不要太高也不要太矮。在本例中，图形总体高度为5厘米看起来是很适合的，因为这样一来，每厘米条形高度代表次数为2是非常方便的。
- 条形的高度与次数是成比例的。例如，每1厘米高度代表次数为2，代表次数4的条形高度就是2厘米。
- 由于是定性数据，条形的宽度没有什么特定含义，也没有理由让这些条形相互连接。因此我们绘制条形时保持均衡一致的条宽和间距。

图形标记特别重要，没有正确的标记，一个图形将是无意义的。下面关于图形标记的总结适用于几乎所有图形，注意这些规则是如何在图3-1中应用的。

重要的图形标记

- 图题：图形应该有一个说明呈现什么内容的图题，如果可以的话应列出数据来源。
- 纵坐标的刻度和名称：沿纵轴画出对应精确数值的刻度线，纵轴旁边的数字应该清楚地标明刻度，数字应该与刻度线对齐。
- 横坐标的刻度和名称：沿着横坐标应该清楚地标出各个类别标志出来 (刻度线对于定量数据是必要的，但是对于定性数据就不是必要的)，并且在横坐标上要标注分类名称。
- 图例：如果是多元数据在一个图中呈现，用图例说明每一种不同的数据。

一个点柱图 (dotplot) 是条形图的变式，在点柱图中以点来代替条形表达次数。每一个点代表一个数据，例如，堆积起来的4个点意思是指次数是4。图3-2呈现了论文测验成绩的点柱图。当使用原始数据作图时，点柱图是非常方便的，因为你可以通过对每一个原始数据打点来进行计数，然后在正式的报告中再转换成条形图。

图3-3a呈现的是美国5个最大城市的人口数，在本例中，5个城市是分类变量，次数是人口数，城市按字母顺序排列。图3-3b呈现同样的数据，但条形是按降序排列的。一个条形图中的条形是按照次数顺序排列的，这种条图就

叫做帕累托图。注意只有当数据是在称名变量水平测量的定性数据时，这种条形顺序的重新排列才是有意义的，比如在本例中分类变量是城市。例如，图 3-1 的顺序水平的成绩数据制成帕累托图就是无意义的，因为将条形按照次数多少的顺序排列就形成了 C、B、A、D、F 这样的排列顺序。

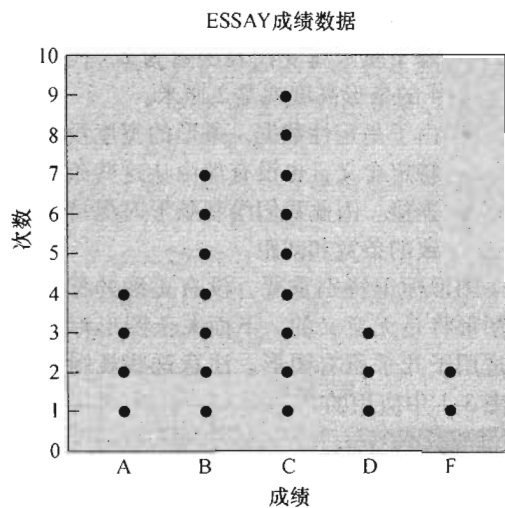


图 3-2 对应表 3-1 数据的论文测验成绩点柱图

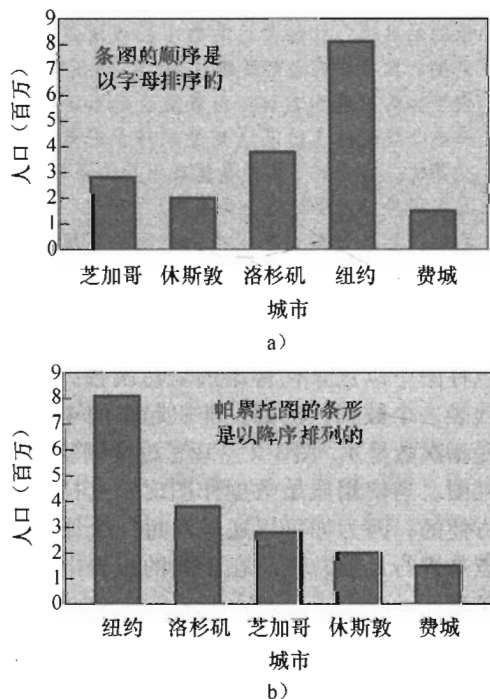


图 3-3 美国最大的 5 个城市的人口
a) 条形图 b) 帕累托图

资料来源：U. S. Census Bureau.

思考题

将图 3-3 中的人口数据制成一个点柱图是否可行？将有关 SAT 成绩的数据制作成帕累托图是否有意义？请说明。

定义

- ◆ **条形图 (bar graph)** 由一些表示特定类别次数 (或者相对次数) 的长条构成，条形的长度与次数成正比。
- ◆ **点柱图 (dotplot)** 与条形图含义相同，只是以点来代表每个个体的数据。
- ◆ **帕累托图 (pareto chart)** 是按次数多少的顺序排列条形的条形图，只有数据是在称名变量水平测量的定性数据时才有意义。

相关知识

帕累托图是意大利经济学家维尔弗雷多·帕累托 (Vilfredo Pareto, 1848—1923) 发明的，帕累托以提出一种收入分布分析方法而闻名，但他最重要的贡献则可能是开发了将数学和统计学应用于经济分析中的一种新方法。

例 3-5 二氧化碳排放量

二氧化碳被大量排放进大气层主要是由于矿物燃料 (石油、煤炭、天然气等) 的燃烧造成的。表 3-11 列出了 8 个最主要的每年二氧化碳排放量的大国。分别绘制二氧化碳总排放量和人均排放量的帕累托图。讨论两个图为什么看起来不一样？

表 3-11 世界八大主要的二氧化碳排放国

国家	二氧化碳总排放量 (百万吨碳)	人均二氧化碳 排放量 (吨碳)
美国	1 582	5.4
中国	966	0.7
俄罗斯	438	3.0
日本	329	2.6
印度	280	0.3
德国	230	2.8
加拿大	164	5.2
英国	154	2.6

解答 分类变量是国家，次数是碳排放数据值，由于二氧化碳总排放量的范围是从 154~1 582，因此纵坐标的刻度从 0~1 600 是个很好的选择。要绘制帕累托图，我们按照降序排列条形，每一个条形都对应碳排放数据

值,我们将分类名称(国家)标志于对应的条形下面。图3-4a呈现了二氧化碳总排放量的结果,图3-4b则显示了人均二氧化碳排放量的帕累托图,纵坐标的刻度从0~6包含全部数据值。注意两个帕累托图的国家排序是不同的,这告诉我们二氧化碳总排放量最多的国家不一定是人均二氧化碳排放量最多的国家。例如,中国是第二大二氧化碳总排放量国家,但是人均二氧化碳排放量却在美国和其他发达国家以下。

方法注解

表3-11所给出的排放量是基于排放的二氧化碳中的碳的重量,而不包括氧的重量。

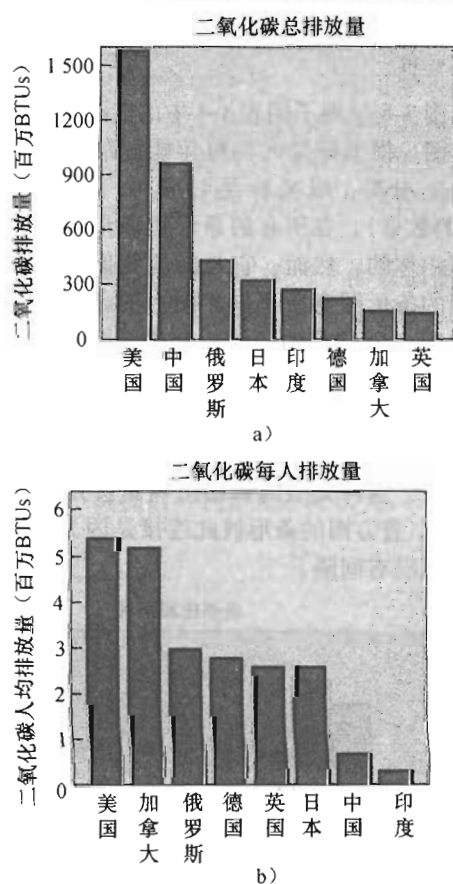


图 3-4 帕累托图

a) 各国二氧化碳总排放量 b) 各国人均二氧化碳排放量

思考题

中国和印度的人口合起来是美国的8倍,但是美国的二氧化碳总排放量却比中国和印度加起来还多。你认为这是为什么?如果中国和印度有着与美国一样的人均二氧化碳排放量,这个世界将会变成什么样呢?

3.2.2 饼图

饼图(pie charts)通常用于相对次数分布,一个圆饼表示相对次数总和100%,每一片扇形则表示不同类别所占的相对次数,饼图是几乎专用于定性数据的。

一个饼图是指一个圆分割成不同的扇形,每一个扇形代表不同类别的相对次数。扇形的大小与相对次数成正比,整个圆饼代表全部相对次数之和100%。

举个简单的例子,分析罗切斯特市的注册选民:25%是民主党人,25%是共和党人,另外50%是独立的。我们用一个饼图来呈现这种党派从属关系。因为民主党人和共和党人各自占选民的25%,那民主党人和共和党人分别占有整个圆饼的25%,或者说1/4扇形,独立人士占有所有选民的一半,所以占有整个圆饼的一半。图3-5呈现了这个结果。一定要注意标志清楚的重要性。

罗切斯特市注册选民

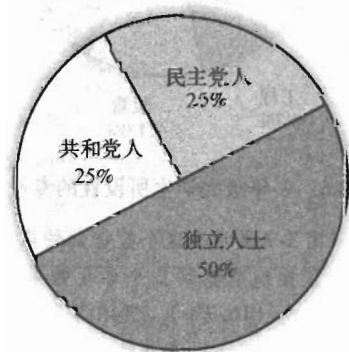


图 3-5 罗切斯特市注册选民的党派从属关系

相关知识

2006年美国全国性选举中,37%的符合年龄要求的投票者说自己认同民主党,31%的认同共和党,32%的认同其他党派或者保持独立(《美国统计年鉴》)。



如图 3-5 那样的饼图是很容易绘制的，因为它的扇形对应的是很简单的分数。更多的复杂饼图，你必须仔细测量扇形角度或者使用软件进行测量。

例 3-6 学生的专业

图 3-6 是一个关于大学一年级新生所设置专业领域的饼图，以同样的数据绘制一个帕累托图。三个最主要的专业领域是什么？评论回答这一问题时应用饼图还是帕累托图相对更为容易回答一些？

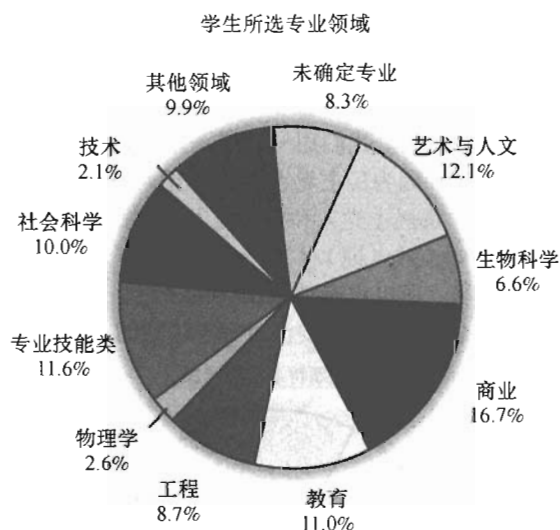


图 3-6 一年级大学生所设置的专业领域

解答 图 3-7 展示了这个数据的帕累托图。这个图能更清楚直接地观察到三个学生数最多的专业领域，分别是商业 (16.7%)、艺术与人文 (12.1%)、专业技能类 (11.6%) (专业技能类包括那些需要有专业执照的领域，比如建筑学、护理、药物学)，相反，饼图则需要我们研究一会儿，才能列出三个学生人数最多的专业。

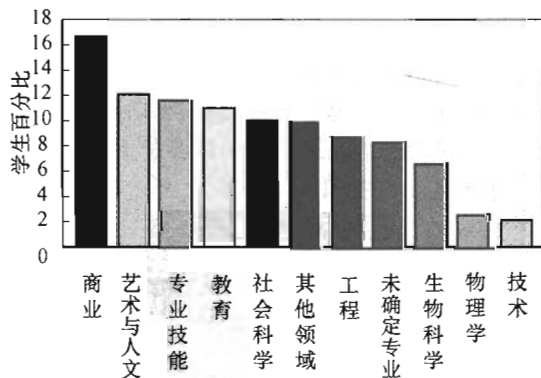


图 3-7 图 3-6 数据所绘制的帕累托图

思考题

例 3-6 讨论了哪些专业领域人数哪最多的数据，帕累托图比饼图更有优势。那你认为饼图与帕累托图相比有什么优势？表现在什么地方？

3.2.3 直方图和茎叶图

方法注解

不同的教科书对术语直方图和条图的定义是不一样的，没有统一公认的定义。在本书中，条图就是任何使用条形的图，直方图是条形图的一种，适用于比率测量水平的定量数据。

图 3-8 呈现了用表 3-4 中的能源数据绘制的条形图，横坐标是人均每年能源消耗量 (百万 BTUs) 分类，纵坐标是对应每一类别的次数 (州的数目)，在所有的条形里条形的长度与次数是相称的。然而，它与前面我们以定性数据绘制的条形图有所不同。图上的条形按照分类值形成一个自然的顺序。此外，图 3-8 中条形的宽度有层特殊的意思，在这种情况下，它告诉我们每一种分类下的能源消耗值范围。条形之间有自然的排序，并且条形的宽度具有一定的意义，这种类型的条形图叫做直方图 (histogram)。直方图的条形彼此连接是因为对应的分类之间没有间隔。

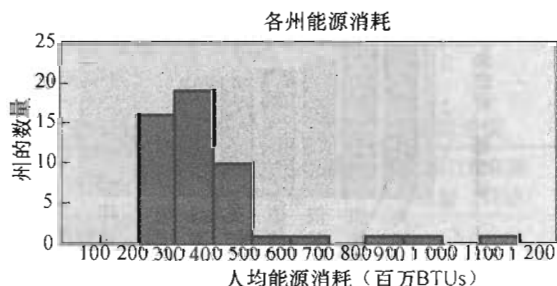


图 3-8 表 3-4 数据所绘制的直方图

能源消耗使用直方图，能够像表 3-4 一样清楚地反映次数变化的趋势，然而，不管是次数分布表还是直方图都不能提供表 3-3 中原始数据的细节。例如，直方图可以告诉我们有一个州的人均每年能源消耗在 800~899 百万 BTUs 之间，但是不能告诉我们是哪一个州以及它的精确能源消耗数值。图 3-9 这样的茎叶图 (stem-and-leaf plot)，

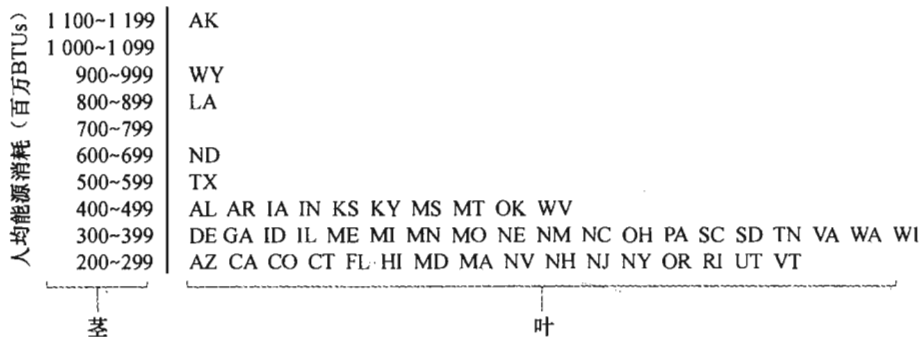


图 3-9 表 3-3 能源消耗数据所绘制的茎叶图

或者叫枝叶图 (stem-and-leaf plot, stemplot), 可以让我们更详细地观察数据的细节。将它旋转一下看起来像直方图, 只是用罗列的数据 (在本例中, 对应每一类的是州名) 代替条形。例如, 我们现在可以看到, 能源消耗在 800 ~ 899 百万 BTUs 之间的州是路易斯安那州。

另一种类型的茎叶图排列个体数据值。例如图 3-10 呈现了一个茎叶图, 是关于表 3-11 中人均二氧化碳排放量的数据。在这个例子里, 每一个数据被表示为一个茎 (包括第一个数字和小数点, 以及一个叶 (包括余下的数字), 我们可以直接从图中读出数据值。例如, 第一行中显示两个数据值分别是 0.3 和 0.7。

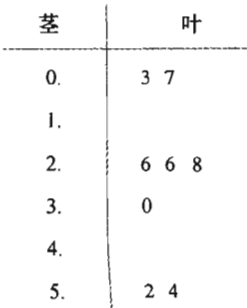


图 3-10 显示数值的茎叶图, 数据来自表 3-11 中的人均碳排放量

例 3-7 奥斯卡金像奖最佳女主角

表 3-12 显示了 1970 ~ 2007 年之间奥斯卡金像奖最佳女主角获奖时的年龄, 以这个呈现数据来制作一个直方图, 讨论其结果。

表 3-12 1970 ~ 2007 年奥斯卡金像奖最佳女主角获奖时的年龄

年龄	女演员人数
20 ~ 29	8
30 ~ 39	18
40 ~ 49	8
50 ~ 59	0
60 ~ 69	2
70 ~ 79	1
80 ~ 89	1

解答 最多的次数是 18 位女演员, 因此我们选择 20 作为直方图的高, 在任何一个直方图中, 一个条形的高度表示所对应类别的次数, 每一个条形的宽度是 10 岁, 条形之间彼此相连。图 3-11 是所制作的

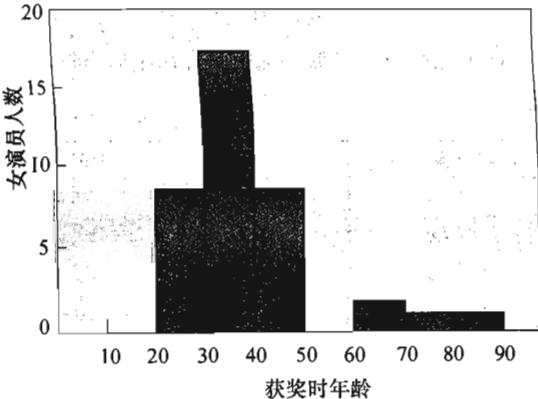


图 3-11 1970 ~ 2007 年奥斯卡金像奖最佳女主角获奖年龄直方图

定义

- ◆ 直方图 (histogram) 是呈现定量数据 (等距和测量的比率水平数据) 分布的一个条形图, 条形有一个自然的排序, 条形的宽度具有特定的含义。
- ◆ 茎叶图, 就像直方图旋转过来, 只是用排列的数字来代表条形。

直方图。我们看到大多数女演员都是在比较年轻时获得奥斯卡金像奖的，相比较而言，男演员获得奥斯卡金像奖时则年龄大一些（参考 3.1 节中的练习题 13），许多女演员相信这个事实反映了一种微妙的差别，那就是好莱坞很少制作这样的电影，在其中由老年女演员扮演具有鲜明性格特征的主要角色。 ■

■ 思考题

你需要增加什么信息才能制作一个关于奥斯卡金像奖最佳女主角获奖年龄的茎叶图？这个茎叶图看起来像什么？

3.2.4 线图

■ 方法注解

由一组分布次数生成的线图经常被叫做次数多边形图，因为它由许多形状各异的分段的直线组成。

像直方图一样，线图（line chart）也显示了定量数据的分布。它将一系列点用直线相连，纵坐标上的点代表次数，点刚好处于直方图条形的正上方中间。图 3-12 显示了用表 3-4 中的能量消耗数据制作的线图。为了比较，线图与前面所做的直方图（见图 3-8）合并在一个图中。

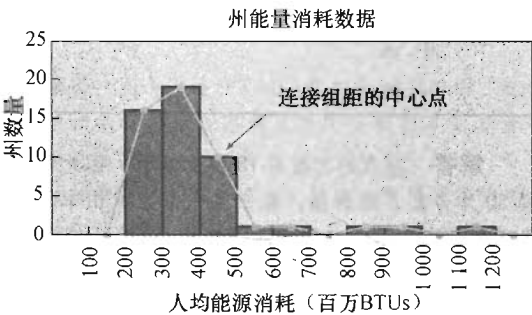


图 3-12 能源消耗数据的线图,且合并直方图进行比较

注意在解释线图时重要的一点：点的横轴位置代表的是组距的中心点。例如，表达 300 ~ 399 百万 BTUs 之间范围的点位于 349.5 百万 BTUs 处。这样，如果你还是不太明白的话，你可以这样认为，这个点的意思是指 19 个州人均每年能源消耗全部是 349.5 百万 BTUs，实际上 19 个州的人均每年能源消耗分布在 300 ~ 399 百万 BTUs 这个范围之间。

■ 定义

◆ 线图将一系列点用直线相连呈现了定量数据的分布。每一个点的横坐标刚好处于所属范围中间，纵坐标位置对应所属类别的次数值。

例 3-8 年龄分布

表 3-13 呈现了按照年龄分类的美国人口分布。制作此数据的直方图和线图。假如你制作一个从现在开始 20 年后的同样的年龄分布图，你想它们会有什么不同？

解答 数据分组以 5 岁为一个组距（除了最后一组年龄 100 岁以上的），我们通过以条形高低代表每一类别的次数（人口数）制作一个直方图。图 3-13 呈现了制图的结果。附加在直方图上的是一条线图，每一个点都位于每一类别所对应条形的上方中间。例如，表示 25 ~ 29 岁的点横坐标对应于 27.5。与此一致，我们制图时将“100 岁以上”看做 100 ~ 104 岁，但是横轴上做标记时仍然按实际上的意思。许多次数分布表包括非封闭结尾这样的类别（比如 100 岁以上），因为存在少量的极端数据会使数据范围扩展到很大。

假如在 20 年后制作一个同样的图，我们会预期这两个图有两点主要的不同，首先，因为人口增长所有的次数都将会增加；其次，由于医疗技术的提高会使得人们更长寿，高年龄组的人口增长要比低年龄组的人口增长明显。 ■

表 3-13 美国人口的估计年龄分布

年龄范围（岁）	人口（百万人）	年龄范围（岁）	人口（百万人）	年龄范围（岁）	人口（百万人）
0 ~ 4	21.0	35 ~ 39	19.0	70 ~ 74	9.0
5 ~ 9	19.4	40 ~ 44	20.4	75 ~ 79	7.2
10 ~ 14	19.9	45 ~ 49	22.2	80 ~ 84	5.6
15 ~ 19	21.7	50 ~ 54	22.0	85 ~ 89	3.5
20 ~ 24	21.2	55 ~ 59	19.2	90 ~ 94	1.6
25 ~ 29	19.8	60 ~ 64	16.3	95 ~ 99	0.6
30 ~ 34	19.0	65 ~ 69	12.2	100 以上	0.1

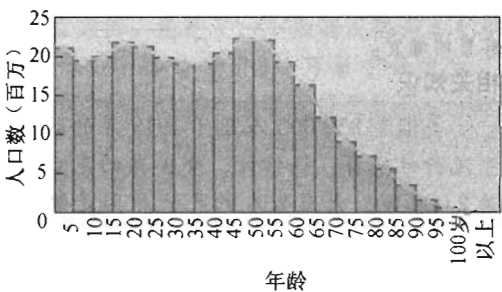


图 3-13 表 3-13 中年龄数据的直方图和线图

3.2.5 时间序列图

表 3-14 显示了美国凶杀率是如何随着时间而变化的，分类是年份，数据是凶杀率（每 100 000 人被凶杀致死的人数），我们可以用直方图或者线图来表示这些数据。图 3-14 呈现了一个线图，由于本例中横轴表示时间，我们称这种图叫做时间序列图（time-series diagram）。

表 3-14 1960 年以来美国每 100 000 人的凶杀率

年份	凶杀率（每 100 000 人）	年份	凶杀率（每 100 000 人）	年份	凶杀率（每 100 000 人）
1960	5.1	1976	8.8	1992	9.3
1961	4.8	1977	8.8	1993	9.5
1962	4.6	1978	9.0	1994	9.0
1963	4.6	1979	9.7	1995	8.2
1964	4.9	1980	10.2	1996	7.4
1965	5.1	1981	9.8	1997	6.8
1966	5.6	1982	9.1	1998	6.3
1967	6.2	1983	8.3	1999	5.7
1968	6.9	1984	7.9	2000	5.5
1969	7.3	1985	7.9	2001	5.6
1970	7.9	1986	8.6	2002	5.6
1971	8.6	1987	8.3	2003	5.7
1972	9.0	1988	8.4	2004	5.5
1973	9.4	1989	8.7	2005	5.8
1974	9.8	1990	9.4		
1975	9.6	1991	9.8		

资料来源：United States Bureau of Justice Statistics.

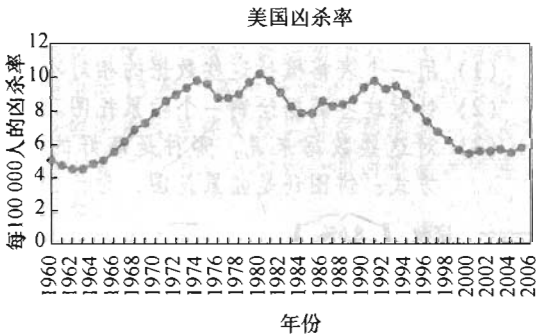


图 3-14 表 3-14 中凶杀率数据的时间序列图

思考题

查找表 3-14 和图 3-14 在 2005 年以后新扩展的数据，你能看出一些凶杀率的发展趋势吗？

例 3-9 降低死亡率

图 3-15 呈现了 1900 年以来美国死亡率（每 1 000 人死亡人数）的时间序列图（例如，1905 年死亡率为 15，意思是指 1905 年从年初到年底每 1 000 人就有 15 个人在这一年里死亡）。讨论和分析其中的一般趋势，特别考虑 1919 年的一个尖峰，如果有人告诉你这个尖峰归因于第一次世界大战战场的死亡，你相信吗？说明理由。

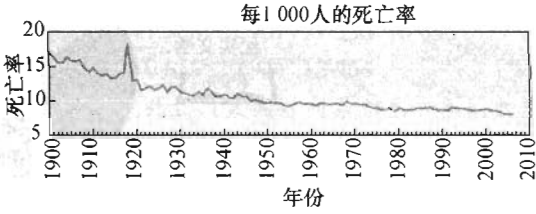


图 3-15 美国历史上每 1 000 人的死亡率

资料来源：National Center for Health Statistics.

解答 死亡率的一般趋势无疑是向下的，推测起来的原因是医疗水平的提高。例如，早在1900年诸如肺炎这样的细菌感染性疾病是主要杀手，但是随着抗生素的应用这类疾病是可治愈的。1919年的一个尖峰与第一次世界大战结束是同时发生的，如果这个尖峰能归因于第一次世界大战战场的原因，那我们也可以预期第二次世界大战结束后的几年内也会有一个同样的尖峰。这样的特征并不存在，提示我们不能将1919年的尖峰归因于第一次世界大

战。实际上，死亡率尖峰的原因在于西班牙型流行性感冒的爆发。

相关知识

美国有850 000人死于1919年的西班牙型流行性感冒，据估计，全世界的死亡人数为2 000万，大多数人的死亡原因是继发性细菌感染，而这在今天使用抗生素很容易治愈。

3.2 练习题

统计素养和批判性思维

1. 分布 通过数据的分布我们可以知道些什么？为什么说直方图比样本数据罗列更有助于理解数据的分布。
 2. 较少的一组数据 如果一组数据仅有5个数值，为什么我们毫不费力地绘制一个直方图？
 3. 时间系列图 哪一种类型数据需要绘制一幅时间系列图，时间系列图揭示了数据的什么内容？
 4. 直方图和茎叶图 假定用一组数据绘制一幅直方图和一幅茎叶图，如果仅使用直方图，能够再现原始数据值吗？如果只使用茎叶图，能够再现原始数据值吗？茎叶图比起直方图来有什么有利的方面吗？
- 判断正误：**判断练习题5~8中的陈述是有意义的（表述正确）还是无意义的（表述错误），并做出解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。
5. 直方图 直方图经常被用来描述按照被试眼睛颜色进行分类的样本的身高。
 6. 时间系列图 一幅时间系列图可以用来呈现2007年不同计算机制造厂商的利润。

7. 帕累托图 一个质量控制检验员想强调质量缺陷产品的最重要原因，所以他用帕累托图来呈现不同质量缺陷导致原因的次数。
8. 饼图 饼图的一个非常重要的不足之处就是它缺乏一个合适的刻度。

概念与应用

- 最适宜的直观呈现图形：**练习题9~12分别描述了一种数据，但是并没有给出实际的数值。针对每一种数据，说明哪种类型的图形对于呈现这种数据是最适宜的。对你的选择做出解释。
9. SAT成绩 统计专业学生的SAT成绩。
 10. 眼睛颜色 女演员的眼睛颜色。
 11. 死亡原因 去年一年去世人们的死亡原因。
 12. 超级碗比赛 每一年的超级碗比赛进行到一半时间时的电视商业价值，第一届超级碗比赛开始于1967年。
 13. 人们在阅读什么 图3-16的饼图呈现了关于人们阅读内容的调查结果。
 - (1) 用一个表格概括这些数据的相对次数。
 - (2) 针对这些数据绘制一个帕累托图。
 - (3) 对这些数据来说，哪种是最好的表达方式：饼图还是帕累托图，为什么？

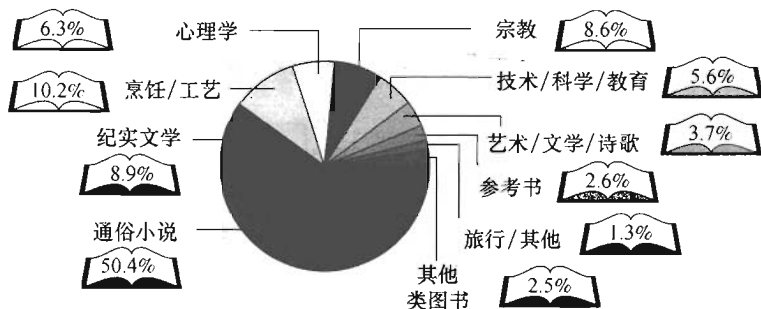


图 3-16 人们的阅读书籍

资料来源：Book Industry Study Group.

14. 直方图 图 3-17 是用 STATDISK 软件生成的直方图, 描述了一个抽烟人群样本可替宁水平 (微克/每毫升)。可替宁是尼古丁的代谢产物, 这意味着当人体吸收尼古丁时就会产生可替宁。数据来源与第 3 次联邦健康和营养调查研究。

- (1) 在直方图里一共呈现了多少被试?
- (2) 多少名被试的可替宁水平低于 400?
- (3) 多少名被试的可替宁水平高于 150?
- (4) 在直方图中呈现的被试的可替宁的最高可能水平是多少?

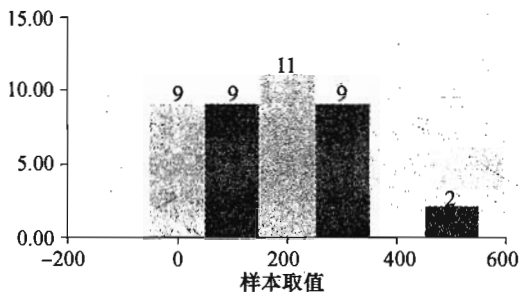


图 3-17 STATDISK 软件生成的直方图

15. 可口可乐的重量 3.1 节练习题中的 11 题要求针对 36 听普通可口可乐的重量 (磅) 编制一个次数分布表。使用这个次数分布表绘制对应的直方图。
16. 健怡可乐的重量 3.1 节练习题中的 12 题要求针对 36 听健怡可乐的重量 (磅) 编制一个次数分布表。使用这个次数分布表绘制对应的直方图。
17. 奥斯卡电影金像奖最佳男演员 3.1 节练习题中的 13 题, 要求针对近年来奥斯卡电影金像奖最佳男演员获奖时年龄编制一个次数分布表, 使用这个次数分布表绘制对应的直方图。
18. 身体温度 3.1 节练习题中的 14 题, 要求针对所列出的随机选择的被试身体温度编制一个次数分布表, 使用这个次数分布表绘制对应的直方图。
19. 求职 进行了一项研究是关于人们是怎样进行求职的。下表列出了 400 个随机选择被试的数据。数据来源于联邦职业策略研究中心的。根据给出的数据绘制一个帕累托图。如果有人想得到一份工作, 哪一种方法看起来是最有效的?

调查对象的工作源	次数
招聘广告	56
猎头公司	44
网络和广播电视	280
群发邮件	20

20. 工作资源 参考练习题 19 所给出的数据, 绘制一个饼图。将饼图与帕累托图进行比较。在显示工作资源的相对重要性时, 你能否确定哪一种图更有效?
21. 列车出轨的原因 关于列车出轨事故的一份分析显示 23 次列车出轨事故是由于铁轨损坏, 9 次是因为仪器出现问题, 12 次是人为事故, 6 次是其他原因 (数据来源于联邦铁路管理局)。根据给出的数据绘制一个饼图。
22. 列车出轨原因的分析 参考练习题 21 所给出的数据, 绘制一个帕累托图, 将帕累托图与饼图进行比较。在呈现铁路出轨事故原因的相对重要性上, 你能否确定哪一种图更有效?
23. 散点图 参考 3.1 节练习题 19 中的打字机式键盘的数据, 绘制一个散点图。
24. 散点图 参考 3.1 节练习题 19 中的德沃夏克键盘的数据, 绘制一个散点图。将 24 题中所绘制的散点图与 23 题所绘制的散点图进行比较。基于比较的结果, 哪一种键盘设计得更好一些? 并做出解释。
25. 关于道琼斯平均工业指数的时间系列图 下列数据是自 1980 年开始道琼斯平均工业指数 (DJIA) 每年的最高点。数据是按行呈现的, 绘制一个时间系列图, 并且确定这里面是否有一个发展趋势, 一个投资者怎样从这种趋势中获利。
 1 000 1 024 1 071 1 287 1 287 1 553
 1 956 2 722 2 184 2 791 3 000 3 169
 3 413 3 794 3 978 5 216 6 561 8 259
 9 374 11 568 11 401 11 350 10 635
 10 454 10 855 10 941 12 464
26. 交通事故死亡人数的时间系列图 下列数据是自 1980 年开始美国每年交通事故死亡的人数。数据是按行呈现的, 绘制一个时间系列图, 并且确定这里面是否有一个发展趋势。如果存在, 请提供一个可能的解释。
 51 091 49 301 43 945 42 589

44 257 43 825 46 087 46 390
 47 087 45 582 44 599 41 508
 39 250 40 150 40 716 41 817
 42 065 42 013 41 501 41 717
 41 945 42 196 43 005 42 884
 42 836 43 443

27. 茎叶图 根据如下的测验分数绘制一个茎叶图, 67, 72, 85, 75, 89, 88, 90, 99, 100。茎叶图是怎样呈现这些数据的分布?

28. 茎叶图 下面列出了有趣的儿童电影的时间长度(分钟)。绘制一个茎叶图。茎叶图能否呈现这些数据的分布? 如果能够呈现的话, 又是如何做到的?

83 88 120 64 69 71 76 74 75
 76 75 75 79 80 78 78 83 77
 71 83 80 73 72 82 74 84 90
 89 81 81 90 79 92 82 89 82
 74 86 76 81 75 75 77 70 75
 64 73 74 71 94

网络和远程项目

请在网站 www.aw.com/bbt 上打开第3章的有用链接“links for internet projects”。

29. 二氧化碳排放量 在美国能源信息管理局(EIA)发布的《国际能源年鉴》网页上寻找并下载国际的二氧化碳排放量数据, 生成一个包括最近数据的统计图。讨论你的统计图中所呈现的特点。

30. 能源表 在美国能源信息管理局的网站上浏览能源表的全部数据。选择你所感兴趣的一个表并且根据它的数据绘制一个统计图。你可以任意选择一种在本章讨论的统计图类型。说明你是怎样绘制你的统计图的, 并且简要讨论我们从中能学到什么。

31. 统计摘要 寻找《美国统计摘要》的网页,

浏览“次数分布表”部分。选择一个你感兴趣的表并且根据它的数据绘制一个统计图。你可以任意选择一种在本章讨论的统计图类型。说明你是怎样绘制该统计图的, 并且简要讨论我们从中能学到什么。

32. 肚脐数据 根据3.1节练习题23中的肚脐数据, 绘制一个恰当的统计图。并且讨论这个数据分布中的某些特定比率。

新闻中的统计学

33. 条形图 寻找近期的一篇新闻报道, 包括关于定性数据分类的条形图。

(1) 简要解释这个条形图呈现了什么内容, 并且讨论它是否有助于我们了解新闻报道的观点。图中标志是否清晰?

(2) 简要讨论这个条形图能否变成一个点图。

(3) 这个条形图是否已经是一个帕累托图? 如果是的话, 说明为什么我们认为这样的图形是一个帕累托图。如果不是的话, 图中的条形如何重新排列就成了一个帕累托图, 并加以解释。

34. 饼图 寻找一则近期的新闻报道, 其中包含一个饼图。简要讨论这个饼图的效果。例如这个数据用条形图呈现是否会比饼图更好。这个饼图能在其他一些方面得到改进吗?

35. 直方图 寻找一则近期的新闻报道, 其中包含一个直方图。简要说明这个直方图呈现了什么内容, 它是否能帮助我们更容易地了解文章的观点。图形的标志是否清晰? 这个直方图是一个时间系列图吗? 说明理由。

36. 线图 寻找一则近期的新闻报道, 其中包含一个线图。简要说明这个线图呈现了什么内容, 它是否能帮助我们更容易地了解文章的观点。图形的标志是否清晰? 这个直方图是一个时间系列图吗? 说明理由。



3.3 媒体中的图表

到目前为止, 我们所学习的基本图形, 仅仅是众多的直观描述数据方法的开始。在这一节, 我们将探讨一些复杂的经常用于媒体的图表类型。

3.3.1 复式条图和复式线图

复式条图(multiple bar graph)是单一条图的简单扩展, 它包括两个及两个以上相对比的条形, 可用来比较两组及两组以上数据。所有的数据组都有相同的分类标志以便于它能够呈现在同一张图上。图3-18是一个复式条图, 呈现了儿童从他们的父母那得到钱的数目。3组

数据表示了3种不同的每周从父母那得到钱的来源,零花钱、奖励、家务报酬(注意:中位数是在4.1节所要讨论的内容,是一组数据中间的数值,例如,9~10岁儿童的零花钱的位数为3美元,是指一半儿童得到的零花钱少于3美元,另一半儿童得到的零花钱大于3美元)。

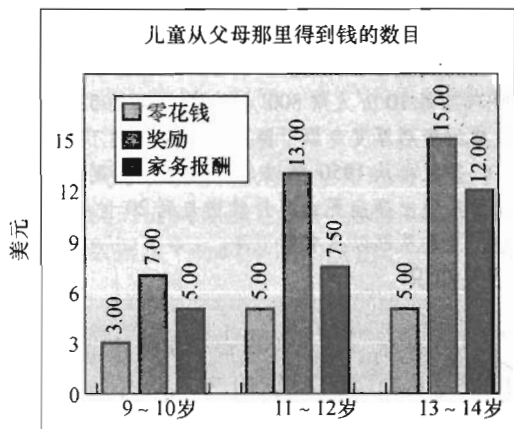


图 3-18 复式条图

资料来源: Consumer Reports and Ziutions

复式线图 (multiple line chart) 与复式条图有着相同意义,但它是以线的形式呈现数据而不是用条。图 3-19 呈现了 12 周时间内的股票、基金和黄金的价格变化。这种类型的图也是一种时间系列图,因为它是以时间顺序呈现数据的。

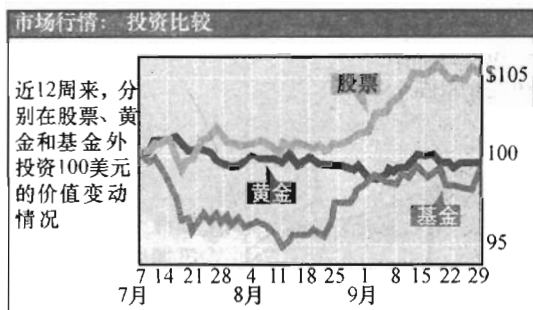


图 3-19 复式线图

资料来源: New York Times.

例 3-10 阅读投资图

思考一下图 3-19,假定在 7 月 7 日,你投资了 100 美元在一个股票上,100 美元在一个基金,100 美元投资黄金。如果你在 9 月 15 日将全部投资都卖出,你最后的结果是赔钱还是赚钱?

解答 图中显示 100 美元的股票在 9 月 15 日价值是 105 美元,100 美元的基金投资将会下降至大约 99 美元,黄金投资将持平价值还是 100 美元。在 9 月

15 日,你的全部投资现在价值是

$$105 + 99 + 100 = 304 \text{ (美元)}$$

你的全部 300 美元的投资最后收益是 4 美元。 ■

相关知识

黄金被认为是非常稳健的投资,并且是投资组合中重要的一部分。然而,黄金的价格在 1996~2006 年非常萧条,2006 年年底,黄金的价格仅为 630 美元/盎司,远低于 1980 年的 2 000 美元/盎司 (经过通货膨胀率调整后的价格)。

例 3-11 图形转换

图 3-20 是一个复式条图,数据是美国家庭使用计算机和应用互联网的数字,采用复式线图重新绘制图形,简要地讨论图形中所呈现出来的趋势。

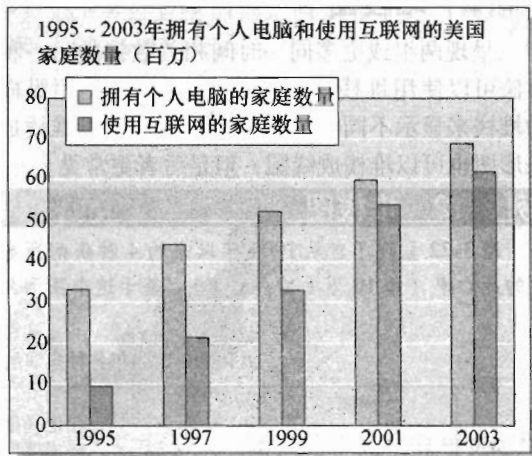


图 3-20 家用电脑发展趋势的复式条图

资料来源: Statistical Abstract of the United States.

解答 将复式条图转换成复式线图,我们必须将两组数据的条形转换成线形,我们在每一类所对应的条形上方中间绘制一个点,然后将这些点用同 1 颜色的线条连接起来,如图 3-21 中显示。

美国 1995~2003 年拥有个人电脑和使用互联网的美国家庭数量

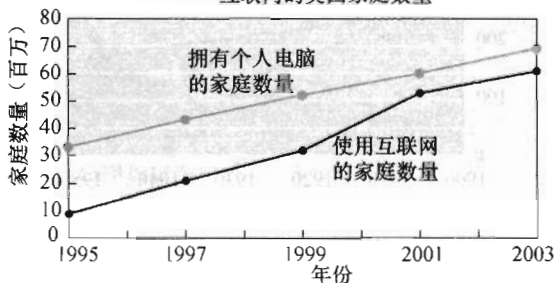


图 3-21 呈现图 3-20 数据的复式线图

最明显的趋势是两组数据都随着时间增长。我们通过比较每一年的条图来看一看第二个趋势，1995年，使用互联网的家庭（大约10 000 000）不足使用计算机家庭的1/3（大约3 300万）。到2003年，使用互联网的家庭（大约6 200万）大约是拥有计算机家庭的90%（大约7 000万）。这告诉我们越来越多的计算机使用者开始应用互联网。如果这种趋势进一步延伸，那么似乎所有应用计算机的家庭都将会使用互联网，而且这个人数也越来越接近于美国家庭总数。

相关知识

1993年，全世界仅有300万人能够连接到互联网。2007年，美国使用网络的人数超过2.1亿，而全世界的网络用户则超出10亿人。

3.3.2 堆栈图

呈现两组或更多同一时间相关数据的另一种方法可以使用堆栈图（stack plot）。它是用纵向的堆栈来显示不同的数据组。尽管既可以堆栈成条形图也可以堆栈成线图，但是后者更常见。

例3-12 堆栈图

图3-22呈现了自从1900年以来的4种疾病所导致的死亡率（每10万人的死亡率）。基于这个图形，

计算1980年因为心脏疾病所导致的死亡率是多少？讨论图形中所呈现出的一般趋势。

解答 每一种疾病都有自己的颜色标注区域或楔形，注意图例的重要性。某一特定时间的楔形的厚度告诉我们那时的取值。例如1980年，心脏病的楔形纵坐标是180~620，所以它的厚度是440，这告诉我们1980年心脏疾病死亡率大约是440人每10万人。这个图形呈现了几种重要的趋势。首先，最上面楔形向下的斜坡显示4种疾病的死亡率是持续下降，从1900年的每10万人中800人到2003年的525人。肺结核病的楔形厚度急剧下降说明这种疾病不再是主要杀手，并且自从1950年以后接近消除，同时，癌症的楔形呈现出癌症死亡率持续增长到20世纪90年代中期，但是从那以后有所下降。

相关知识

20世纪，癌症的死亡率在持续上升，在1995年达到205人每10万人这样的一个高峰。从那以后，癌症死亡率已经下降了10%，2006年实际的癌症死亡人数第一次开始下降（不是每10万人的死亡率）。科学家将这种下降主要归因于吸烟人数的下降（因为据估计，吸烟导致了所有癌症中的1/3）以及对癌症的早期觉察和治疗。

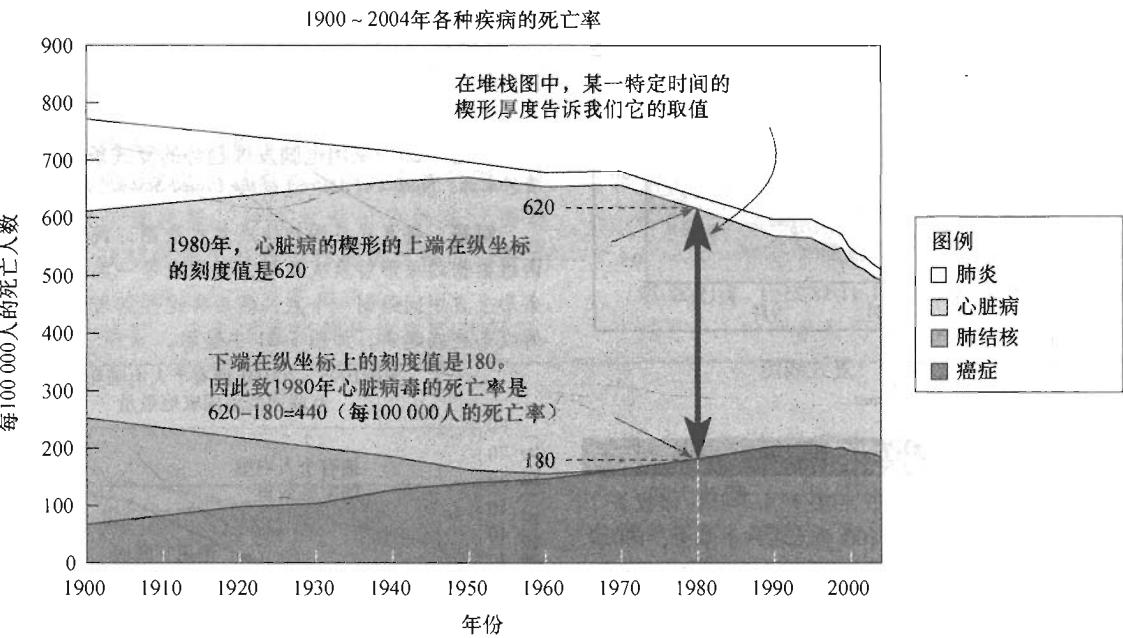


图3-22 应用堆栈分界线的堆栈图

资料来源：National Center for Health Statistics, American Cancer Society.

3.3.3 地理数据图

表 3-3 的能源消耗数据是一个关于地理数据 (geographical data) 的例子, 原始数据对应于不同的地理区域。前面我们用这些数据制作一个次数分布表 (见表 3-4), 一个直方图 (见图 3-8), 以及一个茎叶图 (见图 3-9)。但是无论如何, 这些图形都不能给我们一个有关数据的地理方面的概念。例如, 我们不能从图中得知, 能源消耗相同的州那一个在最北边。图 3-23 呈现了一种展示地理分布的方法。不同类别用图例的方式以颜色来标志的, 每个州都按照他的数据水平标以不同的颜色。

思考题

你能否从图 3-8 的直方图上, 了解到哪些从图 3-23 所呈现的地理图所不容易展示的内容? 反之, 你能否从地理图中看到哪些直

方图中所不能展示的内容? 你能否从图 3-23 中发现一些令人惊奇的地理分布?

图 3-23 非常适用于展示能源消耗数据, 因为每一个州数据都是唯一的。因为穿越不同地理区域的某些数据的变化是连续不断的, 因此应用等高图会更方便。图 3-24 呈现了在某一特定时期全美国温度的等高图。每一条等高线都将具有相同温度的地方连接在一起。例如, 所有温度为华氏 50 °F 的地方连成一条等高线, 所有温度为华氏 60 °F 的地方连成一条等高线, 在两条等高线之间的区域温度就介于华氏 50 °F ~ 60 °F 之间。注意比较大的温度差异意味着比较小的等高空间。例如, 东北部紧密相连的两个等高线说明这个温度存在于一段很小的距离内, 为了让图形更容易阅读, 两条相邻的等高线之间用不同的灰度标注。

州能源使用情况

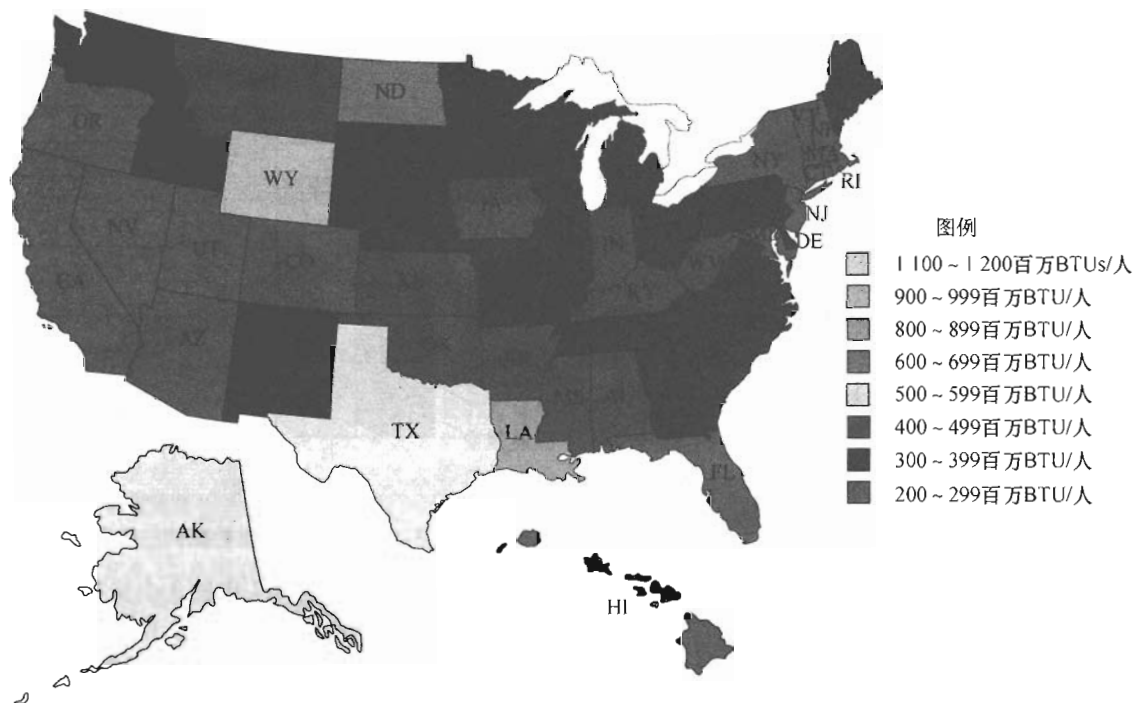


图 3-23 应用不同灰度标志的地理数据图

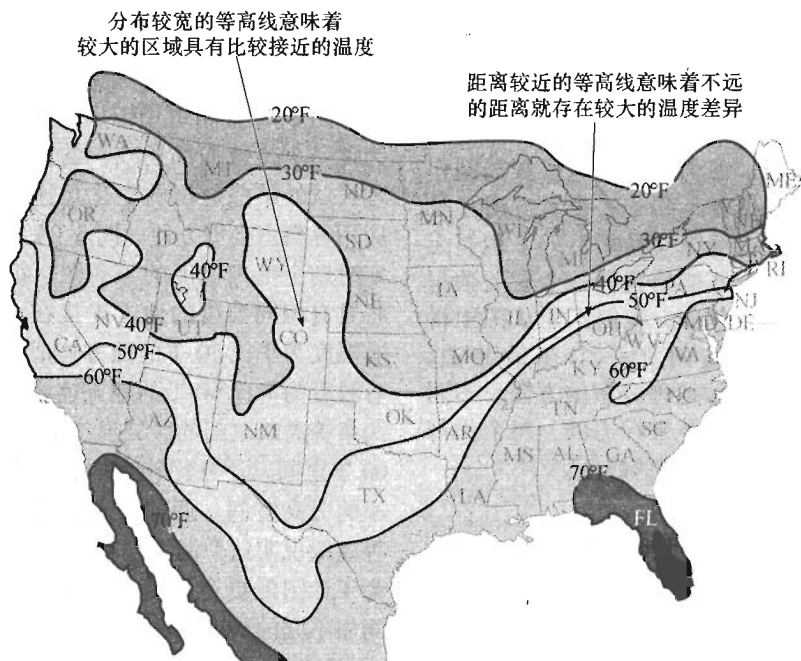


图 3-24 连续变化的地理数据，诸如温度，可以用等高图呈现

例 3-13 一个俯视等高图

等高图也可以用来呈现地理的俯视状况，图 3-25 呈现了一个科罗拉多州波尔得市的俯视等高图。讨论图中呈现的几个关键特征。

解答 图中的标志指明了关键的特征，东部的等高线之间空间比较广阔，那里的地势比较平坦，从空中俯视的地理高度是相对恒定的。再往西看，由于从平原逐渐变成山峰，等高线就变得越来越稠密，图中中间位置封闭的等高线是环绕山峰的地图中心。

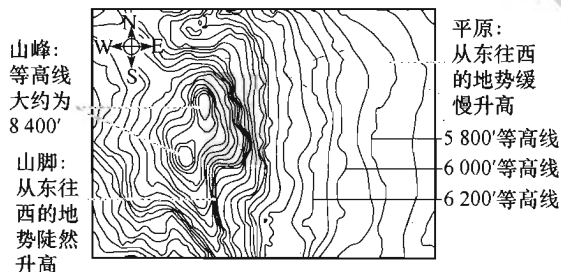


图 3-25 波尔得市的俯视等高图

3.3.4 三维图

今天，计算机软件将任何一种图形以三维图的样式呈现是非常容易的。例如，图 3-26 再次呈现了图 3-1 的条形图，对这个图装扮了一种三维的外观，这样会看起来更好看一些，但是这个三维图只是为纯粹的美观起见，并没

有提供任何多余的信息，仅是图 3-1 数据的再现。

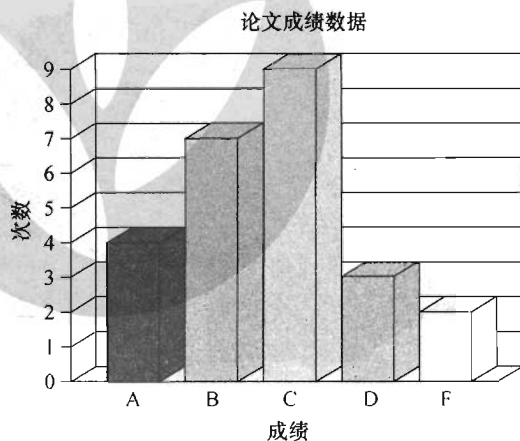


图 3-26 呈现了两维数据的三维图

与之相对比，图 3-27 中的 3 个坐标轴则分别表示了独特的信息，这是一个真正意义上的三维图形。研究者分析了鸟类（北美的食米鸟）的迁徙模式，他们对整晚飞越纽约 7 个城市的鸟类进行数字统计。如图 3-27 所示，7 个城市按着从东到西的顺序排列，方便研究者了解鸟类在晚上的什么时间段飞越纽约哪一个地区，注意 3 个坐标轴上分别是鸟类的数量、晚上的时间和东西部的位

声波映射法跟踪鸟类迁徙

1993年8月28日和29日夜晚，分布于纽约州的传感器记录了夜间飞行的食米鸟每一次的叫声，以此追踪它们在秋天的迁徙。通过计算机处理，下面数据呈现了食米鸟越过纽约东部各地时的时间和数量。

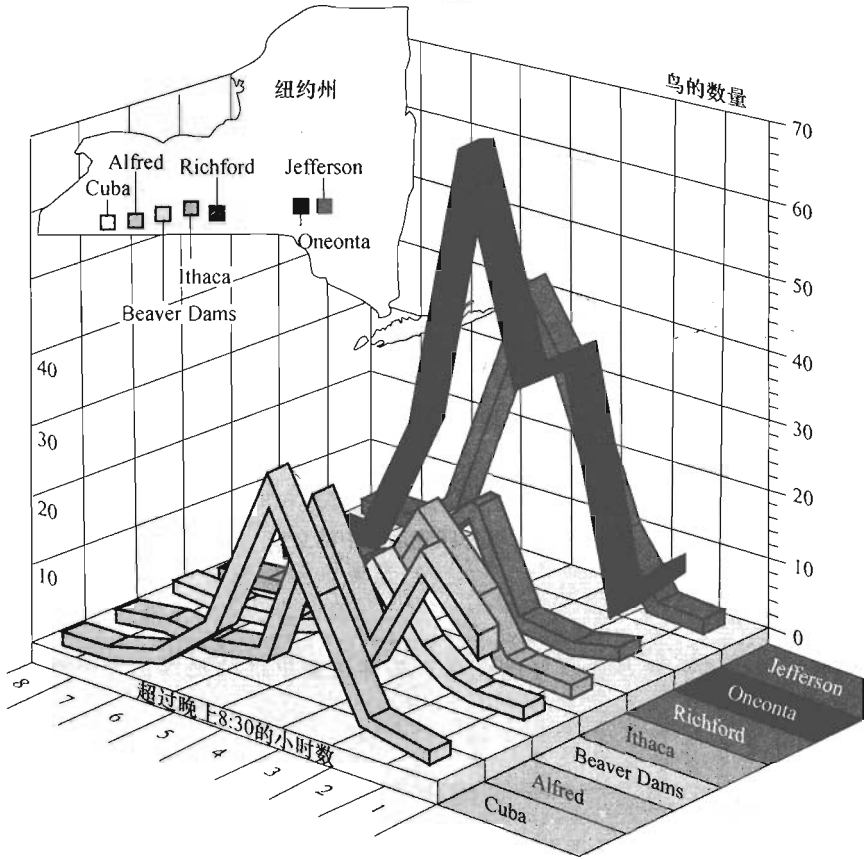


图 3-27 真正呈现了三维数据的三维图

资料来源：Bill Evans/Cornell Laboratory of Ornithology, New York Times.

例 3-14 鸟类的迁徙

基于图 3-27，大约什么时间有最大数量的鸟群穿越以 7 个城市为标记的东西线？大多数鸟类穿越纽约州的哪一个区域？大约晚上 12 点半时大概有多少鸟飞过奥尼安他地区？

解答 在所有城市所探测到的鸟群的数量峰值时间段是在夜里 12 点半到 2 点半。飞跃最西部的两个城市奥尼安他和杰弗逊的鸟群比起其他靠近西部的城市更多一些，这说明大多数鸟类飞过纽约州的最西部。为了回答有关奥尼安他的特殊问题，注意晚上 8 点半以后的 4 个小时是午夜 12 点半。在图形中，这个时间在奥尼安他的线图上刚好是两峰之间的低谷。再看一下鸟群数量这个坐标轴，我们可以看到在这个时间穿越奥尼安他的鸟类是 25~30 只。

3.3.5 联合图形

到目前为止，我们所学习的各种种类的图

形都是很常见的，并且容易制作。但是现在媒体经常会采用更复杂的图形来涵盖多种变量。例如，图 3-28 呈现了与妇女参加夏季奥运会有关的数据。在这个图形上合并了多条线图、多个饼图和复杂的数据资料，尽管这是一个图形，但是它比千言万语的信息量还要多。

例 3-15 奥林匹克运动会的妇女

描述图 3-28 中所呈现的三种趋势。

解答 线图显示了参加夏季奥运会的女性人数在稳步增长，尤其是在 1960 年以后，在 2004 年奥运会中女性参赛者更是接近 5 000 人。饼图显示了参加奥运会的女性比例也在逐步增加，在 2004 年奥运会中达到了 44%，底部的浅红色数字显示了女性参加的竞赛项目数也在急剧增长，在 2004 年奥运会中达到了 135 项。

夏季奥林匹克运动会中人数日益增长的女性选手

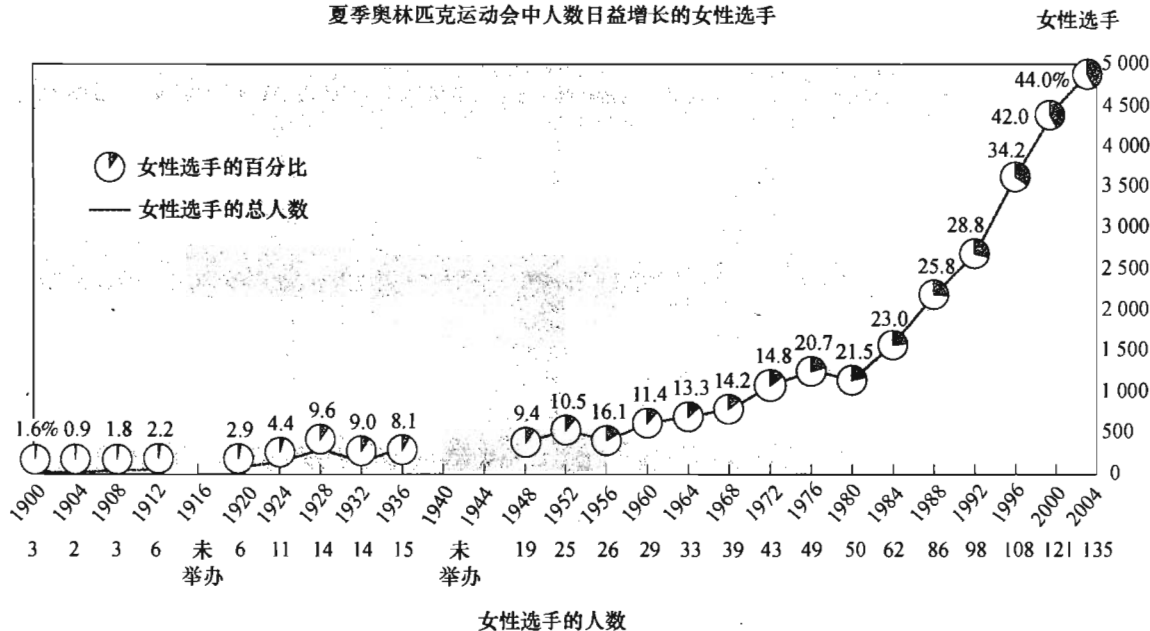


图 3-28 奥林匹克运动中的女性

资料来源: Adapted from the New York Times.

思考题

图 3-28 中呈现的哪一种趋势有可能会在下届奥运会中继续延续？哪一种不会？请说明理由。

3.3 练习题

统计素养和批判性思维

1. 三维统计图 一本杂志上采用三维图中的石油桶大小来描述不同国家石油的生产量。非常明显，石油桶是三维物体，但是不同国家的石油生产量是一个三维数据的样本吗？石油桶能否适宜地代表这些数据？如果不能的话，这些数据怎么样表示才好？
2. 堆栈图 堆栈图是什么？它有什么作用？
3. 地理数据 什么是地理数据？说明呈现地理数据的至少两种图形？
4. 等高图 等高图上的等高线是指什么？当等高线封闭起来意味着什么？当等高线之间距离很远时又意味着什么？

判断正误：判断练习题 5~8 中的陈述是有意义的（表述正确）还是无意义的（表述错误），并做出解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. 三维图 苏伊声称说他需要一个三维图，呈

现萨克拉门托市一年中每一天的最高气温。

6. 堆栈图 一幅堆栈图能够用于呈现，在过去 10 年中每一年你所在大学的男生和女生人数。
7. 等高图 一幅等高图可以用来呈现美国主要城市的人口数。
8. 地理图 一位在杂志社工作的图形艺术家，通过使用不同高度的条形描述了美国主要城市的人口数，图形是通过在美国地图上的城市位置标注了条形而完成的。

概念与应用

9. 学生的性别 图 3-29 的分段条图显示了不同年份的高中毕业生的男生和女生人数，资料来源于美国国家教育统计中心。
 - (1) 用文字描述图中所揭示的趋势。
 - (2) 采用多线图重新绘制图形。简要讨论两种不同图形表达同样数据各自的利弊。

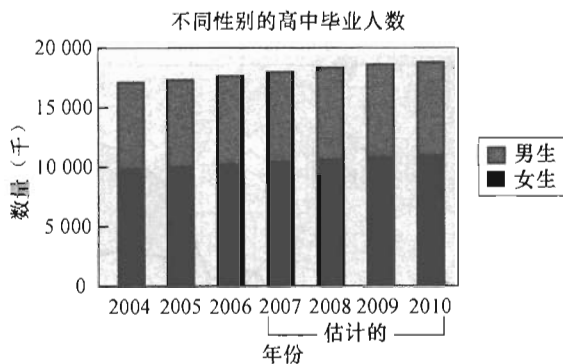


图 3-29 不同年份的高中毕业生男女生人数分段条图

10. 不同地区的住房价格 图 3-30 中的图形呈现了美国不同区域的住房价格。注意这个数据没有经过通货膨胀率的调整。

- (1) 用文字描述适用于美国所有地区住房价格的一般趋势。
- (2) 用语言描述你所注意到的不同区域的所有差别。
- (3) 用语言描述如果数据经过通货膨胀率的调整以后看起来会有什么不同。

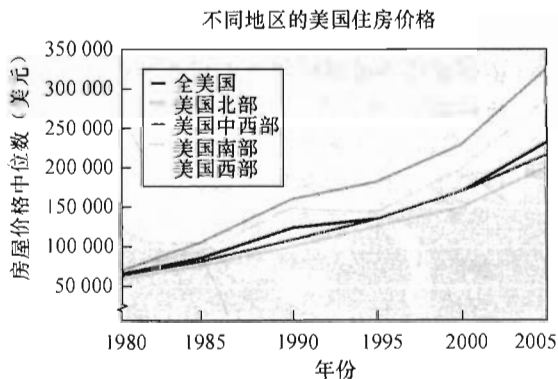


图 3-30 美国不同区域的住房价格

11. 性别和收入 思考图 3-31 所显示的不同年

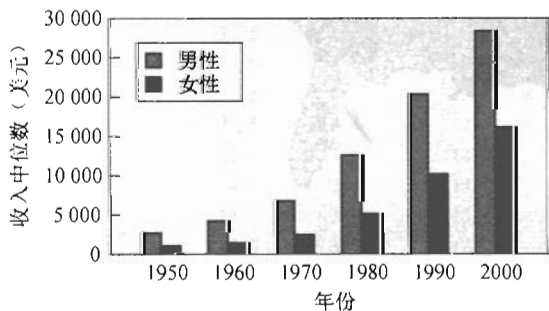


图 3-31 不同年代的男性和女性的收入中位数

代的男性和女性收入的中位数 (美国普查局)。

(1) 这个图形传达了一个什么故事? 男性和女性的收入比看起来是下降的吗? 为什么?

(2) 采用多线图重新绘制图形。简要讨论两种不同图形表达同样数据时各自的利弊。

12. 结婚率和离婚率 图 3-32 描述了自从 1900 年以来所选择的一些年份美国的结婚率和离婚率。结婚率和离婚率都是以人口中的每 1 000 人来计算的 (美国健康和人类服务部)。

(1) 为什么这些数据用结婚率和离婚率, 而不用结婚和离婚的总数? 评论你观察这些结婚率和离婚率所发现的趋势, 并且给出这些趋势可能的历史和社会原因的解释。

(2) 用这些结婚率和离婚率的数据绘制一个分段条图。在每个条形中, 将离婚率置于结婚率之上。请问哪一种图形更容易用于比较, 是分组条图还是分段条图? 并加以解释。

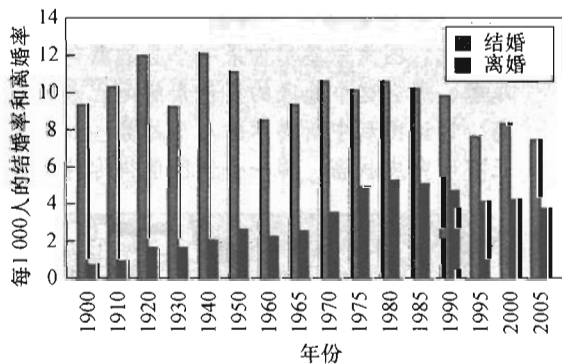


图 3-32 不同年份美国的结婚率和离婚率

13. 联邦预算 图 3-33 中的堆栈图呈现了联邦

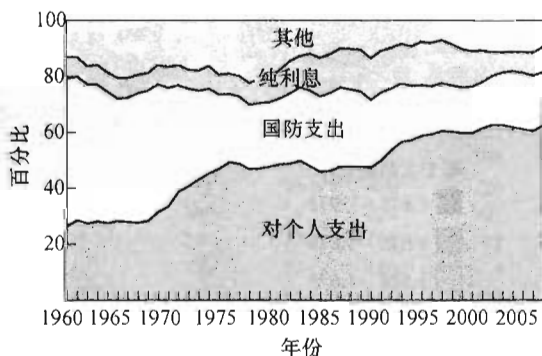


图 3-33 联邦预算的各项主要支出的百分比变化

资料来源: Office of Management and Budget.

预算的各类主要支出的变化。注意“净利息”表示美国国债的利息支出。“其他”类包括如下项目，诸如教育、环境净化、科学研究等。解释这个堆栈图，并且讨论它所揭示的某些趋势。

14. 大学学位 图 3-34 中的堆栈图显示了过去所获大学学位的男性和女性人数。
- (1) 分别估计 1930 年和 2000 年获得大学学位的男性和女性人数。
 - (2) 比较 1980 年和 2000 年获得大学学位的男性和女性人数。
 - (3) 在哪一个 10 年获得大学学位的总数增长是最多的？
 - (4) 比较 1950 年和 2000 年获得大学学位的总人数。
 - (5) 你认为堆栈图是呈现这种数据的有效方法吗？简要讨论可以替代这种图形的其他方法。

15. 黑色素瘤的死亡率 图 3-35 显示了全美国各县的黑色素瘤（一种皮肤癌）的死亡率的变化。图例中显示如果一个县的颜色越深越黑，那么这个地区的黑色素瘤死亡率就越高。讨论图形中所揭示的几个趋势。如果你正在研究皮肤癌，哪一个地区值得你特别关

注并加以研究？为什么？

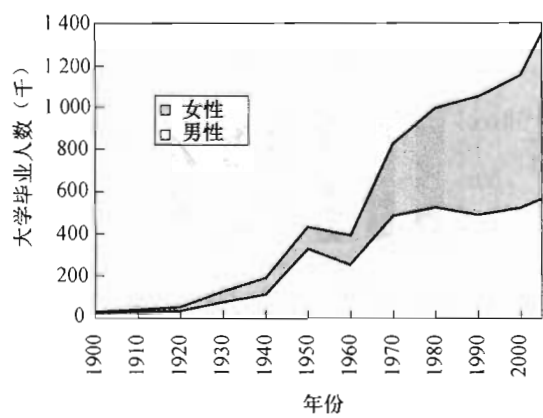


图 3-34 过去所获大学学位的男性和女性人数

16. 学校的种族隔离 测量种族隔离的一种方法是调查一个黑人学生的同学有多少人可能是白人。《纽约时报》研究发现，通过这种测量方法，自从 1990 年以来种族隔离的现象大幅度下降。图 3-36 显示了最近 10 年美国各个县一个黑人学生的同学是白人的概率。在这个图形中呈现了显著的地区差异吗？你能指出在乡村和城市地区之间的差异吗？讨论你在这个图形中所发生的几个趋势的可能的解释。

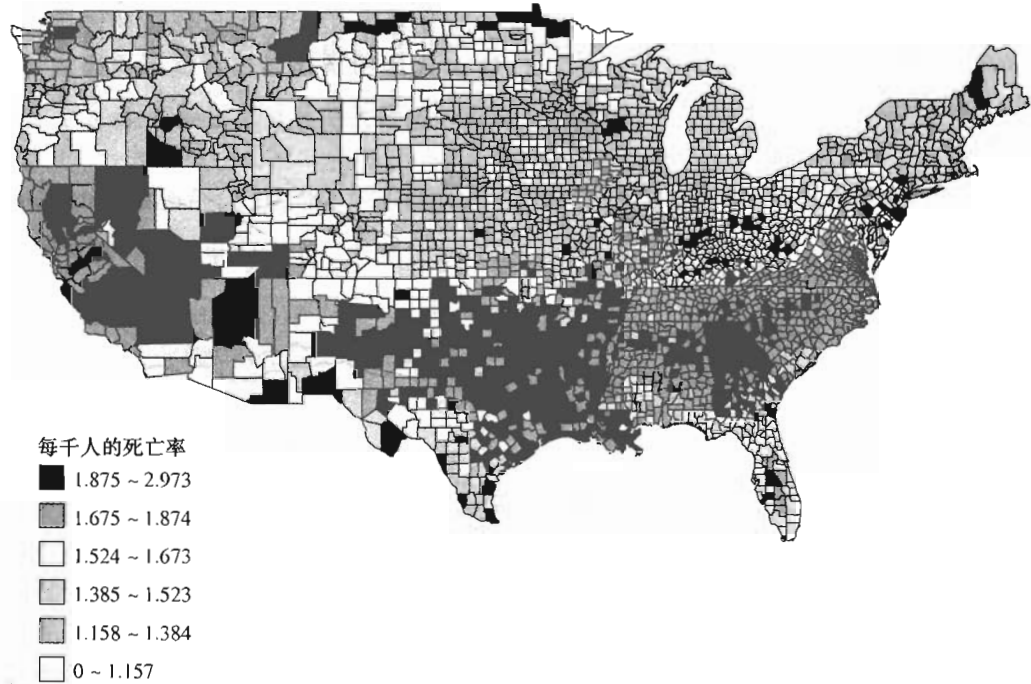


图 3-35 美国各县的黑色素瘤的死亡率

一个黑人学生有白人同学的概率

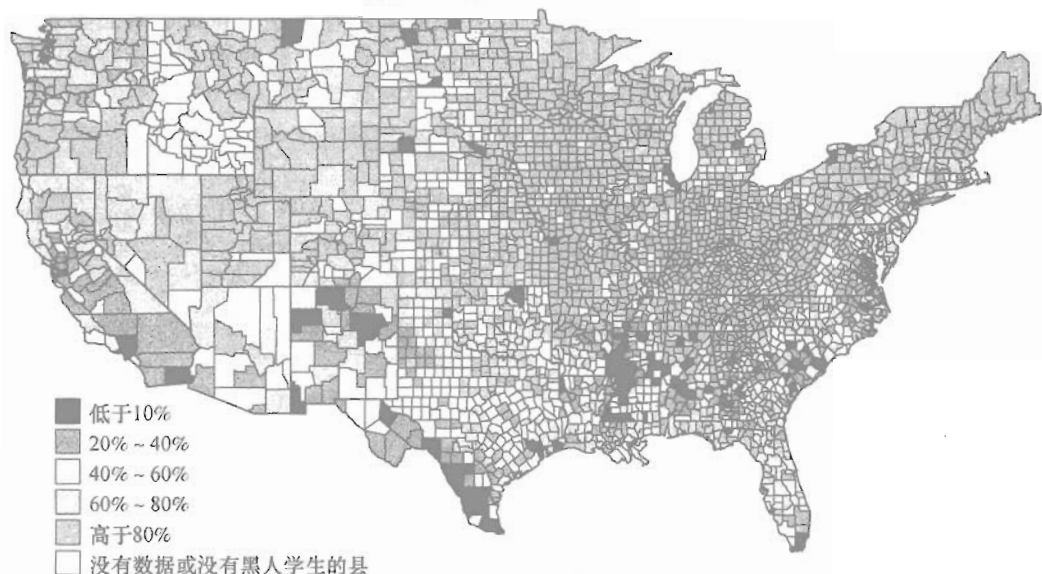


图 3-36 美国各县一个黑人学生的同学是白人的概率

资料来源: New York Times.

绘制图形: 练习题 17~20 给出了一些实际中的数据表格。你可以选择一些你认为合适的图形去呈现这些数据。除了绘制图形之外, 简单写几句话解释你为什么选择这种类型的图形, 并用几句话概括数据中的令人感兴趣的模式。

17. 酒驾 下表给出了交通事故致死的一些数字: (1) 没有酒后驾驶的原因, (2) 中等酒精浓度水平的酒后驾驶 (血液中的酒精含量在 0.01~0.09 水平), (3) 高酒精浓度水平的酒后驾驶 (血液中的酒精含量超过 0.1 水平)。所有数据值是指每 1 000 人的死亡率 (数据来源于美国联邦高速公路交通安全管理局)。

年份	没有饮酒	中量饮酒	过度饮酒
1982	18.7	4.8	20.4
1984	20.5	4.8	19.0
1986	22.0	5.1	18.9
1988	23.5	4.9	18.7
1990	22.5	4.4	17.7
1992	21.4	3.6	14.2
1994	24.1	3.5	13.1
1996	24.8	3.8	13.4
1998	25.5	3.7	12.2
2000 (估计)	25.1	3.9	12.8

18. 日报 右上表给出了自 1920 年以来选择的一些年份里日报的份数和它们的发行量

(数据来源于《编辑与出版》)。

年份	日报的数量	发行量 (百万)
1920	2 042	27.8
1930	1 942	39.6
1940	1 878	41.1
1950	1 772	53.9
1960	1 763	58.8
1970	1 748	62.1
1980	1 747	62.2
1990	1 611	62.3
2000	1 485	56.1
2003	1 456	55.2

19. 火灾 下表概括了近年来不同国家因火灾而导致的死亡人数 (数据来源于阻止枪支暴力联盟)。

国家	火灾死亡 总人数	故意纵火 杀人人数	火灾 自杀人数	重大火灾事 故死亡人数
美国	35 563	15 835	18 503	1 225
德国	1 197	168	1 004	25
加拿大	1 189	176	975	398
澳大利亚	536	96	420	20
西班牙	396	76	219	101
英国	277	72	193	12
瑞典	200	27	169	4
越南	131	85	16	30
日本	93	34	49	10

20. 参加工作的妈妈 下表列出了具有两个不同年龄组的孩子并且参加工作的妈妈的百分比 (数据来源于美国劳动统计局)。

年份	6~17岁	6岁以下
1955	38.4	18.2
1965	45.7	25.3
1975	54.9	39.0
1985	69.9	53.5
1995	76.4	62.3
2005	77.5	62.2

网络和远程项目

请在网站 www.aw.com/bbt 上打开第3章的有用链接“links for internet projects”。

21. 天气地图 许多网站提供了近期天气数据的等高图。例如，你可以登录雅虎网站。天气数据可以生成许多不同种类的等高天气图。请至少绘制两种天气等高图，并且讨论它们呈现了什么内容。
22. 联邦预算 登录美国政府管理预算局(OMB)网站，寻找与美国联邦预算有关的一些图表。选择两种你非常感兴趣的图形，并且讨论它们呈现了什么内容。
23. 容易误解的图形 参考图3-37，在图中百分比被表示成在一个人头部所占的体积部分。这种形式的数据表达能够使其更容易理解和比较吗？这种数据表达方式不会产生误导吗？可否有一种更好的方式表达同样的信息？如果可以的话，绘制你自己的图形以便更好地表达给出的信息。

新闻中的统计学

24. 复式条图 在近期的一则新闻报道中寻找复式条图或者复式线图的例子，评论这种图形表达的效果，能否有另一种图形表示方式用于描述同样的数据？

25. 分段条图 在近期的新闻报道中寻找一个分段条图的例子，评论这种图形表达的效果。能否有另一种图形表示方式用于描述同样的数据？

以年龄分组的CEO

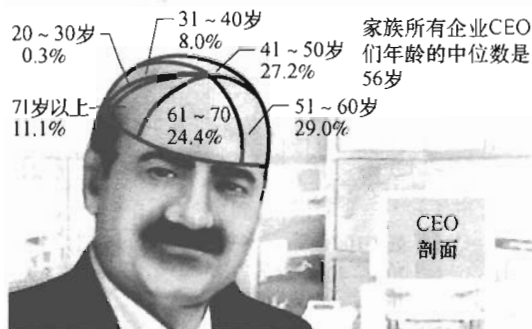


图 3-37 容易误解的图形

资料来源：Data from Arthur Anderson/Mass Mutual Family Business Survey 1997.

26. 地理图 在近期的新闻报道中寻找一个地理图的例子，评论这种图形表达的效果，能否有另一种图形表示方式用于描述同样的数据？
27. 三维图 在近期的新闻报道中寻找一个三维图的例子，这个三维图是必要的还是为了表达一些戏剧的效果？评论这种图形表达的效果，能否有另一种图形表示方式用于描述同样的数据？
28. 想象一种新式图形 在新闻中寻找由两种或两种以上基本图形类型联合而成的统计图的例子，简要解释这个图形想呈现什么内容，图形呈现的效果怎样。

相关知识

19世纪后期，德国的研究者研究了许多类型的图表，图3-38和图3-39这种类型的歪曲是很常见的，以致它们有自己单独的命名，大致意思是“通过视觉直观的平方将使效应变得更加强烈”。

3.4.1 错觉

很多图形的绘制容易让我们产生错觉，图3-38给我们呈现了一种最常见的错觉类型。

3.4 有关图表的几个注意事项

正如前面我们所看到的，图形可以清楚和有意义地对统计数据进行总结。然而，即使制作精良的图形，如果我们不认真地解读它，也会导致误解，而粗制滥造的图形几乎总是误导人们。更有甚者，有一些人会故意地歪曲图形的意思。在这一节中，我们将讨论一些让我们曲解图形的常见方式。

美元形状的条形图经常用来呈现美元随着时间的贬值情况。条形的宽度表示数据，但是我们的眼睛却会关注条形的面积。例如，右边的条形假定用来表示2006年的1美元，它的价值相当于1980年1美元的41%，它的宽度的确是左边条形的41%，但是它的面积要小得多（大约是左边条形面积的17%）。这给我们造成的知觉就是美元贬值的情况比实际上要大得多。

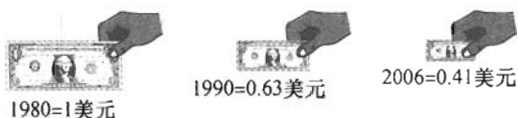


图3-38 美元的长度对应它们的购买力，但是我们眼睛常常会关注比长度下降得多的面积

以宽度作为重要的测量尺度的时候，呈现一个立体的图形甚至会产生更大的错觉。图3-39用电视机来表示1980年和2005年开通有线电视的家庭数量。在这段时间内，这一数量增长了3倍，开通有线电视家庭从1800万增长到7300万，如图3-39电视的高度所表示的。然而，我们的眼睛却被电视机的体积所吸引，两者之间的差别约是 $4^3 = 64$ ，这样，这个数量的增长看起来要比实际上大得多。

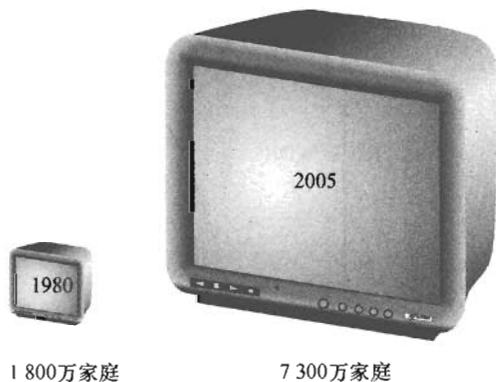


图3-39 开通有线电视家庭数量

电视机的高度对应重要的测量数字，但是我们眼睛却关注它们的体积。

相关知识

近年来有线电视的缓慢增长是由于卫星电视服务的竞争。当将来电话线和网络也可以提供电视节目时，这种竞争还会加剧。



3.4.2 观察刻度值

图3-40呈现了一个关于美国家庭自有住房的条形图，第一眼看上去，在1960~2000年这一时期，拥有自己住房的人口比例增长速度要比近年来快得多。但是再仔细看横坐标的刻度尺：前5个类别表示的年份是10年一个间距，而后面两个类别表示的年份是2003年和2005年。如果在一个完整的10年内，每年都重复从2003~2005年这样很小的增长幅度，那么这个增长速度与其他10年期的趋势并没有什么不同。这个图形第一印象所造成的错觉是由于横坐标轴采用了不统一的刻度值。

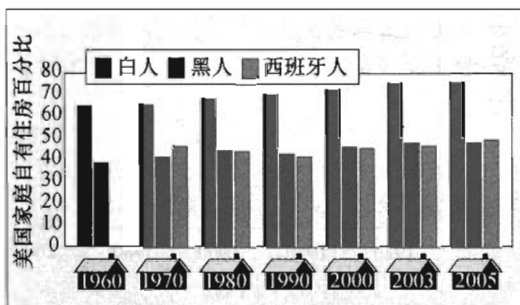


图3-40 美国不同族裔家庭的自有住房百分比（注意横坐标的数据刻度）

资料来源：U. S. Department of Commerce, U. S. Census Bureau; adapted from the *Wall Street Journal Almanac*.

相关知识

到2005年，美国拥有自有住房的家庭将占到全部家庭的69.0%，达到历史上的最高点。



思考题

基于图3-40和你自己的直觉，预测2020年美国家庭私有住房比率是多少（依种族分类或者西班牙裔），说明你预测背后的原因。

纵坐标也会发生同样的刻度值问题。图3-41a呈现了从1910~2005年之间女性大学生的百分比。第一眼看上去，大约从1950年开始

这个百分数有一个很大程度的增长，但是纵坐标不是从0开始也不是结束于100%。如果我们重新绘制一个纵坐标刻度值为0~100%（图3-41b）的图形，会发现实际上是增长着的但这个增长并不是那么急剧。从数学的角度来看，刻度值不从零开始是完全可靠的，它可以让我们更容易地看到数据在很小的刻度值上的变化趋势。尽管如此，如同例子中显示的，如果你不仔细观察刻度值，有可能导致视觉错误。

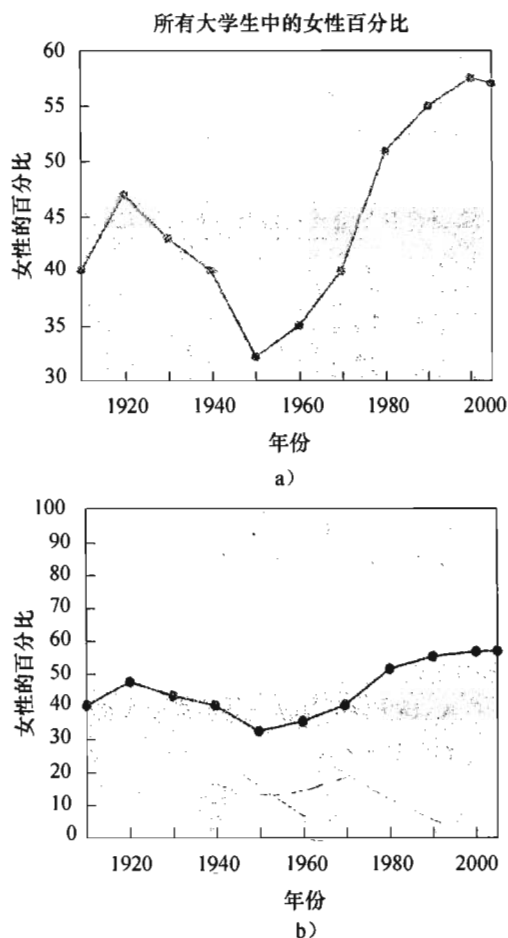


图3-41 大学生中女性百分比
呈现相同数据但是看起来截然不同的两个图（因为纵坐标有着不同的刻度值）。

a) 纵坐标不是从0开始 b) 纵坐标刻度值为0~100%

有时刻度值本身并不模糊，但是如果你不仔细检查仍然有可能导致误解。看一下图3-42a，图中呈现的是随着时间而增长的计算机运算速度。第一眼看上去，计算机的运行速度是线性增长的。例如，看起来，1990~2000年，计算机运算速度的增长与1950~1960年的增长

是相同的。然而，如果我们仔细去看，我们会发现纵坐标轴上的刻度，每1格代表10倍速度的增长。现在我们看到从1950~1960年计算机的运算速度从大约每秒一次运算增长到每秒100次运算，1990~2000年，计算机的运行速度从大约每秒1 000 000 000次运算增长到每秒100 000 000 000次运算。这种类型的刻度值叫做指数刻度（或者叫对数刻度），因为从10的各次幂增长到10的更高次幂叫做指数（例如，10的3次幂记做 10^3 ）。以图3-42a的指数图 and 图3-42b的原始数据图进行对比，通过比较你就会发现，指数刻度值呈现在一个巨大范围之内变化的数据是非常有用的。因为你计算机的运行速度增长得如此之快，如果采用原始刻度值，在这个图形中就会发现，如果想看到计算机运行速度早期的细节是根本不可能的。

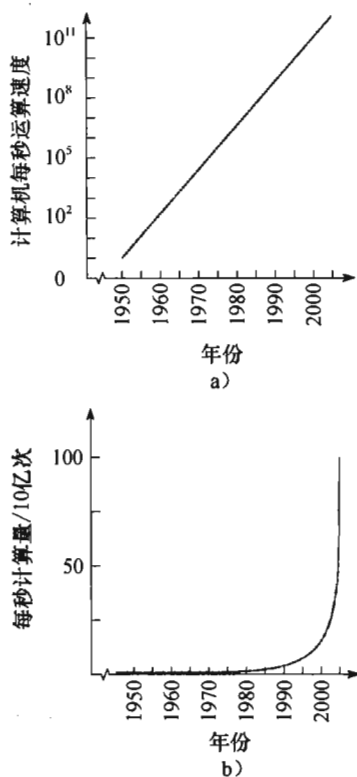


图3-42 计算机运算速度
呈现相同数据的两个图（左边的图采用对数刻度值）。

a) 对数刻度值 b) 实际数量刻度值

相关知识

1965年，英特尔的创建者戈登·E·摩尔（Gordon E. Moore's）指出技术的发展将

使计算机芯片的功能大约每间隔2年增长1倍,这个观点现在被称做“摩尔法则”,自从摩尔首次表述它以来,它已经被实践证实。

思考题

基于图3-42a,你能预测2015年计算机的运行速度有多快吗?基于图3-42b,你还能做同样的预测吗?说明原因。

案例研究

来自小行星的威胁

小行星存在极小的可能性会撞击地球。一些小行星在大气层中就被燃烧殆尽,还有一些落到地面形成撞击圆形坑。但是较大的一些小行星可能会带来实质的灾害。大约6500万年以前,一个直径大约10公里的小行星撞击了地球,在墨西哥尤卡坦半岛的海岸上形成了200公里宽的环形坑。科学家估计那次撞击,使得那时生活在地球上3/4的物种消失,包括所有的恐龙。

很明显,对我们人类来说一次同样的撞击将是一个灾难性的消息。我们还想知道类似的事件发生的可能性是多少。图3-43呈现了小行星或彗星这样的星体的大小和撞击地球的频率;两个坐标轴都是指数刻度。横坐标轴显示了撞击星体(小行星或彗星)的大小,每一个刻度值代表10的幂,纵坐标轴显示了撞击的频率,纵坐标轴往上移动对应着频率更高的撞击事件。在这个双指数图上,我们可以看到更明显的趋势。例如,直径大约1米的小星体每天都会撞击地球,但是只有很小的危害,在另一端,导致大量生物灭绝的足够大的星体撞击地球,这种情况数千万年才仅发生一次。

介于中间的最可能发生的情况才是最令人烦恼的。图中显示,那些能导致“大面积灾害”的星体撞击,诸如一个人口大城市被彻底摧毁,这样的事件有可能几千年就会发生一次。这是我们要采取一些预防措施的依据之所在。最近,天文学家努力尝试预测这样的星体什么时候可能撞击地球。如果想要发现这样的将要撞击地球的星体,科学家必须找到一条预防这种撞击的方法。

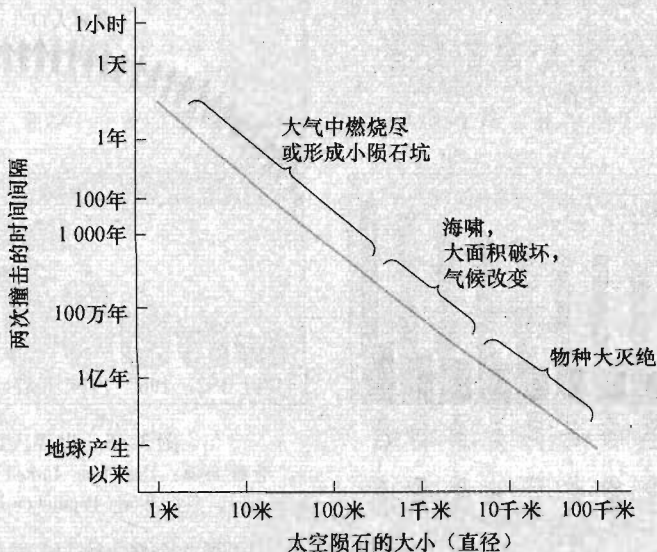
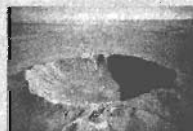


图3-43 小行星或彗星这样的星体的大小和撞击地球的频率(注意越小的星体撞击地球的可能性大)

相关知识

地球上已经发现了超过150个陨石坑,保存得最好的一个是美国亚利桑那州的陨石坑(如右侧照片),它形成于大约50000年以前,直径大约50米的小行星撞击了地球,释放的能量相当于20颗百万吨级氢弹,最后留下一个大约1000米宽200米深的陨石坑。



3.4.3 百分比变化图

大学学费是变得更昂贵了还是更廉价？快速浏览一下图 3-44a 可能会给你一个解释，私立大学的花费一直很稳定，而公立学校花费在 2006 年经历了一段时期的增长以后开始回落。但是仔细地看下，就会发现根本不是这个情况。图中的纵坐标表示的是花费的百分比增长。平面图仅仅意味着每年花费增长的百分比是相同的，但并不意味着花费是保持稳定的。同样，公立大学 2006 年的下降意味着那年花费的增长率要比前一年少。

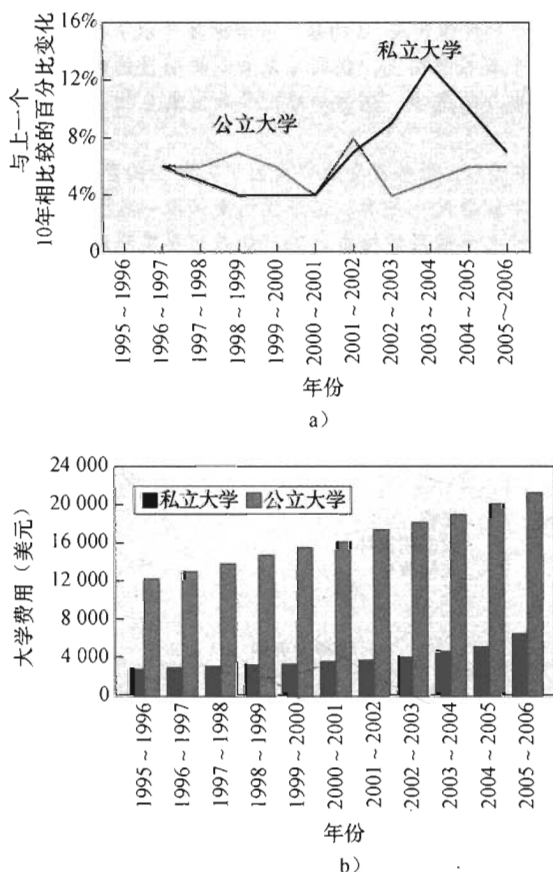


图 3-44 大学费用的发展趋势

a) 每年的百分数变化 b) 实际花费钱数

资料来源：The College Board.

事实上，对于公立大学和私立大学而言，实际的花费（没有经过通货膨胀调整）都是随着时间增加的，如图 3-44b 所示。更进一步来说，因为通货膨胀率（以 CPI 为指标）低于大学花费的增长率，实际上公立大学的花费是稳

步上升的，用来显示百分比变化的图形非常常见，尤其用于经济数据，虽然它们是非常可靠的，如果你不仔细地解读它们，很有可能被误导。

3.4.4 象形图

象形图是一种经过艺术加工的图形。这种艺术加工使得图形更直观，但是这种图形也容易让人分散注意力或者被误导。图 3-45 是一个象形图，呈现了从 1804 ~ 2054 年世界人口的增长（未来数年的人口基于美国的推测）。条形的长度正确对应于每一年的世界人口，然而，图形的艺术加工在几个方面让人困惑。例如，你的眼睛可能会注意到图形上方站在地球之上的人们。因为从象形图的左边到中心人们的位置是上升的，在从中心到右边人们的位置是下降的，他们可能给我们一种暗示未来的人口数是下降的。实际上，排成一队的人群只是一种艺术装饰而没有任何实际意义。

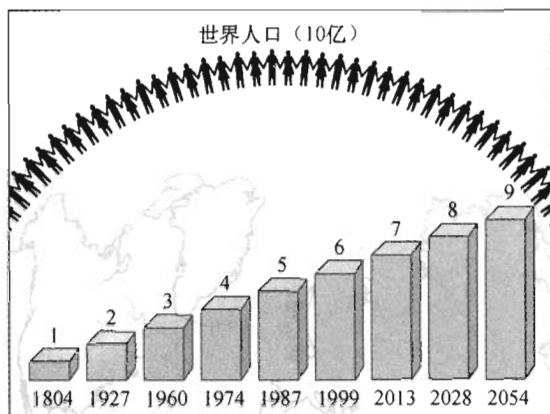


图 3-45 世界人口的象形图

资料来源：Data from United Nations Population Division, World Population Prospects.

也许这个象形图最严重的问题是它使得我们看起来世界人口是线性增长的，然而，注意横坐标中在每一个条形之间的时间间距不是等同的。例如，从 10 亿人口增长到 20 亿，这个时间间距是 123 年（1804 ~ 1927 年），但是从 50 亿人口增长到 60 亿人口，仅用了 12 年（1987 ~ 1999 年）。

象形图是很常见的。如同这个例子中所呈现的，不管怎样，你必须认真地研究它们，以提取其中潜在的信息，而不被装饰的图形所迷惑。

相关知识

象形图经常被用于人口成倍增长的时间，人口增长到原来的两倍需要的时间。在20世纪后期，人口成倍增长的时间大约是40年，如果继续以这个速度增长，世界人口将在2100年达到340亿，在2200年达到1920亿，到2650年，人口总数将大到即使人们挤在一起站着，地球上也挤不下。

3.4 练习题

统计素养和批判性思维

1. 纵坐标轴 当纵坐标轴的刻度值不是从零开始时统计图是怎样引起误解的？
2. 象形图 什么是象形图？简要说明象形图的优点和缺点。
3. 指数刻度值 什么是指数刻度值？什么情况下在统计图里应用指数刻度值是有利的？
4. 百分比变化图 解释图中百分比显示条形图或线看似是下降的，而此时关注变量本身却是增长的。

概念与应用

5. 刹车距离 图3-46显示了相同条件下所测量的不同汽车的刹车距离。讨论这样呈现为什么具有欺骗性。本田讴歌的刹车距离比沃尔沃S80的刹车距离多多少？制作一个更公正的统计图。

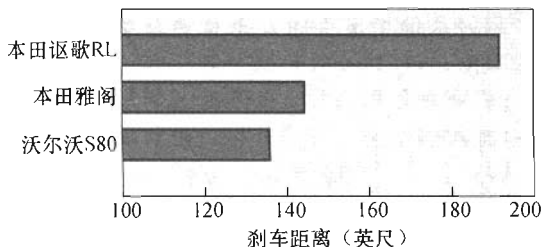


图3-46 刹车距离

6. 收入比较 思考图3-47中所显示的条形图，比较了男性和女性年收入的中位数。说明其易误导的方面，制作一个更客观公正的统计图。

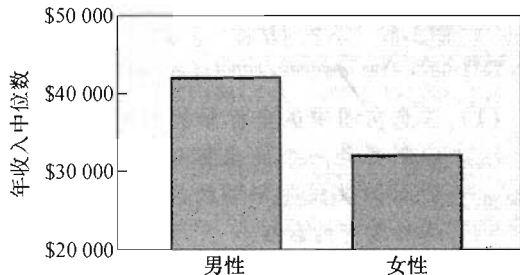


图3-47 男性和女性的年收入

资料来源：U. S. Census Bureau.

7. 统计图 图3-48描述了美国和日本每天的汽油消耗量。说明其是否准确地描述了数据？为什么？

每天燃油消耗量（百万桶）

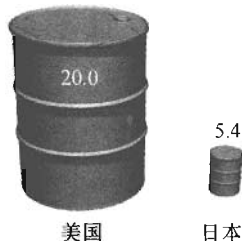


图3-48 美国和日本每天的汽油消耗量

8. 象形图 根据练习题7中图3-48制作一个条形图以便更公正和客观地表示数据。
9. 三维饼图 图3-49中的饼图描述了1990年和2050年（预期的）美国人三个年龄分组人数所占的百分比。

1990年美国人的年龄分布

2050年美国人的年龄分布

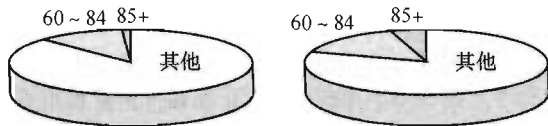


图3-49 1990年和2050年（预期的）美国人3个年龄分组人数所占的百分比

资料来源：U. S. Census Bureau.

- (1) 思考1990年的年龄分布。1990年3个年龄分组实际的百分比是87.5%（其他年龄）、11.3%（60~80岁）、1.2%（85岁以上）。三维饼图是否能够真实地反映这些数值？并做出解释。
- (2) 思考2050年的年龄分布。2050年3个年龄分组实际的百分比是80.0%（其他）、15.4%（60~84岁）、4.6%（85岁以上）。三维饼图是否真实地反映这些数值？并做出解释。
- (3) 用第1小题和第2小题中实际的百分比做一个平面的饼图，来说明这些数据。

说明为什么平面饼图比起三维饼图来，能够更真实地显示数据。

(4) 根据两个饼图评论它们的一般发展趋势。

10. 香烟象形图 考虑 3-50 中的象形图，简要描述这个图形呈现了什么？讨论其是否有效地表达了目标，还是具有某种方式的欺骗性？

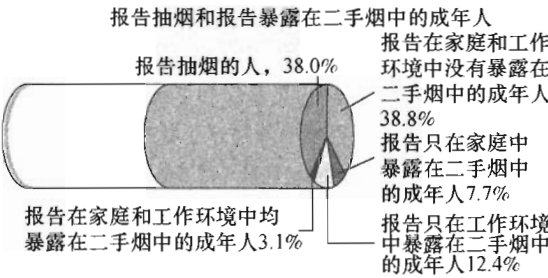


图 3-50 香烟象形图

资料来源：Centers for Disease Control and Prevention, Wall Street Journal Almanac.

11. 手机用户 下面的图表显示美国的手机用户数量。

年份	数量	年份	数量
1985	340	1997	55 312
1987	1 231	1999	86 047
1989	3 509	2001	128 375
1991	7 557	2003	158 722
1993	16 009	2005	207 900
1995	33 786		

(1) 制作这些数据的一个时间序列图，在两个坐标轴上使用一致的刻度。

(2) 制作一个指数图，纵坐标轴上的手机用户单位是 1, 10, 100, 1 000, 10 000, 100 000, 1 000 000。

(3) 比较第 1 小题和第 2 小题中所做的统计图。

12. 摩尔定律 1965 年，英特尔公司的创始人戈登·摩尔发现了摩尔定律：集合电路每平方英寸晶体管的数量大约每 18 个月会翻一倍。下表列出了几个不同的年份晶体管的数量（以千为单位）。

年份	晶体管数量	年份	晶体管数量
1971	2.3	1993	3 100
1974	5	1997	7 500
1978	29	1999	24 000
1982	120	2000	42 000
1985	275	2002	220 000
1989	1 180	2005	410 000

(1) 使用上述数据制作一个时间序列图，

在两个数轴上采用一致的刻度。

(2) 使用这些数据制作一个指数图，纵坐标轴上的单位是 1, 10, 100, 1 000, 10 000, 100 000, 1 000 000。

(3) 比较第 1 小题和第 2 小题中所做的统计图。

13. CPI 百分比变化 图 3-51 显示了连续 17 年 CPI 百分比的变化情况。哪一年 CPI 变化最大？哪一年 CPI 变化最小？比较 1991 年和 1990 年的真实价格。根据这个图，你能总结出这段时期价格是怎样变化的吗？

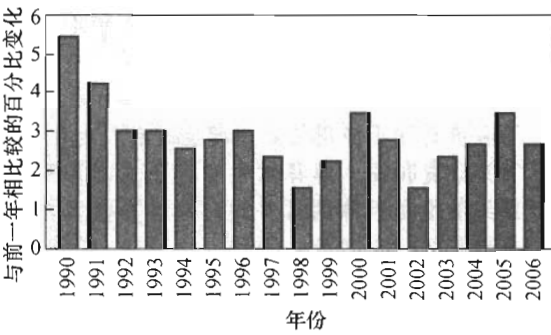


图 3-51 连续 17 年美国 CPI 百分比变化

资料来源：U. S. Bureau of Labor Statistics.

14. 季节对精神分裂症的影响 图 3-52 显示在不同的月份里出生的人患精神分裂症的相对风险。

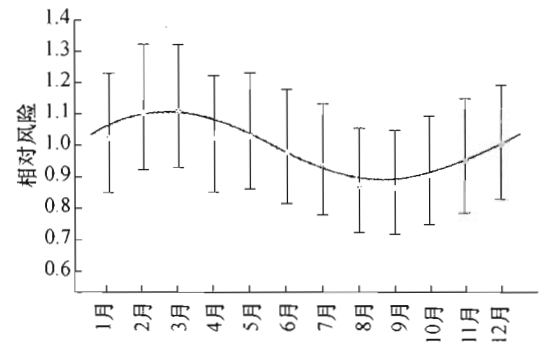


图 3-52 季节对精神分裂症的影响

资料来源：New England Journal of Medicine.

(1) 注意在图中纵坐标轴的刻度不包括 0。绘制另外一个纵坐标轴刻度从 0 开始的图形以描述相同的风险曲线。评论这种改变的影响。

(2) 每月的相对风险值都用一个最可能代表的点来表示，还有一个“误差幅度条形”来表示该数值可能存在的区间。研

究结论说“出生季节与患精神分裂的风险存在显著相关”。结合误差幅度的大小,这种说法合理吗?(是否可能绘制一条平顶线穿过所有误差幅度条形呢?)

15. 坚挺的美元 图 3-53 呈现了美国的最低保障工资以及购买力,数据以 1996 年美元购买力作为参考值进行了通货膨胀率的调节。图中呈现了从 1955~2006 年之间的数据变化。请概括图中所显示的内容。

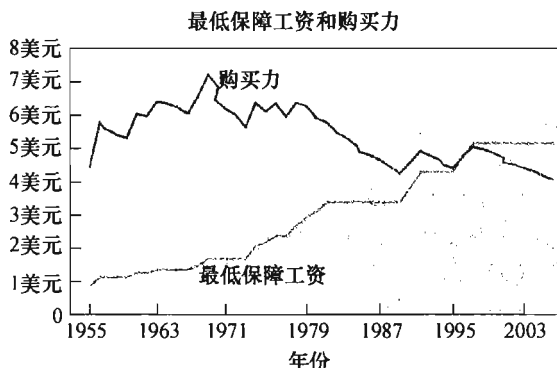


图 3-53 美国的最低保障工资以及购买力

资料来源: U. S. Department of Labor.

16. 双线图 图 3-54 呈现了 1946~1964 年和 1977~1994 年两段时期内美国新生儿的数量。第一段时期内什么时间新生儿的数量达到顶峰?第二段时期内什么时间新生儿的数量达到顶峰?你认为为什么设计者分两个时期来展示数据,而不是用一条从 1946 年贯穿到 1994 年的曲线展示数据?

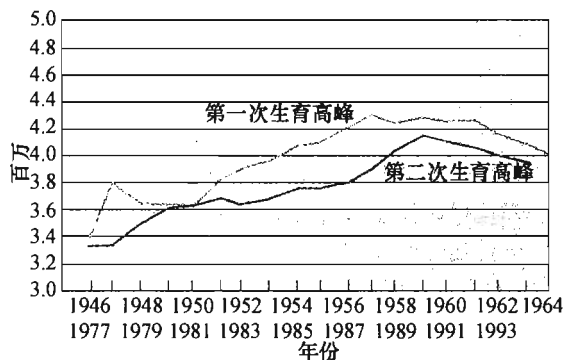


图 3-54 美国两段时期内新生儿的数量

注: 第一次生育高峰中出生的一代人,他们出生于 1946~1964 年,在 1977~1994 年产生了规模较小的第二次生育高峰。

资料来源: National Center for Health Statistics,
Wall Street Journal Almanac.

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 3 章的有用链接“links for internet projects”。

17. 美国印象 《今日美国》提供一个叫做“美国印象”的日常象形图,找到关于《今日美国》本周主题的一个象形图,简要讨论其目标 and 作用。
18. 绘制统计图 当前,即使从事相同的工作,女性的收入也会少于男性,计算男性收入 1 美元时女性会收入多少,制作一个统计图能够客观地反映信息,然后制作另外一个夸大这些差异的统计图。

新闻中的统计学

19. 被曲解的新闻 在近期的新闻报道中找到一个感觉被歪曲了的统计图,解释这种歪曲所产生的影响,并说明怎样才能做到公正地制作统计图。
20. 新闻中的刻度问题 在近期的新闻报道中找到一个统计图,其纵坐标不是从 0 开始的,猜想为什么这样制作统计图,并且讨论这个图是怎么样让人们产生误解的。
21. 新闻中的经济图 在近期的新闻报道中,找到一个呈现过去经济数据的统计图。这组数据经过通货膨胀率的调整了吗?讨论图形的意义,并说明图形在哪些方面有可能会欺骗你。
22. 新闻中的象形图 在近期的新闻报道中找到一个象形图的例子。讨论象形图试图展示什么,并且讨论其艺术设计是有助于还是掩盖了这个目标。
23. 新闻中视觉效果明显的统计图 从近期的新闻报道中找到一个统计图,从你的观点来看,这个图从视觉上展示数据的效果是非常优秀的。讨论统计图要呈现什么,并解释你为什么认为它呈现数据的效果明显。
24. 新闻中效果不显著的统计图 从近期的新闻报道中找到一个统计图,从你的观点来看,它试图以有意义的方式直观具体展示数据的效果是失败的。讨论这个统计图想展示什么,并解释它为什么效果是失败的。说明怎样改进效果更好一些。

第3章复习题

下面所列的数据是听装普通百事可乐和无糖百事可乐抽样测量的净重（磅）。用这些数据完成复习题1~3。

普通百事可乐

0.825 8 0.815 6 0.821 1 0.817 0
0.821 6 0.830 2 0.819 2 0.819 2
0.827 1 0.825 1 0.822 7 0.825 6
0.813 9 0.926 0 0.822 7 0.838 8
0.826 0 0.831 7 0.824 7 0.820 0
0.817 2 0.822 7 0.824 4 0.824 4
0.831 9 0.824 7 0.821 4 0.829 1
0.822 7 0.821 1 0.840 1 0.823 3
0.829 1 0.817 2 0.823 3 0.821 1

无糖百事可乐

0.792 5 0.786 8 0.784 6 0.793 8
0.786 1 0.784 4 0.779 5 0.788 3
0.787 9 0.785 0 0.789 9 0.787 7
0.785 2 0.775 6 0.783 7 0.787 9
0.783 9 0.781 7 0.782 2 0.774 2
0.783 3 0.783 5 0.785 5 0.785 9
0.777 5 0.783 3 0.783 5 0.782 6
0.781 5 0.779 1 0.786 6 0.785 5
0.784 8 0.780 6 0.777 3 0.777 5

1. (1) 制作关于普通百事可乐净重的次数分布表，使用以下分组区间：
- 0.813 0~0.817 9
0.818 0~0.822 9
0.823 0~0.827 9
0.828 0~0.832 9
0.833 0~0.837 9
0.838 0~0.842 9
- (2) 制作关于无糖百事可乐净重的次数分布表，使用以下分组区间：
- 0.774 0~0.777 9
0.778 0~0.781 9
0.782 0~0.785 9
0.786 0~0.789 9
0.790 0~0.793 9
- (3) 比较第1小题和第2小题中的次数分布表，它们有什么明显的不同？怎样解释这些明显的不同？

2. (1) 制作关于普通百事可乐净重的次数分布表，使用以下分组区间：
- 0.813 0~0.817 9
0.818 0~0.822 9
0.823 0~0.827 9
0.828 0~0.832 9
0.833 0~0.837 9
0.838 0~0.842 9
- (2) 制作关于普通百事可乐净重的累加次数分布表。
3. (1) 用复习题1第(1)小题中的结果制作关于普通百事可乐净重的直方图。
(2) 用复习题1第(2)小题中的结果制作关于无糖百事可乐净重的直方图。
(3) 比较第(1)小题和第(2)小题中的直方图，它们的相似点和不同点分别是什么？
4. 饼图 近些年来，5 524 名人员因工伤致死。以下是其死因列表，制作一个饼图来呈现这些数据。

死亡原因	死亡人数（人）
交通事故	2 375
碰触物体或仪器	884
袭击或暴力行动	829
高空坠落	718
暴露在有害的物质或环境中	552
火灾或爆炸	166

资料来源：美国劳动统计局。

5. 帕累托图 根据复习题4中所给出的数据制作一个帕累托图。比较帕累托图和饼图，哪一个图能更有效地显示工伤致死的原因，并做出解释。
6. 条形图 图3-55中显示了2000年和2005年去美国留学的中国学生的数量。这个图存在哪些错误？制作一个统计图能更公正和客观地描述这个数据。

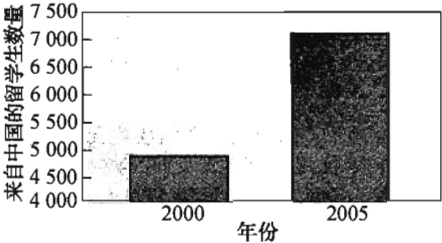


图3-55 2000年和2005年去美国留学的中国学生数量

第3章小测验

1. 从参加国家健康调查的人员中, 随机抽取一个样本, 得到不同颜色眼睛的次数, 哪一种图能够更恰当地表示这组数据: 直方图、条形图、复式条形图、分段条图?
2. 为了调查性别和眼睛颜色的关系, 从参加国家健康调查的人员中, 随机抽取一个男性样本和一个女性样本, 得到不同颜色眼睛的次数。哪一种图能够更恰当地表达这组数据: 直方图、条形图、复式条形图?
3. 从参加国家健康调查的人员中随机抽取一个样本, 测量其腕部宽度。哪一个图能恰当地表示这组数据: 直方图、条形图、复式条形图、分段条图?
4. 在一个次数分布表中, 其相对应的次数是 7。数值 7 说明了什么?
5. 在一个相对次数分布表中, 最低组是 50 ~ 59, 其对应的相对次数是 0.2。数值 0.2 说明了什么?
6. 下面茎叶图中的数值分别说明了什么?

0	5
0	
0	
1	0
1	22
1	4

7. 图 3-56 中的条形图显示了 2000 年和 2004 年美国出生的双胞胎的数量。这个图如何容易引起误解?

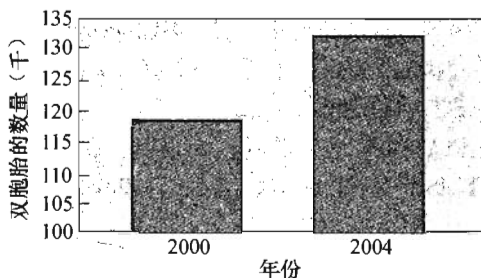


图 3-56 2000 年和 2004 年美国出生的双胞胎的数量

8. 图 3-57 中的点柱图最大的数值说明了什么?

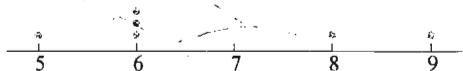


图 3-57 点柱图

9. 图 3-58 中显示从 1982 ~ 2000 年美国酒后驾驶机动车致死的人数, 1982 年酒驾导致死亡的人数是多少? 2000 年是多少?

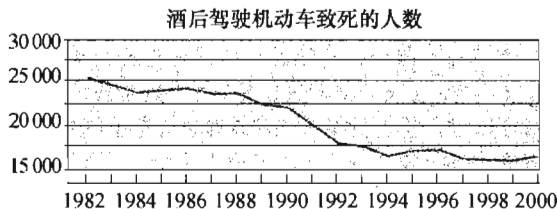


图 3-58 1982 ~ 2000 年美国酒后驾驶机动车致死的人数

资料来源: National Highway Traffic Safety Administration

10. 一家报纸的记者想制作一个关于近些年来美国每年的汽油消耗量的图。为什么不能根据一个特殊的年份 (例如 1996 年) 作参考值调整每年的数量?

技术应用

强大的软件包能有效地制作一系列图表, 其详细的说明有的很简单, 有的非常复杂, 所以我们下面提供一些相关说明。

SPSS

用 SPSS 可以制作条形图、线图、饼图、帕累托图、直方图、时间序列图。首先打开数据库, 然后点击 [Graphs], 进而选择想要的图的类型。例如, 为了制作直方图, 选择菜单中的 [Histogram], 点击一个数据单元, 点击 [variable], 然后点击 [OK]。注意: 学生版

本 SPSS 限制不能超过 1 500 个数值。

Excel

尽管它也能制作直方图、条形图、线图、饼图、散点图、复式线图 and 复式条形图, 但是 Excel 很难使用。可以把 DDXL add-in 装到 Excel 中, 它很容易制作直方图、条形图、点柱图和饼图。DDXL add-in 可以在 www.aw.com/bbt 下载, 在 Excel 中安装 DDXL 后, 步骤如下:

首先进入 A 栏样本数据, 然后点击 [add-in

DDXL], 选择 [Chart and plots], 点击 [function type], 选择菜单中想要的图标。例如为了得到直方图, 从菜单栏中选择 [Histogram], 然后点击 [pencil icon] 进入一个数据单元。

STATDISK

STATDISK 很容易制作直方图。进入数据的 STATDISK 数据窗口, 点击 [Data], 点击

[Histogram], 然后点击 [Plot]。(如果你想自己设置组距或设置起点, 在点击 [Plot] 之前点击 [User defined]。)

为了制作饼图, 首先进入 STATDISK 数据窗口第一列名单分类, 然后进入第二列相关的次数; 选择 [Data], 点击 [Pie Chart], 然后点击 [Chart]。

| 聚焦历史 |

战争可以用图表描述吗

一场战争可以使用一张图表进行描述吗? 图 3-59 是查理·约瑟夫·米纳德在 1869 年所绘制的。它是如此引人注目。这个图表告诉我们关于拿破仑在 1812 年和俄罗斯的一场战争中悲惨的命运, 因此有时它又被称做拿破仑的死亡进军之旅。

下面的米纳德图画中的地图, 显示了从波兰边境上的尼曼河到莫斯科之间大约 500 公里的一片土地。黑色的区域描述了拿破仑军队的进军路线。在米纳德原版的图画中, 1 厘米宽度代表 6 000 个人。这个副本比原版缩小了很多。这个进军的路线从最左边开始, 那时黑色区域是最宽阔的。在 1812 年 6 月 24 日开始向莫斯科进军时, 军队的全部人数是 422 000 人。它是当时所能调动起来的最庞大的军队。

随着进军路线向莫斯科延伸, 黑色区域越来越狭窄, 说明军队在不断减员。(黑色的分支代表被派往其他方向的军队。) 拿破仑仅仅携带了很少的一部分食物供应, 炎热的夏天伴随着瓢泼大雨给部队带来了猖獗的疫病。饥饿、疾病和战役每天要杀死成千上万的士兵。一直到军队在 9 月 14 日进入莫斯科的时候, 部队减员只余下 100 000 人, 然而更大的灾难还在后面。

让拿破仑感到沮丧的是, 俄国人撤出了莫斯科, 以此来对抗法国军队的到来。丧失了和俄罗斯军队进行决战的机会, 如果继续朝俄罗斯首都圣彼得堡进军的话, 拿破仑感觉到他的军队条件实在是太恶劣了, 因此他率领他的军队从莫斯科向南方撤离。图中下面的一部分区域(灰色部分)表达了这次撤退。靠近图中底部的深灰色线条呈现了冬天到来时晚上的温度。我们看到 10 月 18 日时温度已经降到零度以下。

在接下来 11 月温度持续下降一直在零度以下。在 11 月 28 日左右下面的灰色区域突然变窄了, 呈现的是 22 000 名士兵在贝尔齐纳河岸死亡了。3/4 的幸存者在接下来的几天里也冻死了不少, 许多人死在了 12 月 6 日晚上的严寒之中。到 12 月 14 号军队到达波兰的时候, 出发时的 422 000 人仅仅剩下了 10 000 人。

在关于图表技术的一个著名的分析中, 作者爱德华·塔夫特描述了米纳德的图表可能是“历史上绘制的最好的统计图表”。但是另一个引人注目的陈述来自米纳德同时代的人, 艾提安·朱尔斯·马莱, 他认为这张图表“让所有法兰西人潸然落泪”。

问题讨论

1. 讨论这张图表如何帮助你更客观地面对在一场战争中许多人死亡的客观事实, 你自己体验到了哪一种心理冲突?
2. 注意到这张图表有 6 个变量: 两个方向变量(南北和东西), 军队的人数, 军队驻扎地, 军队运动的方向和撤退中的温度。你认为米纳德通过这几个变量可以得到什么观点? 为什么?
3. 讨论你怎么样制作一张类似的图, 用于描述其他的历史和政治事件。

建议阅读

1. Tufte, Edward R. *The Visual Display of Quantitative Information*. Graphics Press, 1992.
2. Wainer, Howard. *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot*. Copernicus (New York), 1997.



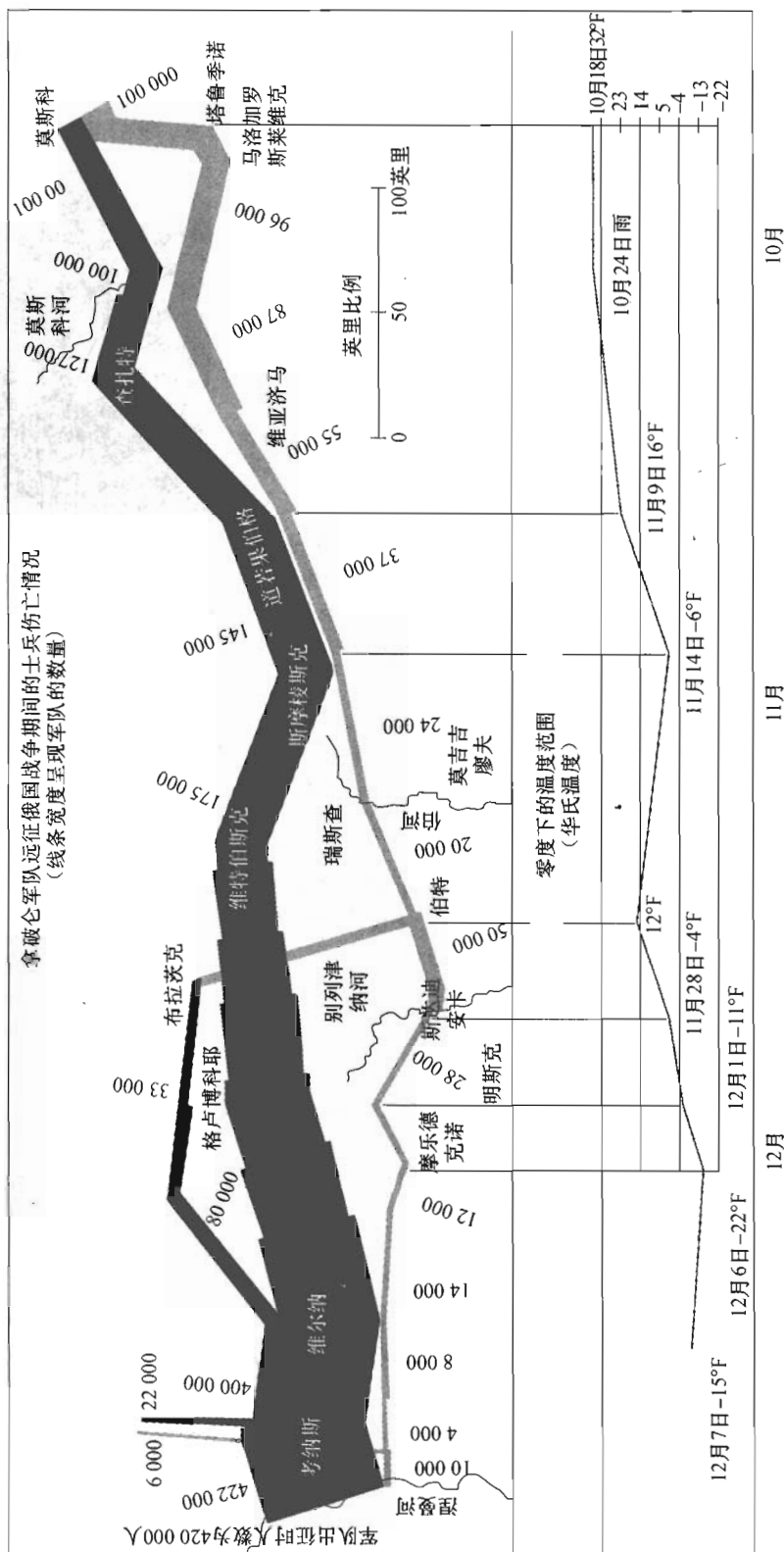


图3-59 拿破仑的死亡进军之旅

资料来源: Edward R. Tufte, *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 1983). Reprinted with permission.

| 聚焦环境 |

我们怎样才能直观呈现全球变暖

全球变暖确实是我们这个时代讨论最多的主题。但是这个问题到底有多么严重，我们怎样才能确定全球变暖是由于人类的活动还是地球气候的自然循环？这些问题都可以用统计学来解答，如同许多其他的统计观念，通过统计图表的直观呈现是最佳的一种方式。

关于全球变暖的科学研究最明显的出发点就是，全球变暖实际上发生了多少。首先你可以猜测，我们通过将报纸上报道的某个地方过去的气温和现在的温度进行比较，从而了解气温发展的简单趋势。尽管如此，全球变暖所指的是我们整体星球平均气温的上升。的确，即使将地球作为一个整体气温上升了，但我们仍然可以预计某些地方会变得越来越冷。如果我们想了解地球是否正在变暖，我们需要一些数据，可以告诉我们地球的平均气温（整个地球的平均气温）是怎样随时间变化的。

现在，同步卫星提供给我们一些数据，由于它可以做到对于全球的一个概览，这样就可以让我们非常准确地掌握好地球的平均气温。我们可以通过遍布在全世界的超过 7 000 多座气象站的“地球实况”测量，同时通过轮船进气阀收集的水温测量数据所得到的海洋气温，来证实这些卫星的记录。结果发现，近 40 年来可信的气温数据与卫星观测地球所得到的数据是大体相同的。但如此一来，在卫星时代之前的更多年以前的数据却多少有些不可信。因为我们只能看到某些特殊地区的数据，这些地区的数字对于我们进一步回顾过去的数个世纪的情况无疑显得太少了。此外，过去的气温记录更多是关于城市的，由于某些独立于全球变暖之外的原因，在这些地方气温会更暖和一些（参考 2.2 节的例 2-4）。科学家经常会考虑“城市热岛效应”，但即使那样，那些数据也只包括大陆的气温，很少有包括地球 3/4 面积的海洋数据记录。科学家在过去几年尽了很大的努力去详细探测过去的气温数据。通过数据的统计分析，重新建构过去两个世纪以来的一个气温历史是可能的，虽然当我们回顾过去的温度历史时，这种不确定性会变得更大。

图 3-60 呈现了自从 1860 年以来地球平均温度的历史再现。尽管数据有些不确定，但建立在这个基础上的结论是清晰的：在过去的一个世纪里，地球平均温度上升了 0.8℃（1.4 ℉）。更进一步说，大多数变暖的情况（大约 0.6℃）发生在过去的 30 年，这个数据非常可信。变暖的趋势看起来像在加速。例如，记录中最热的 10 年中其中 9 个发生在最近 10 年，这个可以在呈现的图中看到。

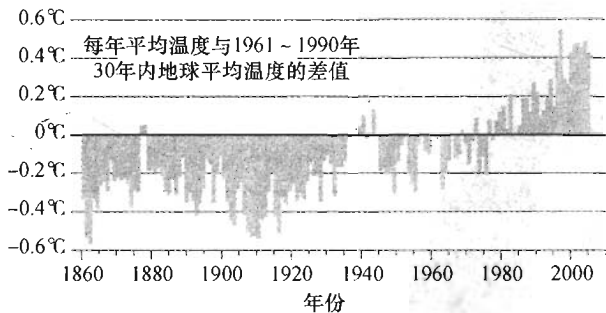


图 3-60 1860 年来每年的地球平均温度与 1961 ~ 1990 年这 30 年的地球平均气温的差值
（注意，在过去的几十年间地球变暖的趋势变得越来越清晰。）
资料来源：National Center for Atmospheric Research.

伴随着新数据显示了一个清楚的变暖趋势，那么下一个问题是我们预计未来的温度会增长多少。就像我们在后面第 7 章中将讨论的，气温的增长与大气中的二氧化碳浓度以及其他温室气体有关。这样，如果我们想预测未来气温将上升多少。我们需要最近的大气中二氧化碳浓度的上升数据。如同过去的二氧化碳浓度和气候温度变化的数据一样。

针对二氧化碳浓度近期的上升，从 1958 年以来，工作在莫纳罗亚山（夏威夷）的科学家搜集到的异乎寻常的数据。这些数据，呈现在图 3-61 的右边，代表了直接测量的二氧化碳浓度的变化。

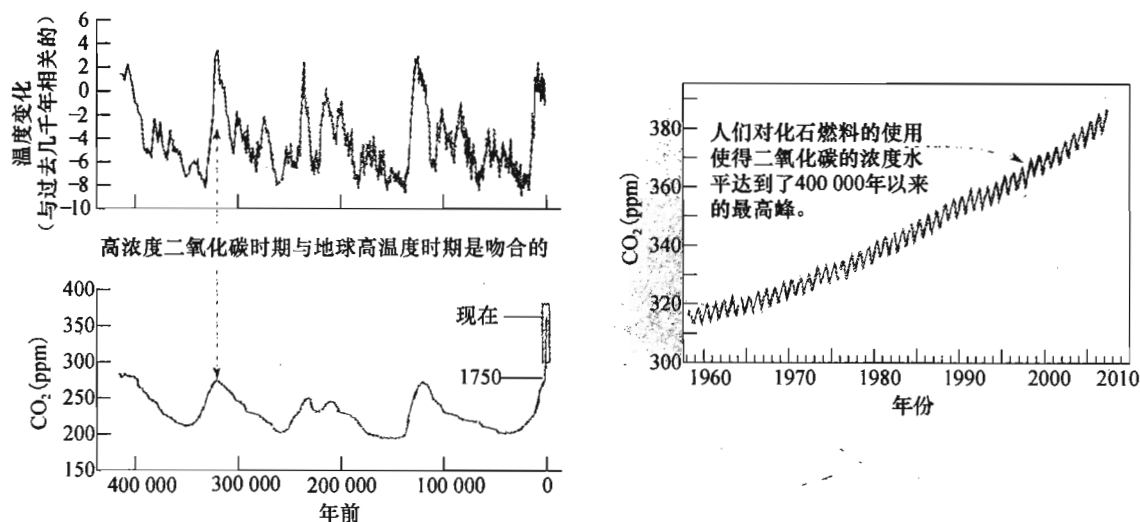


图 3-61 大气中的二氧化碳浓度和通过冰核及其他数据重新建构的过去 400 000 年间地球平均温度，右图中近期的二氧化碳浓度数据是来自在夏威夷莫纳罗亚山直接测量的结果

注意到二氧化碳浓度单位是百万分率 (ppm)。例如，320ppm 意味着空气中的每 100 万个分子中包含 320 个二氧化碳分子。每一年的图中位数据的上下波动呈现出二氧化碳浓度随着季节的变化。尽管存在着这些波动，但总体趋势是上升的。

为了解我们预期气温上升多少，我们想要一些过去的呈现了二氧化碳浓度和气温直接相关的数据。对于过去的几千年，科学家可以通过检测树木的年轮获得这些数据，通过研究这些数据可以揭示气温和大气中二氧化碳浓度的记录。通过钻探南极冰层，这个记录可以进一步扩展到对冰核的研究。冰核是由远古时代的雪压缩而成。因为下雪是有季节性的，每一小薄层冰代表一年。如同树的一个年轮代表树生长的一年。通过研究陷入每一层冰中的气泡，科学家能够再现气温和二氧化碳浓度过去的历史。(二氧化碳浓度能够通过测量气泡所得到，当气温较低时空气中较重的同位素不容易得到转化，所以较重的同位素相对比较轻的同位素被研究者用来估计远古时代的气温。)

结果是令人震惊的，当你看图 3-61 中的图形左侧下边时，冰核的数据提供了超过 400 000 年之间的气温和二氧化碳浓度的变化记录。这是地球走过冰河时代来到暖期的一段时间。当研究这个图形时至少可以得到三个结论。

1. 在气温和二氧化碳浓度之间存在一个相关：气候温度越高，二氧化碳的浓度越高。虽然这并不能证明二氧化碳浓度的提高导致了气候变暖，但是仍然能确定，这两者的变化趋势是一致的。这就意味着现在二氧化碳浓度的提高将会伴随着相应的气候温度的升高，这种联系是有意义的。

2. 气温和二氧化碳浓度随着时间会产生大幅度而且自然的变化。地球平均气温在最近的数十万年已经数次上下波动超过 10℃ (大约 18 ℉)，二氧化碳浓度也在 200 ~ 300ppm 之间波动。

3. 当将图形右边近期的二氧化碳数据与左边的历史数据进行比较时，发现我们完全进入了一个新的区域：现在的二氧化碳浓度远大于过去 40 万年之间我们地球上的任何时期的水平。的确，现在的数据已经初步扩展到通过钻探冰核掌握的超过 100 万年前的数据，因此我们也发现了同样的观点：自从进入工业化时代以来，我们已经大幅度地提升了地球大气层中的二氧化碳浓度，现在二氧化碳浓度水平远超过暖期以及过去 100 万年间冰河时代的任何时候的水平。

即使没有对地球气候未来预期变化的精确估计，这些数据也足以提供给我们一个清醒的信息。冰河时代大部分的美国区域被冰川覆盖，现在则是相当温暖的，气温只有几摄氏度。并且冰核的数据告诉我们，在过去曾经发生过随着二氧化碳浓度的变化，尽管这个浓度变化比我们今天所发生的浓度变化还小，但也引起了巨大的气候温度的变化。如果过去的发展趋势象征着我们所预计的未来的话，那么，我们的子孙将会生活在一个气候完全不同于我们今天的时代里。

问题讨论

1. 仔细地研究图 3-61，今天的二氧化碳浓度与 1750 年如何进行比较？1750 年的二氧化碳浓度如何与过去的

400 000年进行比较？从这些问题的回答中你可以得到什么结论？

2. 讨论一些影响未来大气中二氧化碳浓度的因素。你认为应该怎样做才可能减缓或阻止二氧化碳浓度的升高？
3. 如果全球变暖的话，会出现什么结果？讨论你从新闻报道中听到过什么，考虑哪一种结果看起来极度悲观和让人惊恐。
4. 仅在10年以前，关于全球变暖是否真的会发生，并且问题是否严重仍然存在广泛的争论。今天，这个争论基本上结束了。讨论一下新的数据在澄清全球变暖的风险这个问题上的作用，清楚的图形对我们理解这些新的数据起到了重要作用。

建议阅读

1. 你可以找到大量科学网站，上面有最近的全球变暖数据的更新。一些非常优秀的网站包括政府间气候变化专门委员会（IPCC）、美国大气研究中心（NCAR）、美国环境保护局（EPA）。
2. Bennett, Jeffrey, and Shostak, Seth. *Life in the Universe*, 2nd ed. Addison Wesley, 2007 (see Section 10.5).
3. Flannery, Tim. *The Weather Makers*. Atlantic Monthly Press, 2005.



试图对人们进行概括是没用的，一个人必须接受的指示，既不是他全部所说的，也不是他全部所做的。

——弗吉尼亚·伍尔夫

第 4 章 数据的描述

学习目标

■ 4.1 什么是集中量数

理解平均数、中位数和众数之间的差别，每一个都是如何受异常值影响的。理解容易混淆的不同类型的“集中量数”，如何正确使用加权平均数。

■ 4.2 分布形态

应用众数、峰度和偏度等数值来描述常见的分布形态。

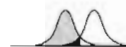
■ 4.3 差异量数

理解和解释常见的差异量数：全距、五数概括法[最小值、第一四分位数 (Q1)、中位数 (Q2)、第三四分位数 (Q3)、最大值] 和标准差。

■ 4.4 统计悖论

研究几个常见的统计悖论，比如通不过“90% 准确性”测谎器实验的，他们中的大多数人事事实上说的是真话。

在第 3 章，我们讨论了用图表呈现数据分布的方法。现在我们开始学习，怎样用比率和数字来概括分布的特征。我们将讨论最常用的描述数据的集中、形态、离散等的方法，就像你所看到的，这些方法是数据分析的重点，也会应用于你在新闻中见过的几乎每个统计研究中。在本章的结尾，我们分析几个即使你仔细地查看数据，也可能会偶然产生的令人惊奇的情况。



4.1 什么是集中量数

术语“平均水平”经常出现于新闻和其他报导中，但并不总是同一个意思。就像你在本节所看到的，最恰当的平均水平的定义有赖于情境。

4.1.1 平均数、中位数和众数

表 4-1 呈现了根据 5 部畅销科幻小说而各自拍摄的影片数量（原作、续集和前传）。这一系列影片的平均数是多少？回答这个问题的一个

办法是计算平均数 (mean)，平均数的正式名称是算术平均数，通常简称均值。我们通过将影片总数除以 5（因为所列出的数据是 5 部序列电影）得到均值。

$$\text{均值} = (4 + 3 + 6 + 10 + 6) / 5 = 5.8$$

表 4-1 五部科幻电影序列

系列名称	影片数量 (截至 2007 年)
异形	4
回到未来	3
人猿星球	6
星际争霸	10
星球大战	6

换句话说，这5种序列电影平均每种拍摄了5.8部影片，更进一步，我们得出任何一组的数据均值是所有数据值的总和除以数据个数。大多数人认为平均水平就是均值。从本质上而言，均值代表了一个定量数据分布的平衡点，如图4-1所示。

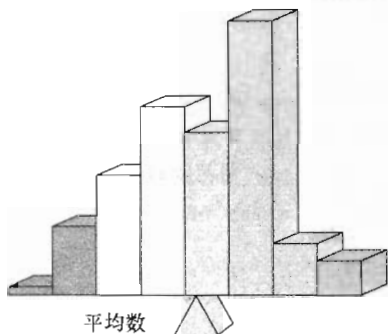


图4-1 在均值位置平衡的柱状直方图

我们也可以通过计算一组数据的中位数来描述影片数量的平均水平，我们将数据从小到大（或者从大到小）排序，出现超过一次的数据将其重复排列，如果数据的个数是单数，序列中最中间的那个数值就是中位数（median），如果数据的个数是偶数，在序列正中间有两个数，这两个数和的一半就是中位数，将表4-1中的数据以升序排列，结果是3, 4, 6, 6, 10，因为序列中间的数值是6，所以影片数量的中位数就是6。

众数（mode）是一组数据中出现最多的一个或者几个数值，在影片数量的例子中，众数是6，因为在这组数据中，6出现了两次，其他数值仅仅只有一次。一组数据可以有一个众数，也可以有多个众数，或者没有众数。有时，众数指的是一组相连的间断数值而不是某个单个数值。众数更常用于定性数据而非定量数据，而均值和中位数更多地用于定量数据。

定义

一个分布中的集中量数

- ◆ 平均数经常也被称做均值，它通过下面的计算得出：

$$\text{平均数} = \frac{\text{全部数据值总和}}{\text{总的数据次数}} \quad (4-1)$$

- ◆ 中位数是指一组排序数据正中间位置的数值（当数据个数是偶数中位数是中间两个数值和的一半），简称中数。
- ◆ 众数是一组数据中出现次数最多的数值（或者几个数值）。

当取整数时，我们将对本章所讨论的计算应用以下规则。

统计计算的取整规则

陈述答案时精确度要比原始数据多保留一位小数。例如：2, 3 和 5 的平均数是 3.333 3...，我们最后保留结果为 3.3，因为原始数据是整数，我们保留位数到 0.1 水平，作为一般的处理方式，只是在得出最后答案采用四舍五入，而并不能用于中间计算过程。

注意在影片数量的例子中我们应用了这样的规则，表中的原始数据全是整数，而表述的平均数是 5.8。

方法注解

如果集中量数的值和原始数据有着相同位数的有意义数字，你可以使用 0 向下扩展一位小数位数或者使用没有扩展小数位数的准确结果。例如，2 和 4 的平均数可以表达成 3 或者 3.0 都可以。

例 4-1 价格数据

8 家食品杂货店分别以下面的价格来出售 PR 能量棒（美元）：

1.09 1.29 1.29 1.35
1.39 1.49 1.59 1.79

计算平均数、中位数和众数。

解答 平均价格是 1.41：

$$\begin{aligned} \text{平均数} &= (1.09 + 1.29 + 1.29 + 1.35 + 1.39 \\ &\quad + 1.49 + 1.59 + 1.79) / 8 = 1.41 \end{aligned}$$

为了得到中位数，我们首先将数据升序排列：

1.09, 1.29, 1.29, 1.35, 1.39, 1.49, 1.59, 1.79

因为一共有 8 个价格（一个重复数值），在序列中间有两个数值：一个是 1.35，另一个是 1.39。中位数是序列中间两个数和的一半，我们将这两个数加起来除以 2 来计算中位数：

$$\text{中位数} = (1.35 + 1.39) / 2 = 1.37$$

应用保留位数规则，我们表述平均数为 1.410，中位数为 1.370，众数为 1.29，因为这个价格出现的次数比其他价格多。

相关知识

美国芝加哥大学的一项研究发现青少年的零用钱的中位数是 50 美元/周，这个研究估计青少年每周接受的零用钱超过了 10 亿美元。

例 4-2 海洋

表 4-2 列出了世界主要海洋的面积（平方千米），计算这些海洋面积的平均数、中位数和众数。

解答 仔细核对一下，你可以计算表中列出的世界 17 个海洋的面积总和是 343 408 600 平方千米。我们将总和除以 17 得出平均数：

$$\begin{aligned}\text{平均数} &= 343\,408\,600 \text{ 平方千米} / 17 \\ &= 22\,893\,907 \text{ 平方千米}\end{aligned}$$

表 4-2 已经按降序排列了，这样很容易计算中位数，

表中列出了 15 个项目，第 8 个数据值就是中位数，因为 7 个数值在它之上，7 个数值在它之下。表中第 8 个数值是墨西哥湾，因此中位数是 1 592 800 平方千米。

这些数据中没有众数，因为没有有一个值出现的次数多于其他数值，尽管如此，如果我们从数据组来看，15 个海洋中的 4 个面积在 100 万 ~ 200 万平方千米，因此我们可以说最常见的海洋面积在 100 万 ~ 200 万平方千米。 ■

表 4-2 世界主要海洋面积

洋/海	面积（平方千米）	洋/海	面积（平方千米）	洋/海	面积（平方千米）
太平洋	165 760 000	加勒比海	2 718 000	日本海	1 007 800
大西洋	82 400 000	白令海	2 291 900	安达曼海	797 700
印度洋	65 527 000	墨西哥湾	1 592 800	北海	575 200
北冰洋	14 090 000	鄂霍次克海	1 589 700	红海	438 000
地中海	2 965 800	哈得逊湾	1 232 300	波罗的海	422 200

资料来源：TIME Almanac.

4.1.2 异常值效应

浏览一下平均数、中位数和众数之间的差异，设想 5 个某大学篮球队里的大学四年级学生，收到的全美职业篮球联赛（NBA）第一年打球的签约合同价值如下（0 意味着运动员没有接到签约合同）。

0 0 0 0 3 500 000 美元

签约合同的平均数是

$$\begin{aligned}\text{平均数} &= (0 + 0 + 0 + 0 + 3\,500\,000 \text{ 美元}) / 5 \\ &= 700\,000 \text{ 美元}\end{aligned}$$

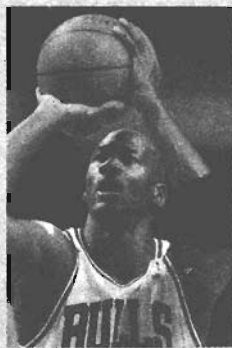
大学篮球队里的四年级学生收到 NBA 的签约合同平均数为 700 000 美元，这样说合适吗？

当然不合适。这个问题在于单个的运动员收到了比其他人多出很多的最大一笔价值的签约合同，如果我们忽略这个运动员只看其余的 4 个人，签约合同价值的平均数就是 0，因为这一笔签约合同价值为 3 500 000 美元，与其他人相比差异极大，我们称之为异常值（outlier），或者叫孤立值（outlying value），如我们例子中所呈现的，一个异常值会使平均数显著地增大（或者减小），这样使得平均数不能很好地代表全部数据。

相关知识

一个研究曾经发现南卡罗来纳州大学地理专业学生的平均起薪要高于其他大学地理专

业的同学。原来这个高平均数的原因来自一个异常值，一位来自南卡罗来纳州大学地理专业的篮球超级明星乔丹。



定义

◆ **异常值**是指一组数据中比绝大多数其他数据值大得多或小得多的数据值。

当异常值使得平均签约合同价值明显增高，但它对中位数却没有影响，5 个人的合同值的中位数为 0。通常，异常值对中位数没有影响，因为异常值不改变一组数据中间位置的数值。当然，异常值也不影响众数（当然，如果删除异常值，中位数会改变，因为改变了一组数据的个数）。表 4-3 是平均数、中位数和众数的特征概括，包括异常值对每一种集中量数的影响。

表 4-3 平均数、中位数和众数的比较

量数	定义	是否常用	个数	是否所有数据值参与运算	是否受异常值影响	优点
平均数	数据值总和/数据总个数	大多数	1	是	是	容易理解；在许多统计方法中应用良好
中位数	中间值	经常	1	否	否	在有异常值时，比平均数有更好的代表性
众数	次数最多的值	偶尔	0, 1 或多个	否	否	适用于定性数据（见 2.1 节）

思考题

对于 5 个人签约合同价值，使用中位数来代替平均数合适吗？为什么？

在统计中如何控制异常值是非常重要的问题。有时，就如同篮球队员的例子，必须要掌握一个异常值是合理的数据，这样才能正确解释平均数和中位数，而有些时候，在一组数据中异常值可能是差错，判断一个异常值是合理而重要的，还是仅仅只是一个简单的差错，这是非常不容易的。

例 4-3 差错

一位田径教练想知道她的运动员在训练时适宜的心率，她选择了她的 5 位运动员，要求他们在训练时佩戴上心率监视仪，在训练中间，她读出 5 位运动员以下的心率值：130, 135, 140, 145, 325，在这个情况下，平均数还是中位数是更合适的集中量数？为什么？

解答 5 个心率值中的 4 个是互相接近的，看起来是合理的训练当中的心率水平，最大值 325 是一个异常值，这个异常值看起来像是一个差错（很可能是心率监视仪的原因），因为任何一个人如果这么高的心率的话会引起心脏骤停的，如果教练使用平均数，她要包括这个异常值，这意味着这里面包含记录错误，如果她使用中位数作为集中量数，这会是一个更合理的数值，因为中位数不受异常值得影响。

4.1.3 集中量数的混淆

“平均水平”的不同含义能够引起混淆，有时，因为我们不知道“平均水平”是指平均数、中位数，还是众数，如果怎样计算得出结果的信息不够充分，这种混淆就会增强。下面的例子会表明这种情况。

例 4-4 薪水争议

一家报纸调查了高科技公司工人的薪资情况，报告说每小时平均薪资为 22 美元。一家大公司的工人

们立刻提出加薪的要求，说他们和其他公司的人们一样努力工作，但是平均薪资仅是每小时 19 美元，经理反驳了他们的要求，说他们的工资是高于平均水平的，他们实际上的平均薪资是每小时 23 美元？谁说的正确？请解释。

解答 两方面都是正确的，只是他们使用了不同的集中量数，在这个案例中，工人可能使用了中位数，而经理使用的是平均数。例如，假设这家公司仅有 5 名工人，他们的薪资（美元）分别是 19、19、19、19 和 39，5 个人薪资的中位数是 19（如工人们宣称的），但是平均数是 23（如同经理所宣称的）。

例 4-5 哪一种集中量数

在一个很小的大学里，总共有 100 名一年级大学生，开设 3 门核心研究课程。两门课采用大型讲座的形式，所有 100 名学生都在一个班级，第 3 门课分成 10 个小班，每班 10 人。学生和教授关于班容量是否过大发生了分歧。学生们说核心研究课程的平均班容量是 70，而教授说平均班容量仅为 25，这两种说法谁的正确？请说明。

解答 两种说法都是正确的，但是采用不同的平均数。学生计算的班容量平均数是基于个人所听课的班级，每一个学生听课的班级，分别为两个班的容量都是 100，另一个班的容量为 10，所以每一个学生所听课的平均班容量为

$$\begin{aligned} & \text{学生听课的总人数} / \text{所听课的班数} \\ &= (100 + 100 + 10) / 3 = 70 \end{aligned}$$

教授计算的平均班容量是基于所有班级的，两个班各是 100 名学生，另有 10 个班各是 10 名学生，学生的总人数是 300，而班级总数是 12，平均班容量就是

$$\text{听课学生注册数} / \text{开班数} = 300 / 12 = 25$$

两种关于平均班容量的说法都是正确的，但是学生和教授所说的平均数是非常不一样的，学生计算的班容量是针对每个学生而言的，而教授所说的班容量是针对每班而言的。

思考题

在例4-5中,教授怎么样重新安排才能使所有的班都是25个学生?讨论这样一个变化的利弊。

4.1.4 加权平均数

假定你某门课程的成绩基于4次小测验和期末测验考试,每次小测验占15%的比例,期末考试占40%比例,你的4次小测验分数分别是75,80,84和88,期末考试分数为96,你的总评成绩是多少?

因为期末考试所占比例大于平时小测验,5个分数的简单的平均数并不能代表总评成绩。在这里,我们必须对每一个分数进行加权(指出其相对重要性)。在本例中,我们给予每次小测验分数的权重分别是15%,给予期末考试分数的权重是40%,最后我们将每个分数和它对应权重的乘积相加,然后除以权重之和,从而得出加权平均数:

$$\begin{aligned}\text{加权平均数} &= (75 \times 15 + 80 \times 15 + 84 \times 15 + 88 \times 15 \\ &\quad + 96 \times 40) / (15 + 15 + 15 + 15 + 40) \\ &= 8745 / 100 = 87.45\end{aligned}$$

由于正确地考虑了小测验和期末考试的不同权重,因此加权平均数是87.45。依照保留位数的规则,我们取总评分数为87.5。

当不同数据值的重要性不一致时,加权平均数是适宜的,使用下面的公式我们可以计算出加权平均数。

定义

加权平均数考虑不同数据值的相对重要性,每一个数据值都被赋予相应的权重,加权平均数计算如下:

$$\text{加权平均数} = (\text{每一个数据值} \times \text{它的权重}) \text{的总和} / \text{所有权重之和} \quad (4-2)$$

相关知识

对于许多不同类别的运动员或者球队,根据其运动成绩所进行的运动统计通常使用加权平均数。例如棒球运动的投手防御率(ERA)和长打率,橄榄球中的四分卫率及其排名。

思考题

因为在课程成绩的例子中权重是百分数,我们也可以认为权重是0.15和0.40,而不是15和40,我们使用0.15和0.40作为权重计算加权平均数,结果是否一样?为什么?

例4-6 平均绩点

兰德尔有38学分成绩为A,22学分成绩为B,7学分成绩为C,他的平均绩点(GPA)是多少?成绩A对应的绩点是4.0,成绩B对应的绩点是3.0,成绩C对应的绩点是2.0。

解答 成绩A、B、C对应的绩点是4.0,3.0和2.0,学分数就是权重。数值4对应的权重是38,B代表的数值3对应的权重是22,C代表的数值2对应的权重是7,加权平均数就是

$$\begin{aligned}\text{加权平均数} &= (4 \times 38 + 3 \times 22 + 2 \times 7) / (38 + 22 + 7) \\ &= 232 / 67 = 3.46\end{aligned}$$

依照保留位数的规则,兰德尔的平均绩点(GPA)约等于3.5(3.46的四舍五入)。

例4-7 股权投资

公司选举投票经常要考虑每一个股票持有者,假定一个公司由5个股东投票决定,公司是否举办一次新的广告活动。投票如下(Y=同意,N=不同意)

股东	持有股票数	投票
A	225	Y
B	170	Y
C	275	Y
D	500	N
E	90	N

按照公司章程,做出决策需要60%的股权通过。这个提议能通过吗?

解答 我们可以将Y视做数值1,N视做数值0,股票数量就是股东投票的权重,股东A投票代表数值为1,权重为225;股东B投票代表数值为1,权重为170,依次类推。投票结果的加权平均数是:

$$\begin{aligned}\text{加权平均数} &= (1 \times 225 + 1 \times 170 + 1 \times 275 + 0 \times 500 \\ &\quad + 0 \times 90) / (225 + 170 + 275 + 500 + 90) \\ &= 670 / 1260 \approx 0.53\end{aligned}$$

加权平均数是53%(或者0.53),少于需要的60%,提议不能通过。

4.1.5 带累加和和符号的平均数(可节选)

许多统计学公式,包括平均数公式,可以

更简约地用叫做求和符号的数学符号来表达。符号 Σ （希腊字母西格玛）叫做累加符号，是指一组数据依次加起来。我们用符号 x 表示一组数据中的每个数值，这样我们将所有数据值的和写成：

$$\text{所有数据值的和} = \Sigma x \quad (4-3)$$

例如，如果一个样本包括 25 个测验分数， Σx 表示所有 25 个分数之和。同样，如果一个样本包括 10 000 个家庭的收入， Σx 表示 10 000 个家庭的收入之总和。

我们以 n 表示样本中的数据个数，这样，平均数的公式一般写成：

$$\begin{aligned} \bar{x} &= \text{样本平均数} = \text{数据值总和} / \text{数据总次数} \\ &= \frac{\Sigma x}{n} \end{aligned} \quad (4-4)$$

符号 $\bar{x}$ 是样本平均数的标准符号，当涉及一个总体平均数而不是样本时，使用希腊字母 μ （读作“谬”）作总体平均数的符号。

求和符号使加权平均数的公式表述起来也变得容易，我们还是使用 x 表示每一个数值，使用 ω 表示每一个数值的权重，每个数据值和它所对应的权重值的乘积的总和表示为 $\Sigma x \times \omega$ ，权重之和表示为 $\Sigma \omega$ ，这样的话，加权平均数的公式就成为：

$$\text{加权平均数} = \frac{\Sigma x \times \omega}{\Sigma \omega} \quad (4-5)$$

方法注解

累加经常往往使用指针的形式来书写，用于说明如何一步步进行求和。例如，符号 x_i ，指的是数据组中第 i 个数据，字母 i 就是

指针，我们将所有的总和写作：

$$\sum_{i=1}^n x_i$$

读作“开始于 $i=1$ ，结束于 $i=n$ ， n 是这组数据的总数”，采用这样的符号，平均数写作：

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

4.1.6 分组数据的平均数和中位数（可节选）

这一节的主要内容是简要说明分组数据，在分组数据中，以每一组的组中值来代表本组每个数据。例如，下表是分组后的 50 个数据。

组限	次数
0 ~ 6	10
7 ~ 13	10
14 ~ 20	10
21 ~ 27	20

第一组的组中值是 3，我们假定 3 这个值出现了 10 次，同理，我们可以得出表中 50 个数据的总和是：

$$(3 \times 10) + (10 \times 10) + (17 \times 10) + (24 \times 20) = 780$$

这样，平均数就是 $780/50 = 15.6$ 。一共 50 个数据，中位数就是第 25 个和第 26 个数据和的一半，这两个数据都落在 14 ~ 20 之间，所以我们称这一组叫做中位数所在组（median class）。众数在次数最多的一组，在本例中也就是 21 ~ 27 这一组。

4.1 练习题

统计素养和批判性思维

1. 异常值 什么是异常值？异常值是怎样被精确定义的从而使它们可以被清楚和客观地识别出来？请加以解释。
2. 平均数和中位数 一个统计课的班级包括 24 名学生，他们中的一部分受雇承担了工资较低的钟点工，另一部分则没有工作，班级中也包括薪水非常高的一位教授，描述班级里 25 个人（包括教授）收入的方法是哪一种：平均数还是中位数？为什么？
3. 平均通勤时间 一位社会学家想计算所有美

国居民从家到工作地点上下班途中的平均通勤时间。他知道调查数百万工作人们中的每一个是不可行的，所以他进行了一个网络搜索寻找到 50 个州中的每个州的平均通勤时间。她将这 50 个通勤时间累加起来并且除以 50。这个结果对于全部在美国工作中的人们而言，是一个平均通勤时间好的估计吗？为什么？

4. 称名数据 在第 2 章里，我们注意到称名测量水平的数据只包括姓名或者标签。一位新英格兰爱国者橄榄球队的球迷，记录了一场

超级碗比赛中每一位爱国者队球员运动服的编号, 计算这些数字的平均数有意义吗? 为什么?

判断正误: 判断练习题 5~8 中的陈述是有意义的 (表述正确) 还是无意义的 (表述错误), 并做出解释。并非所有的陈述都有明确的答案, 所以你的解释比判断更重要。

5. 平均数 一组数据值它们的平均数是 65.2 和 72.3。
6. 众数 一组数据值它们的众数是 65.2 和 72.3。
7. 平均数、中位数和众数 一位研究者计算一组数据的结果是, 平均数、中位数和众数的值完全相同。
8. 加权平均数 研究者计算了一组数据平均数, 并且编制了次数分布表, 然后又计算次数分布表的平均数。研究者得出结论说, 这里面有误差, 因为得到了两个不同的结果。

恰当的集中量数: 练习题 9~12 列出了我们可能想知道的一些集中量数。在每一种情况下, 说明这个平均数或者中位数是不是最恰当的集中量数, 并说明你的理由。

9. 收入 一个大城市里所有成人的平均收入。
10. 重量 一个大箱子里所有柑橘的平均重量。
11. 工作变化 在人们的职业生涯中平均的工作变动次数。
12. 丢失的行李 每一家航空公司的一次飞行中平均丢失行李的数量。

概念与应用

平均数、中位数和众数: 练习题 13~20, 每一个题都列出一组数字。在每道题中, 计算所列出数字的平均数、中位数和众数。

13. 知觉反应时 统计课的学生们参与了一个测验, 让他们判断 1 分钟 (60 秒) 时间是多长, 所记录下来他们所反应的实际时间。

53 52 75 62 68 58 49 49

14. 身体温度 随机选择的正常和健康的成年人的身体温度 (华氏温度)。

98.6 98.6 98.0 98.0 99.0

98.4 98.4 98.4 98.4 98.6

15. 血液中酒精浓度 下列是驾驶员血液中的酒精浓度过高导致机毁人亡, 并且被判入狱服刑。(数据来源于美国司法部)

0.27 0.17 0.17 0.16 0.13 0.24

0.29 0.24 0.14 0.16 0.12 0.16

16. 加州老实泉 美国黄石国家公园老实泉两次喷射之间的时间间距 (以分钟计)。

98 92 95 87 96 90

65 92 95 93 98 94

17. 果蝇 一个雄果蝇样本的胸腔宽度 (以毫米计)。

0.72 0.90 0.84 0.68 0.84 0.90

0.92 0.84 0.64 0.84 0.76

18. 美国总统的年龄 下面是选取的一些美国总统就任时的年龄。

57 61 57 57 58 57 61 54

68 51 49 64 50 48 65

19. 糖果的重量 下面是随机挑选的公司筒装糖果的重量 (以克计)。

0.957 0.912 0.842 0.925 0.939 0.886

0.914 0.913 0.958 0.947 0.920

20. 25 美分的硬币 下面是随机挑选的 25 美分的硬币重量 (以克计)。

5.60 5.63 5.58 5.56 5.66 5.58

5.57 5.59 5.67 5.61 5.84 5.73

5.53 5.58 5.52 5.65 5.57 5.71

5.59 5.53 5.63 5.68

21. 按字母排序的各州 下表给出了以字母排序的, 以字母 A~C 开始的几个州的总面积 (陆地和水面)。

州	面积/平方千米
亚拉巴马	52 200
阿拉斯加	615 200
亚利桑那	114 000
阿肯色	53 200
加利福尼亚	158 900
科罗拉多	104 100
康涅狄格	5 500

- (1) 计算这些州面积的平均数和中位数。
- (2) 在数据的大数段哪一个州的面积是异常值? 如果你去除这个州的面积, 那么这组数据的新平均数和中位数是多少?
- (3) 在数据的小数端哪一个州的面积是异常值? 如果你去除这个州的面积, 那么这组数据的新平均数和中位数是多少?

22. 可口可乐的异常值 一听可口可乐的重量变化非常微小。这里有 7 听可口可乐的测量重量(磅)。
- 0.816 1 0.819 4 0.816 5 0.817 6
0.790 1 0.814 3 0.812 6
- (1) 计算这些重量的平均数和中位数。
- (2) 你认为哪一个重量值可能是异常值？如果有，并加以解释。
- (3) 如果将异常值排除在外的话，这些重量的平均数和中位数是多少？
23. 提高你的分数 假如你在一个数学课堂上的几次小测验中分别得的分数是 70, 75, 80, 70。
- (1) 这些分数的平均数是多少？
- (2) 如果你最后的总平均成绩要达到 75 分的话，那么在下次小测验中你要得多少分？
- (3) 如果小测验的最大分数是 100 分，在第 5 次小测验以后你的平均成绩有可能达到 80 分吗？请说明理由。
24. 提高你的成绩 假如你在一个社会学课堂里的几次小测验中得到的分数分别是 60, 70, 65, 85, 85。
- (1) 这些分数的平均分是多少？
- (2) 如果使最后的总平均成绩要达到 75 分，你下次测验成绩需要达到多少分？
- (3) 如果小测验的最大分数是 100 分，在下次小测验以后你的平均成绩最多可能是多少分？请说明理由。
25. 新的平均数 假如你 6 次小测验以后，你的平均分数是 80 (不可能超过 100 分)。如果你在下次小测验中得到 90 分，你新的平均数是多少？在下次小测验以后，你可能的最大平均分是多少？在下次小测验以后，你可能的最小平均分是多少？
26. 新的安打率 假如一个棒球运动员在 30 次击球以后，平均的安打率是 0.300。如果她在下次击球时得到一个安打，她的新的安打率是多少？(安打率就是安打的次数除以总的击球数。)
27. 平均数比较 假如某地区的教育官员宣称这个地区四年级学生的平均阅读分数是 73 (最高 100 分)。从理论上来说，你知道你的一些四年级学生有如下的成绩：55, 60, 68, 70, 87, 88, 95。你能否说你的学生分数超过了地区平均成绩？并加以解释。
28. 平均数比较 全美职业篮球联盟 (NBA) 报告了篮球运动员的平均身高是 6.8 英尺，作为一个教练，你知道你的球队始发阵容的身高分别是 6.5 英尺、6.6 英尺、6.6 英尺、7.0 英尺、7.2 英尺。你能否声称你球队的始发阵容平均身高超过 NBA 的平均身高。
29. 桃子的平均数 一个杂货商有 3 筐桃子：1 筐有 50 个桃子，重 18 磅；1 筐有 55 个桃子，重 22 磅；1 筐有 60 个桃子，重 24 磅。所有桃子加在一起的平均重量是多少？请加解释。
30. 平均数的困惑 一位指导老师的第一期学生有 25 个，他们在期中考试中的平均分数是 86%；第二期学生有 30 个，他们在同样的期中考试中的平均分数是 84%。那么两个班合在一起的平均分数是不是 85%？为什么？
31. 哪一个平均数 在一所高中里所有的 300 名学生开设了同样的 4 门课，其中的 3 门课被分成 15 个教学班，每班 20 名学生，第 4 门课被分成 3 个教学班，每班 100 名学生。请解释一下，为什么按照学校的说法，平均班容量是 25 名学生，可是恼怒的家长却说每个班容量是 40 个学生。你认为哪一个平均数对于班容量而言是比较公正的描述呢？
32. 最终成绩 你的一门课程的最终成绩是基于以下 4 个方面来计算的，期中测验占最终成绩的 15%，一个课程项目占最终成绩的 20%，一项家庭设计作业占最终成绩的 40%，期末考试占最终成绩的 25%。你的期中测验成绩是 75 分，课程项目是 90 分，家庭作业分数是 85 分，期末考试成绩是 72 分。你的课程最终成绩是多少？
33. 安打率 回顾一下安打率的计算方法，就是将全部击球成功的次数除以击球的总次数 (不计走步、失误以及一些其他特殊情况)。一个运动员在第一节比赛中的安打率是 $\frac{2}{4}$ (在 4 次击球中成功两次)，在第二节比赛中的安打率是 $\frac{0}{3}$ ，在第三节比赛中的安打率是 $\frac{3}{5}$ 。他在这场比赛中的安打率是多少？它是以什么方式计算的平均数？

34. 安打率的平均数 假如一个运动员在许多场比赛中的安打率超过 0.400, 在下一场比赛中, 他的安打率是 $2/4$, 也就是在这场比赛中他的安打率是 0.500, 那么他的新的安打率是否可以做如下的计算: $(0.400 + 0.500)/2 = 0.450$? 请加以解释。
35. 鸡蛋的百分数 一位农场技术员发现一家农场的鸡蛋 8% 被检验出含有沙门氏菌, 而另一家农场的鸡蛋 12% 被检验出含有沙门氏菌。是否可以这样计算, 两家农场的鸡蛋中总共检出含有沙门氏菌的比例是 10%? 为什么?
36. 强打率 除了安打率之外, 另外一个衡量棒球运动员技术水平的测量指标是强打率。在计算强打率的方法中, 一垒安打价值为 1 点, 二垒安打价值为 2 点, 三垒安打价值为 3 点, 本垒打为 4 点。一个运动员的强打率是指它总共的得分点除以所有击球的总数 (不计走步、失误以及一些其他特殊情况)。一个棒球运动员在第一节比赛里在 5 次击球中得到 3 个一垒安打, 在第 2 节比赛里 4 次击球中得到 1 个三垒安打和 1 个一垒安打, 在第 3 节比赛里 5 个击球中得到 1 个二垒安打和 1 个本垒打。
- (1) 这名运动员的安打率是多少?
 - (2) 这名运动员的强打率是多少?
 - (3) 强打率有可能超过 1 吗? 请加以解释?
37. 股票拥有者的投票表决权 某公司有 4 个股东。第 1 个股东拥有 400 份股票, 第 2 个股东拥有 300 份股票, 第 3 个股东拥有 200 份股票, 第 4 个股东拥有 100 份股票。他们投票表决一个新的广告方案。第 1 个股东投赞成票, 其他 3 个股东投反对票。解释投票的结果怎样用一个加权平均数来表达。这个投票的结果是什么?

38. GPA 一种常用的计算成绩等级的方法就是平均绩点, A 被设置为 4 点, B 被试为 3 点, C 被设置为 2 点, D 被设置为 1 点。如果一个学生 GPA 的成绩中 1 门 5 学分制课程是 A, 另外 3 门 3 学分制课程成绩分别为 B、C、D。
39. 豌豆的性状 进行了一项试验以确定土壤中二氧化碳量不足是否会影响豌豆的性状。下面列出了豌豆性状的编号。1 = 表皮光滑且是黄色, 2 = 绿色表皮光滑, 3 = 黄色表皮有皱折, 4 = 绿色表皮有皱折。可以通过测量得到以下这些数值吗? 这个结果有意义吗?
- | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 1 | 2 | 2 | 1 | 2 |
| 3 | 3 | 2 | 3 | 1 | 3 | 1 | 3 | 1 | 3 | 2 | 2 | |
40. 美国人口中心 设想有一张巨大的平面美国地图, 将一些标志放置在地图上以表示居住人口, 使地图保持平衡的那一个点就叫做人口的平均中心。图 4-2 呈现了 1790 ~ 1990 年人口平均中心位置发生的移动。简要说明地图上所呈现的图形有一个什么样的模式。

网络和远程项目

请在网站 [www. aw. com/bbt](http://www.aw.com/bbt) 上打开第 4 章的有用链接 “links for internet projects”。

41. 薪资数据 许多网页提供了不同职业的薪资数据, 寻找你所关注的薪资数据。这种职业的薪资平均数和中位数是多少? 将这个职业的工资与其他你感兴趣的职业的工资怎样进行比较?

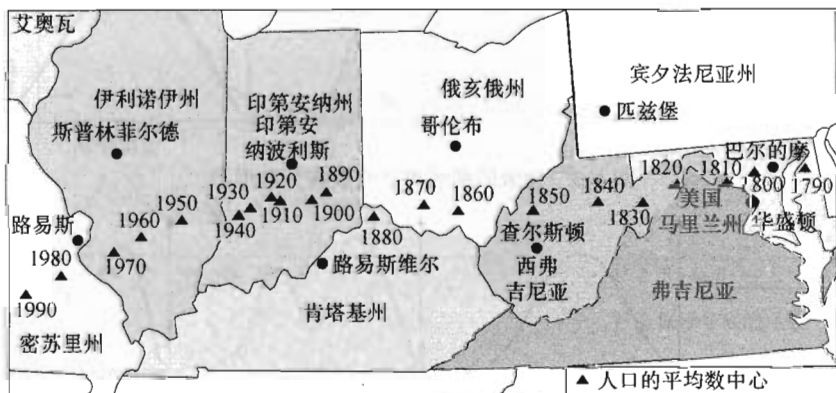


图 4-2 美国人口的平均中心

资料来源: Statistical Abstract of the United States.

42. 中位数是信息吗？阅读这篇文章《中位数不是消息》，作者是生物学家史蒂芬·杰·古尔德，文章可以在网页中找到。写一小段话描述古尔德试图让大家理解的信息是什么。对于其他被诊断为癌症的患者来说，为什么这个信息是重要的？
43. 肚脐数据 将 3.1 节练习题 23 中所收集的数据进行分组，并绘制分布的直方图。分布的平均数是多少？分布的中位数是多少？一个古老的理论说，就平均而言，一个人的肚脐比是符合黄金分割率的： $(1 + \sqrt{5})/2$ 。依据你的观察这个理论看起来准确吗？

新闻中的统计学

44. 日常生活中的平均数 引用 3 个你处理自己日常生活的集中量数的例子（比如平均成绩和安打率）。在每个例子里，说明所使用的集中量数是平均数、中位数还是其他类型的集中量数？简要描述这个集中量数对你来说是怎样起作用的。
45. 新闻中的统计学 寻找三篇近期引用了某些集中量数的新闻报道。说明其中所使用的集中量数是平均数、中位数还是其他类型的集中量数。



4.2 分布形态

到目前为止，本章我们讨论了如何描述定量数据分布的集中趋势，比如计算平均数、中位数等。现在开始将我们的注意力转移到如何描述全部数据的分布形态上来。我们已经在图形上看到完全的分布形态。我们现在的目标是用词汇描述分布。虽然这种描述得到的信息要比图形少的多，但它仍然是很有用的。我们将关注分布的三个特征：众数的个数、对称性（偏度）或离散度。

因为我们首先对常见的分布形态感兴趣，而检验一个由光滑的曲线构成的图形要比由原始数据构成的图容易得多，图 4-3 呈现了这种想法的 3 个例子，两个图以直方图呈现它的分布，而另外一

个则由线图构成，上面光滑的曲线更适宜表现分布形态，但是并不能显示所有的细节。

4.2.1 众数的个数

描述分布形态的简单方式就是峰的个数，或者说众数的个数。图 4-4a 呈现的分布，叫做均匀分布（uniform distribution），因为没有众数，所有取值的次数将是相同的。图 4-4b 呈现了一个众数或者一个峰的分布，它叫做单峰分布（single-peaked 或者 unimodal）。按照惯例，分布中的任何峰也被看做一个众数，即使不是所有的峰都有相同的高度。例如，图 4-4c 可以说有两个众数，即使第 2 个峰比第 1 个要低，它依然是一个双峰分布（bimodal distribution），同样地，图 4-4d 的分布可以说有三个众数，是一个三峰分布（trimodal distribution）。

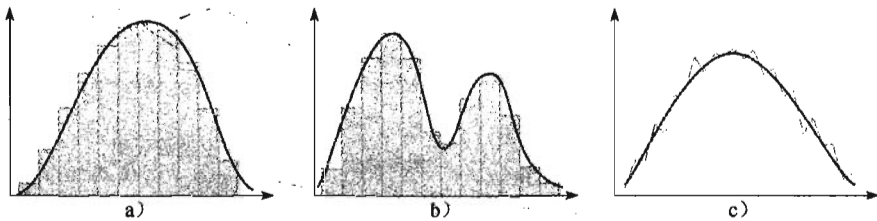


图 4-3 光滑的曲线更适宜表现分布形态

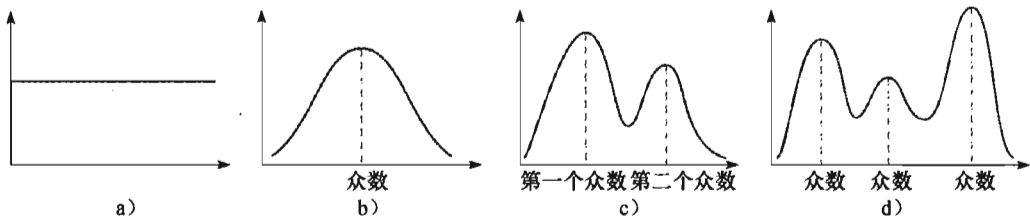


图 4-4 分布形态

a) 没有众数的均匀分布 b) 有一个众数的单峰分布 c) 有两个众数的双峰分布 d) 有三个众数的三峰分布

例 4-8 众数的个数

你认为下面每一个分布有多少众数？为什么？为每一个分布绘制一个粗略的曲线，并清楚地标记坐标轴。

- (1) 1000 名随机选择的成年女性的身高。
- (2) 1000 名随机选择的成年美国人，1 月份花费在观看电视上足球赛的时间。
- (3) 儿童服装零售店贯穿全年的每周销量。
- (4) 社会保障号最后一位是某个特定数字（从 0 到 9）的人口数。

解答 图 4-5 呈现了分布的略图。

- (1) 女性身高的分布是一个单峰分布，因为许多女性的身高在平均数左右，只有较少的女性身高高于或低于平均数。
- (2) 1000 名随机选择的美国成年人，花费在观看电视上足球赛的时间更像是一个双峰分布（两个众数），一个众数代表男性看电视的时间平均水平，另一个众数代表女性的平均水平。
- (3) 儿童服装零售店贯穿全年的每周销量，可能有几个众数。例如，春季可能有一个众数是准备换夏季服装，后面夏天有一个众数是返校时的销售，还有冬天的一个众数是寒假销售。
- (4) 社会保障号最后一位的数字本质上随机的，所以每个不同的末位数字相应的人口数应该是相同的，也就是，大约 10% 的人社会保障号尾数为 0，10% 尾数是 1，依次类推，因此这是一个没有众数的均匀分布。

4.2.2 对称性或偏态

第二个简单的描述分布形态的方式是对称性或者偏态，如果一个分布形态的左半边和右半边是互为镜像的话，那么它就是**对称的**（symmetric），图 4-6 中呈现的分布都是对称的，图 4-6a 是一个单峰钟形分布，这就是统计学中著名的正态分布，它是非常重要的分布形态，所以我们将第 5 章专门学习它。

如果一个分布不是对称的话，那么它的值在一边一定比另一边分布的更多些。在这种情况下，我们说这个分布是**偏态的**（skewed）。图 4-7a 呈现了一个分布，在这个分布中大多数数据值分布远离图形左边，意思是指在低端的数值会是异常值，我们称这样的分布为**左侧分布**（left-skewed），或者叫**负偏态**（negatively skewed）。如果你想象这个分布左侧拖着个长尾巴则更容易理解这个分布。图 4-7b 呈现了一个分布，在这个分布中大多数数据值分布远离图形右边，意思是指在高端的数值会是异常值，我们称这样的分布为**右侧分布**（right-skewed），或者叫**正偏态**（positively skewed）。如果你想象这个分布右侧拖着个长尾巴则更容易理解这个分布。

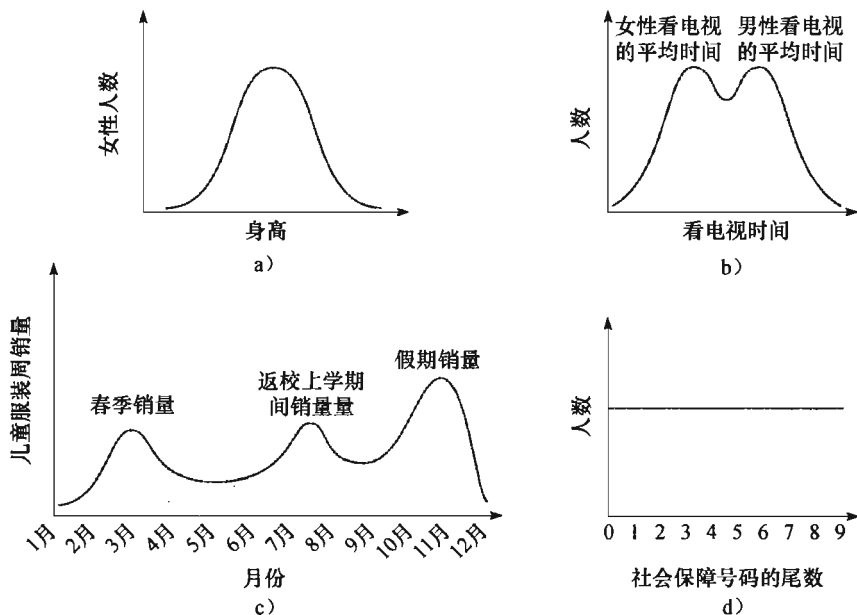


图 4-5 例 4-8 的略图

a) 身高分布 b) 看电视时间 c) 儿童服装周销量 d) 保障号码尾数

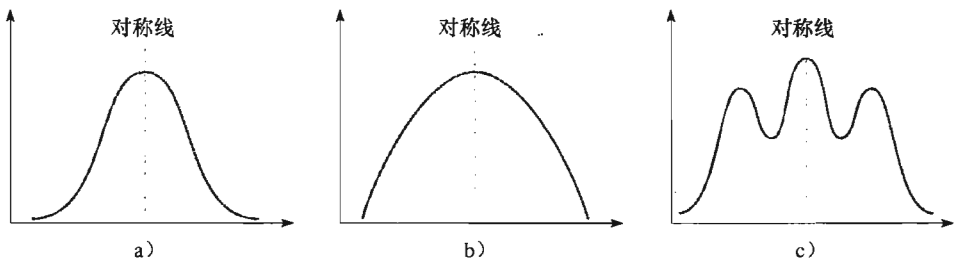


图 4-6 左半边和右半边互为镜像的对称分布
a) 单峰正态分布 b) 单峰非正态分布 c) 三峰对称分布

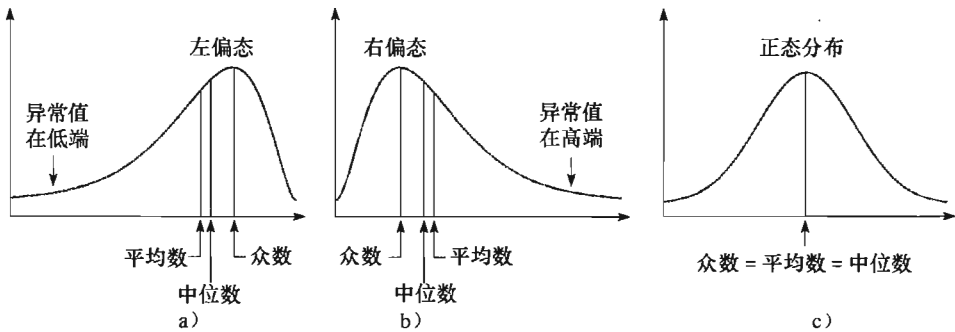


图 4-7 分布的偏态
a) 左侧分布 (平均数和中位数小于众数) b) 右侧分布 (平均数和中位数大于众数)
c) 对称分布 (平均数、中位数、众数三者相等)

图 4-7 呈现了峰度如何影响平均数、中位数和众数之间的相对位置。通过定义，在单峰分布中众数就是最高峰的位置。左侧分布的平均数和中位数靠左，它们的值小于众数，此外，低端的异常值使得平均数小于中位数（见表 4-3）；右侧分布中，平均数和中位数靠右，它们的值大于众数，在这种情况下，高端的异常值会使平均数大于中位数。当分布是对称且单峰时，平均数、中位数和众数三者刚好是相等的。

定义

- ◆ 一个分布是对称的形态，那么它的左半边和右半边就是互为镜像。
- ◆ 一个分布是左侧分布，大多数的数据值分布在远离图形左边。
- ◆ 一个分布是右侧分布，大多数的数据值分布在远离图形右边。

思考题

哪一个偏态分布中“集中量数”的最好指标（或者是分布的中心）：平均数还是中位数？为什么？

例 4-9 偏态

针对下面每一种情况，说明你认为该数据分布是对称的、左偏态的还是右偏态的？并加以解释。

- (1) 100 个女性的身高样本。
- (2) 美国的家庭收入。
- (3) 当巡逻车使用雷达测速时，公路上的汽车的行驶速度。

解答

- (1) 女性的身高分布是对称的，因为身高高于平均水平的和矮于平均水平的女性数量大致是相等的，离平均数远的两极端身高值都是非常少见的。
- (2) 家庭收入的分布属于右侧分布（正偏态），大多数家庭是中产阶层，所以家庭收入分布的众数是一个中产阶层家庭的收入（在美国这个收入水平在 50 000 美元左右）。但有少数家庭会使平均家庭收入靠右，将家庭收入的分布伸展到右侧（高值端）。
- (3) 当司机注意到巡逻车使用雷达测速时，他们通常会放慢行驶速度。很少有司机会超越速度限制，另外少数一些司机会放慢速度较多以保证驾驶速度远低于限制速度。这样，这个速度分布是一个左侧分布（负偏态），众数接近于所限制的速度，有一部分汽车的速度会低于限制速度。

相关知识

速度杀手,在美国,平均每12分钟就有一人死于交通事故。意外的死亡事故中有1/3涉及速度杀手。

思考题

在一般的英语情境中,术语“skewed”经常表示的意思是有些事物变形了或者以一种不均匀的方式来描述。在统计学中使用“skewed”与它的本义是如何关联的?

4.2.3 变异性

描述一个分布的第三种方式是采用变异性(variation)的概念,这是关于数据分布离散程度的测量指标。一个分布中的变异性较小时,大部分数据是聚集在一起的,如图4-8a所示,这样的分布会有一个比较突出的尖峰。当一个分布中的变异性较大时,数据在中心水平两边分布得比较分散,峰会显得低阔,图4-8b呈现了一个中等水平变异性的分布,图4-8c呈现了一个变异性大的分布。在下一节,我们将讨论变异性量化的方法。

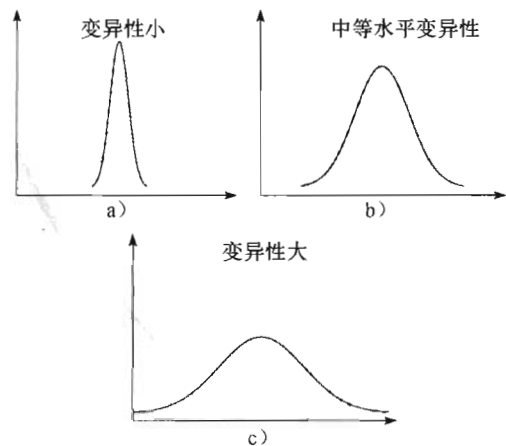


图4-8 三种分布的变异性从左到右逐渐增大

4.2练习題

统计素养和批判性思维

1. 对称性 当我们说一个图形是对称的时候意味着什么?
2. 分布 供给每一个家庭的电压在123.0 ~ 125.0伏特变化。所有的电压水平是近似相等的。下面哪一个术语能够最好地描述这个

定义

◆ 变异性描述了数据围绕这组数据中心分布的离散水平。

例4-10 马拉松比赛完成时间的变异性

你预计奥运会马拉松和纽约市马拉松比赛完成时间的变异性有什么不同?并加以解释。

解答 奥运会马拉松比赛的参与者只能是高水平的运动员,他们完成比赛的时间大多会更接近于世界纪录。纽约市马拉松比赛允许所有人报名参加,完成比赛时间的范围会非常宽泛,全距很大(从接近于世界纪录的两小时到更长的时间)。这样,纽约市马拉松比赛完成时间的变异性要远远大于奥运会。

相关知识

1997年,医生告诉25岁的兰斯·阿姆斯特朗,平均来说,扩散到脑的睾丸癌有不足50%的存活率。他不仅生存下来了,而且创造了连续7次赢得环法自行车大赛的纪录,他的故事说明,就个人层面来说,变异性比平均水平更重要。



分布:偏态分布、双峰分布、均匀分布和单峰分布?并加以解释。

3. 智商分数 思考一下一个统计学课程班里大学生的智商分数,与随机挑选的成人的智商分数进行比较。哪一组人群中的智商分数变异性较小?在这些智商分数的分布图形中较

小的变异性会产生什么效应？

4. 偏度 一个统计图中的偏度是指什么？

判断正误：判断练习题 5~8 中的陈述是有意义的（表述正确）还是无意义的（表述错误），并做出解释。并非所有的陈述都有明确的答案，所以你的解释比判断会更重要。

5. 对称性 因为一组数据有两个众数，所以它是不对称的。

6. 对称性 一组测验数据显示了是对称的，平均数是 75.0，中位数是 80.0。

7. 均匀分布 一组测验数据显示它是均匀分布，并且是单峰分布。

8. 分布 一组测验数据显示它是左侧分布并且是单峰分布。

概念与应用

9. 加州老实泉 图 4-9 的直方图呈现了黄石国家公园中的老实泉 300 次喷射时间的样本（意味着存在 299 次喷射间隔时间）。在这个直方图之上，绘制一条光滑曲线呈现它的一般特点。按照它的众数、对称性和偏态等特征对它的分布进行分类，并且用一句话概括你的结果。

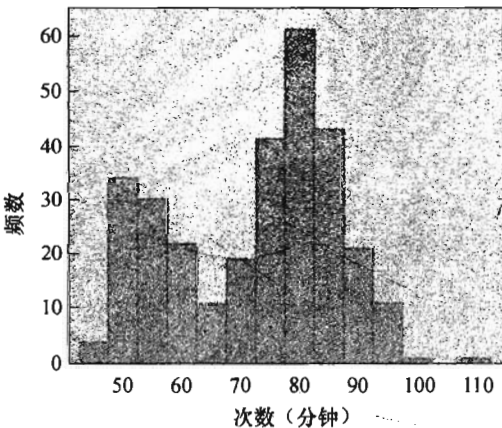


图 4-9 加州老实泉两次喷射的间隔时间
资料来源：Hand et al., *Handbook of Small Data Sets*.

10. 芯片失效 图 4-10 的直方图呈现了一个 108 块计算机芯片的失效时间。在这个直方图之上，绘制一条光滑曲线呈现它的一般特点。按照它的众数、对称性和偏态等特征对它的分布进行分类，并且用一句话概括你的结果。

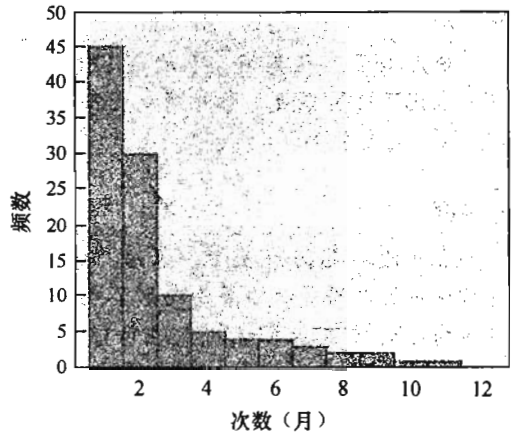


图 4-10 计算机芯片的失效时间
资料来源：Hand et al., *Handbook of Small Data Sets*.

11. 英式橄榄球运动员的体重分布 图 4-11 的直方图呈现了 391 名英式橄榄球运动员的体重。在这个直方图之上，绘制一条光滑曲线呈现它的一般特点。按照它的众数、对称性和偏态等特征对它的分布进行分类，并且用一句话概括你的结果。

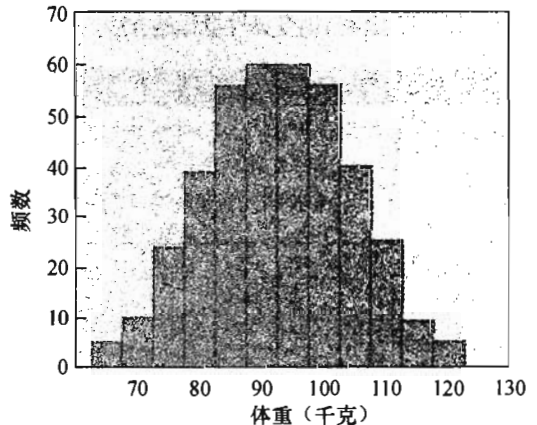


图 4-11 英式橄榄球运动员的体重
资料来源：Hand et al., *Handbook of Small Data Sets*.

12. 硬币重量 图 4-12 的直方图呈现了 72 枚硬币的重量（克）。在这个直方图上，绘制一条光滑曲线呈现它的一般特点。按照它的众数、对称性和偏态等特征对它的分布进行分类，并且用一句话概括你的结果。这个图的特色反映了下面的事实，其中 35 枚硬币是 1983 年前制造，包含了 97% 的铜和 3% 的锌。另外 37 硬币是 1983 年之后制造的，包含了 3% 的铜和 97% 的锌。

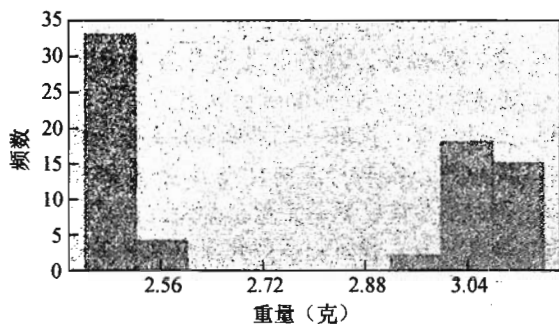


图 4-12 72 枚硬币样本的重量

资料来源: Measurements by Mario F. Triola.

13. 家庭收入 假如你研究了随机选择的 300 个家庭的收入, 调查结果概括如下: 家庭收入的平均数是 41 000 美元; 家庭收入的中位数是 35 000 美元; 家庭收入的最大值和最小化值分别是 25 万美元和 2 400 美元。

- (1) 绘制一个收入分布的粗略概图, 坐标轴要明确标示, 使用对称性、左偏态或者右偏态来描述这个分布。
- (2) 样本中有多少家庭收入少于 35 000 美元? 解释你是怎么知道的。
- (3) 基于给出的数据, 你能够确定有多少家庭收入大于 41 000 美元吗? 为什么?

14. 波士顿的降雨量 最近一年波士顿的每日降雨量 (英寸) 包括 365 个数值以及它们的比率: 每天的平均降雨量是 0.083 英寸; 每天降雨量的中位数是 0; 每天降雨量的最小值和最大值分别是 0 英寸和 1.4 英寸。

- (1) 在 365 个取值中, 最小值是 0 英寸, 中位数也是 0 英寸, 这可能吗?
- (2) 使用对称性、左偏态或者右偏态来描述这个分布。
- (3) 你能确定那些下雨天的总数字吗? 你能得出关于下雨天数的任何结论吗? 请加以解释。

描述分布: 描述练习题 15~30 中每一个题的分布, 回答下述问题:

- (1) 你预期这个分布有几个众数? 说明理由。
 - (2) 你预期这个分布是对称性、左偏态还是右偏态的? 说明理由。
15. 收入 在一个统计课程班里全部人们的年收入, 包括指导教师的。
16. 测验分数 参加一个公平测验的 200 名学生

的平均成绩是 75%。

17. 学生体重 100 名八年级学生的体重。
18. 足球运动员的体重 100 名职业足球运动员的体重。
19. 溜冰者的体重 一个滑冰场上午对职业花样滑冰选手开放, 下午和晚上则对职业冰球选手开放。所有在这个滑冰场的运动员的体重。
20. 交通工具的重量: 停放在代理商处的一些新汽车的重量, 存货清单中的车, 一半是小型汽车, 一半是运动型多用途车。
21. 起飞延误 飞机从一个大型飞机场起飞延误的时间 (以分钟计)。
22. 速度 驾驶员开车穿过一片学校区段的速度。
23. 参观者年龄 一个艺术博物馆参观者的年龄 (不包括学校学生的团体参观)。
24. 游览者年龄 游览一个主题娱乐公园的人们的年龄。
25. 网球运动员 参加女子网球锦标赛双打比赛的运动员的人数。
26. 销售量 圣地亚哥一家商店一年内游泳衣的每月销售量。
27. 收入 在超级碗比赛中坐在豪华包厢内的观众的收入。
28. 收入 在电视中收看超级碗比赛的人们的收入。
29. 薪资 棒球职业运动联盟的运动员薪资。
30. 安打率 棒球职业运动联盟的运动员的安打率。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 4 章的有用链接 “links for internet projects”。

31. 纽约马拉松比赛 浏览纽约马拉松比赛的网页, 找到最近马拉松比赛完成时间的次数分布数据。研究这个数据, 绘制一下这个数据分布的粗略概图, 然后用语言描述这个数据分布的形态。
32. 税收统计 美国国税局 (IRS) 网站提供了所得税纳税申报表、退款和其他统计数据。从网页中选择一组统计数据并且研究其分布。用语言描述其数据分布, 并且讨论你了解到的相关国家税收政策。
33. 社会保障数据 调查下面一个学生样本, 询

问每个人社会保障号码最后一位，并且也询问每个人社会保障号码的第5位。分别绘制最后一位数字和第5位数字分布的图形，将这两个图形加以比较，有什么值得注意的差异是非常明显的。

新闻中的统计学

34. 新闻中的分布 在近期的新闻报道中，寻找三个用直方图或线图呈现数据分布的例子，在每一个数据分布图上面绘制一条能显示它

一般特征的光滑曲线，然后根据它的众数多少、对称性、偏度以及变异性对分布进行分类。

35. 三峰分布 找出一个你希望有3个众数的真实分布，绘制一条关于这个分布的粗略概图，然后在坐标轴上做出标示。

36. 偏态分布 找出一个你期望它是左偏态或者右偏态的实际例子，绘制一条关于这个分布的粗略概图，然后在坐标轴上做出标示。



4.3 差异量数

在4.2节中，我们看到了如何定性描述数据的变异性，在这一节，我们开始学习差异量数的计算。

4.3.1 为什么存在变异性

设想你在两家不同的银行观察到客户在排队等待柜员服务，在至大银行，一个客户可以任选3条不同柜员通道中的一条排队等候；至佳银行也有三个柜员通道，但是所有客户都按单线排队等待被叫号，轮至下一个空闲柜员再接受服务，下列的数据是以分钟为单位的等待时间，每家银行均有11位客户，等候时间以升序排列：

至大银行（三条通道）：

4.1 5.2 5.6 6.2 6.7

7.2 7.7 7.7 8.5 9.3 11.0

至佳银行（单线排队）：

6.6 6.7 6.7 6.9 7.1

7.2 7.3 7.4 7.7 7.8 7.8

你可能会发现在至大银行的客户会比在至佳银行的客户更不愉快，但这并不是因为平均等待时间更长，实际上，你可以自己发现两家银行等待时间的平均数和中位数都是7.2分钟，客户满意度的差异来自两家银行等候时间的离散程度不同。至大银行的等候时间变化幅度更大，所以有一些客户等候的时间更长，更容易感到恼怒。相反，至佳银行的等候时间的离散程度是很小的，所以所有客户感觉到他们是被一视同仁的。图4-13用直方图呈现了这两种离散程度的差异，在直方图中，数据分组的组距

是1分钟。

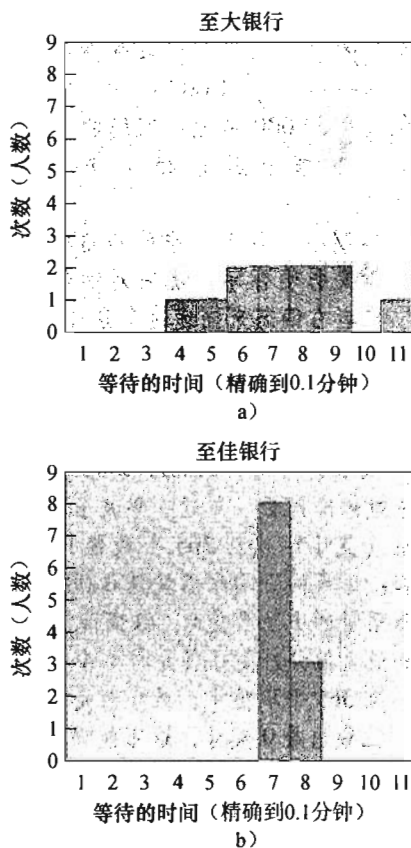


图4-13 在至大银行和至佳银行等待时间的直方图、呈现的数据分组精确到0.1分钟

相关知识

排队等候的时间不仅对于每个人而且对于数据来说都是重要的，特别对于互联网上海量的数据流动而言，大公司经常聘请统计学家帮助他们以确保大量的数据平稳地通过他们的服务器和网页，而不会产生瓶颈效应。

思考题

解释为什么拥有三条独立通道的至大银行会比至佳银行在等候时间上有更大的离散程度。考虑你通常排队等候的几个地方，比如：食品杂货店、银行、售票代理处或者一家快餐店。这些地方采用一条通道排队还是采用多条通道排队，如果有一个地方采用多条通道排队，你认为采用一条会不会更好？请说明原因。

4.3.2 全距

描述一组数据离散程度的最简单（但不一定是最好的）方法，是计算它的全距（range）。它的定义是：最低值（最小值）和最高值（最大值）之间的差别，以两个银行的例子来说明：至大银行的等候时间为4.1~11.0分钟，所以它的全距是 $11.0 - 4.1 = 6.9$ 分钟。至佳银行的等候时间为6.6~7.8分钟，所以它的全距是 $7.8 -$

$6.6 = 1.2$ （分钟）。至大银行的全距更大一些，反映了更大的离散程度。

$$\text{全距} = \text{最高值(最大值)} - \text{最低值(最小值)} \quad (4-6)$$

虽然全距易于计算和方便使用，但是它偶尔也会被误解。就像下一个例子所显示的。

例4-11 被误解的全距

考虑下面两组关于9个学生的测验分数。哪一组数据全距更大，你能说这组数据离散程度更高吗？

测验1: 1, 10, 10, 10, 10, 10, 10, 10, 10

测验2: 2, 3, 4, 5, 6, 7, 8, 9, 10

解答 测验1的全距是 $10 - 1 = 9$ 分，测验2的全距是 $10 - 2 = 8$ 分，这样，测验1的全距更大一些。然而，除去一个单个的低分（一个异常值），测验1根本没有什么差异，因为每个学生都得到了10分。相反，在测验2中没有两个学生得到相等的分数，成绩遍布所有可能的分数，测验2的离散程度实际上更高，尽管测验1全距较大。



资料来源：© 1998 Scott Adams, Inc. Distributed by United Feature Syndicate. Reprinted with permission. All rights reserved.

4.3.3 四分位数和五数概括法

方法注解

统计学没有统一公认的四分位数的计算流程，不同的计算流程会导致结果值有轻微的差异。

描述变异的一个较好的办法是，除了考虑几个最大值和最小值之外的中间值，一个最常用的办法是要求出四分位数（quartiles），也就是将数据分布分割成4段的数值。下述数据重复列出了两家银行的等候时间，同时也用黑体列出了四分位数。注意中间的那个四分位数将数据分成两半，也就是中位数。

	下四分位数 (Q_1)	中四分位数 (Q_2)	上四分位数 (Q_3)		
至大银行:	4.1	5.2	5.6	6.2	6.7
7.2	7.7	7.7	8.5	9.3	11.0
至佳银行:	6.6	6.7	6.7	6.9	7.1
7.2	7.3	7.4	7.7	7.8	7.8

定义

- ◆ 下四分位数 (lower quartile), 也叫做第一四分位数, 或者 Q_1 , 将这组数据下端的 $1/4$ 和上端 $3/4$ 分开, 它也是这组数据低端这一半数据的中位数 (如果这个中位数值本身就是这组数据中间的一个数据, 也包括这组数据的中位数)。
- ◆ 中四分位数 (middle quartile), 也叫做第二四分位数 (second quartile), 或者 Q_2 , 是全部数据的中位数。
- ◆ 上四分位数 (upper quartile), 也叫做第三四分位数 (third quartile), 或者 Q_3 , 将全部数据下端的 $3/4$ 和上端的 $1/4$ 区分开来, 它是这组数据高端这一半数据的中位数, (如果这个中位数值本身就是这组数据中间的一个数据, 也包括这组数据的中位数)。

一旦我们知道了四分位数, 我们就可以用五数概括法 (five-number summary) 来描述一组数据分布。这 5 个数值包括最小值、第一四分位数、中位数、第三四分位数和最大值。对于两家银行的等候时间, 五数概括法包含下述数据:

	至大银行	至佳银行
最小值	= 4.1	= 6.6
第一四分位数	= 5.6	= 6.7
中位数	= 7.2	= 7.2
第三四分位数	= 8.5	= 7.7
最大值	= 11.0	= 7.8

我们可以用一个叫做箱形图 (boxplot), 有时也叫做箱须图 (box-and-whisker plot) 的图形来呈现五数概括法。我们使用一个参考数轴, 首先从第一四分位数到第三四分位数绘制一个箱形图, 把这之间的数值框起来, 然后从中位数的位置画一条竖线贯穿箱形, 最后在最大值和最小值的位置绘制两条竖线, 如同全图的两根胡须一般。图 4-14 呈现了银行等候时间的箱须图。至大银行的箱须图明显地比至佳银行的要宽阔, 说明至大银行的等候时间具有更大的离散性。

绘制箱形图

- 第 1 步: 画一条涵盖所有数据值得数轴;
- 第 2 步: 从第一四分位数到第三四分位数之间用箱形图封闭起来;
- 第 3 步: 在中位数位置画一条竖线贯穿箱形;
- 第 4 步: 在最大值和最小值处绘制两条如须状的竖线。

方法注解

本书中所呈现的箱须图叫做骨架箱形图。一些箱须图画出异常值并标注 (*), 并且须只扩展到非异常值的最大值和最小值, 这种箱须图也叫校准箱须图。

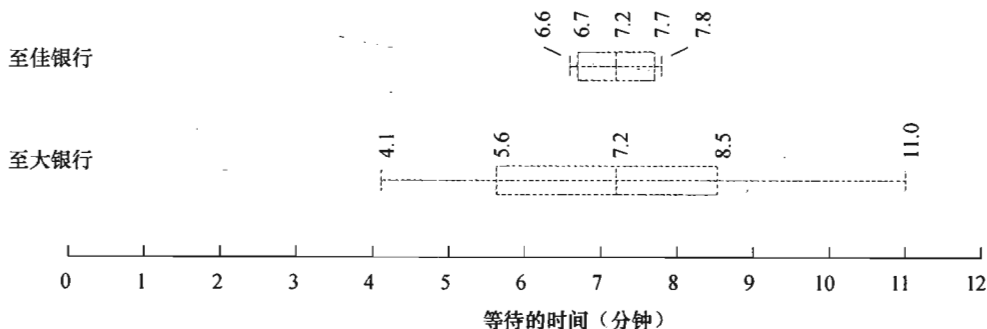


图 4-14 箱形图显示在至大银行等待时间的变异性比在至佳银行等待时间的变异性大

例 4-12 主动和被动吸烟

一个研究吸烟的方法是：测量血液中的血清可替宁的水平，它是人们在吸烟后人体吸收尼古丁后的代谢产物，表 4-4 列出了 50 个吸烟者（主动吸烟者）以及 50 个不吸烟者的样本，不吸烟者的样本是自己不抽烟但是却暴露在家庭和工作场所的吸烟环境中（被动吸烟者）。样本数据为他们血液中的血清可替宁水平，采用五数概括法和箱形图比较这两组数据（吸烟者和非吸烟者），并且讨论你的结果。

表 4-4 主动吸烟者和被动吸烟者各 50 人样本的血清可替宁的水平（毫微克/毫升血液），数据值以升序排列

编号	抽烟者	非吸烟者	编号	抽烟者	非吸烟者
1	0.08	0.03	26	34.21	0.82
2	0.14	0.07	27	36.73	0.97
3	0.27	0.08	28	37.73	1.12
4	0.44	0.08	29	39.48	1.23
5	0.51	0.09	30	48.58	1.37
6	1.78	0.09	31	51.21	1.40
7	2.55	0.10	32	56.74	1.67
8	3.03	0.11	33	58.69	1.98
9	3.44	0.12	34	72.37	2.33
10	4.98	0.12	35	104.54	2.42
11	6.87	0.14	36	114.49	2.66
12	11.12	0.17	37	145.43	2.87
13	12.58	0.20	38	187.34	3.13
14	13.73	0.23	39	226.82	3.54
15	14.42	0.27	40	267.83	3.76
16	18.22	0.28	41	328.46	4.58
17	19.28	0.30	42	388.74	5.31
18	20.16	0.33	43	405.28	6.20
19	23.67	0.37	44	415.38	7.14
20	25.00	0.38	45	417.82	7.25
21	25.39	0.44	46	539.62	10.23
22	29.41	0.49	47	592.79	10.83
23	30.71	0.51	48	688.36	17.11
24	32.54	0.51	49	692.51	37.44
25	32.56	0.68	50	983.41	61.33

注：包括“编号”一列是为了更便利地读表。
资料来源：National Health and Nutrition Examination Survey, National Institutes of Health.

相关知识

主动吸烟对儿童伤害很大，香烟中的毒素对发育的身体要比已经完全成熟的成年人影响更大，同样的效应也发生在其他毒素中，这也是限制儿童暴露于有毒化学品中是特别重要的原因。

解答 两组数据已经按升序排列，这样应用五数概括法就很容易。每一组数据都有 50 个数据值，所以中位数位于第 25 个和第 26 个数据中间，是这两个数据和的一半。对于吸烟者来说，第 25 个和第 26 个数据分别是 32.56 和 34.21，所以中位数是 $(32.56 + 34.21)/2 = 33.385$ ；对非吸烟者来说，第 25 个和第 26 个数据分别是 0.68 和 0.82，所以中位数是 $(0.68 + 0.82)/2 = 0.75$ 。第一四分位数是低端这一半数据的中位数，也就是各组数据中的第 13 个数据值；第三四分位数是高端那一半数据的中位数，也就是各组数据中的第 38 个数据值。对于这两组数据而言应用五数概括法得到如下结果：

主动吸烟者
最小值 = 0.08ng/ml
第一四分位数 = 12.58ng/ml
中位数 = 33.385ng/ml
第三四分位数 = 187.34ng/ml
最大值 = 983.41ng/ml
被动吸烟者
最小值 = 0.03ng/ml
第一四分位数 = 0.20ng/ml
中位数 = 33.13mg/m
第三四分位数 = 30.75ng/ml
最大值 = 61.33ng/ml

图 4-15 显示了这两组数据的箱形图。在箱形图上，我们观察每组数据的关键特征会更容易，例如，我们马上能够清楚地看到，吸烟组的血清可替宁水平的中位数值更大，并且离散程度也更高。我们可以得出结论，吸烟者比起暴露在吸烟环境中但没有抽烟的那些人而言，吸收了更多的尼古丁。虽然如此，被动抽烟者的血清可替宁水平，比起那些吸烟并且没有暴露在抽烟环境中的人更高（可由并不显示在这里的其他数据来证明）。更进一步来说，被动吸烟的量如果处于较高水平的话，甚至会比那些中等水平的吸烟者吸收更多的尼古丁。我们可以得出结论说，二手烟可以让非吸烟者吸收明显数量的尼古丁。基于吸烟的这些明显危害，这些结果可以告诉我们，被动吸烟与健康是有密切关系的。

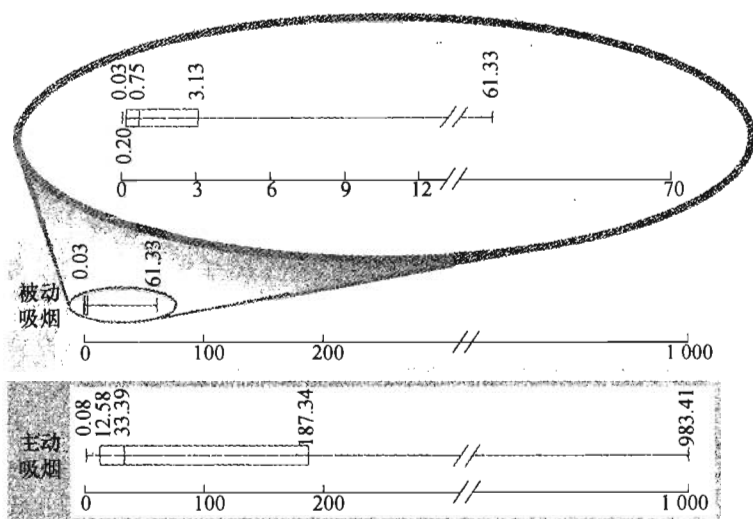


图 4-15 表 4-4 中数据的箱线图

4.3.4 百分位数

方法注解

与四分位数一样，统计学家和不同的统计软件包可能使用不同的计算过程来计算百分位数，结果也会有细微的差异。

四分位数将一组数据分成四段，将这一组数据分成更多段也是可能的，例如，五分位数将一组数据分成 5 段，十分位数将一组数据分成十段，更常见的是将一组数据分成 100 段，这就是我们所采用的百分位数（percentiles）。大体来说，以第 35 百分位数为例，它是一个将一组数据分成低端 35% 和高端 65% 的数值（更准确地说，第 35 百分位数，大于等于至少 35% 的数据值，并且小于等于至少 65% 的数据值）。

如果一个数据值位于两个百分位数之间，通常来说，将这个数据值归属于较低的百分位数。例如，大学入学测验的分数，如果你的分数值高于 84.7% 的参加测验的人，我们就说你的分数是第 84 百分位数。

定义

◆ 第 n 百分位数 (n th percentiles) 将一组数据分成低端的 $n\%$ 和高端的 $(100 - n)\%$ ，一个数据值位于两个百分位数之间，通常来说，归属于较低的百分位数。你可以用下面的公式计算任何数据值的百分位数：

某个数据值对应的百分位数

$$= \frac{\text{小于这个数据值的数据个数}}{\text{数据的总个数}} \times 100 \quad (4-7)$$

计算一个给出的百分位数所对应的原始数据值则是另外一个过程，适用的方法是：寻找第 L 个数据值 (L 是百分位数和样本量的乘积)，例如，对于 50 个样本数据值，第 12 百分位数是 $0.12 \times 50 =$ 第 6 个数值。

例 4-13 抽烟的百分位数

回答下列与表 4-4 中位数据有关的问题：

- (1) 对于抽烟者来说，数据值 104.54ng/ml 对应的百分位数是多少？
- (2) 对于非吸烟者来说，数据值 61.33ng/ml 对应的百分位数是多少？
- (3) 对于抽烟者来说，第 36 百分位数对应哪个数据值？对于非抽烟者来说，第 36 百分位数对应的数据值为多少？

解答

- (1) 对于抽烟者来说，104.54ng/ml 的数据值在整个数据值中排在第 35 位，也就是 34 个数据值位于它下面，这样，它的百分位数是：

$$\frac{\text{少于 104.54ng/ml 的数据值个数}}{\text{这组数据的总个数}} \times 100 \\ = 34/50 \times 100 = 68$$

换句话说，第 35 个数据是第 68 百分位数。

- (2) 数据值 61.33ng/ml 是非吸烟者中的第 50 个数据，也是最大值，意味着 49 个数据位于它下面，这样它的百分位数是：

$$\frac{\text{少于 61.33ng/ml 的数据值个数}}{\text{数据总个数}} \times 100 \\ = 49/50 \times 100 = 98$$

换句话说，这组数据的最大值位于第 98 百分位数。

- (3) 因为这组数据有 50 个数值, 第 36 百分位数是 $0.36 \times 50 = 18$, 对于抽烟者, 第 18 个数值是 20.16ng/ml, 对于非抽烟者, 这个数值是 0.33ng/ml。■

4.3.5 标准差

五数概括法很好地描述了变异的特征, 但是统计学上更希望用一个数字来描述变异程度。经常用来描述离散程度的这个数字叫做标准差 (standard deviation)。

标准差是描述数据分布离散程度的一个差异量数, 它是一组数据中各个数据值围绕平均数的距离是否宽泛的一个度量值。我们要想计算标准差, 首先要计算出平均数, 然后计算出每一个数值距离平均数远近的“离均差” (deviation from the mean), 或者叫“离差” (deviation)。以银行数据为例, 每家银行的平均等候时间都是 7.2 分钟, 8.2 分钟的等候时间它的离差就是 8.2 分钟 - 7.2 分钟 = 1.0 分钟, 意思是指, 它比平均等候时间多 1 分钟。5.2 分钟的等候时间, 它的离差是 5.2 - 7.2 = -2 (分钟), 它比平均数少两分钟。

方法注解

当使用从一个样本得来的数据来计算标准差时, 计算的一部分包括将离均差的平方和除以数据值的总数减去 1。当用的是全部的总体时, 就不必再减去 1。在本书中, 我们只采用针对样本的公式。

标准差是所有数据与平均数之间离差的平均度量值。离差的实际平均数是 0, 因为正离差和负离差总是平衡的。为了避免得到一个 0 值, 我们首先计算离差的平方, 然后取平方根, 最后得到标准差 (因为技术的原因, 我们将平方和除以总的次数 - 1)。

标准差的计算过程

计算任何一组数据的标准差, 都经过以下步骤:

第 1 步, 计算这组数据的平均数, 然后通过每个数据值减去平均数得到每个数据的离差, 这样对每个数据来说:

$$\text{离差} = \text{数据值} - \text{平均数} \quad (4-8)$$

第 2 步, 计算所有离差的平方值;

第 3 步, 将所有离差的平方值累加起来;

第 4 步, 将这个累加的总和除以总数据的个数减 1;

第 5 步, 上述的商值再取平方根就是标准差。

概括地说, 标准差公式如下:

$$\text{标准差} = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad (4-9)$$

方法注解

第 4 步的结果叫做分布的方差。换句话说, 标准差是方差的平方根。虽然方差应用于高级的统计计算中, 但本书中并不使用方差。

注意, 因为我们在标准差的计算过程中, 第 3 步是将离差平方, 然后在第 5 步中取它的平方根, 所以标准差的单位和原始数据的单位是一样的。例如, 原始数据的单位是分钟, 标准差的单位也是分钟。

理论上, 标准差公式是比较容易应用的, 但是实际计算起来, 除了非常简单的数据, 一般都是非常枯燥繁琐的, 实际上, 我们经常会采用计算器或计算机来计算。尽管如此, 如果通过一些例子进行实际详细计算的话, 你会发现标准差的公式非常容易理解。

例 4-14 标准差计算

计算至大银行和至佳银行各自等候时间的标准差

解答 依照上述计算标准差的 5 个步骤, 表 4-5 显示了如何完成前 3 个步骤。第 1 列列出了每个银行的等候时间, 第 2 列列出了离差 (第 1 步), 第 3 列列出了离差的平方 (第 2 步), 我们将所有的离差平方累加起来就可以得到第 3 列最下面的总和 (第 3 步)。注意, 因为我们早已经知道平均等候时间是 7.2 分钟, 所以我们可以计算离差。对于第 4 步来说, 我们将第 3 步的总和除以总的数据个数减 1, 因为一共有 11 个数据值, 所以除以 10。

$$\text{至大银行: } 38.46/10 = 3.846$$

$$\text{至佳银行: } 1.98/10 = 0.198$$

最后, 第 5 步告诉我们, 标准差是第 4 步结果的平方根。

$$\text{至大银行: 标准差} = \sqrt{3.846} = 1.96 \text{ 分钟}$$

$$\text{至佳银行: 标准差} = \sqrt{0.198} = 0.44 \text{ 分钟}$$

换句话说, 虽然两家银行都有相同的平均数, 但是在至大银行中距离平均等候时间上下, 最典型的等候时间大约是 1.96 分钟, 而在至佳银行, 这个时间是 0.44 分钟, 我们又一次看到至大银行等候时间的变异程度更大, 这也解释了为什么至大银行的客户会比至佳银行的客户感到更恼怒。■

表 4-5 计算标准差

至大银行			至佳银行		
等候时间	离差 (等候时间 - 平均数)	离差的平方	等候时间	离差 (等候时间 - 平均数)	离差的平方
4.1	$4.1 - 7.2 = -3.1$	$(-3.1)^2 = 9.61$	6.6	$6.6 - 7.2 = -0.6$	$(-0.6)^2 = 0.36$
5.2	$5.2 - 7.2 = -2.0$	$(-2.0)^2 = 4.00$	6.7	$6.7 - 7.2 = -0.5$	$(-0.5)^2 = 0.25$
5.6	$5.6 - 7.2 = -1.6$	$(-1.6)^2 = 2.56$	6.7	$6.7 - 7.2 = -0.5$	$(-0.5)^2 = 0.25$
6.2	$6.2 - 7.2 = -1.0$	$(-1.0)^2 = 1.00$	6.9	$6.9 - 7.2 = -0.3$	$(-0.3)^2 = 0.09$
6.7	$6.2 - 7.2 = -0.5$	$(-0.5)^2 = 0.25$	7.1	$7.1 - 7.2 = -0.1$	$(-0.1)^2 = 0.01$
7.2	$7.2 - 7.2 = 0.0$	$(0.0)^2 = 0.0$	7.2	$7.2 - 7.2 = 0.0$	$(0.0)^2 = 0.0$
7.7	$7.7 - 7.2 = 0.5$	$(0.5)^2 = 0.25$	7.3	$7.3 - 7.2 = 0.1$	$(0.1)^2 = 0.01$
7.7	$7.7 - 7.2 = 0.5$	$(0.5)^2 = 0.25$	7.4	$7.4 - 7.2 = 0.2$	$(0.2)^2 = 0.04$
8.5	$8.5 - 7.2 = 1.3$	$(1.3)^2 = 1.69$	7.7	$7.7 - 7.2 = 0.5$	$(0.5)^2 = 0.25$
9.3	$9.3 - 7.2 = 2.1$	$(2.1)^2 = 4.41$	7.8	$7.8 - 7.2 = 0.6$	$(0.6)^2 = 0.36$
11.0	$11.0 - 7.2 = 3.8$	$(3.8)^2 = 14.44$	7.8	$7.8 - 7.2 = 0.6$	$(0.6)^2 = 0.36$
		和 = 38.46			和 = 1.98

思考题

仔细地观察例 4-14 中表 4-5 的离差，是不是两组数据的标准差看起来都像是离差的集中量数？请说明原因。

方法注解

另一个解释标准差的方法是使用叫做切比雪夫定理的数学规则。它的表述如下，对于任何数据分布，所有数据值中至少 75% 在平均数两个标准差以内，至少 89% 在平均数三个标准差以内，虽然我们在本书中不使用这个定理，如果你要学习另一门统计课程的话你可能会用到它。

4.3.6 标准差的理解

更进一步理解标准差的一个好方法是：考虑下面概括的一种适用原则，叫做全距的大拇指法则（range rules of thumb），或者称为全距经验法则。

全距的大拇指法则

标准差与全距近似的关系可以通过全距的大拇指法则建立：

$$\text{标准差} \approx \text{全距} / 4 \quad (4-10)$$

如果我们知道全距的分布，我们可以用这个原则估计标准差。反过来，如果我们知道标准差，我们也可以使用这个原则估计最小值和最大值。

$$\text{最小值} \approx \text{平均数} - (2 \times \text{标准差}) \quad (4-11)$$

$$\text{最大值} \approx \text{平均数} + (2 \times \text{标准差}) \quad (4-12)$$

当最大值和最小值是异常值时，这个原则不适用。

在数据值均匀分布的时候，这个原则的应用是合理的，当最大值和最小值是极端的异常值时，这个原则是不适用的。在某一特定情况下，你必须判断，然后决定这个原则是否适用，另外也要记住这个原则是大致近似的，并不是绝对精确的。

例 4-15 应用全距的大拇指法则

应用全距的大拇指法则估计至大银行和至佳银行等候时间的标准差。与例 4-14 中计算出的真实值进行比较。

解答 至大银行的等候时间是 4.1~11.0 分钟，意味着全距是 $11.0 - 4.1 = 6.9$ 分钟，至佳银行的等候时间是 6.6~7.8 分钟，全距是 $7.8 - 6.6 = 1.2$ 分钟，这样根据全距的大拇指法则可以得出下面的标准差估计值：

$$\text{至大银行标准差} \approx 6.9 / 4 = 1.7$$

$$\text{至佳银行标准差} \approx 1.2 / 4 = 0.3$$

例 4-14 中计算的实际标准差分别是 1.96 和 0.44，对于这两个例子而言，应用全距的大拇指法则的估计值稍微小于真实值，尽管如此，估计值还是给我们一个恰当的近似值，说明这个法则是简单可用的。

相关知识

很多技术比如触媒转换器可以帮助降低汽车污染物的释放，但减少使用汽油是降低导致气候变暖的碳排放量的唯一方法。这也是为什么汽车制造商大力发展高里程数的混合动力汽车和零排放量汽车（使用电力或燃料电池）的主要原因。

例4-16 全距的估计

研究BMW汽车在各种不同路况下的汽油公里数,结果显示每加仑汽油平均的公里数为22英里,标准差为3英里。考虑在一般路况中,估计汽油典型的最大公里数和最小公里数。

解答 依据全距的拇指原则,每加仑汽油的最大公里数和最小公里数近似为:

$$\text{最小公里数} \approx \text{平均数} - (2 \times \text{标准差})$$

$$= 22 - (2 \times 3) = 16 \text{ 英里}$$

$$\text{最大公里数} \approx \text{平均数} + (2 \times \text{标准差})$$

$$= 22 + (2 \times 3) = 28 \text{ 英里}$$

BMW汽车的汽油公里大概从最小值16英里到最大值28英里。 ■

4.3.7 应用累加符号表示标准差(可节选)

应用前面所介绍的累加符号更容易书写标准差的公式。回顾一下 x 表示一组数据中的每个个体数据, $\bar{x}$ 代表这组数据的平均数。这样,我们可以将每个数据与平均数的离差书写成如下形式:

$$\text{离差} = \text{数据值} - \text{平均数} = x - \bar{x} \quad (4-13)$$

我们也可以将离差的平方和写成:

$$\text{离差的平方和} = \sum (x - \bar{x})^2 \quad (4-14)$$

下面一步是将离差的平方和除以 $(n-1)$ 并开方取正根就计算出标准差了。你也可以将前面几步概括成:

$$s = \text{标准差} = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad (4-15)$$

符号 s 是样本标准差约定俗成的符号,对于总体标准差,统计学家一般使用希腊字母 σ (西格玛),而样本标准差中的 $n-1$ 也被 n 代替。因此,根据你假定数据代表的是一个样本还是总体,计算的标准差会稍有不同。

方法注解

$$\text{方差的公式是 } s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

方差的标准符号 s^2 反映了方差是标准差平方的事实。

4.3练习题

统计素养和批判性思维

1. 变异性 为什么标准差被看做变异性的一个测量指标?用你自己的话描述采用标准差作为测量指标的一组数据的特征。
2. 变异性比较 下面哪一种情况的变异性更大:一个物理课堂上30名学生的智商分数,一个电影院里30位观众的智商分数?为什么?
3. 正确地陈述 在《别让统计图表糊弄你》这本书中,作者写道“标准差经常用来呈现最大值、最小值分别与平均数的差异,最大值与平均数的差异是正数,最小值与平均数的差异是负数。例如,平均数是1,最大值是3,最小值是-1,标准差是 ± 2 。”这样的陈述正确吗?为什么?
4. 四分位数 对一个庞大的数据组,第一四分位数的计算结果是93.2。当我们说93.2是第一四分位数时,我们的意思是什么?

判断正误:判断练习5~8中的陈述是有意义的(表述正确)还是无意义的(表述错误),并解释。并非所有的陈述都有明确的答案,所以你的解释比判断更重要。

5. SAT分数 一个SAT分数是中位数并且它处于第60百分等级。
6. 家庭收入 《今日美国》的一篇报导陈述说,这几年来美国家庭的年收入的中位数是43 200美元,平均数是70 700美元。
7. 收入 50位全职统计学教师年收入的的标准差少于50位全职物理学家年收入的的标准差。
8. 样本容量 50位随机挑选的全职统计学教师年收入标准差少于200位随机挑选的全职物理学家的年收入标准差。

概念与应用

- 全距和标准差:**练习题19~12每道题列出了一组数字。在每道题中,计算全距和标准差(同样的一组数据用于4.1节练习题13~20)。
9. 知觉反应时 统计课的学生参与一个测验,让他们判断1分钟(60秒)时间是多长,所记录下来的他们反应的实际时间如下所示。

53 52 75 62 68 58 49 49

10. 身体温度 随机选择的正常和健康的成年人的身体温度(华氏温度)。

98.6 98.6 98.0 98.0 99.0

98.4 98.4 98.4 98.4 98.6

11. 血液中酒精浓度 下列是酒驾导致交通事故并被判入狱服刑时驾驶员血液中的酒精浓度。(数据来源于美国司法部)

0.27 0.17 0.17 0.16 0.13 0.24

0.29 0.24 0.14 0.16 0.12 0.16

12. 加州老实泉 美国黄石国家公园老实泉两次喷射之间的时间间距(以分钟计)。

98 92 95 87 96 90

65 92 95 93 98 94

13. 果蝇 一个雄果蝇样本的胸腔宽度(以毫米计)。

0.72 0.90 0.84 0.68 0.84 0.90

0.92 0.84 0.64 0.84 0.76

14. 美国总统的年龄 下面是选取的一些美国总统就任时的年龄。

57 61 57 57 58 57 61 54

68 51 49 64 50 48 65

15. 糖果的重量 下面是随机挑选的公司的简装糖果的重量(以克计)。

0.957 0.912 0.842 0.925 0.939 0.886

0.914 0.913 0.958 0.947 0.920

16. 25美分的硬币 下面是随机挑选的25美分的硬币重量(以克计)。

5.60 5.63 5.58 5.56 5.66 5.58 5.57 5.59

5.67 5.61 5.84 5.73 5.53 5.58 5.52 5.65

5.57 5.71 5.59 5.53 5.63 5.68

变异性的比较：在练习题17~20中，计算每道题中所呈现的两个样本的全距和标准差，并将两个结果加以比较。

17. 瓢泼大雨 统计学有时经常使用作者的不同著作进行比较和识别。下面列出了两本著作前言中的前20个单词的长度，一本是田纳西·威廉斯《热铁皮屋顶上的猫》，另外一本是苏斯博士的《戴帽子的猫》。

《热铁皮屋顶上的猫》：

2 6 2 2 1 4 4 2 4 2

3 8 4 2 2 7 7 2 3 11

《戴帽子的猫》：

3 3 3 3 5 2 3 3 3 2

4 2 2 3 2 3 5 3 4 4

18. 偷渡者的年龄 “玛丽女王号”在英格兰和美国之间航行，在船上经常会发现一些偷

渡者。这些从西往东或者从东往西的偷渡者的年龄在下面被罗列出来(数据来源于肯纳德轮船有限公司)。比较这两种数据的变异性。

从西往东的偷渡者年龄：

24 24 34 15 19 22 18 20 20 17

从东往西的偷渡者年龄：

41 24 32 26 39 45 24 21 22 21

19. 天气预报的准确性 在对天气预报准确性的分析中，经常要用到的是，一个是实际的最高温度与前一天预报的最高温度的比较差值，另一个是实际的最高温度与前5天预报的最高温度的比较差值。下面是陈列的是纽约市达奇斯地区预测最高温度和实际最高温度连续数天的误差。是否如我们预期的那样，标准差告诉我们前一天预测的温度要比前5天预测的温度更准确一些？

(实际高温) - (1天前预测的高温)

2 2 0 0 -3 -2 1

-2 8 1 0 -1 0 1

(实际高温) - (5天前预测的高温)

0 -3 2 5 -6 -9 4

-1 6 -2 -2 -1 6 -4

20. 实验处理效果 宾夕法尼亚州立大学的研究者进行了一项关于白杨树的实验。下面给出了没有给予处理的白杨树的重量和给予灌溉和施肥的白杨树的重量(千克)。这两组标准差存在差异吗？

没有实验处理的：

0.15 0.02 0.16 0.37 0.22

给予灌溉和施肥的：

2.03 0.27 0.92 1.07 2.38

21. 计算百分等级 本书的一位作者花了很多时间将一包465颗糖果中的每一颗糖果都称了一下重量。

(1) 其中的一颗糖果重0.776克，它比另外25颗糖果要重一些，这个特定的重量值百分等级是多少？

(2) 其中的一颗糖果重0.876克，它比另外322颗糖果重，这个糖果重量值百分等级是多少？

(3) 其中的某一个糖果重0.856克，它比另外224颗糖果重，这个糖果重量值的百分等级是多少？

22. 计算百分等级 一组数据包括 76 位获得奥斯卡最佳女演员奖项时的获奖年龄。
- (1) 一位女演员获奖时的年龄是 34 岁, 获奖时的年龄要比其他 38 位女演员获奖时的年龄大, 34 岁对应的百分等级是多少?
 - (2) 一位女演员获奖时的年龄是 29 岁, 获奖时的年龄要比其他 20 位女演员获奖时的年龄大, 29 岁对应的百分等级是多少?
 - (3) 一位女演员获奖时的年龄是 60 岁, 获奖时的年龄要比其他 71 位女演员获奖时的年龄大, 60 岁对应的百分等级是多少?
23. 对标准差的理解 下列 4 组数据都包含 7 个数, 平均数都是 9。
- 第一组数据: 9 9 9 9 9 9 9
- 第 2 组数据: 8 8 9 9 9 10 10
- 第 3 组数据: 8 8 8 9 10 10 10
- 第 4 组数据: 6 6 6 9 12 12 12
- (1) 绘制每一组数据的直方图。
 - (2) 计算出五数概括法的数值, 并针对每一组数据绘制箱须图。
 - (3) 计算每一组数据的标准差。
 - (4) 基于你计算的结果, 简要介绍一下为什么标准差提供了一个每组数据值变异性的单数的概括。
24. 标准差的理解 下列 4 组数据都包含 7 个数, 平均数都是 6。
- 第一组数据: 6 6 6 6 6 6 6
- 第 2 组数据: 5 5 6 6 6 7 7
- 第 3 组数据: 5 5 5 6 7 7 7
- 第 4 组数据: 3 3 6 6 9 9 9
- (1) 绘制每一组数据的直方图。
 - (2) 计算出五数概括法的数值, 并针对每一组数据绘制箱须图。
 - (3) 计算每一组数据的标准差。
 - (4) 基于你计算的结果, 简要介绍一下为什么标准差提供了一个每一组数据值变异性的单数的概括。
- 变异性的比较:** 练习题 25 ~ 28, 完成以下任务。
- (1) 计算每题中两组数据各自的平均数、中位数和全距。
 - (2) 计算每题中两组数据各自的五数概括法的数值, 并针对每一组数据分别绘制箱须图。
 - (3) 计算每题中两组数据各自的标准差。
 - (4) 应用全距的拇指法则估计每题中两组数据各自的标准差。这个法则应用于哪一组数据较好? 简要讨论这个法则应用的适宜性。
 - (5) 基于你计算的结果, 比较并且讨论两组数据他们的集中趋势和变异性。
25. 下列数据给出了波特兰大学的教师停车场和学生停车场所停放的汽车已使用年数。
- 教师停车场: 2, 3, 1, 0, 1, 2, 4, 3, 3, 2, 1
- 学生停车场: 5, 6, 8, 2, 7, 10, 1, 4, 6, 10, 9
26. 下面的数据分列给出了穿过校园区和穿过闹市区的前 9 辆汽车的行驶速度 (英里/时)。
- 穿过校园区: 20, 18, 23, 21, 19, 18, 17, 24, 25
- 穿过闹市区: 29, 31, 35, 24, 31, 26, 36, 31, 28
27. 下列数据给出了美国前七位总统 (从华盛顿到杰克逊) 和最近七位总统 (从尼克松到布什) 的在宣布就职时的年龄。
- 前七位总统: 57, 61, 57, 57, 58, 57, 61
- 后七位总统: 56, 61, 52, 69, 64, 46, 54
28. 下列数据给出了贝多芬的 9 首交响曲和马勒的 9 首交响曲近似的时间长度 (以分钟计)。
- 贝多芬的交响曲: 28, 36, 50, 33, 30, 40, 38, 26, 68
- 马勒的交响曲: 52, 85, 94, 50, 72, 72, 80, 90, 80
29. 比萨饼的送货 在记录了两家不同的比萨饼店的送货时间后, 你得出结论说, 一家比萨饼店的平均送货时间是 45 分钟, 标准差是 3 分钟; 另外一家比萨饼店的平均送货时间是 42 分钟, 标准差是 20 分钟。解释这两种不同的数字。如果你对两家比萨饼店的比萨饼同样喜欢, 你更愿意从哪一家比萨饼店订货? 为什么?
30. 处理顾客投诉 你经营一家小的冰激凌商店, 你的雇员手工挖取冰激凌。一天晚上, 你汇总了员工所卖的冰激凌总数和钱款总数。你发现一天晚上是一个名叫山姆的雇员在工作, 他销售每品脱冰激凌的平均售出价格是 1.75 美元, 标准差是 0.05 美元。另一天晚上是一个名叫凯文的雇员在工作, 他销售一品脱冰激凌的平均售出价格是 1.70 美元, 标准差是 0.35 美元。

哪一个雇员所销售的冰激凌更容易引起投诉？请加以解释。

31. 证券投资组合的标准差 兹维·博迪、亚历克斯·凯恩和艾伦·马库斯三个人所著的《投资》一书中，宣称证券投资组合中，单一股票收益的标准差是 0.55，32 只股票证券投资组合中其收益的标准差是 0.325。解释标准差是怎样测量了两种不同类型的证券投资组合的风险。
32. 安打率的标准差 在过去的 100 年里，棒球职业联盟的安打率持续稳定地保持在 0.260 的水平。尽管如此，安打率的标准差从 1870 年的 0.049 下降到现在的 0.031。这个能告诉我们关于棒球运动员安打率的什么方面？依据这个事实，你能否期望安打率超过 0.350 现在比过去更有可能出现？请加以解释。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 4 章的有用链接“links for internet projects”。

33. 二手烟 浏览美国肺脏学会和美国环境保护署的网页，寻找与二手烟（被动吸烟）对健康影响的统计数据。写出一份简短的总结，内容包括关于你的发现，以及针对健康主题是否和怎么样得到政府的处理。

34. 儿童和媒体 恺撒家庭基金会的一项近期的研究关注了媒体（例如电视、书籍和计算机）在儿童生活中的作用。这个报告发布在凯萨家庭基金会的网页上，给出了许多相关的分布数据。例如，儿童在各种媒体中所花费时间是多少。至少研究你感兴趣的报告中所提到的三种数据分布。用语言概括每一种分布，并且讨论这些发现的社会后果的观点。
35. 变异性的测量 全距和标准差采用不同方法测量了一组数据中的变异性。虚拟两组不同的数据分布，第一组数据的全距要大于第二组数据的全距（提示第一组数据的变异性更大），但是第一组数据的标准差要小于第二组数据的标准差（提示第一组数据均值变异性更小）。

新闻中的统计学

36. 新闻中的全距 寻找两篇近期新闻报道中的数据分布的例子，可以用图也可以用表呈现数据。在每一种情况下，说明分布的全距，并且解释它在新闻报道内容中的意思。应用全距的拇指规则去估计标准差。
37. 新闻数据的概括 在最近的一篇新闻报道中寻找用表格形式呈现的数据分布。应用五数概括法并且绘制分布的箱须图。



4.4 统计悖论

政府计划将测谎仪应用于敏感的安全工作新领域。一般认为测谎仪有 90% 的准确性，即测谎仪能发现 90% 的说谎者，也能肯定 90% 说真话的人。但通不过测谎仪的人中有多少其实说的是真话？大多数人猜测通不过测谎仪的人大约 10% 是错误识别。但是实际上错误判断的识别率要比这个高得多。这是怎么回事呢？

我们稍后再讨论这个回答，但是这个故事道德意味是非常清楚的。其实我们看到本章的前 3 节所讨论的基本原理都是基于认真的描述数据的，我们也可能会得出令人惊讶的结论。在我们讨论测谎仪问题之前，先让我们从一些令人惊讶的统计学问题开始。

4.4.1 每个组成部分都是完美的，但是合成总体后却很差

假设一个医药公司生产一种治疗痤疮的新药，为了检验新药是否比旧药的疗效好，公司对 90 个患者应用旧药，110 个患者使用新药。一些患者痤疮较轻，另外一些患者痤疮非常严重。表 4-6 呈现了 4 周治疗以后的结果。按照给予的治疗是旧药还是新药以及患者痤疮是较轻还是严重进行分类，如果你仔细地研究下表，你将会注意到以下关键事实。

- 在所有痤疮较轻的患者中：
 - 10 位患者接受旧药治疗，2 名治愈，治愈率是 20%。
 - 90 位患者接受新药治疗，30 名治愈，治愈率是 33%。
- 在所有痤疮严重的患者中：
 - 80 位患者接受旧药治疗，40 名治愈，

治愈率是50%。

20位接收新药治疗的患者,12名治愈,治愈率是60%。

表4-6 新旧两种药痤疮治疗的效果

	痤疮较轻		痤疮严重	
	治愈	没有治愈	治愈	没有治愈
旧药	2	8	40	40
新药	30	60	12	8

请注意无论是较轻的痤疮患者(新药的治愈率是33%,老药的治愈率是20%)还是严重的痤疮患者(新药的治愈率是60%,老药的治愈率是50%),新药都有较高的治愈率,公司宣称他们的新药要比旧药要好,这样是公平合理的吗?

首先,这个宣称看起来是有意义的,但是如果不再分别看较轻的患者和严重的患者,让我们看总体的结果。

- 总数为90名患者接受旧药治疗,42名治愈(10名较轻患者中2人治愈,80名较重患者中40名治愈),所以总体治愈率是 $42/90=46.7\%$ 。
- 总数为110名患者接受新药治疗,42名治愈(30名较轻患者中30治愈,20名较重患者中12名治愈),所以总体治愈率是 $42/110=38.2\%$ 。

总体来说,旧药的治愈率更高,尽管新药在较轻的患者中和严重的患者中治愈率都高于旧药。

这个例子表明,分成两组或者多组进行比较表现更好的那一种情况,合成总体比较时反而表现更差些,类似这种情况是有可能发生的。如果你仔细观察,你会发现这种情况的发生是因为总体是否被分成了大小不相同的组(在本案例中,痤疮较轻的患者和痤疮严重的患者)。

相关知识

一组数据分成几组进行比较与合在一起的结果不一样,这就是著名的辛普森悖论。这样命名是由于爱德华·辛普森在1951年描述的。但是,同样的观点早在1900年,苏格兰统计学家乔治·尤尔就描述过。

例4-17 谁投篮更准

表4-7给出了两个运动员在一场篮球比赛中的上下半场各自的投篮情况,沙奎尔·奥尼尔在上半场(40%对25%)和下半场(75%对70%)投篮率的

百分数都较高,你能说全程比赛中沙奎尔·奥尼尔投篮率更准吗?

解答 不是的,我们可以通过整体比赛的统计看到这是为什么。沙奎尔·奥尼尔全程比赛14投(上半场10投,下半场4投),7中(上半场投中4个,下半场投中3个),总体命中率是 $7/14=50\%$ 。文斯·卡特在全程比赛中14投(上半场4投,下半场10投),8中(上半场投中1个,下半场投中7个),总体命中率是 $8/14=57.1\%$ 。令人惊奇的是,即使沙奎尔·奥尼尔上半场和下半场的命中率都高于文斯·卡特,但是文斯·卡特在整场比赛的总体命中率却更高。

表4-7 投篮命中率

球员	上半场			下半场		
	投中	投篮	百分数	投中	投篮	百分数
沙奎尔·奥尼尔	4	10	40%	3	4	75%
文斯·卡特	1	4	25%	7	10	70%

4.4.2 一个乳房X线照片的阳性是否意味着癌症

我们经常将肿瘤和癌症联系起来,但是大多数肿瘤并不是癌症。在医学上,任何一种不正常的隆起或者组织增生都可以看做一个肿瘤。由癌症引发的肿瘤被称做是恶性的,其他的叫做良性的。

设想你是一个医生和护士,正在治疗一个胸部长了肿瘤的患者。患者因为不理解而显得非常紧张。你可以给她一些安慰,告诉她100位胸部肿瘤患者中,只有一位可能是恶性的。但是为了保险起见,你建议她去作一个乳房X线照片,以确定她的肿瘤是否属于那1%恶性的。

现在,假定乳腺X光照片的检测报告是阳性的,提示肿瘤是恶性的。乳腺X光照片不是百分之百准确的,这个阳性的检测报告并不意味着你的患者一定是乳腺癌。更进一步说,让我们假定乳腺X光照片检测的准确率是85%:它能正确判定85%的恶性肿瘤是阳性的,也能正确判断85%的良性肿瘤是阴性的。当你告诉你的患者她的乳腺X光照片是阳性的,那么应该如何告诉她患癌症的实际概率是多少。

相关知识

乳腺X光照片和测谎仪的例子说明条件概率容易混淆。正确掌握条件概率的方法是由贝叶斯(Reverend Thomas Bayes, 1702—1761)发现的,经常被叫做贝叶斯法则。

因为乳腺 X 光照片检测准确率是 85%，大多数人会猜测阳性的结果意味着患者很可能患的是癌症。

研究表明大多数医生也是这样认为的，他们会告诉患者准备癌症治疗。但是更为仔细的分析会告诉我们不同的结论。实际上患者得癌症的机会是相当小的，大约只是 5%。我们通过分析一些数字来看下为什么是这样。

思考下面这样一个研究，对 10 000 名长有胸部肿瘤的妇女进行乳腺 X 光照片检测，假定 1% 的肿瘤是恶性的，实际上只有 $1\% \times 10\,000 = 100$ 个妇女真的是癌症。其余的 9 900 名妇女只不过是良性肿瘤。表 4-8 呈现了乳腺 X 光照片的检测结果。注意以下几点：

(1) 乳腺 X 光照片能够正确识别 100 个恶性肿瘤的 85%。这样，它能判定 85 个恶性肿瘤是阳性的，这些叫做真阳性 (true positives)，另外 15 例恶性肿瘤，检测结果是阴性的，即使这些妇女实际上真的患有癌症，这些叫做假阴性 (false negatives)。

(2) 乳腺 X 光照片能够正确识别 9 900 名良性肿瘤中的 85%，这样，它能判定 $85\% \times 9\,900 = 8\,415$ 个良性肿瘤是阴性的，这些叫做真阴性 (true negatives)，其余的 $9\,900 - 8\,415 = 1\,485$ 个妇女，检测结果是阳性的，但乳腺 X 光照片却是不正确的，这些叫做假阳性 (false positives)。

总体来说，乳腺 X 光照片检测得到的阳性结果是下面这样的，85 个实际上患有癌症的妇女和 1 485 名并没有患癌症的妇女，她们的检测结果是阳性的。阳性结果的总数是 $85 + 1\,485 = 1\,570$ 。因为其中仅仅只有 85 名是真的阳性 (其余的是假阳性)，阳性的结果意味着是真的癌症的机会，仅仅是 $85/1\,570 = 0.054$ ，或者是 5.4%。由此看来，当你的患者乳腺 X 光照片检测结果是阳性的时候，应该消除她的疑虑，她患癌症的机会仍然是很小。

例 4-18 假阴性

假定你是一个医生，看到一个患者胸部长了肿瘤。她的乳腺 X 光照片检测报告是阴性的，依照表 4-8 中的数字，她患癌症的概率是多大？

表 4-8 10 000 名妇女案例的概括 (实际上 100 例肿瘤是恶性的，9 900 例肿瘤是良性的)

	肿瘤是良性的	肿瘤是恶性的	合计
阳性检测结果	85 真阳性	1 485 假阳性	1 570
阴性检测结果	15 假阴性	8 415 真阴性	8 430
总计	100	9 900	10 000

解答 依照表 4-8 概括的 10 000 名妇女的案例，乳腺 X 光照片是阴性的，其中 15 名妇女真的患有癌症，而 8 415 名妇女的肿瘤是良性的，阴性结果的总数是 $15 + 8\,415 = 8\,430$ 。这样，在阴性检测结果中患有癌症的妇女的分数是 $15/8\,430 = 0.0018$ ，或者说，在 1 000 个人中少于两人。换句话说：乳腺 X 光照片检测结果是阴性的妇女，实际上患有癌症的概率是非常少的，每 1 000 人中大约有两人。

思考题

尽管乳腺 X 光照片检测结果是阴性的妇女患有癌症的概率是相当少，但它并不是 0。这样来说，为保险起见，得对所有的肿瘤作生物活检看起来是一个好主意，然而，生物活检是一项外科手术，这意味着是疼痛且花费比较昂贵，基于这些事实，你认为对所有肿瘤进行生物活检应该是例行工作吗？说明你的观点。

4.4.3 测谎仪和药物实验

现在我们回到在这一节开始所问的问题，关于一个 90% 准确度的测谎仪，如何能导致一个令人惊讶的不实指控的数字。

这里的解释与乳腺 X 光照片所用到的是非常一致的。假定政府对 1 000 名申请敏感的安全工作的人进行测谎仪测试。更进一步假定 1 000 人中的 990 人在进行测谎仪测试的时候说的是真话，只有 10 个人说谎话。测谎仪的准确度是 90%，这样我们会得到以下结果：

- 对于 10 位说谎的人来说，测谎仪正确识别率是 90%，这意味着其中 9 人未通过 (他们被识别出是说谎者)，另外一个人则通过。
- 对于 990 位说真话的人来说，测谎仪的正确识别率是 90%，这意味着 $90\% \times 990 = 891$ 位说真话的人通过测谎仪的测试，另外有 $10\% \times 990 = 99$ 人不能通过测谎仪的检测。

图 4-16 概括了上述结果，这样通不过测谎仪检测的总人数是 $9 + 99 = 108$ ，这里面只有 9 人是真正说谎的，其他 99 人并没有说谎，也就是说，108 人中的 99 人，或者 $99/108 = 91.7\%$ ，这些人没有通过测谎仪的检测但他们实际上说的是真话。

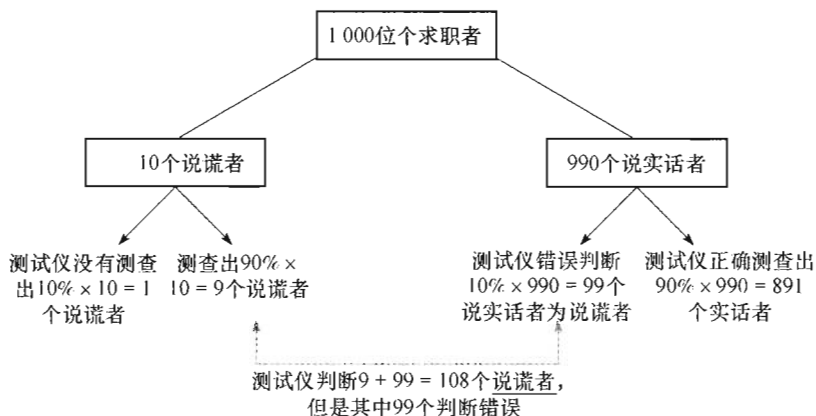


图 4-16 使用准确性为 90% 的测试仪对其中只有 10 人说谎的 1 000 人进行测验的结果的一个概括树形图

在实际情况下，被不实指控的人数比率依赖于检测准确度和说谎人数的比率。虽然如此，基于给出的这些数字，我们得到了一个令人惊讶的结果：假如政府拒绝了不能通过测谎仪检测的人们的申请，那么，这些被拒绝的申请中大约 92% 的人说的是真话，他们是很有资格获得这项工作的。

相关知识

乳癌风险检查的准确性正在快速提高，新的技术包括数字乳腺 X 光照片和超声波，检测的准确率接近 98%。最清晰的乳腺癌检测是生物活检技术，当然机械式生物活检技术如果操作不细的话也可能会误诊。如果你的检测结果是阴性但仍然感觉有些不正常，最好再做第二次检测。它会挽救你的生命。

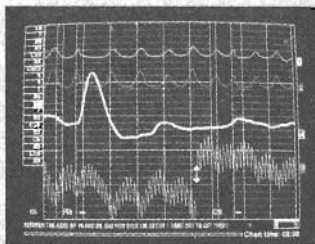
思考题

假如由于不实指控，你被卷入一项犯罪案件。警方提议说，如果你确实是无辜的，你应该同意进行一项测谎仪检测。你将要怎么做呢？为什么？

相关知识

多导生理仪，又叫“测谎仪”，测量包括心率、皮肤温度、血压等多种身体功能，测谎仪的操作者寻找人们说谎时这些身体功能的微妙变化。然而，测谎仪的结果并不能在法庭中作为证据。首先，90% 的准确率对法律裁决来说太低了，另外，

研究显示，测试仪容易被受过训练的人所欺骗。



例 4-19 高中生的药物检测

参加地区高中生田径锦标赛的所有运动员必须提供一个尿样进行兴奋剂检测。那些尿样检测呈阳性的运动员将会被取消本次比赛的资格并且在下一年禁止参加比赛。研究显示，在指定的药物检测实验室，检测的正确率是 95%。假如运动员中有 4% 实际上是使用兴奋剂的。通不过药物检测的运动员有多少比率是属于不实指控，实际上他们并没有服用兴奋剂。

解答 回答这个问题最简单的方法是应用一些样本数据。假如在这次比赛中有 1 000 名运动员参加。其中 4%，或者说 40 名运动员，实际上是使用了兴奋剂的，而其余的 960 名运动员没有服用药物，在这种情况下，准确率为 95% 的药物检测将会得到下列结果：

- 40 名服用兴奋剂的运动员，其中 95%，即 $0.95 \times 40 = 38$ 名运动员，通不过药物检测，另外两名运动员虽然使用了兴奋剂但是通过了药物检测。
- 960 名没有使用药物的运动员，其中 5%，即 $0.05 \times 960 = 48$ 名运动员，不能通过药物检测。

这样，通不过药物检测的运动员的总数是 $38 + 48 = 86$ ，但是其中 48 名运动员，虽然不能通过药物检测但是确实没有服用药物。尽管药物检测率的准确度是 95%，然而有超过一半的被禁赛的运动员是无辜的，他们并没有服用药物。

4.4 练习题

统计素养和批判性思维

- 1. 假阳性和假阴性 当某人接受药物使用测试时，什么是假阳性？什么是假阴性？
- 2. 阳性测试结果 简略地解释一下，为什么癌症检查比如乳房 X 线照片得出的阳性测试结果，并不能说明一个人一定患了癌症。
- 3. 测试准确性 测谎仪检测准确性非常高但同时也会导致很大比例的错误报告，这个可能性是如何产生的？
- 4. 上下半场得分率都高，而整场得分率低 当两个橄榄球队进行比赛时，一个 4 分卫球员能否做到在上下半场中传球率都较高，同时又在整场比赛中传球率较低呢？

判断正误：练习题 5~8，判断其表述是否正确。并做详细解释。并不是所有这些表述都有明确的答案，所有你的解释比你选择的答案更加重要。

- 5. 课程平均分 安和布雷特都选择了相同的统计学课程，该课程的最后成绩由作业和测验来判定。在作业方面，安的平均分高于布雷特的平均分。在测验成绩上，安的平均分也高布雷特的平均分。所以得出结论安在该课程的总体平均分高于布雷特。
- 6. 安打率 安在赛季的上半部分安打率比布雷特的安打率高，在整个赛季下半部分的安打率也比布雷特高。由此得出结论安在整个赛季的安打率比布雷特的安打率高。
- 7. 测试结果 如果你的测试结果是阳性，那么你感染链状球菌的概率就等同于你已经感染后测试结果呈阳性的概率。
- 8. 测试准确性 如果药物测试的准确性是 90%，那么也就是说，在测试结果呈阳性的受测者中，有 90% 的人确实是药物使用者。

概念与应用

- 9. 安打百分比 下面的表格记录了两名棒球运动员在一个赛季的上半部分和下半部分的安打得分。谁在上半部分的安打得分率较高？谁在下半部分的安打得分率较高？谁的总体安打得分率较高？并说明这些结果如何解释辛普森悖论。

运动员	上半个赛季		
	安打数	自由击球数	安打率
约什	50	150	0.333
尤德	10	50	0.200

运动员	下半个赛季		
	安打数	自由击球数	安打率
约什	35	70	0.500
尤德	70	150	0.467

- 10. 传球率 下面的表格给出了两名 4 分卫球员在一场橄榄球赛上半场和下半场的传球情况记录。在上半场比赛中谁传球成功的百分比比较高？在下半场比赛中谁的传球成功的百分比比较高？整场比赛中谁传球成功的百分比比较高？并说明这些结果如何解释辛普森悖论。

运动员	上半场		
	传球成功数	传球数	百分数
艾伦	8	20	40%
艾伯纳	2	6	33%

运动员	下半场		
	传球成功数	传球数	百分数
艾伦	3	6	50%
艾伯纳	12	25	48%

- 11. 测试成绩 下表给出了内布拉斯加州和新泽西州的八年级学生数学测试的成绩。成绩按照学生的族裔分类，表中也给出了所有族裔学生的州平均分。

	白种人	非白种人	所有族裔平均
内布拉斯加州	281	250	277
新泽西州	283	252	272

资料来源：National Assessment of Educational Progress, from *Chance* magazine.

- (1) 哪个州在两个族裔中都有较高得分？哪个州在所有族裔中有较高的总体平均分？
- (2) 说明为什么一个州在每个族裔分类中的得分都较低，却仍然有较高的总体平均分。
- (3) 思考下表，表中给出了每个州白种人和非白种人的百分比，用这些百分比来证实内布拉斯加州的总体平均分是 277。

	白种人	非白种人
内布拉斯加州	87%	13%
新泽西州	66%	34%

- (4) 运用族裔百分比来证实新泽西州的总体平均分是 272, 与本题中第一个表中所呈现的是一致的。
- (5) 用你自己的话简要解释一下, 在这个例子中是如何体现辛普森悖论的。
12. 测验成绩 思考下表, 比较高中生在 1988 年及 1998 年的 GPAs 等级平均绩点和对应的 SAT 数学成绩。

GPA	学生百分比		SAT 成绩		变化
	1988 年	1998 年	1988 年	1998 年	
A +	4	7	632	629	-3
A	11	15	586	582	-4
A -	13	16	556	554	-2
B	53	48	490	487	-3
C	19	14	431	428	-3
总平均分			504	514	+10

资料来源: Cited in *Chance*, Vol. 12, No. 2, 1999, from data in *New York Times*, September 2, 1999.

- (1) 一般情况下, 以五等级进行分类的每组学生 SAT 成绩在 1988 ~ 1998 年是如何变化的?
- (2) 从 1988 ~ 1998 年, SAT 成绩的总体平均分是如何变化的?
- (3) 这个例子如何解释辛普森悖论?
13. 肺结核死亡率 下表给出了 1910 年纽约市和弗吉尼亚州里士满市的肺结核死亡率。

种族	纽约市	
	人口	肺结核死亡人数
白种人	4 675 000	8 400
非白种人	92 000	500
总体	4 767 000	8 900

种族	弗吉尼亚州里士满市	
	人口	肺结核死亡人数
白种人	81 000	130
非白种人	47 000	160
总体	128 000	290

- (1) 计算出纽约白色人种的死亡率、非白色人种的死亡率以及纽约所有居民的肺结核死亡率。
- (2) 计算出里士满市白人的死亡率、非白

色人种的死亡率以及里士满所有居民的肺结核死亡率。

- (3) 解释这个例子如何体现辛普森悖论以及悖论是如何产生的。
14. 重量训练 两个越野赛跑队参与了一项研究(假设的), 在研究中, 每个队有一部分人接受重量训练作为跑步练习的补充。另一部分队员则不接受重量训练。到赛季结束, 赛跑时间的平均提高水平在下表加以记录。

	平均提高水平(秒)		
	重量训练	无重量训练	全队平均
瞪羚队	10	2	6.0
猎豹队	9	1	6.2

- 说明辛普森悖论在表中如何体现的, 通过计算每个队接受重量训练的百分比来解决悖论。
15. 篮球比赛记录 考虑下表所给出的斯佩尔曼学院和莫哈斯学院的篮球比赛记录(假设的)。

	斯佩尔曼学院	莫哈斯学院
主场	10 胜, 19 负	9 胜, 19 负
客场	12 胜, 4 负	56 胜, 20 负

- (1) 请给出数字依据来支持以下的说法, 斯佩尔曼学院的篮球队水平要高于莫哈斯学院篮球队。
- (2) 请给出数字依据来支持以下的说法, 莫哈斯学院的篮球队水平要高于斯佩尔曼学院篮球队。
- (3) 你认为哪一种说法是正确的? 为什么?
16. 更有效的药物 两种药物 A 和 B, 给 2 000 名病人进行测试, 2 000 名病人中男女各占一半。900 名病人接受药品 A 治疗, 1 100 名病人接受药品 B 治疗, 测试结果如下表。

	男性	女性
药物 A	100 个患者 5 个治愈	800 个患者 400 个治愈
药物 B	900 个患者 11 个治愈	200 个患者 196 个治愈

- (1) 给出数字依据来证实药物 B 比药物 A 更有效。
- (2) 给出数字依据来证实药物 A 比药物 B 更有效。
- (3) 你认为哪一种说法是正确的? 为什么?
17. 测谎器测验 假设一台测谎仪的准确性是

90%（它能准确地检测出 90% 说谎的人，同时也能正确地检测出 90% 的人说的都是事实）。一家公司的 2 000 名员工接受了测试，他们被要求回答是否使用了药物。事实上，有 1% 的员工确实服用药物，而他们当时都拒绝承认。假设测谎仪操作员发现一个人不诚实，那么此人就被认定为说谎。

(1) 核实下表中的每个数据都是从所给信息中而来的，解释每个数据。

	服用 药物者	没有服用 药物者	总计
测谎仪发现 员工说谎话	18	198	216
测谎仪检测 员工说真话	2	1 782	1 784
总计	20	1 980	2 000

(2) 有多少员工被指控为说谎？在这些员工中，有多少人是确实在说谎，有多少人说的是实话？在被指控为说谎的人当中，被错误指控的百分比是多少？

(3) 有多少员工被认为是诚实的？在这些员工中，有多少人确实是在说实话？在被认定为诚实的员工当中，被正确断定的百分比是多少？

18. 疾病测试 假设一项疾病测试能准确检测出 80% 的患病患者（真阳性）和 80% 没有患病的人（真阴性）。在一个有 4 000 名患者的样本中，这种疾病的发病率与全国平均水平相吻合，为 1.5%。

(1) 核实下表中的每个数据都是从所给信息而来，并证实其总体发生率为 1.5%，解释说明。

	患病	没有患病	总计
检验阳性	48	788	836
检验阴性	12	3 152	3 164
总计	60	3 940	4 000

(2) 在患病人群中，测试结果呈阳性的百分比是多少？

(3) 在测试结果呈阳性的人群中，患有此病的患者百分比是多少？将此结果与第 2 小题中的结果相比较，并说明为什么两者存在差异。

(4) 假设一位病人的测试结果呈阳性，当

一位医生以此表中的数据为依据，你如何说明该病人实际患有此病的概率？将此数据与该病的总体发病率作比较。

进一步应用

19. 雇用统计学 一家公司计划扩大规模，所以开了一间工厂，有 455 个职位。其中的 70 个白领职位就有 200 名男性和 200 名女性申请。在所有女性求职者中，有 20% 被雇用，而在男性求职者中只有 15% 被雇用。在申请蓝领职位方面，400 名男性求职者中有 75% 被录用，100 名女性求职者中有 85% 被录用。如何看出在白领和蓝领职位上，两者对于雇用女性的偏好？全部的数据是否能说明该公司更偏好于雇用女性？解释说明该例子如何体现辛普森悖论并阐明悖论该如何解决。

20. 药物试验 一家公司对两种疾病进行了治疗试验。在第一个试验中，A 疗法治愈了 20% 的病例（200 个病例中的 40 个），而 B 疗法治愈了 15% 的病例（200 个病例中的 30 个）。在第二个试验中，疗法 A 治愈了 85% 的案例（100 个中的 85 个）而疗法 B 治愈了 75% 的案例（400 个当中的 300 个）。在分别的两次试验中，哪种治疗方法拥有更高的治愈率？哪种疗法拥有更高的总体治愈率？解释说明该例子如何体现辛普森悖论，并阐明悖论该如何解决。

21. 艾滋病风险 纽约卫生署估计在高危人群中的艾滋病（HIV）发病率为 10%，在一般人群中艾滋病风险为 0.3%。艾滋病测试在诊断真阴性和真阳性两方面的准确率都为 95%。下表给出了随机选取的 5 000 名高危人群和 20 000 名一般人群的测试结果。

	高危人群	
	检验阳性	检验阴性
感染	475	25
未感染	225	4 275
	一般人群	
	检验阳性	检验阴性
感染	57	3
未感染	997	18 943

(1) 证实一般人群和高危人群的艾滋病发病率分别为 0.3% 和 10%，证实艾滋病测试对于高危人群和一般人群的检测准确率达到了 95%。

- (2) 关注高危人群。在患有艾滋病的人群中，测试结果为阳性的百分比是多少？在测试结果为阳性的人群中，真正患有艾滋病的百分比是多少？解释说明为什么两者的百分比不同。
- (3) 假设在高危人群测试中有一位病人的测试结果为阳性。以此表中的数据为依据，请问如何说明该病人实际患有此病的概率？将此数据与该病的总体发病率作比较。
- (4) 关注一般人群。在患有艾滋病的人群中，测试结果为阳性的百分比是多少？在测试结果为阳性的人群中，真正患有艾滋病的百分比是多少？解释说明为什么两者的百分比不同。
- (5) 假设在一般人群测试中有一位病人的测试结果为阳性。以此表中的数据为依据，请问如何说明该病人实际患有此病的概率？将此数据与该病的总体发病率作比较。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第4章的

有用链接“links for internet projects”。

22. 测谎仪测试的争论 浏览网上关于支持或反对测谎测试运用的观点。总结概括两方的观点，尤其关注假阴性比例在讨论中的作用。
23. 药物测试 探究在工作环境或田径比赛中的药物测试问题。讨论这些情境下药物测试的合法性以及目前普遍实施的测试的准确性。
24. 癌症筛检 调查关于某些类型癌症的例行检查的一些建议（比如，乳腺癌、前列腺癌或结肠癌）。解释癌症检查的准确率是如何测量的。检测是如何起作用的？检测结果又是如何误导我们的？

新闻中的统计学

25. 测谎仪测试 找出近期的一篇文章，文章中指出测谎测试能判断一个人是否诚实。根据你所了解的测谎测试，你认为测试结果有意义吗？为什么？
26. 药物测试及运动员 找出一篇关于药物测试及运动员的新闻报道。概括药物测试是如何进行的，并讨论药物测试是否可靠。

第4章复习题

1. 以下为随机挑选的 M&Ms 巧克力豆的重量（克）。

红色 M&Ms 巧克力豆：

0.751 0.841 0.856 0.799 0.966

0.859 0.857 0.942 0.873 0.809

0.890 0.878 0.905

绿色 M&Ms 巧克力豆：

0.925 0.914 0.881 0.865 0.865

1.015 0.876 0.809 0.865 0.848

0.940 0.833 0.845 0.852 0.778

0.814 0.791 0.810 0.881

- (1) 分别求出两列数据的平均数和中位数。
- (2) 分别求两列数据的全距及标准差。
- (3) 分别给出两列数据的五数概括法并绘制箱须图。
- (4) 运用全距法则分别估算两列数据的标准差，该法则如何很好地体现？简单讨论原因。
- (5) 根据你计算的所有结果，比较并讨论两列数据的集中趋势和离散趋势。红

色 M&Ms 巧克力豆的重量和绿色 M&Ms 巧克力豆的重量之间是否存在显著差异？

2. 针对第1题的数据样本并作如下计算：
 - (1) 重量为 0.845 克的巧克力糖果的百分位数。
 - (2) 众数。
3. (1) 一个样本容量为 50 的样本，如果所有的值是相同的话其标准差是多少？
 - (2) 下面两种汽车电池，你会选择哪一种？为什么？

A：平均使用寿命为 48 个月，标准差为两个月。

B：平均使用寿命为 48 个月，标准差为 6 个月。
 - (3) 如果样本容量为 50 的样本中包括一个异常值，这个异常值对样本平均数会有什么影响？
 - (4) 如果样本容量为 50 的样本中包括一个

异常值，这个异常值对样本中位数会有什么影响？

- (5) 如果样本容量为 50 的样本中包括一个异常值，这个异常值对样本的全距会有什么影响？

什么影响？

- (6) 如果样本容量为 50 的样本中包括一个异常值，这个异常值对样本标准差会有什么影响？

第4章小测验

- 当你将 3, 5, 8, 12, 20 相加，再除以数列的个数，其结果为 9.6。下面哪个是对该数值的最佳说明：平均数、均值、中位数、众数、标准差？
- 为测量样本的离散程度，应该运用下面哪种统计值：平均数、均值、中位数、众数、标准差？
- 当全距法则被用于计算样本的标准差时，其结果总是标准差的真值吗？
- 对于一组数据值，下面哪种说法是正确的？
 - 这组数据没有众数。
 - 这组数据有两个众数。
 - 这组数据有 3 个众数。
- 指出下列说法是否适用于一组由 1 000 个不相同的数值组成的数据。
 - 第 20 百分位数比第 30 百分位数大。
 - 中位数比第一四分位数大。
 - 第三四分位数比第一四分位数大。
 - 平均数与中位数相等。
 - 全距为 0。
- 设计的一项关于焦虑水平的标准化测试，其平均数为 50，标准差为 10。按照全距的拇指法则，其最低值和最高值可能是多少？
- 一系列数据为 2, 4, 5, 8, 10，运用全距法则计算标准差。
- 一组数据是 5.8, 5.8, 5.8, 5.8, 5.8，其标准差是多少？
- 一组数据是 2.0, 3.7, 4.9, 5.0, 5.7, 6.7, 8.5, 9.0，全距是多少？
- 确定一组数据的五数字概括法的结果。

技术应用

许多电脑软件程序允许你输入一组数据并运用某种操作来得出一些不同的样本统计量，这些被称为描述性统计。描述性统计通常包括平均数、中位数、全距和标准差。下面的一些程序就能得出这样的结果。

下列程序的重要注意事项：

- SPSS 和 Excel 提供“改进的”箱须图，图表能显示出代表异常值的特殊点，这些异常值与其他数值相差甚远。
- 关于计算四分位数，没有要求说必须使用某一种特定的程序，但是不同的计算机程序计算结果会有所差异。例如，如果你使用一组数据 1, 3, 6, 10, 15, 21, 28, 36 计算四分位数，你会得到以下结果：

	Q ₁	Q ₂	Q ₃
SPSS	3.75	12.5	26.25
EXCEL	5.25	12.5	22.75
STATDISK	4.5	12.5	24.5
Minitab	3.75	12.5	26.25
SAS	4.5	12.5	24.5
T1-83/84 PLUS	4.5	12.5	24.5

SPSS

描述性统计：首先输入一组数据值，或打开已有的数据文件。点击 [Analyze]，选择 [Descriptive Statistics]，再选择 [Frequency]。双击左侧的变量或数据列。点击 [Statistics] 键，继续选择你想要的与统计量相对应的按钮。你可以选择四分位数、标准差、全距、最小值、最大值、平均数、中位数，等等。单击 [Continue] 按钮，再单击 [OK] 按钮，就可以得到已选择的统计值。

箱须图：输入或打开数据集。单击 [Graphs] 并选择 [Boxplot]。在 Boxplot 窗口中，选择 [Summaries of separate variables] 和 [Simple]，再单击 [Define] 按钮。之后选择左侧的变量或数据列，单击窗口中间的按钮，最后点 [OK]。

Excel

描述性统计：Excel 的数据分析插件必须先安装。如果程序没有出现数据分析窗口，你就必须安装数据分析插件。如果你使用的是 Excel

2003 版, 单击 [Help], 输入 [Data Analysis Toolpak], 再按 [ENTER] 键。选择 [Load or unload add-in programs] 选项, 然后会显示使用说明。如果你使用的是 Excel 2007 版, 就点击问号图标, 然后输入 [Data Analysis] 并选择 [Load the Analysis Toolpak] 选项, 然后会出现使用说明。

在 A 列输入样本数据。单击 [Tools] 再选择 [Data Analysis]。如果你使用的是 Excel 2007 版, 就单击 [Data] 再单击 [Data Analysis]。在出现的窗口中, 选择 [Descriptive Statistics], 再单击 [OK]。在对话框中输入全距, 选择 [Summary Statistics], 再单击 [OK]。描述性统计包括均值、中位数、标准差、全距、最小值和最大值。

箱须图: 尽管 Excel 原本并不能生成箱须图, 但可以运用本书补充的 DDXL 插件来实现。

| 聚焦股票市场 |

什么是道琼斯平均工业指数

随着“股指”开始流行, 它就是很杰出的一个。你可以在新闻中随时听到它的最新动态, 很多人每天都花大量时间去追踪它的走势。它也是到目前为止衡量股票市场表现的最著名指数。当然了, 我们说的正是道琼斯工业平均指数, 缩写为 DJIA。但是它到底是什么呢?

了解道琼斯指数最简单的方法就是回顾它的历史。在 19 世纪末, 随着现代工业时代的来临, 很多人都将股票看做一项风险很强的投机性很高的投资, 一个原因是缺乏必要的规则来监督富有的投机商、缺乏职业道德的经理人及公司的蓄意收购者来操纵股票价格, 但是还存在另外一个原因, 由于每天股票交易很复杂, 即使华尔街的从业者也很难说出股票市场是涨(牛市)是跌(熊市)。查尔斯 H. 道琼斯, 《华尔街日报》的创建者(与爱德华 D. 琼斯一起)及其第一任主编, 他相信自己可以设置一种反映整个股票市场的“平均指数”来解决这个问题。如果指数上升, 股票市场涨; 指数下降, 股票市场跌。

为使道琼斯指数简单些, 道琼斯选择了 12 家大公司进入他的指数。1896 年 5 月 26 日, 他将 12 家大公司的股票价格取平均值, 计算出平均股价为 40.94 美元。这就是最早的道琼斯指数。正如道琼斯所希望的, 只需要不断比较他的指数, 公众就可以容易地把握市场的方向。

虽然道琼斯指数的股票名单上现在已经是 30 家而不是当初的 12 家(见表 3-5), 但是它背后的基本理念还是一致的, 而且道琼斯指数也不再是计算这 30 家公司的平均股票价格, 而是把这 30 家股票的价格除以一个特别的因子数, 由于这个因子数, 我们现在将道琼斯指数看做帮助我们跟踪股票价值的指数, 而不是一个实际的股票价格的平均值。

首先在 A 列输入数据。选择 [Charts and Plots], 在函数类型中选择 [Boxplot]。在对话框中选择铅笔图标, 输入数据全距。然后单击 [OK]。结果会出现一个改良的箱须图, 图中显示出了轻微异常值和极端异常值。五数概括法所计算的值也会得到。

STATDISK

描述性统计: 在数据窗口中输入数据, 单击 [Data], 选择 [Descriptive Statistics]。再选择 [Evaluate] 得到各种描述性统计量, 包括均值、中位数、全距、标准差和五数概括法的值(最小值, 第一四分位数, 中位数或第二四分位数, 第三四分位数, 最大值)。

盒状图: 在数据窗口中输入数据, 选择 [Data] 再选择 [Boxplot]。单击你想要包含在内的选项, 最后单击 [Plot]。



这个因子数是道琼斯指数为了描绘股票潜在价值时保持连续性而设计的，只要名单上的 30 家公司有所改变或者其中任何一家公司进行股票分割时，这个因子数就得改变。举个简单的例子，假设道琼斯指数只包含 2 只股票（而不是 30 个）：A 股票价格为 100 美元，B 股票价格为 50 美元。两个股票的平均价格是 $(\$50 + \$100)/2 = \$75$ ，现在，假定我们用 C 股票来代替 B 股票，而 C 股票的价格是 \$200，新的平均价格为 $(\$100 + \$200)/2 = \$150$ 。所以，只要替换名单上的 1 只股票，平均价格就从 \$75 提高到了 \$150。当股票名单改变时，为了保持道琼斯指数的稳定性，我们必须把新的平均价格 \$150 除以 2，这样一来，在名单改变前后道琼斯指数都是 75，我们却不再认为这个 75 还是以美元为单位的股票平均价格。

分析一下为什么股票分割也会改变因子数，依然假设道琼斯指数只包括两只股票：X 股票价格 \$100，Y 股票的价格 \$50，其平均价格是 \$75，现在假定将 X 股票每 1 股分割为 2 股，因此新 X 股票的每股价格是 \$50，这样两支股票价格都是 \$50，这样 X 股分割以后平均价格也变成了 \$50。换句话说，即便股票的拆分不影响公司的市值（它只影响了股份的数目和价格），我们依然发现平均价格从 \$75 掉到了 \$50。这种情况下，我们为了道琼斯指数的连续性，就必须把 50 除以 2/3（或者乘以 3/2），道琼斯指数在拆分前后仍然是 75。

这只是一些简单的例子，实际上每当股票名单变动或者发生股票分割时，真正的因子数总是会变化的，所以自从查尔斯 H. 道琼斯第一次用实际上的平均价格计算道琼斯指数以来，它已经变化了很多次了，最新的因子数都被刊登在每天的《华尔街日报》上。

鉴于如今的股票市场有将近 10 000 只上市的股票，因此用一个只有 30 只股票的样本就能反映整个市场的成交情况还是很杰出的。在今天，应用电脑很容易通过很多的其他方法来计算股市的“平均水平”，我们可以通过研究历史数据看出，道琼斯指数真的可以反映整体股票市场交易的大趋势。图 4-17 显示了道琼斯指数的历史数据。

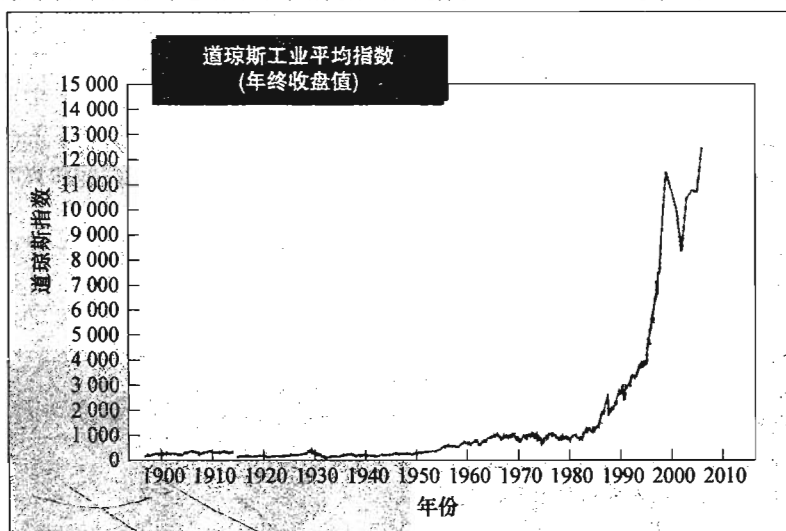


图 4-17 道琼斯指数自产生至 2006 年的每年年终指数值

如果你仔细研究图 4-17，你也许会得出这么一个结论：你可以通过观察图形走势来预测未来市场的价值。遗憾的是，没有人找到可以有效预测的方法，并且绝大多数经济学家认为这样的预测是不可能的。

被人称为“价值投资”之父的本杰明·格雷厄姆教授的事例向我们阐明了为什么这些预测是徒劳的。1951 年春天，他的一个学生向他寻求一些投资建议。格雷厄姆指出，现在道琼斯指数处于 250 点，但是它不久就会跌落到 200 点以下，因为自 1896 年起，道琼斯指数每年都会至少这么波动一次，由于 1951 年道琼斯指数还没有跌破过 200 点，所以格雷厄姆建议他的学生直到道琼斯指数跌破过 200 点之后再买进。格雷厄姆教授也许严格依照了他的预测进行投票操作，可是他的学生却没有这样。相反，他的学生当即往股市投资了“10 000 美元”。结果，股市从 1951 年至今，一直再也没有跌落回 200 点以下，而格雷厄姆的学生也早就成了亿万富翁，他就是沃伦·巴菲特。

问题讨论

1. 相比储蓄和债券，股市依然被认为是风险更大的投资方式。然而，现在的情况与 20 世纪初有很大不同，理财专家普遍会建议你至少持有几只股票。你认为道琼斯指数在建立投资者信心上面起着什么样的作用？
2. 道琼斯指数只是广泛应用的股票指数的其中一种。其他指数，比如 S&P500 指数、罗素 2 000 指数、纳斯达克

克指数,你认为这些指数和道琼斯指数有什么不同?它们是否比道琼斯指数更能有效地预测市场?请说出理由。

3. 道琼斯指数选用30只股票来反映超过10 000只上市股票的整体走势。但是,这不是随机选择的,而是由特定人员基于特定理由精心挑选的,所以某种程度上带有个人观点。假设随机选择30只股票,并且追踪他们的价格,你认为可以和道琼斯指数一样可靠么?请说出理由。
4. 列出包含你想拥有的10只股票的投资组合,假设你分别有100股,计算组合的市值,并且在接下来的一个月追踪股票价格,月底计算出你的股票组合的市值。比较这个月里你的组合与道指的市场表现。如果你能真的拥有上述股票,你会继续持有还是会抛售?请解释。
5. 图4-17显示了道琼斯指数自产生至2006年的每年年终指数值,找到从2006年以后到现在每年年终道琼斯指数值的数据,并添加进表格。从中找出自1995年到现在每年的道琼斯指数的最高点与最低点,并将每年的高点 and 低点绘制成单独的曲线。道琼斯指数的最高点和最低点与年终曲线图相比有什么特点?你认为年终指数数据可以追踪道指的历史变化吗?

建议阅读

1. 你可以在《华尔街日报》、《纽约时报》、*Money*、《商业周刊》和其他众多的期刊中寻找关于道琼斯指数定期的信息。
2. Morris, K. M., and Siegel, A. M. *The Wall Street Journal Guide to Personal Finance*. Fireside, 2000.
3. Prestbo, John (ed.). *The Market's Measure: An Illustrated History of America Told Through the Dow Jones Industrial Average*. Dow Jones & Company, 1999.

| 聚焦经济 |

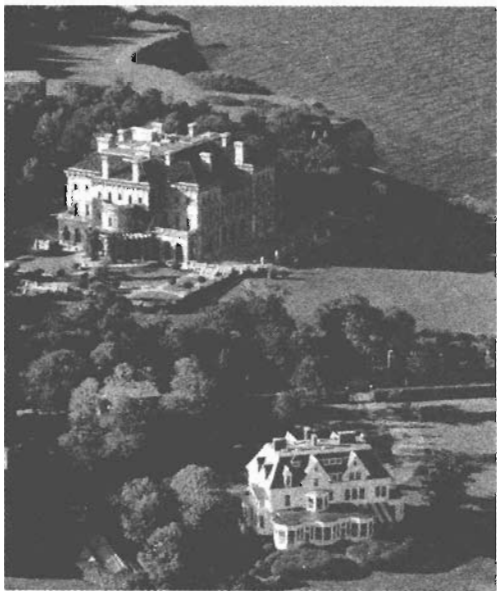
富人变得更富了吗

在“互联网泡沫”的最高峰时期,科技股的价格飙升到了一个前所未有的水平,微软的创始人比尔·盖茨以超过1 000亿美元的身家成为世界首富。除了最富裕的18个国家,比尔·盖茨的财富超过了全世界余下国家的国民生产总值之和。不久盖茨的身家就缩水了,一方面是由于微软股票价格的下跌,另一方面由于他把他所拥有的股份捐给了慈善机构。然而,媒体仍然在持续不断地曝光那些巨富的故事,似乎富人变得更富有了,而我们远远地被落在后面。事实是这样么?

如果我们希望得出一个关于普通人与富人生活水平比较的结论,那我们必须考虑全部收入的分配情况。经济学家创立了一个用来描述收入分配是否平等的数字,称做基尼系数。基尼系数值的范围被定义为从0~1:当基尼系数为0时,说明财富的分配是完全公平的;而基尼系数为1时,意味着完全不公平的,相当于一个人拿走了全部财富,而其他什么人什么也得不到。图4-18显示了1947年以来美国的基尼系数,注意从1947~1968年,整个基尼系数是在下降的,说明这段时期收入分配是越来越公平的,而从1968年到现在基尼系数则是不断上升,说明富人的确是更富有了。

虽然基尼系数提供了一个简单的数字来概括收入分配上的不平等,但是数字本身却是很难清楚解释(与计算的)。另一种可替代的收入分配的研究方法就是5分法,也就是按收入高低区分为5组人群。通常,在收入最高的一组中进一步区分出收入最高的5%的群体与其他人群进行比较。

图4-19显示了美国各个十年期以收入划分的5组人群及最富有的5%人群在全部收入中占的份额。直条图的高度(顶部标有数字)表示在全部收入中所占的份额。举例来说,收入最低的20%人群条形上的数字3.4意味着2005年这1/5的人群只享有全部收入的3.4%,同样,收入最高的20%群体直条上面50.4的数字意味着他们



享有全部收入的 50.4%，而其中最富的 5% 群体占有全部收入的 22.2%，几乎是贫穷端 40% 群体收入的两倍。如果你仔细研究这个条形图，你会发现，除了最富裕的 20% 的群体之外的全部群体，所享有的社会收入份额从 1967 年以来一直在下降，与此同时，最富裕的 20% 的群体所占的收入份额却在稳步增长，与那 5% 的最富有的群体一样。换句话说，这个条形图确认了这个事实：与普通公众相比，富人的确是变得更富了。

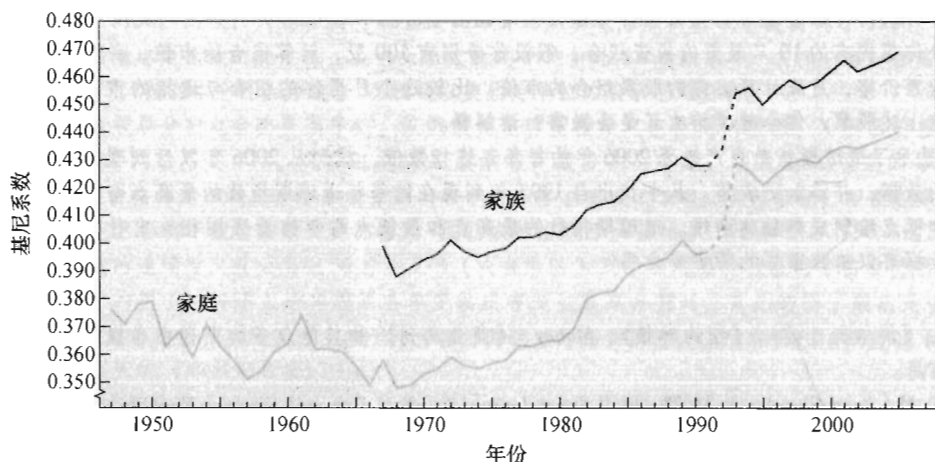


图 4-18 1947~2005 年家族和家庭的基尼系数。家庭数据包括单身的人以及住在一起但并不是家族成员的人，家庭数据从 1967 年开始采集。图中 1993 年处的虚线是指数据采集方法改变了，因此基尼系数的上升可能部分或全部归因于此，而不是收入分配越来越不平等了

资料来源：U. S. Census Bureau.

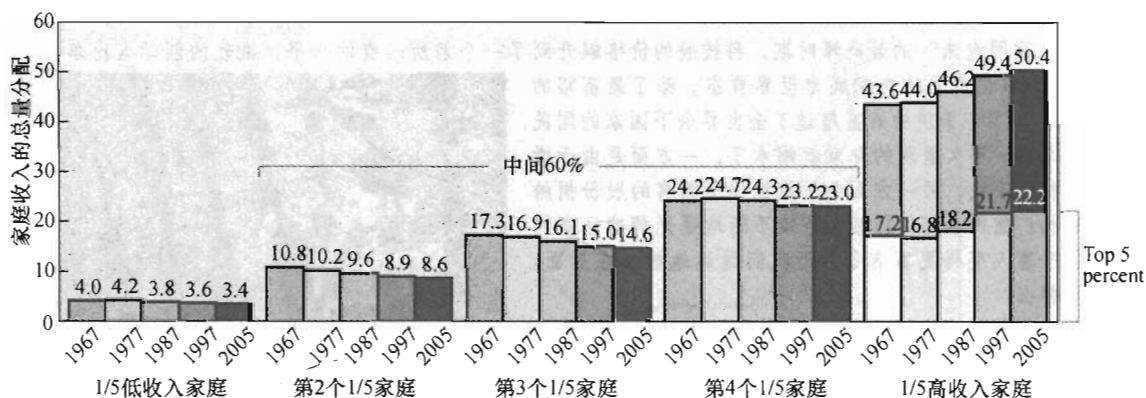


图 4-19 按收入高低划分的 5 组人群（包括最高收入 5% 的人群）分别占家庭总收入的百分比（1967 年、1977 年、1987 年、1997 年、2005 年）

资料来源：U. S. Census Bureau.

既然我们确定富人变得更富了，接下来的问题是，这会对我们产生影响吗？大多数人，包括大多数经济学家，普遍认为持续的收入分配不公平会对民主产生很大影响，但依然有一小部分经济学家（政治倾向既有左翼也有右翼的），他们却认为近十年来的变化是和过去不一样的。

首先，这种变化符合了被广泛接纳的伦理学原则：帕累托法则。它是意大利经济学家维尔弗雷多·帕累托提出并以其名字命名的（帕累托图也以他的名字命名）。它认为，只要改变在让一部分人变得更好的同时并没有让其他人变得更糟，那么这种改变就是好的。帕累托法则似乎很好地解释了当前的情况：不断增长美国经济几乎让所有人受益。换句话说，尽管绝大多数人群在收入份额中所占的比例要比过去要小，但是每个人的绝对财富以及生活条件要比过去要好。

其次，今天的富人也与过去不同。例如，在离现在较近的 1980 年，60% 的福布斯财富排行榜上的富人（最富有的 400 个人），他们所拥有的大部分财富是继承的，而到 2005 年，这个比例小于 20%。这可以给我们一个启示，过去你想成为有钱人，只能靠出身，而现在，你可以通过接受良好的教育以及努力工作获得财富。

因此，这可以促进人们更好地受教育和努力工作。

最后，当整个财富分配的不平等在上升的同时，性别与种族收入上面的分配不平等却在下降。换句话说，这对非裔、拉丁裔与女性来说，获得与白人雇员同等收入要比过去要容易，这些对我们促进民主价值确实是一件好事，当然我们在消除财富分配不公平方面还有很长一段路要走。

当然，即使近期内财富分配上面的不平等是对美国有过好处，但是如果一直持续下去弊端就会显现出来。遗憾的是，没有人能准确知道，导致财富的不平等分配持续上升的原因是什么，当然也就不会有人知道如何扭转这种趋势。

问题讨论

1. 比较图 4-18 与图 4-19 中呈现的这几种不同的分析数据的方法。例如，一种方法是否比另一种方法看起来呈现出收入分配上的更大的不平等？你能想到其他一些可能来呈现这些收入数据，并描绘出与给出的图形有所不同的图形吗？
2. 你同意以帕累托法则来评估经济变化是否符合伦理性，是一种好的方式吗？请说出理由。
3. 总的来说，你认为持续的收入分配不平等对美国的发展是好事还是坏事？如果这种趋势持续下去会如何？论证你的观点。
4. 尽管经济数据表明，绝大多数美国人要比过去几十年要过得好，但是赤穷的人群经济条件仍然非常差，你认为有什么可以做的或者应该做的来帮助这些穷人？你的建议在实施过程中不会伤害到总体经济的运行？请说明之。

建议阅读

1. You can find extensive data and reports on the distribution of income by going to the “Income” area of the U. S. Census Bureau Web site at www.census.gov.
2. Arrow, Kenneth, et al. (eds.). *Meritocracy and Economic Inequality*. Princeton University Press, 2000.
3. Firebaugh, Glenn. *The New Geography of Global Income Inequality*. Harvard University Press, 2006.
4. Nasar, Sylvia. “Is the U. S. Income Gap Really a Big Problem?” *New York Times*. April 4, 1999.



生活并没有什么可怕的，只是需要理解它。

——玛丽·居里

第 5 章 正态的世界

学习目标

■ 5.1 什么是正态

理解正态分布的含义，了解正态分布可能出现的情况。

■ 5.2 正态分布的比例

掌握依据 68-95-99.7 规则、标准分和百分数来解

释正态分布。

■ 5.3 中心极限定理

理解中心极限定理背后的理念以及它在统计学中的重要作用。

当你进入一家商店时，你怎么知道这家商品是否物美价廉？当你运动时你的心率上升了，你怎么知道上升的幅度并不会影响你的身体健康？如果你的 12 岁女儿在 5 分钟完成了 1000 米长跑，那她将来有希望参加奥运会吗？这些问题看起来不一样，但是从统计学的观点来看，这些问题是一致的：每一个问题都在问一个特定的数字（价格、心率、跑步时间）是否不寻常。在本章，我们将讨论基于钟形的正态分布，怎么样回答这样的问题。



5.1 什么是正态

假定你的朋友是个孕妇，预产期是 6 月 30 日，那么你会建议她安排一个时间是 6 月 16 日的重要商业洽谈日程吗？在预产期前两个星期。回答这个问题需要知道，孩子降生时间早于预产期 14 天的可能性是多少。为此，我们需要调查预产期和实际出生日期关系的数据。

图 5-1 是在普罗维登斯纪念医院 300 个孩子自然出生日期的一个直方图，数据是虚拟的，但却是基于没有医学干预的自然出生的日期分

布。横坐标表示出生日期是在预产期之前或之后多少天，负数表示出生日期早于预产期，正数表示出生日期晚于预产期。左边的纵轴呈现了 4 天一组的出生人数。例如，出生日期在 $-2 \sim -2$ 天对应的次数是 35 的条形高度，它说明在总数为 300 个孩子的出生日期中，在预产期前后两天出生的孩子为 35 人。

为了回答前面的孩子出生早于预产期 14 天有多少可能性这个问题，分析相对次数是更有效的。数据的次数除以总次数就是这组数据的相对次数（见 3.1 节）。图 5-1 中右侧的纵坐标呈现了相对次数，例如， $-14 \sim -10$ 天这一组（颜色加深条形）的相对次数是 0.07 或者 7%，也就是说 300 名孩子中大约 7% 在预产期前 14 ~

10 天之间出生。

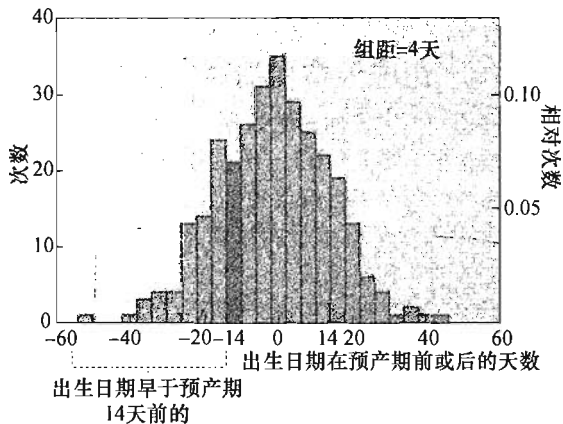


图 5-1 出生日期与预产期关系的次数（左轴）和相对次数（右轴）直方图（数据是虚拟的）

注：负数表示出生日期早于预产期，正数表示出生日期晚于预产期，每一组组距离是 4 天，例如，颜色加深条形所在组是指出生日期早于预产期 10~14 天。

现在我们可以得到出生日期在预产期 14 天之前所占的比例：我们将出生日期在预产期 14 天之前的所有组的相对次数累加，你从图中可以得到这个相对次数大约是 0.21，这样我们可以说，在这组数据中预产期 14 之前出生的约占 21%。我们也可以说你的朋友在商业洽谈日程那天及之前，有 1/5 的机会她的孩子会出生。如果这个商业洽谈是重要的，她可以安排得时间更提前一些。

思考题

假定你的朋友计划在孩子出生后休 3 个月的产假，基于图 5-1 的数据，假定预产期是 6 月 30 日，她应该答应在 10 月 10 日重返工作岗位吗？

5.1.1 正态分布形态

出生日期分布有一个客观的分布形态，我们很容易在直方图上覆盖一条光滑的曲线（见图 5-2）。对于我们现在的目标，这条光滑的曲线分布有三个重要的特征：

(1) 分布是一条单峰分布，它的众数，也就是指次数分布最多的出生日期，是预产期。

(2) 分布是围绕着单峰对称的，这样，中位数、平均数与众数相同，中位数是预产期因为在此之前和之后的出生日期的次数是相同的，

平均数是预产期，是因为在预产期之前每天出生的人数，与预产期之后同每天出生的人数是相同的。

(3) 分布向两端延伸，看起来像个钟形，我们称为“钟形分布”。

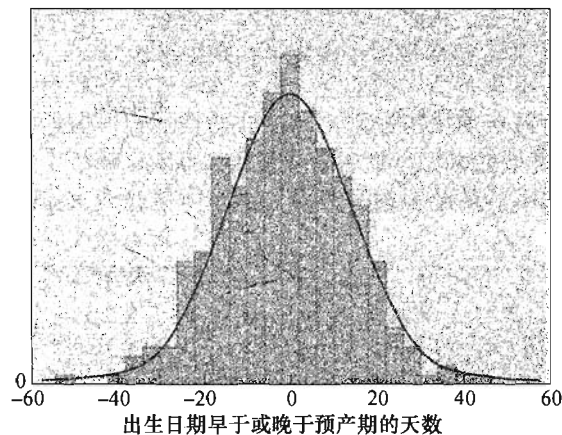


图 5-2 在图 5-1 直方图上方绘制的一条光滑正态分布曲线

思考题

图 5-2 的直方图，是基于一个自然出生日期，客观上是对称的，今天，如果一个孕妇生产日期太晚的话，医生经常会帮助她提前生产。如果提前助产的话这个直方图将会怎样改变。

图 5-2 带有上述三个特征的光滑曲线分布，叫做正态分布（normal distribution）。（注意所给出的出生日期并不是完全的正态分布。）所有正态分布都具有同样的钟形特征，但是它们的平均数和离散度可以不同。图 5-3 呈现了两种不同的正态分布，它们有相同的平均数，但是图 5-3a 的离散度更大一些。我们要在下一节讨论正态分布的标准差（见 4.3 节），它会告诉我们需要了解的离散度的详细情况。一个正态分布只需要两个数字就可以完整地进行描述了：平均数和标准差。

方法注解

平均数 μ 和标准差 σ 已知的正态分布公式为：

$$y = \frac{e^{-\frac{1}{2}[(x-\mu)/\sigma]^2}}{\sigma \sqrt{2\pi}}$$

本书中并不应用这个公式，它用代数方法描述了正态分布的形态。

相关知识

使用来自法国和苏格兰战士的数据，比利时的科学家阿道夫·凯特莱 (Adolphe Quetelet) 在 19 世纪 30 年代认识到人类的身高和胸围等特征符合正态分布，在观察的基础上，他提出了“平均人”的概念。凯特莱是美国统计学会的第一个外籍会员。

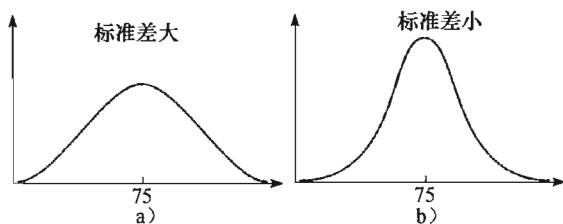


图 5-3 两个图都是正态分布，它们有着相同平均数，但是左侧分布的标准差较大

例 5-1 符合正态分布吗

图 5-4 呈现了两个分布，图 5-4a 大概是在 1846 年搜集的，5 738 位著名的苏格兰国民自卫队员的胸围组成的一组数据，图 5-4b 是美国 50 个州的人口密度数据。哪个分布属于正态分布，说明理由。

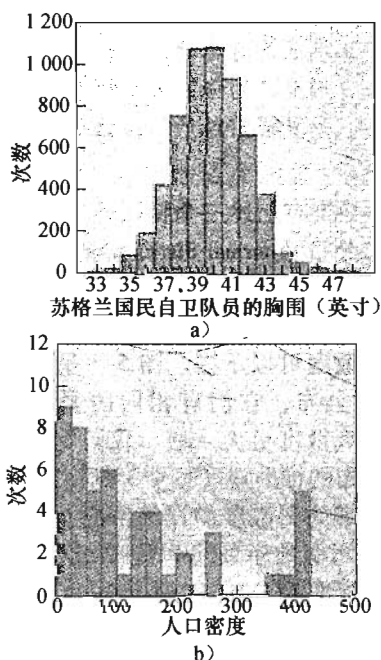


图 5-4 哪个图属于正态分布

a) 苏格兰国民自卫队员的胸围分布 b) 美国 50 个州的人口密度

资料来源：Adolphe Quetelet, *Lettres à S. A. R. le Duc Régnant de Saxe-Cobourg et Gotha*, 1846.

解答 图 5-4a 是近似对称的，平均数在 39~40 英寸，它是接近正态分布的。图 5-4b 显示了大多数州的

人口密度是较低的，只有少数几个州的人口密度较高，这个分布是右偏态的，不属于正态分布。

5.1.2 正态分布和相对次数

回想一下，任何一组数据的相对次数的总和都必须是 1 (见 3.1 节)。现在考虑图 5-2 中正态分布曲线。尽管不再有单独的直条，我们仍然可以将曲线的高度与相对次数联系起来。实际上，相对次数的总和是 1 成了一个条件，那就是正态曲线涵盖下的区域面积也必须是 1。

图 5-5 呈现了我们如何解释正态曲线的一个例子。关键的原则是：一段范围数据值的相对次数就是正态曲线覆盖这段数据值的区域面积。例如，一个精确的计算显示从左边到 -14 天所占的区域，大约占到曲线覆盖全部面积的 18%。我们可以定义：数据 -14 天的相对次数为 0.18，也就是说出生日期在预产期 14 天之前的占到总数的 18%。

同样，生产日期在预产期之后 18 天以上的占了总数 12%。所以，我们可以得出结论说，生产日期在预产期之后 18 天以上的相对次数是 0.12。总之， $18\% + 12\% = 30\%$ 的人生日期早于预产期 14 天以上晚于预产期 18 天以上。(注意，相对次数正好是图 5-5 的正态分布，比较接近图 5-1 的原始数据。)

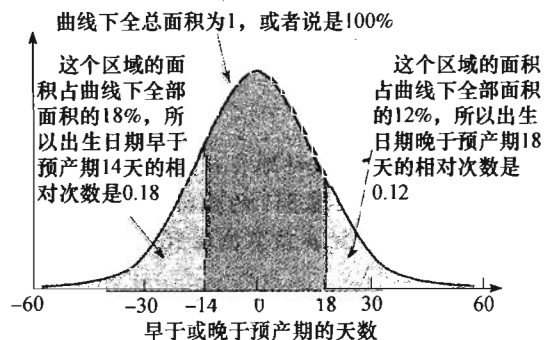


图 5-5 曲线下任何区域的面积能够告诉我们这个区间数据取值的相对次数

相对次数和正态分布

- 与横坐标上一段值域对应的正态分布曲线下包含的面积是这段值域在总体中的相对次数。
- 因为相对次数一定是 1，所以正态分布曲线下的总面积一定是 1，或者 100%。

思考题

按照图 5-5, 多少百分比的出生日期在早于预产期 14 天晚于预产期 18 天之间的, 请解释。(提示: 在曲线之下的总面积是 100%。)

相关知识

苏格兰政治家约翰·辛克莱 (John Sinclair, 1754—1835) 是第一位经济、地理和农业数据的收集者, 公认他把原来在德语中指事件陈述的 “statistics” 和 “statistical” 引进英语词汇。

例 5-2 估计面积

再看一下图 5-5 显示的正态分布。

(1) 估计出生日期在预产期之后 0~60 天的孩子所占百分数。

(2) 估计出生日期在预产期之前 14 天到预产期之后 14 天的孩子所占百分数。

解答

(1) 0~60 天范围的区域大约占到曲线覆盖以下面积的一半。这个意思是指样本中大约 50% 的孩子出生日期是在预产期之后 0~60 天。

(2) 图 5-5 显示出大约 18% 的孩子出生日期是在预产期 14 天前及更早, 因为分布是对称的, 大约 18% 的孩子出生日期是在预产期 14 天之后。这样, 大约总数 $18\% + 18\% = 36\%$ 的孩子出生日期是在距预产期 14 天之前或者超过预产期 14 天之后。这个问题同样也回答了余下的区域, 也就是距预产期前后不到 14 天的时间内, 有 $100\% - 36\% = 64\%$ 的孩子出生日期是在这个时间段内。

5.1.3 什么情况下我们能预期一个正态分布

正态分布是许多有实践意义的变量非常好的近似分布。诸如体重、身高、血压和反应时这样的生理指标基本符合正态分布。诸如 SAT 或者 IQ 这些标准分也是正态分布。运动统计学, 比如安打率、游泳速度、田径运动会中的跳高, 这些也倾向于正态分布。的确, 许多统计是基于正态分布的。尽管如此, 并不是所有变量都符合正态分布, 理解何时能应用正态分布, 何时不适用是非常重要的。

通过分析正态分布图, 我们可以看出正态分布的要素包括什么:

(1) 大量的数据集中在平均数附近, 因此它是单峰分布, 或者只有一个众数。

(2) 数据值围绕着平均数展开, 分布是对称的。

(3) 离平均数越远, 数据出现的机会越少, 具有钟形分布形态。

更进一步来说, 被许多不同因素影响的定量数据倾向于正态分布, 生理特征受许多不同的遗传和环境因素的影响, 标准化测验分数反映了很多个体面对大量的各种测验问题的回答。运动统计包含了不同运动水平的运动员在不同条件下的表现。

相关知识

正态分布曲线也叫做高斯曲线, 高斯 (Carl Friedrich Gauss) 是 19 世纪的德国数学家, 美国逻辑学家查尔斯·皮尔士 (Charles Peirce) 大约在 1870 年介绍了术语正态分布。

例 5-3 符合正态分布吗

你认为下面哪个变量属于正态分布或者近似正态分布?

- (1) 一次非常容易的测验分数。
- (2) 一个随机选取的成年妇女样本的身高。
- (3) 100 个完整藤编筐子中每个筐子里苹果的数量。

解答

- (1) 测验可能的最高分数 (100%) 限制了数据取值的范围。如果测验很容易, 平均数会比较高, 并且许多分数会比较接近最高可能分数。较少的低分数会散落在平均数下方。这样, 我们预计这个分布属于左侧分布而非正态分布。
- (2) 身高是受很多因素影响的 (父母遗传基因、环境因素和成熟)。我们可以预期样本的身高平均数比较接近于众数 (最常见的身高), 我们也可以预计女性身高在平均数以上和平均数以下的大致相等, 非常高的和非常矮的女性比较少, 也就是说身高接近正态分布。
- (3) 不同藤编筐子苹果的数量与苹果的大小有关, 我们预计这个分布有一个接近每筐苹果平均数的众数, 多于平均苹果数量的筐子和少于平均数的筐子是比较接近的, 这样, 我们预计不同藤编筐子苹果的数量接近一个正态分布。

思考题

你预计一个中等难度的测验分数符合正态分布吗? 说出两种以上你估计符合正态分布的定量数据。

5.1 练习题

统计素养和批判性思维

- 1. 正态分布 当我们说到分布的时候，词汇与日常生活语言意思相同吗？或者说它在统计学上有特殊的含义吗？准确来说正态分布是什么？
- 2. 正态分布 对一个正态分布进行非正式的和简略的描述时，我们称其为“钟形”的概率分布。什么是“钟形”分布？
- 3. 随机数字 当进行一项调查时可以应用计算机来随机产生电话号码数字。这些数字的分布是正态分布吗？为什么？
- 4. 面积 在正态分布图上，如果你确定从左到某一特定值的面积是 0.2，那么从右到这个值的面积是多少？

判断正误：判断练习题 5~8 中的陈述是有意义的（表述正确）还是无意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断会更重要。

- 5. SAT 分数 SAT 分数的一个随机样本是正态分布，平均数 1518，中位数 1490。
- 6. 体重 作为国家健康研究的一部分，对一个随机样本中的每个人都测量了体重，这些体重属于一个正态分布，有两个众数。
- 7. 智商分数 一组符合正态分布的智商分数平均数为 100，60% 的智商数分数超过了 105。
- 8. 百事可乐的容量 一个质量控制分析者测量了 100 听普通百事可乐的容量，指出容量的分布符合正态分布。

概念与应用

- 9. 哪一个属于正态分布 指出图 5-6 中的分布，哪个不符合正态？两个正态分布中，哪一个的标准差较大？

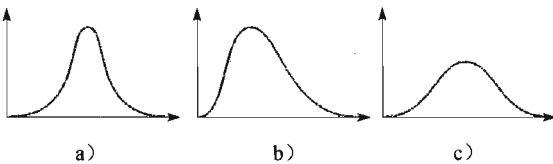


图 5-6 哪一个正态分布

- 10. 哪一个属于正态分布 指出图 5-7 中哪一个

不属于正态分布？两个正态分布中，哪一个的标准差较大？

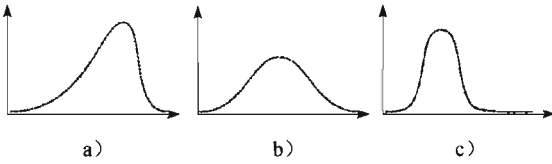


图 5-7 哪一个正态分布

正态变量：对于练习题 11~18 中的每一组数据，你认为它是否符合正态分布。解释你的理由？

- 11. CD 的重量 索尼 CD 的一个随机样本的准确重量。
- 12. 收入 随机挑选的美国成人的收入。
- 13. 抛掷骰子 1 枚均匀的骰子抛掷 500 次的结果。
- 14. SAT 分数 去年，所有人的 SAT 分数。
- 15. 握力强度 一个男性随机样本握力强度的测量。
- 16. 航班延误 民用航班延误的时间长度。
- 17. 等候时间 如果公交车是十分钟一趟，你在任意一个时间到达公交车站等候公交车的时间。
- 18. 一个 1 000 辆汽车的随机样本中，违规停车罚单的数量。
- 19. 电影长度 图 5-8 呈现了 60 部电影时长的直方图，平均电影时长是 110.5 分钟，这个分布接近于正态吗？这个变量属于正态分布吗？为什么？

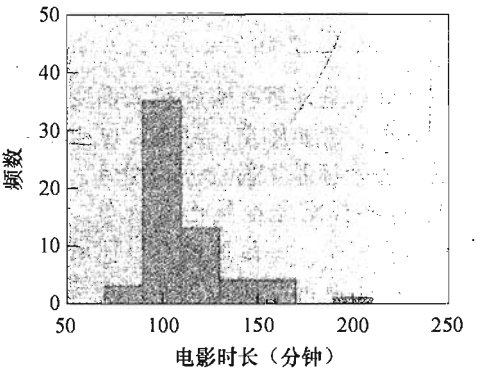


图 5-8 60 部电影时长直方图

- 20. 心率 图 5-9 呈现了 98 个学生心率的直方

图, 平均心率是每分钟 71.2 跳, 这个分布接近于正态吗? 这个变量会符合正态分布吗? 为什么?

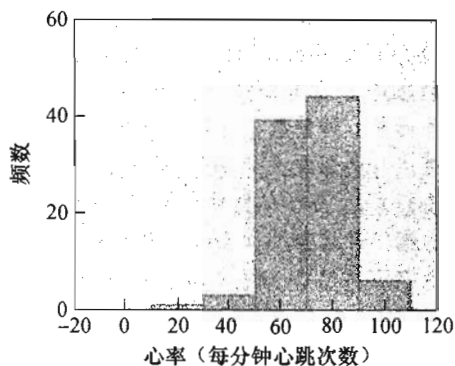


图 5-9 98 个学生心率的直方图

21. 25 美分硬币的重量 图 5-10 呈现了 50 个随机选择的 25 美分硬币重量的直方图, 平均重量是 5.62 克。这个分布接近于正态吗? 这个变量符合正态分布吗? 为什么?

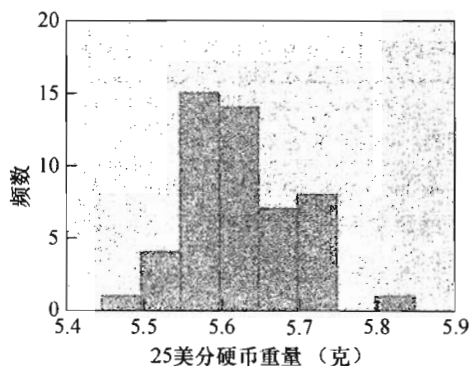


图 5-10 25 美分硬币重量的直方图

22. 阿司匹林的重量 图 5-11 呈现了 30 个随机挑选的阿司匹林药片的重量直方图, 平均重量是 665.4 毫克。这个分布接近于正态吗? 这个变量符合正态分布吗? 为什么?

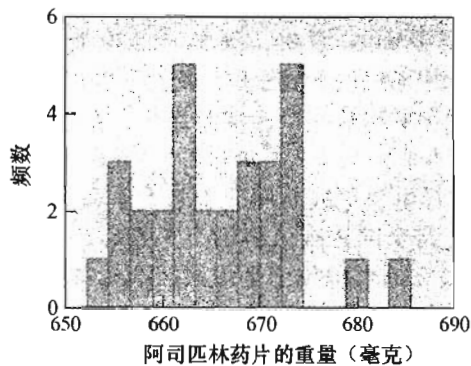


图 5-11 阿司匹林药片重量直方图

23. 面积和相对次数 思考一下图 5-12 中的正态分布曲线, 是关于男性身高分布的相对次数, 分布的平均数是 69.6 英寸, 标准差是 2.8 英寸。

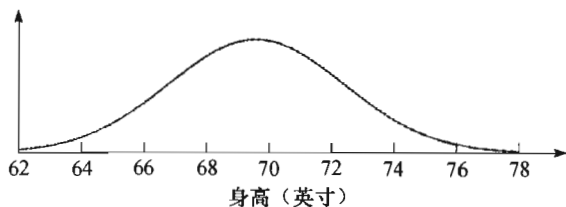


图 5-12 正态分布曲线

- (1) 曲线下的总面积是多少?
 - (2) 估计小于 67 英寸的相对次数。
 - (3) 估计大于 67 英寸的相对次数。
 - (4) 估计 67 ~ 70 英寸的相对次数。
 - (5) 估计大于 70 英寸的相对次数。
24. 面积和相对次数 思考图 5-13 中的正态曲线, 图中呈现了智商分数的相对次数, 分布的平均次数是 100, 标准差是 16。

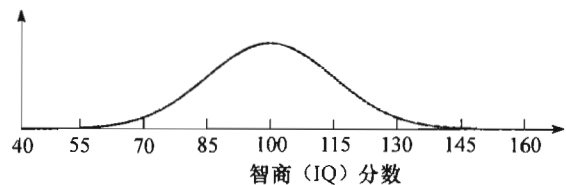


图 5-13 智商分数的正态分布曲线

- (1) 曲线下的总面积是多少?
 - (2) 估计智商分数小于 100 的相对次数。
 - (3) 估计智商分数大于 110 的相对次数。
 - (4) 估计智商分数小于 110 的相对次数。
 - (5) 估计智商分数在 100 ~ 110 的相对次数。
25. 估计面积 考虑图 5-14 的正态分布曲线, 图中显示了女学生样本的收缩压分布的相对次数, 分布的标准差是 14。

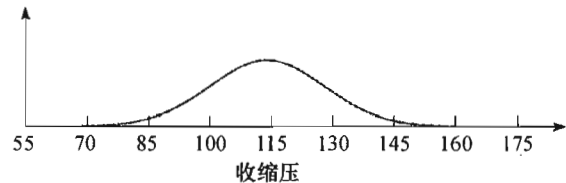


图 5-14 收缩压的正态分布曲线

- (1) 分布的平均数是多少?
- (2) 估计收缩压小于 100 的学生的百分比。
- (3) 估计收缩压在 110 ~ 130 学生的百分比。
- (4) 估计收缩压高于 130 的学生的百分比。

26. 估计面积 考虑图 5-15 中的正态分布曲线，图中给出了一个男学生样本体重分布的相对次数。

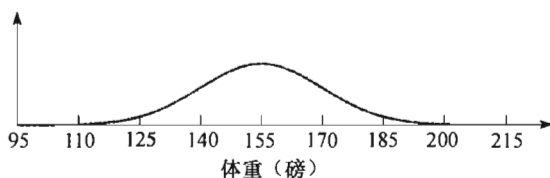


图 5-15 体重的正态分布曲线

- (1) 分布的平均数是多少？
- (2) 估计体重小于 140 磅的学生的百分比。
- (3) 估计体重超过 170 磅的学生的百分比。
- (4) 估计体重在 140~160 磅的学生的百分比。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 5 章的有用链接“links for internet projects”。

27. SAT 分数分布 某学院网站上的公告栏给出了 SAT 分数的分布（通常分组组距是 50 分）搜集这些数据并且就测验的每一部分绘制一个直方图，讨论实际的 SAT 分数分布

是否符合正态分布。

28. 寻找正态分布 采用前文中给出的指导意见，选择你认为接近正态分布的变量，搜集至少 30 个数据，并且绘制成直方图。评论这个分布是否接近于正态分布。在哪些方面它和正态分布有所不同，解释这个差异。
29. 电影时长 搜集数据支持或反对关于电影的如下观点，电影在未来 10 年内会越来越短，进一步绘制从 1940 年到现在每 10 年电影时长的直方图，计算每个样本的电影时长平均数。评价这个分布是否为正态分布。讨论这个结果，并且就你所观察到的趋势给出可能的原因。

新闻中的统计学

30. 正态分布 很少有新闻报道说明变量的实际分布，或者说明这个变量是正态分布，尽管如此，新闻报道中所关注的变量一定符合某些分布。寻找两个你认为它们接近正态分布的新闻报道的变量，解释你的理由。
31. 非正态分布 寻找两个你认为它们不符合正态分布的新闻报道中的变量，说明你的理由。



5.2 正态分布的比例

考虑一个消费调查报告，研究参与者被问到他们的上一台电视机到更换时使用了多久。关注变量是电视机的使用时间，基于这个调查，电视更换时间的平均数是 8.2 年，我们以 μ 表示；标准差是 1.1 年，以 σ （希腊字母西格玛）表示。我们合理地假设电视更换时间接近于正态分布，就可以如图 5-16 那样，将其分布绘制成图形。

希腊字母 σ （西格玛）代表标准差，在 -1σ 与平均数 (μ) 之间的区域表示代表平均数以下 1 个标准差的数据值分布

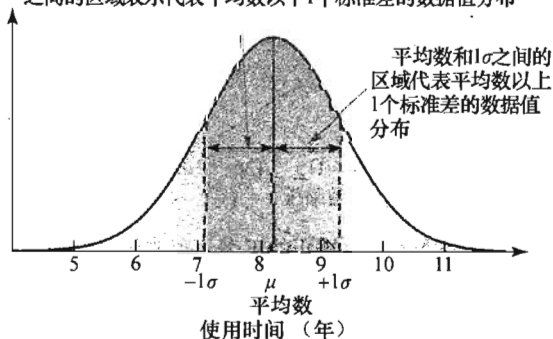


图 5-16 电视机更换时间的正态分布（平均数为 8.2 年，标准差是 1.1 年）

思考题

应用正态分布的四个标准（见 5.1 节）解释电视更换时间的分布为什么接近于正态分布？

方法注解

一个正态分布的平均数可以取任意值，标准差可以是任意正数。术语“标准正态分布”特指平均数为 0，标准差为 1 的正态分布。

因为所有正态分布都有同样的钟形，知道分布的平均数和标准差可以让我们说出大量的数据值分布在哪里。例如，如果我们测量图 5-16 中曲线以下的面积，我们会发现大约 2/3 的面积在平均数一个标准差之内，在本例中就是介于 $8.2 - 1.1 = 7.1$ 年和 $8.2 + 1.1 = 9.3$ 年之间。这样电视更换时间介于 7.1 年和 9.3 年之间的占到全部调查人群的 2/3 左右。同样地，大约 95% 的面积位于平均数两个标准差之内，在本例中就是介于 $8.2 - 2.2 = 6.0$ 年和 $8.2 + 2.2 = 10.4$ 年之间。我们得出结论说，在所调查的人群之中大约有 95% 的人，他们的电视更换

时间在 6.0 ~ 10.4 年。

一个简单的法则叫做 **68-95-99.7 法则** (68-95-99.7 rule)，它是一个有关预测的指导原则，代表任何正态分布位于平均数 1, 2 和 3 个标准差之内的数据所占的百分比，下面就是关于这条法则的文字陈述，图 5-17 是其图形直观呈现。

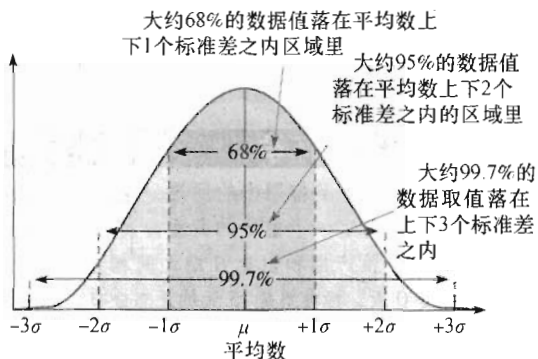


图 5-17 正态分布的 68-95-99.7 法则

正态分布下的 68-95-99.7 法则

1. 大约 68% (更精确的是 68.3%)，或者说刚好超过 2/3 的数据值落在离平均数一个标准差之内。
2. 大约 95% (更精确地是 95.4%) 的数据之落在离平均数两个标准差之内。
3. 大约 99.7% 数据落在离平均数 3 个标准差之内。

方法注解

本章结尾部分的“聚焦心理学”中所讨论的，某一次特定的 SAT 平均分与依据测验和年份所给定的平均分 500 分是不一样的。

例 5-4 SAT 分数

由语言 (阅读理解) 和算术组成的测验 SAT (以及 GRE、LAST、GMAT)，其分数呈正态分布，平均数 = 500，标准差 = 100。解释这个表述。

解答 按照 68-95-99.7 法则，大约 68% 的学生分数在平均数 500 的一个标准差 (100) 之内，也就是 68% 的学生分数在 400 ~ 600 分之间。大约 95% 的学生分数在平均数两个标准差 (200) 之内，也就是在 300 ~ 700 分之间。大约 99.7% 个学生分数在平均数三个标准差之内，也就是 200 ~ 800 分之间。图 5-18 呈现了图形的形象解释，注意横坐标显示了

实得分数，又显示了距离平均数的标准差数值。

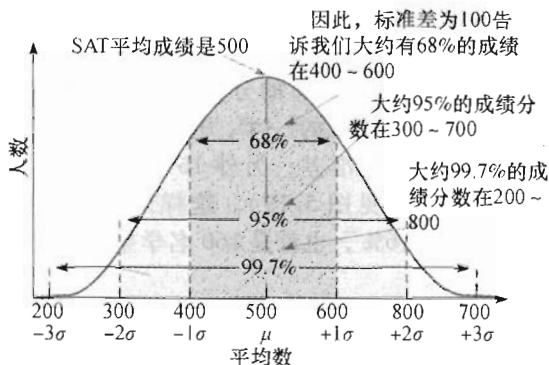


图 5-18 SAT 测验分数的正态分布 (呈现了 1, 2 和 3 个标准差各自对应的百分数)

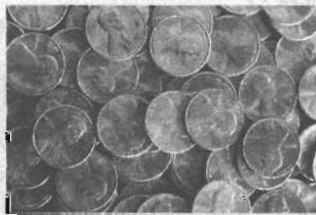
例 5-5 假币检测

自动售货机经常需要调整，以便拒收高于或低于某一特定重量的硬币。0.25 美元的法定重量为一个正态分布，平均数为 5.67g，标准差为 0.0700g。如果一个自动售货机调整成拒收超过 5.81g 和少于 5.53g，那么合法的 0.25 美元硬币将会被这台自动售货机拒收的百分比是多少？

解答 5.81g 超过平均数 0.14g，或者两个标准差，5.53g 低于平均数 0.14g，也是两个标准差。这样，只接受重量在 5.53g 和 5.81g 之间的硬币，自动售货机接受的硬币在平均数两个标准差之内，拒收的硬币在平均数两个标准差之外。根据 68-95-99.7 法则，真硬币中有 95% 被接受，5% 将被拒收。

相关知识

美国每月大概要铸造 10 亿枚 1 美分硬币，但是，按照美国审计局的报告，据估计这些硬币的流通率大约是 34% (25 美分的流通率为 88%)。这意味着大约 2/3 的 1 美分硬币铸造出来后就进了存钱罐或硬币盒里。



5.2.1 68-95-99.7 法则的应用

我们可以应用 68-95-99.7 法则来确定数据值是否位于平均数 1, 2 和 3 个标准差之内。例如，假设 1000 个学生参加了一次测验，分数是正态分布，平均数 = 75，标准差 = 7。

82 分高于平均分 7 分，或者说是高出一个标准差。68-95-99.7 法则告诉我们大约 68% 的分数在平均数一个标准差之内，也可以说大约 $100\% - 68\% = 32\%$ 的分数超出平均数一个标准差之外。32% 的一半，也就是 16%，高于平均数一个标准差，另外 16% 低于平均数一个标准差（见图 5-19）。这样，1 000 名学生中的大约 16%，也就是 160 名学生，分数在 82 分以上。

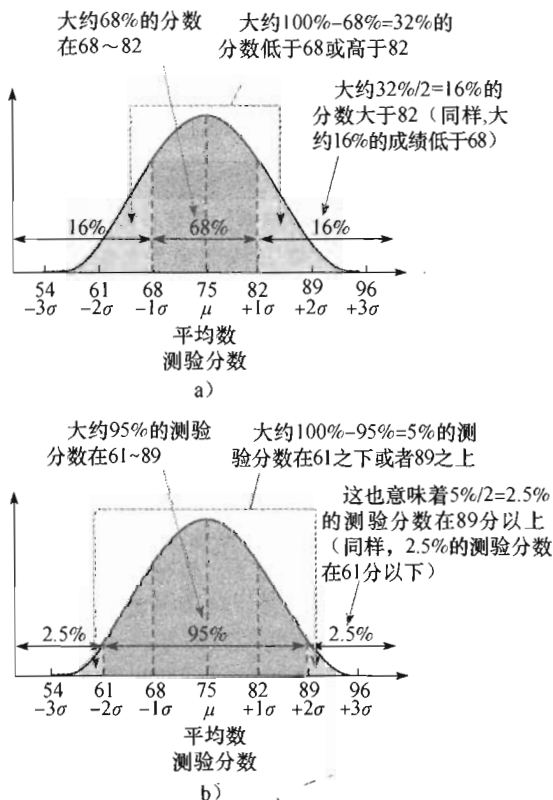


图 5-19 一个测验分数的正态分布（平均数 75，标准差 7）

a) 68% 的分数在平均数 1 个标准差以内 b) 95% 的分数在平均数 2 个标准差以内

同理，分数 61 低于平均数 14 分，或者说低于两个标准差。根据 68-95-99.7 法则，告诉我们大约 95% 的分数在平均数两个标准差之内，也就是大约 5% 的分数在平均数两个标准差以外。5% 的一半，也就是 2.5% 的分数低于平均数两个标准差以下（见图 5-19b），这样 1 000 名学生中的 2.5%，分数低于 61 分。

因为 95% 的分数落在 61~89 分之间，我们会说这个范围之外的分数是很偶然的，因为它们相当少见。

判断偶然的結果

在统计学上，我们经常需要区分哪些数据值是典型的，或者说是“常见的”，哪些数据值是很少见的。通过应用 68-95-99.7 法则，我们可以发现正态分布中大约 95% 的数据值位于平均数两个标准差之内，这也意味着，大约 5% 的数据值位于平均数两个标准差之外。我们可以用这个比率来判断哪些数据值是非常少见的：少见的数据值是那些位于平均数两个标准差之外的数据值。

例 5-6 旅行和怀孕

再次考虑下那个孕妇朋友的问题，是否建议她在预产期两周之前约定一个重要的商业洽谈。实际上的数据提示我们，预产期和出生日期之间相差的天数，属于平均数是 0 天，标准差是 15 天的正态分布。怎样帮助您的朋友作决策？在预产期前两周出生的是否少见的？

解答 假如你朋友的生产日期不是在预产期之前 14 天，也不是在一个标准差前。因为这位结果在平均数两个标准差之内，它并不是很少见的。依据 68-95-99.7 法则，大约有 68% 的孕妇在平均数一个标准差之内生产的，或者说是预产期前 15 天和后 15 天之间。这就意味着大约 $100\% - 68\% = 32\%$ 的生产日期超过预产期前后 15 天（见图 5-20）。你可以告诉你的朋友大约 16% 的孕妇会在预产期之前 15 天生产。如果你的朋友愿意从概率的角度想一想，你可以告诉她有 0.16 的机会（大约 1/6）在她进行商业洽谈那天或者之前孩子就会出生。

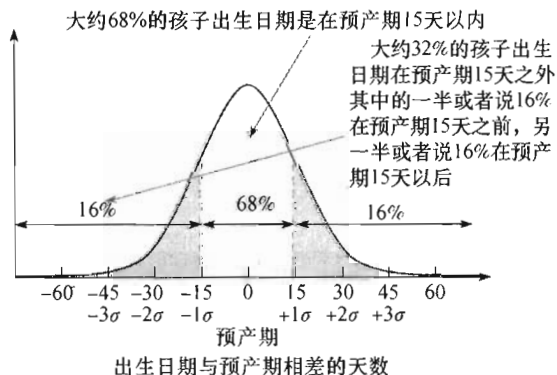


图 5-20 预产期 15 天以前出生的婴儿大约占 16%

例 5-7 正常心率

你测量一下你每天中午的静息心率，坚持一年并且记录下这些数据，你会发现这些数据符合一个正态分布，平均数是 66，标准差是 4。那么这些心率在每分钟 58 以下的时候是多少天？

解答 心率 58 是低于平均数 8 分 (或者两个标准差), 按照 68-95-99.7 法则, 大约 95% 的数据是平均数两个标准差以内。也就是说, 2.5% 的数据值低于平均数两个标准差, 另有 2.5% 的数据值高于平均数两个标准差。一年 365 天的 2.5%, 或者说大约 9 天, 你测量的心率在每分钟 58 以下。 ■

思考题

如例 5-7 中所提到的, 某个体静息心率的多次测量值是正态分布。你预期许多个体的静息心率的平均值还是正态分布吗? 你设想一下哪一个分布会有比较大的标准差? 为什么?

例 5-8 计算百分等级

在医生的办公室里, 医生说你四年级的女儿的体重高于她这个年龄和性别的孩子的平均数一个标准差, 她体重的百分等级是多少? 假定四年级女孩子的体重是正态分布。

解答 回忆一下如果有 $n\%$ 个数据值等于或低于它, 那么这个数据值就位于分布的第 n 百分等级 (4.3 节)。按照 68-95-99.7 法则, 68% 的女孩体重在平均数一个标准差之内, 因此, 34% 的女孩体重 (68% 的一半) 在离平均数 0~1 个标准差之间。我们也可以知道, 因为分布是对称的, 所有女孩体重的 50% 位于平均数以下。这样, $50\% + 34\% = 84\%$ 的女孩的体重低于平均数 1 个标准差 (见图 5-21)。在所有四年级女孩子的体重中你的女儿处于第 84 百分等级。

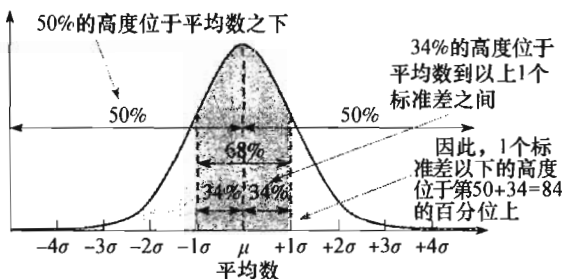


图 5-21 正态分布曲线显示低于平均数 1 个标准差的分数为 84% ■

5.2.2 标准分

68-95-99.7 法则只适用于离平均数的位置在 1, 2 和 3 个标准差位置的数据, 如果可以精确地知道离平均数某一标准差位置的数据值, 我们就能推广这个法则。一个数据值距离平均数的位置是上还是下, 有几个标准差大小, 这个数值叫做标准分 (standard score), 常用字母 z 表示, 例如:

1. 平均数的标准分为 $z = 0$, 因为它离平均数的距离为 0 个标准差。

2. 高于平均数 1.5 个标准差位置的数据值, 标准分为 $z = 1.5$ 。

3. 低于平均数 2.4 个标准差位置的数据值, 标准分为 $z = -2.4$ 。

标准分的计算:

$$z = \text{标准分} = (\text{数据值} - \text{平均数}) / \text{标准差} \quad (5-1)$$

标准分是正值时, 数据值大于平均数, 标准分为负值时, 数据值小于平均数。

例 5-9 计算标准分

斯坦福-比奈智力测验使用的分数是标准分, 平均数是 100, 标准差是 16。计算智商为 85, 100 和 125 的标准分。

解答 我们计算这些智商的标准分, 将平均数 100 和标准差 16 带入标准分公式中。

(1) 智商 85 的标准分: $z = (85 - 100) / 16 = -0.94$

(2) 智商 100 的标准分: $z = (100 - 100) / 16 = 0$

(3) 智商 125 的标准分: $z = (125 - 100) / 16 = 1.56$

我们这样解释这些标准分: 智商 85 位于平均数以下 0.94 个标准差的位置, 智商 100 等于平均数, 智商 125 位于平均数以上 1.56 个标准差的位置。

图 5-22 呈现了这些智商分数在分布中的位置。

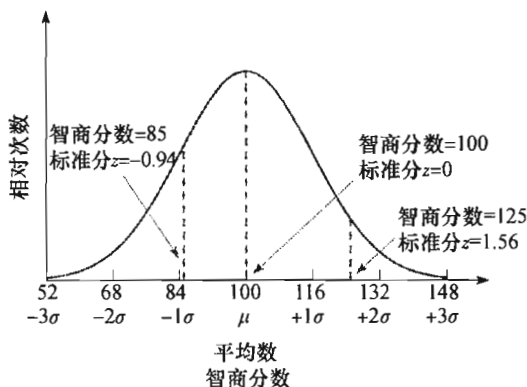


图 5-22 智商为 85, 100 和 125 的标准分 ■

5.2.3 标准分和百分等级

一旦我们知道一个数据值的标准分, 正态分布的比率可以让我们计算它在分布中的百分等级。这个过程可以应用标准分数表完成, 如表 5-1 (附录 A 有更详细的标准分数表)。对于正态分布中的每一个标准分数, 这个表可以给出分布中小于这个数值的百分数。例如, 表中显示出 55.96% 的正态分布中的数值的标准分小于 0.15。换句话说, 标准

分是0.15 的数据值位于第55 百分等级。

例 5-10 胆固醇水平

18~24 岁男性的胆固醇水平符合正态分布，平均数是 178，标准差是 41。

(1) 20 岁的男士其胆固醇水平是 190，他的百分等级是多少？

(2) 第 90 百分等级对应的胆固醇水平是多少？在哪个水平，治疗是必要的吗？

解答

(1) 胆固醇水平为 190，其标准分是：

$$z = \frac{\text{标准分}}{\text{标准差}} = \frac{(\text{数据值} - \text{平均数})}{\text{标准差}} \\ = \frac{(190 - 178)}{41} \approx 0.29$$

表 5-1 显示标准分 0.29 大约在第 61 百分等级。

(2) 表 5-1 显示所有数据的 90.32%，其标准分小于 1.3。这样，第 90 百分等级大约在平均数以上 1.3 个标准差处。已知胆固醇的平均水平是 178，标准差为 41，位于平均数以上 1.3 个标准差处的胆固醇水平就是：

$$\frac{178}{\text{平均数}} + \frac{(1.3 \times 41)}{1.3 \text{ 个标准差}} = 231.3$$

大约 231 的胆固醇水平位于第 90 百分等级。

例 5-11 IQ 分数

IQ 分数属于正态分布，平均数 100，标准差 16

(见例 5-9)。在 IQ 测验中，第 75 和第 40 百分等级的人们其 IQ 分数是多少？

解答 表 5-1 显示，第 75 百分等级位于标准分数 0.65~0.70 之间，我们可以估计它的标准分数大约是 0.67，这个对应于 0.67 个标准差，或者说大约位于平均数 100 以上 $0.67 \times 16 = 11$ 点。这样，一个位于第 75 百分等级的人智商是 111。第 40 百分等级对应的标准分数接近于 -0.25，或者说是平均数 100 以上 $-0.25 \times 16 = 4$ 点，这样，第 40 百分等级的智商为 96。

例 5-12 军队中的女性

18~24 岁的美国女性的身高符合正态分布，平均数为 65 英寸，标准差是 2.5 英寸，为了在美国军队中服役，女性身高必须在 58~80 英寸之间。基于这个身高，百分之多少的女性是不合格的？

解答 军队里女性的最大值 80 英寸和最小值 58 英寸的标准分数分别是：

$$58 \text{ 英寸: } z = (58 - 65) / 2.5 = -2.8$$

$$80 \text{ 英寸: } z = (80 - 65) / 2.5 = 6.0$$

表 5-1 显示标准分数 -2.8 对应的百分等级是 0.26，标准分数 6.0 对应的百分等级在表中没有列出，意思是百分等级在 99.98 以上（表中呈现的最高百分等级）。这样，所有女性的 0.26% 因为身高太矮不能在军队中服役，另有 0.02% 的女性因为太高也不能入伍。合在一起，大约 0.28% 的女性，或者说大约 1/400 的女性，是因为身高原因不能在美军中服役的。

表 5-1 正态分布中的标准分数和对应的百分等级（从右边开始计算）

标准分数	百分等级	标准分数	百分等级	标准分数	百分等级	标准分数	百分等级
-3.5	0.02	-1.0	15.87	0.00	50.00	1.1	86.43
-3.0	0.13	-0.95	17.11	0.05	51.99	1.2	88.49
-2.9	0.19	-0.90	18.41	0.10	53.98	1.3	90.32
-2.8	0.26	-0.85	19.77	0.15	55.96	1.4	91.92
-2.7	0.35	-0.80	21.19	0.20	57.93	1.5	93.32
-2.6	0.47	-0.75	22.66	0.25	59.87	1.6	94.52
-2.5	0.62	-0.70	24.20	0.30	61.79	1.7	95.54
-2.4	0.82	-0.65	25.78	0.35	63.68	1.8	96.41
-2.3	1.07	-0.60	27.43	0.40	65.54	1.9	97.13
-2.2	1.39	-0.55	29.12	0.45	67.36	2.0	97.72
-2.1	1.79	-0.50	30.85	0.50	69.15	2.1	98.21
-2.0	2.28	-0.45	32.64	0.55	70.88	2.2	98.61
-1.9	2.87	-0.40	34.46	0.60	72.57	2.3	98.93
-1.8	3.59	-0.35	36.32	0.65	74.22	2.4	99.18
-1.7	4.46	-0.30	38.21	0.70	75.80	2.5	99.38
-1.6	5.48	-0.25	40.13	0.75	77.34	2.6	99.53
-1.5	6.68	-0.20	42.07	0.80	78.81	2.7	99.65
-1.4	8.08	-0.15	44.04	0.85	80.23	2.8	99.74
-1.3	9.68	-0.10	46.02	0.90	81.59	2.9	99.81
-1.2	11.51	-0.05	48.01	0.95	82.89	3.0	99.87
-1.1	13.57	0.00	50.00	1.00	84.13	3.5	99.98

注：此表只显示了标准分数在 -3.5 ~ +3.5 的百分等级（其他更详细的见附录 A 表 A-1）。

5.2.4 与概率的关系

假定随机选择一个婴儿，询问这个婴儿是否出生于预产期15天以前。因为出生日期围绕着预产期成正态分布，标准差是15天。我知道16%的出生日期在预产期15天前（见例5-6）。对于一个随机选择的婴儿个体而言，大约有0.16的机会（大约1/6）这个婴儿在预产期之前15天出生。换句话说，正态分布的比例可以

让我们对一个个体做概率描述。在这种情况下，我们可以表述为，一个婴儿早于预产期15天以前出生的概率是0.16。

这个例子显示正态分布的比例可以用来表述概率的意思。实际上，我们要做的大部分工作，在余下的章节中将与概率密切相关。由于这个原因，在第6章，我们将重点学习概率基础，但是首先在下一节，我们将用以前所讨论的基本观念来介绍统计学中一个最重要的概念。

5.2 练习题

统计素养和批判性思维

1. 标准分数 一组数据值符合正态分布，平均数的标准分数是多少？
2. 标准分 从左边到某一特定标准 z 分数的百分比是75%；从右到同一 z 分数，对应的百分比是多少？
3. 分布 抛掷1枚骰子，平均结果是3.5，我们能够用68-95-99.7法则得出结论说95%的结果会落到距离3.5两个标准差之内吗，为什么？
4. 标准分和百分数 表5-1包含了标准分数和百分等级， z 分数可以是一个负值吗？百分等级可以是一个负值吗？说明原因。

判断正误：判断练习题5~8中的陈述是有意义的（表述正确）还是无意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断会更重要。

5. 心理测试 一个心理测试中的分数都是正数，并且符合正态分布，平均数是55，标准差是65。
6. 出生重量 美国人的出生重量符合正态分布，平均重量是3420克，标准差是495克。
7. SAT分数 SAT分数符合正态分布，平均数是1428，标准差是325。
8. SAT分数 SAT分数符合正态分布，平均数是1428，标准差是325。马克的分数1840，是出乎寻常得高。

概念与应用

9. 应用68-95-99.7法则 假定一组测验分数符合正态分布，平均数是100，标准差是20，应用68-95-99.7法则计算下列数量：
 - (1) 分数小于100的百分比。

(2) 分数小于120的相对次数。

(3) 分数小于120的百分比。

(4) 分数小于80的百分比。

(5) 分数小于60的相对次数。

(6) 分数大于120的百分比。

(7) 分数大于140的百分比。

(8) 分数大于80的相对次数。

(9) 分数在80~120之间的百分比。

(10) 分数在80~140之间的百分比。

10. 应用68-95-99.7法则 假设样本的心率检测符合样本分布，平均数是70，标准差是15，应用68-95-99.7法则计算下列数字：

(1) 心率小于70的百分比。

(2) 心率小于55的百分比。

(3) 心率小于40的相对次数。

(4) 心率小于85的百分比。

(5) 心率小于100的相对次数。

(6) 心率大于85的百分比。

(7) 心率大于55的百分比。

(8) 心率小于40的相对次数。

(9) 心率在55~85之间的百分比。

(10) 心率在70~100之间的百分比。

11. 应用68-95-99.7法则 在一个面部表情的研究中，控制组的人们在5分钟的时间段内，对目光接触的时间进行机试，这个机试符合正态分布，平均数是184.0秒，标准差是55.0秒，应用这个法则去计算下述要求的数字：

(1) 计算在距离平均数184.0秒上下浮动55.0秒之内的机试的百分比。

(2) 计算平均数184.0秒上下浮动110.0秒之内的机试的百分比。

(3) 计算距离平均数184.0秒上下浮动

165.0 秒之内的百分比。

- (4) 计算 184.0 ~ 294.0 秒之间的机试的百分比。
12. 应用 68-95-99.7 法则 在一个新款汽车上设计安装了一个 CD 播放器，工程师必须考虑司机向前可以抓到的地方，女性向前可以抓到的地方符合正态分布，平均数是 27.0 英寸，标准差是 1.3 英寸。应用 68-95-99.7 法则计算下述数字：
- (1) 计算女性向前抓握在 24.4 ~ 29.6 英寸之间的百分比。
- (2) 计算女性向前抓握范围小于 30.9 英寸的百分比。
- (3) 计算女性向前抓握范围在 27.0 ~ 28.3 英寸之间的百分比。

智商分数：练习题 13 ~ 24 应用智商分数的正态分布，平均数是 100，标准差是 16，使用表 5-1 计算所要求的数字，注意：表 5-1 呈现了 -3.5 ~ 3.5 的标准分，在下述问题中，标准分超过 3.5 的百分等级是 99.99%，标准分低于 -3.5 的百分等级是 0.01%。

13. IQ 低于 100 的百分比。
14. IQ 低于 84 的百分比。
15. IQ 高于 116 的百分比。
16. IQ 低于 76 的百分比。
17. IQ 高于 132 的百分比。
18. IQ 低于 80 的百分比。
19. IQ 低于 65 的百分比。
20. IQ 低于 129 的百分比。
21. IQ 在 84 ~ 116 之间的百分比。
22. IQ 在 76 ~ 124 之间的百分比。
23. IQ 在 80 ~ 120 之间的百分比。
24. IQ 在 92 ~ 115 之间的百分比。

女性身高：练习题 25 ~ 36 应用女性身高的正态分布，平均数是 162 厘米，标准差是 6 厘米。参照表 5-1 求相应的数字注意表 5-1 呈现了 -3.5 ~ 3.5 的标准分，在下述问题中，标准分超过 3.5 的百分等级是 99.99%，标准分低于 -3.5 的百分等级 0.01%。

25. 身高高于 162 厘米的百分比。

26. 身高低于 168 厘米的百分比。

27. 身高高于 156 厘米的百分比。

28. 身高高于 171 厘米的百分比。

29. 身高低于 177 厘米的百分比。

30. 身高低于 147 厘米的百分比。
31. 身高低于 144 厘米的百分比。
32. 身高高于 179 厘米的百分比。
33. 身高在 156 ~ 168 厘米之间的百分比。
34. 身高在 159 ~ 165 厘米之间的百分比。
35. 身高在 148 ~ 170 厘米之间的百分比。
36. 身高在 146 ~ 156 厘米之间的百分比。
37. 硬币重量 仔细观察下表，表中呈现了 5 种美国硬币的官方平均重量和估计标准差。假定一个自动售货机被设计成，拒绝所有重量超过标准平均重量两个标准差以上的硬币。对于每一种的硬币，计算自动售货机接受的重量范围。在每种情况下的，合法硬币被自动售货机拒绝的百分比是多少？

硬币	重量（克）	估计标准差（克）
1 美分	2.500	0.03
5 美分	5.000	0.06
10 美分	2.258	0.03
25 美分	5.670	0.07
50 美分	11.340	0.14

38. 怀孕时间 孕妇怀孕时间符合一个正态分布，平均时长为 268 天，标准差是 15 天。
- (1) 怀孕时间少于 250 天的妇女百分比是多少？
- (2) 怀孕时间长于 300 天的妇女百分比是多少？
- (3) 如果怀孕时间少于 238 天，新生儿就被看做早产儿，早产儿的百分比是多少？
39. SAT 分数 基于大学布告栏的数据，SAT 分数符合正态分布，平均数为 1 518，标准差是 25。
- (1) 计算 SAT 分数超过 2 000 分的人数百分比。
- (2) 计算 SAT 分数少于 1 500 分的人数百分比。
- (3) 计算 SAT 分数在 1 600 ~ 2 100 分之间的人数百分比。
40. GRE 分数 假定美国研究生入学考试 (GRE) 成绩符合正态分布，平均分是 497 分，标准差是 115 分。
- (1) 一个研究生学院录取分数线是 GRE 分数为 650 分，对应的上线人数百分比是多少？

- (2) 一个研究生学院录取分数为第95百分等级以上,这对应的实际 GRE 分数是多少?
41. 气压计刻度 一个工厂气压计(测量大气压的量具)制造商的监测员在每天的同一时间查看50个气压计的读数。平均读数是30.4英寸汞柱,标准差是0.23英寸汞柱,气压计读数基本呈正态分布。
- (1) 气压计读数超过31英寸汞柱的百分比是多少?
- (2) 气压计读数低于31英寸汞柱的百分比是多少?
- (3) 公司决定如果读数超出或低于平均数1.5个标准差的气压计就被判定为不合格产品。那么低于平均数到什么读数就被判定为不合格产品,高出平均数到什么读数就被判定为不合格产品?
- (4) 在查看这些气压计的读数后,真正的大气压值是多少?请加以解释。
42. 60个女孩儿的平均分是71分,标准差是6分;而50个男孩儿的平均分是66分,标准差是5分。这些学生,分数高于75分被认为是合格的。女孩儿的所占的百分比是多少?
43. 成为一个海员 按照来自美国国家健康调查中心的数据,成年男性的身高是符合正态分布的,平均身高是69.0英寸,标准差是2.8英寸。美国航运公司要求男性的身高在64~78英寸之间。美国成年男性符合海员身高标准的百分比是多少?
44. 电影时间长度 基于一个电影时间长度的随机样本,平均长度是110.5分钟,标准差是22.4分钟。假定电影时间长度符合正态分布。
- (1) 电影时间长度超过两小时的比例是多少?

- (2) 电影时间长度少于1.5个小时的比率是多少?
- (3) 随机挑选一部电影时间长度少于2.5个小时的概率是多少?

网络和远程项目

请在网站 www.aw.com/bbt 上打开第5章的有用链接“links for internet projects”。

45. 数据和故事图书馆 浏览数据和故事图书馆的网页,寻找一个正态分布的例子。写出一页你所发现的说明。包括一个原因变量的详细说明、分布的细节以及你期望这个变量接近正态分布的解释。
46. 各州生命统计 美国国家健康统计中心在各州数据表格中提供了各州(或者各县)的生命统计。寻找一个变量(例如出生率、死亡率或者婴儿死亡率),制作一个呈现各州这个变量的直方图。评价这个变量在各州之间是否符合正态分布?
47. 网页上的正态分布动画演示 在互联网上以关键词“正态分布”进行网络搜索,寻找一个关于正态分布的动画演示。描述你所观察到的动画演示的进程和有用的细节。
48. 估计一分钟时间长短 调查一些被试让他们在不看钟表的情况下估计一分钟,对每个被试说开始让他们开始估计,当他们认为一分钟到的时候自己说停下。(另外,你也可以在不看钟表的情况下自己重复估计一分钟时间,然后观察你估计的时间和正确时间的对比。)使用一个秒表记录实际的时间,绘制一个估计时间的图形。这个图形接近于正态分布吗?平均估计时间接近一分钟吗?



5.3 中心极限定理

一个高中英语教师有100个毕业生要参加大学入学测验。这个测验被设定为平均分是500分,标准差是100分。应用本章所学的方法,她预计个体分数达到多少才能够达到入学标准,比如说600分。但是关于这100名中学生的成绩她能预测什么呢?例如,这一组学生平均成

绩在600分以上的可能性是多少?这种类型的问题,是想要知道在一个非常大的总体中,抽取一组或者一个样本的平均数是多少,中心极限定理可以回答这个问题。

5.3.1 中心极限定理的实验

在我们探讨中心极限定理本身之前,我们通过分析掷骰子的问题加深思考。假定我们抛掷一个骰子1000次,记录每一次抛掷的结果,抛掷的结果以数字1,2,3,4,5,6来表示。图5-23

呈现了结果的直方图。所有六种结果大致来说都有相同的相对次数。因为抛掷骰子的六种可能结果每种出现的机会都是相同的。这样，直方图呈现了一个近似均匀分布（参考 4.2 节）。应用第 4 章所描述的方法，我们能够计算这个分布的平均数和标准差。它显示出图 5-23 中的分布平均数为 3.41，标准差是 1.73。

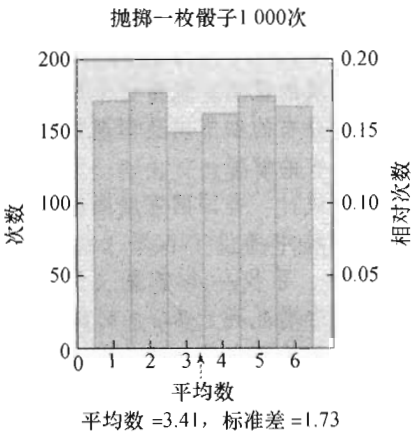


图 5-23 抛掷一枚骰子 1 000 次结果的次数和相对次数分布

假如我们抛掷两枚骰子 1 000 次，每一次抛掷之后记住两个数字的平均数（图 5-24）。先看每

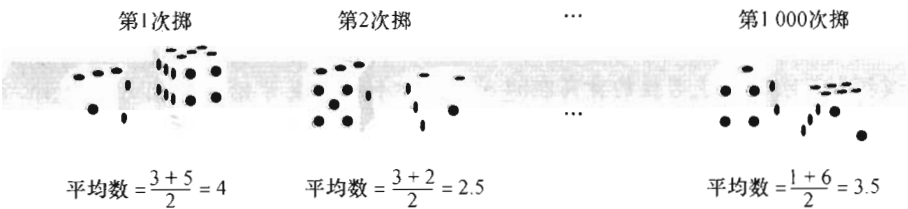


图 5-24 抛掷两枚骰子 1 000 次并记录每一次平均数结果的图示，平均数就是两枚骰子结果总和除以 2，平均数的全距从 $1(1+1)/2 =$ 到 $(6+6)/2 = 6$

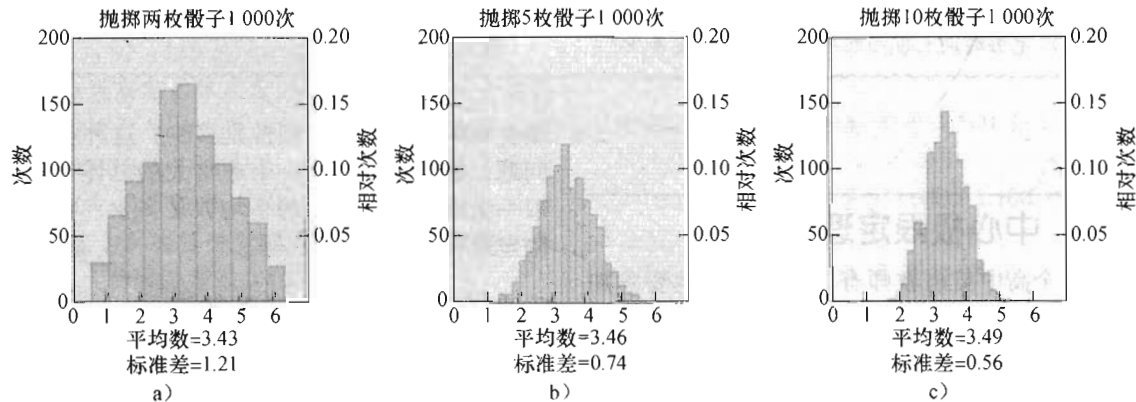


图 5-25 抛掷骰子结果平均数的次数和相对次数分布

a) 抛掷两枚骰子 1 000 次 b) 抛掷 5 枚骰子 1 000 次 c) 抛掷 10 枚骰子 1 000 次

一枚骰子的数字，然后将这两个数字相加并且除以 2。例如，如果两枚骰子的结果分别是 3 和 5，这一次抛掷的平均数就是 $(3+5)/2=4$ ，这样抛掷两枚骰子的可能平均数就是 1.0，1.5，2.0，…，5.0，5.5，6.0。

图 5-25a 呈现了抛掷两枚骰子 1 000 次的结果，这个分布中最常见的数值是中心值 3.0，3.5 和 4.0。这些数值常见是因为它们以几种方式产生。例如，出现平均数 3.5 可以是两枚骰子分别是 1 和 6，2 和 5，3 和 4，4 和 3，5 和 2，或者 6 和 1，较高的平均数和较低的平均数很少发现是因为它们产生的方式很少。例如，抛掷了两枚骰子出现平均数是 1.0 时，只有两枚骰子都是 1 的时候才有可能。并且，计算这个分布的平均数和标准差是可能的，它们分别是 3.43 和 1.21。

如果我们增加骰子数量会出现什么结果？假如我们抛掷 5 枚骰子 1 000 次，并且记录每一次抛掷结果 5 个数字的平均数，这个实验的结果直方图呈现在图 5-25b 中，我们又一次看到围绕中心值 3.5 左右出现次数最多。但是这个分布比前两次分布的狭窄了。计算这个分布的平均数和标准差，它们的数值是平均数 3.46 和标准差 0.74。

如果我们更多地增加骰子的数量，抛掷 10 个骰子 1 000 次，我们看到图 5-25c 的直方图，这个分布变得更狭窄，计算这个分布的平均数和标准差，它的平均数是 3.49，标准差是 0.56。

表 5-2 概括了我们已经描述的 4 种试验结果，表的第 2 列和第 3 列，列出了分布的平均数和标准差。在每种抛掷骰子的试验中，我们记录了 1 000 次抛掷中的每一次平均数（也就是，每一次抛掷一枚骰子的平均数，每一次抛掷两枚骰子平均数，每一次抛掷 5 枚骰子的平均数，每一次抛掷 10 枚骰子的平均数）。这样，所有的 1 000 次抛掷的平均数就是平均数分布中的平均数（第 2 列）。同样，每一次试验中的 1 000 次抛掷的标准差就是平均数分布的标准差（第 3 列）。

表 5-2 抛掷骰子实验结果概括

一次抛掷骰子 的数量	平均数分布 的平均数	平均数分布 的标准差
1	3.41	1.73
2	3.43	1.21
5	3.46	0.74
10	3.49	0.56

从这 4 个实验中得到的一个合理看法浮现出来了，抛掷 $n = 1$ 枚骰子 1 000 次可以被看做，从所有可能的抛掷骰子的总体中，选取的样本容量为 $n = 1$ 的 1 000 个样本，抛掷 $n = 2$ 个骰子 1 000 次可以被看做选取的样本容量为 $n = 2$ 的 1 000 个样本，抛掷 $n = 5$ 枚骰子 1 000 次可以被看做选取的样本容量为 $n = 5$ 的 1 000 个样本，抛掷 $n = 10$ 个骰子 1 000 次可以被看做选取的样本容量为 $n = 10$ 的 1 000 个样本。据表 5-2 显示，随着样本容量的增长，平均数分布的平均数越来越接近的值为 3.5，标准差也变得越来越小（使分布越来越狭窄）。更为重要的是，随着样本容量的增加分布越来越像一个正态分布。后面的实际情况让人看起来更为惊讶，因为我们是从一个均匀分布中抽取的样本（抛掷 1 枚骰子的结果呈现在图 5-23 中），而不是从一个正态分布中选取的样本。不管怎样，只要样本容量足够大，平均数的分布会越来越接近一个正态分布。这个事实就是中心极限定理（central limit theorem）所表达的结果。

中心极限定理

假定我们从任何一种分布中随机选取样本容量为 n 的样本（不一定是从正态分布中），将每一个样本的平均数作分布记录，会有下述结果：

- (1) 如果样本容量足够大，平均数的分布会越来越接近正态分布。
- (2) 随着样本容量增大，平均数分布的平均数越来越接近于总体平均数 μ 。
- (3) 随着样本容量增大，平均数分布的标准差接近 $\sigma/\sqrt{n}$ ， σ 是总体标准差。

注意第三条是非常重要的，当以一个样本来代替一个个体的时候：

样本平均数的标准差不是总体标准差 σ ，而是 $\sigma/\sqrt{n}$ ， n 是样本容量。

思考题

确认表 5-2 中 $n = 2, 5, 10$ 各种情况下平均数分布的标准差，是否与中心极限理论所预测的一致。已知总体标准差 $\sigma = 1.73$ （图 5-23 中所呈现的总体标准差），例如， $n = 2$ 时， $\sigma/\sqrt{n} = \sigma/\sqrt{2} = 1.22 \approx 1.21$ 。

方法注解

(1) 从实际意义上讲，如果样本容量大于 30，样本平均数的分布就接近正态分布；(2) 如果原始总体为正态分布，任意大小的样本容量下，样本平均数的分布都符合正态分布；(3) 从理论上讲，样本平均数的分布是由所有可能的样本平均数形成的，样本平均数分布的平均数等于原总体平均数 μ ，样本平均数分布的标准差等于 $\sigma/\sqrt{n}$ 。

我们总结一下中心极限理论的构成要素，我们从某一个特定的变量开始，比如抛掷骰子的结果和人们的体重，这些变量都是通过在一个总体中随机抽取，变量有一个确定的总体平均数 μ 和总体标准差 σ 。这个总体参数我们也可能知道，也可能不知道。变量可以是任何一种分布，不一定是正态分布。现在我们从这些变量中选取一些样本，每一个样本中有 n 个个体，计算每个样本的平均数（比如抛掷骰子的平均数和 n 个人组成的样本体重的平均数）。如果我们绘制一个许多样本平均数的直方图，我们将会看到这个分布比较接近正态分布。样本容量越大，平均数的分布就越接近正态分布。仔细地研究图 5-26，可以帮助我们巩固这个重要的观念。

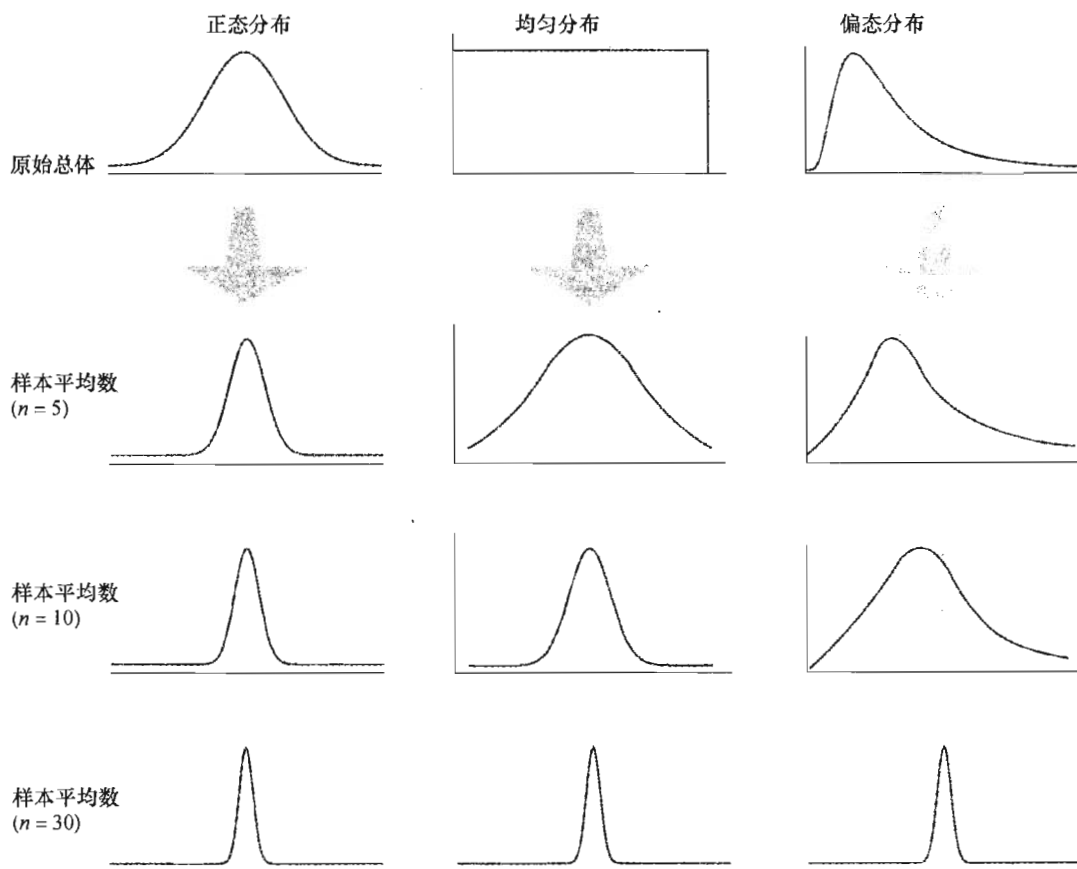


图 5-26 不管原始分布是什么样的，当样本容量增加时 ($n=5, 10, 30$)，样本平均数的分布越来越接近正态分布。样本容量越大，样本平均数分布的标准差越小。

例 5-13 预测测验分数

你是一个中学校长，100 个八年级中学生准备要参加一个联邦标准化测验，这个测验设计的平均数是 400，标准差是 70。假定测验分数是正态分布。

(1) 从这些八年级学生中随机选择一位，测验成绩在 375 分以下的可能性是多大？

(2) 作为一个当事人，你的业绩有赖于全部的八年级小学生的这次测验的成绩是好还是坏。那么这 100 位八年级学生平均分数在 375 分以下的可能性是多少？

解答

(1) 涉及每一个学生的分数，我们应用 5.2 节里所讨论的标准分的方法。已知平均分为 400，标准差是 70，那么 375 分的标准分是：

$$z = \frac{(\text{数据值} - \text{平均数})}{\text{标准差}} = \frac{(375 - 400)}{70} = -0.36$$

按照表 5-1，标准分为 -0.36 对应的百分等级大约是 36，也就是全部学生的 36% 成绩期望在 375 分以下。这样随机抽取一个学生成绩在 375 分以下大约有 0.36 的机会，注意我们需要知道

这个分数是否符合正态分布以便于计算，因为标准分数表仅仅适用于正态分布。

(2) 关于一群学生平均分的问题，必须采用中心极限定理来处理。按照这个定理，如果我们随机选择一个样本，样本容量是 $n=100$ 个学生，计算每组样本的平均测验分数，这个平均数分布是接近于正态的。并且，这个分布的平均数 $\mu=400$ ，标准差是 $\sigma/\sqrt{n} = 70/\sqrt{100} = 7$ ，在这个平均数和标准差的前提下，测验平均分 375 所对应的标准分是：

$$z = \frac{(\text{数据值} - \text{平均数})}{\text{标准差}} = \frac{(375 - 400)}{7} = -3.57$$

表 5-1 显示出标准分 -3.5 对应的百分等级是 0.02，这种情况下的标准分是相当低的。换句话说，随机选择一个由 100 名学生组成的随机样本平均分低于 375 分的机会少于 0.02%。这样说来，选择一个由 100 名学生组成的随机样本平均分低于 375 分的机会小于 0.0002，或者说 5 000 次中大约只可能出现 1 次。注意这个关于样本的平均数的计算并不依赖于样本中的个体分数是否符合一个正态分布。

这个例子给我们上了重要的一课，个体的分数在375分以下的可能性超过了1/3（36%），但是由100个学生组成的群体，其平均分低于375分的可能性低于1/5 000（0.02%）。换句话说，个体的分数要比一个群体的平均分有更大的变异性。

例5-14 薪资的公平性

在Holley.com工作的9 000名雇员的平均薪资为26 400美元，标准差为2 420美元。一名调查员从这些雇员中随机抽取了400名，发现这个样本的平均薪水是26 650美元。这名调查员得到这个结果是否偶然？或者说两个数据不同是否意味着调查员的结果是令人怀疑的？

解答 与400个雇员的平均薪资有关的问题，属于一个中心极限定理的案例。这个定理告诉我们，如果我们选择很多由400个雇员所组成的样本，然后计算每个样本的平均数。这个平均数的分布将非常接近于一个正态分布，平均数 $\mu = 26\,400$ 美元，标准差为 $\sigma/\sqrt{n} = 2\,420 \text{ 美元} / \sqrt{400} = 121 \text{ 美元}$ 。在这个平均数的分布里，26 650美元的平均薪水对应的标准分是：

$$z = \frac{\text{数据值} - \text{平均数}}{\text{标准差}} = \frac{(26\,650 \text{ 美元} - 26\,400 \text{ 美元})}{121 \text{ 美元}} = 2.07$$

换句话说，如果我们假定的样本是随机选取的，样本的平均工资要高于全公司的平均工资两个标准差以上。按照表5-1，2.07的标准分对应的百分等级是98。这样，这个样本的平均薪资，比较起所有的由400

个工人组成的可能样本的平均工资来，这个样本的平均薪资，要高于其中98%的样本平均薪资。也就是，选择一个由400个工人所组成的样本，其平均工资超过26 650美元的可能性大约为2%。这个样本的平均薪资高得让人惊奇，很可能这个调查是有缺陷的。

思考题

一个工人的薪资为26 650美元是处于百分等级98以上还是以下？说明理由。

5.3.2 中心极限定理的意义

如果我们知道总体平均数和总体标准差，中心极限定理可以告诉我们有关样本平均数的很多事情。这是非常有用的，但是更为重要的是它的反向应用。

统计学的两个主要任务是估计总体平均数和对总体平均数进行假设检验。假设我们并不知道一个变量的总体平均数，如果我们仅知道一个小一些的样本平均数，对总体平均数进行良好的估计是可能的吗（比如所有互联网使用者的平均收入）？像你能猜到的那样，能够回答这种类型的问题，统计样本就成了核心，尤其是在民意测验和调查中。中心极限定理提供了回答此类问题的关键。我们将在第8章讨论这个主题。

5.3练习

统计素养和批判性思维

- 方便样本 因为一个上统计课的学生作一个项目可拖延到了最后，她只有一点时间只能从她的女性朋友和女性亲戚中搜集她们的身高数据。她计算了样本中女性的平均身高。假定这些女性的身高符合正态分布，平均数是63.6英寸，标准差是2.5英寸，她能够使用中心极限定理分析样本中的平均身高吗？
- 符号 在这一节中，代表样本平均数的标准差的符号是 $\sigma/\sqrt{n}$ ，在这种表述中， σ 代表什么？ n 代表什么？
- 中心极限定理 一个过程包括如下重复的操作：从一些总体中随机挑选两个数值然后计算这两个数值的平均数。依据中心极限定理，这个样本平均数会符合正态分布吗？为什么？
- 中心极限定理 一个过程包括如下重复操作，

从一个已知标准差的正态分布的总体中随机挑选4个数值，样本平均数的标准差是多少？

概念与应用

- 智商分数和中心极限定理 智商分数符合正态分布，平均数是100，标准差是16，假定从一个非常大的人口总体中抽取许多容量为 n 的样本，每组样本的平均智商分数。
 - 如果样本容量是 $n = 64$ ，计算样本平均数分布的平均数和标准差。
 - 如果样本容量是 $n = 100$ ，计算样本平均数分布的平均数和标准差。
 - 为什么第1小题和第2小题的标准差是不同的？
- SAT分数和中心极限定理 基于大学公告栏上的数据，假定SAT分数符合正态分布，平均数是1 518，标准差是325。假如从一个大

量的学生总体中随机抽取许多容量是 n 的样本，计算每个样本的平均分数。

- (1) 假如样本容量 $n=100$ ，计算样本平均数分布的平均数和标准差。
 - (2) 假如样本容量 $n=2\,500$ ，计算样本平均数分布的平均数和标准差。
 - (3) 为什么第 1 小题和第 2 小题的标准差是不同的？
7. 12 面的骰子和中心极限定理 抛掷 1 枚均匀的 12 面的骰子，产生了一个结果为数字从 1~12 的均匀分布，平均数是 6.5，标准差是 3.452。假如 n 个 12 面的骰子被抛掷许多次，针对每次 n 个骰子的结果计算它的平均数。
- (1) 计算 $n=81$ 时的结果样本平均数分布的平均数和标准差。
 - (2) 计算 $n=100$ 时的结果样本平均数分布平均数和标准差。
 - (3) 为什么第 1 小题和第 2 小题的标准差是不同的？
8. 10 面的骰子和中心极限定理 抛掷一枚均匀的 10 面骰子产生了一个结果为数字从 1~10 的均匀分布，平均数是 5.5，标准差是 2.872。假如 n 个 10 面的骰子被抛掷许多次，针对每次抛掷骰子的结果计算它的平均数。
- (1) 计算 $n=49$ 时的结果样本平均数分布的平均数和标准差。
 - (2) 计算 $n=400$ 时的结果样本平均数分布平均数和标准差。
 - (3) 为什么第 1 小题和第 2 小题的标准差是不同的？

飞机使用年限：在练习题 9~12 中，假定民用航空器的使用年限符合正态分布，平均数是 13.0 年，标准差是 7.9 年（数据来自航空数据服务中心）。

9. 飞机使用年限超过 15 年的百分比是多少？假如随机挑选由 49 架飞机组成的样本，计算样本的平均使用年限。样本平均数超过 15 年的百分比是多少？
10. 飞机使用年限少于 10 年的百分比是多少？假如随机挑选由 84 架飞机组成的样本，计算样本的平均使用年限。样本平均数少于 10 年的百分比是多少？
11. 飞机使用年限在 10~16 年之间的百分比是多少？假如随机挑选由 81 架飞机组成的样本，计算样本的平均数。样本平均数在 10~16 年之间的百分比是多少？
12. 飞机使用年限在 12.5~13.5 年之间的百分比是多少？假如随机挑选由 400 架飞机组成的样本，计算样本的平均使用年限。样本平均数在 12.5~13.5 年的百分比是多少？
13. 可乐的容量 假如一听可乐充满后实际上的容量是符合正态分布的，平均数是 12.00 盎司，标准差是 0.11 盎司。
 - (1) 一个由 36 听可乐组成的样本平均数至少是 12.05 盎司的可能性是多少？
 - (2) 如果第 1 小题的结果是已知的，那么 1 听装满后的可乐实际上平均容量是 12.00 盎司，你认为这是合理的吗？如果这个平均数不是 12.00 盎司，那么消费者被欺骗了吗？
14. 设计闪光灯 飞机上的闪光灯被设计成闪光间隔时间是符合正态分布的，平均数是 3.00 秒，标准差是 0.40 秒。
 - (1) 某次闪光间隔时间大于 4.00 秒的可能性是多少？
 - (2) 60 次随机挑选的闪光间隔时间平均数大于 4.00 秒的可能性是多少？
 - (3) 我们知道闪光灯是用来帮助其他飞机看见这架飞机的，哪一种结果用于评估闪光灯安全性更准确，是第 1 小题的结果还是第 2 小题的结果？为什么？
15. 摩托车头盔的设计 工程师在设计男性摩托车头盔时，必须要考虑男性头围的大小。男性的头围符合正态分布，平均数是 6.0 英寸，标准差是 1.0 英寸（数据来源是戈登、丘吉尔等人的人体测量调查数据）
 - (1) 如果随机挑选一位男士，他的头围小于 6.2 英寸的可能性是多少？
 - (2) 安全卫士头盔公司计划生产 100 个头盔样品，100 个随机挑选的人头围平均数小于 6.2 英寸的可能性是多少？
 - (3) 生产经理看到第 2 小题中的结果后，推理出所有男性头盔应该被制作成头围小于 6.2 英寸的，因为这样除了少数一些人之外对所有人都适合。这个推理错误在哪里？
16. 避免没有热水 为了满足顾客的热热水需求，

Luxurion 酒店的经理计算出顾客每天淋浴的平均时间是 11.4 分钟（基于观点研究公司的数据）。假定淋浴时间是正态分布，标准差是 2.7 分钟。

- (1) 计算淋浴时间超过 12 分钟的旅客百分比。
 - (2) 酒店安装了一个能提供足够热水的系统，可以使得 84 位旅客淋浴的平均时间少于 12 分钟。如果酒店刚好有 84 位旅客，热水不够用的可能性是多少？这个系统看起来是有效的吗？
17. 弹射座椅的重新设计 当女性被允许成为战斗机的飞行员时，工程师需要重新设计弹射座椅，因为以前的设计都是针对男性的。ACES-II 型弹射座椅是根据男性的体重在 140~211 磅之间设计的，女性体重的总体符合正态分布，平均数是 143 磅，标准差是 29 磅（基于美国国家健康调查的数据）。
- (1) 女性体重在 140~211 磅之间的百分比是多少？
 - (2) 如果随机挑选 36 位女性，她们的平均体重在 140~211 磅之间的可能性是多少？
 - (3) 战斗机弹射座椅的重新设计是为了更加适合女性飞行员。哪种可能性与之更相关，是第 1 小题中的结果还是第 2 小题中的结果？为什么？
18. M&M 公司的标签 M&M 公司的普通糖果重量符合正态分布，平均数是 0.856 5 克，标准差是 0.051 8 克（基于本书一位作者的自行测量结果）。从总数为 465 颗糖的包裹中随机选取 100 颗 M&M 公司的普通糖果，作为一个样本，包裹上标签的净重为 396.9 克。（如果每个包裹都有 465 颗糖果，每颗糖果的平均重量就是 $396.9/465 = 0.8535$ ，因为包裹的净重不得以少于 396.9 克）
- (1) 如果一颗普通糖果是随机挑选的，那么它的重量大于 0.853 5 克的可能性是多少？
 - (2) 如果 465 颗普通糖果被随机挑选出来，它们的平均重量至少是 0.853 5 克的可能性是多少？
 - (3) 根据以上结果，推测火星糖果公司是

按照标签上注明的重量给 M&M 公司的消费者提供的糖果吗？

19. 自动售货机 一般来说，25 美分的硬币重量符合正态分布，平均数是 5.670 克，标准差是 0.062 克。一个自动售货机被设定为只接受重量在 5.550~5.790 克之间的 25 美分硬币。
 - (1) 如果 280 枚硬币被塞进自动售货机，你预期有多少硬币会被拒绝？
 - (2) 如果 280 枚硬币被塞进自动售货机，你认为它们的平均重量落在 5.550~5.790 克之间的可能性是多大？
 - (3) 如果你自己拥有一个自动售货机，哪一种结果与你更有关系，是第 1 小题的结果还是第 2 小题的结果？为什么？
20. 飞行器安全标准 按照老的联邦飞行指导规则，航空公司估计每个乘客的重量是 185 磅。（这个说明是针对冬季的成人旅客，包括 20 磅的托运行李。）最近的规则需要估计每位乘客的重量是 195 磅。男性的体重是符合正态分布的，平均数是 172 磅，标准差是 29 磅。
- (1) 如果随机挑选一个成年男性，假定他有 20 磅托运行李，那么它的总重量超过 195 磅的可能性是多少？
 - (2) 如果一架波音 767-300 飞机满载着 213 位成年男性乘客，假定每一位有 20 磅托运行李，那么旅客的平均体重（包括 20 磅托运行李）超过 195 磅的可能性是多少？飞行员必须要符合这个体重限制吗？
21. 遗传变量 假定一个变量，比如体重、血压，在大量的人群中符合正态分布，平均数是 μ ，标准差是 σ 。
- (1) 假如我们抽取了很多样本，每个样本容量 $n=100$ ，并且计算每个样本平均数。样本平均数分布标准差与总体标准差相比较是什么样的？为什么？
 - (2) 假如我们抽取了很多样本，每个样本容量 $n=1000$ ，并且计算每个样本的平均数，样本平均数分布的标准差与第 1 小题的样本平均数分布的标准差比较，有什么不同？这个样本平均数的标准差与总体标准差相比较又是怎样的？

解释你的答案。

- (3) 我们抽取越来越大的样本容量时，样本平均数分布的标准差将是什么样的？
22. 女性的血压 年龄在 18~24 岁之间的女性的血压是符合正态分布的，平均数是 114.8 毫米汞柱，标准差是 13.1 毫米汞柱。
- (1) 某个女性血压超过 125 毫米汞柱的可能性是多少？
- (2) 假如抽取一个随机样本 $n = 300$ 个女性，计算样本的平均血压。样本平均血压超过 125 毫米汞柱的可能性是多少？
- (3) 假如抽取一个随机样本 $n = 300$ 个女性，计算样本的平均血压，样本平均血压低于 114 毫米汞柱的可能性是多少？

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 5 章的有用链接“links for internet projects”。

23. 互联网上的中心极限定理 在互联网上以“中心极限定理”为关键词进行搜索，就会发现很多网页与此有关。寻找一个包含中心极限定理的动画演示的网页。用自己的话描述你所观察到的，动画中怎么样说明中心极限定理。
24. 互联网上的五数概括法 在网络中搜索，寻找一个包含五数概括法的动画演示网页。描述五数概括法，并且解释它是怎样说明中心极限定理的。
25. 抛掷骰子 用抛掷骰子说明中心极限定理，如同本节中所讨论。给你班级里的每个人发许多枚骰子。让他们抛掷一枚骰子并且绘制结果的直方图。然后让他们每个人抛掷两枚骰子并且绘制每次抛掷结果平均数的直方图。如果时间允许尽可能增加抛掷骰子的数目。评价每一种情况抛掷结果直方图的样子。

第5章复习题

- 对于下面每种情况，说明数据的分布是否是一个正态分布。给出你所判断结果的简要解释。
 - 一个轮盘赌上旋转到的数字结果（有 38 个相等的扇区，分别标注的数字 0, 00, 1, 2, 3, ..., 36）。
 - 成年金毛猎犬的体重。
 - 18 岁驾驶员的刹车反应时。
- ACT 测验的分数符合正态分布，平均数是 21.1，标准差是 4.8。
 - 应用 68-95-98.7 法则，计算 ACT 测验分数在平均数 21.2 上下浮动 9.6 范围内的百分比。
 - 应用 68-95-98.7 法则，计算 ACT 测验分数在平均数 21.2 上下浮动 14.4 范围内的百分比。
 - 35.6 这个分数很少出现吗？为什么？
- 假定健康成年人的身体温度是符合正态分布的，其平均温度是 98.20°F ，标准差是 0.62°F （基于马里兰大学的研究数据）。
 - 如果你的身体温度达到了 99.00°F ，对应的百分等级是多少？
 - 将 99.00°F 转换成标准分数。
 - 身体温度达到 99.00°F 是不寻常的吗？为什么？
 - 随机挑选 50 个成年人。他们的平均体温是 97.98°F 或者更低一些的可能性是多少？
 - 发现一个人的体温是 101.00°F ，这个结果极其罕见吗？为什么？我们将得出什么结论？
 - 第 95 百分等级对应的身体温度是多少？
 - 第 5 百分等级对应的身体温度是多少？
 - 纽约市的 Bellevue 医院采用 100.6°F 作为发烧的身体温度最低标准。正常和健康成年人被看做发烧的百分比是多少？这个百分比提示 100.6°F 的发烧标准线是恰当的吗？
 - 如果我们假定身体的平均温度不是 98.20°F ，而是 98.60°F （如同许多人相信的那样），那么随机挑选 106 个人，则他们的平均体温是 98.20°F ，或者更低的可能性是多少（继续假定身体温度的标准差是 0.62°F ）？马里兰大学的研究者确实得到了一个这样的研究结果，我们应该对此会得出什么结论？

第5章小测验

- 下面哪种陈述是正确的？
 - 正态分布并不是一种很少出现的分布。
 - 正态分布的形状是钟形的。
 - 如果总体符合正态分布，那么平均数和中位数是不相等的。
 - 在一个正态分布中，标准差为0是可能的。
 - 正态分布的图形是对称的。
- 一个知觉测验被设计成分数符合正态分布，标准差为50，标准差为10。应用68-95-98.7法则，计算在平均数50分上下20分范围内的百分比。
- 一个总体符合正态分布，平均数为50。随机挑选样本容量为100的一个样本，这个样本平均数的平均数是多少？
- 一个总体符合正态分布，标准差为5，随机挑选样本容量100的一个样本，这个样本平均数的标准差是多少？
- 一个总体符合正态分布，平均数为50，标准差为10。那么70对应的标准分数是多少？
- 一个总体符合正态分布，平均数为50，标准差为10。那么40对应的标准分数是多少？
- 一个测验分数的总体符合正态分布，平均数为50，标准差为10。那么测验分数高于50分的百分比是多少？
- 一个测验分数的总体符合正态分布，平均数50，标准差为10。如果97.72%的测验分数低于70。那么分数大于70的百分比是多少？
- 一个测验分数的总体符合正态分布，平均数50，标准差为10。如果97.72%的测验分数低于70。那么测验分数少于30分的百分比是多少？
- 下面哪种分布更接近于正态分布？
 - 一枚骰子被抛掷很多次以后出现的结果。
 - 两枚骰子被抛掷很多次以后，计算每次抛掷结果的平均数后，产生的平均数分布。
 - 5枚骰子被抛掷很多次以后，计算每次抛掷结果的平均数后，所产生的平均数分布。

技术应用

表5-1包括与按一定间距所给出的标准分数相对应的百分等级，但是一般的统计软件包能够应用所有的标准分数。表5-1是基于从左向右的累加数据所对应的百分等级。一般的统计软件包也采用同样的公式。例如，表5-1呈现出当标准分数 $z=2.0$ 时，所对应的累加百分等级是97.72。统计软件包产生的从左向右累加的面积（或者概率）是0.9772。

SPSS

- 在SPSS软件中，为了计算某一给定数值所对应的从左到右的累加面积（或者概率），可以按照以下步骤进行。
- 进入SPSS软件的数据编辑窗口，找到变量 x ，然后按下[ENTER]键。
 - 在SPSS软件的数据编辑窗口，点击菜单栏中的[Transform]。
 - 点击下拉栏中的[Compute]。
 - 在“Compute variable”窗口，进行以下两步操作：
 - 将“prob”输入“Target Variable”框
 - 将“IDF. normal (p , mean, standard de-

框中。

- 将“cdf. normal (x , mean, standard deviation)”输入“Numeric Expression”栏中。（如果使用标准分数，输入平均数为0，标准差为1。）

5. 点击[OK]。

- 从左向右到 x 值的概率值（或者对应的面积区域）将被呈现出来，点击这个呈现的值可以看到更多的小数位数。

- 为了计算从左向右的累加面积所对应的 x 值，可以进行以下操作：

- 进入SPSS软件的数据编辑窗口，找到从左向右到 x 值的概率（或者面积），然后按下[ENTER]键。
- 在SPSS软件的数据编辑窗口，点击菜单栏中的[Transform]。
- 点击下拉栏中的[Compute]。
- 在“变量计算”窗口，进行以下两步操作：
 - 将“x”输入“Target Variable”框。
 - 将“IDF. normal (p , mean, standard de-

vation)”输入“Numeric Expression”框中，每一个输入都对应特定的值。 p 代表从左向右直到期望值的累加面积。（如果使用标准分数，输入平均数0和标准差1。）

5. 点击[OK]。

6. x 值将被呈现出来，点击这个呈现的值可以看到更多的小数位数。

EXCEL

1. 计算从左边开始到这个特定的值的累加面积，点击 fx 然后选择了Statistical, NORMDIST。在对话框中，输入 x 变量数值（比如2），输入平均数和标准差，并且在“cumulative”栏中输入1。（如果使用标准分数，输

入平均数0，标准差1，并且在“cumulative”栏中输入1。）

2. 计算一个已知面积所对应的数值，点击 fx 然后选择Statistical, NORMDIST。然后在对话框中进行以下操作（如果使用标准分数，输入平均数0，标准差1）。输入概率值时，对应的是从左到某个给定值的面积。

STARDISK

选择“Analysis, Probability Distributions, Normal Distributions”。输入任意一个标准分数可以计算所对应的累加面积（或概率），或者输入任意一个从左向右的某一面积来计算对应的标准分数。输入完一个数值后，点击“Evaluate”。

| 聚焦教育 |

我们能从 SAT 考试的发展趋势中了解到些什么

从1941年11000个学生参加测验开始，学习能力测验（SAT）就成了美国高中生进入美国大学的标准入学考试。现在，每年有将近150万高中生参加这个测验，有时甚至会有全美大约一半的高中毕业生参加这个测验。2006年以前，它分为两大部分：语言部分和数学部分。2006年，语言部分变成评判性阅读部分，另外增加了写作部分。数学部分的测验变化也一样。

SAT的分数最开始是为平均分500，标准差100，最大值和最小值分别是800和200，一般情况下，分数在3个标准差范围内。（如果分数超过平均数3个标准差或低于平均数3个标准差，成绩会记作800分或200分。）

但是，每年的平均分数并不总是500。每年的测验都和前一年会有一些相同的内容，大学委员会（负责测验设计的组织）会用这些相同的问题来把本年度和前一年度的学生做个比较。大学委员会认为500分代表着相同的测验成就水平，不管今年的测验是怎么样。一年又一年的变异代表着每年的真实变化。比如，如果某年平均数高于500分，它表明这一年的学生比平均数为500的那一年的学生成绩更好些。

SAT分数的变化趋势已经被广泛地应用到评估美国全国教育上。图5-27表明了从1972~2006年间语言/评判性阅读以及数学部分的平均分。分类标志“语言/评判性阅读”指的是2006年以前的语言部分和2006年的评判性阅读部分。请注意两部分的分数在20世纪90年代早期就有所下降，近年来又有所回升。写作部分的分数结果没有显示，因为这部分资料太少，还显示不出趋势来。

如果这些分数真的是能够比较的话，那么这个趋势显示最近参加测验的学生比十几年前的学生成绩差，和这些分数相关的分数也是这样。但是在我们接受这个观点之前，我们必须回答下面两个重要的问题：

1. 高中学生样本的趋势能代表整个高中学生总体的情况吗？
2. 除了一小部分相同的问题部分外，测验内容每年都在变。这一年的分数能和其他年的分数作比较吗？



很可惜的是，这两个问题都没有一个明确的肯定回答。比如，许多人认为 SAT 分数的长期下降是由于参加测试的样本发生变化。在 20 世纪 70 年代，只有 1/3 的高中毕业生参加了测验，到了 2006 年有将近一半的高中毕业生参加。这些样本代表着高中学生最好的那批，平均分的下降仅仅只反映了好学生的能力。SAT 的各州资料都说明了这个问题。一般来说，分数最高的州也是参加测验人数比率最低的那个州。对于这种趋势的一种推论解释是这种解释有更多资料支持，当一个州参加考试的学生很少时，这些学生一般都是这个州最好的一些学生。

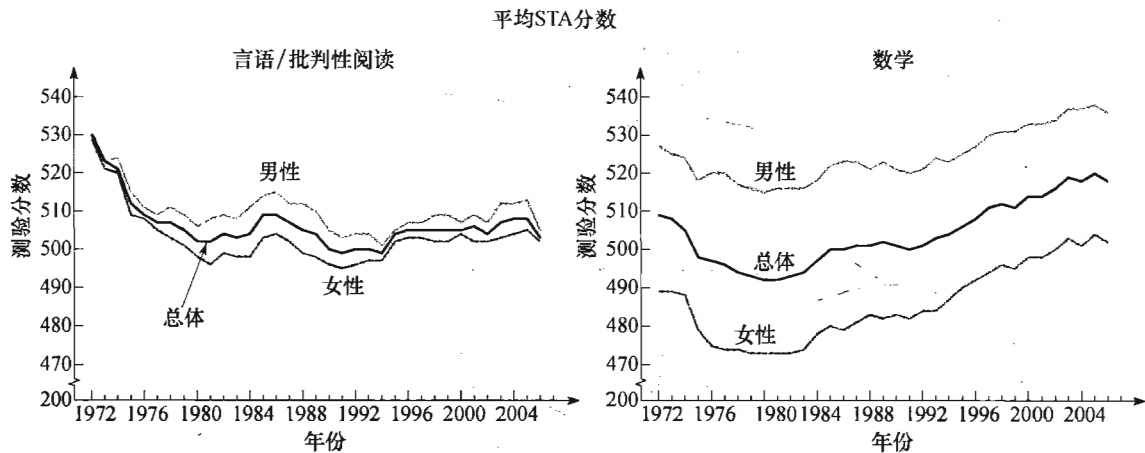


图 5-27 1972~2006 年 SAT 测验中的语言/评论性阅读和数学分数

注：在 1996 年之前的分数显示的是调整后的分数，而不是原始分。

资料来源：The College Board.

第二个难题可能更困难。即便一些测验问题在下一年还出现，在高中更难的数学题会被当做重点，因此这就改变了本年度和下一年度学生回答这些问题的状况。

更复杂的情况出现在科目发生改变的时候。比如，1994 年第一次在数学考试中被允许使用计算器，还增加了考试时间（语言部分也有了明显的变化）。那一年数学分数上升了只是一个巧合吗？或者说考试变得更容易了。相似地，在 2006 年评判性阅读和数学分数在近 30 年内，下滑得非常厉害。评论者认为分数的下降是由于测验中增加写作之后更长更难了。而大学委员会认为下降是由于重复参加考试的人变少了。（重复参加考试的人会让平均分提高，因为这些学生在第 2 次或者第 3 次参加考试中能获得高分。）很明显，测验结构的改变使得通过分数的变化来看教育水平的变化非常困难。

其实，SAT 分数非常明显的变化，从 1996 年就开始了。在那时，分数非常明显地从 500 均分下降了，语言部分的分数下降到 420 分，而数学部分的分数下降到 470。因为分数的变化范围是 200~800，这些下降说明最小值和最大值不再能反映距离平均数远近的标准差数量了。这增加了统计分析的难度。大学委员会因此决定把所有平均分以 1996 年为参考标准重新调整到 500。这个调整影响了不同的百分等级，事实上，在语言分数方面增加了 80 分，数学方面增加了 30 分。图 5-27 呈现了这次调整的结果，1996 年之前，语言部分有 80 分的变动，数学方面有 30 分的变动。比如，如果图 5-27 显示 1994 年的语言均分是 500 分，如果你看 1994 年的报告的话，你可以看到 1994 年的语言均分是 420。

大学委员会认为关于 SAT 的统计分析是非常仔细严谨的，尽管期间有调整 and 变化。但 SAT 的变化趋势还是如实地反映了教育水平的变化。批评者则基本同意统计分析做得不错，但是认为需要考虑其他一些因素，比如参加测验的人口变化以及教育系统的变化，这样才能做一个有效的比较。对于学生来说，最关键的事实在于许多大学一直沿用 SAT 测验分数作为评价入学申请者的一个标准。只要这种情况还存在，SAT 就会一直是全美国最重要的测验，而对于它价值的怀疑也会一直持续下去。

问题讨论

1. 你认为比较两个年度的 SAT 分数是有意义的吗？如果超过十年呢？你同意纵向发展趋势说明现在的学生语言能力比以前的学生差，数学能力比十年前的学生好的观点吗？请加以说明。
2. 语言能力的下降趋势是大学委员会增加写作测验的一个主要原因。他们希望新测验能让学生在高中时期重视写作，因而能提高写作和口头表达能力。你觉得呢？你认为这么做会成功吗？
3. 注意男生的分数一直是高于女生的。你认为这种情况是为什么呢？你认为我们教育体系的改革能改变这种

差距吗？请加以解释。

4. 和你的同学谈谈你的 SAT 测验的知识。根据你个人所了解的知识，你觉得 SAT 测验考察的是哪些方面呢？你认为 SAT 测验作为考核学生是否可以获得大学的申请许可是否合理呢？为什么？

建议阅读

1. Arenson, Karen W. "Scores on Reading and Math Portions of SAT Show Significant Decline." *New York Times*, August 30, 2006.
2. College Board. "The Effects of SAT Scale Recentering on Percentiles." Research Summary RS-05, 1999, available at http://www.collegeboard.com/repository/rs05_3962.pdf.
3. Crouse, James, and Trusheim, Dale. *The Case Against the SAT*. University of Chicago Press, 1988.

| 聚焦心理学 |

我们比父母更聪明吗

大多数小孩倾向于认为他们比自己的父母更聪明，但这正确吗？如果你相信智力测验的结果，那么就会发现，不仅我们比我们的父母聪明，而且我们的父母也比他们的父母也就是我们的祖父母聪明。实际上，如果生活在 100 年以前，我们很可能被看做天才。当然，在我们开始把我们自己当做爱因斯坦一般来吹捧之前，我们最好研究一下背后存在的道理。

智商的概念代表的是一个人的智力商数，它是由法国的心理学家阿尔弗雷德·比纳提出来的。他编制了一个测验用来测量那些在学校里需要特别帮助的学生。他测试了非常多的儿童，并且通过比较每个孩子的心理年龄和生理年龄来计算每个孩子的智力商数。例如，一个 5 岁的孩子，他的分数和一个 6 岁的孩子相同，那么，他的 IQ 就是 6 除以 5 再乘以 100，也就是 120。值得注意的是，通过这样的定义，智力测验对于孩子来说是合理的。然而，后来的研究者，特别是美国军队的心理学家，将智商的概念运用到成人领域。

今天智商被定义为平均数是 100，标准差是 16 的正态分布。按照 68-95-99.7 法则，68% 的人智商分数在 84 ~ 116 分之间，95% 的人智商分数在 68 ~ 132 分之间，99.7% 的人智商分数在 52 ~ 148 分之间，从传统意义上讲，心理学家将智商低于 70 分的（大约低于平均数 100，两个标准差）看做智力低下，那么分数在 130 分以上（高于平均分 100 两个标准差）的则被看做智力超常。

你可能意识到智力测验的矛盾性，由此引发了两个重要的争论：

1. 智力测验能测查出智力或者其他相关因素吗？
2. 如果智力测验能测量智力，测量的是天生遗传的成分还是被环境教育所塑造出来的成分呢？

关于这些争论的讨论有很多，但没有定论。近来，智力测验呈现出一条新的发展思路。正如我们看到的，智力测验的分数后面还有很多东西。现在有几种不同的智力测验，大多数都是在不断地更新和变化着。但是，在所有的智力测验测试中，平均数都是 100，标准差是 16。但是，智力测验的分数呈曲线状态。因为调整的原因，智商的均分一直是 100，这让测量智商分数下降还是上升变得不太可能。

然而，少数智力测验没有随时代而更新，包括一些军队中使用的，另外一些测验虽然更新了但也保留了一些以前的题目。20 世纪 80 年代早期，政治学教授詹姆斯·弗林开始关注到这种在更新的题目和测验中没有调整的原始分，结果令人大吃一惊。

詹姆斯·弗林博士发现，智力测验原始分一直在增长，包括各种类型的测试。增长最高的测试是测量抽



象推理能力的，比如瑞文测试。对于这些测试，他发现现代社会以来，每10年，智商分数会增长6%。一个人在2000年得100分，相当于在1990年的106分，1980年的112分。100年以来，提高了60分，这表明现在被认为智力低下的70分在20世纪初是被认为智力超常的130。

智力测验分数持续上升的趋势现在被叫做弗林效应。现在所有类型的智力测验都是这样，但是程度上和抽象推理测试不一样。比如，图5-28所显示的1932~1997年间，智力测验的结果是如何大范围变化的。值得注意的是，如果按照没有调整的分数，每年的平均分要增长20分。另外，如果一个人在1997年的测试中得100分，他在1932年可能得120分。正如图中所显示的，1997年1/4的测试者会在1932年的测试中被评为智力超常。也有资料显示近来这种增长趋势变缓了。当然，这些资料也是备受争议的。

许多其他科学家也发现过弗林效应。对于这种长期趋势的真实性也有一点怀疑。这是很清楚的：无论是什么类型的智力测验，现在的人们都比几十年以前的人分数高。如果测试能测量智力，那么它意味着一般来说我们比我们的父母聪明，父母比祖父母聪明。

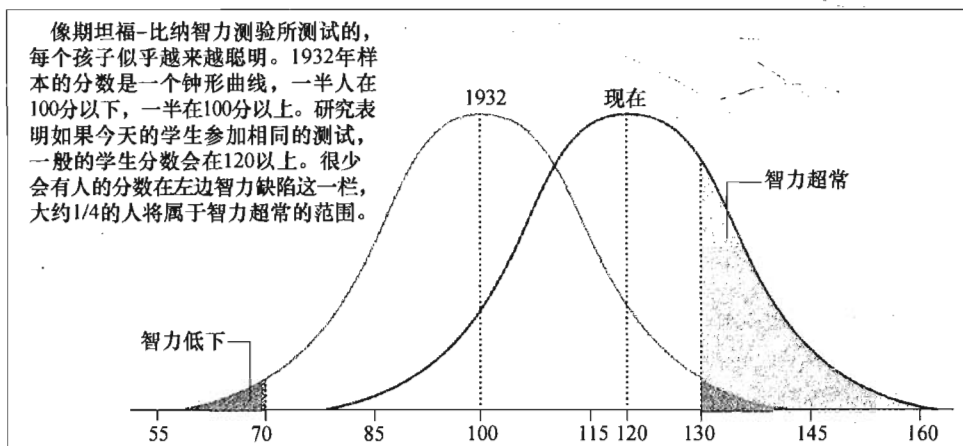


图 5-28 曲线变得更高了

资料来源：New York Times.

当然，如果智力测验不仅能测量智力，而且能测量其他类型的能力，增长的分数也说明今天的孩子比过去的孩子技能更强。事实上最大的增长在抽象思维能力方面。这支持了这种观点。这些测试包括解决问题、寻找最佳处理方式方面的内容。这些问题和过去相比都更容易被儿童解决。

但是弗林效应并不能回答智力测验是否能测验智力，它只能告诉我们一个重要的事情。如果智商真的增长了，智商就不是一个内在的遗传性的特质，因为内在遗传的特质不会在几十年里发生很多变化。因此，如果智力测验测量的是智力，智力能被环境因素塑造。相反地，如果智力是遗传的，那么智力测验就不能测量它。

弗林博士的发现已经改变了心理学家对智力测验的看法。它被认为是将来几十年重点关注的问题。而且很多现代科学会应用到智力测验中。弗林效应可能有深远的社会和政治影响。回到我们最初的问题，我们是否比我们的父母聪明，我们真的不能这么说，但是我们希望是这样的，因为这有利于人类发展。

问题讨论

1. 你喜欢对弗林效应的哪一种解释：是人们变得更聪明还是人们在智力测验中的技能更高了。说明你的理由。
2. 智力测验的成绩上升，比如 SAT，近几十年来很明显。请想想可能的原因，给出一个最可能的解释。
3. 在不同的民族之中，智力测验的结果是不一样的。一些人利用这个结果，来证明一些民族比另外的民族更聪明。这种观点能被弗林效应支持吗？给出你的充分意见。
4. 讨论一下使用智力测验的广泛性。你认为智力测验应该有目的地被使用吗？弗林效应改变了你对智力测验的看法吗？解释一下。

延伸阅读

1. Hall, Trish. "I. Q. Scores Are Up, and Psychologists Wonder Why." *New York Times*, February 24, 1998.
2. Neisser, Ulric (ed.). *The Rising Curve: Long-Term Gains in IQ and Related Measures*. American Psychological Association, 1998.



只有当它变好时你才能理解
什么是运气的变化。

——奥格登·纳什

第 6 章 统计中的概率

学习目标

■ 6.1 概率在统计学中的作用：统计显著性

理解统计显著性的概念，以及概率在定义这个概念时不可或缺的作用。

■ 6.2 概率基础知识

掌握如何通过概率定理和相对次数法计算概率，理解怎样建构基本概率分布。

■ 6.3 概率的大数定律

理解大数定律并应用它掌握和计算期望值，了解

因为对大数定律的误解而导致的赌徒谬论。

■ 6.4 风险观念和预期寿命

计算并解释旅行、疾病和寿命预期等方面风险变化的概率。

■ 6.5 组合事件的概率（补充部分）

区分独立事件和非独立事件、同时事件和非同时事件，并计算“和”与“或”的概率。

大多数统计研究旨在从一个较小的样本中发现其所在总体的一些规律，一项统计研究的关键问题是从一个样本推论到总体是否有效。为了回答这个问题，我们必须掌握已知样本的数据适用于总体的可能性或者概率。在本章我们将重点学习统计学常用的概率基础知识，同时也将了解概率本身的应用。



6.1 概率在统计学中的作用： 统计显著性

为了更好地理解概率在统计学中的重要性，让我们从一枚硬币抛掷的简单例子开始。假设我们想检验一枚硬币是否做假，就要看抛掷它时正面朝上和反面朝上落地的机会是否相等。比如抛掷一枚硬币 100 次，结果 52 次正面朝上，48 次反面朝上，能得出结论说这枚

硬币做假了吗？不能，尽管我们知道正常情况下正面和反面朝上的机会是相等的，抛掷硬币 100 次，预期的结果大致应该是 50 次正面朝上，50 次反面朝上。但是我们也应该认识到每 100 次抛掷硬币试验的结果是有变化的。与 50 正 - 50 反的结果有一些小的偏差并不能说明这枚硬币是做假的，因为这些小的偏差总会偶然发生的。

假设现在抛掷 1 枚硬币 100 次，结果是 20 次正面朝上，80 次反面朝上，这与 50 正 - 50 反有较大差异，这看起来不像一枚正常的硬币

抛掷随机发生的。换句话说,这枚硬币更像是做假的。当观察的现象与预期结果之间的差异不能用随机来解释时,我们就可以说这种差异具有统计显著性(statistically significant)。

例 6-1 不确定事件

1. 底特律的一个侦探发现,在过去的一周里发生的 62 起枪击案中,有 25 起用的枪是在同一家枪支商店里买的。这个发现具有统计显著性。因为在底特律地区有很多家枪支商店,62 起枪击案中就有 25 起的用枪来自同一家商店不大可能是偶然的。

2. 20 世纪,全球平均温度最高的 5 年都出现在 1990~1999 年这段时间,这是具有统计显著性的。通常情况下,一个世纪里最热的 5 年是随机出现的,不大可能出现在同一个 10 年当中。这个统计显著性说明全球正在变暖。

3. 一个具有最差比分记录的篮球队赢了联赛卫冕冠军。这场胜利并不具备统计显著性,因为我们预料到这样的球队会输掉大多数比赛,同样我们也会预料这样的球队也会偶尔取得胜利,甚至战胜联赛卫冕冠军。

6.1.1 从样本到总体

让我们来看一个民意测验的例子:假设随机抽取 1000 人,其中 51% 支持总统。一个星期之后,再另外抽取 1000 人,结果只有 49% 支持总统。我们能否得到这样的结论,一个星期内美国人对总统的态度发生了变化?

你可能猜到答案是没有变化。民意测验的结果是样本统计量:第一个样本中 51% 的人支持总统。我们可以根据这个样本统计量去估计一个总体参数:全部美国人支持总统的人数比例。如果第一次民意测验实施得比较成功,我们就可以推测美国人支持总统的人数比例接近 51%。同样,第二次测验得到的结果可以推测美国人支持总统的人数比例接近 49%。因为这两个样本统计量之间的差异很小(51% 对 49%),很可能美国人支持总统的人数比例实际上并没有变化,这两次民意测验的结果反映了合理的和预期中的两个样本间的差异。

相反,假设第一次民意测验样本中 75% 的人支持总统,而一个星期后,第二次测验只有 30% 的人支持总统。假如民意测验实施得很精细,两组随机抽取的 1000 人之间的差异不可能是偶然发生的。在这种情况下,我们就要去寻找另外的解释,也许一星期内美国人对总统的

态度真的发生变化了。

就统计显著性而言,在第一组民意测验中,从 51% 到 49% 的变化是不具有统计显著性的,我们可以合理地把两个样本间的变化归因于随机误差。然而,第二组测验中,从 75% 到 30% 之间的变化是显著的,因为这种差异不大可能是偶然发生的。

方法注解

对于一项典型的调查来说,49% 和 51% 的差异是不具有统计显著性的。但如果调查了一个非常大的样本的话,这个差异则可能显著的。通常情况下,如果样本容量足够大的话,任何差异都可能具有统计显著性。

例 6-2 实验中的统计显著性

一个研究者设计了一个双盲实验来检验一个新的草药方是否对预防感冒有效。在 3 个月中,随机抽取 100 个人作为实验组,让他们服用这种草药;按同样方法再抽取 100 个人作为控制组,让这些人服用安慰剂。实验结果显示,实验组 30 人感冒,控制组 32 人感冒。我们能否得出结论说这个草药方对预防感冒有效?

解答 一个人在 3 个月的时间里会不会感冒要受很多不可预测的因素影响。因此,我们不能预期任意两组 100 人的群体患感冒的人数完全相等。在本实验中,实验组 30 人感冒和控制组 32 人感冒的差异很小,完全可以解释成是偶然发生的。这种差异不具有统计显著性,因此我们并不能得出草药方有效的结论。

6.1.2 统计显著性的量化

在例 6-2 中,我们说实验组 30 人感冒和控制组 32 人感冒,两组间的差异是不显著的。这个结论是显而易见的,因为差异非常小。假设实验组 24 个人感冒,控制组 32 人感冒。24 人与 32 人之间的差别是不是大到足够具有统计显著性呢?到目前为止,我们使用的统计显著性的定义对于回答这个问题还是太模糊。我们需要一个量化统计显著性的方法。

通常,我们应用概率来量化一个事件偶然发生的可能性,以此来界定统计显著性。因此我们会问这样一个问题:观察到的差异发生的概率是否小于或等于 0.05? 如果是,我们就认为差异在 0.05 水平上显著;如果不是,我们就

认为这种差异的发生实属偶然，差异不具有统计显著性。

选择 0.05 水平多少有点武断，但它是统计学经常用到的数字。不过有时也会应用其他概率水平，比如 0.1 和 0.01。0.01 比 0.05 的统计显著性水平高，0.05 比 0.1 统计显著性水平高。

统计显著性的量化

- 如果观察到的差异偶然发生的概率 $P \leq 0.05$ ，则这种差异在 0.05 水平显著。
- 如果观察到的差异偶然发生的概率 $P \leq 0.01$ ，则这种差异在 0.01 水平显著。

运用统计显著性时应当注意，20 次试验中不超过 1 次我们就可以得出统计结论说它在 0.05 水平显著，当然这个事件真的可能是偶然发生的。从某种角度上说，统计显著性在 0.05 水平——或者在任何水平，都不能保证一个重要的效应或者差异真的存在。

6.1 练习题

统计素养和批判性思维

1. 统计显著性 一个实验旨在通过性别选择的方法增加女孩的出生率，得到的结果是，20 对夫妇生了 11 个女孩和 9 个男孩。实验公司的发言人声称，生女孩的概率超出了 50% 正是这种方法有效的证据。你同意这样的说法吗？为什么？
2. 统计显著性 统计显著性这个术语是指结果是显著的，是不是同样也意味着结果非常重要？解释它的意思。
3. 统计显著性 如果某一结果在 0.05 水平显著，那么它一定在 0.01 水平显著吗？为什么？
4. 统计显著性 如果某一结果在 0.01 水平显著，那么它一定在 0.05 水平显著吗？为什么？

判断正误：判断练习题 5~8 中的陈述是有意义的（表述正确）还是无意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. 酒驾 酒驾致死的概率具有统计显著性，因为它在好多人身上得到了证实。
6. 统计显著性 在一个检验性别选择方法的试验中，100 对夫妇生了 80 个女孩和 20 个男孩。因为结果具有统计显著性，所以不可能

思考题

假定实验发现采用新的草药疗法的人比服用安慰剂的人更少感冒，并且这个结果在 0.01 水平上具有统计显著性。这个实验能否证明草药疗法有效果，为什么？

例 6-3 脊髓灰质炎疫苗的显著性

在索尔克脊髓灰质炎疫苗实验中，实验组 200 000 个儿童中患麻痹性脊髓灰质炎的 33 人，控制组 200 000 个儿童中患麻痹性脊髓灰质炎的 115 人，计算结果表明两组间的差异偶然发生的概率小于 0.01。请你描述结果蕴含的意义是什么？

解答 脊髓灰质炎疫苗实验的结果在 0.01 水平上显著，这意味着实验组与控制组之间的差异偶然发生的概率只有 0.01 或者小于 0.01。因此我们可以相当确定实验组中疫苗的使用是有效的。（实际上，索尔克实验的结果偶然发生的概率比 0.01 更小，所以研究者更相信疫苗是有效的：我们在第 9 章还会讨论，这种概率叫做 P 值。）

是偶然发生的。

7. 性别选择 一个检验性别选择技术的试验，100 个刚出生的孩子中至少包括 80 个女孩。因为 100 个孩子中至少包括 80 个女孩这个事件只有十亿分之一的概率，所以结果是显著的。
8. 临床实验 在一个减少背部疼痛疗法的临床实验中，结果显示，治疗组和对照组（不治疗）的差异显著，这表明这种方法对于治疗每个人的背部疼痛肯定都有效。

概念与应用

主观判断显著性：练习题 9~16 中，判断每个事件实际发生的和理论上随机发生之间的差异是否显著，并讨论其意义。

9. 抛掷硬币 在 500 次抛掷硬币试验中，255 次反面朝上。
10. 抛掷硬币 在 500 次抛掷硬币试验中，400 次反面朝上。
11. 掷骰子 在 60 次掷骰子试验中，3 点朝上从未出现。
12. 转轮盘 轮盘上有 38 个扇格，分别记为数字 0, 00, 1, 2, 3, ..., 36。5 次转动轮盘，数字 5 连续 5 次出现在相同的扇格。

13. 调查 在美国实施的一次成人调查, 一个民意测验专家声称他随机选择了 20 个人都是女子。
14. 统计课 一个老师在她的第一节统计课上发现所有女生的身高都在 6 英尺以上。
15. 陪审团的组成 一个拒交子女抚养费的案件中, 陪审团的组成人员包括 6 名男子和 6 名女子。
16. 临床实验 在治疗过敏症的新药临床实验中, 实验组 80 个被试 5 人出现头疼症状, 控制组 160 个被试 8 人出现头疼症状。
17. 燃料测试 选择 30 辆同样的小汽车进行燃料测试, 其中一半的车加平常的汽油, 另一半加新的实验燃料。第一组的车每加仑行驶 29.3 米, 第 2 组的车每加仑行驶 35.5 米。讨论这种差别是否具有统计显著性。
18. 腕骨神经综合征治疗 进行一个实验证明分别采用手术与夹板固定术治疗腕骨神经综合征是否有差异。手术的 73 例治愈率是 92%, 夹板固定术 83 例治愈率 72%。讨论实验得到的差异是否具有统计显著性。
19. 性别选择 遗传和试管婴儿学会进行了一次性别选择方法的临床实验。在写这本书的时候, 325 名婴儿出生了, 他们的父母都曾应用光谱仪方法, 这种方法设想能够增加女孩出生概率, 其中 295 名婴儿是女孩。讨论这个结果是否具有统计显著性。
20. 驱虫蚊帐和疟疾 在肯尼亚进行的一个随机对照实验, 用来验证使用驱虫蚊帐是否能够减少疟疾的发生。343 名婴儿使用驱虫蚊帐, 15 个人得了疟疾。294 名婴儿没有使用驱虫蚊帐, 27 个人得了疟疾。假设驱虫蚊帐没有预防作用, 这个结果偶然发生的概率是 0.015。出现这个结果是否具有统计显著性? 使用驱虫蚊帐预防疟疾是否有效?
21. 人体温度 美国马里兰大学的研究者进行了一项研究, 他们测量了 106 个人的体温, 平均体温是 98.20°F , 公认的人体温度值是 98.60°F 。样本均数和公认常值之间的差异在 0.05 水平上显著。
 - (1) 讨论这个案例中显著性水平的意义。
 - (2) 如果我们假设人体温度的公认值是 98.60°F , 样本平均数是 98.20°F 或者更低些温度的概率是 0.000 000 001。

解释这个概率的意义。

22. 安全带和儿童 一项儿童在车祸中受伤害的研究中发现, 使用安全带的儿童平均要在重症监护室看护 0.83 天, 没有使用安全带的儿童平均要在重症监护室看护 1.39 天。在 0.0001 水平两组间的差异是显著的。解释这个结果。
23. SAT 预备课程 一项针对参加了 SAT 预备课程的 75 个人的研究 (*American Education Research Journal*, Vol. 19, No. 3) 显示他们 SAT 的平均成绩提高了 0.6 个百分点。假定预备课程没有效果的话, 而成绩提高 0.6 个百分点实属偶然的概率是 0.08。讨论预备课程是否显著地提高了 SAT 成绩。
24. 体重的年龄差异 一项全国健康调查, 结果显示 804 位 25 ~ 34 岁的人平均体重为 176 磅, 1 657 位 65 ~ 74 岁的人平均体重为 164 磅。两组体重在 0.01 水平上差异显著。请解释此结果。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 6 章的有用链接 “links for internet projects”。

25. 卫生统计的显著性 浏览一个卫生统计的网页 (比如美国人口普查局或者联邦健康统计中心), 选择诸如下列的一个问题:
 - (1) 各个月的出生人数差异显著吗?
 - (2) 一周的各天中自然死亡率差异显著吗?
 - (3) 被抽选的各州之间婴儿的死亡率差异显著吗?
 - (4) 在被抽选的各州之间某种特殊疾病的发生率差异显著吗?
 - (5) 各州的结婚率差异是否显著?
26. 河流的长度 在一本年鉴或者网络中, 找到世界主要河流的长度。将河流长度的前置数字排成数字表。看看这些特别的数字是否比其他数字更经常出现? 看看差异是否显著并试着解释。

新闻中的统计学

27. 统计显著性 找到一篇近期报纸上登载的关于统计研究的文章, 文章内容中涉及了统计显著性的概念, 写出一页纸的概要, 包括研究摘要、具有统计显著性的结果以及你是否相信这个统计显著性结果的简要讨论。
28. 有效的试验 寻找近期的一则关于统计研究

的新闻故事，内容是采用实验的方法来验证一些新的治疗方法是否有效。依据可利用的信息，简要讨论你能从结果的统计显著性推断出什么。依据给出的显著性，你认为新的治疗方法有效吗？请解释。

29. 个人生活的统计显著性 描述一个你个人生活中发生的事件，这个事件是没有达到你预期的、异常的或者可能是偶然发生的。你能说这个事件具有统计显著性吗？你把这件事的发生归因于什么？



6.2 概率基础知识

在6.1节中，我们看到概率是统计的基础，特别是统计显著性的概念。我们会在第7~10章接着谈统计显著性这个主题。现在，我们需要探讨概率的本质属性和日常生活中概率的应用。

让我们从抛掷两枚硬币的思考开始。图6-1显示硬币的落下有4种可能的情况。我们看到每种可能情况都会是抛掷两枚硬币的不同结果(outcome)。这是抛掷两枚硬币最基本的可能结果。假设我们只关心硬币正面向上的次数，图6-1中间的两种结果都只有一次正面向上，我们说这两种结果代表着相同的事件(event)。一个事件描述了一种或多种具有同样属性的相同可能结果。图6-1表明抛掷两枚硬币会有4种可能结果，但只有3类可能事件：0枚正面向上、1枚正面向上、2枚正面向上。



图6-1 同时抛掷两枚硬币的4种可能结果，中间两个结果是1枚硬币正面向上的同样事件。

定义

- ◆ 结果是观察到的或实验中出现的最基本的可能情况。
- ◆ 事件是具有同一种属性的一种或多种结果的集合。

从数学角度来讲，我们用0~1之间的数字来表示概率。比如，1枚硬币正面向上的概率是1/2，或0.5。如果事件是不可能的，我们就用0来表示它的概率。比如，遇到一个已结婚的单身汉的概率是0，在另外一头，我们用1来表示必然会发生的事。比如，来自本杰明·富

兰克林的一句名言“死亡和交税的概率是1”。我们用 $P_{\text{事件}}$ 代表概率，用下标符号来表示事件的内容。比如，用H来表示扔的硬币正面朝上。这样，1枚硬币正面朝上的概率，我们可以写作 $P_{(H)}=0.5$ 。

概率的表示

一个事件的概率(probability)表示为 $P_{\text{事件}}$ ，常常在0~1之间。概率为0意味着事件不可能发生，概率为1意味着事件必然发生。

图6-2显示事件发生的可能性与概率值的对应关系。它有助于我们理解概率值，比如0.95表明事件非常可能发生，而不是必然发生。100次里面会有95次发生的可能。与之相对，概率值为0.01意味着事件非常不可能发生，100次里面只有1次发生的可能。

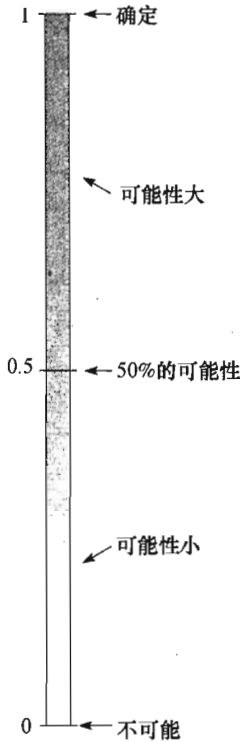


图6-2 不同概率值表达的不确定性水平

6.2.1 理论概率

相关知识

数学家佩尔西·戴康尼斯的一个精细分析显示一副新扑克牌需要经过7次洗牌, 扑克牌所有可能的排列顺序出现的机会才能相等。

当我们说抛掷一枚硬币正面向上的概率是 $1/2$ 时, 我们认为硬币正面向上和反面向上的可能性是一样的。从本质上来说, 我们说 $1/2$ 的概率是来自理论上的。因为也有另外的可能: 硬币滚动。如果有6种结果可能平均出现见图6-3, 那么出现每种结果的理论概率就是 $1/6$ 。

只要所有的结果可能平均出现, 我们就能用下面的公式来计算理论概率。

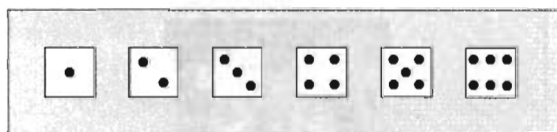
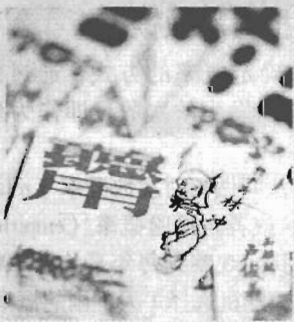


图 6-3 抛掷一枚6面骰子的6种可能结果

相关知识

早在公元前3600年的中东地方, 圆形的骨头就被当做骰子一样用在机遇性的游戏中, 我们熟悉的立方体6面骰子出现在公元前2000年的埃及和中国。扑克是10世纪中国人发明的, 在14世纪传入欧洲。



结果可能性相等条件下理论概率的计算方法

第1步, 计算可能结果的总数。

第2步, 在所有可能的结果中, 计算我们关注的事件A发生的次数。

第3步, 计算概率值。

$$P_{(A)} = \frac{\text{事件A发生的次数}}{\text{可能结果的总次数}} \quad (6-1)$$

例6-4 猜生日

假定我们从一个集会中的大量人群中随机选择一个人, 被选中的人生日在7月份的概率是多少? 假定一年365天。

解答 我们假定出生在每1天的机会是相同的, 我们可以通过三步计算理论概率。

第1步, 每一个可能的生日表示为一个结果, 一共有365个可能结果。

第2步, 7月有31天, 7月生日这个事件在365个可能结果中占有31个。

第3步, 随机挑选一个人是7月生日的概率是:

$$P_{(7\text{月生日})} = 31/365 = 0.0849$$

比 $1/12$ 稍多一点。

相关知识

一年中的生日并不是随机分布的, 在计算了一年不同月份的长度后, 2月的出生率最低, 6月和7月的出生率最高。



结果计数

假如抛掷两枚硬币并想计算结果的总数, 第1枚硬币有两种可能的结果, 正面向上(H)和反面向上(T), 第2枚硬币也有两种可能的结果, 第1枚硬币的两种结果和第2枚硬币的两种结果可以任意结合, 因此两枚硬币抛掷的结果总数就是 $2 \times 2 = 4$ 种, 分别是HH, HT, TH, TT (见图6-4a)。

思考题

解释一枚硬币先后抛掷两次和两枚硬币同时抛掷的结果为什么是一样的?

我们这样扩展我们的思路, 如果我们抛掷3枚硬币, 会得到可能结果是 $2 \times 2 \times 2 = 8$ 种: HHH, HHT, HTH, HTT, THH, THT, TTH, TTT (见图6-4b)。

这种思路是后面计数规则的基础。

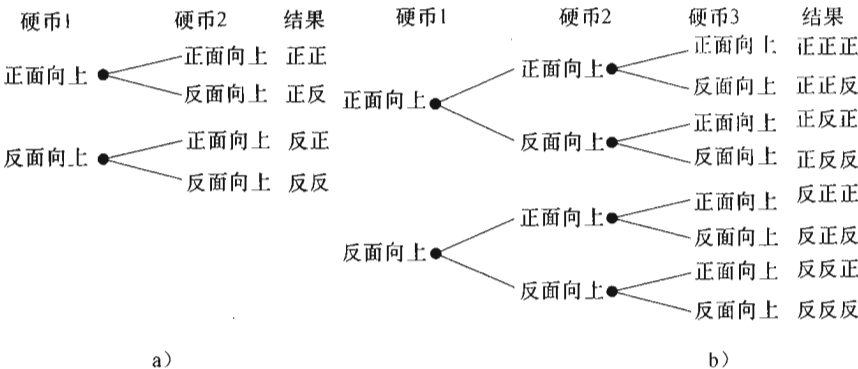


图 6-4 抛掷两枚硬币和 3 枚硬币的结果图示

假设事件 A 有 a 种可能结果，事件 B 有 b 种可能结果，A 事件和 B 事件相互独立，两个事件同时发生的可能结果数目是 $a \times b$ ，这种思路是所有事件计数的规则。例如，如果事件 C 有 c 种可能结果，3 个事件同时发生的可能结果数是 $a \times b \times c$ 。

例 6-5 计算

- (1) 如果你同时抛掷 1 枚骰子和 1 枚硬币，会有多少种可能结果？
- (2) 同时抛掷两枚骰子，结果是两个 1 点的概率是多少？

解答

- (1) 第 1 步，掷 1 枚骰子，有 6 种可能结果 (1, 2, 3, 4, 5, 6)；第 2 步，掷 1 枚硬币，有两种可能结果 (H, T)，接下来，将前两步合在一起，有 $6 \times 2 = 12$ 种可能结果 (1H, 1T, 2H, 2T, ..., 6H, 6T)
- (2) 掷一个单独的骰子，有 6 种可能的结果，两个骰子同时抛掷，就会有 $6 \times 6 = 36$ 种可能结果，在这 36 种结果中，只有一个结果是符合事件要求的 (两个 1 点)，所以抛掷两个 1 点的概率就是：

$$P_{(\text{两个1点})} = \text{两个1点发生的结果数} / \text{总的可能结果数} = 1/36 = 0.0278$$

例 6-6 儿童出生顺序计算

在一个随机选择的有 3 个孩子的家庭，老大是男孩，老二是女孩，老三也是女孩的概率是多少？假设生男生女的机会是相同的。

解答 每一个孩子的出生都有两种可能的结果，男孩或女孩，对于一个三个孩子的家庭来说，可能的出生顺序是 $2 \times 2 \times 2 = 8$ 种 (男男男, 男男女, 男女女, 女男男, 女男女, 女女男, 女女女)。这个问题是问一种特殊的出生顺序 (男孩-女孩-女孩)，是 8 种可能出生顺序的一种，这种出生顺序的概率是 $1/8 = 0.125$ 。

相关知识

男孩和女孩的出生率是不同的，自然状态下，男孩和女孩的出生率比约为 105 比 100，然而，男孩的死亡率要高于女孩，因此成年人数量中女性高于男性。



思考题

如果考虑出生顺序，4 个孩子的家庭有多少种可能性？1 对夫妇有 4 个孩子全是女孩的概率是多少？

6.2.2 相对次数概率

第二种确定概率的方式，通过多次观察然后计算事件 A 发生的次数，即作为事件 A 发生概率的近似值。

这种方法叫做**相对次数概率** (relative frequency method)，或者叫**经验概率** (empirical method)。例如，如果我们观察到每年平均有 100 天会下雨。我们可以说，随机选择 1 天下雨的概率是 $100/365$ 。对这种方法我们可以采用一个常用的法则。

相对次数概率

第 1 步，重复这个观察过程许多次，并且记录我们所关注的事件 A 发生的次数。

第 2 步，通过以下公式估计 $P(A)$

$$P(A) = \text{事件 A 发生的次数} / \text{所有观察的总次数} \quad (6-2)$$

例 6-7 500 年一遇的洪水

地理记录显示有一条河在过去的 2 000 年里曾经发生过 4 次特别大的洪水。那么这条河在下一年发生特别巨大洪水的概率是多少？

解答 基于上述数据，这条河在任一年中发生特别巨大洪水的概率是

$$\frac{\text{发生洪水的年数}}{\text{总共的年数}} = \frac{4}{2\,000} = \frac{1}{500}$$

因为这样巨大的洪水平均每 500 年发生 1 次，所以它叫做“五百年一遇的洪水”。则任一年中发生这样巨大洪水的概率是 $1/500$ ，或者说是 0.002。 ■

6.2.3 主观概率

确定概率的第三种方法是应用经验或直觉估计主观概率 (subjective probability)。例如，一个朋友在下一年要结婚的概率，或者能够在统计学课程上取得良好成绩的概率。

三种计算概率的方法

(1) 理论概率是基于所有可能发生的结果是相等的假设，用一个事件发生的次数除以所有可能结果发生的总次数，所计算的结果就是理论概率。

(2) 相对次数概率是基于观察和实验，通过所关注事件的相对次数来确定概率。

(3) 主观概率 (subjective probability) 就是基于经验和直觉对概率的一个估计值。

相关知识

理论概率也叫做先验概率，词语“priori”是拉丁语，意思是指“事实之前”或者“经验之前”。

例 6-8 采用哪一种方法来计算概率

面对下述情况确定计算概率的方法是什么。

(1) 你在下一年结婚的概率是 0。

(2) 基于政府报告的数据，交通事故中死亡率是 $1/7\,000$ (每年)。

(3) 抛掷一枚 12 面的骰子，结果是 7 的概率。

解答

(1) 这是一个主观概率，因为它基于一种最近时期的感觉。

(2) 这是一个相对次数概率，因为它基于过去的交通事故的观察数据。

(3) 这是一个理论概率，因为它基于一种假设，这个 12 面的骰子结果为每 1 面的概率是完全相同的。 ■

相关知识

计算概率的另一种方法，叫做蒙特卡洛方法，使用计算机模拟，这种技术的本质是计算相对次数概率，在这样的情况下，观察由计算机完成。

例 6-9 飓风的概率

图 6-5 呈现了弗洛伊德飓风接近美国东南海岸情况的地图。地图显示在飓风经过的三个不同的地区“飓风侵袭概率”，解释术语“飓风侵袭概率”并且讨论这些概率值以及它们是怎么被估计出来的。

解答

飓风侵袭概率告诉我们飓风眼穿越某一特定地区的可能性。例如，对于一个人来说，如果飓风侵袭概率为 20% ~ 50%，也就是他有 0.2 ~ 0.5 的概率被飓风击倒。了解这些概率可以帮助人们做出预防或者是撤离的决策。概率是通过计算机模型计算出来的，对飓风的预测也是基于飓风常发地区的天气类型、有关飓风的大气物理以及过去发生的飓风数据做出的。 ■

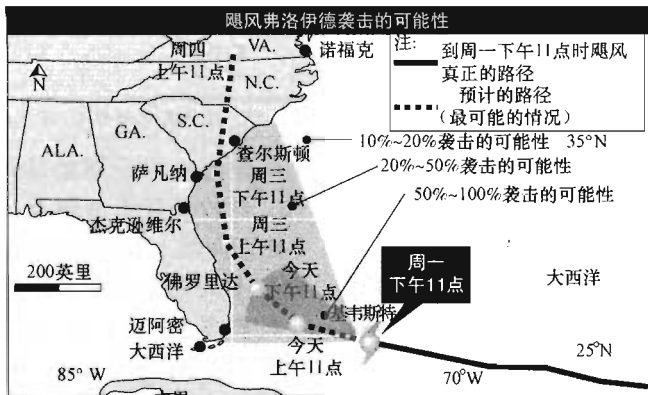


图 6-5 弗洛伊德飓风接近美国东南海岸的实际路径 (直到地图绘制时间周一上午 11:00)，它也呈现了可能路径的概率，最可能的路径以虚线表示。

资料来源: Virginia Pilot.

6.2.4 未发生事件的概率

假如我们关注某一特定事件或者没有发生的事件结果的概率。例如，考虑有 5 个被选答案的选择题回答错误的可能性。如果随机猜测回答正确的概率是 1/5，那么回答错误的概率就是 4/5。注意，回答正确与回答错误的概率之和一定是 1。这样我们可以推导出以下观点：

一个未发生事件的概率

如果一个事件 A 概率是 $P(A)$ ，那么 A 事件没有发生的概率是 $P(\text{非} A)$ 。因为以 一个事件要么发生要么不发生，我们可以这样来表述：

$$P(A) + P(\text{非} A) = 1 \text{ 或者 } P(\text{非} A) = 1 - P(A)$$
(6-3)

注意：非 A 事件叫做事件 A 的互补事件；“非”经常用字母上面一杠来表示，因此 $\bar{A}$ 意思是非 A 。

相关知识

抽签指的是对一个人的安排或定位，使用一个物体或标记来做随机选择，抽签在特洛伊战争被用于决定谁先开始投枪，而在《旧约》中用于分割征服的土地。

例 6-10 对于特价商品光电扫描仪的准确性是一样的吗

在一项关于光电扫描结账系统的研究中，购买的样品经常要和过去价格进行扫描对比。表 6-1 概括了 819 项目样本的结果。基于这个数据，常规价格项目出现扫描错误的概率是多少？海报促销特价商品出现扫描错误的概率是多少？

表 6-1 扫描的准确性

	常规价格商品	广告特价商品
扫描价格低于标价	20	7
扫描价格高于标价	15	29
扫描价格正确	384	364

解答 我们用字母 R 来代表常规价格商品扫描正确的。因为 419 项常规价格商品中的 384 项被正确扫描。

$$P(R) = 384/419 = 0.916$$

扫描出现错误的事件与扫描价格正确的事件是互补事件，所以常规价格出现扫描错误

(索价低于或者高于) 的概率是：

$$P(\text{非} R) = 1 - 0.916 = 0.084$$

现在我们以字母 A 来表示广告促销特价商品扫描正确的。样本中一共有 400 件特价商品，其中 364 件扫描正确。这样：

$$P(A) = 364/400 = 0.910$$

一件促销特价商品被扫描错误的概率是：

$$P(\text{非} A) = 1 - 0.910 = 0.090$$

两种商品扫描错误率是基本相同的。然而，注意到表 6-1 中大多数海报促销特价商品所出现的错误是不利于消费者。而常规价格商品出现的错误半数以上是有利于消费者。

6.2.5 概率分布

从第 3~5 章，我们涉及了次数分布和相对次数分布，例如，年龄或者收入的分布。大多数概率学和统计学上的基本理念之一就是概率分布。顾名思义，一个概率分布是指关注变量与概率有关的分布。

假如你同时抛掷两枚硬币。结果是两枚硬币各自正面向上或反面向上组合的变化。因为每一枚硬币都有两种可能的结果（正面向上或者反面向上），两枚硬币具有 $2 \times 2 = 4$ 种不同的结果，表 6-2 的第 3 列显示了 4 种结果中的每一种。

表 6-2 抛掷两枚均匀硬币的结果

硬币 1	硬币 2	结果	概率
H	H	HH	0.25
H	T	HT	0.25
T	H	TH	0.25
T	T	TT	0.25

注：H 代表正面向上，T 代表反面向上。

注意 4 种不同的结果表达了 3 个不同事件：两枚都是正面向上，两枚都是反面向上，1 枚正面向上和 1 枚反面向上，这样，两枚正面向上的概率是 $P = 1/4 = 0.25$ ；两枚反面向上的概率 $P = 1/4 = 0.25$ ，1 枚正面向上和 1 枚反面向上的概率 $P = 2/4 = 0.50$ 。这个概率结果就如表 6-3 中或者如图 6-6 呈现的一个概率分布。注意所有概率之和一定是 1（因为这些可能的结果一定会发生其中的一个）。

相关知识

另一个表达可能性的通常方法是使用胜率，反对一个事件的胜率代表这个事件不发生的概率与发生的概率的比值。例如，反对抛掷一枚6面骰子为6的胜率为 $(5/6)/(1/6) = 5:1$ 。赌博中使用的胜率也叫赔率，表达了赌博中的净收益。例如，假定赛马比赛中一匹马的赔率是3:1。这意味着你每用1美元打赌这匹马胜，如果这马真的赢的话，你将赢得3美元（而且也拿回你自己投入的1美元）。

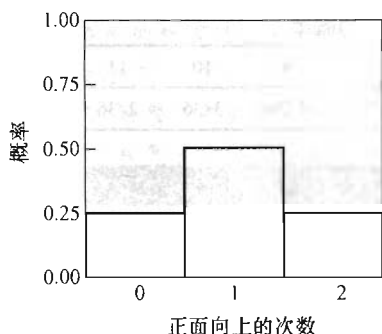
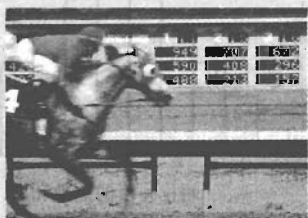


图 6-6 抛掷两枚硬币结果的概率直方图

表 6-3 抛掷两枚硬币

结果	概率
2 枚正面向上, 0 枚反面向上	0.25
1 枚正面向上, 1 枚反面向上	0.50
0 枚正面向上, 2 枚反面向上	0.25
总计	1

整理一个概率分布

一个概率分布代表了所有可能事件的概率。按照下面的步骤可以展示一个概率分布。

第1步，列出所有可能的结果。使用图和表会很有用的。

第2步，归纳表示同类事件的结果，计算每一种事件的概率。

第3步，绘制一个表，一栏列出每一种事件，另一栏列出它的概率。概率的总和一定是1。

例 6-11 抛掷3枚硬币

当3枚硬币同时被抛掷的时候，记录其正面向上的次数，整理它的概率分布。

解答 我们应用上面的3步过程。

第1步，当3枚硬币同时被抛掷的时候，出现的不同结果的个数是 $2 \times 2 \times 2 = 8$ 。图 6-4b 显示了我们怎样得到的8种可能的结果。它们分别是：

HHH, HHT, HTH, HTT, THH, THT, TTH, TTT。

第2步，这里有4种可能事件：0次正面向上，1次正面向上，2次正面向上，3次正面向上。再看一看8种可能的结果，我们发现以下：

- 8个结果中只有1种表达了0次正面向上的事件，所以它的概率是 $1/8$ 。
- 8个结果中有3种表达了1次正面向上的事件，所以它的概率是 $3/8$ 。
- 8个结果中有3种表达了2次正面向上的事件，所以它的概率是 $3/8$ 。
- 8个结果中只有1种表达了3次正面向上的事件，所以它的概率是 $1/8$ 。

第3步，我们制作一个表格，右边一栏表示4种事件，左边一栏表示它们的概率。表 6-4 显示了这个结果。

表 6-4 抛掷3枚硬币

结果	概率
3 枚正面向上, 0 枚反面向上	0.125
2 枚正面向上, 1 枚反面向上	0.375
1 枚正面向上, 2 枚反面向上	0.375
0 枚正面向上, 3 枚反面向上	0.125
总计	1

思考题

当你抛掷4枚硬币的时候，有多少种不同的可能结果？如果你记录正面朝上的次数，有多少种可能情况？

例 6-12 两枚骰子的分布

记录抛掷两枚骰子的点数做成概率分布，将这个分布编制成一个表和一个直方图。

解答 因为每一枚骰子抛掷以后，会出现6种不同的点数。这样抛掷两枚骰子就会出现 $6 \times 6 = 36$ 种结果。当我们将这36种结果呈现在表 6-5 中，一枚骰子的结果呈现在行中，另一枚骰子的结果呈现在列中。在中间的每一格中，显示的是两枚骰子的点数和。

表 6-5 抛掷两枚硬币的结果和点数

	1	2	3	4	5	6
1	1 + 1 = 2	1 + 2 = 3	1 + 3 = 4	1 + 4 = 5	1 + 5 = 6	1 + 6 = 7
2	2 + 1 = 3	2 + 2 = 4	2 + 3 = 5	2 + 4 = 6	2 + 5 = 7	2 + 6 = 8
3	3 + 1 = 4	3 + 2 = 5	3 + 3 = 6	3 + 4 = 7	3 + 5 = 8	3 + 6 = 9
4	4 + 1 = 5	4 + 2 = 6	4 + 3 = 7	4 + 4 = 8	4 + 5 = 9	4 + 6 = 10
5	5 + 1 = 6	5 + 2 = 7	5 + 3 = 8	5 + 4 = 9	5 + 5 = 10	5 + 6 = 11
6	6 + 1 = 7	6 + 2 = 8	6 + 3 = 9	6 + 4 = 10	6 + 5 = 11	6 + 6 = 12

点数和的可能事件从 2 ~ 12。这是问题中所关注的事件。我们计算每一个事件的概率，计算方法是每个点数和所包括的所有可能结果数除以全部的结果数 36。例如，在表中点数和为 8 的时候一共有 5 种可能的结果，所以点数和为 8 的概率是 5/36。表 6-6 显示了全部的概率分布，图 6-7 显示了这个分布的直方图。

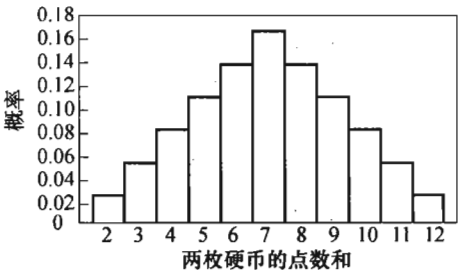


图 6-7 抛掷两枚硬币所得点数和的概率分布图

表 6-6 抛掷两枚硬币所得点数的概率

事件（点数）	2	3	4	5	6	7	8	9	10	11	12	总和
概率	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

6.2 练习题

统计素养和批判性思维

- 1. 符号 $P(A)$ 表示什么意思？符号 $P(\text{非 } A)$ 或者也写作 $P(\bar{A})$ 表示什么意思。
- 2. 概率的解释 当我们说“赢得伊利诺伊州福利彩票大奖的概率是 1/20 358 520”这句话的时候，意味着什么？获得这个大奖是不同寻常的吗？为什么？
- 3. 下雨的概率 当我们写出明年 7 月 4 日波士顿地区下雨的概率时，一家报纸的记者说下雨的概率是 1/2，因为要么下雨要么不下雨，他的推理正确吗？为什么？
- 4. 主观概率 运用主观判断去估计下面这个事件的概率，下一次你乘坐电梯时电梯卡在楼层之间的概率是多少？

判断正误：判断练习题 5 ~ 8 中的陈述是有意义的（表述正确）还是无意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

- 5. 不可能事实 因为抛掷两枚普通的骰子，出现总点数是 14 的情况是不可能的，所以两枚

- 骰子总点数为 14 的概率为 0。
- 6. 汽车的碰撞事故 一家保险公司说明某一种汽车一年内发生汽车碰撞的概率是 0.14，那么这个汽车一年内不会发生碰撞事故的概率是 0.82。
- 7. 相反事件 如果 $P(A) = 0.25$ ，那么 $P(\text{非 } A) = 0.75$ 。
- 8. 确定事件 如果 $P(A) = 1$ ，那么事件 A 一定发生。
- 9. 闪电击中 杰克估计他下一年中被闪电击中的主观概率是 1/2。
- 10. 闪电击中 琼斯估计他下一年中被闪电击中的主观概率是 1:1 000 000。

概念与应用

- 11. 结果和事件 对于下面所描述的观察结果，说明是否有一种办法或者多种办法告诉我们观察的结果会发生什么（每一种情况都是一个事件）。
 - (1) 抛掷一枚硬币正面向上。
 - (2) 抛掷 3 枚均匀的硬币两枚正面向上。

- (3) 抛掷一枚公平的骰子点数是6。
- (4) 抛掷一枚公平的骰子点数是偶数。
- (5) 抛掷两枚公平的骰子点数和是7。
- (6) 从一副普通的扑克牌中抽出一对王。

12. 结果和事件 对于下面所描述的观察结果,说明是否有一种办法或者多种办法告诉我们观察的结果会发生什么(每一种情况都是一个事件)。

- (1) 抛掷一枚均匀的硬币结果是反面向上。
- (2) 抛掷两枚均匀的硬币一枚正面向上,一枚反面向上。
- (3) 抛掷两枚公平的骰子至少一枚骰子的点数是5。
- (4) 抛掷一枚骰子出现的点数是奇数。
- (5) 从一副普通扑克牌中抽出一张J。
- (6) 从一副普通的扑克牌中抽出3张A。

理论概率:对于练习题13~20使用理论概率的方法计算给出的结果或事件的概率并且陈述你所需要的假设。

- 13. 骰子 抛掷一枚骰子得到的点数小于3。
 - 14. 骰子 抛掷一枚骰子点数大于6。
 - 15. 轮盘赌 当你旋转一个轮盘赌时得到结果是0或者00(一个轮盘赌在不同扇区表上是0, 00, 1, 2, 3, ..., 36)。
 - 16. 美国总统的生日 计算下一任美国总统出生在星期六的概率。
 - 17. 生日 计算你遇到的下一个和你同一天生日的概率(忽略闰年)。
 - 18. 出生性别 计算在阿拉斯加州下一个出生的是女孩的概率。
 - 19. 出生性别 假如一对夫妇他们有两个孩子并且都是男孩,计算下一个孩子是女孩的概率。
 - 20. 测验 在一个选择题中进行随机猜测恰巧是正确答案的概率,这个选择题的备选答案项是5择1。
- 相反事件:**在练习题21~28中,确定所给出的反向事件的概率,并表明你所采用的假设。
- 21. 骰子 抛掷一枚公正的骰子,得到的结果不小于3的概率。
 - 22. 骰子 抛掷一枚公正的骰子,得到的结果不大于6的概率。
 - 23. 轮盘赌 当你旋转一个轮盘赌时得到结果不

是0或者00?(一个轮盘赌在不同扇区表上是0, 00, 1, 2, 3, ..., 36。)

- 24. 美国总统的生日 计算下一任美国总统出生在星期六的概率。
- 25. 篮球 一个罚球命中率是55%的球员在下次罚球中罚失的概率。
- 26. 测验 在一个选择题中进行随机的猜测恰巧是错误答案的概率,这个选择题备选项是5择1。
- 27. 棒球 一个棒球运动员安打命中率是0.280,那么在他下一次击球中打不出安打的概率。
- 28. 有缺陷的保险丝 当从一个保险丝生产线上随机选择一个保险丝是残次品的概率是2%,那么从生产线上挑选另外一个保险丝是正品的概率是多少?

理论概率:对于练习题29~32,使用理论概率的方法计算给出的结果或事件的概率,并且陈述你所需要的假设。

- 29. M&Ms公司的糖果包里包含10颗红色的,15颗蓝色的,20颗黄色的,从中拿出一个是红色的概率是多少?拿出一个是蓝色的概率是多大?拿出一个是黄色的概率是多大?不是黄色的概率是多少?
- 30. 测验问题 新英格兰医学院使用了一套选择题的入门测验,每一道题有5个选项,只有1个是正确的,如果你对每个题都随机猜测,预期你最可能得到的分数是(百分制)。
- 31. 有三个孩子的家庭 假如你随机挑选一个有三个孩子的家庭,假设男孩和女孩的出生机会是相等的,计算下面各种情况的概率。
 - (1) 三个女孩。
 - (2) 两个男孩和一个女孩。
 - (3) 出生顺序是一个女孩、一个男孩然后还是一个男孩。
 - (4) 至少一个女孩。
 - (5) 至少两个男孩。
- 32. 有四个孩子的家庭 假如你随机挑选一个有四个孩子的家庭,假设男孩和女孩的出生机会是相等的,计算下面各种情况的概率。
 - (1) 有多少种可能的出生顺序,将它们罗列出来。

- (2) 这个家庭有四个男孩的概率是多少，这个家庭有四个女孩的概率是多少？
- (3) 这个家庭的出生顺序是一个男孩，一个女孩，一个男孩再一个女孩的概率是多少？
- (4) 这个家庭有两个男孩两个女孩的概率是多少？

相对次数概率：使用相对次数的方法，估计练习题 33~36 的概率。

33. 天气预报 记录你当地的天气预报持续 30 天，你发现正确预报的次数是 12 次，下一次天气预报正确的概率是多少？
34. 洪水 百年一遇的洪水在今年发生的概率是多少？
35. 篮球 在赛季过了一半的时候，一个篮球运动员的罚球命中率是 86%，他下一次罚球成功的概率是多少？
36. 手术 在一个腕管综合征患者进行手术的临床试验中，73 个实施手术的患者中 67 个治愈了，在下次手术中成功的概率是多少？
37. 老年人的概率 2000 年，在美国全部 2.81 亿人口中超过 65 岁的老年人是 3400 万。到 2050 年，预计所有美国人口将达到 3.94 亿，其中 65 岁以上的人口会达到 7890 万。在 2000 年或 2050 年，你随机遇到一位老年人的概率多大？请说明理由。
38. 结婚的年龄 下表给出了男性和女性第一次结婚年龄分布在不同年龄组里的百分数（美国普查局）。

	20 岁以下	20 ~ 24 岁	25 ~ 29 岁	30 ~ 34 岁	35 ~ 44 岁	45 ~ 64 岁	65 岁以上
女性	16.6	40.8	27.2	10.1	4.5	0.7	0.1
男性	6.6	36.0	34.3	14.8	7.1	1.1	0.1

- (1) 随机选择一个女性，年龄在 35~44 岁之间，这个女性在这个年龄第一次结婚的概率是多少？
- (2) 随机选择一个结婚的男士，他在 20 岁之前第一次结婚的概率是多大？
- (3) 绘制一个包括男性和女性第一次结婚年龄的分段条图。
39. 抛掷 4 枚硬币出现的点数之和的概率分布。
- (1) 编制一个与表 6-2 相类似的表，呈现抛掷 4 枚硬币所能呈现的所有结果。

- (2) 编制一个与表 6-4 相类似的表格，呈现以下结果，包括 4 次正面向上、3 次正面向上、2 次正面向上、1 次正面向上和 0 次正面向上的概率分布。
- (3) 同时抛掷 4 枚硬币，两次正面向上和两次反面向上的概率是多少？
- (4) 同时抛掷 4 枚硬币，非 4 枚硬币都是正面向上的概率。
- (5) 哪一种事件最可能发生？

40. 科罗拉多州的福利彩票分布 图 6-8 呈现了 5964 次科罗拉多州福利彩票数字出现次数的直方图（数字的可能取值是 1~42）。

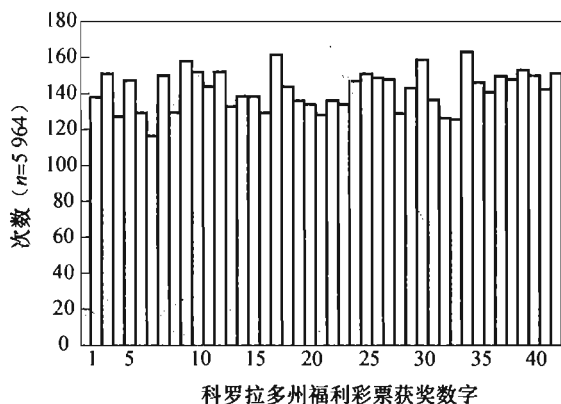


图 6-8 科罗拉多州的福利彩票分布

- (1) 假如福利彩票的数字出现是随机的，你预期任何一个数字出现的概率是多少？
- (2) 基于上述的直方图，最有可能出现的数字的相对次数概率是多少？
- (3) 基于上述直方图，出现次数最少的数字的相对次数概率是多少？
- (4) 评价经验概率和预期的理论概率的差异，这种差异是显著的吗？

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 6 章的有用链接“links for internet projects”。

41. 血型 四种主要血型分别是 A、B、AB 和 O，每一种血型又分别包括两种 RH 血型，即阳性和阴性。使用图书馆的资源或网络资源，寻找血型的相对次数数据，包括 RH 血型。编制一个表格，呈现所遇到的某个人是八种血型之一的概率。
42. 年龄和性别 随着人口年龄的变化，男性和

女性的比例也在发生变化。使用互联网上的近期数据,编制一个表格,呈现遇到一个男性或女性的概率,并且分别按如下年龄进行分组:0~5岁,6~10岁,11~20岁,21~30岁,31~40岁,41~50岁,51~60岁,61~70岁,71~80岁,80岁以上。

43. 图钉的概率 寻找一个标准图钉,并且试着抛掷它落到一个光滑的平面上,注意会有两种不同结果:图钉的尖儿朝上,或图钉的尖儿朝下。

- (1) 抛掷这个图钉50次,并且记录结果。
- (2) 计算两种结果的相对次数概率。
- (3) 如果可能的话,请其他一些人重复这个过程,和你所计算的概率一样的可能性是多大?

44. 三枚硬币实验 同时抛掷3枚硬币50次,记录正面向上的结果,基于你的观察,计算结果的相对次数概率与理论概率是否吻合,解释并讨论你的实验结果。

45. 随机化调查 假如你想进行一项调查,其中包含一个敏感问题,不是所有的参与者都愿意选择诚实回答的(例如,包括的一个问题是偷税或者吸毒)。这里有一个办法,保证调查时能够识别他的回答。我们首先假定这个敏感性问题的回答为“是”或“否”。首先,要求所有被试抛掷一枚均匀的硬币,然后按照下面的指导进行回答:

- 如果你抛掷一枚硬币结果是正面向上,那么回答一个设定的诱饵问题(“是”和“否”):你是出生在一个月的奇数日子吗?

- 如果你抛掷一枚硬币结果是反面向上,那么回答一个真正要调查的问题(“是”和“否”)。

在所有被试回答完这些问题后(“是”和“否”):计算回答(“是”和“否”)的总数。

- (1) 选择一个总体上不可能全部诚实回答的问题,然后使用这个技术在你的班级中做一次调查。
- (2) 给出所有问题回答(“是”和“否”)的总数,解释你是如何估计回答诚实问题时“是”和“否”的数量。
- (3) 在第2小题中估计的结果是否准确?请加以解释。
- (4) 假定这个诱饵问题被其他指导语代替:如果你抛掷一枚硬币是正面向上,然后回答“是”,你是否仍然能够确定回答诚实问题(“是”和“否”)的数量。

新闻中的统计学

46. 理论概率 寻找一篇新闻报道中引用理论概率的文章,给出一段话的评论。
47. 经验概率 寻找一篇新闻报道中引用相对次数概率(或者叫做经验概率)的文章,给出一段话的评论。
48. 主观概率 寻找一篇新闻报道中参考了主观概率的文章,给出一段话的评论。
49. 概率分布 寻找一篇新闻报道中引用概率分布的文章,或者使用了概率分布的研究报告,给出一段话的评论。



6.3 概率的大数定律

当我们对于一个随机事件进行观察和测量时,预测一个单独的结果是不可能的,我们只能对一个单独的结果发生的概率进行描述。然而,当我们对随机现象进行非常多次的观察和测量时,我们期望这些结果的分布会呈现某种模式或者某种规律,在这节中,我们应用概率的观念对大样本做一些有用的描述,在这节的课程中,你将会发现在概率和统计学之间一种最重要的关联。

6.3.1 大数定律

如果你将一枚硬币抛掷出去以后,你不可能预测它落下来到底是什么样子,你只能说这枚硬币正面向上的概率是0.5,如果你抛掷这枚硬币100次,你仍然不能准确预测有多少次正面向上。但是你可以合理地预期正面向上的次数会接近50%,如果你抛掷这枚硬币1000次,同样你可以预期正面向上的比率更接近50%了。一般情况下,你抛掷硬币的次数越多正面朝上的比率就会越接近50%。即使个体事件仍然是不可预测的,大多数事件会呈现某种模式,这样的现象叫做大数定律(law of large numbers),

或者叫**趋中定律** (law of averages)。

大数定律

大数定律 (law of large numbers) 是指当一个事件 A 的概率是 $P(A)$ 时, 并且后面所发生的事件并不依赖于前面发生事件的结果 (它们是相互独立的)。大数定律表述为如果一个现象被多次重复, 事件 A 发生的比率会越来越接近概率 $P(A)$ 这种现象重复得越多, 事件 A 发生的比率越接近概率 $P(A)$ 。

我们可以通过骰子抛掷实验来说明大数定律, 抛掷一枚骰子出现点数为 1 的概率为 $P(1) = 1/6 = 0.167$ 。为了避免单调地抛掷骰子这样多次, 我们可以使用计算机模拟随机抛掷许多枚骰子。图 6-9 呈现了计算机模拟抛掷一枚骰子 5000 次的结果。横坐标显示了抛掷的次数, 曲线的高度显示了点数为 1 的比率, 虽然当抛掷的次数很少时, 曲线的浮动较大, 但是当抛掷的次数越来越多时, 点数为 1 的比率越来越接近概率 0.167, 正如大数定律所预测的那样。

例 6-13 轮盘赌

一个轮盘有 38 个数字, 18 个黑色数字 18 个红色数字, 数字 0 和数字 00 是绿色的。(假定所有的结果, 也就是这 38 个数字每个数字出现的概率是相同。)

- (1) 任何一次旋转出现红色数字的概率是多少?
- (2) 如果一个轮盘赌场的客人旋转这个轮盘 100 000 次, 你预期转出红色数字的次数是多少?

解答

- (1) 轮盘的每一次旋转得到一个红色数字的理论概率是

$$P(A) = \frac{\text{红色数字出现的次数}}{\text{所有结果出现的次数}} = \frac{18}{38} = 0.474$$

- (2) 大数定律告诉我们当这种游戏进行非常多次后, 出现红色数字的次数比率也越来越接近 0.474, 在 100 000 的旋转中, 轮盘停留在红色数字的概率也越来越接近 47.4% 或者说 47 400 次。

相关知识

卡特里娜飓风 (2005 年) 是美国历史上损失最为惨重的自然灾害, 接近 2 000 人死伤, 估计损失至少为 800 亿美元。



6.3.2 期望值

假定全民保险公司要销售一种特殊类型的保险, 这个保险条款规定, 如果因为严重的疾病必须离职回家的话你将获得 10 万美元的赔偿。基于过去的一些数据, 保单持有人因为失去工作而获得赔付的概率是 $1/500$, 如果保险公司销售这种保险, 每个人的价格是 250 美元, 他们能预期获利吗?

如果全民保险公司只销售出这种保险很少, 获益或亏损是不可预测的, 例如, 每份保险为 250 美元, 销售出 100 份保单获得收入是 $100 \times \$250 = \$25\,000$ 。如果这 100 名保单持有人没有一人提出赔付要求保险公司将获得一批可观的收益。另一方面, 如果任意一名保单持有者提出赔付要求, 而全民保险公司必须支付 100 000 美元的话, 那么它将面临一个巨大的亏损。

相反, 如果全民保险公司销售了大量的保险给人们, 大数定律会告诉我们, 提出赔付要求并且获得保险公司赔偿的保单持有者的比率会越来越接近于单个保单持有者 $1/500$ 的这个概率。例如, 如果保险公司销售了 100 万份保单, 可以预期接受赔付 10 万美元的保单持有者会接近于:

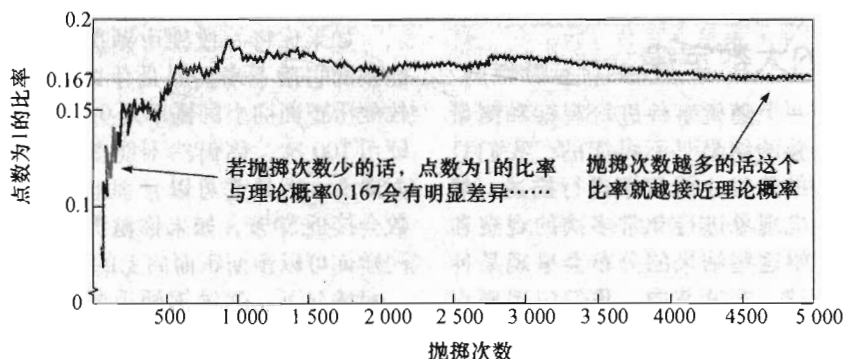


图 6-9 计算机模拟抛掷一枚骰子实验的结果。随着抛掷次数的增多, 点数为 1 的比率越来越接近于理论概率 0.167

$1\,000\,000$ (保单持有者的数字) $\times (1/500)$
(获得 100 000 美元的概率) $= 2\,000$

这 2 000 位保单持有者将获得的赔付要花费
 $2\,000 \times \$100\,000 = \$200\,000\,000$ 。

对于这 100 万位保单持有人来说, 每个人大约平均花费 200 美元, 这也意味着, 卖给每个人的保险是 250 美元, 而公司预期可以获得的收益是每人平均 $250 - 200 = 50$ 美元。我们称这个平均数是每份保单的期望值 (expected value)。注意, 这个期望值只有在公司销售了大量的保险时才成立。

一个变量的期望值 (expected value) 是指它的所有可能事件的加权平均数, 因为它是一个平均数, 所以只有发生大量事件时我们才能计算这个期望值。也就是说, 当大数定律发生作用时, 这个期望值才有意义。

我们可以通过公式来计算同样的期望值, 保险公司的例子包括两个有明显区别的事件, 对于这个公司来说每一个事件都有特定的概率和数值:

(1) 在这个事件中, 一个人购买一份保险, 付给保险公司 250 美元保险费, 这个事件发生的概率是 1, 因为每个人买一份保险就要付 250 美元。

(2) 在这个事件中, 一个人提出要求申请赔付, 对于保险公司而言它的价值是 -100 000 美元, 在这种情况下, 公司要损失 100 000 美元, 所以是负数。这种概率是 $1/500$ 。

现在我们将每个事件的价值和概率相乘并且将这些乘积加起来, 就可以计算出每份保险单的期望值。

期望值 $= 250 \text{ 美元} \times 1 + (-100\,000 \text{ 美元}) \times \frac{1}{500}$

保险	每销售一份	赔付	赔付
销售价格	保险收益 250	的价格	的概率
	美元的概率		

每份保险单的期望收益是 50 美元, 这也是先前问题的回答, 注意如果卖出 100 万份保险, 那么这个收益值就是 5 000 万美元。

方法注解

下面是可供选择的计算期望值方法。在保险的例子中, 我们可以用事件 1 代表没有要求赔偿的保单, 事件 2 代表要求赔偿 10 万美元的保单。事件 1 的值是付给保险公司 250 美元 (保单的价格), 概率为 $499/500$ 。事件 2 的值是 250 美元减去保险公司支付的 10 万美元就是 -99 750 美元, 概率为 $1/500$ 。注意这个备选方法仍然可以得出同样的期望值 50 美元: $250 \text{ 美元} \times (499/500) + (-99\,750 \text{ 美元}) \times (1/500) = 50 \text{ 美元}$ 。一些统计学家更愿意选择这个公式, 因为它的概率和为 1。

计算期望值

考虑到两个事件, 每一个事件都有自己的值和可能性, 那么期望值就是:

$$\text{期望值} = (\text{事件 1 的值}) \times (\text{事件 1 的概率}) + (\text{事件 2 的值}) \times (\text{事件 2 的概率}) \quad (6-4)$$

这个公式可以扩展为任意多的事件的期望值总和。

思考题

保险公司可以期望在每个人的保单中获得收益 50 美元吗? 可以预期 1 000 张保单获得收益 50 000 美元吗? 解释理由。

例 6-14 彩票的期望值

假定一张 1 美元的彩票具有如下的概率: $1/5$ 的概率再赢得一张免费彩票 (价值 1 美元), $1/100$ 的概率赢得 5 美元, $1/100\,000$ 概率赢得 1 000 美元, $1/10\,000\,000$ 的概率赢得 1 000 000 美元。彩票的期望值是多少? 讨论它的意义。(注意: 购买彩票者不会得回他花在购买彩票上的 1 美元。)

解答 这类问题的方法是制作一张如下的表, 将所有相关事件的值和概率列出来的。这样我们就可以计算彩票的期望值, 注意彩票的购买价格是一个负值因为它花费了你的钱。赢得的奖励是正数。

事件	价格 (美元)	概率	价值 $\times$ 概率
购买彩票	-1	1	$(-1) \times 1 = -1.00$
中奖为免费送张彩票	1	$1/5$	$1 \times 1/5 = 0.20$
中奖 5 美元	5	$1/100$	$5 \times 1/100 = 0.05$
中奖 1 000 美元	1 000	$1/1\,000\,000$	$1\,000 \times 1/1\,000\,000 = 0.01$
中奖 1 000 000 美元	1 000 000	$1/10\,000\,000$	$1\,000\,000 \times 1/10\,000\,000 = 0.10$
			最后一列总和: -0.64

期望值是所有取值和概率乘积的总和，表格中最后一栏的底部显示期望值是 -0.64 美元。这样，如果购买了许多彩票以后平均算来，你可以预期购买一张彩票会失去 0.64 美元。如果你购买 1 000 张彩票，预期你会失去 $1\,000 \times 0.64 = 640$ 美元。

思考题

许多州利用彩票作为公园、休闲娱乐和教育的支出来源。彩票也可以让这些州的税收保持在较低水平。但是在另一方面，研究发现购买彩票的人大多都是低收入阶层。你认为彩票是一项好的社会政策吗？你认为彩票是一项好的经济政策吗？

6.3.3 赌徒的谬论

思考一项包含抛掷硬币的一个游戏样本：如果你抛掷硬币的结果是正面向上时，你将赢得 1 美元，如果是反面朝上时，你将赔出 1 美元。假如你抛掷硬 100 次，最后结果是 45 次正面向上 55 次反面朝上，结果是你输掉了 10 美元，你连胜的好运气快来了吗？

你可能认识到这个答案是否定的：你过去的坏运气不会影响你未来的机会。然而，许多赌徒，特别是狂热的赌徒，会得出相反的答案。他们相信他们的坏运气已经过去了，好运气时机到了。这个错误的信念经常被称做赌徒的谬论 (gambler's fallacy)，或者称做赌徒的破产 (gambler's ruin)。

定义
◆ 赌徒的谬论是指连续的坏运气会给一个人带来连续的好运气这样的一个错误信念。

人们深信赌徒的谬论的一个原因就是対大数定律的错误理解。在抛掷硬币的游戏中，大数定律告诉我们，随着抛掷次数的增多，正面向上的比率也越来越接近于 50%。但这并不意味着可以弥补先前的抛掷结果，我们来看看为什么，研究表 6-7 中呈现了计算机模拟抛掷大量硬币的结果。注意当抛掷次数增加时，正如大数定律所预测的那样，正面向上的百分比确实越来越接近 50%。然而，最后一栏显示，正面向上的次数和反面向上的次数的差别，这意味着在某一个特定的模拟实验中，即使正面向上的比率接近 50% 时，硬币抛掷的结果（正面向上的次数和反面向上的次数的差别）也同样变得差别巨大。

表 6-7 硬币抛掷试验的结果

硬币抛掷次数	正面向上次数	正面向上的百分数 (%)	正面向上和反面向上的次数差
100	45	45	10
1 000	470	47	60
10 000	4 950	49.5	100
100 000	49 900	49.9	200

图 6-10 显示了计算机模拟一枚硬币抛掷 1 000 次的图示结果。虽然在 1 000 次抛掷中正面向上的比率接近于 50%，我们也看到实际结果正面向上和反面向上的平均值有较大的偏离。

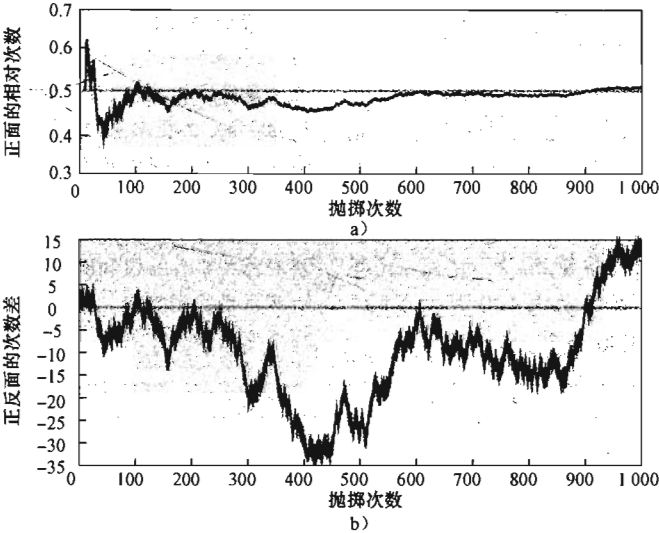


图 6-10 计算机模拟一枚硬币抛掷 1 000 次的结果

- a) 呈现正面向上次数随着抛掷次数增加的变化（注意比率接近 0.5，与我们依据大数定律预计的一致）
- b) 呈现正面向上和反面向上次数差随着抛掷次数增加的变化（尽管正面向上的次数比率接近 0.5，但是次数差越来越大）

例 6-15 持续的硬币抛掷

你正在做一项硬币抛掷游戏，如果正面向上将赢得 1 美元，如果反面向上将输掉 1 美元。在 100 次抛掷以后，因为 45 次正面向上和 55 次反面向上结果你输掉了 10 美元。你继续参与这个游戏直到你抛掷硬币 1 000 次，结果是 480 次正面向上和 520 次反面向上。这个结果和大数定律所预期的是一致的吗？你有赢回一些你所输掉的吗？请解释。

解答 在 100 次硬币抛掷中正面向上的比例是 45%，在 1 000 次硬币抛掷中，正面向上的比率增加至 $480/1\,000$ ，或者说是 48%。因为正面向上的比率越来越接近于 50%，所以这个结果和大数定律所预期的是一致的。然而，你现在赢得 480 美元（480 次正面向上），输掉了 520 美元（520 次反面向上），总共输掉了 40 美元。这样，尽管正面向上的比率越来越接近于 50%，但是你输的钱数是增加的。 ■

6.3.4 连胜

另外一个通常导致赌徒谬论的误解则包含了对连胜的预期。假如你抛掷一枚硬币 6 次结果是 HHHHHH（6 次正面向上）。然后你再抛掷硬币 6 次的结果是 HTTHTH，大多数人都会说后一次的结果是正常的，而前一次连续 6 次正面向上的结果是让人惊异的。但实际上，这两种结果出现的可能性相等。抛掷硬币 6 次的所有可能结果是 $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$ ，出现每一种结果的概率都是 $1/64$ 。

更进一步来说，假如你抛掷硬币 6 次刚好都是正面向上，现在你要打赌下一次抛掷硬币的结果。你可能会想，正面向上出现的次数太多，在下次抛掷中会出现反面向上

的机会。但是在下一次抛掷中正面向上和反面向上的概率都是 50%，硬币没有前一次抛掷的记忆。

思考题

如果一个家庭有 6 个男孩，那么下一个孩子还是男孩的概率是增加了还是减少了，如果一个篮球运动员已经连续 20 次因为冲撞而被罚球，那么下一次还是因为冲撞而被罚球的概率是更多了还是更少了？某一天的天气和下一天的天气是否是独立的（像例 6-16 中所假定的）？说明原因。

相关知识

关于峰的同样一种想法应用于选择彩票中，没有一种特定的彩票数字集合比其他数字集合更可能被抽中，尽管如此，人们还是选择某一种特定的数字集合（比如 1, 2, 3, 4, 5, 6）多于普通的数字组合。因此，如果你以一个普通的数字集合获得大奖，你与其他人平分奖金的可能性更少，净效果是你的奖金会更多。

例 6-16 雨天的计划

一个农场主知道他所在地区一年的这个时候，在某一天里下雨的概率是 0.5。已经连续 10 天没有下雨了，现在他需要确定是否开始灌溉。他能证明下一天将是一个雨天而决定延期灌溉是正确的吗？

解答 连续 10 天的干旱是非预期的，如同一个赌徒一样，农场主现在正处在一个“连输”的状态中。然而，如果我们假定天气的情况每一天和下一天都是相互独立的，那么，希望下一天下雨的概率大于或小于 0.5 的想法是个谬误。 ■

6.3 练习题

统计素养和批判性思维

1. 大数定律 用你自己的话描述大数定律。
2. 期望值 一个遗传学家计算了 5 个出生的婴儿中，女孩的数量最后的结果是 2.5 个。他四舍五入最后确定为 3 个女孩，因为他推理说出生的 5 个孩子中有 2.5 个女孩是不可能的。他的推理正确吗？为什么？
3. 赌博策略 一个职业赌徒在威尼斯人赌场中玩 21 点，在前 10 副牌中每副都输掉了。他准备加大赌注玩下一副。因为他前面获胜的

比率为 0，现在他获胜的比率要大大提高了，因为这样才能接近于平均获胜的次数。他的赌博策略听起来怎么样？他的推理正确吗？为什么？

4. 赌徒的谬论 用你自己的话描述什么是赌徒的谬论。

判断正误：判断练习题 5~8 中的陈述是有意义的（表述正确）还是无意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断会更重要。

5. 轮盘赌策略 斯蒂夫在轮盘赌中连续 40 次压注在数字 13 上，结果没有赢过一次却输掉了 200 美元。他推理说，按照大数定律，他的连输应该结束了，这样才能让他获胜的比率接近 $1/38$ ，这一次他决定在下一次的轮盘赌中加大赌注以挽回他前面所输掉的金钱。
6. 彩票 凯立 在一家便利店工作，离彩票自动售卖机非常近。她推理到如果购买每一张所有可能的数字联合的州彩票，那么她一定会中大奖，所以她计划为买这些彩票准备足够的钱。
7. 彩票 凯姆 购买了一张州彩票，选择的数字是 1, 2, 3, 4, 5 和 6，他推理到这个数字联合和其他任何数字联合获奖的机会是一样的。
8. 硬币抛掷 当一个均匀的硬币被抛掷 10 次时，10 次都是正面向上的结果从不会发生的。

概念与应用

9. 理解大数定律 假如你抛掷一枚均匀的硬币 10 000 次，你预期会得到一个刚好是 5 000 次正面向上的结果吗？为什么？大数定律会告诉你最可能得到的结果是什么？
10. 超速驾驶司机 一个人有超速驾驶的习惯，但是从来没有发生过交通事故，也没有得到交通罚单。这是否意味着大数定律捕捉不到他？实际上是这样吗？说明原因。
11. 你应该怎么玩 如果有人和你打赌，你用两枚骰子不可能抛掷出两个偶数，如果你成功了会赢得 5 美元，如果你失败了就要输掉 1 美元。你认为这个游戏的期望价值是什么？第一次游戏时，你预期是会赢得还是输掉呢？如果游戏进行 100 次你期望的结果是什么样的？说明理由。（表 6-5 会帮助您计算所需要的概率。）
12. 肯塔基州的选择 4 彩票 如果你花费 1 美元购买肯塔基州的选择 4 彩票，你可能会失去 1 美元或者获得 4 999 美元（获奖者的奖金是 5 000 美元，但是你购买彩票的 1 美元是不返还的，所以最后的获奖是 4 999 美元）。这个游戏是在 0000 ~ 9999 之间的 4 位数中选择 一个。获胜的概率是多少？如果你购买了一张彩票的数字是 1234，你获奖的期望值是多少？
13. 橄榄球比赛中的附加分 在橄榄球触地以后球队有一个得 1 分或者得 2 分的选项。把球踢过门柱可以得 1 分，跑过去并且球穿越的

- 球门线的可以得 2 分。根据全美橄榄球联盟某年的统计数据，得 1 分的机会是 94%，得 2 分的机会仅仅是 37%。同样地，如果失败的话就只能得 0 分。计算得 1 分和 2 分的两种不同尝试的期望值。基于这些期望值，哪一种选择在大多数情况下是更有意义的？你认为在何种情况下一个球队应该做出一个不同于期望值所建议的选择吗？说明理由。
14. 保险单的赔付 一家保险公司的精算师依据现有的 1 000 美元的保单进行估计，平均算来，每 100 份保险单持有者就会有 1 份提出 20 000 美元的赔偿要求，每 200 份保险单持有者就会有 1 份提出 50 000 美元的赔偿要求，每 500 份保险单持有者就会有 1 份提出 100 000 美元的赔偿要求。
- (1) 公司每卖出 1 份保单它的期望值是多少？
- (2) 如果公司卖出 100 000 份保险单，预期可以赢利吗？说明这种计算的假设。
15. 预期的等候时间 假如你随机到达一个公共汽车候车站，所有到达时间可能性都是相等的。公共汽车没有延误每 30 分钟到达一次（也就是说，到达时间是在半小时或者是整小时）。你等候时间的期望值是多少？说明你是怎么得到的答案？
16. 强力球彩票 多个州的强力球彩票广告总是强调大奖的金额以及只花 1 美元就有获得大奖的可能性。假定每周大奖价值 3 000 万美元。注意，获得某个金额的奖金的方式不止一种（例如获得 100 美元的奖金有两种方式），下表给出了可以获得的各种奖金数量及其概率。只购买一张彩票获奖的期望值是多少？如果你每年在彩票上花费 365 美元，预计你是赢利多少还是赔掉多少？

奖金（美元）	获奖概率
大奖	1/80 089 128
100 000	1/1 953 393
5 000	1/364 042
100	1/8 879
100	1/8 466
7	1/207
7	1/605
4	1/188
3	1/74

17. 大乐透 多个州的大乐透彩票广告总是强调大奖的金额以及只花1美元就有获得大奖的可能性。大奖是不确定的,但是可以假定它有一个300万美元的平均值。注意两种不同概率的结果可能会获得相同的奖金。只购买一张彩票获奖的期望值是多少?如果你每年在彩票上花费365美元,预计是赢利多少还是赔掉多少?

奖金(美元)	获奖概率
大奖	1/76 275 360
150 000	1/2 179 296
5 000	1/339 002
150	1/9 686
100	1/7 705
5	1/220
5	1/538
2	1/102
1	1/62

18. 轮盘赌的期望值 当你在拉斯维加斯的威尼斯人赌场中,对轮盘赌赌中数字7下了5美元的赌注,你有37/38的概率输掉这5美元,有1/38的概率赢得175美元(奖金是180美元,但是你的5美元赌注是不返还的,所以最后收益是175美元)。如果你的5美元赌注是针对结果都是奇数的,那么你输掉5美元的概率是20/38,而获得5美元的概率是18/38。(最后结果是一个奇数并且赢了,你可以获得包括赌注在内的10美元,所以最后的收益是5美元。)

- (1) 如果你下赌注在数字7上,您的期望值是多少?
- (2) 如果你下赌注在奇数上,你的期望值是多少?
- (3) 那一种选项是最佳的,是下赌注在数字7上,还是下赌注在奇数上,还是不赌博?为什么?

19. 赌场骰子的期望值 当你在赌场一个骰子游戏中下注5美元,赌结果是“过线”。你输掉5美元的概率是251/495,赢得5美元的概率是244/495。(如果你赢了,通常给你5美元并且你可以继续下注,所以你的收益是5美元。)你的期望值是多少,在一个长时间的游戏过程中,每1美元赌注你会输掉

多少?

20. 平均家庭人口 据估计57%美国人家庭人口是1~2人,32%的家庭人口是3~4人,11%的家庭人口是5个及5个以上。说明你将怎么计算美国家庭人口的期望值。这个期望值与平均家庭人口有什么关系?

21. 心理期望值 1953年,一位叫莫里斯·阿莱的法国经济学家进行了一项人们如何评估风险的调查。他设计了两种情境,要求人们在两种可选择的情境中进行选择。

决策一

选项1:100%的机会获得1 000 000美元。

选项2:10%的机会获得2 500 000美元,89%的机会获得1 000 000美元,1%的机会什么也得不到。

决策二

选项1:11%的机会获得1 000 000美元,89%的机会什么也得不到。

选项2:10%的机会获得2 500 000美元,90%的机会什么也得不到。

阿莱发现在决策一中,大多数人们选择选项1,在决策二中大多数人选择选项2。

- (1) 对于每种决策,计算每种选项的期望值。
- (2) 已知调查中的回答与期望值是否一致。
- (3) 对于阿莱调查的结果给出一个可能的解释。

22. 杂志抽奖的期望值 《读者文摘》举行了一次抽奖,获奖金额以及概率列在下面:

1 000 000美元(1/90 000 000的概率),
100 000美元(1/110 000 000的概率),
25 000美元(1/110 000 000的概率),
5 000美元(1/36 667 000的概率),
2 500美元(1/27 500 000的概率)

- (1) 假如抽奖是不需要任何花费的,计算获得奖金的期望值。
- (2) 如果抽奖需要花费一张邮票钱,计算获得奖金的期望值。那么值得加入这个竞争吗?

23. 赌徒的谬论和骰子 假如你和一个朋友玩抛掷骰子游戏,并且遵守如下规则:你每抛掷出的一个偶数,你将从你的朋友那里赢得1美元;你每抛掷出一个奇数,你将输给朋友1美元。

- (1) 一个公正的骰子在一次抛掷中投出偶数的机会是多少？抛掷出奇数的机会又是多少？
 - (2) 假如在前 100 次抛掷中，你投出了 45 个偶数，你将赢得或者输掉多少钱？
 - (3) 假如在第 2 个 100 次抛掷中，你的运气增加了，投出了 47 次偶数，那么在这 200 次抛掷中你共赢得或者输掉多少钱？
 - (4) 假如在下一个 300 次抛掷中，你的运气又增长了，投出了 148 次偶数，那么在总共 500 次抛掷中你共赢得或者输掉多少钱？
 - (5) 在 100 次、200 次和 500 次抛掷中，你投出偶数的百分比分别是多少？解释为什么这个游戏说明了赌徒的谬误？
 - (6) 在下一个 100 次抛掷中你要投出多少个偶数才能够做到不赔不赚？
24. 在硬币抛掷之后：你能追上吗？假如你抛掷一枚均匀的硬币 100 次，结果是 38 次正面向上和 62 次反面向上，反面向上的次数比正面向上的次数多 24 次。
- (1) 说明为什么在你的下一次抛掷中，正面向上的次数和反面向上的次数的差距最可能增长到 25 次或者下降到 23 次。
 - (2) 从第 1 小题的解释中进行扩展，如果你抛掷硬币超过 1 000 次，正面向上的次数和反面向上的次数最终的差异是更可能大于 24 还是小于 24？
 - (3) 假如你继续抛掷这枚硬币，说明为什么如下表述是正确的：如果你随机在任何一次停止抛掷硬币，在总数上看起来正面向上的次数比反面向上的次数会更少。

- (4) 假如你打赌每一次抛掷硬币都是正面向上的，在 100 次抛掷以后，那结果是输了（输掉 62 次赢得了 38 次）。说明为什么。如果你继续打赌，你将很可能会输掉。这个答案与赌徒的谬论是怎样相关联的？

网络 and 远程项目

请在网站 www.aw.com/bbt 上打开第 6 章的有用链接“links for internet projects”。

25. 分析网络上的彩票 浏览所有美国的彩票网站，对各州和多州彩票的奇数和奖金写一个研究总结，选取 5 种彩票并且确定在每种情况下获奖的期望值。并讨论你的结论。
26. 大数定律 使用一枚硬币去模拟 100 个婴儿出生，掷硬币 100 次，记录下结果，并将结果转换成婴儿的性别（反面向上 = 男孩，正面向上 = 女孩）。将这个结果填在如下的表格中，随着样本容量的增长女孩的比率发生了什么变化？它是如何表明大数定律的？

出生人数	10	20	30	40	50	60	70	80	90	100
女孩比率										

新闻中的统计学

27. 个人的大数定律 描述一个你个人应用大数定律的情境，可能正确也可能不正确。为什么你在这种情境下用到大数定律？它对你有什么帮助？
28. 生活中赌徒的谬论 描述一个你或者其他成为赌徒的谬论牺牲品的情境。在这种情境下怎样才能正确控制？
29. 赌徒的毁灭 描述一个你所知道的情境，有人因为赌博失去了几乎所有东西。他的策略看起来是合理的吗？或者说看起来是破坏性加法的结果吗？

会非常喜爱这款产品！它物有所值，它对提高你的生活质量非常有帮助。只是不太有利的地方在于，它最终将会杀死使用它的人。你愿意拥有它吗？

毕竟，不大可能有任何产品好到值得你冒着生命危险拥有它。几周以后，销售员又来见你了：



6.4 风险观念和预期寿命

一位能言善辩但是诚实的销售员来向你推销一种新产品：

我不能透露产品的细节，但是

没有一个人购买这种产品，因此我们做了一些改进。你现在被杀死的机会只有不到 $1/10$ ，你准备购买它吗？

尽管有所改进，但是大多数人仍然会将销售员请出家门，然后直到他再次回来：

好的，这次我们真的将产品完全改善了。现在产品已经变得很安全了，它在 20 年里杀死的人数才和生活在旧金山的居民一样多，它的价格也仅仅是 20 000 美元。

你可能非常惊讶地认识到，如果你像大多数美国人一样，你就会欣然接受这个提议。这个产品就是汽车。它确实在好多方面提高了我们的生活质量，并且 20 000 美元也是汽车非常典型的价格。还有我们也知道大约每年会有 40 000 美国人死于交通事故，在大约 20 年里死于交通事故的人数等于旧金山的人口总数（大约 800 000 人）。

像这个例子中所呈现的，我们经常在和风险和利益之间权衡。在这一节中的，我们可以看到概率的观念如何帮助我们量化风险，也帮助我们在这些权衡之后做出最终的决策。

6.4.1 旅行和风险

汽车还是多用途运动车更安全呢？今天的汽车比 30 年以前更安全吗？如果说你需要穿越美国全境去旅行，是乘坐飞机还是自己驾驶汽车更安全呢？为了回答这类问题，我们必须量化包含在旅行中的风险。然后才能做出适合于我们个人情况的决策。

相关知识

每天大概有 115 人死于交通事故，或者说大概每 13 分钟就有 1 人死于交通事故。交通事故已经成了 6~27 岁人们死亡的主要原因。交通事故死亡率的 85% 是由于车主的责任，另外 15% 则归因于行人和自行车。



旅行风险经常用事故率（accident rate）和死亡率（death rate）这样的术语来表达。例如，假如每年的事故率是每 100 000 人中有 750 起事故。这意味着，在一个 100 000 人的人群中，每年平均有 750 个人发生交通事故。这个陈述本质上是一个期望值，它的意思也表达了一种概率：它告诉我们，一个人在一年内发生事故的概率是 $750/100\,000$ ，或者是 0.0075。

旅行风险的概念是简单的，但是我们仍然必须要注意仔细地解释这个数字。例如，旅行风险有时候被陈述为每 100 000 人，这个和上面的例子是一样的。但有时却陈述为每公里或者每次旅行。如果我们使用每次旅行的死亡率来比较航空和公路交通的风险。类似地，如果我们使用每人的事故率，我们将会忽略这样的事实，那就是大多数交通事故只会造成很小的伤害。

例 6-17 公路交通会越来越安全吗

图 6-11 呈现了 30 年以来的每年的公路交通的总里程数，以及交通死亡率的总人数（全部美国人之中的）。采用每公里行驶路程的死亡率指标，公路交通的风险产生了怎样的变化？

解答 图 6-11a 显示了每年因交通事故而死亡的人数，从 1970 年的 52 000 人下降到 2004 年的 43 000 人。同时，图 6-11b 显示了每年的交通里程数，从 1970 年的大约 10 000 亿公里（ 1×10^{12} ）到 2004 年的 29 000 亿公里（ 2.9×10^{12} ）。这样，这个时期内每公里的死亡率开始和结束的数字分别是：

$$1970 \text{ 年: 死亡 } 52\,000 \text{ 人} / 1 \times 10^{12} \text{ 公里}$$

$$\approx 5.2 \times 10^{-8} \text{ 死亡人数 / 每公里}$$

$$2004 \text{ 年: 死亡 } 43\,000 \text{ 人} / 2.9 \times 10^{12} \text{ 公里}$$

$$\approx 1.5 \times 10^{-8} \text{ 死亡人数 / 每公里}$$

注意到，因为 $10^8 = 10$ 亿， 5.2×10^{-8} 死亡人数/每公里，相当于每 10 亿公里死亡 5.2 人。这样，在 34 年间，每 10 亿公里的死亡率从 5.2 下降到 1.5 人，根据这个测量数据，公路交通在这段时间内应该变得越来越安全。大多数研究者相信这个结果来自汽车安全方面的设计，比如在今天广泛应用的安全带和安全气囊。

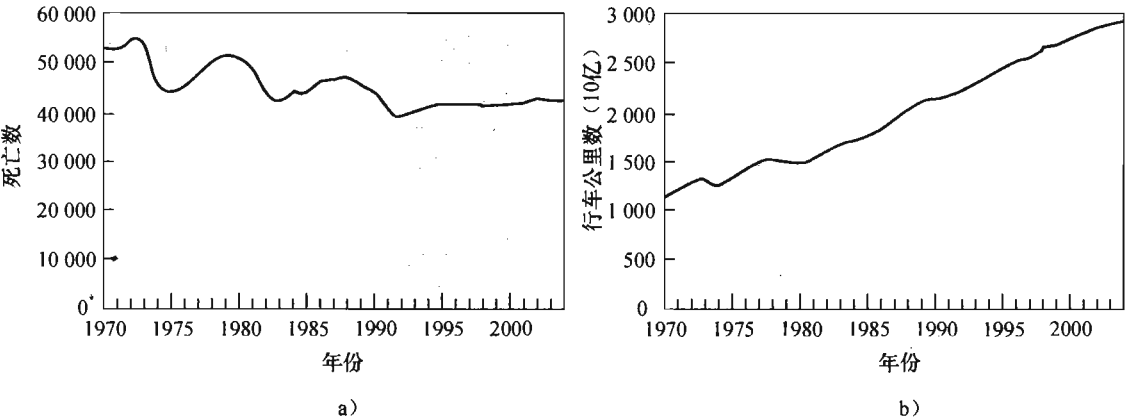


图 6-11 公路交通和死亡率
a) 每年公路交通死亡的人数 b) 每年的交通总里程数

例 6-18 民用航空还是公路交通，哪一个更安全

在美国过去的 20 年间，民用航空的平均死亡人数大约是每年 100 人。（每年实际人数会有明显的变化。）近年来，美国的民航乘客每年大约飞行 80 亿公里。使用这些数据计算民用航空死亡率，比较民用航空的风险和公路交通的风险。

解答 假设每一年民用航空的平均死亡率是 80 亿公里死亡 100 人，飞行的风险就是：

死亡 100 人/ 8×10^9 公里 $\approx 1.3 \times 10^{-8}$ 死亡人数/每公里

这个风险相当于每 10 亿公里死亡 1.3 人，要比公路交通每 10 亿公里死亡 1.5 人（见例 6-17）要稍小一些。注意，因为每一次航空的里程要比一次公路交通的旅程要远得多，每一次旅行中航空的风险要更高一些，虽然航空每公里的风险要小一些。

思考题

假如你需要从亚特兰大到休斯敦进行一次 800 公里的里程。你认为乘坐飞机还是驾驶汽车哪一个更安全？为什么？

6.4.2 人口统计

有关公民的出生和死亡的数据，经常被称为人口统计。它对于权衡风险和收益是非常重要的。例如，保险公司经常要使用人口统计来测量风险和利率。保健专业人士研究人口统计来评价医疗过程，以及决定医疗资源应该集中在哪里。人口统计学使用出生率和死亡率来预测未来的人口趋势。

表 6-8 呈现了一个重要的与死亡原因有关

的人口统计数据。这些数据极其常见，整个表使用年龄、性别和死亡率作为分类栏目。人口统计经常使用每 100 000 人中的死亡人数进行表达，这可以在不同年代之间、不同州之间以及不同国家之间很容易进行比较。

表 6-8 美国人的死亡原因（近期某一年的数据）			
死亡原因	死亡人数	死亡原因	死亡人数
心脏病	684 462	糖尿病	73 249
癌症	554 643	肺炎/流行性感冒	65 681
中风	157 803	老年痴呆症	63 343
肺病	126 128	肾病	42 536
事故	105 695	败血病（血液中毒）	32 243

例 6-19 人口统计的解释

假定美国人口总数是 300 000 000 人，计算并比较每人和每 100 000 人肺炎（和流行性感冒）和癌症的死亡率。

解答 我们将死亡人数除以总人口数，从而计算每人的肺炎和癌症死亡率：

肺炎/流行性感冒：死亡人数 65 681 人/300 000 000 人 = 0.000 22 死亡/每人

癌症：死亡人数 554 643 人/300 000 000 人 = 0.001 8 死亡/每人

我们可以用概率解释这些数字：肺炎和流行性感冒的死亡概率是 0.000 22，癌症的死亡概率是 0.001 8。将它们转换成每 100 000 人的死亡率，我们将每个人的死亡率 $\times 100\,000$ 。这样我们知道每 100 000 人的肺炎/流行性感冒的死亡率是 22 人，癌症的死亡率是每 100 000 人死亡率是 180。癌症死亡率是肺炎或者流行性感冒死亡率的 8 倍。

相关知识

在大学生这个年龄的人群中,酒精消费是影响健康最严重的风险。美国国立卫生研究院估计在大学生中酒精每年导致1 400人死亡,500 000人受伤,与每年70 000起性攻击的案件基本持平。

思考题

表6-8提示,中风的死亡率是50%,远高于事故的死亡率。但是这些数据包括所有的年龄组。你认为中风和事故的风险在年轻人中和在老年人中有什么不同?请加以解释。

6.4.3 寿命预期

我们现在来看预期寿命的观念,经常用它来比较不同时代和不同国家之间总体的健康水平。如果我们来看下死亡率的话,这个概念就会变得越来越清楚。图6-12a呈现了不同年龄组美国人的总体死亡率,以每1 000人的死亡率来计算。注意到,在接近出生的年龄死亡率有一个提升,接下来年龄的死亡率又下到非常低的水平,大约在15岁的时候,死亡率又开始逐渐上升。

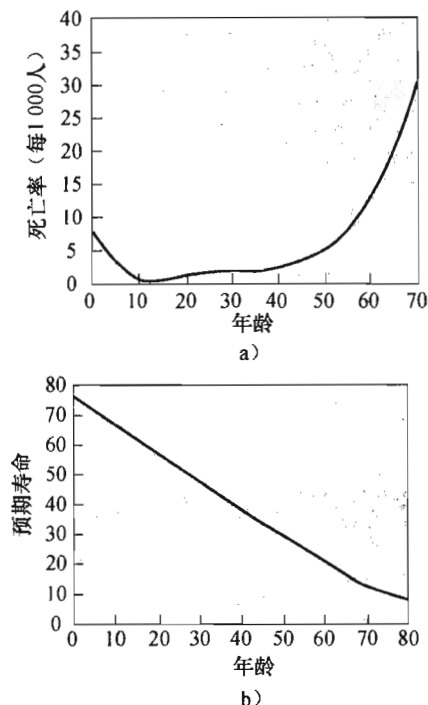


图6-12 死亡率和预期寿命

a) 不同年龄组美国人的死亡率(每1 000人中的死亡数)

b) 不同年龄组美国人的预期寿命

资料来源: U. S. National Center for Health Statistics.

图6-12b 美国人不同年龄的预期寿命 (life expectancy), 它被定义为一个已知其年龄的人预期平均生存的时间。正如我们所预期的, 年轻人的预期寿命较高, 因为就平均而言, 他们的生命还有很长的路要走。在刚出生时, 现在的一个美国人预期寿命是78年(男性是75年, 女性是81年)。

相关知识

预期寿命在世界各地差异很大, 美国的预期寿命(78岁)在许多发达国家之后, 这些国家包括日本(81岁)、澳大利亚(80.5岁)、加拿大(80.2岁)。最低的预期寿命发现于战争和疫病流行的撒哈拉沙漠以南的非洲国家, 比如安哥拉(38岁)和塞拉利昂(40岁)。

定义

◆ 预期寿命 (life expectancy) 被定义为一个已知其年龄的人预期其继续生存的平均年限。

对预期寿命说明的微妙变化来自医学科学和公众健康。预期寿命经常通过近期的死亡率来计算。例如, 我们说一个新生儿预期寿命是78年, 我们的意思是说如果未来的医学科学和公众健康没有大的改变的话, 这个婴儿的寿命平均会是78年。这样, 预计寿命提供了一个现代总体健康的有用测量指标, 它不再被看做未来寿命的一个预测。

实际上, 由于医药科学和公务健康的高速发展, 在过去的一个世纪中预期寿命出现了急剧增长, 这个增长大约是60% (见图6-13)。如果这个趋势持续的话, 现在婴儿平均下来预期寿命会超过78岁。

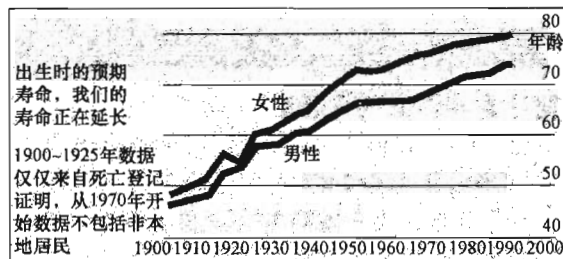


图6-13 20世纪美国人预期寿命的变化

资料来源: New York Times and National Center for Health Science Statistics.

思考题

使用图 6-13 的数据，比较男性和女性的预期寿命，简要讨论这些差别。你认为这对于社会政策有什么启示吗？比如保险利率？请加以解释。

例 6-20 预期寿命

使用图 6-12b 的数据，计算 20 岁的人和 60 岁人的预期寿命。这些数字一致吗？为什么？

解答 图中显示 20 岁人的预期寿命大约是 58 年，

60 人的预期寿命大约是 21 年。这意味着一个 20 岁的人平均下来预期能够再生存 58 年，到 78 岁；一个 60 岁人平均预期能够再生存 21 年，到 81 岁。

首先看起来可能很奇怪，60 岁的人会比 20 岁的预期寿命更长一些（81 岁与 78 岁）。但是要记得预期寿命是基于当前数据的。如果将来的医疗和公共健康条件没有变化的话，一个 60 岁的人活到 81 岁的概率要比 20 岁的人活到 81 岁的概率较大，因为他已经活到了 60 岁。尽管如此，如果医疗和公共健康条件持续改善，那么现在 20 岁的人可能要比现在 60 岁的人活得更长久一些。

案例研究

寿命预期和社会保险

因为美国社会的人口组成有所变化，享受社会保险的退休人数预计在未来会大大增加。工薪阶层支付社会保险的人数保持相对稳定。结果就是，未来社会保障最大的挑战之一是，寻找一条确保能够支付未来退休者社会保险的途径。

近期的统计显示，如果没有明显的变化，社会保障工程将会破产，不能够在 2040 年以后支付全部退休者的保险费用。（这听起来好像很遥远，但是现在的大学生将在那时以社会保险作为主要的收入。社会保险的官员已经提议好多种方法来解决这个问题，包括支付费用的改变、增加社会保险的税收率、延长退休年龄、部分或全部私有化社会保障工程。每个提议都要面对政策障碍。另外，所有提议都基于未来的寿命预期。如果没有退休年龄的对应变化，长寿对社会保险意味着更多的财务支出。

更为特别的是，社会保障官员的一项提议已经假定现在出生的人们预期寿命将在 2030 年提高到 79.3 岁，在 2070 年提高到 81.5 岁（对全部人群）。在更远的将来这些数字将会更悲观。例如，社会保险工程假定美国妇女到 2033 年预期寿命不会达到 82 岁，但是欧洲和亚洲的一些国家女性的预期寿命早已达到这个水平。未来 10 年美国女性预期寿命会达到这个水平吗？一个专家小组提议说，对预期寿命更理性的估计是在 2030 年将达到 81 岁，到 2070 年将达到 85 岁。这意味着到 2070 年社会保障工程将会对每人多提供 4 年财务支付。并且，如果医疗科学成就进一步有所突破，让人们获得更长寿命一些，那么社会保障危机将会更加严重。

思考题

按照一些生物学家的观点，21 世纪，用医学科学成就帮助人们活到 100 岁或超过 100 岁是非常可能的。这样的话，对于社会保障工程会有什么影响？你预想对于社会其他领域会有什么影响？从总体上来说，你认为预期寿命的大幅增长对于社会是好事还是坏事呢？阐述你的观点。

6.4 练习题

统计素养和批判性思维

1. 出生率 以致美国现在的人口出生率是每 1000 人中出生 14.1 人。比较不同国家的人口增长率，为什么使用人口出生率比使用实际的人口出生数字更好一些呢？
2. 人口统计 什么是人口统计？
3. 预期寿命 什么是预期寿命？一个 30 岁的人和 20 岁的人会有相同的预期寿命吗？为

什么？

4. 预期寿命 基于近期的数据，美国 20 岁的人预期寿命是 57.8 年。这意味着什么？

判断正误：判断练习题 5~8 中的陈述是有意义的（表述正确）还是无意义的（表述错误），并做出解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. 飞机失事 飞机空难在媒体中会引起很大关

注，即使空难的死亡者只有一两个人。因为飞机的空难让人如此关注，所以飞机空难的死亡风险要比公路交通的死亡风险大。

- 6. 预期寿命 一位公共健康官员陈述说人们的预期寿命随着年龄增长会逐渐下降。
- 7. 预期死亡年龄 当你变得越来越老时，你的预期死亡年龄会下降。
- 8. 死亡风险 在近一年中，死于交通事故和肺炎的美国人数大体是相当的。这样，交通事故和肺炎的每 100 000 人死亡率是大体相等的。

民用航空的死亡率：练习题 9 ~ 12 中使用下表数据，上面概括了美国 3 个不同年份的民用航空死亡率。

年份	出境	死亡	乘客公里数 (10 亿)	乘客 (百万)
1995	8 062	168	540.7	547.8
2000	9 035	92	692.8	666.2
2004	11 182	14	731.9	697.8

- 9. 对于 3 个年份中的每一年，计算每 1 000 人的死亡率。依据这些死亡率，哪一年是最安全的？为什么？
- 10. 对于 3 个年份中的每一年，计算每 100 亿公里乘客的死亡率。依据这些死亡率，哪一年是最安全的？为什么？
- 11. 对于 3 个年份中的每一年，计算每 1 000 万乘客的死亡率。依据这些死亡率，哪一年是最安全的？为什么？
- 12. 针对 2004 年，计算每公里乘客的死亡率。我们为什么不报告每公里乘客的死亡率呢？

生存表：在练习题 13 ~ 16 中，使用下表数据，上面列出了在美国，年龄在 16 ~ 21 岁之间的美国人的口数。

年龄段 (岁)	死亡 概率	到年龄段开始时 的生存人数	死亡 人数	预期生存时间 (从年龄段开始)
16 ~ 17	0.000 607	98 943	60	61.7
17 ~ 18	0.000 706	98 883	70	60.7
18 ~ 19	0.000 780	98 814	77	59.7
19 ~ 20	0.000 833	98 736	82	58.8
20 ~ 21	0.000 888	98 654	88	57.8

- 13. 预期生存年限 随机挑选一位 19 岁的美国人，从他 19 岁生日那天开始算，他的预期

生存年限是多少？

- 14. 预期生存年限 随机挑选一位 17 岁的美国人，从他 17 岁生日那天开始算，他的预期生存年限是多少？
- 15. 死亡率 人寿保险公司必须非常仔细地监控人口死亡率。在和一个人 19 岁的年轻人签订保险协议之前，公司需要知道这个年龄组的人口死亡率。计算 19 岁年龄组的人群每 10 000 人的死亡率。
- 16. 死亡率 计算 16 岁年龄组的人群每 10 000 人的死亡率。
- 17. 高/低出生率 美国 2003 年的人口出生率最高的和最低的州分别是犹他州和迈阿密州。犹他州报告了大约 2 400 000 人口中新出生了 49 870 人，迈阿密州报告了大约 1 300 000 人口中新出生了 13 861 人。使用这个数据回答如下问题。
 - (1) 犹他州每天出生多少人？迈阿密州每天出生多少人？
 - (2) 犹他州每 1 000 人的人口出生率是多少？迈阿密州的每 1 000 人的人口出生率是多少？
- 18. 高/低出生率 近一年来，加利福尼亚总共死亡了 235 000 人，是美国人口死亡数最多的州，人口死亡数最少的州是阿拉斯加，总共死亡 3 000 人。加利福尼亚州和阿拉斯加州的人口总数分别大约是 35 463 000 人和 648 000 人。
 - (1) 计算加利福尼亚州和阿拉斯加州每 1 000 人的人口死亡率。
 - (2) 基于加利福尼亚州和阿拉斯加州的死亡人数分别是最多和最少的州，你可以认为加利福尼亚州和阿拉斯加州分别是人口死亡率最高和最低的州吗？为什么？
- 19. 美国的人口出生率和死亡率 2006 年，美国人口总数达到了 300 000 000 人。总体的人口出生率估计是每 1 000 人 14.1 人，总体的人口死亡率的估计是每 1 000 人 8.4 人。
 - (1) 2006 年美国大约有多少人口出生？
 - (2) 2006 年美国大约有多少人口死亡？
 - (3) 基于单独的出生率和死亡率（不考虑移民入境和移民出境问题），2006 年，美国的人口增加了多少？
 - (4) 忽略移民入境和移民出境问题，2006

年美国的人口增长率是多少？如果表示为百分比人口增长率是多少？

20. 中国的人口出生率和死亡率 这是中国 2006 年人口估计数值，总人口是 1 313 973 713，出生率估计是每 1000 人 13.25 人，总体人口的死亡率的估计是每 1000 人 6.97 人。
 - (1) 2006 年中国大约有多少人口出生？
 - (2) 2006 年中国大约有多少人口死亡？
 - (3) 基于单独的出生率和死亡率（不考虑移民入境和移民出境问题），2006 年，中国的人口增加了多少？
 - (4) 忽略移民入境和移民出境问题，2006 年中国的人口增长率是多少？如果表示为百分比人口增长率是多少？

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 6 章的有用链接“links for internet projects”。

21. 美国与世界的预期寿命 你可以在网络中的是找到大量关于全世界的人口预期寿命的数据。将美国人的预期寿命与其他发展中国家人口的预期寿命进行比较会怎么样？如何解释你看到的差异，基于你的发现，讨论对美国社会和政府政策的潜在影响。
22. 男性和女性的预期寿命 寻找关于为什么男性和女性的预期寿命会随着时间的变化的数据，以及如何变化的。为什么女性的预期寿命比男性更长一些？我们能够希望男性的预期寿命会赶上女性的预期寿命吗？
23. 非洲国家乌干达的案例研究 寻找过去几十年非洲国家乌干达的预期寿命的数据。你会发现过去 10 年中乌干达的预期寿命有一

个急剧提高。这个提高的解释是什么？这个提高对其他非洲国家会有什么借鉴意义。

24. 预期寿命的计算 你可以在互联网上找到许多可以有效应用的预期寿命计算器。请尝试使用其中的一些，它们能否给出准确的和有意义的结果？尝试一些统计技术用来制作预期寿命表。
25. 里氏震级风险系数 英国皇家统计学会，提出了用于测量地震的风险大小和类似风险因素的里氏震级风险系数系统，到互联网上了解这种风险的测量指标是如何定义和计算的。使用这种指标，讨论不同的活动和事件的风险。
26. 风险理解 彼得·伯恩斯坦在 1996 年出版的 *Against the Gods: The Remarkable Story of Risk* (John Wiley, 1996) 是一本备受赞誉的有关概率和风险评估历史的好书。在图书馆或者书店中寻找这本书（这是一本值得购买和收藏的好书），改变我们对风险理解的某一特定事件。写出有关这一事件的历史和结果的两篇论文。

新闻中的统计学

27. 旅行安全 寻找一则近期的新闻报道，内容是讨论某种类型交通安全的（比如机动车交通或者民用航空器事件的风险，儿童汽车安全座椅的有效性，驾车时接打手机的影响）。总结其中关于风险的统计，并且在讨论基础上说明你自己是如何看待这个安全主题的。
28. 人口统计 寻找一则近期的新闻报道，上面有最近的人口统计或者预期寿命的数据。概括这则报道和其中的统计内容，并且讨论这些新数据对个人和社会的影响。



6.5 组合事件的概率（补充部分）

本章前面所讨论的概率观念，对本书大部分的内容而言已经足够了。尽管如此，概率在统计学和生活中的其他领域有许多更宽泛的用途。在本节中，我们将研究有关概率的另外几个理论以及它们的一些用途。

6.5.1 “和”事件的概率

假如你抛掷两枚公平的骰子，并且想知道

两枚骰子的点数都是 4 的概率。计算这个概率的一种方法是，将抛掷两枚骰子看做一次性抛掷两枚骰子，这样我们可以使用的理论方法来计算概率。因为两枚骰子的点数都是 4 是 36 种可能结果中的一种，所以它的概率是 $1/36$ 。

另外一种可选择的办法，我们可以考虑这两枚骰子分别是独立的。对于每一枚骰子，点数为 4 的概率是 $1/6$ 。我们可以将这两个骰子各自的概率相乘来计算的两枚骰子都是 4 点的概率。

$$P(\text{两枚骰子都是 4 点}) = P(4) \times P(4) \\ = 1/6 \times 1/6 = 1/36$$

通过这样一个办法，抛掷两枚骰子点数都是 4

点的概率是 $1/36$ 。通常,我们将事件 A 和事件 B 同时发生的概率叫做事件的“和”概率 (and probability), 或者叫联合概率 (joint probability)。

概率的乘法定理很容易被扩展到包括两个以上事件的情况中。例如,我们想计算抛掷 10 枚硬币 10 枚都是正面向上的概率,或者计算在同一年生了一个孩子并且同时加薪的概率。然而,在对和事件概率进行计算的时候需要注意一个重要的区别。我们必须能够区分出这些事件是独立事件还是非独立事件,让我们来做分别的探讨。

1. 独立事件

重复抛掷一枚单独的骰子就会产生独立事件 (independent events), 因为每一次抛掷的结果并不影响其他抛掷结果的概率。同样地,抛掷硬币也是独立事件。确定一个事件是否是独立事件是非常重要的,但答案并不总是非常明显的。例如,对一个篮球运动员罚球的成功性进行分析,我们能否确定他的一次罚球和另外一次罚球是相互独立的呢? 只要事件是彼此独立的,我们就可以通过将两个或多个事件的概率进行相乘计算和事件的概率。

相关知识

事件是否独立总是很明显的,思考一下由单独一个人完成的篮球罚球。一些人支持每一次罚球命中的概率都是相同的,一些人认为成功和失败的心理冲突会影响他下一次投篮的准确性。尽管许多统计学家研究了罚球,但是没有一个人能够清楚地说明哪种观点是正确的。



独立事件的和概率

如果一个事件结果并不影响另外一个事件的结果,那么这两个事件就是相互独立的 (independent)。考虑两个独立事件 A 和 B 的概率分别是 $P(A)$ 和 $P(B)$, 那么事件 A 和事件 B 同时发生的概率就是,

$$P(A \text{ 和 } B) = P(A) \times P(B) \quad (6-5)$$

这个法则可以扩展到任何数量的独立事件。例如,独立事件 A 、 B 、 C 的概率分别是 $P(A)$ 、 $P(B)$ 、 $P(C)$, 那么,

$$P(A \text{ 和 } B \text{ 和 } C) = P(A) \times P(B) \times P(C) \quad (6-6)$$

例6-21 3枚硬币

假如你抛掷 3 枚均匀的硬币, 3 枚硬币都是反面向上的概率是多少?

解答 因为抛掷 3 枚硬币是相互独立的, 我们将每一枚硬币反面向上的概率进行相乘:

$$P(3 \text{ 枚硬币都是反面向上}) = P(\text{第 1 枚硬币反面向上}) \times P(\text{第 2 枚硬币反面向上}) \times P(\text{第 3 枚硬币反面向上}) = 1/2 \times 1/2 \times 1/2 = 1/8$$

3 枚硬币都是反面向上的概率是 $1/8$ 。 ■

2. 非独立事件

一批 15 张内存卡其中 5 张是有缺陷的产品。如果你从中随机选择一张, 选择的是残次品的概率为 $5/15$ 。现在, 假如第一次选择了一张残次品, 并且将其放到了你的口袋中。那么你在选第 2 张内存卡时还是残次品的概率是多少?

因为你已经从这批内存卡中抽取了一张残次品, 现在的内存卡只余下了 14 张, 其中 4 张是残次品。这样, 第 2 次从这些卡中抽取到残次品的概率是 $4/14$ 。这个概率要小于第一次抽取时的概率 $5/15$, 因为第一次抽取卡片改变了这一批内存卡的总体数量和残次品的数量。因为第 1 个事件的结果影响了第 2 个事件的结果, 所以这两个事件之间是非独立事件 (dependent events)。

计算非独立事件的概率仍然要将每个单独事件的概率进行相乘, 但是我们要考虑先前的事件是如何影响后续事件的。在前面内存卡的案例中, 我们计算抽取两张残次品的概率, 是将抽取第 1 张残次品的概率 $5/15$ 和继续抽取第 2 张残次品的概率 $4/14$ 相乘:

$$\begin{aligned} P(\text{抽取两张残次品}) &= P(\text{抽取第 1 张残次品}) \\ &\times P(\text{再抽取第 1 张残次品之后抽取第 2 张残次品}) \\ &= 5/15 \times 4/14 = 0.0952 \end{aligned}$$

这样, 计算连续抽取两张内存卡都是残次品的概率是 0.0952。只比在第 2 次抽取内存卡时放回已经抽取的第 1 张内存卡这种情况下的概率 $(5/15) \times (5/15) = 0.111$ 稍小一些。

非独立事件的和概率

如果第一个事件的结果会影响另外一个事件结果的概率，这两个事件就叫做非独立事件。两个非独立事件 A 和 B 同时发生的概率是：

$$P(A \text{ 和 } B) = P(A) \times P(B \text{ 在 } A \text{ 发生后}) \quad (6-7)$$

$P(B \text{ 在 } A \text{ 发生后})$ 的意思是指在事件 A 发生以后事件 B 发生的概率。

这个法则可以扩展到任何数量的非独立事件。例如，非独立事件 A 、 B 、 C 同时发生的概率是：

$$P(A \text{ 和 } B \text{ 和 } C) = P(A) \times P(B \text{ 在 } A \text{ 发生后}) \\ \times P(C \text{ 在 } A \text{ 和 } B \text{ 都发生以后}) \quad (6-8)$$

方法注解

$P(B \text{ 已知 } A)$ 叫做条件概率，在某些书中，它的符号记作 $P(A|B)$ 。

例 6-22 宾戈游戏

宾戈游戏包括从一个箱子里随机地按下有标示的按钮，不允许重复。箱子中一共有 75 个按钮，按钮上分别标有字母 B、I、N、G、O，标有每个字母的按钮分别是 15 个。那么在前两次选择中按下的按钮都是 B 的概率是多少？

解答 宾戈游戏包含的是非独立事件，因为每按下一次按钮就会改变按钮的数量。第 1 次按下按钮为 B 的概率是 $15/75$ 。如果这种情况发生了，在箱子中就余下 74 个按钮，其中 14 个按钮上标志的是字母 B。这样，第 2 次按下标志字母 B 的按钮概率就是 $14/74$ 。因此，在前两次按下的都是标示字母 B 的按钮的概率是：

$$P(B \text{ 和 } B) = P(\text{第 1 次是 } B) \times \\ P(\text{在第 1 次是 } B \text{ 后第 2 次也是 } B) \\ = (15/75) \times (14/74) = 0.0378$$

思考题

不需要计算，从一副扑克牌中抽牌，第 1 种情况是每抽一张牌后要重新洗牌（放回取样），第 2 种情况是每抽出一张牌以后不重新洗牌（不放回取样），比较这两种情况下连续抽出张红心的概率。哪一种情况下的概率较大？为什么？

例 6-23 什么情况下我们将非独立事件看做独立事件

一个民意调查机构通过电话访谈调查了 1 000 人。调查者知道这 1 000 人中有 433 人是民主党的成员。假如给每个人只打一次电话，那么打电话的前两个人

都是民主党成员的概率是多大？

解答 这个问题包括了非独立事件的和概率：一旦给某一个人打过电话以后就不再给他打电话。第一次打电话的那个人是民主党成员的概率为 $433/1\,000$ ，随着这个人从打电话的列表去除以后，第 2 次打电话的那个人是民主党成员的概率就变为 $432/999$ 。这样，打电话的前两个人都是民主党成员的概率就是：

$$(433/1\,000) \times (432/999) = 0.1872$$

如果我们将这两次打电话看做独立事件，两次打电话的对象各自是民主党成员的概率一样是 $433/1\,000$ 。这样前两次打电话的对象都是民主党成员的概率就是：

$$(433/1\,000) \times (433/1\,000) = 0.1875$$

这个计算的结果非常接近于两个事件是非独立事件时的概率。通常，如果从一个非常大的群体中选择相对较少的项目或人数（在本例中，从 1 000 人中选取两人），在这种情况下，非独立事件被看做独立事件的误差就会非常小。一个通常的法则如果是说样本容量小于总体大小的 5%，我们可以将非独立事件看做独立事件。这个操作性原则通常用于调查机构。■

例 6-24 陪审团成员的选定

从一个非常大的男性和女性大致相等的人群中随机挑选一个 9 个人组成的陪审团，所挑选出来的陪审团成员都是男性的概率是多大？

解答 当选择一位陪审员以后，这名陪审员就从人群中去除了。尽管如此，因为我们从一个非常大的人群中选择了很少数的人作为陪审员，这样我们就可以将陪审员的选择看做独立事件。假如人群中男性和女性的人数是相等的，从人群中的随机选取一个人是男性的概率就是 $P(\text{男性}) = 0.5$ 。这样，选择 9 位陪审员都是男性的概率就是：

$$P(9 \text{ 位男性}) = 0.5 \times 0.5 \times 0.5 \times 0.5 \times 0.5 \times \\ 0.5 \times 0.5 \times 0.5 \times 0.5 = 0.5^9 = 0.00195$$

随机选择 9 位陪审员都是男性的概率就是 0.00195，或者说大约为 $2/1\,000$ 。■

相关知识

20 世纪 60 年代后期，在越南战争期间，著名的儿科医生本杰明·斯波克被全部由男性组成的陪审团判决有罪，他的辩护人如果说陪审团中有女性，则会更富有同情心。



6.5.2 “或事件”的概率

到目前为止，我们已经考察了在第一次实验中一个事件发生之后，接下来在第 2 次实验中另一个事件发生的概率。假如我们想知道的

概率是这样的,当进行一次试验时,两个事件中只要有一个事件发生的概率。在这种情况下,我们需要计算的就是“或事件”的概率(either/or probability)。比如生下的一个婴儿是蓝眼睛或者蓝眼睛的概率,再比如在一场火灾或者一次飓风将你的住宅摧毁的概率。

1. 非重叠事件

一枚硬币被抛掷以后,要么正面向上要么反面向上,但是它不可能同时正面向上和反面向上。当两个事件不可能同时发生时,这两个事件叫做**非重叠事件**(non-overlapping events),或者叫做**相互排斥事件**(mutually exclusive events)。我们可以将非重叠事件用维恩图解来表达,每一个事件被表示成一个圆。如果两个圆不重叠,那就意味着对应的事件不会同时发生。例如,我们将抛掷硬币正面向上和反面向上分别表示为两个不重叠的圆,因为一枚硬币不可能同时既是正面向上又是反面向上(见图6-14)。

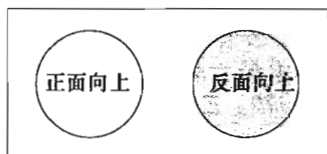


图6-14 非重叠事件的维恩图

假定我们抛掷一枚骰子,想计算结果的点数为1或者2的概率。因为有6个相等的结果,并且问题中的事件对应其中的两个结果。我们可以用理论方法来计算抛掷一枚骰子结果的点数是1或者2的概率, $P(1 \text{ 或者 } 2) = 2/6 = 1/3$ 。作为一种可供选择的方法,我们也可以使用概率的加法定理来计算概率, $P(1) = 1/6$, $P(2) = 1/6$ 。通常,我们计算两个非重叠事件的“或概率”是将两个事件各自的概率进行相加。

非重叠事件的“或概率”

如果两个事件不能同时发生就叫做非重叠事件。如果事件A和事件B是非重叠事件,那么事件A或者事件B发生的概率是:

$$P(A \text{ 或者 } B) = P(A) + P(B) \quad (6-9)$$

这个法则可以扩展到任何数量的非重叠事件中。例如事件A或者事件B或者事件C发生的概率是:

$$P(A \text{ 或者 } B \text{ 或者 } C) = P(A) + P(B) + P(C) \quad (6-10)$$

注意事件A、事件B、事件C都是非重叠事件。

例6-25 抛掷骰子的“或概率”

假如你单独抛掷一枚骰子,结果的点数为偶数的概率是多少?

解答 偶数的结果包括点数为2,4,6,它们是非重叠事件,因为抛掷一枚骰子的点数只可能是其中的一种结果。我们知道 $P(2) = 1/6$, $P(4) = 1/6$, $P(6) = 1/6$,这样,这个组合事件的概率就是:

$$\begin{aligned} P(2 \text{ 或者 } 4 \text{ 或者 } 6) &= P(2) + P(4) + P(6) \\ &= 1/6 + 1/6 + 1/6 = 1/2 \end{aligned}$$

抛掷一枚骰子结果的点数为偶数的概率是1/2。

2. 重叠事件

为了提高从法国到美国的旅游业,两个国家的政府组成了一个由20个人构成的委员会:2名美国男士、4名法国男士、6名美国女士、8名法国女士,如表6-9所示。如果你随机碰见他们中的任何一位,这个人是一位女性或是一位法国人的概率是多少?

表6-9 旅游委员会成员组成

	男性	女性
美国人	2	6
法国人	4	8

20个人中的12个人是法国人,所以遇到一位法国人的概率是12/20。同样地,20个人中的14个人是女性,所以遇到一位女性的概率是14/20。两个概率的和是:

$$12/20 + 14/20 = 26/20$$

这不可能是遇到一位女士或是一位法国人的概率,因为概率不可能大于1。图6-15显示为什么在这种情况下简单的相加是错误的。左边的圆包括12位法国人,右边的圆包括14位女性,美国男性不在其中的任何一个圆里。我们看到,是法国人或者女性的(或者既是法国人又是女性)人数是18位,因为总共的人数是20位,所以遇到一个人是法国人或者女性的概率是18/20=9/10。

如图6-15所呈现的,简单地相加是不正确的,因为两个圆重合的部分包括8个人,他们既是法国人又是女性,如果我们将两个单独的概率相加,这8个人将会被计算两次,一次计算在女性中,另一次计算在法国人中。遇见一位法国女性的概率是8/20,如果我们将这个重复计算的误差从概率中减去,这样遇见一个法国人或者女性的概率就是:

$P(\text{女性或法国人})$
 $= 14/20(\text{女性的概率}) + 12/20(\text{法国人的概率})$
 $- 8/20(\text{法国女性的概率}) = 18/20 = 9/10$

这和我们先前计算的概率是一样的。我们说遇见一位女性和遇见一位法国人是重叠事件 (overlapping events) 或者叫做非相互排斥事件 (non-mutually exclusive events)，因为它们可能同时发生。概括我们在例题中所使用的步骤，可以概括出以下的法则。

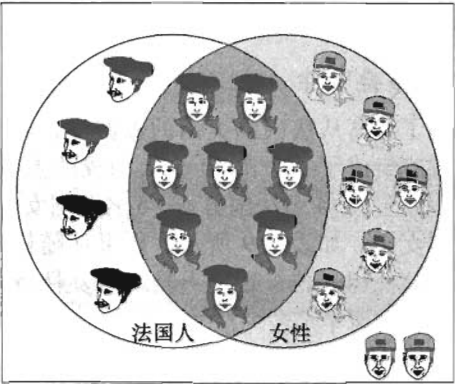


图 6-15 重叠事件的维恩图

重叠事件的或概率

如果事件 A 和事件 B 能够一起发生，它们就叫做重叠事件。事件 A 或事件 B 发生的概率就是：

$$P(A \text{ 或者 } B) = P(A) + P(B) - P(A \text{ 和 } B)$$

(6-11)

上面公式中后面的符号 $P(A \text{ 和 } B)$ 是指事件 A 和事件 B 同时发生的概率，注意使用这个公式不一定是必要的，正确的概率总是要仔细地计算并且要避免重复计数。

思考题

出生在星期四或者出生在拉斯维加斯是重叠事件吗？出生在星期四或者出生在 5 月是重叠事件吗？出生在星期四或者出生在星期五是重叠事件吗？请加以解释。

例 6-26 少数民族和贫困

松鼠希腊是一个典型的美国城镇：它有 2 350 名居民，1 950 名是白人，其中的 11% 或者说是 215 人生活在贫困水平以下。少数民族居民有 400 人，其中的 28% 或者 112 人生活在贫困水平以下（见表 6-10）。如果参观松鼠希腊，你随机遇到的一个人是少数民族或者生活在贫困水平以下的概率是多少？（这个例子是假设的，但是和美国民族分布的百分比是一致的。）

表 6-10 松鼠希腊镇的居民

	贫困水平以下	贫困水平以上
白人	215	1 735
少数民族	112	288

解答 遇到一个少数民族或者遇见一个生活在贫困水平以下的人是重叠事件。使用表 6-10 是非常有用的，它显示了四类居民每类的人数。你可以检查一下表中的数字和给出的数据是否一致，四类的总人数是 2 350 人，因为有 400 名是少数民族居民，所以遇到一位少数民族居民的概率是 $400/2\,350 = 0.170$ 。因为一共有 $215 + 112 = 327$ 人生活在贫困线以下，所以遇到一位居民是生活在贫困线以下的概率为 $327/2\,350 = 0.139$ 。遇到一个人既是少数民族又是生活在贫困线以下的概率是 $112/2\,350 = 0.0477$ 。按照重叠事件的概率法则，遇见一个人是少数民族或者生活在贫困线以下的人的概率是：

$P(\text{少数民族或者贫困线以下})$
 $= 0.170 + 0.139 - 0.0477 = 0.261$

遇见一个人是少数民族或者生活在贫困线以下的概率是大约 1/4，注意减去对应的遇见一个人既是少数民族又是生活在贫困线以下的这个比率的重要性。

6.5.3 总结

表 6-11 提供了计算联合事件概率公式的概括与总结。

表 6-11 联合事件概率计算的总结

独立事件的和概率	非独立事件的和概率	非重叠事件的或概率	重叠事件的或概率
$P(A \text{ 和 } B)$ $= P(A) \times P(B)$	$P(A \text{ 和 } B) = P(A) \times$ $P(B \text{ 在 } A \text{ 发生后})$	$P(A \text{ 或者 } B) =$ $P(A) + P(B)$	$P(A \text{ 或者 } B)$ $= P(A) + P(B) - P(A \text{ 和 } B)$

6.5 练习题

统计素养和批判性思维

1. 独立事件 用你自己的话陈述, 两个事件是独立事件意味着什么?
2. 非重叠事件 用你自己的话陈述, 两个事件是非重叠事件意味着什么?
3. 放回取样的抽样 教授在 25 个学生组成的班级里随机选择了一位学生, 然后又随机选择了第 2 位学生, 如果所有 25 位学生在第 2 次选择中都是有效的, 这个抽样是放回取样的还是不放回取样的? 第 2 次抽样的结果与第一次抽样是相互独立的吗?
4. 对立事件 如果两个事件是对立事件, 比如 A 和非 A , 这两个事件一定是非重叠事件吗? 为什么?

判断正误: 判断练习题 5~8 中的陈述是有意义的 (表述正确) 还是无意义的 (表述错误), 并加以解释。并非所有的陈述都有明确的答案, 所以你的解释比判断会更重要。

5. 彩票 最近一期的彩票大奖数字是 5, 17, 18, 27, 36 和 41, 下一次彩票中它们出现的可能性会减少吗? 还应该选择这些数字吗?
6. 组合事件概率 旋转一枚硬币, 结果是正面向上的概率是 0.5, 从一副扑克牌中选取一张扑克是红色的概率也是 0.5。当旋转一枚硬币和从一副扑克牌中抽取一张牌, 硬币正面向上或者扑克牌是红色的概率是 $0.5 + 0.5 = 1$ 。
7. 或事件的概率 $P(A) = 0.5$ 和 $P(A \text{ 或 } B) = 0.4$ 。
8. 彩票 在彩票的数字组合中, 1, 2, 3, 4, 5 和 6 的数字组合要比其他数字的组合出现的可能性小得多。

概念与应用

9. 出生性别 假设男孩和女孩总数是相等的, 并且一个儿童的性别不会被他的兄弟姐妹的性别所影响。计算一对夫妇有 3 个孩子都是女孩的概率。
10. 猜测 一个小测验包括一道是非题和一道有 4 个选项的选择题。如果两道题的回答都是随机猜测的, 计算两道题都答对的概率。在这样的小测验中, 猜测是一个好的策略吗?
11. 密码 一个新的计算机的所有者设置了一个包括两个字符的密码, 他随机选择了一个字母作为第一个字符, 然后选择了一个数字

(0, 1, 2, 3, 4, 5, 6, 7, 8, 9) 作为第二个字符。他的密码是 K9 的概率是多少? 如果有人试图登录他的计算机的话, 他的密码是一个有效的障碍吗?

12. 穿着橙色衣服的猎人 关于猎人受伤害和穿橙黄色衣服之间关系的一个研究显示, 在 123 起由于失误所导致的伤害中, 6 个人穿着橙黄色的衣服 (数据来自疾病控制中心), 如果下面的一个研究从这 123 个例子中随机挑选一个猎人, 计算选择的前两名猎人都穿着橙黄色衣服的概率是多少?
 - (1) 假定在第 2 个猎人被挑选时, 第 1 个被挑选的猎人被放回人群中。
 - (2) 假定在第 2 个猎人被挑选时, 第 1 个被挑选的猎人不放回人群中。
 - (3) 这种情况下, 哪一种抽样更有意义? 放回抽样还是不放回抽样? 为什么?
13. 收音机上的音乐 一个 MP3 上存储了 60 首音乐曲目, 30 首是摇滚风格, 15 首是爵士风格, 另外 15 首是布鲁斯风格, MP3 设置成随机播放, 乐曲被随机播放并且能够重复。下面每一种事件各自的概率分别是多少?
 - (1) 前 4 首乐曲都是爵士。
 - (2) 前 5 首乐曲都是布鲁斯。
 - (3) 第 1 首乐曲是爵士, 第 2 首是摇滚。
 - (4) 在前 4 首乐曲中, 没有 1 首是摇滚。
 - (5) 第 2 首乐曲和第 1 首是一样的。
14. 电话调查 一个电话调查者有 45 个投票者的姓名和电话号码, 其中 20 人是注册民主党人, 另外 25 名是注册共和党人, 打电话的顺序是随机的。假定你想计算前两个电话都打给共和党的概率。
 - (1) 这是独立事件还是非独立事件? 请加以解释。
 - (2) 如果你把它们看做独立事件, 前两个人都是共和党的概率是多少?
 - (3) 如果你把它们看做非独立事件, 前两个人都是共和党的概率是多少?
 - (4) 比较第 2 小题和第 3 小题的结果。

概率和法院判决: 下表的数据显示了在 1 028 名犯罪嫌疑人的法院判决案例中, 做有罪辩护和无罪辩护的结果。使用这个数据

回答练习题 15 ~ 20 题。

	有罪辩护	无罪辩护
入狱服刑	392	58
无罪释放	564	14

资料来源：Brereton and Casper, "Does It Pay to Plead Guilty? Differential Sentencing and the Functioning of the Criminal Courts," *Law and Society Review*, Vol. 16, No. 1.

- 15. 随机挑选一位犯罪嫌疑人做有罪辩护或者被判入狱的概率是多少？
- 16. 随机挑选一位犯罪嫌疑人做无罪辩护并被判无罪的概率是多少？
- 17. 如果随机选取两位被告人都进入有罪抗辩的概率是多少？
- 18. 如果随机选取两位被告人，他们都被判入狱的概率是多少？
- 19. 如果随机选取一位被告人，这位被告人进入有罪抗辩并且被判入狱的概率是多少？
- 20. 如果随机选取一位被告人，这位被告人进入有罪抗辩但是被宣判无罪的概率是多少？

行人的死亡：对于练习题 21 ~ 26，使用下表，表中总结了 985 位死于交通事故的行人。（数据来源于美国高速公路交通安全委员会。）

		行人醉酒	
		是	否
驾驶员醉酒	是	59	79
	否	266	581

- 21. 如果随机选取一例死于交通事故的行人，计算行人喝酒或者驾驶员喝酒的概率。
- 22. 如果随机选取一例死于交通事故的行人，计算行人没有饮酒或者驾驶员没有饮酒的概率。
- 23. 随机挑选一例死于交通事故的行人，计算行人饮酒或者驾驶员没有饮酒的概率。
- 24. 如果随机挑选一例死于交通事故的行人，计算驾驶员饮酒或者行人没有饮酒的概率。
- 25. 如果随机选择两例不同的行人死亡的交通事故，计算两例事故中驾驶员都饮酒的概率。
- 26. 如果随机选取两类不同的行人死亡的交通事故，计算两例事故中行人都饮酒的概率。
- 27. 药物的检测 一项治疗过敏药物检验，给予 120 人药物，给予另外 100 人安慰剂。一个控制组包括 80 人没有得以任何处理。每一组有所改善的人数呈现在下表中。

	治疗过敏药物	安慰剂	控制组	总计
改善	65	42	31	138
没有改善	55	58	49	162
总计	120	100	80	300

- a. 随机选择研究中的一个人是给予药物或者安慰剂的概率是多少？
 - b. 随机选择的一个人症状没有改善或者有所改善的概率是多少？
 - c. 随机选择的一个人给予药物治疗或者有所改善的概率是多少？
 - d. 随机选择一个人给予药物治疗并且有所改善的概率是多少？
28. 政治党派 在一个典型的城镇里（与人口统计学的分布一致），成年人的政治党派情况如下表所示，所有数字都是百分比。

	共和党	民主党	独立人士	总计
男性	17	15	18	50
女性	14	20	16	50
总计	31	35	34	100

- (1) 随机选择本城镇的一个人是共和党或者是民主党的概率是多少？
 - (2) 随机选择本城镇的一个人是共和党或者是一位女性的概率是多少？
 - (3) 随机选择本城镇的一个人是独立人士或者一位男士的概率是多少？
 - (4) 随机选择本城镇的一个人是共和党并且是一位女性的概率是多少？
29. 遗传和概率分布 某些特征是被一个显性基因 A 和一个隐性基因 a 所控制。假如父母双方具有这些基因的比率是 3:1。也就是父母任何一方具有基因 A 的概率是 0.75，具有基因 a 的概率是 0.25。假如这些基因是从父母那里随机遗传的。为了回答下述问题，设想有 100 次“发生”试验。
- (1) 一个儿童从父母双方那里都遗传了一个基因 A 的概率是多少？
 - (2) 一个儿童的从父母一方遗传了一个基因 A，从另一方遗传了一个基因 a 的概率是多少？注意这种情况能以两种方式发生。

- (3) 一个儿童从父母双方那里都遗传了一个基因 a 的概率是多少?
- (4) 制作一个表格显示所有事件的概率分布。
- (5) 如果组合事件 AA 和 Aa 产生的结果都是同样的显性特征 (比如说棕色头发), aa 产生的结果是隐性特征 (比如说金色头发), 那么一个儿童具有显性特征的概率是多少?
30. 宾戈游戏 宾戈游戏包括从一个箱子中随机按下标志有数字和字母的按键。B 字母的数字是 1~15, I 字母的数字是 16~30, N 字母的数字是 31~45, G 字母的数字是 46~60, O 字母的数字是 61~75。按键在选择以后不可以重复。下面的每一种事件所对应的概率是多少?
- (1) 按下一个 B 按钮。
 - (2) 连续按下两个 B 按钮。
 - (3) 按下一个 B 按钮或者 O 按钮。
 - (4) 先按下一个 B 按钮, 接着按下一个 G 按钮, 然后又按下一个 N 按钮, 以这样的顺序。
 - (5) 在 5 次按下按钮的情况下没有出现一个 B 按钮。

至少一个的问题: 一个通常的概率问题需要计算在给定次数的试验中某一种事件至少出现一次的概率。假定这个特定事件的概率是 P (例如, 从一副扑克牌中抽出一张红心的概率是 0.25)。那么在 N 次试验中, 某个事件至少出现一次的概率是:

$$1 - (1 - P)^N$$

例如, 在 10 次抽取扑克牌的实验中 (放回取样) 至少抽出一张是红心的概率是:

$$1 - (1 - 0.25)^{10} = 0.944$$

使用这个法则解答练习题 31 和 32。

31. 德·梅尔骑士之赌 据说在 17 世纪, 概率理论是用来解释一个贵族名叫德·梅尔骑士的赌博的。

- (1) 在他的第一个游戏中, 骑士打赌, 一枚公正的骰子在 4 次抛掷中至少出现一次 6 点。如果游戏重复进行, 他有多大希望赢得这个游戏?

- (2) 在他的第 2 个游戏中, 骑士打赌说两枚公正的骰子, 在 24 次抛掷中至少出现一次双 6 点。如果游戏重复进行, 他们有多大希望赢得这个游戏吗?

32. 大学生群体中的艾滋病 (AIDS) 假定在某一所大学中的大学生已知携带 AIDS 病毒的百分比是 3%。

- (1) 如果一个学生在一学年中有 6 位性伴侣, 那么他们中至少 1 人携带 AIDS 病毒的概率是多少?
- (2) 如果一个学生在一学年中有 12 位性伴侣, 那么他们中至少 1 人携带 AIDS 病毒的概率是多少?
- (3) 如果一位学生遭遇艾滋病病毒携带者的概率超过 50%, 那么他需要有多少性伴侣?

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 6 章的有用链接 “links for internet projects”。

33. 模拟 一个经典的概率问题涉及一个国王想在他的王国中增加女性的比率。他发布命令说, 在一个妈妈生了一个儿子之后禁止再生孩子。国王推理说一些家庭将会只有一个男孩子同时其他家庭会有几个女孩子和一个男孩子, 因此女孩的比率将会增加。采用抛掷硬币来模拟王国严格遵守这个规定的结果: “在一个妈妈生下一个儿子之后, 她将不再生其他孩子。” 如果这个规定被执行, 女孩的比率会上升吗?

新闻中的统计学

34. 一位《纽约每日新闻》的专栏作家斯蒂芬·艾伦斯沃斯写出了他在纽约州彩票中的选择秘诀。他极力推介使用一种“冷门数字”法的选号系统。冷门数字是指彩票数字在 7 天内只出现一次或者根本没有出现过的数字。他做了如下的表述: “这个选号系统会产生如下的数字组合比如 5-8-9, 7-8-9, 6-8-9, 0-8-9, 3-8-9。这 5 个数字组合在本周里会有极高的被选中的机会。祝所有人都有好运。” 这个选号系统有用吗? 为什么?

第6章复习题

对于复习题1~6,采用如下数据,包含内容是没有怀孕的被试和怀孕被试的临床血液测试结果(“Specificity and Detection Limit of Ten Pregnancy Tests,” by Tiitinen and Stenman, *Scandinavian Journal of Clinical Laboratory Investigation*, Vol. 53, Supp. 216)。

	检查结果阳性 (显示怀孕)	检查结果阴性 (显示没有怀孕)
怀孕的被试	80	5
没有怀孕的被试	3	11

1. 如果随机挑选99位被试中的一位,计算所挑选的被试是孕妇的概率。
2. 如果随机挑选99位被试中的一位,计算所挑选的被试检测是阳性的概率。
3. 如果随机挑选99位被试中的一位,计算所挑选被试是孕妇或者检测是阳性的概率。
4. 如果随机挑选99位被试中的一位,计算所挑选的被试检测是阴性或者她是一位孕妇的概率。
5. 如果随机挑选两位不同的被试,计算她们检测都是阳性的概率。
6. 如果随机挑选两位不同的被试,计算她们都是孕妇的概率。
7. 二进制的计算机公司所制造的计算机芯片应用于DVD播放器上。这些制造出来的芯片27%有效,意味着其中27%产品是好的,而其他产品则有缺陷。
 - (1) 如果随机选择一个芯片,计算这个芯片有缺陷的概率。
 - (2) 如果随机选择两个芯片,计算这两个芯

片都是好产品的概率?

- (3) 如果随机选择5个芯片,其中的好芯片可能有几个?
 - (4) 如果随机选择5个芯片,计算它们全部是好产品的概率是多少?如果选择的5个芯片都是好的,你还能相信成品率是27%吗?为什么?
- 对于练习题8~10,该事件发生的概率小于或等于0.05时为“不寻常”事件。
8. 回答如下问题:
 - (1) 一位指导教师痴迷于十进制,他设计的选择题包含10个可能选项,其中只有一个是正确的。如果随机猜测的话,答对这样一个问题的概率是多少?
 - (2) 通过猜测答对这样一个问题是“不寻常”事件吗?
 9. 回答如下问题:
 - (1) 一个轮盘赌有38个扇区,一个扇区标注数字0,还有一个扇区标注数字00,其他的扇区标注数字1~36。如果你将所有的零用钱打赌旋转的最后结果是数字13,你最终获胜的概率是多少?
 - (2) 在轮盘赌中你获胜是一个“不寻常”事件吗?
 10. 回答如下问题:
 - (1) 一个研究随机选择400架美国飞机,其中344架准点到达目的地。你估计一架美国飞机延误飞至目的地的概率是多少?
 - (2) 一架美国飞机延误飞至目的地是“不寻常”事件吗?

第6章小测验

1. 一个多项选择测验题有6个备选答案,a、b、c、d、e和f,其中只有一个答案是正确的。如果你做随机猜测的话,那么选对的概率是多少?
2. 三道选择测验题,每道题有6个备选答案,a、b、c、d、e和f,其中只有一个答案是正确的。如果你对这3道题的答案进行随机猜测的话,那么3道题全猜对的概率是多少?
3. 如果 $P(A) = 0.45$,求 $P(\bar{A})$ 。
4. 如果抛掷一枚硬币15次,计算正面向上的期

望值。

5. 在一个临床试验中,一些被试采用尼古清戒烟贴进行戒烟,另外一些被试采用安慰剂。结果显示发生副作用的概率是0.279。对于这个结果,偶然性是一个合理的解释吗?

在练习题6~10中,使用下面的一个有关尼古清戒烟贴的临床试验数据,嚼口香糖可以帮助人们戒烟(数据来源于美国梅若道化学药品公司)。

	尼古清戒烟贴	安慰剂
感觉到口腔和喉咙疼痛	43	35
没有感觉到口腔和喉咙疼痛	109	118

6. 如果随机选取试验中的一位被试，计算给予他安慰剂的概率。

7. 如果随机选取试验中的一位被试，计算这名被试使用尼古清戒烟贴并且感觉口腔和喉咙疼痛的概率。
8. 如果随机选取试验中的一位被试，计算这名被试使用尼古清戒烟贴或者感觉口腔和喉咙疼痛的概率。

9. 如果随机选择试验中的两位被试，计算他们都是从安慰剂组抽取的概率。

10. 如果随机选择试验中的3位不同的被试，计算他们都是从尼古清戒烟贴实验处理组中抽取出来的概率。

| 聚焦社会科学 |

彩票公平吗

彩票已经成了美国生活方式的一部分。绝大多数州都有官方彩票，包括一些联邦性的彩票例如“给力”和“大乐透”。联邦统计数据显示，美国人彩票投入人均每年接近200美元。因为很多人从来不买彩票，这也意味着活跃的彩民每年投入大大超过这个数字。

计算彩票的获奖概率就是将那些成功的数字组合统计出来，计算过程可能会复杂一些，但是基本的结论总是不变的：赢得大奖的机会是微乎其微的。彩票广告让人听起来似乎彩票是一种不错的投资，可是实际上彩票可预期的收益却是少得可怜。平均算下来，每个彩民预计的损失大致是投入的一半。

彩票支持者也会指出一些积极的方面。比如，由于有彩票的大笔收入，州政府可以有足够的预算用于改善教育、文娱和环保。这些收入也可以让州政府的税率保持相对较低的水平。彩票支持者还指出，彩民购买彩票完全是自愿和享受的，也代表了社会的某一个方面。事实上，最新的民意测验显示，3/4的美国人赞同政府发行彩票，而只有2/3的人赞同赌博合法。

上述这张讨人喜欢的图片是彩票市场营销与公共关系的一部分。科罗拉多州的官方提供了一组关于彩民与大众的年龄、收入以及教育水平的数据对比（见图6-16）。彩民的年龄分布与大众的年龄分布大体是平行的，只存在很小的百分点的差别。同样地，收入的直方图也直观地显示，除了收入在15 000~25 000美元和25 000~35 000美元的人群外，彩民作为一个整体与典型的市民是一致的，就在人口中占的比例而言，穷人购买彩票会更多一点。

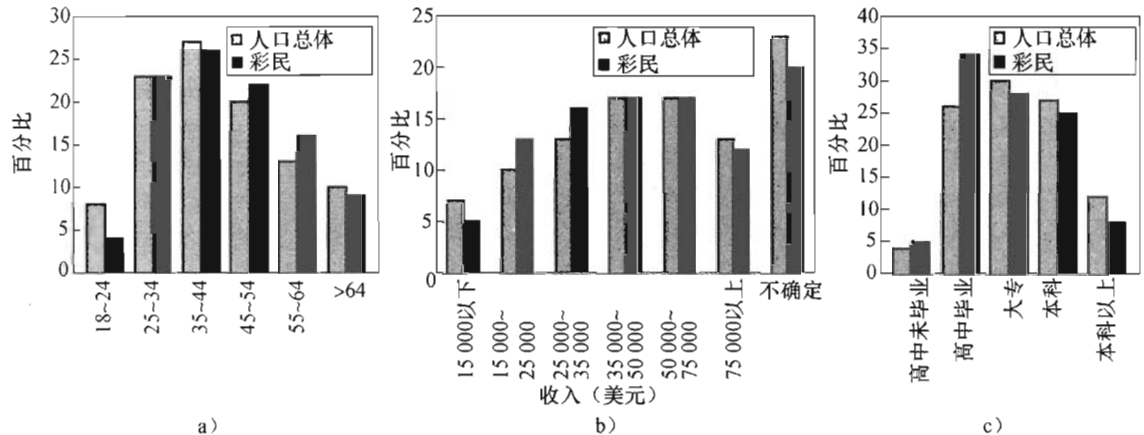


图 6-16 科罗拉多州彩民与一般大众分布对比图
a) 年龄对比 b) 收入对比 c) 教育背景对比

尽管彩票有这些显而易见的好处，仍然有批评者坚持认为，彩票只是一种不公平的税收形式。一些支持的观点来自联邦赌博影响研究委员会近期发布的研究报告以及《纽约时报》在新泽西州的关于彩票的调查。这些研究都关注个人花费在彩票上的总投入。

《纽约时报》调查了 48 875 名来自新泽西州的至少中过 600 美元以上奖金的彩民。这是一个设计精巧的取样，中奖彩民均来自从所有彩民中随机抽取的一个样本，因此，这些中奖者组成的样本完全是随机的。当然，这个样本并不能真正代表全部彩民，因为这些彩票中奖者会比一般彩民买彩票更多些。通过分析彩民的邮政编码地址，研究者可以研究彩民的收入水平、平均受教育程度以及多种人口统计学意义上的特征。最后，《纽约时报》调查的主要结论是，相对而言，彩票投入对那些低收入以及低学历的人影响更大，比如说，下面就是一部分特别的发现：

- 1. 来自低收入社区的人们花费在彩票上的投入占他们收入的比例是高收入社区的 5 倍。低收入社区每 10 000 美元的收入中有 29 美元花费在某种彩票上，仅比高收入社区少 5 美元。
- 2. 低收入社区每 10 000 个人拥有的售彩中心的数字（可以购买彩票的地方）几乎是发达社区的两倍。
- 3. 每 10 000 美元的年收入中用于彩票花费的，低学历社区是高学历社区的 5 倍。
- 4. 彩票的广告和促销也主要针对的是低收入社区。

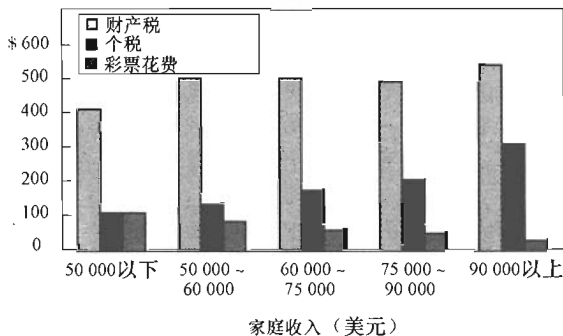


图 6-17 新泽西州中奖者在彩票与税收上的花费（数字来自每 10 000 美元收入）。

《纽约时报》的部分调查结果在表 6-17 中概括了出来。尽管新泽西州有着正向的税收结构（高收入人群的税赋在收入中所占的比例更高），但是“彩票税”却是负向的。更进一步的调查表明，彩票收入贡献比例较高的社区并没有分享相应比例的资金。

在其他州同样的研究也发现类似的结果。总体结论是毫无质疑的：虽然彩票为当地州政府提供了诸多好处，但是这些收入的大部分却来自低收入和低学历人群。事实上，“联邦赌博影响研究委员会”的报告中就把彩票定义为“美国最普遍的赌博”，认为州政府“不负责任地通过以下方式在社会上极大范围内推广赌博……诸如持续的广告宣传以及临近商店无处不在的售彩机器”。

问题讨论

- 1. 研究图 6-16，从年龄分布的角度说，彩民是美国社会典型的横断面吗？从收入上讲呢？从教育水平上说呢？请解释原因。“收入不确定”那类人群怎样影响统计结果？
- 2. 有些彩民用系统的科学方法选号，比如他们咨询专家哪个号码频率高或者注定出现。那些科学方法真的能提高中奖概率么？为什么？你认为科学方法的使用和受教育背景有关吗？
- 3. 基于这篇文章提到的因素以及其他你发现的事件，你认为彩票对那些穷人及低学历的人公平吗？彩票发行是合法好呢，还是应该加以限制呢？
- 4. 找到并研究一则特别的彩票广告，分析它是不是在误导受众？
- 5. 有句在网上传播的名言，“彩票对不懂数学的人来说就是一种税收。”就它所表达的意义和准确性给予点评。

建议阅读

1. Pulley, Brett. "Living Off the Daily Dream of Winning a Lottery Prize." *New York Times*, May 22, 1999.

2. Safire, William. "Lotteries Are Losers." *New York Times*, June 21, 1999.

3. Schwartz, Nelson D., and Nixon, Ron. "Some States Consider Leasing Their Lotteries." *New York Times*, October 14, 2007.

4. Stodghill, Ron, and Nixon, Ron. "For Schools, No Lucky Numbers." *New York Times*, October 7, 2007.
5. Walsh, James. *True Odds*. Merritt Publishing, 1996.

| 聚焦法律 |

DNA 指纹识别可靠吗

DNA 指纹识别（也称为 DNA 指纹分析或 DNA 鉴定）已经成为法律执行的一种主要工具。它被应用于刑事案件、亲子鉴定案例，甚至人体残骸的身份鉴定（在 2001 年世贸中心的“9·11”恐怖袭击事件之后，DNA 指纹识别是鉴别遇难者遗体的主要方法）。

DNA 指纹识别的科学基础已经存在了几十年。然而，这些想法首次被应用在法庭上还是在 1986 年。案件是关于一个 17 岁男孩被指控奸杀英格兰中部地区纳伯勒的两名女学生。在对嫌疑人的审讯过程中，他被要求做血液测试，血样被送到了位于莱斯特大学附近著名遗传学家亚历克·杰弗里斯的实验室里。杰弗里斯运用在亲子鉴定中已有所发展的方法，将嫌疑人的 DNA 与从受害者身上得到的样本进行了对比。测试结果显示，两起强奸案都是同一个人作案，但却不是被羁押的那个嫌疑人。在接下来的一年，经过收集了超过 4 500 份血液样本以后，研究者运用杰弗里斯的方法成功地找出了凶手。

关于这起英格兰案件和杰弗里斯的方法的言论迅速地传播开来。DNA 识别技术迅速地被测试、商业化和宣传推广。毫无例外地，DNA 鉴定技术的使用也立即遭到了反对和争论。

为了探究概率学和统计学在 DNA 鉴定中的关键作用，我们把它看做一个简单的目击者类比。假设你正在寻找一个曾经帮助你摆脱困境的人，而你只记得这个人的三个特征：

1. 这个人是个女性。
2. 她的眼睛是绿色的。
3. 她有着红色的长发。

如果你发现有人符合这个形象，你能得出结论说这个人就是曾经帮助过你的女人吗？为回答这个问题，你需要数据来告诉你自己，在人群中随机选择个体同时具有这些特征的概率。一个人是女性的概率是 $1/2$ 。我们可以假设一个人是绿眼睛的概率是 0.06（世界上 6% 的人口是绿眼睛），一个人是红色长发的概率是 0.007 5。如果我们设想这些特征彼此间独立的，那么随机选取一个人同时拥有 3 个特征的概率则是：

$$0.5 \times 0.06 \times 0.0075 = 0.000225$$

或者说在 10 000 人中大约有两个人。这个概率相对较低，但是可能还没有低到足以得出一个可靠的结论。例如，10 000 人中有两个人符合要求，如果在一个百万人口的城市里会有 200 个人符合条件。显然，如果你想确保你找到了正确的人，你会需要关于那人外貌的额外信息。

DNA 识别就是根据相似的理念，但它的设计比外貌匹配的概率低得多。每个个体的 DNA 是唯一的并且贯穿一生不变，一个单一的生理性状可以由一小片位于染色体上特定位置的 DNA 决定，这一小片 DNA 叫做基因，人类有 23 条染色体和超过 30 000 个基因（见图 6-18）。一个基因可以有两种或更多（通常上百种）不同的形式，称为等位基因。不同的等位基因因性状的不同而变化（例如，不同的头发颜色或不同的血型）。不仅不同的等位基因会出现在同一基因位点，而且相应的 DNA 片段在不同的人身上会有不同的长度（叫做数目可变的串联重复序列，或 VNTRs）。实验室进行收集和基因的证据，包括 5~8 个不同基因位点的等位基因的长度或类型。

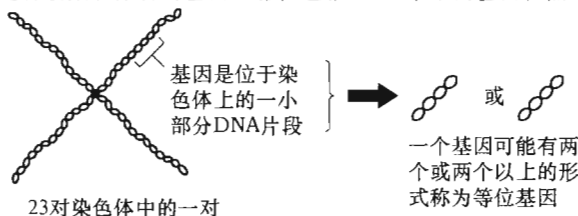


图 6-18 染色体、等位基因、基因和基因位点的关系图



至少从理论上说，收集基因证据（从血液样本、组织、毛发、精液甚至邮票上的唾液）并分析是简单而直接的。然而，这个过程既备受争论，证据来源的误差也大。假设，在我们的类比中，选取到的一个人是“棕红色”头发而不是“红色”头发。因为很多特征属于连续变量（非离散），那你是应该排除这个棕红色头发的人还是假定这个人已经足够匹配了呢？由于这个原因，分组的问题就变得非常重要（见3.1节）。你可以将具有红色头发的所有人都包含在内，或者你可以选择一个更有限的范围（只能是鲜红色头发的人），这样会得到一个更有鉴别力的测试。

同样的问题也出现在基因检测中。在实验室中测量等位基因的类型或长度时，连续变量在测量中有非常大的变异或误差。分组宽度要进行选择，而这个选择是争论的来源。宽度越小使得测试结果更加精密完善，会排除更多的可疑因素，最终会为指控被告提供更有力的证据。

科学争议的另一个来源是关于一个设想，它假设遗传特征是相互独立的。关于这一点，一些专家（遗传学家和分子生物学者）大篇幅地验证也没有达成一致。但存在一个一致观点认为，和其他误差来源比较，特征独立假设并没有引进显著的误差。重要的是理解一个科学难题如何对相关数学计算产生影响：如果遗传特征是相互独立的，那么独立事件运用乘法法则是合理的；如果遗传特征不是独立的，就需要使用另一种不同的概率算法。

争议的第三点涉及参照人群的选择。有些人认为，如果使用了全部人口的遗传数据，那这些数据可能就不能真正地表示民族亚群。例如，相对于整个美国人，在所有意大利人中一个特定的等位基因可能会拥有更多完全不同的次数。这样的—个分歧会改变概率计算的结果，并影响到为案子中被告提供的证据是有利的还是不利的。

图6-19展示了在基因测试中使用参照人群数据的例子。横轴表示测量不同等位基因的30个组，竖轴表示每一组的频率。图中显示的是四个亚洲亚群的频率曲线。使用专门的亚群数据库导致了曲线中的显著变化。然而，这样的数据库十分完整，并且使用专门数据库能为案件提供充足的证据之前，现有的做法还是选择使用全部人群作为参照。

DNA技术开辟了一条新的道路，把科学家和法律学者紧密联系在一起，并改变着我们的社会。毋庸置疑，统计学方法和思想会是改变中不可或缺的部分。

问题讨论

1. DNA检测结果被认为是物证（与环境间接证据相对）。但是与典型的物证相比，例如凶器或者一件衣服，DNA检测就显得更加精密复杂和难以理解。一些人因此认为，当刑事审讯的陪审团成员没有完全理解如何收集和分析DNA检测时，DNA检测结果不应该被作为证据使用。你是如何看待这个分歧的？阐述你的观点。
2. 假设在亚群中的一个等位基因比在所有群体中的一个等位基因拥有更多的频率，解释它是怎样影响为被告提供的证据是有利的还是不利的。假设在亚群中的一个等位基因比在所有群体中的一个等位基因拥有更少的频率，解释它是怎样影响为被告提供的证据是有利的还是不利的。
3. 血液测试的结果有 $1/200$ 的概率可以鉴别出嫌疑犯，DNA测试的结果作为证据指控嫌犯的概率通常是千万分之一。如果你是一个陪审员，你能接受把如此概率的证据作为鉴定嫌犯身份的有效证据吗？
4. 无罪工程利用DNA来试图澄清嫌犯不正当地被定罪，通过DNA检测，确立无罪是不是比被定罪更容易呢？说明原因。
5. 讨论DNA检测的其他一些用途，例如解决亲子关系问题。总的来说，你认为DNA鉴定在将来会对我们的社会产生多大的影响？

建议阅读

1. DNA指纹识别技术正在不断地被完善和更新，你可以在互联网上搜索到大量的信息。
2. 了解更多关于无罪计划（清白项目），请浏览<http://www.innocenceproject.org>。

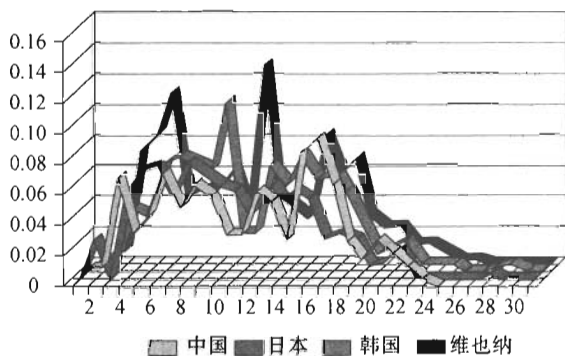


图6-19 四个不同种族人群一个等位基因的分组次数分布数据

资料来源：Kathryn Roeder, “DNA Fingerprinting: A Review of the Controversy,” *Statistical Science*, Vol. 9, No. 2, pp. 222-247.



知道“怎么办”的人总是有事可做。

知道“为什么”的人总是成为他的老板。

——丹妮·里威奇

第 7 章 相关和因果

学习目标

■ 7.1 寻求相关

能够掌握相关的定义，通过散点图认识正相关和负相关，理解作为相关强弱指标的相关系数。

■ 7.2 对相关的解释

了解对相关进行解释的注意事项，尤其是异常值效应和数据分组效应，更为至关重要的事实则是相关并不意味着因果。

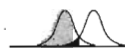
■ 7.3 最佳拟合线和回归

熟悉最佳拟合线的概念，认识这样的最佳拟合线何时具有预测价值，何时没有预测价值，理解相关系数的平方与拟合质量的关系，定性地理解多元回归的用途。

■ 7.4 寻找因果关系

理解从相关关系中建立因果关系是很困难的，调查指导原则可用于帮助我们建立可信的因果关系。

吸烟能导致肺癌吗？低失业率会导致通货膨胀吗？人类使用矿物燃料会使地球变暖吗？许多统计研究的一个主要目标，就是要寻找不同变量之间的相互关系，并使研究者能确定一个变量是不是影响另一个变量的原因。一旦发现了两个变量间的关系，我们就可以尝试去确定这里是否存在一个根本的原因。在这一章中，我们将研究相关关系，探讨相关的重要性，并且要面对更困难的寻找因果关系的任务。



7.1 寻求相关

当我们说吸烟导致肺癌的时候，它的意思指的是什么？如果你只抽一支雪茄，这并不意味着你会罹患肺癌，甚至它也并不意味着，即使你吸烟抽得很厉害，并且抽了很多年，这样你就会肯定罹患肺癌。因为某些严重的吸烟者并没有罹患肺癌。严格来说，这只是一个统计陈述，它的意思是指，如果你吸烟比起不吸烟

来，你罹患肺癌的可能性会更高些。

那么研究者是怎么知道吸烟会导致肺癌呢？这个过程开始于非正式的观察。比如，医生们会注意到，那些罹患肺癌的病人中，他们吸烟的比率高得出人意料，这些观察进一步引起严密的实验设计，研究者比较吸烟者和非吸烟者罹患肺癌的比率。这些研究明显表明严重的吸烟者更有可能罹患肺癌。用更为正式的概念，我们可以说，在吸烟的数量和罹患癌症的可能性这两个变量之间存在着相关（correlation）。相关是变量之间一种特殊类型的关系。也就是

说，当一个变量增长或者减小时，另外一个变量也会相应地增长或者减小。

下面是相关的几个例子：

- 在人们的身高和体重之间存在着相关，也就是说，身材较高的人们比起身材较矮的人们，体重会倾向于更重一些。
- 对苹果的需求和苹果的价格，这两个变量之间也存在着相关，也就是说，当价格上升的时候需求会下降。
- 对于钢琴练习者来说，在练习时间和技能水平之间这两个变量之间存在着相关，也就是说，那些练习时间更多的钢琴演奏者技巧可能会更高。

在两个变量之间存在相关，并不意味着一个变量的变化会引起另外一个变量的变化，认识到上面这一点是很重要的。在吸烟和肺癌之间的相关，它自身并不能证明吸烟能导致肺癌。我们可以假设一下，例如，具有某些遗传倾向的人，可能同时会吸烟并且罹患肺癌。无论如何，要想了解吸烟是否导致肺癌，确定它们之间的相关是至关重要的第一步。在这一章后面，我们将会讨论分析因果这个更难的任务。现在，我们将注意力集中在如何寻找、确定和解释相关上。

思考题

假如真的有一个基因，可以使得人们倾向于既吸烟又会罹患癌症，那么请解释为什么我们还要寻找肺癌和吸烟之间的高相关，但是却不能说吸烟导致肺癌。

相关知识

吸烟与许多严重疾病有关系，除了肺癌以外，也包括心脏病和肺气肿，吸烟也和非致命性健康状态有关，比如皮肤皱纹增加和勃起功能障碍。

7.1.1 散点图

表 7-1 列出了一家珠宝店的钻石样本的数据，包括钻石的价格和几个决定它们价格的常见指标。因为关于钻石的广告经常只是引证它们的重量（以克拉计），我们可以怀疑在钻石的价格和重量之间存在着相关。我们可以通过绘制一个关于钻石价格和重量的散点图（scatter diagram or scatterplot）来寻找它们之间的相关。

表 7-1 来自宝石经销商的 23 颗样本钻石价格和特征

钻石	价格 (美元)	重量 (克拉)	深度比	台宽比	颜色	净度
1	6 958	1.00	60.5	65	3	4
2	5 885	1.00	59.2	65	5	4
3	6 333	1.01	62.3	55	4	4
4	4 299	1.01	64.4	62	5	5
5	9 589	1.02	63.9	58	2	3
6	6 921	1.04	60.0	61	4	4
7	4 426	1.04	62.0	62	5	5
8	6 885	1.07	63.6	61	4	3
9	5 826	1.07	61.6	62	5	5
10	3 670	1.11	60.4	60	9	4
11	7 716	1.12	60.2	65	2	3
12	7 497	1.16	59.5	60	5	3
13	5 170	1.20	62.6	61	6	4
14	5 547	1.23	59.2	65	7	4
15	7 521	1.29	59.6	59	6	2
16	7 260	1.50	61.1	65	6	4
17	8 139	1.51	63.0	60	6	4
18	12 196	1.67	58.7	64	3	5
19	14 998	1.72	58.5	61	4	3
20	9 736	1.76	57.9	62	8	2
21	9 859	1.80	59.6	63	5	5
22	12 398	1.88	62.9	62	6	2
23	11 008	2.03	62.0	63	8	3

注：重量以克拉计（1 克拉 = 0.2 克）；深度比被定义为 100 倍的高径比；台宽比是指台面宽度与直径的百分比；颜色和净度以标准分计，1 是好的。对于颜色：1 = 无色，数字越大说明颜色越黄；对于净度，1 = 无瑕疵，6 说明有肉眼能见的缺陷。

相关知识

术语“克拉”也用来描述黄金（符号记作 K），但是和描述钻石（符号记作 c）的意思是不一样的。一克拉钻石等于 0.2 克，而用在黄金上却是纯度指标，24K 黄金是 100% 纯度，18K 黄金纯度是 75%，12K 黄金纯度则是 50%，依次类推。

定义

◆ 散点图（scatter diagram）是指每一个点代表两个不同变量的取值，这些点呈现在一起构成的图形。

下面的步骤描述了如何绘制图 7-1 这样的散点图。

(1) 我们将两个变量分别安排在纵坐标和横坐标轴上, 数值的范围要适合数据。有时, 坐标轴的选择是主观的, 但常常是一个变量有赖于另外一个变量, 这样我们就把解释变量放在横坐标轴上, 反应变量放在纵坐标轴上。在本例中, 我们设想钻石价格至少部分有赖于它的重量, 这样我们称重量是解释变量 (因为它有助于解释价格), 价格是反应变量 (因为它对解释变量的变化做出反应)。我们设计重量的坐标轴是从 0 ~ 2.5 克拉这样一个范围, 价格的坐标轴是从 0 ~ 16 000 美元这样一个范围。

(2) 对于表 7-1 中每一个钻石来说, 我们绘制一个点, 横坐标对应于它的重量, 纵坐标对应于它的价格。例如, 对第 10 号钻石, 那个点的位置在于横坐标是 1.11 克拉, 纵坐标是 3 670 美元。图 7-1 中的虚线显示了我们怎样定位这个点。

(3) (可选项) 我们可以标注其中一些或者全部的数据点, 如图 7-1 中所标注的钻石 10、16 和 19。

散点图因为这些点散落在图中揭示了变量之间的相关而得名。在图 7-1 中, 我们看到了一个上升的趋势, 说明钻石越重倾向于越贵。相关并不是完全的, 例如, 最重的钻石并不是最贵的, 但是总体趋势是非常清楚的。

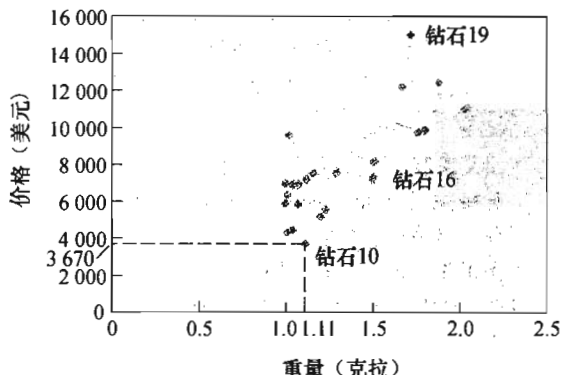


图 7-1 散点图呈现了表 7-1 中两个变量钻石价格和钻石重量的关系。虚线呈现了如何确定 10 号钻石点所在的位置

思考题

确定图 7-1 中代表 3 号钻石、7 号钻石和 23 号钻石的那些点。

例 7-1 颜色和价格

运用表 7-1 的数据制作一个散点图, 以寻找钻石颜色和价格之间的相关。对这个相关发表评论。

解答 我们期望价格有赖于颜色, 所以我们将解释变量颜色安排在横坐标上, 反应变量价格放在纵坐标上, 如图 7-2 所示。(你可以检查其中的几个点, 看看是否与表 7-1 中的数据符合。) 这个图中的点看起来要比图 7-1 中的点分布得更加散乱。尽管如此, 你还是可以注意到一个微弱的趋势, 从左到右钻石的价格是向下的, 这个趋势表明, 钻石的颜色越黄 (数字越高, 颜色越黄), 价格越低。这个趋势和我们所预期的是一致的, 因为透明的钻石看起来更耀眼, 通常也被看做更有价值。

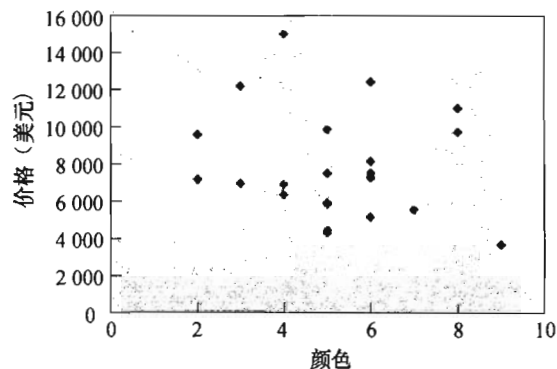


图 7-2 表 7-1 中关于钻石颜色和钻石价格两个变量的散点图

思考题

在工作中获得大额奖金是让人高兴的, 这样就有一笔 6 000 美元的开支可以买钻石戒指。售货员建议你这个价格有两种选择, 一个钻石重 1.20 克拉, 颜色值是 4; 另外一枚重 1.18 克拉, 颜色值是 3。假如这两枚钻石的其他特征都是相同的, 你将选择那一枚? 为什么?

7.1.2 相关的类型

让我们先看看两个相关的例子。图 7-1 呈现了重量和价格之间较强的相关, 图 7-2 呈现了颜色和价格之间较弱的相关。我们现在开始界定各种相关类型。图 7-3 显示了关于变量 x 、 y 的 8 种散点图, 注意这些散点图的以下特征:

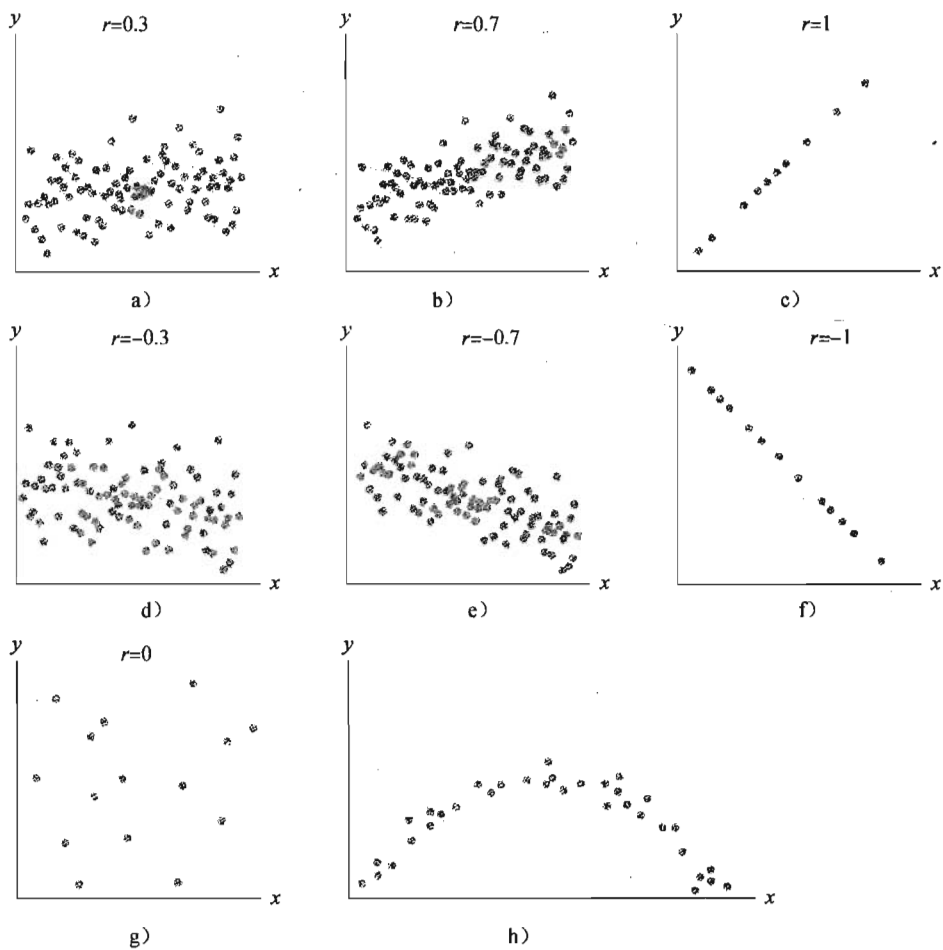


图 7-3 散点图所呈现的各种相关类型

- a) 变量 x 和 y 间弱的正相关
- b) 变量 x 和 y 间强的正相关
- c) 变量 x 和 y 间的完全正相关
- d) 变量 x 和 y 间弱的负相关
- e) 变量 x 和 y 间强的负相关
- f) 变量 x 和 y 间的完全负相关
- g) 变量 x 和 y 间的零相关
- h) 变量 x 和 y 间的非线性相关

- 图 7-3 中的 a、b、c 呈现了正相关。当 x 变量增长时， y 变量的值也趋向于增长。从 a 到 c 相关也越来越强，实际上 c 图呈现了一个完全正相关。在这种情况下，所有的点坐落在一条直线上。
- 图 7-3 中的 d、e、f 图呈现了负相关。当 x 变量的值增长时， y 变量的值倾向于下降。从 d 到 f 相关越来越强，f 图呈现了一个完全负相关。在这个图上，所有的点坐落在一条向下的直线上。
- 图 7-3 中的 g 图显示了变量 x 与 y 之间的无相关。换句话说， x 变量的取值与 y 变量的取值没有任何关系。
- 图 7-3 中的 h 图呈现了非线性相关。 x 和 y

之间的相关是存在的，但是这个相关并不对应于一条直线（线性意思是指一条直线，非线性的意思是指不是一条直线）。

相关类型

正相关 (positive correlations): 两个变量增长或下降的趋势是一样的。

负相关 (negative correlations): 两个变量变化的趋势是相反的，当一个增长时，另外一个下降的。

无相关 (no correlations): 在两个变量之间没有明显的线性关系。

非线性相关 (nonlinear relationship): 两个变量之间是有关系的，但是这种关系在散点图上并不表现为一条直线的形式。

方法注解

在本书中，术语“相关”只是指线性相关，一些统计学家将非线性关系称做“非线性相关”，有许多应用非线性相关的方法，与本书中所描述的线性相关方法是相同的。

例 7-2 预期寿命和婴儿出生率

图 7-4 呈现了 16 个国家的预期寿命和婴儿出生率这两个变量之间的散点图。它呈现的是一种什么类型的相关？这个相关有意义吗？它能用因果关系解释吗？说明原因。

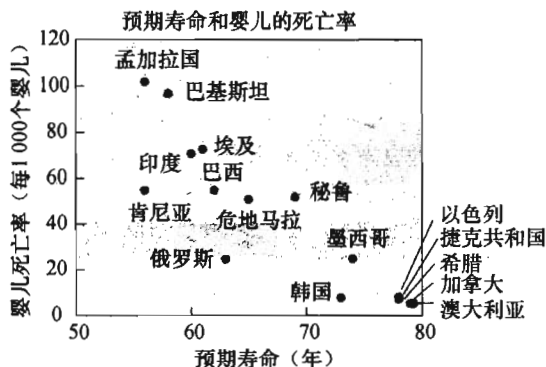


图 7-4 预期寿命和婴儿死亡率两个变量数据的散点图

资料来源：United Nations.

解答 散点图呈现了中等程度的负相关，在那些婴儿出生率较低的国家，倾向于有较高的预期寿命。它是一个负相关，因为两个变量变化方向相反。这个相关是有意义的，因为我们可以预想，那些具有较好健康设施的国家，同时会具有较高的预期寿命和较低的婴儿出生率。当然，它并不能解释婴儿出生率与预期寿命之间的因果关系，我们不能希望降低婴儿出生率能够显著地提高预期寿命，除非这两者都是提高健康护理水平的一部分。降低婴儿出生率将会提高预期寿命，因为婴儿死亡率下降会提高人口死亡的平均年龄。

7.1.3 相关强度的测量

在大多数情景下，表述一个强度的强、弱就可以了。然而，有时以更精确的方式表述相关强度是有必要的。用数字表示的相关强度的统计指标叫做相关系数 (correlation coefficient)，以字母 r 表示。

相关系数很容易解释，虽然它的计算是很繁琐的，使用计算器或计算机容易些。仔细

再看一下图 7-3，你可以注意到，每一个散点图都呈现了相关系数 r 的值。相关系数总是在 $-1 \sim 1$ 之间取值，散点图中的所有点越接近一条向上的直线，相关系数越接近 $+1$ ；同样，这些点越排成一条向下的直线，相关系数越接近 -1 。这些点如果不符合任何一条直线，或者越接近于横坐标的直线（标明 x 值与 y 值无关）相关系数越接近于 0 。

方法注解

对于这一节中的方法，有一个要求就是两个变量应该符合“双变量正态分布”。基本意思是指对于一个变量的定值，另一个变量对应的取值应该符合正态分布。这个要求通常是难以检测的，所以这个检测会降低要求，变成验证两个变量各自符合正态分布。

相关系数 r 的性质

1. 相关系数 r 是相关强度的指标，它的取值是 $-1 \sim 1$ 。
2. 如果没有相关，这些点不适合任何一条向上或向下的直线， r 的值接近于 0 。
3. 如果存在正相关，相关系数是正数 ($0 < r \leq 1$)：两个变量一起增长。完全正相关（在散点图上所有点排成一条向上的直线）的相关系数 $r = 1$ 。 r 的值越接近于 1 ，意味着相关的强度越高， r 值越接近于 0 ，说明相关强度越弱。
4. 如果存在负相关，相关系数是负数 ($-1 < r \leq 0$)。当一个变量增长时，另一个变量下降。完全负相关（在散点图上，所有点排成一条向下的直线）的相关系数 $r = -1$ 。 r 的值越接近于 -1 ，意味着是强的负相关， r 值越接近于 0 ，说明是弱的负相关。

相关知识

1900 年，超过 40% 的美国人口在农场工作，到 2000 年，只有不到 2% 的人在农场工作了。

例 7-3 美国的农场大小

图 7-5 呈现了美国的农场数量和农场大小这两个变量的散点图。每一个点代表 1950 ~ 2000 年之间每一年的数据值。在这个散点图中，过去的年份通常在右边，现在的年份通常在左边。将本图与图 7-3 进行比较，估计其相关系数，并且讨论相关背后的原因。

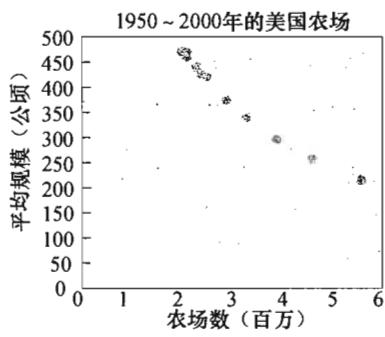


图 7-5 农场大小数据的散点图
资料来源：U. S. Department of Agriculture.

解答 这个散点图显示了一个强的负相关。这个图与图 7-3 中的散点图非常接近，相关系数大约是 $r = -0.9$ 。

这个相关说明，当农场数量下降时，农场的大小在增长。这个趋势反映了美国农场的—个基本变化，1950 年，大多数农场是小型农场，到现在这些小农场逐渐被农业公司所拥有的大型农场所替代。

例 7-4 天气预报的准确性

图 7-6 的散点图呈现了两周以来的数据，比较了每天的实际高温和同一天预报以及 3 天前预报的数据。估计两种数据类型的相关系数，并且讨论这些相关系数说明了天气预报的什么问题。

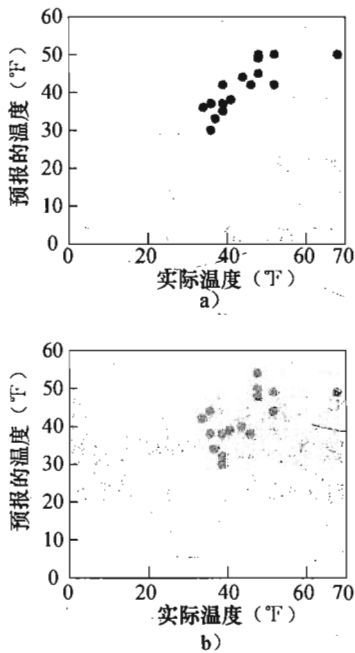


图 7-6 实际最高温度和预测最高温度的比较
a) 当天预测 b) 3 天前预测

解答 如果每次天气预报都是准确的，每一次实

际的温度将会对应于预报温度，结果就是所有点位于一条直线，并且相关系数 $r = 1$ 。在图 7-6a 中，天气预报是由每天开始做出的，这些点比较接近于一条直线，意味着当天的预报与实际的温度更一致。通过与图 7-3 中的散点图相比较，我们可以合理地估计，这个相关系数大约是 $r = 0.8$ 。图 7-6b 中的相关是很弱的，说明三天前的天气预报不能和当天的天气预报那样与实际温度一致，这个相关系数大约是 $r = 0.6$ 。这个结果并不奇怪，因为我们预期的是长时间的天气预报并不是很准确。

相关知识

以百分比来讲，历史上最赚钱的电影是《女巫布莱尔》，是由来自中佛罗里达大学电影项目组的两个朋友花了 35 000 美元，电影票房超过了 140 亿美元，每投入 1 美元就会获得 4 000 美元的收益。

例 7-5 成功的电影

表 7-2 显示了 15 部制作成本和票房收入巨大的科幻电影。制作一个有关制作成本和票房收入的散点图，估计相关系数，并讨论它的意义（票房收入是指这部电影在所有剧院上映的总票房收入）。

表 7-2 科幻电影的最大预算		
电影	电影制作成本 (百万美元)	票房收入 (百万美元)
《金刚》(2005)	207	218
《蜘蛛侠 2》(2004)	200	373
《纳尼亚传奇》(2005)	180	292
《水世界》(1995)	175	88
《范海辛》(2004)	170	120
《极地快车》(2004)	170	172
《终结者》(2003)	170	150
《波塞冬》(2006)	160	52
《蝙蝠侠：侠影之谜》(2005)	150	205
《哈利波特与火焰杯》(2005)	150	290
《世界末日》(1998)	140	201
《星际战警》(2002)	140	190
《蜘蛛侠》(2002)	139	403
《太空战士之灭绝光年》(2001)	137	32
《绿巨人》(2003)	137	132

解答 图 7-7 显示了在散点图中，横坐标代表制作成本，纵坐标代表票房收入（你可以检查其中几个点看是否符合表 7-2 中的数据）。这个散点图与表

示无相关的图 7-3g 非常相似,所以我们可以说这个相关系数接近于 $r=0$ 。换句话说,至少对于这些电影而言,在制作电影花费多少钱的成本和这些电影最后票房收入多少钱之间没有相关。

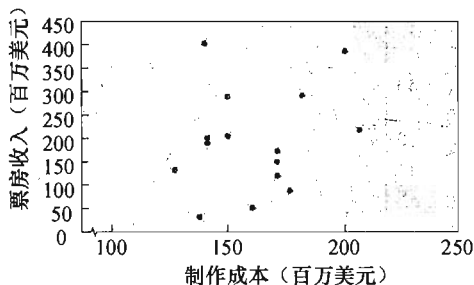


图 7-7 表 7-2 中数据的散点图

思考题

更进一步的练习,直观估计钻石重量和价格(见图 7-1)之间的相关系数,以及钻石颜色和价格(见图 7-2)之间的相关系数。

7.1.4 相关系数的计算(可选节)

线性的相关系数 r 可以用以下几种方式表示,并且它们是等价的,计算的结果也都相同。下面公式更符合相关系数 r 所表示的潜在意义。

$$r = \frac{\sum \left[\frac{(x - \bar{x})(y - \bar{y})}{S_x S_y} \right]}{n - 1} \quad (7-1)$$

式中,被 $n-1$ 除表明 r 是某一种类型的平均数,由于它包含许多对数据值,因此它不会轻易增长。符号 S_x 表示 x 变量的标准差,符号 S_y 表示 y 变量的标准差。在公式中 $(x - \bar{x}) / S_x$ 的含义是标准分数(见 5.2 节)。通过应用 x 和 y 变量的标准分数, r 取值就不会轻易变化。理解相关系数的关键在于 x 变量标准分数和 y 变量的标准分数的乘积。当它们之间是正相关的时候这个乘积是正数;当它们之间是负相关的时候这个乘积是负数;当它们之间没有相关的时候,这个乘积也可能是正数也可能是负数,但是这个值会非常接近零。

下面一个相关系数 r 的可选公式可用做简化计算之用,经常用在手工计算的时候,该公式也可以用于统计软件和计算器的编程。

$$r = \frac{n \times \sum(xy) - (\sum x) \times (\sum y)}{\sqrt{n \times (\sum x^2) - (\sum x)^2} \times \sqrt{n \times (\sum y^2) - (\sum y)^2}} \quad (7-2)$$

这个公式至少在理论上是可以直接使用的:首先计算每一个所需值的总和,然后将这些值带进公式。注意公式中的 $\sum x^2$ 和 $(\sum x)^2$ 是不相等的: $\sum x^2$ 告诉你先平方后累加, $(\sum x)^2$ 告诉你先累加以后再平方。换句话说,先对括号内的公式进行操作。在对另外一个变量 $\sum y^2$ 和 $(\sum y)^2$ 进行操作时也是如此。

7.1 练习题

统计素养和批判性思维

1. 相关 什么是相关?统计学上的术语“相关”与日常生活中的意思一样吗?
2. 相关 在对 5 对数据计算相关系数以后,你推断说它们没有相关。如果在对获得的 100 对以上的类似数据进行计算还能得到一个没有相关的结论吗?为什么?
3. 散点图 如果计算出的相关系数是 $r=0.997$,描述在对应的散点图中点的模式。
4. 叠合相关 一项研究显示人均葡萄酒消费和教师的收入呈显著相关。推断教师在葡萄酒上花费他们的额外收入显然是荒谬的。(为什么?)因此要确定至少一个另外的因素来解释这个相关。

判断正误: 判断练习题 5~8 中的陈述是有

- 意义的(表述正确)还是没有意义的(表述错误),并加以解释。并非所有的陈述都有明确的答案,所以你的解释比判断更重要。
5. 出生人数 一项研究显示,在一个城镇里,当鹤鸟的数量增加时,城镇里的出生人数也在增长。因此,鹤鸟的数量增长导致了出生人数的增长。
6. 阳性效应 一位研究者计划调查不同国家婴儿死亡率的下降和人均收入的关系。因为婴儿死亡率的下降是一个好的效应(或者叫阳性效应),我们知道在不同国家婴儿死亡率和人均收入之间存在一个正相关。
7. 相关 两项研究都发现了在出生体重和免疫功能低下之间的一个相关,第二项研究的样本容量较大,所以说得到的相关一定会更大。
8. 解释相关系数 r 一个研究者使用 20 对数据

计算线性相关系数的值，得到结果是 $r=1.2$ 。

概念与应用

相关类型：练习题9~16列出了成对变量。对于每一对变量，陈述你是否相信这两个变量存在相关。如果你认为它们是相关的，表述这个相关是正的还是负的。说明你的推论。

- 9. 身高与体重 随机选择的年龄在1~21岁之间的50位女性身高和体重。
- 10. 学习时间和成绩 学生学习一门历史课程所花费的时间和他们在这门课程测验中的成绩。
- 11. 重量和燃料消耗 汽车的重量和他们的燃料消耗率（以每加仑行驶的公里数作为测量指标）。
- 12. 智商和帽子大小的尺寸 随机选择的成人的智商分数和他们所戴的帽子大小尺寸。
- 13. 马拉松比赛 马拉松比赛完成时间（以秒计）和最后的比赛名次（第一、第二等）。
- 14. 高度和温度 飞机的飞行高度和外面的空气温度。
- 15. 身高和SAT测验分数 从参加SAT测验的人们中随机选择的被试的身高和SAT测验分数。
- 16. 高尔夫比赛分数和奖金 职业高尔夫选手参加高尔夫比赛分数和他们所获得的奖金。
- 17. 全世界的肉类和粮食生产 图7-8呈现了全世界的肉类生产和谷物生产关系的散点图，都是以每人千克作为测量指标。这些数据点对应于1950~2000年之间的11个不同的年份。估计它们之间的相关并且讨论这个相关背后的原因。

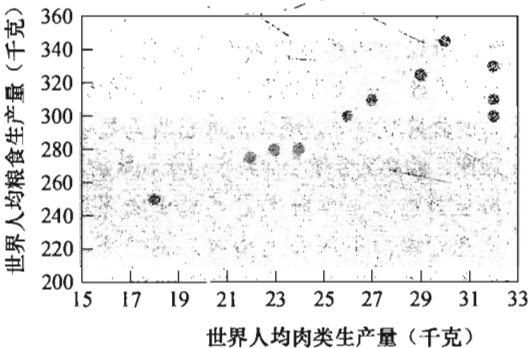


图7-8 11年中肉类和粮食生产量的散点图
资料来源：U. S. Department of Agriculture.

- 18. 两天前的天气预报 图7-9呈现了当天实际的最高温度与两天前的预报最高温度相比较的散点图。估计相关系数并且讨论这些数据

说明了天气预报的什么问题。你认为如果制作另外一个两周的类似数据的散点图，会得到相同的结果吗？为什么？

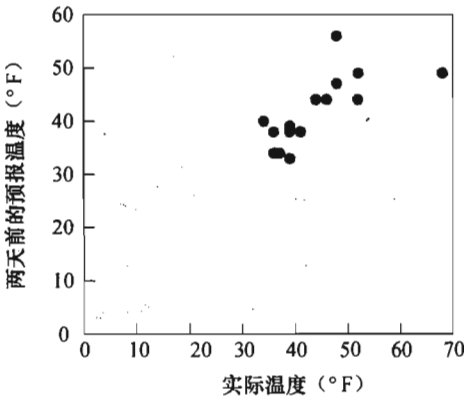


图7-9 天气预报的散点图

- 19. 安全速度 思考下表，下表呈现了20世纪80年代所选择的一些国家交通事故的速度限制和死亡率。

国家	死亡率 (每1亿交通里程数)	速度限制 (公里/小时)
挪威	3.0	55
美国	3.3	55
芬兰	3.4	55
英国	3.5	70
丹麦	4.1	55
加拿大	4.3	60
日本	4.7	55
澳大利亚	4.9	65
荷兰	5.1	60
意大利	6.1	75

资料来源：D. J. Rivkin, *New York Times*.

- (1) 绘制关于上表数据的一个散点图。
 - (2) 使用简练的语言概括相关的特征（例如，强的正相关，弱的负相关），并且估计这些数据的相关系数（或者使用计算器和软件准确地计算这些数据的相关系数）。
 - (3) 在新闻报道中，这些数据是一篇名叫“每小时55公里的速度限制并不是安全的保障”的文章提供的。基于上述数据，你同意这个结论吗？请加以解释。
- 20. 人口增长 思考下表，下表呈现了在10年的一段时期内10个州的人口变化比率和出生率（每千人人口）。

州	人口变化百分数	出生率
内华达	50.1%	16.3
加利福尼亚	25.7%	16.9
新罕布什尔	20.5%	12.5
犹他	17.9%	21.0
科罗拉多	14.0%	14.6
明尼苏达	7.3%	13.7
蒙大拿	1.6%	12.3
伊利诺伊	0%	15.5
艾奥瓦	-4.7%	13.0
西弗吉尼亚	-8.0%	11.4

资料来源: U. S. Census Bureau and Department of Health and Human Services.

- (1) 绘制关于上表数据的一个散点图。
- (2) 使用简练的语言概括相关的特征, 并且估计这些数据的相关系数。
- (3) 总的来说, 出生率可以很好地预测一个州的人口增长率吗? 如果不可以的话, 什么样的其他因素会影响人口增长率?

21. 最有价值的球员 思考下表, 下表呈现了最有价值的棒球运动员所完成的本垒打和安打率 (NL = 全美棒球联盟, AL = 美国棒球联赛)

球员	本垒打	安打率
杰夫·肯特 (2 000NL)	33	0.334
杰森·昂比 (2 000AL)	43	0.333
贝瑞·邦兹 (2 001NL)	73	0.328
铃木一郎 (2 001AL)	8	0.350
贝瑞·邦兹 (2 002NL)	46	0.370
米格尔·特哈达 (2 002AL)	34	0.308
贝瑞·邦兹 (2 003NL)	45	0.341
亚历克斯·罗德里格斯 (2 003AL)	47	0.298
贝瑞·邦兹 (2 004NL)	45	0.362
弗拉基米尔·葛雷洛 (2 004AL)	39	0.337
艾伯特·普荷斯 (2 005NL)	41	0.330
亚历克斯·罗德里格斯 (2 005AL)	48	0.321
莱恩·霍华德 (2 006NL)	58	0.313
贾斯汀·莫洛 (2 006AL)	34	0.321

- (1) 绘制这些数据的一个散点图。
- (2) 用简略的语言概括相关的特征, 并且

估算相关系数。

- (3) 高安打率是本垒打的一个很好的指标吗? 请加以解释。

22. 电影数据 思考下表, 其所呈现的 1990 ~ 2002 年之间全部美国电影总的票房收入和观众人数。

年份	总收益 (10 亿美元)	观众总数 (10 亿)
1990	5.0	1.18
1991	4.8	1.14
1992	4.9	1.17
1993	5.2	1.24
1994	5.4	1.29
1995	5.5	1.26
1996	5.9	1.34
1997	6.4	1.39
1998	7.0	1.48
1999	7.5	1.47
2000	7.7	1.42
2001	8.4	1.49
2002	9.5	1.64

资料来源: Motion Picture Association of America.

- (1) 绘制数据的散点图。
- (2) 用简略的语言概括相关的特征, 并且估算相关系数。
- (3) 自 1990 年以来, 是什么原因影响了持续增长的电影价格? 如果你是一个电影院经理人, 这些数据所得到的一般结论对你有什么重要意义?

23. 电视时间 思考下表, 5 类不同年收入家庭观看电视的平均时间。

家庭收入 (美元)	每周观看电视时间
少于 30 000	56.3
30 000 ~ 40 000	51.0
40 000 ~ 50 000	50.5
50 000 ~ 60 000	49.7
60 000 以上	48.7

资料来源: Nielsen Media Research.

- (1) 绘制数据的散点图。为了定位这些点, 采用每一种收入分组的中点, 使用 25 000 美元这个值代表少于 30 000 美元这一组, 使用 70 000 美元这个值代表超过 60 000 美元这一组。

- (2) 用简略的语言概括相关的特征，并且估算相关系数。
- (3) 推论为什么家庭收入高的观看电视时间少？你认为这些数据是否意味着你可以通过少看电视来提高你的收入？请说明理由。

24. 一月份的天气 考虑下表所呈现的 10 个北半球的城市一月份的平均降水量和平均每日最高温度（联邦海洋和气候委员会）。

城市	一月份平均 每日最高温度 (°F)	一月份平均 降水量 (英寸)
雅典	54	2.2
孟买	88	0.1
哥本哈根	36	1.6
耶路撒冷	55	5.1
伦敦	44	2.0
蒙特利尔	21	3.8
奥斯陆	30	1.7
罗马	54	3.3
东京	47	1.9
维也纳	34	1.5

资料来源：The New York Times Almanac.

- (1) 绘制这个数据的散点图。
 - (2) 用简略的语言概括相关的特征，并且估算相关系数。
 - (3) 从一月份的温度和降水量这些数据中，你能得出一些一般性的结论吗？请加解释。
25. 零售商 思考下表所列出的美国 8 家大零售商一年总的销售额和利润。

公司	总销售额 (10 亿美元)	利润 (10 亿美元)
沃尔玛	315.6	11.2
克罗格	60.6	0.98
家得宝	81.5	5.8
好市多	60.1	1.1
塔吉特	52.6	2.4
星巴克	7.8	0.6
盖普	16.0	1.1
百思买	30.8	1.1

资料来源：Fortune. com.

- (1) 绘制这个数据的散点图。
- (2) 用简略的语言概括相关的特征，并且估算相关系数。
- (3) 讨论你的观察，高销售额一定能转化

为高利润吗？为什么？

26. 卡路里和婴儿死亡率 考虑下表所列出的 10 个国家平均每日卡路里摄入量（所有居民）和婴儿的死亡率（每 1 000 个出生的婴儿）。

国家	平均每日 消耗卡路里	婴儿死亡率 (每 1 000 个出生婴儿)
阿富汗	1 523	154
澳大利亚	3 495	6
布隆迪	1 941	114
哥伦比亚	2 678	24
埃塞俄比亚	1 610	107
德国	3 443	6
利比里亚	1 640	153
新西兰	3 362	7
土耳其	3 429	44
美国	3 671	7

- (1) 绘制这些数据的散点图。
- (2) 用简略的语言概括相关的特征，并且估算相关系数。
- (3) 讨论你所观察到的以及从中得出的一般结论。

相关系数的性质：对于练习题 27 和 28 判断所给出的属性是否正确，并且解释你的答案。

- 27. 相互交换的变量 如果我们相互交换变量 x 和 y ，相关系数保持不变。
- 28. 改变测量单位 如果我们分别或一起改变变量 x 、 y 的测量单位，相关系数保持不变。

网络和远程项目

请在网站 [www. aw. com/bbt](http://www.aw.com/bbt) 上打开第 7 章的有用链接“Links for internet projects”。

- 29. 失业率和通货膨胀 登录美国劳工部的网页，寻找过去一年的每月失业率和通货膨胀率的数据，绘制上述数据的散点图，你会发现什么趋势？
- 30. NFL 的成功率 寻找上赛季 NFL 球队的统计数据，编制一个表呈现每一个队的下述数据：赢球的场次、在每场比赛中的平均进攻得分以及在每场比赛中的防守环节让对方的得分。编制一个散点图，呈现赢球和进攻的相关以及赢球和防守的相关。讨论你的发现，你认为在其他球队的统计中也能发现赢球场次之间的强相关吗？
- 31. 统计摘要 浏览美国统计摘要的次数分布表网

页,选择你感兴趣的数据并且计算至少两个相关。简要讨论你从这些相关中所掌握的信息。

32. 身高和臂长 选择至少8个人的样本,测量每个人的身高和臂长(当你测量臂长时,每个人应该伸直双臂,就像飞机展翅一样)。使用这些成对样本的数据编制一个散点图,并且计算相关系数的值,你会得到什么结论?
33. 身高和心率 选择至少8个人的样本,通过对一分钟内的心跳进行计数的方法记录每个人的心率,也记录每个人的身高。使用这些成对样本的数据编制一个散点图,并且计算相关系数的值,你会得到什么结论?

新闻中的统计学

34. 新闻中的相关 寻找近期的一则新闻报道,并且讨论其中所涉及某种类型的相关,描述这种相关。文章说明了这种相关强度的意义吗?它会提示相关会反映一些潜在的因果吗?简要讨论你是否相信文章中对于相关的暗示?
35. 你自身的正相关 举例说明你所希望的两个变量具有正相关,说明为什么这两个变量是相关的,并且说明为什么相关是重要的。
36. 你自身的负相关 举例说明你所期望是负相关的两个变量,解释为什么这两个变量是相关的,并且说明为什么相关是重要的。



7.2 对相关的解释

研究者通过仔细分析统计数据通常可以得出相关是有意义的。一个新的令人惊异的相关发现经常会招致大量的新闻报道。你仔细回忆一下你所听到的一些相关发现:经常食用燕麦皮与心脏疾病的风险降低是相关的,移动电话的使用与增加交通事故的风险是相关的,食量少与长寿是相关的。可惜的是,解释这样的相关的任务要比发现这些相关还要难。在新闻报告的热潮消退以后很久,我们仍然不能确定这些相关是否显著,也就是说,这些相关能否告诉我们一些实践价值。在这一节中,我们将讨论与相关解释有关的一些常见难题。

7.2.1 注意异常值

仔细检查一下图7-10中的散点图,你会发现,当 x 值增大时 y 值也随之增大,这是一个正相关。的确,如果计算这些数据的相关系数,你会发现这些相关系数高达 $r=0.880$,这是一个强相关。

然而,如果用你的拇指遮挡住图7-10右上角的数据点,很明显的相关消失不见了。实际上,没有这个数据点,这个相关系数是0。换句话说,删除掉这个数据值会让相关系数从 $r=0.880$ 变化到 $r=0$ 。这个例子告诉我们相关对于异常值是非常敏感的。回顾一下异常值,它是一个与这一组数据中其他数据值相比较非常极端的一个数据值(见4.1节)。在解释一个相关以前,我们要仔细地检查一下异常值和它的影

响。一方面,如果这个异常值是数据组中的一个错误,它能导致一个明显的但实际上并不存在的相关,或者它会掩盖一个实际上存在的相关。另一方面,如果异常值确实是实际存在着的真实的数据值,它也会告诉我们在其他时候很难发现的一个相关。

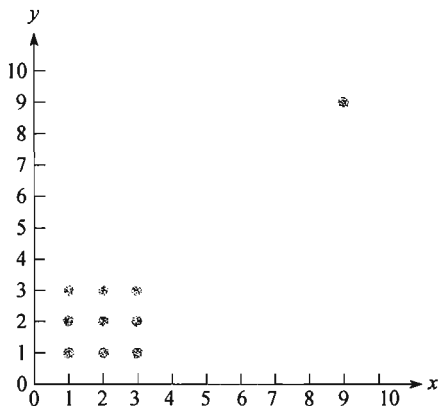


图7-10 异常值如何影响相关

要注意,当我们仔细地检查异常值的时候,除非我们有很明确的理由说明这个数据不属于这一组数据,否则我们是不能删除异常值的。即使在这种情况下,良好的研究原则也需要我们报告这些异常值,说明我们删除这些异常值是合理的。

例7-6 被掩盖的相关

你实施了一项研究来判定一个人一天消费的卡路里数量及其花费在骑自行车上的时间之间的相关。样本包括10位女性骑自行车者,这些人的体重和身高都非常接近。在两周的时间内,你要求每一个妇女记录她每天花费在骑自行车上的时间,以及她每天吃多少

东西。你用她的食品消费记录计算每天的卡路里。图 7-11 呈现了这个散点图，横坐标是每一天花费在骑自行车上的时间，纵坐标是每一天卡路里的摄入量。骑自行车时间愈高是否对应的卡路里摄入量也越高？

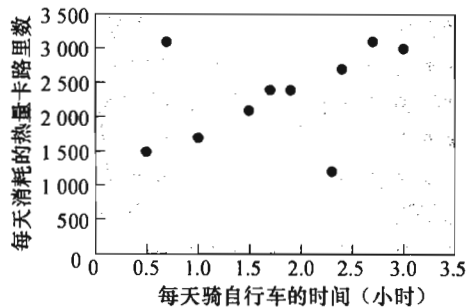


图 7-11 自行车运动数据的散点图

解答 如果你全部地看一下数据，你会发现这里有一个正相关，骑自行车时间越长也倾向于摄入更多的卡路里。但是这些相关很弱，相关系数 $r=0.374$ 。然而，注意两个异常值，一个代表每天只骑行半小时的自行车但摄入超过 3 000 卡路里的热量；另一个代表每天骑行两小时自行车但是摄入卡路里仅仅是 1 200。这两个异常值是难以解释的，已知样本中 10 位女性的身高和体重是相近的。我们可以怀疑这两位女性数据记录是错误的，或者说在这两周时间内并不是她们通常的习惯。如果我们能够证实这种怀疑，我们就有理由删除这两个无效数据。图 7-12 呈现了删除这两个异常值之后的相关，相关是非常明显的，表明每小时的自行车骑行时间大约对应于 500 卡路里的摄入。当然，如果不能证实我们的怀疑，它们仍然是有效的数据，我们不能删除它们，但是我们应该报告不考虑它们的原因。

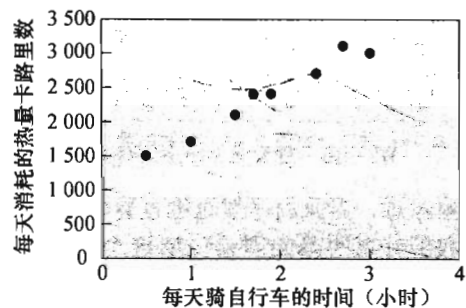


图 7-12 图 7-11 数据去掉两个异常值后的散点图

7.2.2 注意不正确的分组

当数据分组不正确时，相关也会被错误地解释。在有些情况下，分组数据掩盖了相关。考虑一个研究（虚拟），研究者发现了每周观看电视的时间和高中生会考成绩的相关。在表 7-3

中一共收集了 21 对数据。

表 7-3 观看电视时间和高中生 GPA 成绩（虚拟数据）

每周观看电视时间（小时）	GPA 成绩
2	3.2
4	3.0
4	3.1
5	2.5
5	2.9
5	3.0
6	2.5
7	2.7
7	2.8
8	2.7
9	2.5
9	2.9
10	3.4
12	3.6
12	2.5
14	3.5
14	2.3
15	3.7
16	2.0
20	3.6
20	1.9

图 7-13 的散点图显示是没有相关，数据所计算出的相关系数 $r = -0.063$ ，表面上的结论是电视观看习惯与学业成就没有相关。然而，一个机敏的研究者认识到一些学生大多数时候看的是教育节目，另外一些学生更多地观看戏剧和电影。她将这些数据分成两组，一组学生大多数时候观看的是教育节目，其他同学是另一组。表 7-14 呈现了她将这些学生分成两组的研究结果。

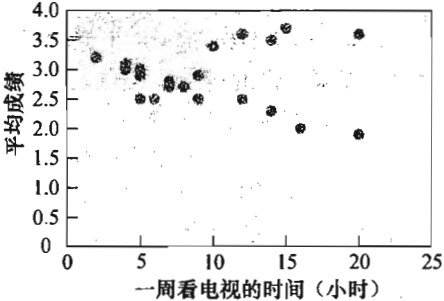


图 7-13 观看电视时间和 GPA 成绩的全部数据直观显示没有相关

表 7-4 观看电视时间和高中生 GPA 成绩分组数据（虚拟数据）

组 1：观看教育节目		组 2：观看平常节目	
每周观看电视时间	GPA 成绩	每周观看电视时间	GPA 成绩
5	2.5	2	3.2
7	2.8	4	3.0
8	2.7	4	3.1
9	2.9	5	2.9
10	3.4	5	3.0
12	3.6	6	2.5
14	3.5	7	2.7
15	3.7	9	2.5
20	3.6	12	2.5
		14	2.3
		16	2.0
		20	1.9

现在我们可以发现非常强的相关是存在的（见图 7-14），收看教育节目的一组同学是强的正相关（ $r=0.855$ ），收看其他节目的一组同学是强负相关（ $r=-0.951$ ）。这个故事的意义在于原始数据隐藏了电视和学业成绩之间的重要相关。收看教育节目和学业成绩是正相关，收看非教育节目和学业成绩与是负相关。只有这组数据被恰当地分成两组，这个相关才能被发现。

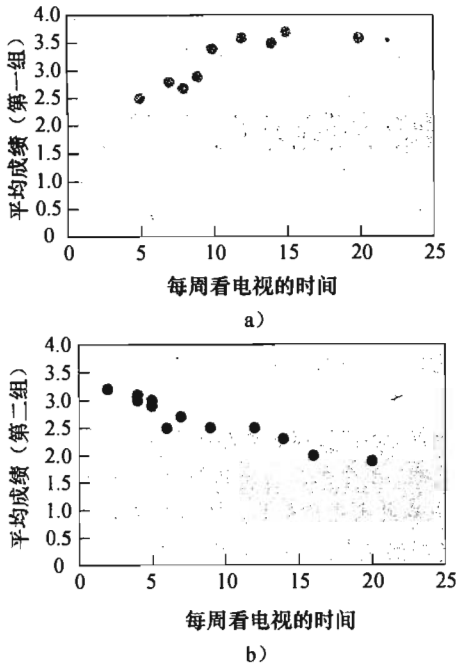


图 7-14 和图 7-13 一样数据的散点图，只是数据按照表 7-4 分为两组

在其他一些情况中，一组数据整体会呈现出比分组更强的相关，思考一下表 7-5 中的数据，这

是搜集的一组消费者的数据，用来研究汽车重量和价格之间相关的。图 7-15 呈现了这个散点图。

全部的数据显示了强相关，相关系数 $r=0.949$ ，然而，更进一步检验发现，数据落入两个差别很大的组分别对应于轻型和重型汽车，如果我们将其分组进行分析，则两组均不显示相关，轻型汽车（表 7-5 中上面 6 个数据）的相关系数是 $r=0.019$ ，重型汽车（表 7-5 中下面 6 个数据）的相关系数是 $r=-0.022$ 。你可以通过观看图 7-15 发现这个问题。全部数据具有明显的相关，由于这两组数据是分离的，但是每一组数据内部并没有相关。

■ 相关知识

2~5 岁的儿童平均每周看电视 26 小时；6~11 岁的儿童平均每周看电视 20 小时（尼尔森媒体研究）；成人电视观众平均每周看电视超过 25 小时。如果用每小时 8 美元薪水的工作来代替看电视，他每年的收入将提高超过 10 000 美元。

表 7-5 汽车重量和价格（虚拟数据）

重量（磅）	价格（美元）
1 500	9 500
1 600	8 000
1 700	8 200
1 750	9 500
1 800	9 200
1 800	8 700
3 000	29 000
3 500	25 000
3 700	27 000
4 000	31 000
3 600	25 000
3 200	30 000

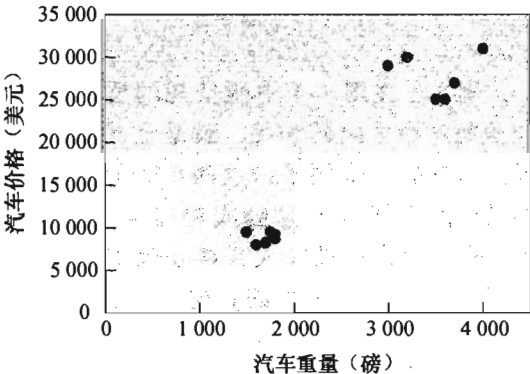


图 7-15 表 7-5 汽车重量和价格数据的散点图

思考题

假如你准备购买一辆小型车。如果你仅仅看图 7-18 中的全部数据和相关系数。你认为汽车的重量是一个对价格有重要影响的因素吗？如果你再看一下轻型车和重型车分开的数据，你会怎么想？说明原因。

案例研究

审查相关

牛津大学的物理学家理查德·佩托在英格兰医学杂志 *Lancet* 上发表了一篇文章，他提出心脏病罹患者如果在心脏病发作的几小时内服用阿司匹林会增加其生存的机会。*Lancet* 的编辑让理查德·佩托将数据分成几个小组，以观察阿司匹林对于不同组罹患者是否作用不同。例如，阿司匹林是否对某一年龄组或者某种饮食习惯的罹患者更有效？

将全部数据分组可以揭示一些重要的因素，比如按照男性女性来分组，可以发现他们对于治疗的反应有什么不同。然而，理查德·佩托认为编辑将他的样本分成了太多的组。于是他拒绝了这个要求，他分辩说这样分组会导致一些纯粹的巧合。《华盛顿邮报》报道这个故事的记者瑞克·韦斯说，“当编辑坚持要这样分组时，佩托妥协了。但是除了编辑所要求的分组之外，他将他的罹患者也按照星座进行分组，并且提出要求说他的所有发现都要发表。今天，像是对所有统计外行的警告信号，可以看到下面令人感到怪异的数字：阿司匹林除了对于双子座和天秤座的心脏病罹患者没有用之外，对于其他星座的人都可以增加生存的机会。”

这个故事的寓意在于对相关的“非法调查”经常导致这样的结果，它并不能使相关更有意义，即使按照标准的统计测量的标准，这些相关是非常显著的。

7.2.3 相关并不意味着因果

也许在说明相关的时候，最重要的警示是我们早已强调过的：相关并不一定意味着因果。通常情况下，出现相关可能是下面三个原因中的任何一个。

相关的可能解释

- (1) 相关可能是一种巧合。
- (2) 两个相关变量可能直接被一些常见的基本因素所影响。
- (3) 相关的两个变量，其中一个变量确实是另外一个变量的原因。但是要注意，即使在这种情况下，它也可能只是几种原因之一。

例如，在图 7-4 中所呈现的婴儿死亡率和预期寿命之间的相关，就是受共同的潜在因素影响的一个例子：这两个变量同时都受到另外一个潜在变量健康护理水平的影响。吸烟和肺癌之间的相关反映了吸烟导致肺癌的事实（参考 7.4 节的讨论）。巧合的相关也是很常见的，下面的例 7-7 将讨论这样一种情况。

许多统计研究的设计是要发现因果关系的

事实，关注因果是尤其重要的。因为这些研究通常是开始于相关的寻找，一旦发现相关，工作也就结束了，这样想来研究是充满吸引力的。可是，当读到 7.4 节讨论因果时，你会发现建立一个因果关系是非常困难的。

例 7-7 如何在股票市场变得更富有

每位财经顾问都有一套预测股票市场发展趋势的策略。大多数人会关注基础经济数据，诸如利率和公司利润。但另外一个可选择的策略是依靠一个令人关注的著名相关，即每年一月份超级碗的获得者和这一年来下时间内股票市场的发展趋势之间的相关。如果超级碗的赢家来自 1970 年之前旧的美国职业橄榄球联盟，股票市场的趋势是上涨的；如果超级碗的赢家不是来自老美的职业橄榄球联盟，股票市场的趋势是下跌的。这个相关与最初的 40 届超级碗比赛中的 31 届比赛结果是匹配的。成功率是 77.5%，远大于预期的偶然结果。2007 年第 41 届超级碗比赛的获胜者，印第安纳波利斯小马队，是 1970 年之前旧的美国职业橄榄球联盟的一支球队。基于这一事实，在 2007 年超级碗比赛之后，你应该将你所拥有的闲置资金（甚至包括你所借的）投资于股票市场吗？

解答 基于所报告的相关，在旧的美国职业橄榄球联盟球队获得冠军以后，在这一年的余下时间股票市场的趋势是上涨的，你可能尝试投资（失利的球队是芝加哥熊队，也是一支美国职业篮球联盟的老球

队，所以无论谁最后获胜，按照超级碗的相关预测2007年的股票市场是上涨的)。然而，如果你相信超级碗的结果实际上导致了股票市场的变化方向，那么

这种投资是没有意义的。非常清楚这是荒谬的，这个相关毫无疑问是一个巧合。如果你打算投资，是不能根据这种相关来投资的。

案例研究

燕麦麸和心脏病

如果你购买一种含有燕麦麸的产品，那将是一个宣传吃燕麦麸的健康效应的好机会。的确，一些研究已经发现吃比较多的燕麦麸与罹患心脏病的低发生率这两者之间存在相关。但这是否意味着每个人应该吃更多的燕麦麸呢？

答案是不必要的，这只是因为，燕麦麸消费与心脏病降低之间存在相关并不意味着它可以降低心脏病的风险。实际上，这个案例中的因果问题是相当有争议的。另外一些研究表明吃较多燕麦麸的人，他们的饮食结构更倾向于健康的。这样，燕麦麸消费和心脏病的风险降低可能只是许多共同潜在因素中的一种：健康饮食习惯导致人们消费更多的燕麦麸，这样也降低了患心脏病的风险。在那种情况下，对一些人来说，在饮食中的增加一些燕麦麸并不是一个好主意，因为它可能会使人们体重增加，而体重增加和心脏病的风险增加是有联系的。

这例子告诉我们，当我们思考相关和因果问题时，应用因果关系的重要性，在医学研究者确切地知道在你的饮食中增加燕麦麸是否可以降低心脏病的发生风险，还有一段长期的研究道路要走。

7.2.4 相关用途的说明

在讨论相关的用途时可能会导致错误的解释，我们已经描述了异常值的作用，不恰当地分组，相关的非法调查，关于相关意味着因果

的不当结论。但是关于相关仍然还有许多有用的和正确的解释。例如，在7.1节中有关如何用相关来确定钻石价格。在另外的应用中，相关也可以用来建构人口数量和作为垃圾被弃置的塑料之间的关系。相关还可以用来建构加州黄石公园老实泉的持续喷射和两次喷射的时间间隔的关系。通常，相关在下述大量的领域中起到了显著和重要的作用，包括气象学、医学研究、商业、经济、市场研究、广告、心理学和计算机科学等。

7.2 练习题

统计素养和批判性思维

1. 相关和因果 为了解释相关，理解没有潜在的因果关系的相关是非常重要的。当我们说“相关并不意味着因果”这时候我们的意思是什么？
2. 婴儿猝死综合征 在《纽约时报》发表的一篇关于婴儿死亡率的文章包括这样的陈述：基于研究结果，让婴儿仰面向上睡觉的姿势可以降低婴儿猝死综合征的机会？这个陈述有什么错误吗？
3. 异常值 什么是异常值？异常值怎样影响相关的结果？
4. 散点图 什么是散点图？它是如何应用的？

判断正误：判断练习题5~8中的陈述是有意义的（表述正确）还是没有意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. 汽车重量和燃料消耗量 基于一个研究所呈

现的一辆汽车的重量和它的燃料消耗率之间的显著相关。我们可以得出这样的结论，由于汽车重量（磅）增加的原因导致它的燃料消耗率（每加仑的公里数）下降。

6. 运动和健康 研究显示运动和健康之间存在着相关。但是相关并不意味着因果。这样，我们不应该运动，因为这样做并不能导致增进健康的结果。
7. 饮酒和驾车 因为研究显示，在喝酒和交通事故之间存在着相关，我们可以得出结论说，喝酒是交通事故的原因。
8. 异常值 已知20对数据，假如我们增加另外一对包含异常值的数据，异常值对相关系数的影响将会非常小，因为21对数据中只有一对异常值。

概念与应用

相关和因果：练习题9~16是对相关的陈述，在每一道题中，明确地表述相关（例如，

在变量 A 和变量 B 之间存在一个正相关)。说明这个相关是更像是巧合呢，还是可能存在潜在的因果关系，或者是直接的因果关系？解释你的回答。

- 9. 枪支和犯罪率 在某一个州里，在过去几年中注册的持枪数量稳定地增长，同时犯罪率也在增长。
- 10. 跑步和体重 研究发现那些通过跑步进行有规律的运动的人们，他们的体重要小于不跑步的人群。那些进行长跑的人们比起跑步距离较短人群体重要轻。
- 11. 高速费和距离 当一名司机在马萨诸塞州收费高速公路上驾车行驶距离越远，所交的高速费用越高。
- 12. 交通工具和等候时间 研究发现随着注册的交通工具有所增长，驾驶员在交通中所花费的等候时间也有所增长。
- 13. 交通信号灯和交通事故 研究发现随着交通信号灯的增加，汽车之间的碰撞事故也在增加。
- 14. 星系 天文学家发现，除了离我们较近的一些星系之外，宇宙中的所有星系正在离我们越来越远。并且，离我们越远的星系，远离我们的速度就越快。也就是说，距离我们越远的一个星系，它离开我们的速度就越快。
- 15. 汽油和驾驶 研究发现随着汽油价格的的增长，交通工具行驶的距离变短了。
- 16. 黑素瘤和纬度 一些研究显示，对于一些人种来说，随着纬度的下降，黑素瘤（最危险的一种皮肤癌）的发生率增加了。
- 17. 异常值效应 思考图 7-16 中的散点图。

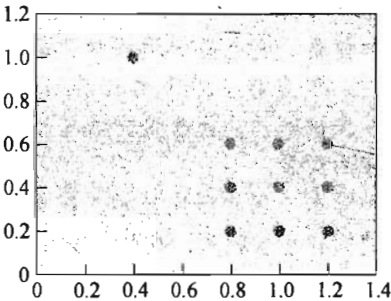


图 7-16 散点图

- (1) 哪一个点是异常值？忽略这个异常值，估计或者计算其余点值的相关系数。
- (2) 现在包括这个异常值，这个异常值对相关系数会产生怎样的影响？估计或

者计算全部数据的相关系数。

- 18. 异常值效应 思考图 7-17 中的散点图。

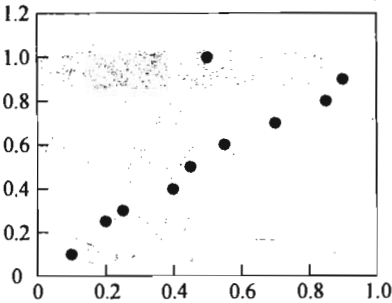


图 7-17 散点图

- (1) 哪一个点是异常值？忽略这个异常值，估计或者计算其余点值的相关系数。
 - (2) 现在包括这个异常值，这个异常值对相关系数会产生怎样的影响？估计或者计算全部数据的相关系数。
- 19. 分组的鞋码数据 下表给出了 10 个人的体重和鞋码数据（包括男性和女性）。
 - (1) 绘制这些数据的数据的散点图，估计或者计算它们的相关系数。基于这个相关系数，关于鞋码和体重的关系你会得到什么样的结论。请加以解释。
 - (2) 你后来了解到表中的前 5 位数据是女性，后 5 位数据是男性。关于鞋码和体重关系的观点你会有什么样的改变？你仍然会推论鞋码和体重是相关的吗？

体重（磅）	鞋码
105	6
112	4.5
115	6
123	5
135	6
155	10
165	11
170	9
180	10
190	12

- 20. 分组的温度数据 下表呈现了世界上 10 个主要城市的一月份的平均最高温度和七月份的平均最高温度。

城市	1 月份最高温度	7 月份最高温度
柏林	35	74
日内瓦	39	77
喀布尔	36	92
蒙特利尔	21	78
布拉格	34	74
奥克兰	73	56
布宜诺斯艾利斯	85	57
悉尼	78	60
圣地亚哥	85	59
墨尔本	78	56

(1) 绘制这些数据的散点图。估计或者计算它们的相关系数。基于这个相关系数，你可以得出结论说在这些城市中1月份和7月份温度是相关的吗？请加以解释。

(2) 请注意，表中的前5个城市是属于北半球的，后5个城市是属于南半球的。这会改变你对于上面相关的观点吗？你现在对这些城市1月份和7月份的温度是否有关是怎样认为的？请加以解释。

21. 出生率和死亡率 图 7-18 呈现了不同国家的出生率和死亡率，出生率和死亡率都是以每 1000 人来作为测量指标的。

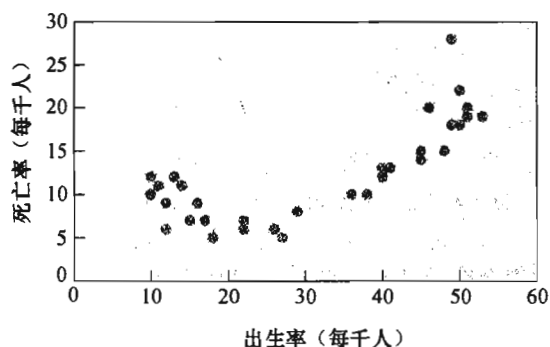


图 7-18 不同国家的出生率和死亡率

资料来源：United Nations.

(1) 估计相关系数并且讨论在两个变量之间是否存在强相关。

(2) 请注意全部数据看起来是由两组数据点组成的。作为一个合理的猜测，这些数据组是怎样构成呢？在哪一组中你可能会找到像瑞典这样相对富裕的

国家？在哪一组中你可能会找到向安哥拉这样的不发达国家？

- (3) 假如你在第 2 小题中的猜测是正确的，在不同组内会有明显的相关吗？请加以解释。你怎样证实你关于分组的猜测。

22. 阅读和测验分数 下表（拟定的）的数据给出了 10 名六年级学生每周阅读的时间，以及他们在一个标准化的语言测验中的分数（百分制）。

每周阅读时间	语言测验成绩
1	50
1	65
2	56
3	62
3	65
4	60
5	75
6	50
10	88
12	38

(1) 绘制这些数据，估计或者计算它们的相关系数。基于计算的相关系数，你可以得出结论说阅读时间和测验分数是相关的吗？请加以解释。

(2) 假如你了解到其中 5 名儿童仅阅读漫画，而另外 5 名儿童会阅读一般书籍。你猜测这些数据点分别会落在哪一组？你怎样证实你的关于分组的猜测。

(3) 假如你在第 2 小题中的猜测是正确的。你会改变关于阅读时间和测验分数之间的相关吗？请加以解释。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 7 章的有用链接“links for internet projects”。

23. 橄榄球股票数据下载 寻找近年来与超级碗比赛的获胜者有关的股票市场的年终变化（上涨或者下跌）。最近的结果仍然支持例 7-7 中所描述的相关吗？请加以解释。

24. 现实中的相关

(1) 描述一种真实的情境，在这种情况下存在着的一种正相关是巧合的结果。

(2) 描述一种真实的情境，在这种情况下存在着的一种正相关可能是潜在因果

的结果。

- (3) 描述一种真实的情境，在这种情况下存在着的一种正相关可能是直接原因的结果。
- (4) 描述一种真实的情境，在这种情况下存在着的一种负相关是巧合的结果。
- (5) 描述一种真实的情境，在这种情况下存在着的一种负相关可能是潜在因果的结果。
- (6) 描述一种真实的情境，在这种情况下存在着的一种正相关可能是直接原因

的结果。

新闻中的统计学

25. 错误解释的相关 寻找一篇近期的新闻报道，你认为其中的相关可能是解释错误的。描述这个相关，报道中对于相关的解释，以及你所发现解释中存在的问题。
26. 准确解释的相关 寻找一篇近期的新闻报道，你认为其中对相关的解释是合理的。描述这个相关，报道中对于相关的解释，以及你为什么说这个解释是有效的原因。



7.3 最佳拟合线和回归

假如你非常幸运地在一次比赛中赢得了1.5克拉的钻石，基于图7-1中所显示的重量和价格之间的相关，预测钻石的大概价格。我们只需要仔细地研究一下图，1.5克拉对应于哪一点是最恰当的。为了做到这些，我们需要通过这些数据绘制一条**最优拟合线**（也叫做回归线），如图7-19所示。这条线是按照统计测量标准（稍后我们简短讨论），感觉上最适合的，我们可以通过这些数据绘制一条直线，比起其他任何一条直线，这些数据点与这条直线的距离是最接近的。

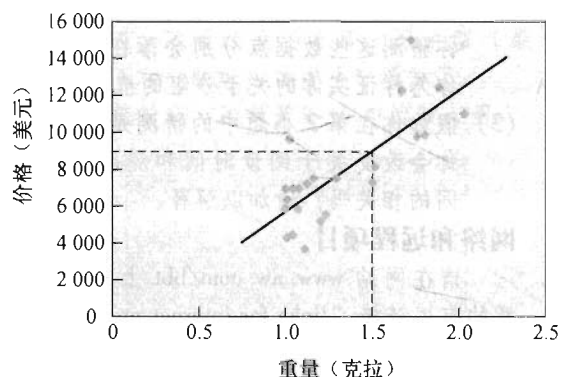


图7-19 图7-1数据的最佳拟合线

定义

- ◆ **最佳拟合线 (best-fit line)**，或者叫回归线，是指在散点图中有一条直线，所有数据点与该直线的距离是最接近的（按照统计测量的接近标准）。

相关知识

术语“回归”来自1877年弗朗西斯·高尔顿的研究，他发现父亲身材过高或过低男孩的身高比其父亲更接近平均水平。因此他说儿童的身高向平均水平回归，我们就得到“回归”的概念，现在即使数据并没有向平均水平回归的趋势，我们也使用这个概念。



在一个散点图中所有可能的直线都可以被绘制出来，您怎么知道哪一条是最佳拟合线？在许多情况下，您可以通过观看数据做一个最优拟合线的估计，然后穿过所有数据点直观形象地绘制出这条直线。这种绘制最优拟合线的方法是“目测”完成的。正如你所猜测的那样，还有一个精确计算最优拟合线的方法（参考这一节最后的可选择主题），许多计算机程序和计算器可以自动地完成这些计算。对我们这本书的目的来说，通过目测直观地绘制一条最优拟合线就可以了。

7.3.1 应用最佳拟合线进行预测

我们可以用图7-19中的最佳拟合线来预测

1.5 克拉的钻石价格。如图中虚线所标志的最优拟合线预测钻石的价格大约为 9 000 美元。注意,在图中对应的两粒 1.5 克拉钻石的实际价格都少于 9 000 美元。这样,虽然 9 000 美元的预测价格听起来是合理的,但它的确并不能保证。在实际中,这个案例中数据点的散落水平告诉我们,我们不能相信最优拟合线所预测的每粒钻石价格都是准确的。相反,这一预测在统计学意义上是有价值的,它告诉我们,如果我们检验非常多的 1.5 克拉的钻石,他们的平均价格大约在 9 000 美元。

这仅仅是几个解释应用最佳拟合线进行预测的重要注意事项之一。第二个注意事项是强调采用最佳拟合线进行预测必须建立在大量有效的数据基础之上。图 7-20 呈现了图 7-4 中婴儿死亡率和预期寿命之间相关的最佳拟合线。依照这条直线,预期寿命超过 80 岁的国家,婴儿死亡率将是一个负数,这是不可能的。

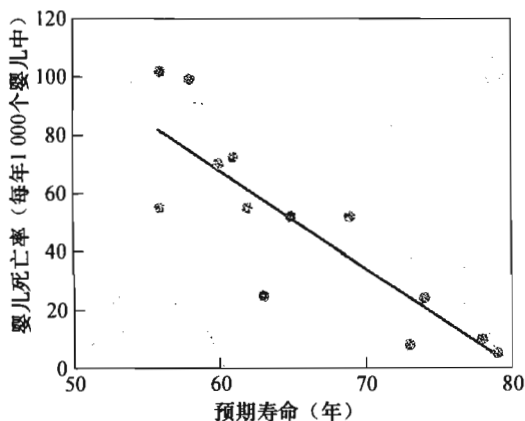


图 7-20 图 7-4 婴儿死亡率和寿命之间相关数据的最佳拟合线

资料来源: United Nations.

第三个注意事项是,避免使用过去老数据的最佳拟合线来预测现在或者将来的结果。例如,经济学家研究历史数据发现,在失业率和通货膨胀率之间的负相关。依照这个相关,当失业率下降至 6% 的时候,通货膨胀率应该有个急剧上升。但是通货膨胀仍然保持较低的水平,显示过去数据所呈现的相关并不适用于现在。

第四个注意事项是,从某一特定总体中抽出的一个样本所发现的相关通常并不是用来预

测别的总体。例如,在一个关于阿司匹林的摄入量和心脏病之间相关的实验中,由于全部被试均为男性,我们不能期望这个研究结果也适用于女性。

第五个注意事项是,要记住即使我们能穿越所有数值点绘制一条最优拟合线,但如果相关是不显著的或者相关是非线性的,那么这条直线也是没有意义的。例如,在鞋码和智商之间没有相关,所以我们就不能用鞋码去预测智商。

应用最佳拟合线进行预测时的注意事项

(1) 不要希望最佳拟合线能给出一个最好的预测,除非相关是非常强的,并且有很多数据点。如果样本的数据点与最佳拟合线非常接近,相关非常强这样预测也就比较准确。如果样本的数据点与最佳拟合线距离比较远,相关较弱预测也就相对不太准确。

(2) 不要应用最佳拟合线去预测那些超出数据范围的值。

(3) 基于过去数据的一条最佳拟合线对现在和将来结果的预测都是无效的。

(4) 不要对与样本所在总体不一样的群体进行预测。

(5) 要记住当相关是不显著的或者相关是非线性时最佳拟合线是没有意义的。

例 7-8 有效的预测?

陈述下面每一个案例中的预测是否值得信任,并解释原因。

(1) 在每天人们锻炼几小时和每天人们消耗的卡路里数量之间的相关基础上,建立了一条最优拟合线。你使用相关预测到一个每天锻炼 18 小时的人大概每天要消耗 15 000 卡路里。

(2) 在 SAT 分数和大学成绩之间有一个著名的但是较弱的相关。你使用此相关依据你的朋友在 SAT 测验中的分数去预测他的大学成绩。

(3) 历史数据显示国家人口出生率和富裕水平之间有一个强的负相关。那就是,富裕水平越高的国家倾向于较低的人口出生率。

(4) 在中国的一项研究发现,在设计中国儿童喜欢的博物馆展览中相关是有用的,博物馆馆长建议采用这个信息设计一个针对亚特兰大地区学生的博物馆展览。

(5) 科学研究显示,儿童的铅摄入量 and 心理发育迟滞值之间存在一个强相关。基于这个相关,含有铅的绘画颜料被禁止了。

(6) 基于一个庞大的数据组,你绘制了一个关于

洋葱消费（每人）和受教育年限关系的散点图。散点图显示没有显著相关，但你还是绘制了一条最佳拟合线。这条直线预测每周消费1品脱洋葱的人，他的受教育年限至少为13年。

解答

- (1) 没有人可以持续不断地每天锻炼18小时，所以这么多的锻炼时间一定超出了所搜集的数据范围。这样，关于一个人每天锻炼18小时的预测将是不可信的。
- (2) 在SAT分数和大学成绩之间弱相关的事实意味着数据中点都是散落的。从结果来看，我们不能期望我们使用这个弱相关对某一个体的预测是准确的。
- (3) 我们不能自行假定这些历史数据仍然适用于今天。实际上，今天的俄罗斯有非常低的人口出生率，但是它的富裕水平也并不高。
- (4) 将在中国的研究所得到的信息用于亚特兰大地区的展览，这就说明将中国研究中所得到的相关用于对亚特兰大进行预测。然而，我们知道中国和亚特兰大地区存在很大的文化差异。所以如果没有更多其他信息支持的话，博物馆馆长的建议不应该被考虑。
- (5) 由于严重的后果和强相关，这个预测和禁令看起来是非常合理的。实际上，后来的研究也支持铅是影响心理发育迟滞这个实际存在的因素，这使得这个禁令更合理。
- (6) 因为两者没有显著相关，所以最佳拟合线和任何预测都是没有意义的。

相关知识

在美国，1978年铅被禁止用于房屋涂料，1991年被禁止用于饮料罐，从1995年开始25年之内要逐步停止使用含铅汽油。然而，1997年的疾病健康控制中心的一份报告估计，仍然有100万6岁以下的儿童血液中含铅过高，足于伤害他们的健康，不断发生的铅风险包括旧房屋的涂料、接近主干道的汽油燃料所排放的铅污染。

例7-9 女性会比男性走得更快吗

图7-21显示了在一公里竞走比赛中男性和女性世界纪录数据和最佳拟合线。基于这些数据，可以预测未来女性的世界纪录相比男性的世界纪录更快。你

同意这个预测吗？

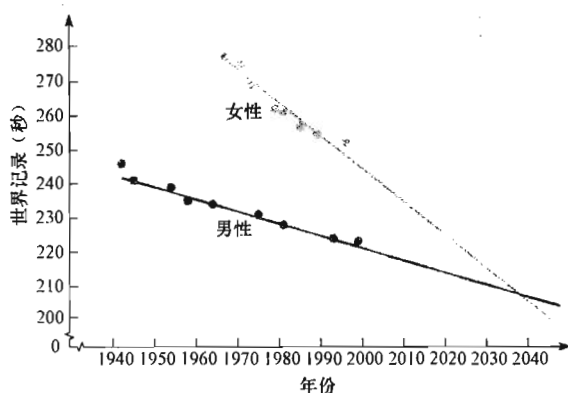


图7-21 男性和女性一公里竞走比赛的世界纪录

解答 如果我们接受所绘制的最佳拟合线，到2040年，女性的世界纪录将与男性的世界纪录相同。然而，这不是一个有效的预测，因为它是基于最优拟合线超出实际数据范围的一个扩展。到2040年，女性的世界纪录将会超过男性，这当然不是毫无疑问的。如果按照同样的推理，男性和女性的世界纪录最终都会等于零，这将意味着当发令枪一开始有些人就会完成比赛。

7.3.2 相关系数和最优拟合线

前面，我们讨论相关系数作为测量相关强度的一种方法。我们也可以用相关系数来说明最佳拟合线进行预测的效度。

由于数学的原因（不在本书中进行讨论），相关系数的平方（或者 r^2 ）是最优拟合线所解释的一个变量变化的比例（或者，更专业地说，最优拟合线所表达的线性关系）。例如，钻石重量和价格数据之间的相关系数（见图7-19）结果是 $r=0.777$ 。如果我们将这个相关系数值平方，得到结果是 $r^2=0.604$ ，我们可以做出如下解释：钻石价格变化中的大约0.6或者60%可以由钻石重量和价格之间的最优拟合线来解释。余下的40%价格的变化可以归因为其他因素，例如深度、色彩、明亮度等，这是为什么图7-19的最优拟合线的预测不是那么精确。

只有在一个完全相关的最优拟合线中才可以给出精确的预测（ $r=1$ 或者 $r=-1$ ），这时计算得 $r^2=1$ ，也就是意味着变量的100%变化可以被最优拟合线所解释。在这种特殊情况下，预测将是完全正确的，除非样本数据并不是总体数据的一个典型代表。

最优拟合线和 r^2

相关系数的平方, 或者 r^2 , 是指最优拟合线所能解释变量的变化比率。

方法注解

统计学经常将 r^2 称做决定系数。

例 7-10 零售企业的招聘

你是一个大型零售店的经理。在这些年里, 你发现在9月份销售量和节日季期间为提高效率而招聘的雇员数量存在强相关, 相关系数是0.950。这一一年里的9月份销售量也是相当的高。基于这个最佳拟合线, 你应该开始打广告招聘员工吗?

解答 在这个案例中, 我们计算 $r^2 = 0.950^2 = 0.903$, 这意味着员工人数峰值90%的变化可以由与9月份销售量线性相关来解释。不能解释的员工人数峰值的变化只余下10%。因为90%是相当高的数字, 我们可以下结论说最优拟合线解释这个数据相当好, 用它来预测这一年中, 节假日所需要的员工数量, 是一个很好的主意。

例 7-11 投票率和失业人数

政治科学家非常想知道是什么因素影响选举中的投票率。其中一个因素是失业率。自从1964年在总统选举年这个数据被采集后发现, 在投票率和失业率之间存在一个较弱的负相关。相关系数大约是 $r = -0.1$ (见图7-22)。基于这个相关, 我们可以用失业率来预测下一年的总统选举中的投票率吗?

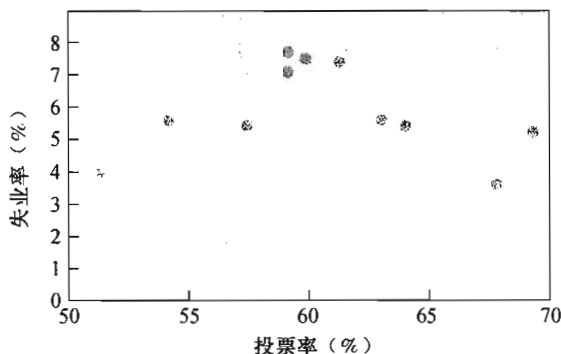


图 7-22 1964 ~ 2004 年美国失业率与投票率数据
资料来源: U. S. Bureau of Labor Statistics.

解答 相关系数的平方是 $r^2 = (-0.1)^2 = 0.01$, 这意味着数据中只有1%的变化是可以由最优拟合线来解释的。数据中几乎所有的变化都是由其他因素来解释的。我们可以得出结论失业率不是投票率可信的预测。

7.3.3 多元回归

如果你曾经购买了一颗钻石, 你可能会惊奇地发现在钻石颜色和价格之间存在如图7-2中的一个弱相关。当然, 如果一个钻石的颜色质量非常差的话价格是不大可能较高的。也许颜色可以帮助解释为什么重量和价格之间的相关并不完全。例如颜色之间的差异可以解释为什么两个相同重量的钻石会有不同的价格。为了检验这个想法, 寻找价格和其他一些变量比如重量和色彩之间的联合相关, 这将是一个好主意。

思考题

检查表7-1中的观点。注意, 例如, 钻石4和钻石5的重量比较接近, 但是钻石4价值仅仅只有4 299美元, 而钻石5却价值9 589美元。它们之间的颜色差异可以解释价格的不同吗? 研究一下表7-1中的其他两个钻石重量相同但是价格不同的例子。总的来说, 你认为如果我们用重量和颜色一起来代替其中单独某个与价格之间的相关, 是否会更强一些? 请说明理由。

有一个方法可以研究一个变量 (比如价格) 和两个或更多联合变量 (比如重量和颜色) 之间的相关, 这个技术叫做多元回归 (multiple regression), 这可以方便我们计算三个或更多变量之间相关的最佳拟合方程。因为它包括超过两个以上的变量, 我们不能绘制一个简单的图表来显示多元回归的最佳拟合方程。然而, 计算这些数据是否适合线性方程的指标仍然是可能的。多元回归中最常见的指标是决定系数, 符号表示为 R^2 , 它告诉我们数据中的这些散点有多少可以被最佳拟合方程解释。如果决定系数 R^2 接近1, 在数据取值范围内用最佳拟合方程来作预测将是非常有用的。如果决定系数 R^2 接近于零, 那么最佳拟合方式所作的预测基本是没有意义的。

定义

◆ 多元回归 (multiple regression) 的用途是通过最优拟合方程的计算, 表达一个变量 (比如价格) 和两个或多个以上的联合变量 (比如重量和颜色) 之间最适合的关系。决定系数 R^2 告诉我们, 数据中可以被最佳拟合方程所解释的散点比率。

在本书中, 我们不再描述通过多元回归计算最佳拟合方程的方法, 然而你可以使用决定系数

R^2 去解释多元回归的结果。例如，价格与重量和颜色两者的联合变量的相关是决定系数 $R^2 = 0.79$ 。这要比我们先计算的价格和重量之间单独的决定系数 $r^2 = 0.61$ 要高。研究钻石价格变化的统计学家知道，他们可以再增加另外一个的变量（比如，深度、清晰度等）使得这个相关变得更强。我们知道每年在钻石上的花费超过数十亿美元，你就能确定统计学家在帮助珠宝商认识珠宝的最大价值中发挥了显著的作用。

例 7-12 校友的贡献

你受雇于你的大学校友联合会，帮助他们合理预计联合会能够获得的新基金增长动力是多少？主管建议使用过去的贡献和校友收入水平之间的相关数据，它们之间的相关系数是 0.6。这是一个好策略吗？你能否提出一个更好的方案？

解答 相关系数 $r = 0.6$ 意味着决定系数 $r^2 = 0.36$ 。所以在贡献和校友收入之间的相关仅能解释 36% 的变化。这样，使用最佳拟合线预测每个校友捐赠钱款的数目将不是一个非常准确的预测。更好的策略是使用多元回归方程，将其他可能影响捐赠的因素，诸如校友的专业、校友现在的居中住地距离大学的远近、毕业年限、是否为校友联合会的成员等。比起仅基于两个变量的相关，恰当的多元回归方程将出现与捐赠更强的相关。■

相关知识

一个关于校友捐款的研究发现，要建立一个多元回归方程，应该用收入、年龄、婚姻状况、兄弟联谊会成员还是妇女联谊会成员、在校友会事务中是否活跃以及全国失业率等变量用作经济指标。（Bruggink and Siddiqui, "An Econometric Model of Alumni Giving: A Case Study for a Liberal Arts College," The American Economist, Vol. 39, No. 2).

7.3.4 最佳拟合线的计算公式(可选节)

最佳拟合线方程的数学计算公式是基于下列基本观念的。如果我们在散点图上任意绘制一条直线，我们可以测量每一个点到直线之间的垂直距离。这条直线是否反映这些数据点的测量指标就是这些点和直线间垂直距离的平方和。这些数据点到直线间的垂直距离的平方和越大，说明这条直线越不适合，这些数据点到某条直线间较小的平方和说明这条直线拟合得比较好。在所有可能性的直线中，最佳拟合线是一条所有数据点到它的垂直距离平方和最小的直线。因为这个特征，最佳拟合线有时又叫做最小平方线。

你也可以将任何一条直线方程写成如下通常的形式：

$$y = mx + b \quad (7-3)$$

m 代表直线的斜率， b 代表直线的截距

$$\text{斜率} = m = r(s_y/s_x) \quad (7-4)$$

$$\text{截距} = b = \bar{y} - (m \times \bar{x}) \quad (7-5)$$

在上述公式中， r 是相关系数， s_x 代表 x 变量的标准差， s_y 代表 y 变量的标准差， $\bar{x}$ 代表 x 变量的平均数， $\bar{y}$ 代表 y 变量的平均数。因为这类公式如果手工计算是非常繁琐的，因此我们通常使用计算器和计算机计算斜率和截距。

当使用软件和计算机计算最佳拟合线的斜率和截距时，结果通常表达为下面的公式：

$$y = b_0 + b_1x \quad (7-6)$$

b_0 是斜率， b_1 是截距，我们要仔细地区分这两个变量。

7.3 练习题

统计素养和批判性思维

1. 最佳拟合线 什么是最佳拟合线（也叫回归线）？最佳拟合线的用途是什么？
2. 多元回归 什么是多元回归？
3. r^2 符号 r^2 代表着什么？对其如何解释？也就是说，它的取值能告诉我们关于变量的什么方面？
4. R^2 符号 R^2 代表什么？它的取值能告诉我们关于变量的什么方面？它的取值接近于 1 时

告诉我们数据的什么方面？

判断正误：判断练习题 5~8 中的陈述是有意义的（表述正确）还是没有意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. r^2 值 一个男性样本 $r^2 = 0.007$ ，每一对数据包括一位男性以英寸为单位的身高和以厘米为单位的身高。
6. r^2 值 一个男性样本 $r^2 = -0.040$ ，每一对数

据包括一位男性的以英寸为单位的身高和 SAT 测验成绩。

7. 身高和体重 数据来源于全美健康调查, 所计算出女性身高和体重的最佳拟合线方程。通过这个方程, 一个身高为 120 英寸的女性身高可预测她的体重是 430 磅。
8. 加州老实泉 成对的样本数据包括加州老实泉的喷射时间(秒)和喷射之间的时间间距(分钟)。计算出 $r^2 = 0.926$ 。这说明通过最优拟合线方程, 大约 93% 的喷射之间的时间间距变化可以用两个变量之间的相关所预测。

概念与应用

散点图上的最优拟合线: 对于练习题 19 ~

12, 完成以下任务。

- (1) 在给出的散点图上绘制一条最优拟合线。
 - (2) 估计或者计算 r 和 r^2 , 基于你计算出的 r^2 值, 确定有多少比重的变量的变化可以用最优拟合线解释。
 - (3) 简要讨论你使用这个最优拟合线方程进行预测是否有效。
9. 应用图 7-2 中的颜色和价格的散点图。
 10. 应用图 7-4 中的预期寿命和婴儿死亡率的散点图。
 11. 应用图 7-5 中的农场数量和农场大小的散点图。
 12. 应用图 7-6 中实际的和天气预报温度的散点图。

最优拟合线: 参考 7.1 节中的表, 对于练习题 13 ~ 20, 完成以下任务。

- (1) 绘制成一个散点图, 基于直观的数据, 采用目测的方法绘制一条最优拟合线。
- (2) 简要讨论相关的强度。估计或者计算 r 和 r^2 , 基于你所计算的 r^2 值, 说明最优拟合线可以解释的变量变化比率。
- (3) 识别散点图中的异常值, 并且讨论它们对相关强弱和最优拟合线的影响。
- (4) 在每道题中, 你认为最优拟合线对超出散点图中数据范围之外的变量的预测可信吗? 请加以解释。

13. 应用 7.1 节中练习题 19 的数据。
14. 应用 7.1 节中练习题 20 的数据。
15. 应用 7.1 节中练习题 21 的数据。
16. 应用 7.1 节中练习题 22 的数据。
17. 应用 7.1 节中练习题 23 的数据。为了定位这些点, 用每一个收入范围的组中值来做代表值: 采用 25 000 美元这个值代表“少于 30 000 美元”这一组, 使用 70 000 美元这个

值代表“超过 60 000 美元”这一组。

18. 应用 7.1 节中练习题 24 的数据。
19. 应用 7.1 节中练习题 25 的数据。
20. 应用 7.1 节中练习题 26 的数据。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 7 章的有用链接“links for internet projects”。

21. 铅中毒 研究铅中毒及其来源和影响。讨论能够帮助研究者理解铅中毒的相关。讨论如何预防铅中毒。
22. 石棉 研究石棉及其来源和影响。讨论能够帮助研究者理解暴露在石棉中影响健康的相关。讨论如何预防这些对健康不利的因素。
23. 世界人口预测

下表给出了挑选出来的 11 个国家的五个人口指标。研究这些数据并找出其中可能存在相关。如果需要的话, 可以作进一步的研究。讨论你所发现的可能相关, 推测这些相关的原因, 讨论是否存在因果关系。出生率和死亡率是每 1 000 人的数据, 生育率是每个女性的数据。

国家	出生率	死亡率	预期寿命	城镇人口百分比	生育率
阿富汗	50	22	43	20	6.9
阿根廷	21	8	72	88	2.6
澳大利亚	15	7	78	85	1.9
加拿大	14	7	78	77	1.6
埃及	29	8	64	45	3.4
萨尔瓦多	30	6	68	45	3.1
法国	13	9	78	73	1.6
以色列	21	7	77	91	2.8
日本	10	7	79	78	1.5
老挝	45	15	51	22	6.7
美国	16	9	76	76	2.0

资料来源: *The New York Times Almanac*.

新闻中的统计学

24. 新闻中的预测 寻找近期的一则新闻报道, 其中包含用于预测的一种相关。估计预测的有效性, 思考在这一节中所讨论的所有因果关系。总的来说, 你认为预测有效吗? 为什么?
25. 新闻中的最优拟合线 虽然新闻中很少使用散点图, 但也不是从来没有。寻找一篇新闻报道中的散点图(近期的或者以往的), 用目测方法绘制一条最优拟合线。讨论通过这

条最优拟合线你可以预测到什么。

26. 你自身的多元回归 用你自己生活中和工作中的一个例子，采用多元回归揭示一个重要

的发展趋势。不必做实际的分析，只需要用语言描述通过这个多元回归会得出什么，答案可能会有什么用处。

7.4 寻找因果关系

相关可能意味着因果，但是相关本身并不能够确定因果关系。需要更多事实才可以说明一个因素是另外一个因素的原因。在前面的内容中，我们发现两个变量之间的相关是下列可能的几个结果之一：①巧合；②一个共同的潜在原因；③一个变量确实对另外一个变量有直接影响，这个因果关系建立的过程本质上是一个排除前两种解释的过程。

在理论上，我们可以通过设计实验来排除前两种解释：

- 我们可以通过多次重复实验或者在实验中采用大量被试来排除巧合，因为巧合是偶然发生的。不可能在许多被试身上或者许多次实验中重复发生。在少数几次实验中因为巧合可以混淆实验结果，但在非常多的实验过程中这种巧合是不可能影响结果的。
- 我们可以通过对实验的控制和随机化排除混淆变量的影响（参考1.3节），这样你可以排除共同的潜在原因。通过这样的方式，在实验组和控制组之间只有关注变量的变化。如果这种实验控制排除了混淆变量，那么余下的效应就是所研究变量的影响。

可惜的是，这些想法在实验中经常是难以实现的。在排除巧合的情况下，大量重复一个实验是费时费力和不经济的。为了排除一个共同的潜在原因，实验必须控制除了关注变量之外的所有事情，这是不可能的。更进一步说，在许多时候实验是不可操作的或者是不符合伦理道德，因此我们只能搜集观察数据。因为观察研究不能建立明确的因果关系，要建立因果关系一定要寻找其他途径。

7.4.1 建立因果关系

假如你发现了一个相关以及一个不确定的因果关系。怎么能够检验你的怀疑呢？让我们回到

吸烟和肺癌的问题中来。在吸烟和肺癌之间的强相关自身并不能证明是吸烟导致肺癌的。理论上，我们可以通过一个控制实验来寻找证据。但是这样一个实验是不符合伦理道德的。因为实验需要强迫一组随机挑选的人们去吸烟。如果吸烟真的是肺癌的影响原因，那又该怎么办呢？

答案包括几条证据：第一，研究者发现吸烟和肺癌之间的相关发生在许多不同组别的人们之中，女性、男性以及不同种族和文化的人们；第二，这些不同组别的人们看起来并没有什么不同，在不吸烟者身上发现的肺癌比率更少；第三，那些吸烟时间更长和吸烟更多的人罹患肺癌的比率更高；第四，当研究者解释肺癌其他可能的潜在原因（比如暴露在氡气或者石棉中），他们发现几乎绝大多数罹患肺癌的个案都发生在吸烟者中间（或者发生在暴露于二手烟的人群之中）。

以上四点证据成为强烈的理由，但是仍然不能完全排除其他一些可能因素的影响，比如遗传可能会使人们倾向既吸烟又可能罹患肺癌。然而，另外两个证据使得这种可能性几乎不存在。一条证据来自动物试验。在严格控制的试验中，动物被随机分成实验组和控制组。实验还是发现了香烟吸入和肺癌之间的相关，至少在动物身上看起来像排除了遗传因素。另一个证据来自生物学家对细胞培养的研究上（取自人类肺组织的一个小样本）。生物学家发现香烟的构成物质能够引起癌细胞突变。这个过程的出现并不依赖于任何特定的遗传因素。暴露于二手烟中与肺癌相联系的事实更进一步反驳了遗传的观点（因为二手烟所影响的是非吸烟者），而这是和香烟的构成物质引起癌细胞突变相一致的。

相关知识

关于吸烟导致肺癌证据的强度，约翰霍布金斯大学的教授戴维·西德兰斯基（David Sidransky）说：“我们在分子水平的证据完全能够做到，以一个癌症患者为例，根据他基因变异的模式来确定吸烟是否导致他这种癌症。”

下面是关于如何建立因果关系观点的总结。通常来说，当下列这些指导条目出现的时候可

以说明存在因果关系的可能性很强。

建立因果关系的指导

如果你怀疑某一个变量是产生某些效应的原因(猜想原因):

(1) 寻找那些与猜想原因发生效应联系的而且其他因素也变化的情境。

(2) 当猜想原因是在不同组别的对象之间存在或者是不存在时,检查它的效应是同样存在还是不存在。

(3) 寻找引起这个效应的大量猜想原因。

(4) 如果这个效应可能是由其他潜在原因引起的(除了您的猜想原因之外),要确定在考虑了这些其他可能原因之后,这个效应仍然存在。

(5) 如果有可能的话,用实验检验这个猜想原因。如果这个实验由于伦理道德原因不能在人类身上操作,考虑进行一些动物实验、细胞培养试验和计算机模型实验。

(6) 尝试解决猜想的原因产生这个效应的生理机制。

相关知识

指导中的前四条叫做穆勒定律,以约翰·斯图亚特·穆勒(John Stuart Mill, 1806—1873)命名的。穆勒是他所在时代的重要学者,是妇女投票权力的一个早期倡议者,在哲学上,这四种方法分别叫做:一致法、差异法、共变法和剩余法。

思考题

关于动物实验是否符合伦理道德存在大量争议,你对动物试验持什么样的观点?说明你的观点。

案例研究

安全气囊和儿童

到20世纪90年代中期,副驾驶一侧的安全气囊成为汽车的标准配置。统计研究显示,在中速或者高速碰撞的过程中安全气囊挽救了许多人的生命。但是另外一种困扰也出现了,至少在某些情况下,年幼的儿童,特别是位于儿童汽车座椅中的婴儿和幼童,在低速碰撞中被安全气囊杀死了。

首先,许多安全倡议者发现这是难以置信的,安全气囊居然是导致死亡的原因。但是按照建立因果关系的前四条指导意见,观察得到的证据很充足的。例如,按照第三条指导意见,坐在儿童汽车座椅中的婴儿是有很大风险的,因为它说明表明接近安全气囊会增加死亡的风险。(儿童汽车座椅安装于汽车的前部,这样就使得儿童,比起其他位置来说和安全气囊更接近。)

为了实际验证这种情况,安全专家进行了一项仿真实验。他们发现,由于儿童身形较小,他们经常坐在安全气囊强力张开容易伤害到的地方。这个实验也显示安全气囊冲击儿童安全座椅的力量足以导致死亡,这样也揭示了儿童死亡的生理机制。

在掌握了生理机制之后,我们就有可能制定一些预防死亡的策略。首先,显而易见的最好的预防策略就是让儿童远离安全气囊,这引出了一条新的指导意见,告诉父母儿童安全座椅不应该被安装在前排,12岁以下的儿童应该尽可能坐在汽车的后排。其次,发现也引出了新的实验建议,告诉我们安全气囊在提供相同安全保护的同时应该具有较小的张力。现在新型汽车都需要打开时张力不是强烈的安全气囊。最后,儿童由于身形较小所导致的危险也提示身材较小的成年人会有同样的危险。这使得我们推荐身材较小的成年司机(比如个子较矮小的女性)的座位尽可能离安全气囊远一些,并且尽量使用配置了张力较小的安全气囊的新型汽车。

相关知识

在许多国家和美国的许多州中都有与儿童乘车相关的法律,例如在比利时,如果后排座位空着的话儿童坐在前排是非法的。



7.4.2 隐藏的原因

迄今为止,我们已经讨论在发现一个相关以后怎样建立因果关系。有时,无论如何,相关甚至没有相关背后隐藏着潜在的因果关系。如同下面的研究案例所呈现的,由于混淆变量的存在经常导致因果关系被掩盖。

案例研究

心脏搭桥手术

那些罹患有严重的动脉阻塞（冠状动脉）导致心脏供血不足的患者需要实施心脏搭桥手术。如果在这些动脉中血流停滞，罹患患者会遇突发心脏病导致死亡。心脏搭桥手术将一条心血管移植到阻塞动脉处，使的血液能够在阻塞的地方重新循环起来。直到 20 世纪 80 年代中期，许多医生还确信这个手术延长了罹患者的生命。

然而，一些关于过去的回溯研究使得这个研究结果让人困惑：就统计来讲，这个手术导致的差异很小。换句话说，平均起来实施这个手术的罹患患者比起没有实施手术的同样罹患患者并没有活得更长久。如果这是实际情况的话，那么它意味着忍受痛苦、冒着风险并且还要承受昂贵的费用去实施这项手术就没有价值。

因为这个结果与许多医生在他们自己病人身上所观察到的完全相悖，因此研究者决定进行更深入的研究。不久，他们发现实施这个手术的罹患患者所罹患的动脉阻塞更严重，很明显是因为医生对这样的罹患患者会更强烈地建议他们实施手术。因为这样的罹患患者情况更严重，所以将他们和其他罹患患者进行生存时间的比较是没有实际意义的。

更重要的是，不久以后研究者搞明白了罹患患者在不同医院施行手术差异的实质。特别是，少数几个医院在心脏搭桥手术方面做得极其成功，他们的罹患患者生存时间会比在其他医院手术或没有实施手术的人更长。很明显，高水平的医院里医生的手术技术是高超的，与其他医院的医生手术技术是有差异的。有关医生差异的研究要能够保证所有医生都要经过同样高水平的技术训练。

概括来说，混淆变量包括堵塞的数量和手术技术，这些证明了早期研究中在心脏搭桥手术和存活时间之间确实存在相关。今天，心脏搭桥手术被认为是延长那些冠状动脉阻塞罹患患者寿命的一种很好的方法。现在也是最常见的手术类型之一，一般情况下，实施这项手术罹患者的寿命能够增加 10 年。

7.4.3 因果关系的可信度

六条指导建议给我们提供了一个检验某种情况下检验因果关系强弱的方法，但是我们经常要在一个因果关系被完全确立之前做出决策。思考一下，例如，比如众所周知的全球变暖问题。到底是不是化石燃料的燃烧导致了全球变暖（参考本章末的“聚焦环境”栏目），所有的疑问都要被证明几乎是不可能的。但是当我们面临一些不确定的因果关系时必须做出决定怎么办，那么在我们做出决定怎么办之前必须要掌握多少信息？

在统计学的其他领域，一些已经被接受的技术可以帮助我们控制这一类不确定性问题，可以帮助我们计算信任度和显著性的数字水平。但是还没有可以使我们能够计算不确定性和因果关系之间数字水平的统计技术。可喜的是，另外一些研究领域可以帮助我们控制数百年的实际因果问题：我们的法律系统。你可以经常见到下面三个表达法律水平上的可信度的大概方法。

相关知识

如你所猜测的，合理怀疑是非常难以定义的。对于犯罪案件来说，美国最高法院认可鲁斯·金斯伯格（Ruth Bader Ginsburg）的观点：“合理怀疑的证据就是指你坚信犯罪人有罪的证据。在这个世界上很少有事物是绝对确定的。法律上的犯罪案件并不要求克服所有可能怀疑的证据，基于你对证据的思考，你坚信对被告的犯罪指控是成立的，你必须寻找他有罪的证据。另一方面，如果你认为他真的可能无罪的话，你必须给他怀疑的权利，并且寻找他无罪的证据。”



因果关系可信度的大概水平

可能原因：我们已经发现了一个相关，但是不能确定这个相关是否意味着因果关系。从法律角度说，可能的原因（比如我们认为某一个嫌疑人可能和某一个案件有关）是开始一个调查的常见原因。

可信的原因：我们有很好的理由去怀疑相关包含因果关系，也许因为符合一些建立因果关系的指导意见作用。从法律的角度说，得到法官批准进行搜查和窃听通常要达到可信的原因这样的标准。

超出合理怀疑的原因：我们发现一个物理模型能够成功解释一种现象是导致另一种现象的原因，

怀疑这个因果关系看起来是不合理的。从法律角度说，超出合理怀疑的原因是一个常用的定罪标准，它通常会要求检方说明嫌疑人怎么样和为什么与案件有联系（特别是物理模型）。注意超出合理怀疑并不意味着超出所有怀疑。

虽然这些大概的水平仍相当模糊，但它至少给了我们一些讨论因果关系可信度的常用语言。如果你学习法律，你会了解更多解释这些术语的微妙之处。尽管如此，因为统计学很少谈及这些，因此我们将不在本书中过多地讨论这些。

7.4 练习题

统计素养和批判性思维

1. 相关 说明对于两个变量之间存在的相关三种不同的解释。
2. 实验的作用 理论上，我们可以通过实验，能够排除两个变量之间三种相关解释的两种。三种解释中的哪一种我们不能排除？为什么我们不能排除它？
3. 混淆变量 什么是混淆变量？混淆变量怎么样使得潜在因果关系被隐藏起来？
4. 相关关系和因果关系 发现两个变量之间相关和建立两个变量之间因果关系有什么不同？

判断正误：判断练习题5~8中的陈述是有意义的（表述正确）还是没有意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. r 的取值 如果两个变量之间的成对数据得到了一个完全相关，相关系数是1，我们可以得出结论说一个变量是另外一个变量的直接原因。
6. r 的取值 在一个不同药剂量所导致副作用的药物实验中，依据药物量和疼痛指标的标准分数所得到的20对数据，计算得出相关系数。如果相关系数 $r=0.013$ ，那么，药物剂量大小就不是导致疼痛的直接原因。
7. 吸烟和可替宁 一个研究显示，暴露于二手烟与身体中的可替宁含量存在一个相关。如果我们不能排除相关解释中的巧合性的话，那么就不能得出暴露在二手烟中是导致可替宁含量上升的原因。
8. 尼古丁 假如关于尼古丁贴片的一个研究有

实际证据表明，尼古丁被身体吸收后的代谢物是可替宁。在此研究中进行的一个实验中有五名被试，相关分析说明尼古丁贴片的数量和身体内可替宁的含量存在一个相关。

概念与应用

物理模型：在练习题9~12中，有关因果关系的表述是否有效。如果因果关系是有效的，请加以解释。

9. 抛物线运动 高尔夫球的抛射距离受到击打此球时球棒速度的影响。
10. 磁疗 在你的腕部佩戴一个磁性手镯可以治疗心脏病。
11. 饮酒和反应时 大量饮酒会降低一个人的反应时。
12. 海拔高度和健康 当人们在没有携带制氧装备爬上较高海拔高度时，他们可能会体验到一些增强的生理反应，比如头疼和缺氧。
13. 确定原因：头痛 你试图确定困扰你好多天的每天下午头痛的原因。对于下列的每一种检测和观察，请解释符合建构因果关系的六条指导原则中的哪一条。基于你所有的观察，然后总结你的一个整体性结论。
 - (1) 只有当你去工作的那些天里才发生头痛。
 - (2) 如果您在午饭时不再饮用可口可乐，头疼还是会发生。
 - (3) 在夏天里，如果你稍微打开办公室的窗户，头痛发生的次数会少一些。在这种情况下，甚至比全部打开办公室的窗户时，头痛的次数还要少。

14. 吸烟和肺癌 在吸烟和肺癌的发生之间存在着一个强相关，并且大多数生理学家认为吸烟会导致肺癌。但是，并不是每个吸烟的人都会罹患肺癌。简要讨论吸烟是怎样导致肺癌的，同时为什么不是所有吸烟的人都会导致肺癌。
 15. 其他导致肺癌的原因 除了吸烟以外，研究显示其他的一些因素也可能是导致肺癌的原因。例如，暴露在石棉中和暴露在氡气中都能导致肺癌，这两种化学物质可能在很多家庭中都会发现。假如所遇到的一个人，他生活在含有高浓度氡气水平并且绝缘材料中含有研究石棉的住所中，他告诉你：“我也吸烟，因为我认为自己注定要得肺癌的。”你对他的话将作何反应？请加以说明。
 16. 管弦乐队指挥的长寿 发表在《医药论坛》上的一项著名研究得出结论说，一些主要的乐队指挥的平均寿命是 73.4 岁，比那个时期全部美国男性的平均寿命要高出 5 岁。作者因此得出结论说音乐生涯可以导致长寿。评估这个结论的因果关系，并且提出指挥家长寿的其他解释。
 17. 年长的母亲 发表在 *Nature* 上的一篇研究报告声称，生育年龄较晚的一些女性倾向于寿命更长。在研究中的 78 位至少 100 岁以上的老年妇女，其中 19% 是在她们过了 40 岁生日之后才有了孩子。在研究中 54 位当时已 73 岁的女性，只有 5.5% 是在 40 岁生日以后才有孩子的。研究者表述说：“如果你的生殖系统老化得比较慢的话，你可以在 40 岁以后再要第一个孩子，它可能是一个很好的预示，说明你余下的生命时光也会老得比较慢。”这是一个观察研究还是一个实验研究？这个研究是否意味着，晚要孩子是导致长寿的原因或者是一个潜在原因？评论研究报告中的结论是怎样让你信服的。
 18. 高压线 假如生活在高压线附近的人们比起那些生活远离高压线的人们有一个较高的癌症发生率。你能得出结论说高压线是导致癌症发生率增加的原因吗？如果不是的话，其他可能的解释是什么？在你得到结论说高压线导致癌症之前你想看到什么样的其他类型的研究？
 19. 枪支控制 那些支持对枪支进行控制的人们经常会指出，在凶杀率和枪支拥有率之间存在一个正相关，以此来支持他们关于枪支控制可以拯救生命的观点。这个相关自身是否能够说明枪支拥有导致了高凶杀率？揭示一些其他可以支持或者反对这个结论的因素。
 20. 输精管切除术和前列腺癌 一篇题目为“输精管切除术会导致前列腺癌吗”(*Chance*, Vol. 10, No. 1) 的文章后报告说，几个大型研究发现男性在输精管切除术后增加了罹患前列腺癌的风险。在缺乏直接原因的前提下，几个研究者将相关归结于检测偏差，那些实施了输精管切除术的男性更有可能去看医生，也就更有可能被医生发现罹患前列腺癌。简要解释检测偏差是怎样影响输精管切除术导致前列腺癌这样结论的。
- ### 网络和远程项目
- 请在网站 www.aw.com/bbt 上打开第 7 章的有用链接“links for internet projects”。
21. 安全气囊和儿童 从美国国家高速公路交通安全委员会的网站上，寻找有关安全气囊安全性的最近研究报告，特别是有关儿童的。写出你所发现的简短的报告总结，并提出在汽车上如何提高儿童安全的建议。
 22. 膳食纤维和冠状动脉心脏病 关于膳食纤维预防女性的冠状动脉心脏病 (CHD) 的大型研究中 (*Journal of the American Medical Association*, Vol. 281, No. 21)，研究者探测到食用较多膳食纤维降低了女性冠状动脉心脏病的风险。寻找这篇研究论文，总结它的发现并且讨论所得到的相关是否具有因果关系。
 23. 咖啡和胆结石 发表在《美国医学会杂志》(第 281 卷第 22 期) 上的一篇研究报告发现了男性的咖啡消费量和胆结石疾病的发生率之间存在一个负相关。找到这篇研究报告，总结它的发现，并且讨论所得到的相关是否具有因果关系。
 24. 酒精和中风 发表在《美国医学会杂志》(第 281 卷第 1 期) 上的一篇研究报告说中等水平的酒精消费和 40 岁年龄的人们中风的风险下降存在一定的相关 (过量饮酒可产生有害的结果)。找到这篇研究报告，总结它的发现，并且讨论所得到的相关是否具有因果关系。
 25. 烟草诉讼 烟草公司有很多件与烟草的危害有

关的诉讼案件，研究一个近期的诉讼案件，原告试图证明什么？他们使用什么样的统计证据？你认为他们所建构的因果关系合理吗？谁赢得了诉讼？用一两页纸总结你的发现。

新闻中的统计学

26. 新闻中的因果关系 寻找近期的一则新闻报道，其中的统计研究得出了因果关系的结论。描述这个研究以及所声称的因果关系。

第7章复习题

对于复习题 1~4，参考表中的烟草数据。所有的测量单位都是每根香烟的微克量，所有香烟都是 100 毫米长，带有过滤嘴，不包含薄荷味和轻型香烟（based on data from the Federal Trade Commission）。

品牌	焦油	尼古丁	一氧化碳
骆驼	16	1.0	17
健牌	13	1.0	13
幸福	13	1.1	13
马里布	15	1.2	15
万宝路	16	1.2	15
Merit	9	0.7	11
Now	2	0.2	3
旧金	18	1.4	18
波迈	15	1.2	15
云丝顿	16	1.1	18

- 根据焦油和尼古丁的成对数据计算的相关系数是 $r=0.962$ ，关于焦油和尼古丁相关的强度你可以得到什么样的结论？通过焦油和尼古丁之间的相关可以解释尼古丁变化量的百分比例是多少？
- 根据一氧化碳和尼古丁的成对数据计算的相关系数 $r=0.909$ ，关于一氧化碳和尼古丁相关的强度你可以得到什么样的结论？通过一氧化碳和尼古丁的相关可以解释尼古丁变化量的百分比例是多少？
- 根据焦油和一氧化碳的成对数据计算的相关系数 $r=0.979$ ，对于焦油和一氧化碳相关的强度你可以得到什么样的结论？通过焦油和一氧化碳相关可以解释焦油的变化量的百分比例是多少？
- 用给出的焦油和尼古丁的数据值绘制一个散点图（使用焦油的数据值作为横坐标）。这个统计图的什么特点说明在焦油和尼古丁之

你认为报道中所声称的因果关系合理吗？请加以解释。

27. 法律上的因果关系 寻找近期的一则有关法律案件的新闻报道，可以是民事诉讼也可以是刑事诉讼，其中所示的因果关系对于结果是非常重要的。简要描述案件中有关因果关系的主题，支持或者反驳因果关系的能力是怎样影响案件结果的。

间是否存在一个相关？

- 基于瑞典的研究，几家报纸报道说：“住所距离高压线比较近会导致儿童易罹患白血病。”什么样的数据可能成为这样一个声明的基础？“邻近高压线会导致儿童易罹患白血病”这个声称的根本性错误是什么？
- 计算 10 对样本数据的相关系数是 $r=-1$ ，你知道它的散点图是什么样的吗？
- 在一个随机选择被试的研究中，研究发现在家庭收入和看牙医的次数之间存在一个强相关。得出“高收入人群看牙医的次数更多”这样的结论是有效的吗？这个相关如何解释的？
- 你思考一下你所购买的最昂贵的东西：你购买的住房。至少确定可能影响住房实际价格的五种不同变量。在这些你所确定的变量中，哪一个变量对房屋价格的影响最大？确定一个对房屋价格影响很小甚至没有影响的变量。
- 一个研究者搜集了成对样本量数据，并且计算了它们的线性相关系数是 $r=0$ 。基于这个数值，他得出结论是两种变量之间没有关联。这个结论错误在哪里？
- 观察一下图 7-23 的散点图并且估计相关系数的值。

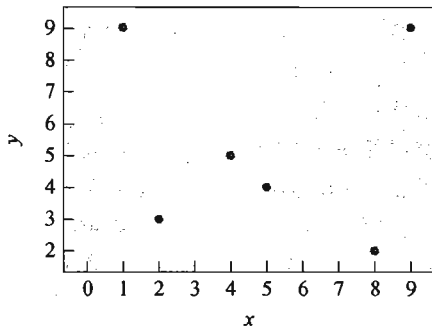


图 7-23 散点图

第7章小测验

- 1. 填空：每一个可能的相关系数的取值一定在_____和_____之间。
- 2. 下列成对数据中哪一个更可能存在相关？
 - (1) 随机选择的被试的 SAT 测验成绩和体重
 - (2) 随机选择的被试的反应时和智商
 - (3) 随机选择的被试的身高和臂长
 - (4) 随机选择的民航飞机的满座率和航空公司的利润
 - (5) 随机选择的汽车拥有者的汽车价格和年收入
- 3. 对于搜集的成对样本数据，计算出的相关系数是 -0.998 。下面哪一种陈述是对于两种变量关系的最好描述。
 - (1) 它们之间没有相关。
 - (2) 它们之间存在一个弱相关。
 - (3) 这是一个强相关的。
 - (4) 一个变量是另外一个变量的直接原因。
 - (5) 这两个变量之间没有直接关系。
- 4. 估计图 7-24 中散点图中数据的相关系数。
- 5. 再次参考图 7-24 中的散点图。在两个变量之间存在一个显著的相关吗？
在测验题 6~10 中，判断所给出陈述的对错。

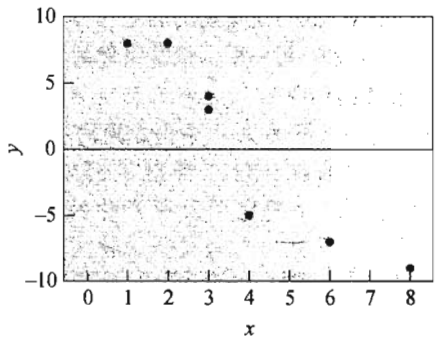


图 7-24 散点图

- 6. 如果两个变量之间存在一个强相关，那么其中一个变量是另外一个变量的直接原因。
- 7. 一个散点图就是所有点都是散落分布的，并且没有任何规律的图形。
- 8. 如果一个散点图所有的点都位于左下角到右上角的一条直线上，那么相关系数 $r=1$ 。
- 9. 如果从随机挑选的被试上，搜集到他们从 1995~2008 年之间收入和税收的成对数据，存在一个相关，说明这种关联可以用来预测 2015 年的基于收入的税收。
- 10. 如果 $r^2=0.09$ ，相关系数 r 的取值有可能是负数。

技术应用

SPSS

散点图：进入 SPSS 数据编辑窗口中。点击 [Graphs]，然后选择菜单项 [Scatter/Dot]，然后点击 [Simple Scatter]，接下来点击 [Define]。在对话框中，点击左侧的变量然后点击中间的按钮，将第一个变量放入相应的“X axis”（横坐标）或者“Y axis”（纵坐标）框中，然后将第二个变量放入相应的“Y axis”或者“X axis”（纵坐标）框中。然后点击 [OK] 键，这样散点图就会被呈现出来。

直线相关系数 r ：进入 SPSS 数据编辑窗口中。点击 [Analyze]，然后选择菜单项 [Correlate]，接下来再选择 [Bivariate]。在对话框中，点击 [Pearson]。点击左侧的变量然后点击中间的按钮将选中的变量移入“Variable”框中。进一步编辑左侧其他的变量然后点击中间按钮将其选入“Variable”框中。然后点击

[OK] 就可以得到相关系数 r 的计算值，并且会得到一个显著性的说明。

最佳拟合线：进入 SPSS 数据编辑窗口中。点击 [Analyze]，然后选择菜单项 [Regression]，接下来选择 [Linear]，点击左侧的变量然后点击中间的按钮将选中的变量选入对应的“Dependent”或者“Independent”，然后点击 [OK]，就会得到几个信息表。寻找到“相关系数表”可以得到最优拟合线的斜率和截距（表中名称为“Constant”）。

Excel

线性相关系数 r ：首先选中成对样本数据 A 和 B 列，然后点击位于主菜单键上的“fx”功能键，然后选择分类功能 [Statistical] 并且选择功能名称 [CORREL]，然后点击 [OK]。在对话框中，输入 x 变量的取值范围，例如 A1:A8，并且输入 y 变量的取值范围，例如 B1:

B8, 这样线性相关系数 r 的取值就会呈现出来。

散点图: 如果使用 Excel 2003, 点击主菜单上的 [Chart Wizard], 然后选择图形类型名称为 [XY (Scatter)]。在对话框中, 输入数据的范围, 比如 “A1: B8”, 然后点击 [Next], 然后继续使用对话框修改你想要的图形。

如果你使用 Excel 2007, 使用鼠标加亮成对数据。然后点击 [Insert], 然后在图表区选择 [Scatter], 将会呈现几个图形, 点击左上角的一个, 散点图就会呈现出来, 可以修改成你想生成的图形样式。

最佳拟合线: 首先选中成对样本数据 A 和 B 列, 如果使用 Excel 2003, 从主菜单上选择 [Tools], 然后选择 [Data Analysis] 和 [Regression], 接下来点击 [OK]。如果你使用 Excel 2007, 先点击 [Data], 然后点击 [Data Analysis], 然后选择 [Regression], 接下来

点击 [OK]。在数据选择对话框中, 输入 x 变量的范围, 比如 “A1: A8”。点击标记为 “Line Fit Plots” 的键, 然后再点击 [OK]。在 Excel 提供的所有信息中, 最优拟合线的斜率和截距可以在名为 “相关系数” 的表格下面发现, 所呈现的图形包括一个原始数据点构成的散点图, 同时还有一个最优拟合线所预测的点分布图。

STATDISK

首先进入 STATDISK 的数据窗口, 选中成对数据列, 然后在主菜单条中选择 [Analysis], 然后进行操作 [Correlation and Regression]。输入一个显著性水平的值, 选中可用数据列, 点击 [Evaluate] 键, STATDISK 就会呈现包括线性相关系数以及最优拟合线方程斜率和截距。通过点击 [plot] 键也可以得到一个散点图。

| 聚焦教育 |

是什么帮助儿童学习阅读

关于怎么样才能最有效地教会儿童阅读, 每个人都有自己的想法。一些人提倡 “语音教学法”, 也就是让学生 “大声读出” 单词。还有一些人提倡 “整体语言教学法”, 也就是教会学生从上下文语境中识别单词。另外还有一些人提倡把这些方法相互结合在一起, 或者其他完全不同的教学方法。如果这些不同的方法仅仅只是一种观点或想法的话, 那么它们并不重要。但对于在教育方面每年花费上万亿美元的一个国家来说, 不同的阅读教学方法就涉及了不同特殊利益集团之间的政治冲突。一个政治上的变革会导致学校教育政策突然的变化。例如, 1998 年加利福尼亚州立法机关通过了相关法律, 即公立学校的基金与学校是否放弃阅读的整体语言教学法挂钩。



巨大的利益关系涉及阅读教学要求统计学方法来测量各种方法的有效性。在政治方面, 至少最重要的教育统计学方法来自美国教育进展评价项目 (简称 NAEP), 通常被更简单地称为 “全国报告卡”。NAEP 是一项持续性的调查学生学业成就的项目, 由全国教育统计资料中心管理, 由美国国会授权和资助。

NAEP 采用分层随机抽样 (参考第 1 章), 根据不同的种族、家庭收入、学校类型等, 选取四年级、八年级、十二年级的学生, 形成有代表性的样本。对选取出的学生进行测试, 旨在测量他们在特定学科领域的学业成就, 比如阅读、数学或历史。既在各州范围内也在全国范围内选取样本。总体来说, 每一项测试中都会包括几千名学生样本。随着报纸报道 NAEP 测试分数的提高或降低, 其结果必然导致文章是受人吹捧还是遭到谴责。NAEP 测试结果同样也有政治影响。例如, 加利福尼亚州学校的 NAEP 测试成绩在 50 个州排名第 45, 之后该州学校才摒弃了整体语言教学法。

然而究竟是什么导致了阅读能力的提高? 研究者开始探询阅读能力和其他因素的相关。有时相关是显著的, 但是却没有指出明确的方向。例如, 父母的受教育程度与阅读成绩有显著相关: 与父母没受教育的儿童相比, 父母接受高等教育的孩子能够更熟练地阅读。但是这个相关不能给学校提供指导, 因为孩子无法更换他们的父母。有些时候这些相关会提示一些改善阅读的方法。比如, 在学校和家庭作业中, 每天阅读量多的学生比阅读量少的学生成绩得分更高。这就说明学校应该布置更多的阅读任务给

学生。

当然，教育方面涉及的巨大利益关系会特别导致教育统计被误解或误用。不妨来思考几个让 NAEP 阅读测试难以解释的问题：

1. NAEP 阅读测试是一套标准化测试，对所有学生的测试内容都是一样的，往往是选择题居多。一些人认为这样的测试必然是有失偏颇的，并且无法测量出学生的真实阅读能力。

2. 因为 NAEP 测试通常不会影响学生的升级情况，所以有些学生就不会很认真地对待这个测试，导致测试结果无法反映出他们的真实阅读能力。例如，一些监考人员就反映，有些学生会随便完成选择题，而不是运用自己的能力认真回答问题。

3. 如果各州组成样本的学生群体之间差异明显的话，那么就无法对各州的测试成绩进行比较。例如，在加利福尼亚州有相当大部分的学生把英语作为他们的第二种语言。一些人认为这才是该州学生测试分数低的原因，而与教学技术无关。

4. 有些证据表明，NAEP 测试中有部分人涉及作弊。比如在选择样本时，并非选择真正有代表性的样本，而是偏向选择那些阅读水平更好的学生。这种作弊的动机可能是由于各个学校、学区及各州都要根据 NAEP 的测试结果进行排名。获得高分的学校，其学校管理层和教师会得到奖励，他们的财政资金会增加，薪水会提高。而相反成绩低的学校会受到各方面的处罚。

你可能会想到其他一些导致 NAEP 测试结果难以解释的问题。因此，阅读方法毫无意外地会继续发展成为一个巨大的政治战场。那么，你自己作为一个体，又能做些什么来帮助儿童学会阅读呢？幸运的是，NAEP 测试研究也发现了一些符合常理的并且不会引起争议并与阅读水平相关的因素。例如，高水平的阅读能力与以下的几个因素存在相关关系：

1. 不管是为了学业成绩还是出于兴趣，只要有更多的阅读量；
2. 相对而言，让学生的阅读有更多选择，也就是允许学生自主选择阅读内容；
3. 更多的写作练习，尤其在拓展到如散文或长信等形式的写作；
4. 与朋友和家人进行更多有关不同阅读材料的讨论；
5. 少看一些电视节目。

以上这些相关因素至少能指导人们如何帮助儿童学会阅读，并且这是一个探讨如何提高读写能力的很好出发点。当然，政客们和特殊利益集团可能会找到某些方法，使这些测试结果符合任何他们可能预设的议程方案。因此，如果你对于教学技能有自己的强烈意见，你也可以加入到这场政治争斗中，而这场斗争可能会持续进行几十年。

问题讨论

1. NAEP 阅读测试的一个显著结论是，私立学校的学生成绩得分比公立学校的学生明显要高。这是不是就暗示着私立学校的教学质量要比公立学校“好”呢？请阐述你自己的观点。
2. 你认为，类似 NAEP 这种标准化测试真的能有效测量出学生的学业成绩吗？为什么？
3. 目前，NAEP 测试只在美国几千所学校上百万名学生中进行了施测。有些人主张自愿或者强制地让所有学生接受同样的测试。你认为这样的“全国标准化测试”是一个好方法吗？为什么？
4. 关于计算机应用和阅读能力之间的关系，目前还没有仔细研究。你认为，使用计算机和网络会有助于儿童学习阅读，还是会阻碍其学习？为什么？
5. NAEP Reading Report Card（可以链接下面的网址得到），其中关于美国阅读教学的最新研究成果是什么？

建议阅读

1. Lemann, Nicholas. “The Reading Wars.” *The Atlantic Monthly*, November 1997.
2. NAEP Reading Report Card. 在线最近版本的网址 <http://nces.ed.gov/nationsreportcard/reading>.
3. Schemo, Diana. “In War over Teaching Reading, a U. S. -Local Clash,” *New York Times*, March 9, 2007.

| 聚焦环境 |

是什么原因导致全球变暖

在第3章聚焦环境中提供的观察数据，可以用来解释全球正处于变暖的过程中。在这部分内容中，开始将我们的注意力转到更深层次的一个问题上：所观测到的全球变暖现象是由于自然原因导致的，还是由于我们

使用化石燃料或其他人类活动而导致的呢？这个问题会产生广泛的政治和经济影响后果。如果全球变暖是自然原因导致的，那么我们就不过分警惕和紧张，因为这种现象可能会很自然地转变，或者我们根本也无能为力。但是如果全球变暖是由于我们人类活动造成的，并且仍然持续我们过去的活动方式，那将会进一步让全球变暖的进程加剧，此时，在这个进程对人类文明造成巨大威胁之前，我们就很有必要进行政治和经济的改变来减缓这个问题的负面影响。

如今，那些研究地球气候的科学家存在一个明确的共识，他们都认可人类活动是导致全球变暖的原因。2007年，政府间气候变化委员会（IPCC）——由来自世界各地的600多名气候学家组成的专家组得出明确的结论说：过去50年观察到的全球气候变暖现象，至少有90%以上确实是由于人类排放的气体诸如二氧化碳造成的，而二氧化碳的持续排放会导致这个世纪的气候更大程度地变暖，并且会造成气候类型实质性的改变以及海平面的上升。那么科学家是如何得出这个结论的呢？

为了更好地理解这些争论，我们首先来看一下关于全球变暖的潜在科学基础：某些大气气体会产生吸收热量的**温室效应**（greenhouse effect）。温室效应的物理机制很简单，数十年以前人们就已经可以很好地理解了。它的运行机制如图7-25所示：太阳光是一种能温暖地球表面的能量，为了能使地球表面的温度保持稳定，被吸收的热量就必须返回到宇宙空间中（否则，持续吸收的太阳光会使地球表面温度迅速地上升到很高的水平），地球表面是以红外线的形式把热量返回到太空中。（我们用肉眼无法看见红外线。）水蒸气、二氧化碳和甲烷这些气体都被称做温室气体，当这些温室气体以自己的方式吸收了向上散发过程中的红外线时，温室效应就产生了。这些气体向各个方向重新释放红外线，因此它们的整体效果是保留了更多的能量集中在地球表面，使之更加温暖。大气中温室气体的成分越高，向外太空散发热量的速度就越慢，地球就会变得越温暖。

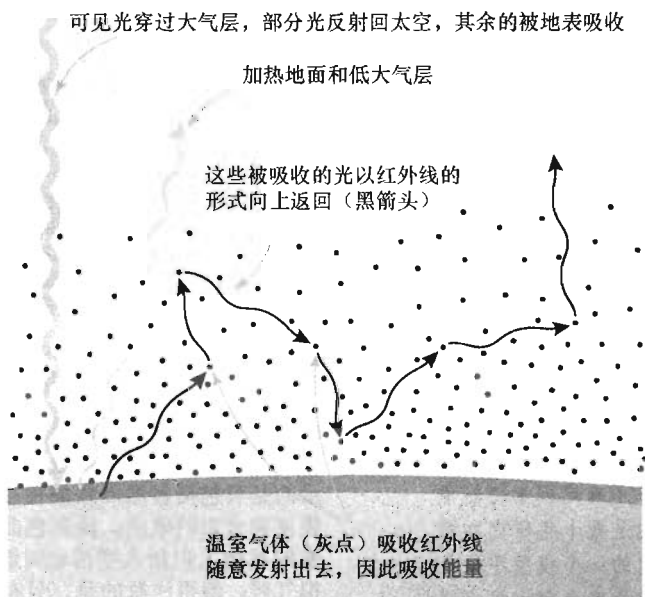


图7-25 这个图呈现了温室效应的基本作用机制由于大量温室气体的存在，逸出的红外线变得缓慢，行星因此变暖

值得注意的是，虽然对于人类活动引起全球变暖的程度仍然有一些不确定性，但是对于温室效应的机制并没有科学质疑。事实上，地球上温室效应的产生是自然现象，并且已经被证实是件好事：如果没

有温室效应，地球表面的平均温度将会接近 0 °F，从而使地球太冷以致对液态水和生命的存在是不利的。但由于自然产生的温室效应，地球的实际平均温度接近 60 °F。对于其他行星的研究进一步证实了我们对温室效应的理解，并且这种研究发现甚至比起一个行星和太阳之间的距离而言，温室效应对这个行星表面温度的影响来得更重要。金星就提供了一个最有利的证据：尽管金星比起地球来，它距离太阳更近，但是金星的云层会不断反射阳光，导致到达金星表面的阳光要比到达地球表面的阳光少得多。因此，如果没有温室效应的情况下，金星应该比地球温度低得多。然而由于金星表面有一层很厚的大气层，并且其中包含着比地球大气层中更多的二氧化碳（大约是地球的 170 000 倍），金星的实际表面温度达到了灼热的 870 °F。自然产生的温室效应对于地球上的生命而言是一件好事，而金星的例子却说明温室效应太大反而不是一件好事。

正是由于温室效应机制科学上的确定性，最先导致我们对人类活动造成全球变暖的可能性的质疑。由于人类活动如化石燃料的燃烧（随着其他活动如砍伐或烧毁森林）等向大气中释放了大量二氧化碳和其他温室气体，我们可以预见大气中的温室气体浓度会增加，进而使我们的地球变得更暖，超过其自然状态下的温室效应水平。事实上，如果我们继续长期地向大气层中排放温室气体，毫无疑问将最终导致更严重的全球变暖，足以威胁我们的生命。现在的问题是，气候变暖的进程是否已经在发展中并且发展的程度是怎么样，我们是否能够区分出温室气体浓度的变化是由于人类活动造成的还是由于自然因素造成的。

为了回答这些问题，美国和其他国家在过去的 20 年里已经投入了数十亿美元，尽了前所未有的努力去研究地球的气候。正是这项研究证实了人类排放的温室气体是导致全球气候变暖的原因。下面的 3 个方面的证据提供了特别强有力的证明。

第一，如我们在第 3 章所讨论的，现有的数据显示了在过去一个世纪里，一个清晰全球变暖的趋势，在过去 30 年中全球变暖加速（见图 3-60）。这些数据有效地解决了全球是否正在变暖的争论，留给我们的问题是全球变暖是由于人类活动还是自然原因造成的。

第二个证据来自对地球大气层中过去的和现在的二氧化碳浓度的精密测量。如果回顾这些数据资料（见图 3-61），你可以看到过去二氧化碳浓度的变化与温度的变化呈显著相关，符合我们关于温室气体浓度越高气温越高的预想。（尽管如此，这些数据本身并不能证实温室气体导致了更高的气候温度；那样的理解来自其物理运行机制。）值得注意的是，虽然过去的数据显示二氧化碳浓度在自然地变化，但是同时也显示出近年来温室气体浓度的增长比在过去几百几千年中的增长要大得多。对于近期的二氧化碳浓度巨幅增加，人类活动是唯一可行的解释。

第三个证据来自实验研究。我们不能对我们所处的地球进行整体的控制实验，但我们可以运用计算机模型进行实验来模拟地球气候的变化方式。地球气候极其复杂，在建立计算机气候模型时很多不确定因素仍然保留了。然而，现有的模型是数十年研究和精炼的结果：每一次当此前的一个模型不能与真实的数据相匹配时，科学家会努力找出模型中遗漏或者错误的部分，再尝试改进模型。现有的模型并不完善，但是与真实的数据能很好地匹配，科学家相信这个模型有很好的预测价值。图 7-26 将模拟数据和真实数

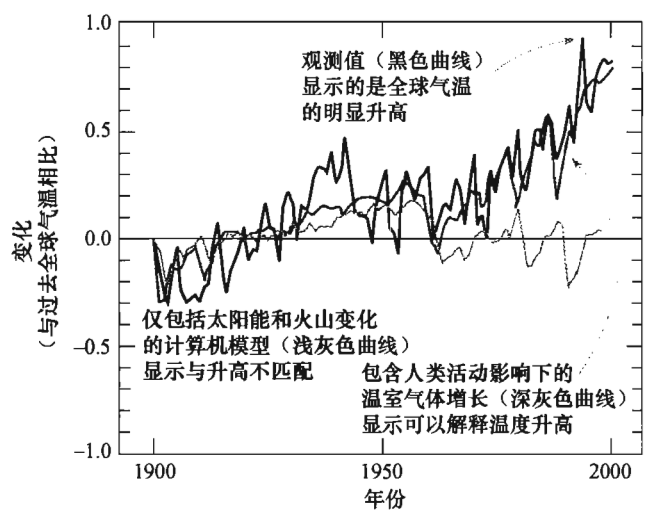


图 7-26 此图将实际观测的气候温度变化（黑色曲线）与气候模型模拟的大气温度进行了比较。深灰色曲线表示的是只包括自然因素的模拟气候，比如太阳光的强度和火山的效应。浅灰色曲线表示的是在自然因素的基础上包括人类活动对温室气体的影响后的模拟气候。值得注意的是，只有深灰色曲线才与实际观测的数据匹配良好，特别是近 10 年来，提供了人类活动是全球气候变暖的强有力证据。（深灰色曲线和浅灰色曲线是许多科学家所建立的相互独立的全球变暖模型的平均数，通常彼此之间的误差只在 0.1 ~ 0.2 °C。）

据进行了对比。更为重要的是,只有将人类向大气中排放温室气体的影响包含在内时,模型才能更好地匹配真实数据。如果只考虑到自然因素的模型则无法很好地反映真实观测到的数据。

结论非常清楚:根据这个气候变化模型、有效的数据以及我们对温室效应机制的理解,我们确信全球变暖主要是由于人类向大气层中大量排放二氧化碳和其他温室气体而导致的。

问题讨论

1. 回顾 7.4 节建立因果关系的六条原则,讨论每个原则是否或如何满足现有的数据并理解全球变暖现象。
2. 政府间气候变化委员会 (IPCC) 得出结论说全球变暖“很有可能”是由于人类活动造成的,但仍然有一些人提出相反的观点。做一个网络调查,找到那些不同意 IPCC 结论的人们的观点,并认真评价每一个观点。
3. 回顾在 7.4 节中讨论的因果关系中的规定信任水平,你认为人类活动是导致全球变暖现在已经到了可能原因这个程度吗?还是其程度只是超越了合理质疑范围?请阐述你的观点。
4. 根据你对于全球变暖的看法,描述一下你认为应该采取什么措施来改善这一现象。

建议阅读

1. Bennett, Jeffrey, and Shostak, Seth. *Life in the Universe*, 2nd ed. Addison Wesley, 2007 (see Section 10.5).
2. Flannery, Tim. *The Weather Makers*. Atlantic Monthly Press, 2005.
3. IPCC. *Climate Change 2007*. The full report is available online at <http://www.ipcc.ch>.



一些人非常厌恶统计学这个名字，但我却发现它是美丽和有趣的，它并不冰冷，只是采用这个高级方法要进行得认真细致，并谨慎地进行解释，它处理复杂事件的能力是无与伦比的。

——弗朗西斯·高尔顿

第 8 章 从样本到总体

学习目标

■ 8.1 抽样分布

理解抽样分布的基本观点，理解样本平均数和样本比率的分布是怎样形成的。掌握表示样本平均数和样本比率的符号。

■ 8.2 总体平均数的估计

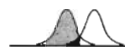
掌握总体平均数的估计和计算相关联的误差幅度

和置信区间的方法。

■ 8.3 总体比率的估计

掌握总体比例的估计和计算相关联的误差幅度和置信区间的方法。

你 是否曾经期望一次联邦选举的结果在投票结束数小时之前就可以被预测？或者一家大零售商仅仅基于一个数百人的调查就可以做出关键市场营销的决策？这些例子表明了统计学的强大影响，具有从一个小样本获得的信息预测更大总体估计的能力。这个过程，叫做推断，这个统计学的分支叫做推断统计。



8.1 抽样分布

考虑下述来自最近的新闻文章或研究报告中的陈述：

- 美国人平均每天的蛋白质消耗是 67 克。
- 在全美国范围内，孕妇在分娩一个婴儿后，在医院停留的时间从 1980 年的平均 3.2 天下降到现在的 2.0 天。
- 30% 的美国高中女生相信结婚比不结婚快乐。
- 全部美国儿童中的大约 5% 和祖父母一起生活。

我们每天能听到和读到像这样的表述。他们调查了大量的个体，但是总体中每个人都被测量和调查仍然是不可能的。例如，第三条陈述是基于一个高中女生的随机样本，但并不是建立在对这个国家的每一个高中女生的调查上。样本中女生的反应通常被看做全部高中女生的代表。同一个随机样本推论出关于总体非一般性结论的可能性是怎样的。这些问题的回答涉及统计学的重要组成部分，即推断统计。

在前面的陈述中注意两种不同类型的声称。前两个声称估计了一个量化的平均数，新蛋白质消耗 67 克，平均住院时间为 3.2 天和 2.0 天，

后两个陈述是关于总体比率的, 30% 的高中生和 5% 的美国儿童。这些平均数和比例是属于全部总体的, 所以他们是总体参数 (见 1.1 节)。本章的主题和统计推断的主要目标, 是学习如何使用样本数据估计总体参数。

8.1.1 样本平均数: 基本理念

本章和下一章的主要工作包括从总体中选择一个样本, 分析这样本, 然后依据样本中的信息得出总体结论, 为了从这一重要过程中获得一些经验, 我们从一个例子开始。

表 8-1 列出了一个职业篮球队的 5 名首发队员 (用字母 A ~ E 表示) 的身高, 为了方便起见, 我们将这 5 名运动员看做全部总体 (平均体重 242.4 磅的), 换句话说, 这是该球队首发阵容的全部总体。从统计学标准来说, 这是一个非常小的总体, 正因为这个总体很小才使我们能够认真地看到这个抽样过程。从 5 个运动员的总体中抽样出来的样本容量从 $n=1$ 到 $n=5$ 。

表 8-1 一个篮球队 5 名首发队员的体重

篮球队员	体重 (磅)
A	215
B	242
C	225
D	215
E	315

当样本容量 $n=1$ 时, 有 5 个不同的样本可以供选择: 每一个运动员都是一个样本。每一个样本容量 $n=1$ 的样本的平均数刚好就是样本里的这名运动员的体重。图 8-1 呈现了样本平均数的直方图, 这也叫做样本平均数的分布 (distribution of sample means), 因它的呈现了 5 个容量为 $n=1$ 的样本平均数。通过一个

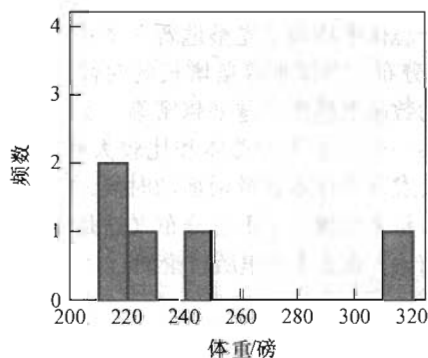


图 8-1 样本容量 $n=1$ 时的抽样分布

抽样分布 (sampling distribution) 的例子说明了样本平均数分布的产生过程。这一术语指的是样本统计量的分布, 比如从特定样本容量中提取的可能的样本平均数, 注意这 5 个样本平均数的平均数刚好是总体的平均数:

$$215 + 242 + 225 + 215 + 315 = 242.4 (\text{磅})$$

这证明了一个基本的法则: 样本平均数分布的平均数就是总体平均数。

让我们将样本容量扩大到 $n=2$, 在每一个样本里包括两名不同的运动员。对于这 5 名运动员来说, 当样本容量 $n=2$ 的时候, 有 10 个不同的样本, 每一个样本都有自己的平均数。例如, 由运动员 A 和 B 组成的样本, 其平均数为 $(215 + 242)/2 = 228.5$ 磅, 表 8-2 列出了这 10 个样本以及它们的平均数, 图 8-2 则呈现了这 10 个样本平均数的分布。另外, 注意这个样本平均数分布的平均数也等于总体平均数 242.4 磅。

表 8-2 当 $n=2$ 时, 篮球队员体重样本的样本平均数

篮球队员	体重 (磅)
AB	228.5
AC	220.0
AD	215.0
AE	265.0
BC	233.5
BD	228.5
BE	278.5
CD	220.0
CE	270.0
DE	265.0
平均数	242.4

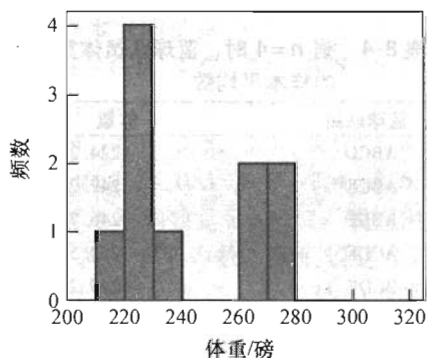


图 8-2 样本容量 $n=2$ 时的抽样分布

总体为 5 个运动员, 当样本容量 $n=3$ 时, 也有 10 个不同的样本, 表 8-3 呈现了这些样本和它们的平均数。图 8-3 呈现了这些样本平均

数的分布。另外这个样本平均数分布的平均数也等于总体平均数 242.4 磅。

表 8-3 当 $n=3$ 时，篮球队员体重样本的样本平均数

篮球队员	体重（磅）
ABC	227.3
ABD	224.0
ABE	257.3
ACD	218.3
ACE	251.7
ADE	248.3
BCD	227.3
BCE	260.7
BDE	257.3
CDE	251.7
平均数	242.4

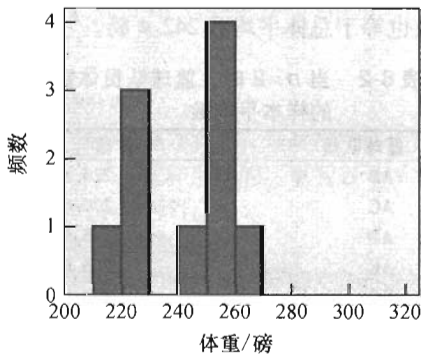


图 8-3 样本容量 $n=3$ 时的抽样分布

当样本容量 $n=4$ 的时候，只有 5 个不同样本。表 8-4 显示了这些样本和它们的平均数，图 8-4 呈现了这个样本平均数的分布。

表 8-4 当 $n=4$ 时，篮球队员体重样本的样本平均数

篮球队员	体重（磅）
ABCD	224.25
ABCE	249.25
ABDE	246.75
ACDE	242.50
BCDE	249.25
平均数	242.4

最后，对于 5 个运动员的总体，当样本容量 $n=5$ 的时候，只有 1 个可能的样本，就是所有运动员都进入样本。在这种情况下，样本平均数的分布只剩下一个单独的条形（见图 8-5）。另外样本平均数分布的平均数就是总体平均数 242.4 磅。

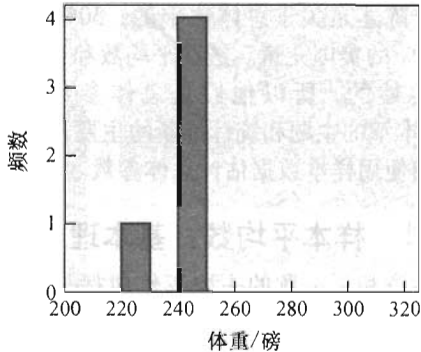


图 8-4 样本容量 $n=4$ 时的抽样分布

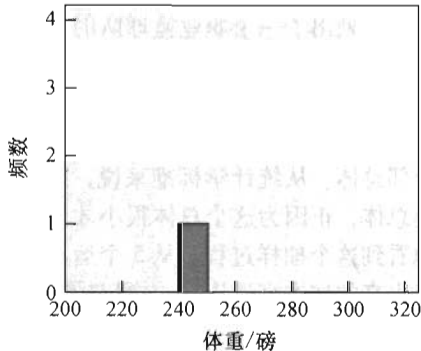


图 8-5 样本容量 $n=5$ 时的抽样分布

方法注解

表 8-2、表 8-3 和表 8-4 包括的样本是不放回抽样的，好处是避免重复，以免同一个项目被抽中两次以上。然而，在统计学上，我们更关注由于以下这些原因放回取样的选择：在大总体中，无论是否放回取样都没有显著差异，但是放回取样的抽样是相互独立的事件，不会受前面结果的影响；独立事件更容易分析，计算它们结果的公式也更简单。

对上述情况进行总结，这一个已知容量的总体所有可能的样本，样本平均数分布的平均数总是等于总体平均数。完整地看一下从图 8-1 到图 8-5 的分布，当样本容量增长的时候，分布围绕着平均数越来越狭窄越来越密集。实际上如果我们观察一个非常大的总体和比较大的样本容量，我们会发现当样本容量增加的时候，样本平均数的分布越来越像一个正态分布（这是中心极限定理的结论，在 5.3 节里所讨论的）。

8.1.2 大总体的样本平均数

在实际的统计研究中，我们很少涉及一个小到只有 5 个个体的总体。为了让探索样本选

取和形成样本平均数分布的过程更符合实际,让我们考虑一些较大的总体。

假定你在一个较小的学院的计算机服务中心工作,为了更好地安排网络使用时间,你调查了学院的全部400名学生,以确定他们每周花费多少小时使用互联网上的搜索引擎。调查结果如下(小时/每周)。

3.4 6.8 6.7 3.4 0.0 5.0 5.4 1.8 0.7 1.6
2.1 3.5 3.4 6.4 7.2 1.8 7.4 3.0 4.0 5.2
1.2 7.8 7.0 0.4 7.2 4.8 3.6 8.0 5.4 6.4
3.5 5.3 4.7 5.4 5.6 3.8 0.1 2.4 0.5 4.0
4.5 8.0 4.2 1.0 6.2 7.1 3.8 0.7 5.5 1.7
2.6 1.6 0.7 1.3 6.5 2.4 3.0 0.3 2.2 0.4
1.9 5.0 2.0 5.3 7.5 5.0 0.3 7.4 6.0 4.3
1.3 0.8 7.2 6.6 0.2 3.4 1.6 2.2 3.0 4.5
5.5 5.3 6.5 0.1 0.3 4.2 2.2 6.2 7.3 3.1
5.4 1.3 6.3 4.5 7.1 5.8 6.1 0.5 0.4 4.1
7.0 6.0 1.1 0.8 1.4 2.9 7.3 0.8 2.7 0.6
3.0 0.7 2.8 6.5 1.9 3.6 1.6 2.6 2.6 6.6
6.8 6.1 3.6 1.4 7.7 5.2 3.8 6.0 2.2 7.5
6.7 4.4 4.1 7.3 5.2 5.7 6.7 2.4 0.6 6.7
1.0 2.3 0.7 1.2 4.5 3.3 4.2 2.1 5.9 3.0
7.2 7.9 2.5 7.1 8.0 6.7 4.1 4.9 0.0 3.1
6.0 0.5 4.2 2.7 0.1 1.4 2.1 2.5 3.9 5.8
5.9 2.7 2.8 3.7 7.3 0.7 6.9 4.4 0.7 1.6
3.1 2.1 7.4 3.6 6.5 2.9 5.4 3.9 3.0 0.8
0.3 0.8 3.3 0.8 8.0 5.6 7.1 1.3 0.2 5.2
7.8 4.7 7.2 0.9 5.1 0.9 1.7 1.2 0.4 6.9
0.6 3.0 3.6 6.1 1.6 6.0 3.8 0.4 1.1 4.0
3.8 4.0 1.8 0.9 1.1 3.9 1.7 1.7 2.6 0.1
4.0 1.4 1.9 0.9 0.2 4.2 4.7 0.2 5.3 2.2
5.8 7.5 5.8 5.2 3.9 3.4 7.3 4.1 0.5 7.9
7.7 7.7 5.0 2.3 7.8 2.3 5.6 6.5 7.9 5.0
2.0 5.5 5.4 6.6 6.7 4.4 7.2 2.5 4.9 7.0
2.1 7.2 4.1 1.2 6.2 3.3 6.3 2.3 4.9 2.2
6.4 7.2 0.1 5.3 3.0 0.7 1.5 1.2 1.1 7.4
5.1 7.2 7.2 3.0 7.1 4.5 6.7 7.2 7.2 0.9
2.9 4.3 2.5 0.7 7.6 3.9 0.7 5.8 6.6 3.4
0.3 6.5 7.5 0.7 6.1 6.1 4.8 1.9 1.9 5.0
1.1 7.8 6.8 4.9 3.0 6.5 5.2 2.2 5.1 3.4
4.7 7.0 3.8 5.7 6.8 1.2 1.7 6.5 0.1 4.3
6.3 1.2 0.8 0.7 0.6 7.0 4.0 6.6 6.9 0.5
4.3 1.0 0.5 3.1 0.9 2.3 5.7 6.7 7.3 0.5
0.3 0.9 2.4 2.5 7.8 5.6 3.2 0.7 5.4 0.0

5.7 0.3 7.2 5.1 2.5 3.2 3.1 2.8 5.0 5.6
3.1 0.7 0.5 3.9 2.6 7.3 1.4 1.2 7.1 5.5
3.1 5.0 6.8 6.5 1.7 2.1 7.3 4.0 2.2 5.6

你可以计算全部400名学生反应的平均数。结果是平均每周3.88小时,这个平均数是真实的总体平均数(population mean),因为它是全部400个学生的总体平均数。我们使用希腊字母 μ 代表总体平均数。同理,我们也可以计算总体标准差结果是 $\sigma=2.40$ 。

在一般的统计应用中,总体是非常巨大的,调查总体中每个个体不太现实或者是非常昂贵的。因此,我们很少知道总体平均数的真实值。这样,考虑使用一个样本平均数去估计全部总体的平均数比较可行。虽然一个样本很容易得到,但是它不可能完全代表总体。这样我们不可能希望从样本平均数中得到的总体平均数的估计是完美的。通过使用一个样本来估计总体引入的误差叫做抽样误差(sampling error)。我们可以通过400个学生互联网使用的样本来分析这个概念。

抽样误差

抽样误差是指一个随机样本被用来估计总体参数时所导致的误差。不包括其他误差来源,比如有偏样本的、不好的调查问题或者记录错误。

思考题

如果样本容量增长的话,你设想抽样误差是增大而减小?请解释原因。

假定你从400个学生的回答中选择一个随机样本 $n=32$,并且计算它们的平均数。例如,这个随机样本可能是:

1.1 7.8 6.8 4.9 3.0 6.5 5.2 2.2
5.1 3.4 4.7 7.0 3.8 5.7 6.5 2.7
2.6 1.4 7.1 5.5 3.1 5.0 6.8 6.5
1.7 2.1 1.2 0.3 0.9 2.4 2.5 7.8

这个样本平均数是 $\bar{x}=4.17$,我们使用标准符号 $\bar{x}$ 来代表这个平均数。我们说这个 $\bar{x}$ 是样本统计量,因为它是来自一个总体中的样本。这样, $\bar{x}$ 叫做样本平均数。

总体和样本平均数的符号

n = 样本容量

μ = 总体平均数

$\bar{x}$ = 样本平均数

我们的目标都是由一个单独的样本平均数来估计总体平均数，但首先应该探讨其他样本。假定你为了研究样本平均数的特点而选择另外一个样本。以下是另外一个这样的样本。

1.8 0.4 4.0 2.4 0.8 6.2 0.8 6.6
5.7 7.9 2.5 3.6 5.2 5.7 6.5 1.2
5.4 5.7 7.2 5.1 3.2 3.1 5.0 3.1
0.5 3.9 3.1 5.8 2.9 7.2 0.9 4.0

这个样本平均数是 $\bar{x} = 3.98$ 。

现在你有两个彼此不相同的样本平均数，并且与总体平均数的真值也不一样。假如你决定选择更多的 32 个人的样本。对于每个样本，都计算样本平均数，为了绘制所有这些样本平均数的图形，可以制作一个显示所有不同样本的样本平均数的直方图。图 8-6 呈现了 100 个不同样本的直方图结果，每一个样本容量都是 32 个学生。注意这个直方图非常接近于正态分布，它的平均数也非常接近于总体平均数。

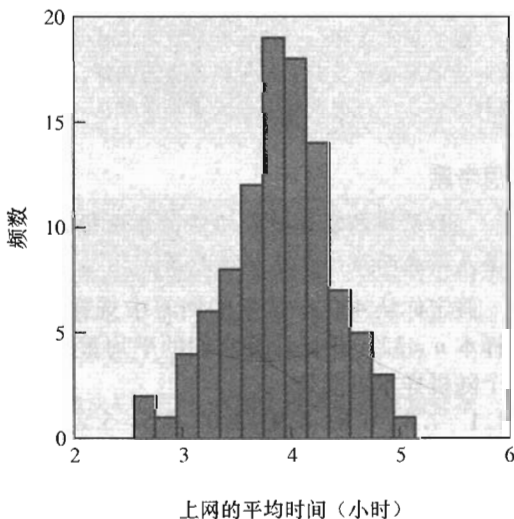


图 8-6 100 个样本平均数的分布，（样本容量 $n = 32$ ）呈现出平均数为 3.88 的正态分布

思考题

假如你只选择容量为 $n = 32$ 的一个样本，按照图 8-6，所选择的样本平均数小于 2.5 或者小于 3.5 的可能性是多大？说明原因。

在前面以 5 个篮球运动员为总体的例子中，

我们可以列出样本容量从 $n = 1$ 到 $n = 5$ 的所有可能的样本。但是对于 400 个人的总体来说，同样的任务是我们所不能完成的，因为你当样本容量 $n = 32$ 的时候，所有可能样本的数量将是一个天文数字（更精确地来说，大约是 2×10^{47} ）。尽管如此，在图 8-6 所呈现的为数不多的样本中，我们也能够看出样本平均数的分布是接近正态分布。进一步说，我们还可以看出样本平均数分布的一个重要性质：样本容量越大，样本平均数的分布越接近于正态分布，样本平均数的平均数等于总体平均数 μ 。

实际上，我们经常只有一个样本可资利用。在这个情况下，样本平均数 $\bar{x}$ 就是总体平均数 μ 的最好估计。幸运的是，如同我们后面所看到的，样本平均数 $\bar{x}$ 与总体平均数 μ 接近的程度是可以用来表示的。

方法注解

一个通常的法则是，如果样本容量大于 30 的话，假定样本平均数的分布是接近正态分布的。

样本平均数的分布

样本平均数的分布是指已知一个样本容量条件下的所有可能的样本，这些样本平均数的分布结果。样本容量越大，样本平均数的分布越接近于一个正态分布。无论何种情况，样本平均数分布的平均数都等于总体平均数。如果只有一个样本可以利用，样本平均数 $\bar{x}$ 是总体平均数 μ 的一个最好估计值。

我们可以扩展这样的观念来看一下正态分布起到什么样的重要作用。再考虑一下图 8-6 中的样本平均数的分布。如果分布包括样本容量 $n = 32$ 的所有可能的样本，那么这个分布将会具有如下这些特征：

- (1) 样本平均数的分布接近一个正态分布。
- (2) 样本平均数分布的平均数是 3.88（等于总体平均数 μ ）。

(3) 样本平均数的标准差有赖于总体标准差和样本容量。总体标准差是 $\sigma = 2.40$ ，样本容量 $n = 32$ ，所以样本平均数的标准差是：

$$\frac{\sigma}{\sqrt{n}} = \frac{2.40}{\sqrt{32}} = 0.42$$

假如我们从前面 400 个学生的反应中随机

选择一个如下的 32 人的样本：

5.8 7.5 5.8 5.2 3.9 3.4 7.3 4.1
0.5 7.9 7.7 7.7 5.0 2.3 7.8 2.3
5.0 6.8 6.5 1.7 2.1 7.3 4.0 2.2
5.6 4.7 5.3 3.5 6.5 3.4 6.6 5.0

样本平均数是 $\bar{x}=5.01$ ，已知样本平均数分布的平均数是 3.88，样本标准差是 0.42，样本平均数 $\bar{x}=5.01$ 的标准分数是：

$$z = \frac{(\text{样本平均数} - \text{总体平均数})}{\text{标准差}} \\ = \frac{(5.01 - 3.88)}{0.32} = 2.7$$

(回忆一下我们用 z 表示标准分数，见 5.2 节。) 我们所选的样本其标准分数 $z=2.7$ ，是指在样本分布的平均数以上 2.7 个标准差。从表 5-1 中我们可以知道，这个标准分对应于第 99.65 百分等级，也就是说选择另外一个平均数小于 5.01 的样本概率是 0.9965。选择另外一个样本平均数大于 5.01 的概率是 $1-0.9965=0.0035$ 。很明显，从这个分布中选择这样的样本是极其罕见的。这个例子显示，如果一个抽样分布符合正态分布的话，我们可以确定这个抽样分布中某个特定样本是很常见还是极其少见。

思考题

假定一个样本平均数正好处于第 95 百分等级。请说明从分布中随机选择另外一个样本，其平均数大于前一个样本平均数的概率是 0.05。

相关知识

美国农场的总数从 1980 年的 2 440 000 个下降到 2000 年的 2 172 000 个，每个农场的平均面积从 1980 年的 426 英亩增长到 2000 年的 434 英亩。



例 8-1 农场抽样

得克萨斯州大约有 225 000 个农场，与美国其他任何一个州都要多。实际上农场的平均面积 $\mu=582$ 英亩，标准差是 $\sigma=150$ 英亩。随机选择一个样本 $n=100$ 个农场，计算出这个样本平均数分布的平均数和标准差。随机选择包括 100 个农场的样本，其平均数大于 600 英亩的概率是多少？

解答 因为样本平均数的分布是一个正态分布，那么它的平均数就和全部总体的平均数是一样，也就是 582 英亩。抽样分布的标准差是 $\sigma/\sqrt{n}=150/\sqrt{100}=15$ 。一个样本平均数为 $\bar{x}=600$ 英亩，那么它的标准分数是：

$$z = \frac{(\text{样本平均数} - \text{总体平均数})}{\text{标准差}} = \frac{(600 - 582)}{15} = 1.2$$

按照表 5-1，这个标准分数处于第 88 百分等级的位置，因此选择一个平均数小于 600 英亩的样本的概率是 0.88，这样，选择一个平均数大于 600 英亩的样本的概率大约是 0.12。

8.1.3 样本比率

我们已经掌握的有关样本平均数分布的更多问题，要留待后面的样本比率的分布来处理。我们可以通过回顾先前所提到的关于 400 名学生的调查来看看类似的问题。假定你的目标是要确定所有 400 名学生拥有汽车的比率。下面的每个学生关于“你是否拥有自己的汽车”这个问题的不同回答“是”(代表拥有)和“否”(代表没有)。

是	否	是	是	否	是	否	否	是	否
是	是	否	否	否	是	是	是	是	否
是	否	是	否	是	是	是	否	是	是
否	是	是	否	否	否	是	是	是	否
是	否	是	否	是	否	是	是	是	是
是	是	否	是	是	否	是	是	否	否
否	是	是	是	否	是	否	是	否	是
否	是	是	是	是	是	否	是	否	否
是	是	是	是	否	是	是	否	否	是
否	是	是	是	是	否	是	否	是	否
是	是	是	否	是	是	否	是	是	否
否	否	是	是	是	否	是	否	是	否
是	否	是	是	是	是	是	否	是	是
否	是	否	否	是	否	是	是	否	否
否	否	否	是	是	是	否	是	否	是

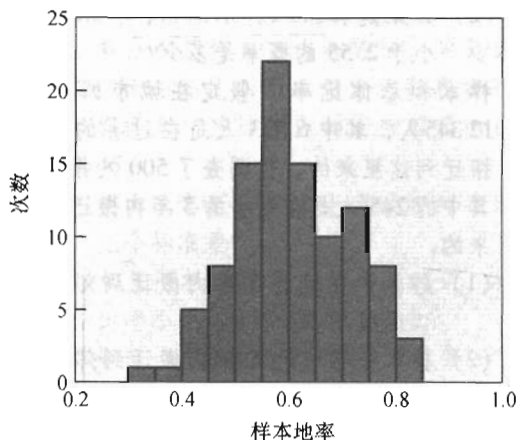


图 8-7 100 个样本比率的分布, 每个样本容量为 32, 形状接近正态分布

样本比率的分布

样本比率的分布是指一个样本容量已知的条件下, 所有可能的样本比率的分布结果。样本容量越大, 分布就越接近于一个正态分布。在所有情况下, 样本比率分布平均数等于总体比率, 如果只有一个可资利用的样本, 样本比率就是总体比率的最好估计值。

例 8-2 一个样本比率的分析

考虑图 8-7 中所呈现的样本比率的分布图。假定它的平均数是 $p=0.6$, 标准差是 0.1。设想你随机选择了如下的 32 人的样本:

是 是 否 是 是 是 是 否
是 是 是 是 是 是 否 是

是 否 是 是 是 否 是 是
否 是 是 否 是 否 是 是

计算这个样本的样本比率 $\hat{p}$, 它离分布的平均数有多远? 如果选择另外一个样本, 其样本比率大于你前面所选择的这个样本的概率是多少?

解答 这个样本中回答“是”的比例是:

$$\hat{p} = 24/32 = 0.75$$

平均数 0.6, 标准差 0.1, 样本统计量 $\hat{p} = 0.75$, 那么这个样本统计量的标准分数是:

$$z = \frac{\text{样本比率} - \text{总体比率}}{\text{标准差}} = \frac{0.75 - 0.6}{0.1} = 1.5$$

样本比率高于分布的平均数 1.5 个标准差, 依照表 5-1, 我们可以知道标准分数 1.5 所对应的百分等级是 93。如果选择另外一个样本, 样本比率小于我们所选择的这个样本, 其概率为 0.93, 这样, 我们选择另外样本比率大于我们所选择样本的概率大约是 $1 - 0.93 = 0.07$ 。换句话说, 如果我们随机选择包含 32 人的 100 个样本, 我们大概可以预期只有 7 个样本比率会高于我们先所选择的那个样本。

方法注解

样本比率的标准差是:

$$\sqrt{\frac{p(1-p)}{n}}$$

如果我们不知道总体比率 p 的取值, 我们可以通过样本比率 $\hat{p}$ 来估计这个标准差。例如, 如果 $\hat{p} = 0.75$, $n = 32$, 我们可

以得到 $\sqrt{\frac{0.75(1-0.75)}{32}} = 0.0765$, 或者约等于 0.1。

8.1 练习题

统计素养和批判性思维

1. 抽样分布 调查者经常采用随机选择的 0~9 数字生成被采访的电话号码。这个随机选择的数字是什么样的分布? 如果我们重复这个过程随机生成 50 个数字, 并且计算它们的平均数, 样本平均数的结果是一个什么样的分布?
2. 最佳估计值 仅仅已知一个样本, 总体平均数的最佳估计值是什么? 仅仅已知一个样本, 总体比率的最佳估计值是什么?
3. 符号 符号 $\bar{x}$ 代表什么? 符号 μ 代表什么? 它们之间有什么差异?

4. 符号 符号 $\hat{p}$ 代表什么? 符号 p 代表什么? 它们之间有什么差异?

判断正误: 判断练习题 5~8 中的陈述是有意义的 (表述正确) 还是没有意义的 (表述错误), 并加以解释。并非所有的陈述都有明确的答案, 所以你的解释比判断更重要。

5. 大样本 托德是一个大学生, 正在完成一个关于估计美国成年人遭遇车祸的比率的项目。它采用了他所在大学的学生的便利样本, 但是他的样本容量很大是 3 000 人。托德认为他的估计是较准确的, 因为他的样本容量比较大足以抵消他所使用便利样本这个情况。

6. 记录误差 当一个调查者不正确地记录被试的反应时, 样本误差就产生了。
7. 大样本容量 使用一个随机样本用来估计总体平均数的时候, 当样本容量增大时, 样本平均数对总体平均数的估计就越来越准确。
8. 民意测验 在盖洛普的一项调查中, 随机选择的 1 012 名成年人中, 其中 901 人表态说, 不应该允许对人类进行克隆, 如果扩大样本容量, 样本比率将很可能大于 $901/1\,012$ (或者 0.890)。

概念与应用

9. 总体比率的估计 在一个对 5 ~ 17 岁儿童的调查中, 从东北部的 9 个州中随机选择了 1 050 名儿童, 在他们的家庭使用的语言不是英语的比率为 0.19。基于这个样本统计量, 对东北部的所有儿童比率的最佳估计是什么? 这个样本比率是美国儿童的最佳估计吗? 为什么?
10. 总体平均数的估计 随机选择 40 位女性, 并检测她们的胆固醇水平, 平均数是 240.9 毫克。基于这个样本统计量, 所有女性的胆固醇平均水平的最佳估计是什么? 如果样本包括对 500 名女性的测量, 你的估计是否会更可信? 请加以解释。
11. 样本量平均数的分布 假定一听可乐充满后的实际容量平均数是 12.00 盎司。随机挑选的 36 听可乐平均容量是 12.19 盎司。假定样本容量为 36 的样本平均数的分布是正态的, 它的平均数是 12.00 盎司, 标准差是 0.02 盎司。
 - (1) 这个样本的平均数距离样本平均数分布的平均数有几个标准差?
 - (2) 通常情况下, 随机选择一个容量为 36 的样本, 其样本平均数至少是 12.19 盎司的概率是多少?
 - (3) 这说明消费者被欺骗了吗? 为什么?
12. 样本平均数的分布 假定你知道容量为 500 的家庭成员数的样本平均数的分布是正态的, 平均数是 2.64, 标准差是 0.06。假如你随机选择了一个样本容量 $n = 500$ 个家庭的样本, 并且计算出这个样本家庭成员数的样本平均数是 2.55。
 - (1) 这个样本平均数距离样本平均数分布平均数有几个标准差?
 - (2) 如果选择第二样本的话, 样本平均数小于 2.55 的概率是多少?
13. 样本和总体比率 假定在城市的郊区有 12 345 人, 其中 6 523 人是在过去的 5 年内搬迁到这里来的。你调查了 500 人并且发现其中的 245 人是在过去的 5 年内搬迁到这里来的。
 - (1) 总体中在过去 5 年内搬迁到郊区的人口比率是多少?
 - (2) 样本中在过去 5 年内搬迁到郊区的人口比例是多少?
 - (3) 你的样本看起来能够代表总体吗? 请加以讨论。
14. 样本和总体比率 假定一所学校里包括 1 608 名学生, 其中 137 名学生是左利手, 你调查了 100 名学生并且在这个样本中发现其中 11 名学生是左利手。
 - (1) 左利手的总体比率是多少?
 - (2) 左利手的样本比率是多少?
 - (3) 这个样本看起来能够代表总体吗? 请加以讨论。
15. 估计总体比率 你从包括 1 608 参加者的医药大会中随机抽取了一个 150 人的样本。在你的样本中, 你发现 73 人是来自国外的。基于这个样本统计量, 估计在这个医药大会中有多少人来自国外。如果你的样本人数变成 300 人, 你的估计是否更可信? 请加以解释。
16. 估计总体比率 从观看迈阿密超级碗比赛的 74 512 名观众中随机选择了 750 个人作为样本。在这个样本中, 420 人支持芝加哥熊队。基于这个样本统计量, 估计观看比赛的人群中支持芝加哥熊队的人数。如果你的样本是 2 000 人, 你的估计是否更为可信? 请加以解释。
17. 样本比率的分布 假定你知道 200 名学生走读生比率的分布是符合正态分布的, 平均数是 0.34, 标准差是 0.03。假如你随机挑选一个包括 200 名学生的样本, 发现样本中走读生的比率是 0.32。
 - (1) 样本比率距离样本比率分布的平均数几个标准差?
 - (2) 如果随机选择第二个样本的话, 样本比率小于 0.32 的概率是多少?

18. 样本比率的分布 假定你知道女性雇员的样本比率分布是符合正态分布的, 平均数是 0.42, 标准差是 0.21。假定你随机挑选一个雇员样本并且计算出的女性雇员的比率是 0.45。

- (1) 样本比率距离样本比率分布的平均数个标准差?
- (2) 如果随机选择第二个样本的话, 样本比率小于 0.45 的概率是多少?

19. 抽样分布 一位橄榄球的四分卫在第一节比赛中打出 1 个截击球, 在第二节比赛中打出 2 个截击球, 在第三节比赛中打出 5 个截击球, 然后就下场休息了。将取值 1, 2, 5 看做一个总体, 假定从总体中随机抽取样本容量是 2 的样本作为总体的代表。

- (1) 列出所有 9 个不同的可能样本, 计算每一个样本的平均数。
- (2) 第 1 小题中所计算出来的样本平均数的平均数是多少?
- (3) 第 2 小题中样本平均数分布的平均数是否等于总体所列出的 3 个数值的平均数? 这两个平均数是否总是相等的?

20. 抽样分布 曾经在军队中任职的一共有 5 位美国总统, 下面是他们就任总统时的年龄: 艾森豪威尔, 62 岁; 格兰特, 46 岁; 哈里森, 68 岁; 泰勒, 64 岁; 华盛顿, 57 岁。假定从上述 5 个就职年龄的总体中随机抽取样本容量是 2 的样本作为总体的代表。

- (1) 列出所有 25 个不同的可能样本的, 计算每一个样本量平均数。
- (2) 第 1 小题中所计算出来的样本平均数的平均数是多少?
- (3) 第 2 小题中样本平均数分布的平均数是否等于总体所列出的 5 个年龄的平均数? 这两个平均数是否总是相等的?

21. 抽样分布的形成 下表是美国的 5 个州和它们的面积 (单位: 1 000 平方公里)。将这 5 个州看做全部总体, 可以从中抽出任何一种样本作为总体的代表。思考样本容量 $n=1, 2, 3, 4, 5$ 时样本平均数的分布。计算各种情况下样本平均数分布的平

均数。样本平均数分布的平均数和总体平均数进行比较。

州	面积 (1 000 平方公里)
亚拉巴马	52
康涅狄格	5
佐治亚	60
缅因	33
俄勒冈	97

22. 抽样分布的形成 在一个练习赛结束以后, 一个曲棍球队的 5 名首发队员得到了如下表的成绩分数。

运动员	成绩
A	101
B	87
C	75
D	66
E	62

思考样本容量 $n=1, 2, 3, 4, 5$ 时样本平均数的分布。计算各种情况下样本平均数分布的平均数。样本平均数分布平均数和总体平均数进行比较。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 8 章的有用链接 “links for internet projects”。

23. 样本平均数的分布 考虑一下学生们花费在互联网搜索引擎上的大量时间, 具体数据已经在本节中列出。讨论从总体中选择一个随机样本的方法, 让班级中的每一个人从总体中随机选择一个样本容量 $n=10$ 的样本, 并且计算他所选择的样本平均数。计算这些样本平均数的平均数并且将它和总体平均数进行比较。然后将样本容量扩大到 $n=20$, 重复上述过程。样本平均数分布的平均数和总体平均数怎么样进行比较?
24. 样本比率的分布 考虑本节中所呈现的 400 个调查问卷 “是” 或者 “否” 的反应 (“你是否拥有一辆汽车”)。讨论从总体中选择一个随机样本的方法。让班级中的每一个人从总体中随机选择样本容量 $n=10$ 的反应样本, 并且计算他的样本中反应为 “是” 的比率。计算这些样本比率的平均数并且将它和总体

比率 0.6 进行比较。然后将样本容量扩大到 $n=20$ ，重复上述过程。样本比率分布的平均数和总体比率怎样进行比较？

新闻中的统计学

25. 新闻中的样本平均数 寻找一种新闻和研究

报告，其中引用了样本平均数。讨论怎么样使用它估计总体平均数。

26. 新闻中的样本比率 寻找一种新闻和研究报告，其中引用了样本比率。讨论怎么样使用它估计总体比率。

8.2 总体平均数的估计

在 8.1 节里，我们看到样本平均数和样本比率的分布是怎样生成的。我们也看到，如果我们从一个总体中随机挑选一个样本，样本平均数就是总体平均数的最佳估计值。但是这个“最佳估计值”是怎么样一个最佳呢？在这一节中，我们来研究如何定量描述，从一个样本平均数中来推断总体平均数的不确定性。

8.2.1 总体平均数的估计：基本原理

动植物会持续生产新的细胞，不仅仅是为了生长，也是为了更替死去的细胞。新细胞的生产过程需要蛋白质，生命组织的基本单元。因此蛋白质是我们饮食中必要的组成部分。那么每天我们应该摄入多少蛋白质呢？许多营养机构和政府机关会提供一个每日膳食建议供给量（RDAS），不仅包括蛋白质，也包括维生素、矿物质、碳水化合物和脂肪。这些每日膳食建议供给量按照来源不同有所区别，并且随着新的研究而随时做出改变。大多数关于蛋白质消费的建议量大约围绕在男性每人每天摄入 55~60 克，女性每人每天摄入 45~50 克（不包括孕妇或者正在哺乳期的女性）。

相关知识

蛋白质的 RDAS（建议饮食量）有时是以每千克体重摄入的克数来衡量的。非运动员和非孕妇的女性一个典型的摄入量数字是 0.8 克/千克体重。

对照给出的我们应该吃多少蛋白质的这些建议，我们非常感兴趣地想知道美国人实际上摄入多少蛋白质。图 8-8 显示了一个有关美国人饮食习惯的大调查的部分结果（Third National Health and Nutrition Examination Survey, or NHANES III, conducted by the National Center for

Health Statistics)。研究的原始数据包括将近 30 000 名参与者，有关他们不同类型的健康和饮食被进行了调查。下面的直方图呈现了平均每天摄入的蛋白质（以克计算），数据来自研究中 $n=267$ 个男性的样本。

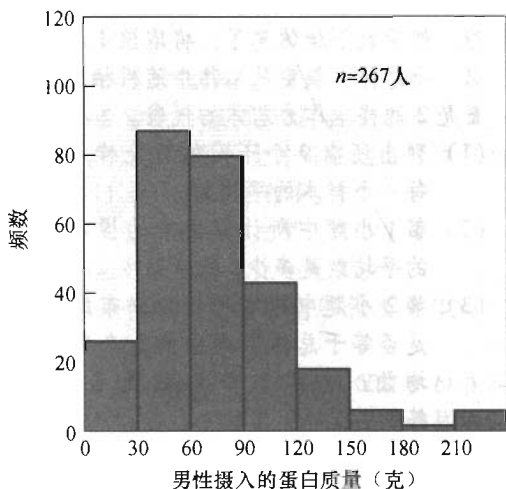


图 8-8 $n=267$ 个男性每天摄入蛋白质的样本直方图
资料来源：National Center for Health Statistics.

我们的目标是使用 267 个男性蛋白质摄入的样本，去推断全部美国男性这个总体的蛋白质摄入的平均水平。这个样本数据的平均数是 $\bar{x}=77.0$ 克，标准差是 $s=58.6$ 克。就像在 8.1 节中所讨论的，当我们只有一个单独的样本时，样本平均数是总体平均数 μ 的最佳估计值。然而，我们不能期望样本平均数等于总体平均数，因为很可能存在一些抽样误差。这样的话，为了推断总体平均数，我们需要一些方法去描述样本平均数对于总体平均数代表性如何。最常见的方法就是使用置信区间的概念。我们在第 1 章讨论了置信区间的用途。现在我们再来看一看置信区间是怎样计算的。

置信区间的概念直接来源于我们所做的抽样分布的工作。一个精确的计算呈现了样本平均数分布是一个正态分布，平均数是 μ ，95% 的

样本平均数位于总体平均数 1.96 个标准差之内,为了与本书的目标相一致,我们将其近似地看做两个标准差。一个置信区间 (confidence interval) 是总体平均数的真值最可能位于的一个值域范围。在本书中,我们主要采用 95% 的置信水平,虽然有时也会采用其他的置信水平 (比如 90% 或者 99%)。我们通过使用误差幅度 (margin of error) 计算样本平均数 95% 的置信区间,就如同下面的内容所定义的一样。

总体平均数的 95% 的置信区间

95% 的置信区间所表达的误差幅度是:

$$\text{误差幅度} = E \approx 2s / \sqrt{n} \quad (8-1)$$

公式中, s 代表样本标准差,我们通过将样本平均数加上和减去误差幅度来计算 95% 的置信区间。也就是说,95% 的置信区间是:

从 (样本平均数 - 误差幅度)

到 (样本平均数 + 误差幅度)

我们可以将置信区间用符号表示为:

$$\bar{x} - E < \mu < \bar{x} + E \quad (8-2)$$

或者简写成:

$$\bar{x} \pm E$$

进行如下仔细的解释:如果我们多次重复选取样本的过程并且建构相似的置信区间,这些置信区间包含总体平均数的可能性是 95%,有 5% 的可能性不包含总体平均数。

例 8-3 误差幅度的计算

计算样本容量 $n = 267$ 位男性的蛋白质摄入量的误差幅度并且计算 95% 的置信区间,样本平均数 $\bar{x} = 77.0$ 克,样本标准差 $s = 58.6$ 克。

解答 样本容量是 $n = 267$, 样本标准差是 $s = 58.6$ 克,因此误差幅度是

$$E \approx \frac{2s}{\sqrt{n}} = \frac{2 \times 58.6}{\sqrt{267}} = 7.2$$

样本平均数是 $\bar{x} = 77.0$ 克,所以 95% 的置信区间的近似值是从 $77.0 - 7.2 = 69.8$ 克到 $77.0 + 7.2 = 84.2$ 克。我们可以将这结果用符号表示成:

$$69.8 \text{ 克} < \mu < 84.2 \text{ 克}$$

或者简写成 77.0 ± 7.2 克。

所有美国男性的蛋白质摄入量,总体平均数的 95% 的置信区间是从 69.8 ~ 84.2 克。值得注意的是,即使是置信区间的下限,也远大于推荐的男性每天蛋白质摄入量 55 ~ 60 克,这说明实际上的蛋白质消费量是显著大于推荐的摄入量。

方法注解

如果我们使用精确公式 (以 1.96 来代替 2), 我们计算出的置信区间的边界值就是 70.0 和 84.0。

方法注解

误差幅度的精确公式是使用 1.96 而不是 2, 同样, 一个中心极限定理的结果样本平均数分布的标准差是 $\sigma / \sqrt{n}$, 但当样本容量足够大时 (一般是 $n > 30$, 但是对样本容量的要求要依据实际的总体分布的性质), 可以用样本标准差 s 代替 σ 。误差幅度的公式也可以扩展到其他置信水平。例如, 对于 90% 的置信水平, 就以 1.645 代替公式中的 2, 对于 99% 的置信水平, 代替 2 的是 2.575。

样本平均数位于置信区间中心, 然后向两个方向各自扩展相同的误差幅度的距离 (见图 8-9)。我们可以说在这个置信区间中包含总体平均数的真值有 95% 的可能性, 这个陈述应该

8.2.2 置信区间的解释

图 8-10 呈现了置信区间的一个直观形象的解释, 假如我们有 20 个不同的样本。每一个样本都有不同的样本平均数, 以及围绕样本平均数的置信区间。我们根本不可能确切地知道哪一个置信区间包括总体平均数。然而就平均而言, 95% 的样本, 也就是 20 个样本中的 19 个, 他们的置信区间捕捉到了总体平均数的真值。偶尔只有一个置信区间 (大约占到这些样本中的 5%) 不能捕捉到总体平均数的真值。

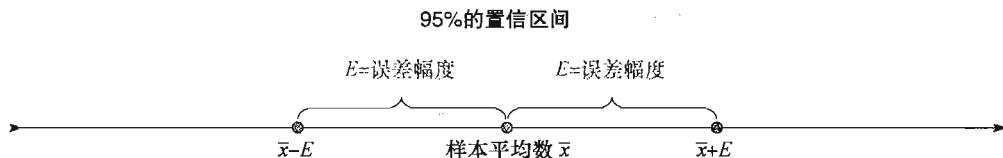


图 8-9 95% 的置信区间是指在样本平均数的两边各扩展一个误差幅度的距离

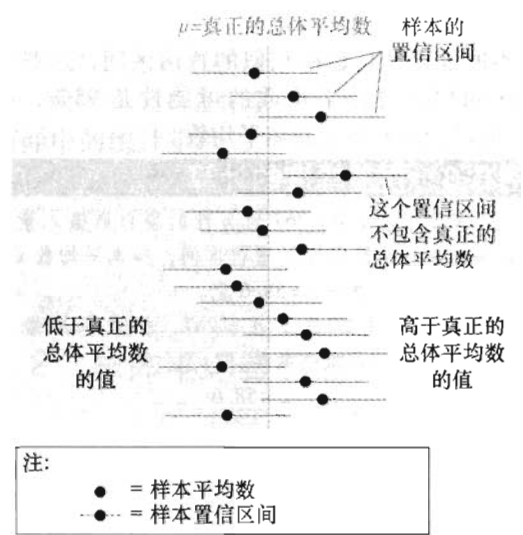


图 8-10 这个图解释了置信区间背后的含义。中心的纵线代表了总体平均数 μ 的真值，20 条横线中的每一条代表了依据每个特定样本所建立的置信区间，置信区间中心的点代表了样本平均数。对于 95% 的置信区间，我们预期所有样本中的置信区间中的 95% 包含总体平均数，如同图中的例子所呈现的，20 个置信区间中的 19 个的确包含了总体平均数，我们预期有 5% 的样本所建立的置信区间不包含总体平均数，在本例中，20 个置信区间中的 1 个（从上面数第 6 个）不包含总体平均数

只要仔细认真些，正确地解释或者表述置信区间是很容易的。看起来是这样的，声称一个 95% 的置信区间意味着总体平均数落进置信区间的概率是 0.95。但是事实上这并不是真实的，即使我们并不知道真正的总体平均数 μ ，它是一个定值而非一个变量，因此没有总体平均数落进置信区间或没有落进置信区间这样的说法。我们不能说总体平均数 μ 落进置信区间的概率。概括地描述一下，95% 的置信区间的正确解释是：如果我们重复选取样本并且建构置信区间的话，从总体上来说，95% 的置信区间将包括总体平均数的真值。

例 8-4 建构一个置信区间

一项研究计算了八年级学生花在看电视上的平均时间是每周 6.7 小时，误差幅度是 0.4 小时（95% 的置信区间）。建构并且解释 95% 的置信区间。

解答 总体平均数的最佳估计值是样本平均数，

$\bar{x} = 6.7$ 小时。我们将样本平均数加上和减去误差幅度计算得到置信区间，所以置信区间的范围是从 $6.7 - 0.4 = 6.3 \sim 6.7 + 0.4 = 7.1$ 小时。这样，我们可以声称八年级学生每周花费在看电视上的平均时间是在 $6.3 \sim 7.1$ 小时，或者说：

$$6.3 \text{ 小时} < \mu < 7.1 \text{ 小时}$$

如果从总体中选取的 100 个样本容量相同的随机样本，我们可以预计其中 95% 的样本的置信区间会包含总体平均数。

例 8-5 女性的蛋白质摄入量

全美国第三次健康和营养调查中，也提供了女性蛋白质摄入量的数据。图 8-11 显示，一个样本容量 $n = 264$ 的女性样本（只是整体研究中很小的一部分）。数据的平均数是 $\bar{x} = 59.6$ 克，样本的标准差 $s = 30.5$ 克，估计总体平均数并且给出 95% 的置信区间，评论这个值域与所推荐的女性每日蛋白质摄入量 45 ~ 50 克相比较的情况。

解答 样本平均数 $\bar{x} = 59.6$ 克，是总体平均数的最佳估计值。为了计算 95% 的置信区间，我们必须首先计算使用样本容量 ($n = 264$) 和样本标准差 ($s = 30.5$ 克) 得到的误差幅度：

$$E \approx \frac{2s}{\sqrt{n}} = E = 2 \times \frac{30.5 \text{ 克}}{\sqrt{264}} = 3.8 \text{ 克}$$

95% 的置信区间是从 $59.6 - 3.8 = 55.8$ 克到 $59.6 + 3.8 = 63.4$ 克，或者

$$55.8 \text{ 克} < \mu < 63.4 \text{ 克}$$

我们也可以说总体平均数（女性蛋白质的每日平均摄入量）95% 的置信区间处于 55.8 ~ 63.4 克，我们可以得出结论说女性每日蛋白质摄入要远大于推荐的女性每日蛋白质摄入量。

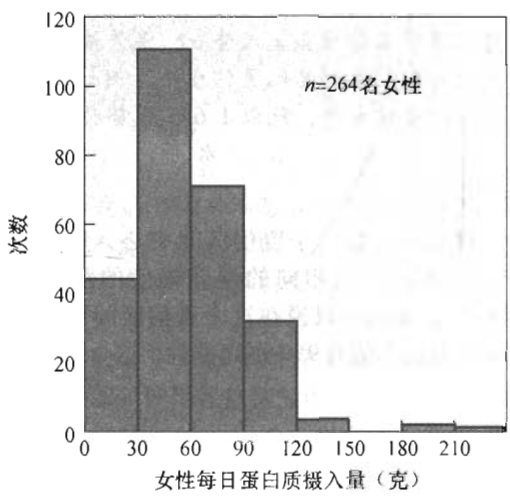


图 8-11 264 位女性每日蛋白质摄入量的直方图
资料来源：National Center for Health Statistics.

思考题

再次回顾男性样本每日蛋白质摄入量的标准差是 $s = 58.6$ 克，几乎是女性样本每日蛋白质摄入量标准差 ($s = 30.5$ 克) 的两倍。这种差异是怎么样影响误差幅度的？推测为什么男女两性标准差是不一样的？

相关知识

按照垃圾工程的数据，大多数家庭垃圾的主要成分是纸（占到垃圾总量的 $1/3 \sim 1/2$ ），接下来就是食品和玻璃，当然，丢弃的塑料和家庭成员的多少是有显著相关的，因此，丢弃的塑料可以用来估计某个地区人口的数量。

例 8-6 垃圾产出

美国亚利桑那州大学一项关于垃圾工程的研究，研究主题是分析 $n = 62$ 个家庭样本的垃圾内容。每个家庭成员人数从 2 ~ 11 人不等。图 8-12 的直方图显示，样本中的这些家庭每周的垃圾产出总量（以磅计）。（全部的研究进一步说明了垃圾的不同细分种类。）样本平均数是 $\bar{x} = 27.4$ 磅，样本标准差是 $s = 12.5$ 磅，估计每周垃圾产出总量的总体平均数 95% 的置信区间，评价这个研究的结论。

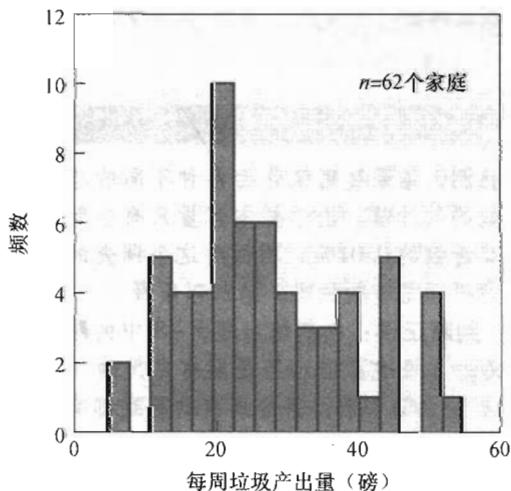


图 8-12 $n = 62$ 个家庭每周垃圾产出总量的直方图

解答 样本平均数 $\bar{x} = 27.4$ 磅，这是总体平均数的最佳估计值。我们使用误差幅度公式来计算：

$$E \approx \frac{2s}{\sqrt{n}} = \frac{2 \times 12.5 \text{ 磅}}{\sqrt{62}} = 3.2 \text{ 磅}$$

这样 95% 的置信区间是从 $27.4 - 3.2 = 24.2$ 磅到 $27.4 + 3.2 = 30.6$ 磅，或者写作

$$24.2 \text{ 磅} < \mu < 30.6 \text{ 磅}$$

在 95% 置信度的条件下，我们可以声称一个美国家庭一周所产出的垃圾数量平均数是 24.2 ~ 30.6 磅。从这个结论中可以观察到两个方面。首先，家庭垃圾产出量的宽泛值域说明了数据中显著的变异性，这种变异性通过大的标准差产生大的误差幅度反映了出来，如果从每个家庭垃圾产出量的角变来看，这个结论可能更有意义。其次，相对较少的样本容量（62 个家庭）也会产生较大的误差幅度，一个更可信的结论应该来自一个更大的样本。

例 8-7 平均体温

美国马里兰大学的研究者所进行的一项试验检测了 $n = 106$ 名被试的身体温度。数据组所得到的样本平均数是 $\bar{x} = 98.20$ °F，样本标准差是 $s = 0.62$ °F。估计身体温度的总体平均数 95% 的置信区间。

解答 样本平均数 $\bar{x} = 98.20$ °F，是身体温度总体平均数的最佳估计值，误差幅度是

$$E \approx \frac{2s}{\sqrt{n}} = \frac{2 \times 0.62 \text{ °F}}{\sqrt{106}} = 0.12 \text{ °F}$$

这样 95% 的置信区间的值域范围从 $98.20 \text{ °F} - 0.12 \text{ °F} = 98.08 \text{ °F}$ 到 $98.20 \text{ °F} + 0.12 \text{ °F} = 98.32 \text{ °F}$ ，或者是

$$98.08 \text{ °F} < \mu < 98.32 \text{ °F}$$

我们对此结果作如下解释：如果我们选取了许多样本容量 $n = 106$ 的不同样本，并且计算所有样本的置信区间。我们预测其中 95% 的置信区间将包含总体平均数的真值。注意通常所引证的人体平均体温（98.6 °F）并不包含在这个置信区间中。基于这样的样本，我们所接受的人体平均温度看起来很可能是错误的。

8.2.3 样本容量的选择

在设计统计调查和实验时，我们经常预先知道可能接受的误差幅度。例如，我们可能想要估计一辆新汽车的平均价格，误差范围在 200 美元以内。只需要通过误差幅度的公式 ($E \approx 2s/\sqrt{n}$) 来确定样本量 n ，通过简单的数学转换，我们可以计算出：

$$n \approx (2s/E)^2 \quad (8-3)$$

准确的样本容量公式需要知道总体标准差 σ ，以此来代替样本标准差 s 。在实际操作中，我们很少会知道总体标准差，因为我们只能研究样本。这样，为了精确地使用样本容量公式，我们经常需要用先前的研究、预备试验或者有根据的推测来估计总体标准差。任何实际的样本容量必须是一个整数，所以我们可以将样本容量公式的结果，取一个最

接近的整数。任何一个样本容量如果大于满足给定的误差幅度要求的样本容量时，误差幅度就会相同或变小。

样本容量的选择

为了以某一种不超过 E 的误差幅度来估计总体平均数，样本容量至少应该是

$$n \approx (2\sigma/E)^2 \quad (8-4)$$

在公式中， σ 是总体标准差（经常用样本标准差 s 来做估计值）

方法注解

给出的样本容量的公式假定我们是以95%的置信水平计算的，如前面所讨论的那样，更精确的公式是使用1.96代替2。对于这一节的方法，如果总体标准差未知，我们可以使用样本标准差 s 来作为总体标准差的估计值。当然，只有当样本容量足够大时这个结果才理想。如果是取自非正态分布的小样本，使用 s 来代替 σ 会导致非常差的结果。

例 8-8 房屋的平均价格

你想通过在不同的地区（代表性地区）抽取最近的房屋销售情况来研究一个国家的房屋平均价格。你的研究目标是提出房屋价格估计值的一个95%的置

信区间。先前的研究提示说总体标准差大约是7 200美元。在下列不同误差幅度情况下，样本容量应该是多少（最少数量）？

(1) 误差幅度在总体平均数真值的500美元以内？

(2) 误差幅度在总体平均数真值的100美元以内？

解答 (1) 当 $E=500$ 美元，总体标准差估计值是 $\sigma=7\,200$ 美元时，满足要求的最少样本容量应该是：

$$\begin{aligned} n &\approx (2\sigma/E)^2 = (2 \times 7\,200/500)^2 \\ &= (28.8)^2 = 829.4 \end{aligned}$$

因为样本容量必须是一个整数，所以我们得出结论说样本容量至少应该是830。

(2) 当 $E=100$ 美元，总体标准差估计值是 $\sigma=7\,200$ 美元，满足要求的最少样本容量应该是

$$\begin{aligned} n &\approx (2\sigma/E)^2 = (2 \times 7\,200/100)^2 \\ &= 144^2 = 20\,736 \end{aligned}$$

值得注意的是，当误差幅度减少到1/5(100~500美元)时，我们必须将样本容量增加到原来的25倍。这就是为什么要提高准确率需要付出很高的成本。

思考题

如果你想得到一个置信区间中较小的误差幅度，你应该增加还是减少样本容量？请加以解释。

8.2 练习题

统计素养和批判性思维

1. 置信区间 基于一个对女性进行测量的随机样本，我们计算出来所有女性的平均胆固醇水平95%的置信区间是：

$$183.3 \text{ 毫克} < \mu < 298.5 \text{ 毫克}$$

解释这个置信区间的含义。

2. 误差幅度 基于一个对男性进行测量的血压的随机样本，样本平均数是73.2mmHg，95%置信区间的误差幅度是2.8mmHg。请确定血压的置信区间。
3. 媒体中的置信区间 下面是一个媒体中的典型陈述：“基于最近的研究，硬币的平均重量是2.5克，误差幅度是0.006克。”这个陈述中遗漏了什么重要的相关信息？使用术语“平均”是否合适？
4. 样本容量 国家健康检测包括对25 000人的

检测，结果被用来作为各种不同的总体平均数的估计值。由于样本容量只有全部美国人口总数的0.01%，因而对这个调查的结果进行批评是有效的吗？请加以解释。

判断正误：判断练习题5~8中的陈述是有意义的（表述正确）还是没有意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. 平均数的估计 一个调查随机选取了200名读者，它用来计算置信区间为47 200美元。
6. 误差幅度 估计高中数学老师的平均收入是48 213美元，误差幅度是5%。
7. 误差幅度 当样本数据被用来估计总体平均数的数值时，当样本容量增加时误差幅度也随之增加。
8. 最佳估计值 当样本数据被用来估计所有硬

币的平均重量时,得到的95%的置信区间是:

$$2.495 \text{ 克} < \mu < 2.505 \text{ 克}$$

基于这个结果,总体平均数的最佳估计值是2.500克。

概念与应用

计算误差幅度和置信区间: 对于练习题9~

12,假设从所描述的样本来估计总体平均数。在每种情况下,用样本的结果近似计算误差幅度和95%的置信区间。

9. 样本容量=100,样本平均数=75.0,样本标准差=10.0。

10. 样本容量=64,样本平均数=2.50,样本标准差=1.08。

11. 样本容量=1 068,样本平均数=46 025美元,样本标准差=24 000美元。

12. 样本容量=692,样本平均数=155千克,样本标准差=27千克。

样本容量: 对于练习题13~16,假定你想计算总体平均数估计值的95%的置信区间。计算估计95%的置信区间条件下特定的误差幅度所需要的样本容量。样本标准差是已知的。

13. 误差幅度=1.0厘米,标准差=3.5厘米。

14. 误差幅度=2.0厘克,标准差=16厘克。

15. 误差幅度=0.01千米,标准差=0.11千米。

16. 误差幅度=24毫米,标准差=416毫米。

17. 电视调查的样本容量 尼尔森媒体研究中心希望估计高中学生每周花费在观看电视上的平均时间。期望的误差幅度是0.25小时。过去的研究表明总体标准差1.7小时是比较合理的。估算对总体平均数进行估计时符合要求的准确度所需要的最小样本容量。

18. 房屋价格的样本容量 政府实施了一项调查来估计大都市的房屋平均价格,期望的误差幅度是10 000美元。预备研究表明总体标准差是65 500美元。估算对总体平均数进行估计时符合要求的准确度所需要的最小样本容量。

19. 估计统计学学生平均智商所需要的样本容量 韦克斯勒智力测验被设计成为:正常成年人的平均智商是100,标准差是15。计算对特拉华州居民的平均智商分数进行估计所需要的最少样本容量。期望在95%的置

信水平上,样本平均数与真值的误差幅度在2个智商分数以内。假设 $\sigma = 15$,计算需要的样本大小。

20. 估计收入所需要的样本容量 某位经济学家想要估计,大学生第一年参加工作后的平均年收入,这些大学生智慧超常并且曾经学习过一门统计课程。如果他想要在95%的置信水平上,样本平均数与总体平均数真值的误差幅度在500美元以内,估计这些大学生的年收入是多少?假定先前的研究已经表明收入的总体标准差 $\sigma = 6\,250$ 美元。

21. 25美分的重量 你想要估计流通中的25美分的重量。一个40枚25美分硬币构成的样本,测得其平均重量是5.639克,标准差是0.062克。采用单个点值来估计所有25美分的平均重量。同时计算95%的置信区间。

22. SAT测验分数 在一所大型大学里的250名大学一年级学生组成的样本,他们的SAT数学测验的平均成绩是1 722,标准差是280。采用单个点值来估计这所大学的全部一年级新生的SAT数学的测验成绩,同时计算95%的置信区间。

23. 毕业时间 来自联邦教育统计中心4 400名大学毕业生的数据显示,毕业并且获得学士学位所需要的时间平均为5.15年,标准差是1.68年,采用单个点值来估计所有大学生毕业所需要的时间,同时计算95%的置信区间。

24. 垃圾产出 基于62个家庭的样本,所丢弃的塑料垃圾的平均重量是1.91磅,标准差是1.07磅(数据来自亚利桑那州大学的垃圾研究项目),采用单个点值来估计所有家庭丢弃的塑料垃圾的平均重量,并且计算95%的置信区间。

25. 熊的重量 黄石国家公园的有关熊的总体健康状况,是通过对熊进行麻醉以后定期测量进行监控。下面是关于熊的重的量的一个样本。计算所有这些熊的重的量的总体平均数估计值的95%的置信区间。

80	344	416	348	166	220	262	360	204
144	332	34	140	180	105	166	204	26
120	436	125	132	90	40	220	46	154
116	182	150	65	356	316	94	86	150

26. 吸烟者的可替宁水平 当人们吸烟时,所吸

收的尼古丁在身体内被转化为可替宁，这是可以测量的。下面列出的是40个吸烟者样本体内的可替宁水平。计算所有吸烟者的平均可替宁水平估计值的95%的置信区间。

1 0 131 173 265 210 44 277
32 3 35 112 477 289 227 103
222 149 313 491 130 234 164 198
17 253 87 121 266 290 123 167
250 245 48 86 284 1 208 173

27. 家庭人口数 你从邻居中随机选择了一个 $n=31$ 个家庭的样本，并且得到下面每个家庭的成员数量（这个家庭的人口数量）：

2 3 6 5 4 2 3 3 1 2 3
2 3 4 5 3 1 3 3 4 7 3
2 3 2 2 3 4 1 5 2

- (1) 这个样本中家庭成员人数平均数是多少？
- (2) 这个样本中家庭成员人数标准差是多少？
- (3) 对于所有美国家庭而言，家庭成员总体平均数的最佳估计值是多少？
- (4) 这个估计的95%的置信区间是多少？
- (5) 评价这个估计的可信度。

28. 电视机 一个样本容量 $n=31$ 个家庭的随机样本，调查这些家庭每家有几台电视机。调查结果如下：

1 0 2 3 2 3 4 2 1 1 2
4 3 2 3 3 0 1 0 1 3 2
4 3 2 1 4 0 1 2 3

- (1) 这个样本中每个家庭平均拥有电视机的数量是多少？
- (2) 这个样本中家庭拥有电视机的标准差是多少？
- (3) 对于所有美国家庭而言，家庭平均拥有电视机数量的最佳估计值是什么？
- (4) 这个估计的95%的置信区间是多少？
- (5) 评价这个估计的可信度。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第8章的有用链接“links for internet projects”。

29. 汽车驾驶年龄 假如你想要估计你所在大学大学生的平均驾龄。先前的研究显示大学生驾龄的标准差是3.7年。如果在95%的置信水平下，样本平均数与总体平均数的误差在一年之内，满足上述条件下随机选择的大学生驾龄的样本容量应该是多少？采用上述的样本容量，选取你自己的样本数据，包括你所在大学大学生的汽车驾龄。然后使用本节所学的方法计算95%的置信区间。然后写下你研究结果的概述。
30. 民意测验机构 美国三个主要的公众民意测验机构分别是盖洛普、哈里斯民意测验调查所、杨凯洛维奇伙伴关系公司。浏览上述机构的网页，描述每一个机构的历史以及它们所提供的民意测验服务。在它们的网页中对它们的民意调查方法说明得最清楚的是哪一个机构？
31. 网络调查 所有主要的电视和网络机构都有很多对各种各样的主题所进行的调查。至少浏览一家主要网络机构并且搜集其某一特定的包括对总体平均数估计内容的调查，并且确定包含样本容量、误差幅度和置信区间这些所有的信息。
32. 国际性贪污腐败 国际透明组织采用调查的方式编制了行贿受贿指数和贪腐印象指数，用来测定世界上许多国家的贪污腐败水平。浏览国际透明组织的网页并且回顾这个组织所采用的扩展性文件描述方法。讨论这些结果以及它们的有效性。

新闻中的统计学

33. 总体平均数的估计 寻找一则新闻报道中的文章，其内容是从一个样本来对总体平均数进行估计的。这篇文章应该包括误差幅度和置信区间。讨论研究中所采用的方法以及研究结论是怎样得出来的。



8.3 总体比率的估计

在这一节中，将我们的注意力转向总体比

率的估计。许多著名的民意测验和调查都基于下面我们所讨论的技术。例如，尼尔森媒体调查公司针对某一广播或者电视节目所进行的总体比率估计，美国劳动统计局所发布的每月的失业者指数就是美国人中失业的人数比率，还

有主导美国人政治的民意调查是对支持某一个特定候选人和政党的人口比例的估计。

8.3.1 总体比率估计的基本原理

美国劳工部统计局基于每月对 60 000 个家庭的调查估计全美的失业率（参考 1.1 节的例 1-2）。这个样本的失业率就是样本比率（失业人数所占样本中全部人数的比率，符号记作 $\hat{p}$ ）。样本比率是总体比率（失业人数在总体全部人数中所占的比率，符合记作 p ）的最佳估计值。

如同总体平均数一样，如果我们谈论起它的准确值，总体比率的估计值也很容易理解。此外，我们也使用误差幅度和置信区间。唯一有别于总体平均数的估计是，有关总体比率估计误差幅度的定义。图 8-13 呈现了我们怎么样解释总体比率估计的置信区间。

总体比率估计的 95% 置信区间

对于总体比率而言，95% 置信区间所对应的误差幅度是：

$$E \approx 2 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (8-5)$$

在公式中， $\hat{p}$ 是样本比率。

95% 的置信区间，其值域范围是

从 $\hat{p} - \text{误差幅度}$ 到 $\hat{p} + \text{误差幅度}$

我们也可以将这个置信区间用公式表述为：

$$\hat{p} - E < p < \hat{p} + E \quad (8-6)$$

方法注解

误差幅度的精确公式是使用 1.96 来代替 2。这个公式也可以进一步扩展到其他置信水平。对于 90% 的置信水平，代替 2 的是 1.645，对于 99% 的置信水平，代替 2 的是 2.575。这里给出的误差幅度公式要求 $np \geq 5$ 和 $n(1-p) \geq 5$ ，这个条件在实际应用很容易满足（涉及适用小样本的方法属于更高级的统计学内容。）

例 8-9 失业率

美国劳动统计局在 $n = 60\,000$ 个人组成的样本中失业人数是 2 160 人。请估计总体失业率并且给出 95% 的置信区间。

解答 样本比率就是样本中的失业率：

$$\hat{p} = 2\,160 / 60\,000 = 0.036$$

这是总体比率的最佳估计值。误差幅度是：

$$E \approx 2 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 2 \sqrt{\frac{0.036(1-0.036)}{60\,000}} = 0.0015$$

（因为这是一个大样本，所以近似值是有效的。）95% 的置信区间其值域范围是从 $0.0360 - 0.0015 = 0.0345$ 到 $0.0360 + 0.0015 = 0.0375$ ，或者 $0.0345 < p < 0.0375$

我们可以说包含总体比率真值的范围是 3.45% ~ 3.75%，得出此结论的置信度是 95%。我们可以对此结果作如下解释：如果选取样本容量都是 $n = 60\,000$ 的很多样本，并且计算其置信区间，我们可以预计这些置信区间中的 95% 会包含总体比率的真值。

相关知识

下面是美国劳动统计局在失业率调查中如何描述不确定性的。“一个样本不是全部，调查也不可能产生如同对全部人口进行访谈般同样的结果。90% 的机会根据样本得来的每月失业率估计值会在总普查可得到 230 000 个可能的图形之内，由于最近的年份里每月的失业率总共的排列方式在 600 万 ~ 1 100 万，抽样可能的误差结果并不会大到歪曲总体失业率图形的程度。”

例 8-10 尼尔森电视收视率

尼尔森电视收视率采用随机的家庭样本。一项尼尔森调查的结果是关于英式足球女子世界杯比赛的，其整体收视率是 72.3%。假定样本包括 5 000 个随机选择的家庭。计算收视率估计值的误差幅度和 95% 的置信区间。

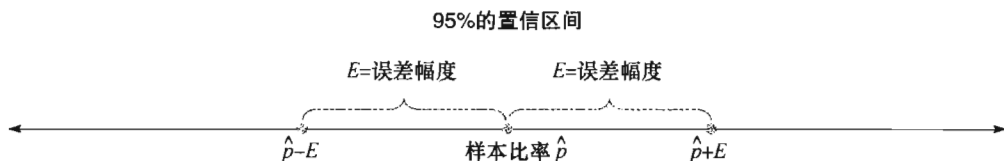


图 8-13 在样本比率两边分别扩展相同大小的误差幅度就构成了置信区间

解答 样本比率 $\hat{p} = 72.3\% = 0.723$ ，是对总体比率的最佳估计值。误差幅度是：

$$E \approx 2 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= 2 \sqrt{\frac{0.723(1-0.723)}{5000}} = 0.013$$

95% 置信区间是： $0.723 - 0.013 < p < 0.723 + 0.013$ 或者写作：

$$0.710 < p < 0.736$$

在 95% 的置信水平上，我们可以得出结论说，英式足球女子世界杯比赛的整体收视率是从 71.0% ~ 73.6%。

例 8-11 盖洛普民意测验

自 1935 年以来，盖洛普调查公司就一直是有关公众态度、观点和行为测量和分析领域的先驱。虽然盖洛普调查公司是以盖洛普民意测验最为著名，但是盖洛普调查公司也为大公司提供市场和经营研究。在近期的一项调查中，盖洛普公司随机选取了 1016 名成年人并且询问他们如下的问题：

“众所周知，前棒球大联盟运动员彼得·罗斯是没有资格入选棒球名人堂的，因为他卷入了棒球比赛的赌球事件。你认为彼得·罗斯是否应该获准进入棒球名人堂？”

在对调查的反应中，59% 的人认为彼得·罗斯应该入选棒球名人堂。样本人群中的一部分 628 人认为自己是“棒球狂热爱好者”。在这群棒球狂热的爱好者中，62% 人认为彼得·罗斯应该有资格入选棒球名人堂。计算每一个样本的误差幅度和置信区间。对于棒球狂热爱好者调查的误差幅度引证说“不超过 5 个百分点”。这个说明与样本容量是否一致？

解答 样本容量 $n = 1016$ ，样本比率 $\hat{p} = 0.59$ ，误差幅度是：

$$E \approx 2 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= 2 \sqrt{\frac{0.59(1-0.59)}{1016}} = 0.031$$

或者说是 3 个百分点。这样，95% 的置信区间其值域范围是从 $59\% - 3\% = 56\%$ 到 $59\% + 3\% = 62\%$ 。

样本容量 $n = 628$ ，样本比率 $\hat{p} = 0.62$ ，误差幅度是：

$$E \approx 2 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= 2 \sqrt{\frac{0.62(1-0.62)}{628}} = 0.039$$

或者说是 4 个百分点。这样，95% 的置信区间其值域范围是从 $62\% - 4\% = 58\%$ 到 $62\% + 4\% = 66\%$ 。

注意所引证的误差幅度“不超过 5 个百分点”，实际上，是估计过高的。

8.3.2 样本容量的选择

调查或民意测验的设计者经常设定结果准确性的某一水平。例如，预期使用 95% 置信区间以及不超过 1.5 个百分点的误差幅度来估计总体平均数。在这种情况下，为了保证准确性，确定样本容量的大小是必需的。当我们使用 95% 的置信度时，我们可以用下面关于误差幅度的简单的近似公式。

$$E \approx 1/\sqrt{n} \quad (8-7)$$

这个公式对于误差幅度的估计是相对保守的（高于必要值）。计算样本容量需要知道误差幅度 E ：

$$n \approx 1/E^2 \quad (8-8)$$

任何大于或等于这个数字的样本容量都可满足对于样本容量的要求。

样本容量的选择

为了在 95% 的置信水平和某一特定的误差幅度 E 条件下估计总体比率，样本容量应该至少是

$$n = 1/E^2$$

方法注解

你可以从前面给出的精确误差幅度公式，用可能最大值 0.25 代替 $p(1-\hat{p})$ 导出 $E \approx 1/\sqrt{n}$ 。这个近似公式高估了实际的误差幅度，并且当 p 接近于 0.5 时，才更准确。

例 8-12 调查的最小样本容量

你设计了一个调查来估计你所在校园里经常使用手机的学生比率。如果你想在 95% 的置信水平和误差幅度不超过 4 个百分点的条件下，在你的样本中至少应该有多少名学生？

解答 注意 4 个百分点意思是指误差幅度为 0.04。根据前面的公式，最小样本容量应该是：

$$n = 1/E^2 = 1/0.04^2 = 625$$

答案是你应该至少调查 625 名学生。

例 8-13 杨凯洛维奇民意测验

杨凯洛维奇伙伴公司是一家国际知名的态度和市场调查公司。公司为《时代周刊》和美国有线电视新闻网络 (CNN) 定期进行调查，调查结果每月被发布在《杨凯洛维奇视野》上。近期的一次调查得出结论说全部家庭中拥有计算机的占到 61%，误差幅度是 3.5 个百分点。请问，在这个调查中样本容量近似

是多少?

解答 误差幅度 3.5 个百分点 (或 0.035), 这样达到这样的目标, 样本容量应该是:

$$n = 1/E^2 = 1/0.035^2 = 816.3$$

因为样本容量应该是整数, 所以我们得出结论所调查的家庭近似为 817 个。

8.3 练习题

统计素养和批判性思维

1. 置信区间 基于癌症研究学会 (ICR) 调查研究所进行的一次调查, 得到以下关于总体比率 p 的 95% 置信区间: $0.457 < p < 0.551$ 。解释这个置信区间的意义。
2. 误差幅度 在一项关于 8 411 次通用航空器紧急迫降的样本中, 飞行驾驶员死亡的比率是 0.052。如果计算所有类似航空事故的总体比率 95% 的置信区间, 误差幅度是 0.005。请确定这个置信区间。
3. 媒体报道中的置信区间 在媒体报道中会有下面这样的典型陈述: “据民意测验结果, 38% 的所有可能的投票者将会投票给迪亚兹。民意测验的误差幅度是 3 个百分点。”在这个陈述中, 遗漏了什么样的重要相关信息。
4. 便利样本 纽波特大学的学生会主席对本校的 3 250 名大学生进行了调查。基于调查结果, 她计算得到所有美国大学生中酗酒者比率在 95% 的置信区间。她声称, 较大的样本容量可以补偿所有被试都是在同一所大学的事实。她这个说法有效吗? 为什么?

判断正误: 判断练习题 5~8 中的陈述是有意义的 (表述正确) 还是没有意义的 (表述错误), 并加以解释。并非所有的陈述都有明确的答案, 所以你的解释比判断更重要。

5. 置信区间 《金士顿纪事报》发表的一篇文章所表述的调查数据计算所得到的置信区间是 0.45。
6. 媒体报道中的置信区间 《金士顿纪事报》发表的一篇文章, 所表述的内容为, 基于调查结果, 82% 的奥兰治县居民反对增加税收, 误差幅度是 4 个百分点。一位读者将这个置信区间表示为 $0.78 < p < 0.86$ 。
7. 样本容量 95% 置信区间其值域范围是 $0.200 < p < 0.400$, 其样本容量为 500。如果样本容量增加, 置信区间就会变小 (或者变得狭窄)。
8. 样本容量 《金士顿纪事报》的一位记者声称, 任何一种好的置信区间其基础是, 样本

容量应该不少于总体容量的 5%。

概念与应用

误差幅度和置信区间: 对于练习题 9~12, 假定总体比率是从所描述的样本估计得来的。在每种情况下, 计算近似的误差幅度和 95% 的置信区间。

9. 样本容量 = 100, 样本比率 = 0.25。
10. 样本容量 = 400, 样本比率 = 0.40。
11. 样本容量 = 1 068, 样本比率 = 0.228。
12. 样本容量 = 1 492, 样本比率 = 0.377。

样本容量: 对于练习题 13~16, 估计达到已知误差幅度要求的最小样本容量。

13. $E = 0.01$
14. $E = 0.03$
15. $E = 0.06$
16. $E = 0.025$
17. 尼尔森电视收视率 尼尔森媒体研究中心使用 5 000 个家庭的样本对电视节目进行排序。尼尔森中心报告说, 15% 的电视观众收看《60 分钟》这档电视节目。这个结果 95% 的置信区间是多少?
18. 尼尔森电视收视率 重复第 17 题的研究, 只是样本容量增加到 10 000 个家庭。这会导致尼尔森调查增加双倍的人物财力的消耗, 样本容量的增长是否与可信度的增加相匹配?
19. 运动员被欺负的现象 阿尔弗雷德大学所进行的一项研究得出结论说, 所有的学生运动员其中 80% 受过某种形式的欺负。这个研究基于对 1 400 名运动员的调查。这个研究的误差幅度和 95% 的置信区间是多少?
20. 学生们的观点 加州大学洛杉矶分校和高等教育研究院每年都要进行一项有关一年级大学生的调查, 调查了大约 276 000 名大学生, 询问他们关于不同事物的态度。按照最近的调查结果, 51% 的一年级大学生认为堕胎应该是合法的 (从 1990 年的 65% 下降到现在的 51%), 40% 的学生认为婚前性行为是可以接受的 (从 1975 年的 50% 下降到现在的

- 40%)。这些估计的误差幅度和95%的置信区间是多少?
21. 各州的生活习惯 疾病控制和预防中心进行了一项调查,比较各州居民的生活习惯和态度。针对下面每一种不同的结果,计算95%的置信区间。
 - (1) 在所调查的样本中,科罗拉多州的居民17.2%是习惯久坐的(在全美中是最低的):样本容量=1 500。
 - (2) 在所调查的样本中,犹他州居民吸烟的比率是13.2%(在全美中是最低的):样本容量=2 500。
 - (3) 在所调查的样本中,威斯康星州的酗酒者的比率为22.9%:样本容量=3 500。
 22. 重要的主题 在美国广播公司/《华盛顿邮报》的一项调查中,随机抽取的1 526名美国人,要求被调查者列出最近选举中最重要的一些主题。结果教育(79%)、经济(74%)和财政预算(74%)被列为最重要的主题。调查的误差幅度是3个百分点。这个误差幅度和样本容量是否一致?
 23. 电影中的毒品 斯坦福大学的研究者为美国国家药物控制政策办公室和卫生部所做的一项研究结果发现,98%的热门租赁电影包含吸毒、饮酒和吸烟的内容。假定这些研究是基于400部热门租赁电影的数据。
 - (1) 使用样本的结果估计所有电影包含吸毒、饮酒和吸烟内容的比率。
 - (2) 95%的置信区间是多少?
 - (3) 你认为这400部热门租赁电影代表了一个随机样本吗?请加以解释。
 24. 青少年压力 一项美国教育部委托进行的研究,结果发现44%的青少年感到学习成绩是他们最大的压力来源。这些研究基于对1 015名青少年的调查。95%的置信区间是多少?
 25. 预测选举结果 在一个大城市中对随机抽取的1 600人进行调查,结果发现在即将到来的选举中其中的900人支持现任市长。基于这个样本数据,你能否认定现任市长将会在接下来的选举中胜出?请加以解释。
 26. 预测选举结果 在一个大城市中对随机抽取的2 500人进行调查,结果发现在即将到来的选举中,其中的1 300人支持现任市长。基于这个样本数据,你能否认定现任市长将会在接下来的选举中胜出?请加以解释。如果在250个人的样本中其中130人支持现任市长,前述的结论能否扩展到这个样本中。
 27. 选举前的民意测验 在全美的参议员大选之前,实施了三个民意测验。在第一个民意测验中,1 500名投票人中的780名支持候选人马丁内斯;在第二个民意测验中,2 500名投票人中的1 285名支持候选人马丁内斯;在第三个民意测验中,3 500名投票人中的1 802名支持候选人马丁内斯。分别计算三个民意测验中95%的置信区间。基于这些民意测验讨论马丁内斯竞选的前途。
 28. 失业率调查 美国劳动统计局通过每月调查60 000人来估计全美的失业率。
 - (1) 在第一个月,发现被调查的60 000人中的3.4%处于失业状态。计算这个估计的误差幅度。这个精确度是否合理(10个百分点)?请加以解释。
 - (2) 假如所调查的个体人数增加到4倍(增加到240 000人),误差幅度将会发生什么样的变化?
 - (3) 假如所调查的个体人数减少到原来的1/4(减少到15 000人),误差幅度将会发生什么样的变化?
 29. 态度调查 一个态度调查发现总体中的54%支持总统的工作计划,调查的误差幅度是4%(假设是在95%的置信水平)。
 - (1) 赞成总统工作计划的总体百分数真值的95%的置信区间是多少?
 - (2) 这个调查的样本容量是多少?
 30. 夹藏武器 Talmey-Drake 调查战略公司对626位科罗拉多州居民进行了调查,其中2/3(或者66.7%)说他们支持立法机构的一个待表决的法案,这个法案将授予枪支拥有者随身携带武器的许可。这个法案将强迫当地执法机关授予此类许可证。报道中说这个调查的误差幅度为4个百分点。
 - (1) 报道中所提到的误差幅度与样本容量是否一致?
 - (2) 如果要求误差幅度为2个百分点,样本容量应该是多少?

网络和远程项目

请在网站 www.aw.com/bbt 上打开第8章的有用链接“links for internet projects”。

31. 谁是副总统？假定你想估计你所在大学的大学生能够正确说出美国副总统名字的人数比率。如果在95%的置信水平并且样本比率与总体比率的误差在0.1之内，你必须随机选取多少大学生？采用这样的样本容量，随机选取你所在大学的大学生进行调查并得到你自己的样本数据。使用本节所学习的方法计算95%的置信区间。并且写出有关你的调查结果的概述。
32. 尼尔森媒体研究方法 浏览尼尔森调查机构的网页，并且报告实际上用来估计总体尼尔

森电视收视率及计算其置信区间的方法。

33. 网络调查 所有主要的电视和网络机构都会在大量的主题上定期进行调查。浏览主要网络机构的网页并且搜集某一个特定调查的结果，内容包括对总体比率的估计，并且确认包括以下所有信息，诸如样本容量、误差幅度和置信区间，了解实际调查过程的所有细节。

新闻中的统计学

34. 总体比率的估计 寻找一则近期的新闻报道文章，内容包括从一个样本对总体比率进行估计。文章应该包括一个误差幅度和一个置信区间。讨论在研究中所使用的方法，并且分析是如何得出结论的。

第8章复习题

1. 在对药物 Ziac 进行的一项临床研究中，221名 Ziac 的使用者感到头晕（数据来源于美国立达公司实验室）。
 - (1) 使用这样的结果，建立 Ziac 的使用者感到头晕的总体比率估计值的95%水平的置信区间。
 - (2) 写出第1小题置信区间的正确解释。
 - (3) 误差幅度是多少？
 - (4) 样本比率0.032来自由221名被试组成的一个特定样本。假定可以选取都是由221名被试组成的许多不同的样本，并且得到每个样本中感到头晕的人数比率。那么这个样本比率的抽样分布是一个什么形态的分布？
 - (5) 如果增加样本容量到大于221名被试，这将会对置信区间的范围产生什么样的影响？这个范围将变得更宽泛还是更狭小，或是没有发生变化呢？
 - (6) 使用另外一个不同的样本数据，我们计算得到的95%水平的置信区间是 $0.400 < p < 0.500$ 。下面这个解释错在哪里：“总体比率落在 $0.400 \sim 0.500$ 之间的机会是95%”？
2. 我们想针对大学生总体进行斯坦福-比奈智力测验平均智商分数的估计。我们知道对于一般人群的总体，斯坦福-比奈测验智商分数的标准差是16。
 - (1) 使用标准差16，如果我们想在95%的置

信水平上，样本平均数与总体平均数的误差在3个智商分数以内，我们应该随机选择多少名大学生进行斯坦福-比奈智力测验？

- (2) 如果样本容量大于所必需的，这对我们估计大学生的平均智商会产生什么样的影响？如果样本容量少于所必需的，又将产生什么样的影响？
- (3) 在实际测量中，大学生的智商分数标准差更可能等于16，还是大于16，或者是小于16呢？请加以解释。如果我们采用实测值来代替第1小题中所用到的智商标准差16，这对于第1小题的答案会产生什么样的影响？所需要的样本容量是同样的，或者是少一些，还是更多一些？
3. 随机选取包含50名被试的样本，并且测量每名被试的白细胞数量。所得结果平均数是7.1，标准差是2.5。
 - (1) 使用这个样本结果来建立总体平均数估计值的95%水平的置信区间。
 - (2) 写出第1小题置信区间的正确解释。
 - (3) 误差幅度是多少？
 - (4) 如果标准差是2.5，如果我们想在95%的置信水平并且要求样本平均数与总体平均数的误差不超过0.25，我们应该选取多少被试？
4. (1) 你被英特尔公司雇用来调查那些计划升

级一个新操作系统的计算机拥有者的特点。假定你想在 95% 的置信水平上并且要求样本比率在总体比率真值的误差在 0.02 以内，你必须至少要调查多少人？

(2) 假如与在第 1 小题中进行的调查所描述的一样，你发现被访问的调查者有一半

拒绝回答调查问题，因为他们认为你想给他们推销东西。如果你通过对许多人打两次电话以便保证样本容量足够大，这样的调查结果是否恰当？请加解释。

第8章小测验

1. 如果从一个成年女性心率的总体中抽取许多样本容量为 100 的不同随机样本，样本平均数的分布将会是什么形态？
2. 如果从一个大学生总体中抽取许多样本容量为 500 的不同随机样本，女大学生比率的分布将会是什么形态？
3. 符号 $\hat{p}$ 代表什么？
4. 一篇期刊上的文章以 0.60 ± 0.08 这样的格式提供了置信区间。采用 $a < p < b$ 这样的格式来表达置信区间。（也就是，用某一特定值来替代公式中的 a 和 b 。）
5. 假如我们想要估计美国成年男性的握力平均水平。如果我们得到一个关于成年男性握力的随机样本，下面哪一个统计量是总体平均数的最佳估计值？
 - (1) 样本的中位数
 - (2) 样本平均数
 - (3) 样本标准差
 - (4) 样本全距
 - (5) 样本比率
6. 计算与 95% 水平的置信区间 $0.240 < p < 0.280$ 相对应的误差幅度。
7. 计算与 95% 水平的置信区间 $240 < p < 280$ 相对应的误差幅度。
8. 当建立总体平均数的 95% 水平的置信区间时，如果一个样本平均数是 4.60，误差幅度是 0.15。请确定总体平均数 95% 水平置信区间。
9. 当建立总体比率 95% 水平的置信区间时，如果一个样本比率为 0.600，误差幅度是 0.200。请确定总体比率 95% 水平的置信区间。
10. 说明总体比率 95% 水平的置信区间为： $65.0 < p < 70.0$ ，它的错误在哪里？

技术应用

为了描述计算总体平均数估计值置信区间的方法，本章主要只涉及了正态分布应用的案例。你会发现第 10 章介绍总体平均数估计值的置信区间时，会应用到 t 分布。在第 10 章的结尾你会发现应用这些方法的过程。

SPSS

平均数的置信区间：SPSS 并没有一个过程是用来在总体标准差已知的条件下，得到总体平均数估计值的一个置信区间。在第 10 章的结尾处，你可以看到当总体标准差未知的条件下，如何进行总体平均数估计值置信区间计算方法应用的介绍。

比率的置信区间：SPSS 并没有一个过程是用来可以得到总体比率估计值的一个置信区间的。

Excel

总体平均数估计值的置信区间：使用数据

工作台 Data Desk/XL，将它安装于 Excel。

首先进入样本数据列 A。如果使用 Excel 2003，点击 [DDXL]。如果你使用 Excel 2007，请点击 [Add-in]，然后再点击 [DDXL]。选择 [Confidence Intervals]。在功能键中，选择 [1 Var z Interval]。然后点击铅笔图标，输入数据列，比如 A 数据列的 A1:A12，然后点击 [OK]。在对话框中，选择置信水平，并且输入已知的总体标准差的数值，然后点击 [Compute Interval]，这样置信区间就会呈现出来。

总体比率估计值的置信区间：使用数据工作台 Data Desk/XL，将它安装于 Excel。

首先在单元格“A1”输入成功的次数，然后在单元格“B1”中输入实验的总次数。如果使用 Excel 2003，点击 [DDXL]。如果你使用

Excel 2007, 请点击 [Add-Ins], 然后再点击 [DDXL]。然后选择 [Confidence Intervals], 选择 [Sumn 1 Var Prop Interval] (这是一种“使用一种变量的概括数据来计算总体比率置信区间”的小型公式)。然后点击内容为“Num Successes”(“成功次数”) 铅笔图标并且输入“A1”, 然后点击内容为“Num Trials”(“试验总次数”) 并且输入 [OK]。在对话框中, 选择置信水平, 然后点击 [Compute Interval],

这样, 置信区间就会呈现出来。

STATDISK

选择 [Analysis] 然后再选择 [Confidence Intervals]。然后选择 [Mean-One Sample] 或者 [Proportion-One Sample]。在出现的对话框中, 首先输入一位小数作为显著性水平。输入 0.95 或者 95% 作为置信水平。接下来输入其他所需要的项目。然后点击 [Evaluate], 置信区间就会呈现出来。

| 聚焦历史 |

统计学的起源

许多学科的起源在古代就已经失传了, 但是统计学的根源能够由一些明确的事实而得以确定。保存系统的记录开始于 1532 年的伦敦, 是每周一次收集的死亡数据。同一年代的后期, 法国官方开始收集关于出生、死亡和婚姻的数据。1608 年, 瑞典开始收集类似的重要统计数据。而在 1666 年加拿大则进行了第一次的官方普查。

当然, 统计学不仅仅是收集数据。如果存在一个统计学的创始人的话, 那么此人必须能够以灵活和系统的方式使用这些数据, 并且能够得出在此之前并不明显的结论。许多专家认为一位名叫约翰·格兰特 (John Graunt) 的英国人应得到统计学创始人的头衔。

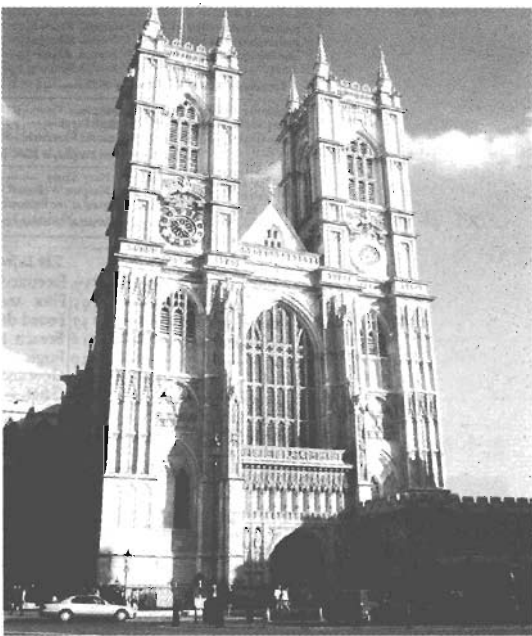
约翰·格兰特 1620 年出生于伦敦。在他的大家庭里, 他以老大的身份接管了他父亲 (一位服装和纺织品商人) 的生意。他生命的大部分时候是一位杰出的伦敦公民, 直到他在 1666 年的伦敦大火中失去他的房子和财产。8 年之后, 他死于贫困之中。

不知道从何时开始, 约翰·格兰特对每周的出生和死亡的记录 (著名的死亡率表) 非常感兴趣, 自 1563 年开始他的这些记录就在伦敦被保存下来。在《自然与政治观察: 死亡率表》一书的序言中, 他注意到其他人“几乎不使用这些数据”, 并且很想知道“与此相同的一些知识能给世界带来什么好处。”在他的书于 1662 年出版之前, 他已经在他的统计课题上花费了数年时光。

约翰·格兰特的主要工作是对死亡率表进行每年度的年终总结。图 8-14 是 1665 年的年度表 (疫病大流行的一年)。表上面 1/3 部分的内容呈现了每个教区的埋葬和接受洗礼的人数, 埋葬和接受洗礼的总人数则标注于表中间的 1/3 部分, 以及被单独分离出来的死于瘟疫的记录。表下面的 1/3 部分则呈现了由于各种各样的原因死亡的人数, 以及男性和女性的死亡总数。

约翰·格兰特知道因为收税的目标而对伦敦人口总数所做的粗略估计, 但是他怀疑伦敦人口在 1661 年有 600 万或者 700 万的这个估计数字。使用上述这个年度表, 比较死亡和出生人口、估计伦敦家庭的密度 (平均每个家庭有 8 名成员), 他通过这 3 种不同的方法估计伦敦人口有 460 000 人, 这与 600 万或 700 万这个数字有很大的不同。他也发现当农村人口下降的同时, 伦敦市区的人口却在增加, 呈现了一个早期城市化的趋势。并且他高度关注了婴儿的高死亡率, 同时反驳了伴随新国王继位会降临瘟疫的人口理论。

约翰·格兰特最重要的贡献或许是他构建了第一个生命表。虽然详细的死亡年龄数据并没有什么用处, 但是他知道关于 100 个新生儿的数据, “他们中的 36 个孩子在 6 岁前就死去了, 但或许有一位到 76 岁依然健在。”他用这样的数据点填充了如表 8-5 所示的期间各个年份, 但是方法他没有做过多解释。



在对各个死亡年龄估计的基础上，约翰·格兰特制作了如表 8-6 的生存年限表。尽管当代一些统计学家对构建这些表的方法曾表示怀疑，但约翰·格兰特好像更重视这些表的价值，并且期望将来的人寿保险公司可以使用保险预测精算表。但是直到 1693 年，埃德蒙·哈雷（Edmund Halley），与彗星同名的一个人，才使用以年龄为基础的死亡率所构成的生命表。

表 8-6 格兰特的不同年龄生存人数表

年龄（岁）	生存者数量	年龄（岁）	生存者数量
16	40	56	6
26	25	66	3
36	16	76	1
46	10	80	0

问题讨论

- 1. 你会对准确估算的 1660 年伦敦人口是无效的感到惊奇吗？在那个时候得到这样的估计，你有什么建议？
- 2. 你认为埋葬和接受洗礼的记录能得出实际关于出生和死亡的准确数目吗？为什么？
- 3. 如果表 8-5 关于死亡的估计是准确的，那么表 8-6 生存年限的估计和它是一致的吗？注意表 8-6 的数量总和不是 100，请加以解释。
- 4. 根据约翰·格兰特生命表，估计 1660 年人们的平均预期寿命，并解释你的理由。

建议阅读

1. Johnson, N., and Kotz, S. (Eds.) .*Leading Personalities in the Statistical Sciences*. John Wiley and Sons, 1997.

2. Sutherland, I. "John Graunt; A Tercentenary Tribute." *Journal of the Royal Statistical Society*, Vol. 126, Pt. 4, 1963, pp. 537-557.

| 聚焦文化 |

莎士比亚总共认识多少单词

想象一下你参加一个为国际学生举办的东方式聚会。在聚会上你遇到了 12 个瑞典人、9 个中国人、6 个法国人、4 个以色列人、3 个韩国人和 1 个意大利人。你也知道有一些其他国家的人，只是没在这个聚会上中遇到。如果仅仅以你遇到的人为基础，有没有可能会估计到出席聚会人员的来自国家总数（其中包括你没有遇到的国家的人）？正因为有了抽样的观念，所以这个回答是肯定的。

聚会的问题意义可能不太明确。但是，当牛津大学海洋生物学家查尔斯·帕克斯顿（Charles Paxton）想知道到底有多少没有被发现的“海怪”（体型超过 2 米的海底生物）时，提出的是与上述聚会的例子本质相同的问题。在这种情况下，参加聚会人员来自的不同国家的数字相当于“海怪”的类型数量。运用统计方法，除了已知的大概 220 多种“海怪”以外，查尔斯·帕克斯顿能够估计出尚未被发现的“海怪”有 47 种。

同样的方法已经被用来分析莎士比亚总共认识多少个单词。统计学家布拉德利·埃弗龙（Bradley Efron）和罗纳德·齐斯泰兹（Ronald Thisted）想知道莎士比亚总共认识单词的确切数量，这些单词一定比在他写作中用到的单词数量多。在这里，参加聚会人员来自的不同国家数量，相当于莎士比亚戏剧和诗歌中出现的不同单词数量。在聚会上收集的数据可以看成是第一个样本。就莎士比亚来说，第一个样本由莎士比亚一部全部公开发表的作品



构成，更具体的数据就是单词在这些作品中被使用了1次、2次、3次等出现的次数。表8-7显示了第一个样本的一（小）部分数据。例如，此表显示出在莎士比亚的作品出现几次的单词量，准确地说，14 376个单词只出现了一次，4 343个单词出现了两次等。（整个表格更完整，一直继续到出现10次以上的单词数量。）

表8-7 莎士比亚所有作品中使用1~10次的单词数量

使用次数	单词数量	使用次数	单词数量
1	14 376	6	837
2	4 343	7	638
3	2 292	8	519
4	1 463	9	430
5	1 043	10	364

依据这个表格作为第一个样本，现在我们可以假设一个问题。假定选取莎士比亚不同于第一个样本的新的第二个作品样本，它和第一个样本的容量同样大小。我们预期可以找到多少没有在第一个样本中出现的新单词？我们预计在第二个样本中会有比较少的新单词出现，因为在第一个样本中每个第一次出现的单词都是新的，即使是像“这”这样特别常用的单词。在第二个样本中这些特别常用的单词就不再是新的。布拉德利·埃弗龙和罗纳德·齐斯泰兹估计有11 430个没有出现在第一个样本中的单词出现在第二个样本中。

如果他们重复选取第三个样本、第四个样本、第五个样本等。在每一个新的样本中，新单词的数量在减少，但是所有样本中使用单词的总数却在增加。布拉德利·埃弗龙和罗纳德·齐斯泰兹最终发现新单词的数量大约是35 000个。这意味着除了莎士比亚认识并使用的31 534个单词之外，还有大约35 000个单词他认识但并没有使用。因此，他们估计莎士比亚大约认识66 500个单词。

布拉德利·埃弗龙和罗纳德·齐斯泰兹这种必要而先进的分析方法在1976年开始使用。10多年后，当发现了一首莎士比亚时期的不知道作者的新十四行诗时，这种方法被应用于分析作者是否是莎士比亚。如前面所描述的一样，莎士比亚全部作品中的单词被当做第一个样本。这里，第二个样本就是包括429个单词的这首十四行诗。这种统计方法被用于预测这首诗的作者是谁，如果这是莎士比亚的作品，这首诗中应该有7个单词在莎士比亚的全部作品中没有出现过。实际上，这首新诗有9个单词在以前的莎士比亚的作品中没有出现过。同理，这种方法预测这首新诗应该有4个单词在莎士比亚的全部作品中只出现一次。事实上，有7个单词在莎士比亚的全部作品中只出现1次。预测在全部作品中只出现2次的单词有3个，事实上有5个这样的单词。最后作者认为在这首十四行诗中，预测单词所出现的次数和实际上单词所出现的次数比较相一致，这足以把这首诗归为是莎士比亚的作品。

这些统计方法用于区分，莎士比亚与伊丽莎白一世时代的像马洛、唐纳和琼森等这些作家的作品。它们还被用于强化詹姆斯·麦迪逊确实是《联邦党人文集》的作者身份的这个信念，甚至将此统计方法用于确定柏拉图作品的顺序。

这个例子说明统计学在解决表面上看起来没有相关的学科问题上，具有非凡的适用性。也许更重要的是，它能证明像文学和海洋生物学这两个遥远的学科，也可以通过普遍适用的统计思维将其联系起来。

问题讨论

1. 评论你是否相信统计分析可以促进文学发展。例如，你会由于知道莎士比亚认识多少单词而更加佩服莎士比亚吗？
2. 你认为统计分析的方法对鉴定“匿名”作者的作品是有用的吗？
3. 建议其他学科（除以生物学和文学以外），比如那些在聚焦部分所涉及的，采用统计方法来估计不确定的数量。

建议阅读

1. Efron, B., and Thisted, R. "Estimating the Number of Unknown Species: How Many Words Did Shakespeare Know?" *Biometrika*, Vol. 63, No. 3, 1976, pp. 435-437.

2. Morin, Richard. "Unconventional Wisdom: Statistics, the King's Deer, and Monsters of the Sea." *Washington Post*, March 7, 1999.

3. Snell, Laurie. *Chance News* 8.03.

有教养的标志是你能够接纳一个你并不接受的观点。

——亚里士多德 (Aristotle)

第9章 假设检验

学习目标

■ 9.1 假设检验的理论基础

理解假设检验的基本目标,掌握一个假设检验的基本结构,包括如何设置虚无假设和备择假设。如何界定假设检验的可能结果,如何在这些可能的结果之间做出统计结论。

■ 9.2 总体平均数的假设检验

理解和说明总体平均数的单侧和双侧假设,认识和避免推断统计中假设检验中的统计错误(包括 I 型错误和 II 型错误)。

■ 9.3 总体比率的假设检验

理解和说明关于总体比率的假设检验。

每个人都会做出一些声明:广告商会声称他们的产品是多么优秀,大学会声称他们的课程是如何精良,政府声称他们的项目计划是如何有效,律师会证明他们为之辩护的嫌疑人是有罪的或者是清白的,临床诊断会告诉被检查者有无疾病,制药公司会声称他们的药物是如何有效。但是我们怎么才能知道这些声明是真实的?统计学提供了一个检验方法,这是一种非常强大的技术,我们称之为假设检验。

相关知识

性别选择产品包括印刷的指导材料和体温表,它依据的理论是一个婴儿的性别可以通过仔细定时的概念决定,没有证据证明这种方法真的有效,并且它也不再被应用。然而,市场上还有其他产品,看起来有些效果,由此引出了道德问题,性别选择是否应该或者何时可以成为父母的一种选择。



9.1 假设检验的理论基础

一个名叫 Pro Care 的公司,曾经声称它的产品能够增加一个妇女受孕生女儿的机会——可以使得生女儿的机会“增长到 80%”。让我们来看看这个声称,它声称某种产品可以使生女儿的机会增加,大于通常情况预期的近似 50% 的机会。怎样才能检验这个改变性别选择概率的声称是真实的?

一种方法就是研究随机样本，比如说，使用这种性别选择产品的女性产下了 100 个婴儿。如果这种产品没有效果，我们可以预期这些婴儿之中有一半是女孩；如果这种产品有效，我们可以预计这些婴儿中女孩的比例明显超出一半。关键问题是，什么才是“显著地超出”。如果样本中 100 个婴儿包括 97 个女孩，我们可能都同意女孩人数明显地超出一半，这种产品是有效的。如果样本中 100 个婴儿包括 52 位女孩，我们可能会认为女孩人数非常接近一半，因而没有理由相信这种产品是有效的。但是当情况是这样时，比如说，100 个婴儿的样本中包括 64 个女孩，能否说女孩人数“显著地超出了”总婴儿数的一半，同时能否认为这种药物真的有效？

在统计学中，我们通过“假设检验”方法来回答这样的问题。在我们探讨假设检验的特定术语和步骤之前，先让我们来看一下，使用这种产品女性所产下 100 名婴儿的样本中，其中 64 名是女孩，我们是如何得出关于性别选择结论的过程。

1. 我们先假定性别选择没有效果，它并不能增加女孩的百分比。如果这是真实的，我们可以预计使用这种产品的女性会产下大约 50% 的女孩。也就是说，女孩的出生比率等于 0.50。

2. 现在使用样本（100 个婴儿中包含 64 个女孩）来检验上述的假设。我们通过计算随机抽取 100 个婴儿其中 64 个是女孩的可能性（也就是说，从一个女孩比率为 50% 的婴儿总体中，选取的女孩比率大于或等于 64% 的概率），来建立这个检验过程。

3. 如果计算得出，从一个随机样本中选取 64%（或者更多）的女孩的可能性比较大，那么我们就不能证明性别选择发生了效果。然而，如果计算得出，几乎不可能从一个随机样本中选取至少 64% 的女孩，那我们就可以得出结论说，在这个进行了性别选择的样本中，可能存在某些原因导致女孩出生的机会发生了改变，这也意味着产品可能是有效的。值得注意的是，尽管如此，这种情况并没有证明产品是有效的，仍然有可能是其他原因导致的结果（比如我们选择了一个不寻常的样本，或者可能包含我们没有考虑到的混淆变量）。

思考题

假如性别选择是有效的，你随机选择了一个使用这种产品的女性所产下的 1 000 个婴儿的新样本。基于前面的样本（100 个婴儿之中包括 64 位女孩），你预计在新样本中会有多少个女孩？接下来，假定性别选择是无效的，在同样的随机选择的使用这种产品的女性所产下的 1 000 个婴儿的新样本中，你预计在这个新样本中会有多少个女孩？

9.1.1 假设检验的公式

检验性别选择有效性的其他关键问题在于，我们既不能描述怎样来计算抽取一个随机样本的可能性，比如这个 100 个婴儿样本中包括 64% 的女孩的可能性到底是多少。不仅如此，当我们问到这样随机抽出的一个样本是“非常可能”还是“不可能”时，我们也不能说明它的准确定义。这一章的大部分内容是主要用来掌握解决这些问题的正式步骤。解决这些问题的第一步是要准确定义我们所要检验的是什么。

一个假设是有关总体参数的一个声称。例如，在性别选择的案例中，总体参数是使用这种产品的女性所产下女孩的比率。假设检验则是指通过一个有效的证据来检验某一个特定的声称是否得到支持。

定义

- ◆ 假设 (hypothesis) 是关于总体参数的一个声称，比如总体比率 (p) 或者总体平均数 (μ)。
- ◆ 假设检验 (hypothesis test) 是一个检验有关总体参数声称的标准过程。

在任何一个假设检验中，里面至少有两个假设。在性别选择的案例中，这两个假设分别是：①性别选择方法可以有效提升通常预计的女孩 50% 的出生比率；②性别选择方法在提高女孩出生比率方面没有效果。如我们先前所讨论的，假设检验的出发点实际上是从第二个假设开始的，这个假设是说性别选择方法无效，不能提高女孩出生比率。我们将这个假设检验的出发点称做虚无假设 (null hypothesis)，写作 H_0 ；另外一个假设叫做备择假设 (alternative hypothesis)，写作 H_a 。下面针对性别选择的例子进行概括。

(1) 虚无假设是性别选择无效的声称, 在本例中, 使用产品的女性所产女孩的总体比率是 50%, 或者是 0.5。使用 P 来代表总体比率, 我们将虚无假设写作:

$$H_0(\text{虚无假设}): P = 0.5$$

(2) 备择假设是指性别选择有效的声称, 在本例中, 使用产品的女性所产下女孩的总体比率应该大于 50%。我们将备择假设写作:

$$H_a(\text{备择假设}): P > 0.5$$

在这一章中, 虚无假设总是包括相等的条件, 比如性别选择例子中的虚无假设就包括等式, $P=0.5$ (我们将在第 10 章中看到几个其他类型的虚无假设的例子)。但是, 性别选择例子中的备择假设 ($P>0.5$), 这是我们在这一章所讲到的三种常见的备择假设中的一种:

总体参数 < 声称值

总体参数 > 声称值

总体参数 $\neq$ 声称值

如我们将在 9.2 节中所看到的, 假设检验中三种不同类型的备择假设在计算中会有稍微的不同。因为这个原因, 我们需要给予它们不同的名称。如果采用的假设是第一种形式 (“<”), 我们称为“左侧检验”(left-tailed), 因为它需要检验总体参数是否位于声称值的左侧(值域范围的低端)。同理, 如果采用的假设是第二种形式 (“>”), 我们称为“右侧检验”(right-tailed), 因为它需要检验总体参数是否位于声称值的右侧(值域范围的高端)。如果采用第三种形式的假设 (“ $\neq$ ”), 我们称之为“双侧检验”(two-tailed), 因为它需要检验总体参数是否显著的远离声称值的高低两端。

相关知识

词语“null”来自拉丁语“nullus”, 意思是指“没有”, 虚无假设经常被陈述为没有特殊效应或者差异。

虚无假设和备择假设

虚无假设 (null hypothesis), 或者 H_0 , 是假设检验的开始假设。对于本章中的假设检验的类型, 虚无假设总是声称总体参数的某一个特定值, 并且采用等式的方式来表达。

(虚无假设): 总体参数 = 声称值

备择假设 (alternative hypothesis), 或者 H_a , 是声称总体参数不同于虚无假设的取值。它可以采用如下形式表达。

(右侧检验): 总体参数 < 声称值

(左侧检验): 总体参数 > 声称值

(双侧检验): 总体参数 $\neq$ 声称值

例 9-1 确定假设

在每一个例子中, 确定所声称的总体参数是什么, 描述假设检验中的虚无假设和备择假设, 并且假设检验采用的是右侧检验、左侧检验还是双侧检验。

(1) 一种新型的混合动力汽车制造商发布广告说, 他们公司生产的汽车燃料消耗量为每加仑燃料行驶 62 公里。而一组消费者声称这款汽车每加仑燃料平均行驶里程不足 62 公里。

(2) 俄亥俄州卫生部门声称, 俄亥俄州医院的平均分娩时间多于全联邦 2.0 天的平均水平。

(3) 一个工作在非洲热带草原的野外生物学家声称, 这个地区的雌性斑马比率与公认的 50% 的比率有差异。

解答 (1) 本例所声称的总体参数是总体平均数 (μ), 即这种新型的混合动力汽车的平均燃油公里数。虚无假设必须表述为总体平均数等于某个特定值。因此我们认识到虚无假设, 就是广告所声称的这种汽车平均燃油消耗量是每加仑 62 公里。备择假设就是消费者所声称的真实燃油消耗量小于广告声称的, 概括表述如下。

$$H_0(\text{虚无假设}): \mu = 62 \text{ 公里/加仑}$$

$$H_a(\text{备择假设}): \mu < 62 \text{ 公里/加仑}$$

因为备择假设是小于的形式, 所以这个假设检验为右侧检验。

(2) 总体是所有的近期在俄亥俄州分娩的妇女, 总体参数是她们在分娩后停留在俄亥俄州医院的平均时间。虚无假设必须是等式, 所以在本例中就是声称在医院停留的平均时间等于全联邦的平均时间 2.0 天。备择假设就是卫生部门所声称的在俄亥俄州医院停留的平均时间大于联邦的平均时间, 概括表述如下。

$$H_0(\text{虚无假设}): \mu = 2.0 \text{ 天}$$

$$H_a(\text{备择假设}): \mu > 2.0 \text{ 天}$$

因为备择假设是大于的形式, 所以这个假设检验为左侧检验。

(3) 在本例中, 声称是关于总体比率 (P) 的, 即本地区的斑马总体中雌性斑马的比率。公认的总

体中雌性斑马的比率为 $P=0.5$ ，这就是虚无假设。野生生物学家声称，实际上总体中雌性斑马的比率是不同于虚无假设的数值的。因为“有差异”可以是多于也可以是少于，备择假设就是“不等于”的形式。

H_0 (虚无假设): $P = 0.5$

H_a (备择假设): $P \neq 0.5$

“不等于”形式的备择假设采用的检验属于双侧检验。

9.1.2 假设检验的可能结果

一个假设检验总是开始于“虚无假设是真的”这样的假定。然后检验看看数据是否能给我们其他任何理由否定虚无假设。接下来我们可以看到，一个假设检验通常只有两种可能结果，下面是对这两个结果的概括。

对一个假设检验来说，有两种可能的结果：

- 拒绝虚无假设 H_0 ，这种情况下我们有证据支持备择假设。
- 不能拒绝虚无假设 H_0 ，这种情况下我们没有足够的证据支持备择假设。

值得注意的是，“接受虚无假设”并不是可能的结果，因为虚无假设总是开始的假定。假设检验可能没有给出我们理由拒绝开始的假设，但是它本身并没有给出理由，让我们得到“开始的假设是真的”这样的结论。

只有两种可能的结果，使得在不偏不倚的方式下选择虚无假设和备择假设非常重要。尤其值得注意的是，在从总体中抽出一个样本进行检验之前，两个假设要用公式明确表达出来。否则，来自样本的数据可能会在选择待检验的假设时产生不恰当的偏差。

思考题

“假设检验并不能导致我们接受虚无假设”的观念只是一个古老格言的例证。这个格言是“缺乏证据并不代表证据的缺乏”。作为这种观念的一个例证，请解释为什么从理论上证明一些传说中的动物（比如大脚，传说中生存在北美洲西北部太平洋沿岸森林中的野人；或者尼斯湖水怪）存在是容易的，但是证明它们不存在几乎是不可能的。

相关知识

第一个记载使用假设检验的例子归功于苏格兰人约翰·阿巴斯诺特（John Arbuthnot 1667—1735），他研究了82年的数据，注意到每年接受洗礼的男性数量要稳定地高于女性。了解到基于性别的洗礼并没有偏差，他认为这样一个模式不能被偶然性解释，一定有“神的旨意”在里面。意思是指男孩出生的比例超出50%多一点，在现代方法中，就是拒绝数据可以被偶然性单独解释的虚无假设。

例9-2 假设检验的结果

针对例1中的三个案例，描述对其进行假设检验的可能结果，以及我们如何解释这些结果。

解答 (1) 回顾这个小题，它的虚无假设是广告商声称新汽车的平均公里数为 $\mu = 62$ 公里/加仑。备择假设是消费者所声称的真实的公里小于广告商所声称的公里数，表达为 $\mu < 62$ 公里/加仑。可能的结果如下。

A. 拒绝虚无假设： $\mu = 62$ 公里/加仑。在这种情况下，我们有证据支持消费者的声称，真实的公里数小于广告商所声称的公里数。

B. 不能拒绝虚无假设： $\mu = 62$ 公里/加仑。在这种情况下，我们缺少证据支持消费者的声称，值得注意的是，尽管如此，但这并不能代表广告商所声称的公里数是真实的。

(2) 此小题虚无假设是全国孕妇生产后平均停留在医院的时间是2.0天。备择假设是俄亥俄州卫生部门所声称的，在俄亥俄州医院产妇产生产后平均停留在医院的时间大于全国的平均时间。可能的结果如下：

A. 拒绝虚无假设： $\mu = 2.0$ 天，在这种情况下，我们有证据支持俄亥俄州卫生部门的声称，产妇生产后在州医院的平均停留时间大于全国平均水平。

B. 不能拒绝虚无假设： $\mu = 2.0$ 天，在这种情况下，我们缺少证据支持俄亥俄州卫生部门的声称，值得注意的是，尽管如此，但这并不能代表在俄亥俄州医院产妇产生产后均停留在医院的时间等于全国的平均时间。

(3) 虚无假设是雌性斑马的比率就是公认的50%的总体比率 ($P=0.5$)。备择假设是生物学家的声称公认的比率是错误的，这意味着实际上的雌性斑马的总体比率不等于50%（它可能大于也可能小于50%）。可能的结果是：

A. 拒绝虚无假设 $P=0.5$ ，在这种情况下，我们有证据支持生物学家所说的“公认的雌性斑马的总体比率是错误的”。

B. 不能拒绝虚无假设 $P=0.5$ ，在这种情况下，我们没有证据支持该生物学家的声称。 ■

9.1.3 假设检验的结论

让我们回顾性别选择的例子，设想我们随机抽取了100个婴儿的样本（使用了性别选择产品的孕妇所生产的婴儿），并且发现其中64个婴儿是女孩。我们怎样才能确定这个样本结果，导致我们是否拒绝虚无假设呢？归根结底，答案来自虚无假设真实的前提下，样本结果是更可能发生，还是发生的可能性很小呢？

回顾这个例子中的虚无假设，是说使用性别选择工具的总体中女婴的真实比率是50%，或者说是 $P=0.5$ 。使用第8章介绍的公式，我们所研究的这个样本，样本容量是100，样本比率 = 0.64。将问题明确，如果真实的总体比率是 $P=0.5$ （如同虚无假设所声称的那样），那么在一个样本容量 $n=100$ 的样本中，样本比率至少是0.64的概率是多少？如果这个概率很小，我们得到这样的样本的机会可能性是很小的；这样我们就有理由拒绝虚无假设。如果观察到这个样本结果的机会是中等或者较高水平的，我们得到这种样本的机会是非常有可能的，这样我们就不能拒绝虚无假设。

1. 统计的显著性水平

我们在6.1节介绍了统计显著性水平的概念，回想一下，如果某一特定结果的概率小于或者等于0.05，我们就可以说这个结果在0.05水平上具有统计显著性；如果这个结果的概率小于或等于0.01，我们就可以说这个结果在0.01水平上显著。这意味着0.01水平的统计显著性要高于0.05水平的统计显著性。下面的内容是有关如何将这些概念直接用于统计假设检验中的。

基于一定统计显著性水平的假设检验的统计决断
我们通过将实际的样本结果（平均数或者样本比率），与虚无假设为真的前提预期的结果进行比较之后，从而对假设检验的结果做出统计决断。

为了做出统计决断，必须选择某一个统计显著性水平。

- 作为一种观察的结果而言，如果得到样本结果的机会小于1%（或者0.01）的话，取得的这个假设检验的统计显著性水平就在0.01水平，这提供了一个拒绝虚无假设强有力的证据。
- 作为一种观察的结果而言，如果得到样本结果的机会小于5%（或者0.05）的话，取得的这个假设检验的统计显著性水平就在0.05水平，这提供了一个拒绝虚无假设中等水平的证据。
- 作为一种观察的结果而言，如果得到样本结果的机会大于所选择的显著性水平0.01或者0.05的话，我们就不能拒绝虚无假设。

例9-3 假设检验的统计显著性水平

考虑到性别选择的例子，其样本容量是 $n=100$ ，样本比率是0.64。使用我们在这一章所讨论的技术，在假定虚无假设（ $P=0.5$ ）是真的前提下，计算随机抽取这样一个样本的概率是可能的（或者一个更极端的样本比率 >0.64 的样本），抽出这样一个样本的结果概率是0.0026。基于这个结果，我们应该拒绝还是不拒绝虚无假设？

解答 0.0026的概率意味着，如果虚无假设是真的（意思是指真实的总体比率是 $P=0.5$ ），抽出一个样本，其样本比率至少是0.64的概率会小于3/1000。因为这个概率小于1/100，所以这个结果的统计显著性在0.01水平。这给了我们一个强有力的理由拒绝虚无假设，也就是说，提供了针对备择假设“性别选择工具可以提高女孩的出生比率，使其超过50%”的一个支持。 ■

2. P 值

在上面的例子中，我们得出结论说，样本的结果给了我们一个拒绝零假设的理由，因为这个结果在0.01水平上具有统计显著性。实际上，这个结果甚至可以更进一步，计算出的概率值是0.0026，大约是0.01的1/4。这样精确的概率，相对于只陈述一个统计显著性水平，给了我们更多的信息。

公开发表的假设检验的概要或新闻报道经常只陈述一个统计显著性水平，但是确定这个水平总是需要我们计算一个更精确的概率。这个概率就叫做 P 值（概率值），注意字母 P ，经常要避免写成小写字母 p ，而 p 是用来表示总体比率的。换句话说，在上面的例子中，抽出一

个样本比率至少在 64% 以上的概率是 0.002 6, 我们说假设检验的 P 值是 0.002 6。我们将要在 9.2 节和 9.3 节讨论 P 值的实际计算, 在此, 我们仅仅关注它们的解释。

基于 P 值 (P -Value) 的假设检验的统计决断关于声称一个总体参数值的假设检验的 P 值 (概率值), 是指假定虚无假设为真的前提下, 随机抽取一个样本作为一个被观察的样本, 至少等于某个极端值的概率。

- 一个小的 P 值 (比如小于或等于 0.05), 说明抽取这样的样本结果是不大可能的, 这样就提供了一个拒绝虚无假设的理由。
- 一个大的 P 值 (比如大于 0.05), 说明这样的样本结果抽出机会是比较大的, 所以我们不能拒绝虚无假设。

例 9-4 均匀的硬币

你猜想一个硬币可能不均匀, 因为它被抛掷落地时, 更多时候都是反面向上的次数多于正面向上的次数。为了检验这个猜想, 将这枚硬币抛掷了 100 次。结果是正面向上的次数是 40, 反面向上次数是 60。通过计算说明, 一枚均匀的硬币抛掷 100 次, 40 次以下正面向上的概率是 0.022 8。说明结果的 P 值和统计显著性水平。你能够得出结论, 这枚硬币是不均匀的, 抛掷以后反面向上的机会更大一些。

解答 虚无假设是这枚硬币是均匀的, 在这种情况下, 抛掷以后正面向上的比率应该是 0.5 (或者 50%), 虚无假设可以表述为: $P = 0.5$ 。备择假设就是你的猜想, 这枚硬币是不均匀的, 抛掷以后更容易出现反面向上的情况。在这种情况下, 投掷以后正面向上的比率应该少于 0.5 (或者 50%), 备择假设可以表述为, $P < 0.5$ 。抛掷这枚硬币 100 次代表了一个随机样本, 样本容量 $n = 100$, 40 次正面向上的结果说明这个假设检验的样本比率 = 0.4。这个假设检验的 P 值就是计算至多不超过所得到的样本比率情况下的概率, 假定这枚硬币是均匀的, 正面向上的总体比率是 0.5。在这种条件下发生至多 40 次正面向上的概率是 0.022 8, 0.022 8 就是这个检验的 P 值。因为 P 值小于 0.05, 所以这个检验的统计显著性水平是 0.05 水平。因为这个 P 值大于 0.01, 所以这个结果在 0.01 水平并不显著。0.05 水平的统计显著性给出一个中等程度的理由去拒绝虚无假设, 并且得出结论说这枚硬币是不均匀的, 抛掷它反面向上的机会更大一些。

9.1.4 信息汇总

除了实际的计算过程, 我们现在已经掌握

了假设检验的所有基本观念, 我们将在 9.2 节讨论总体平均数的假设检验计算, 在 9.3 节讨论总体比率的假设检验计算。下面的内容就是关于假设检验步骤的概括。

假设检验的步骤

第一步, 采用公式表达虚无假设和备择假设, 每一个假设均是关于总体参数的一个说明, 比如总体平均数 (μ) 或者总体比率 (p); 这个说明要在抽取一个样本或者搜集数据之前完成。应基于备择假设的形式, 来确定你需要的检验是左侧检验、右侧检验还是双侧检验。

第二步, 从总体中选取一个样本, 并且计算样本统计量, 包括样本容量 (n)、相关的样本统计量, 比如样本平均数 ($\bar{x}$), 或者样本比率 ($\hat{p}$)。

第三步, 在假定虚无假设是真实的前提下, 计算你所观察到的样本统计量 (平均数或比率) 至少作为极端情况能被抽取出来的概率。这样一个观测的准确概率就是样本结果的 P 值。

第四步, 做出统计决断, 在选定的统计显著性水平 (通常是 0.05 或者 0.01 水平, 在某些情况下, 两种显著性水平都会使用) 的前提下, 确定拒绝虚无假设还是不拒绝虚无假设。

此外, 在使用字母 p 的时候注意避免混淆以下三种不同的用法:

- 小写字母 p 代表总体比率, 也就是说在一个完全总体中真实的比率。
- 小写字母 $\hat{p}$ 代表样本比率, 也就是说, 在一个总体中所抽取出来的样本比率。
- 大写字母 P 代表 P 值的概率。

例 9-5 车辆公里数的平均租赁价格

在美国, 一辆汽车每年平均驾驶里程数大约是 12 000 公里。一家大型汽车租赁公司的老板, 猜想自己的车辆平均行驶公里数远大于 12 000 公里。他从自己的车辆中随机选取了 $n = 250$ 辆汽车的这样一个样本, 计算的样本平均公里数 = 12 375 公里。计算结果显示, 如果你假定全国车辆的每年平均公里数是 12 000 公里, 从中选取 225 辆汽车的一个样本, 每年平均公里数至少是 12 375 公里的概率是 0.01。基于这些数据, 描述建立一个统计假设并且做出统计决断的步骤。

解答 按照上述统计假设检验的四个步骤, 本例可以表述如下。

第一步, 本例中所关注的总体参数是总体平均数, 也就是指车辆租赁公司所有汽车的每年平均公里数这个总体。虚无假设就是指这个车辆租赁公司所有

汽车的年平均公里数均等于全美国汽车的年平均公里数 12 000 公里。备择假设就是汽车租赁公司的老板所声明的车辆租赁公司年平均公里数大于全美国的年平均公里数——12 000 公里，具体表达如下。

因为备择假设是大于的形式，所以确定需要的检验是右侧检验。

第二步，这一步要求从总体中选取一个样本，并且计算样本统计量，已知条件是样本容量 ($n=225$)、样本平均数 ($\mu=12\,375$ 公里)。

第三步，在假定虚无假设是真实的前提下，也就是指假定汽车租赁公司的年平均公里数真的是 12 000 公里，在这个条件下计算选取一个样本平均数至少是 12 375 公里的概率。在这里我们不需要具体的计算。因为已知条件告诉我们概率是 0.01 (具体的计算方法参考 9.2 节)。这个概率就是 P 值。如果虚无假设真实的话，从这个总体中抽取出样本平均数至少是 12 375 公里的概率仅仅是 0.01。

第四步， P 值是 0.01。这告诉我们这个结果的统计显著性水平为 0.01。这样我们可以拒绝虚无假设，并且可以得出结论说，对于接受备择假设而言，这个假设检验提供了一个强有力的支持。这意味着这家汽车租赁公司的年平均公里数大于全国平均水平 12 000 公里。 ■

9.1.5 关于假设检验的法学类比

一个法律方面的类比可以帮助我们澄清有

关假设检验的观念。在美国法庭中，基本的原则是如果没有证据的话就假定被告无罪。因为开始的假定是无罪的，所以虚无假设表达如下。

H_0 虚无假设：被告是无罪的

H_a 备择假设：被告是有罪的

检察官的任务就是提供证据使得陪审团发现被告有罪，不得不拒绝虚无假设。如果检察官不能提供足够的证据，陪审团将不会拒绝虚无假设，被告将被宣告“无罪”。值得注意的是：宣告一个人无罪 (接受虚无假设) 并不是一种选择权，一个无罪的裁定事实意味着证据不充分，不足以证明被告有罪，但是它并不能证明被告无罪。

思考题

考虑下面两种情况。第一种情况，你作为一项案件的陪审员，在这个案子中被告可能会被判决罚款 2 000 美元。另外一种情况，你同样作为一项案件的陪审员，在这个案子中被告可能会被宣判死刑。比较你在两种情况下的显著性水平。在每种情况下，拒绝虚无假设会产生什么样的后果？

9.1 练习题

统计素养和批判性思维

1. 假设检验 什么是假设检验？
 2. 假设 什么是虚无假设，虚无假设使用什么符号表示，什么是备择假设，备择假设使用什么符号表示？
 3. 备择假设 一个研究者想要检验一个有关健康成人的平均体温的说法：健康成人的平均体温是 98.6°F 。说明用作备择假设的三种可能的不同表达方式。
 4. P 值 什么是假设检验的 P 值？
- 判断正误：**判断练习题 5 ~ 12 中的陈述是有意义的 (表述正确) 还是没有意义的 (表述错误)，并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。
5. 媒体报道 《纽波特编年史》发表的一篇文章有如下陈述：“研究证明，未通过药物测试的工作申请人的百分比 = 6%”。
 6. P 值 一个研究者确信他能够证明一种新药

在降低 LDL 胆固醇方面是有效的。他声明，支持降低 LDL 胆固醇平均水平的说法 P 值是 0.001。

7. 虚无假设 为了检验平均 LDL 胆固醇水平少于 130 毫克/毫升的说法，研究者将虚无假设表述为 $\mu < 130$ 毫克/毫升。
8. 虚无假设 在建立一个假设检验之后，一个研究者得出了一个最初的结论，有足够的样本证据可以支持虚无假设成立。
9. 药物治疗 在一个检验中所声称的内容是，经过药物 Ziac 治疗的患者感到头晕的比例少于 0.06，备择假设假设就是 $p \geq 0.06$ 。
10. 样本平均数 一个研究被设计成确定体重超过 200 磅的男性比率，所以样本平均体重必须可以通过计算得来。
11. P 值 在一种新药的临床测试中，项目的负责人非常乐于知道药物的有效性被描述成 P 值是 0.001。

12. 心率 在一个假设检验中，假设运动的人平均心率少于每分钟 72 次，虚无假设是 $\mu = 74$ ，备择假设是 $\mu < 74$ 。

概念与应用

13. 显著性水平是多少？ 假定你驱车上学，每周五下午你会加 7 加仑汽油，加满你的油箱。

(1) 一周内你正常驾驶驱车到校，结果发现汽车需要 8 加仑汽油，你会认为这个增长具有统计显著性，还是会将它归因为随机的波动？

(2) 在另一周正常驾驶驱车到校之后，你发现汽车需要 12 加仑汽油，你会认为这个汽油消耗量的增长具有统计显著性，还是将它归因为随机波动？

(3) 这种汽油消耗量增长多少，你就需要考虑它具有统计显著性并且寻找合理的解释？请加以说明。

14. 什么情况下具有统计显著性？ 假定你的长途电话费大约是每月 7 美元。

(1) 在电话正常使用一个月之后，你的长途电话费账单是 8.00 美元，你将考虑这个变化具有统计显著性，并且寻找一个合理的解释？还是将它归因于一种偶然的变化？

(2) 在电话正常使用一个月之后，你的长途电话费账单是 40 美元，你将考虑这个变化具有统计显著性，并且寻找一个合理的解释？还是将它归因于一种偶然的变化？

(3) 这种长途电话费增长多少你就需要考虑它具有统计显著性并且寻找合理的解释？请加以说明。

用公式表达假设：在练习题 15 ~ 22 中，用公式表达一个假设检验的虚无假设和备择假设。用明确的术语陈述对应于已知声明的两种可能的结论。

15. 可乐的净重 一个消费者抱怨一家超市所售卖的标志上注明 1 升的可乐实际净重总是少于 1 升。

16. SAT 分数 一个高中校长声称，他所在学校的高年级学生的 SAT 分数低于全国平均分 1518。

17. 性别选择 一家医药制造公司的首席运营官

声称，他们的药物会增加产妇生女孩的概率，使用这种药物后生女孩的比率超过 0.7。

18. 质量控制 一家跑步机制造公司的质量监控经理声明，他们制造的跑步机中残次品的比率低于 0.02。

19. 自动售货机 一个销售代理商声明，他们的自动售货机所售卖的咖啡平均容量等于 10 盎司。

20. 阿司匹林 食品和药物委员会声称，一家制药公司所生产的阿司匹林片剂所含的阿司匹林平均含量少于 350 毫克。

21. 犹太人大屠杀 一个高中老师说，他的大多数学生不知道大屠杀这个术语是指什么。

22. 吸烟 一个教育者声称，吸烟的大学毕业生所占比率低于 20%。

P 值和婴儿的性别 假设生下男婴和女婴的机会是相同的，下表呈现了一个 100 名婴儿的随机样本中男婴出生数量不同的概率。在练习题 23 ~ 28 中使用下表的信息，假设我们检验一个男婴出生概率的偏差。

100 个婴儿中的男孩	概率
35 个及以下	0.002
40 个及以下	0.028
45 个及以下	0.184
48 个及以下	0.382

23. 一个 100 名婴儿的随机样本中有 48 名男婴，这个结果在 0.05 水平上有统计显著性吗？这个结果的 P 值是多少？

24. 一个 100 名婴儿的随机样本中有 45 名男婴，这个结果在 0.01 水平上有统计显著性吗？这个结果的 P 值是多少？

25. 100 名婴儿的随机样本中有 40 名男婴，这个结果在 0.01 水平上有统计显著性吗？这个结果的 P 值是多少？

26. 100 名婴儿的随机样本中有 40 名男婴，这个结果在 0.05 水平上有统计显著性吗？这个结果的 P 值是多少？

27. 100 名婴儿的随机样本中有 35 名男婴，这个结果在 0.01 水平上有统计显著性吗？这个结果的 P 值是多少？

28. 100 名婴儿的随机样本中有 32 名男婴，这个结果在 0.01 水平上有统计显著性吗？这个

结果的 P 值是多少?

网络和远程项目

请在网站 www.aw.com/bbt 上打开第9章的有用链接“links for internet projects”。

29. 专业杂志 许多专业杂志, 比如《美国医药联合会杂志》上面的许多文章包括正式的假设检验信息, 寻找这样的一篇文章, 并且确定虚无假设和备择假设, 用简单的术语陈述这个假设检验的对象和最后得出的结论。
30. 抛掷硬币 选择一枚特定的硬币, 检验这个说

法: 它落地时更容易出现正面向上。表述出虚无假设和备择假设, 然后抛掷这个硬币 100 次。根据你的感觉, 关于硬币抛掷以后更容易正面向上这个说法, 你得到什么样的结论?

新闻中的统计学

31. 新闻中的假设检验 寻找一篇新闻报道或研究报告, 上面的内容是关于总体平均数或总体比率的假设检验 (可能内容不是很详细), 概括这篇文章所使用的方法, 并将文章摘录下来。



9.2 总体平均数的假设检验

在上一节, 我们概括地描述了假设检验步骤的基本轮廓。回忆一下所有的例子, 假设检验的最终目标是关于是否拒绝虚无假设做出统计决断, 虚无假设也是这个检验的最开始的假设。在这一节中, 我们将描述用于总体平均数假设检验的计算 (我们将在下一节讨论总体比率的假设检验)。我们首先探讨单侧检验的步骤 (右侧检验或者左侧检验), 然后再讨论单侧检验步骤与双侧检验的细微的差异。

9.2.1 单侧假设检验

考虑下面的假想情况。哥伦比亚大学广告宣传他们的毕业生平均的起薪是 39 000 美元。广告真实委员会, 一个独立的机构, 猜想这个宣传是言过其实的, 并且决定建立一个假设检验来寻找支持此怀疑的证据。

注意到, 这个关注的参数是所有哥伦比亚大学毕业生总体的平均起薪, 所以这个假设检验将与总体平均数有关。虚无假设是哥伦比亚大学所宣传的毕业生平均起薪是 39 000 美元。备择假设是委员会所声称的上述宣传言过其实的, 哥伦比亚大学的毕业生平均起薪低于 39 000 美元。采用公式表达的话, 虚无假设和备择假设分别如下。

$$H_0: \mu = 39\,000 \text{ 美元}$$

$$H_a: \mu < 39\,000 \text{ 美元}$$

因为备择假设是“小于”的形式, 所以我们应该采用左侧假设检验。左侧检验和右侧检验的一般步骤是相同的, 所以我们将它们一同

称做单侧检验。

为了检验这个假设, 广告真实委员会从哥伦比亚大学近期的毕业生中随机挑选了一个包括 100 个学生的样本。在这个样本中, 毕业生的平均起薪是 37 000 美元。换句话说, 样本容量 $n = 100$, 样本平均数 = 37 000 美元。

如果你回顾上一节中所描述的假设检验的四个步骤, 你会发现, 前两步已经完成。所关注的总体参数——所有哥伦比亚毕业生的平均起薪已经确定, 虚无假设和备择假设已经表述, 样本被抽取出来, 并且已经计算了样本容量和样本平均数。现在我们开始进行第三步和第四步, 我们假定虚无假设是真的, 然后来确定样本统计量能否给出一个理由来拒绝这个假定。

方法注解

从 2006 年起, 来自联邦教育统计中心的数据说明, 所有获得学士学位的大学毕业生起薪的中位数是 35 000 美元。已知中位数高于平均数, 因为薪水分布倾向于偏态的, 专业不同造成的差异很大, 工程和商学这两个专业通常起薪最高。

1. 抽样分布

假设检验步骤的第三步是, 在假定虚无假设是真的的前提下, 计算所观察到的样本平均数作为一个极端例子时的可能性。对于哥伦比亚大学毕业生的例子来说, 这个问题就变成了: 当总体平均数是 39 000 美元时, 我们从中抽取一个样本 (样本容量 $n = 100$), 样本平均数等于或者少于 37 000 美元的可能性是多少? 为了回答这个问题, 根据本书第 8 章抽样分布的内

容，我们需要一个重要的观察结果：观察到的样本平均数（37 000 美元）仅是样本平均数分布的一个点。更进一步来说，假定虚无假设是真的，样本平均数的分布近似于正态分布，其平均数 = 39 000 美元。

为了理解这个观察样本，设想一下，广告真实委员会不仅抽取了这一个样本（样本容量 $n = 100$ ），还抽取了许多个同样大小样本容量的样本。每一个样本都有一个自己的样本平均数，我们将所有这些样本平均数的分布绘制成一个图形。如同在 8.1 节中所讨论的，因为样本容量 $n = 100$ 相对较大，抽样分布的结果将接近于正态分布，并且样本平均数分布的平均数等于总体平均数。因为虚无假设说的是总体平均数 = 39 000 美元，那么如果虚无假设正确的话，样本平均数抽样分布的峰值将在这个数值上。

图 9-1 呈现了以上观点的图示。黑色曲线代表了假定虚无假设为真的前提下的抽样分布。这个曲线是根据我们预想中的抽取许多样本平均数的分布所绘制的。但实际上我们只有一个样本平均数（比如哥伦比亚毕业生例子中的样本平均数 = 37 000 美元），它仅仅只是代表了这条曲线上的一个点。如果这个点比较接近曲线的峰值，它将告诉我们样本平均数接近虚无假设中所预期的总体平均数。在这样的情况下，得到这样一个样本平均数的概率并不小，没有理由拒绝虚无假设。相反，如果样本平均数位于虚无假设所声称的总体平均数很远的位置，如果在虚无假设是真实的前提下，得到这样一个样本平均数的概率就比较小。因此我们可以得出结论说真实的总体平均数可能并不是虚无

假设所声称的数值，在这种情况下，我们就有理由拒绝虚无假设。

2. 标准分数的计算

如图 9-1 中所说明的，做出统计决断是否拒绝虚无假设有赖于样本平均数与虚无假设所声称的总体平均数间的距离是“远”还是“近”。因为抽样分布符合或者接近正态分布，那么关于样本平均数的标准分数就代表了样本平均数和所声称的总体平均数之间距离的量化指标。回顾 5.2 节中，正态分布中一个数据值的标准分数（或者 z 分数），它是指这个数值位于平均数分布上端或下端几个标准差的位置。根据 5.3 节所说的中心极限定理，样本平均数分布的标准差就是 $\sigma/\sqrt{n}$ ，在公式中 σ 代表总体标准差， n 代表样本容量。将上述内容联系在一起，我们就可以采用公式计算样本平均数的标准分数：

$$z = \frac{(\text{样本平均数} - \text{总体平均数})}{\text{抽样分布的标准差}} \\ = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (9-1)$$

余下的问题就是我们并不知道总体标准差。我们暂时使用样本标准差来近似地代替总体标准差，样本中 100 个人薪水的标准差是 $s = 6\,150$ 美元（在 10.1 节中，我们将讨论在总体标准差未知的情况下，有没有更好的办法）。在这样的情况下，我们将总体标准差 = 6 150 美元代入公式中，就可以得到样本平均数分布的标准差是：

$$\sigma/\sqrt{n} = 6\,150 \text{ 美元} / \sqrt{100} = 615 \text{ 美元}$$

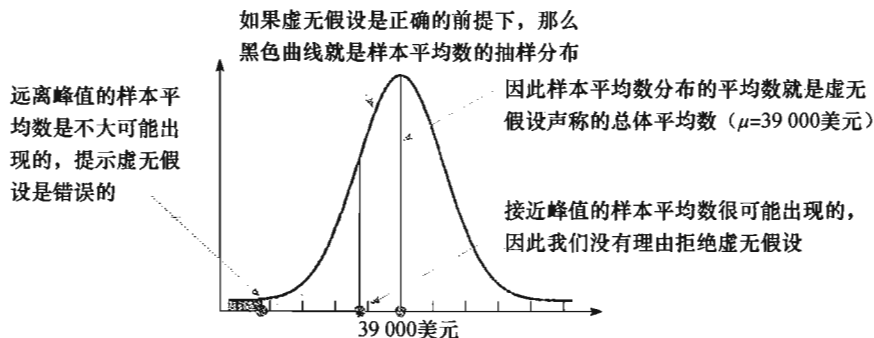


图 9-1 哥伦比亚大学毕业生假设检验的图示说明

注：如果虚无假设为真，那么总体平均数 = 39 000 美元。在这种情况下，如果我们从总体中抽取许多样本，样本平均数的分布将接近正态分布，平均数是 39 000 美元。在这样的假定条件下，某一样本平均数与总体平均数之间的差值让我们决定是否拒绝虚无假设。

将这个值代入前面所提到的标准分数的公式, 这样在总体平均数 = 39 000 美元的抽样分布中, 我们就可以得到样本平均数 = 37 000 美元所对应的标准分数是:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$= (37\,000 \text{ 美元} - 39\,000 \text{ 美元}) / 615 \text{ 美元}$$

$$= -3.25$$

换句话说, 样本平均数 = 37 000 美元位于抽样分布平均数左端 3.25 个标准差的位置。

图 9-2 呈现的是, 标准分数 -3.25 说明样本结果位于分布左侧远端的位置。说明如果总体平均数真的是虚无假设所声称的 39 000 美元的话, 这是一个非常罕见的样本结果。这意味着虚无假设是不正确的, 应该拒绝虚无假设。尽管如此, 这不是拒绝虚无假设直观的唯一理由, 让我们来看一看通过公式分析标准分数并且做出统计决断的过程。

在假设检验中计算样本平均数的标准分数

当我们抽取一个样本进行假设检验时, 我们可以想到这是抽样分布中许多可能样本中的一个。已知样本容量 (n), 样本平均数 ($\bar{x}$), 总体标准差 (σ), 并且也知道所声称的总体平均数 (μ), 我们就可以进行以下计算:

$$\text{样本平均数分布的标准差} = \sigma / \sqrt{n} \quad (9-2)$$

$$\text{样本平均数的标准分数 } z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad (9-3)$$

值得注意的是: 我们很少有机会能够知道总体标准差; 参考 10.1 节的内容, 就能够找到怎样处理这种情况的办法。

方法注解

样本平均数的标准分数经常被叫检验统计量。

3. 统计显著性的临界值

回顾一下, 如果将计算出所观察的样本的结果作为临界值, 其概率等于或者小于 0.05 (假定虚无假设为真的前提下), 假设检验就是在 0.05 水平上进行。对于一个左侧检验而言, 那么我们要寻找一个位于或者低于抽样分布 5% 位置的标准分数。从表 5-1 中可以看到, 第 5 百分等级所对应的标准分数在 $z = -1.6$ 和 $z = -1.7$ 之间。查询更精确的标准分数表 A-1, 可以知道 5 百分等级所对应的标准分数 $z = -1.645$ 。因此, 如果样本平均数的标准分数小于或等于 $z = -1.645$, 这个左侧检验就在 0.05 水平上显著。在左侧检验中, 这个标准分数代表 0.05 水平统计显著性得到临界值。

同样的道理可以适用于右侧检验, 其备择假设的形式是: $\mu >$ 声称值。在这样的情况下, 0.05 水平的显著性要求样本平均数位于或者大于第 95 百分等级, 或者说要求样本平均数的标准分数大于或等于 $z = 1.645$ 。图 9-3 说明了左侧检验和右侧检验的这些临界值。为了计算 0.01 显著性水平的临界值, 我们可以计算第 1 和第 99 百分等级的标准分数 (替换第 5 和第 95 百分等级)。表 A-1 显示出这些临界值分别是 -2.33 和 2.33。

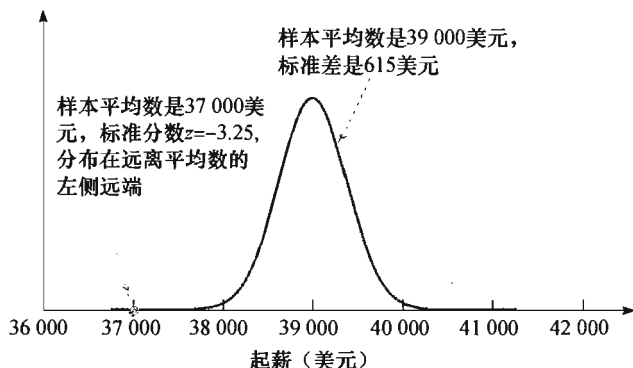


图 9-2 假如虚无假设为真, 哥伦比亚大学生例子中的样本平均数分布, 其平均数是 39 000 美元, 标准差是 615 美元。在这种情况下, 样本平均数 = 37 000 美元的标准分数是 -3.25

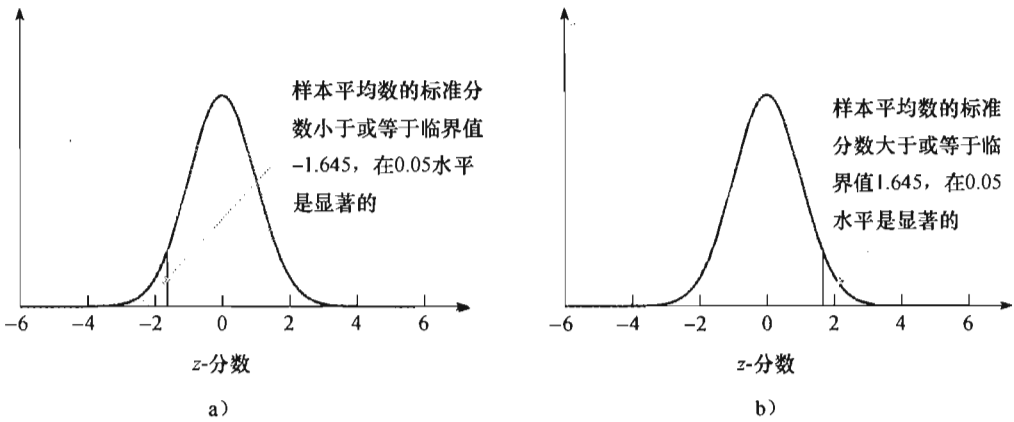


图 9-3 图中说明了在 0.05 显著性水平上的单侧假设检验的平均数的临界值。左侧检验临界值对应第 5 百分等级，右侧检验对应第 95 百分等级。
a) 左侧检验 $H_a: < \text{声称值}$ b) 右侧检验 $H_a: > \text{声称值}$

基于某种显著性水平的单侧假设检验的统计决断

我们将样本平均数的标准分数和临界值 (critical values) 进行比较，从而做出统计决断，拒绝还是拒绝虚无假设。表 9-1 概括了 0.01 显著性水平和 0.05 显著性水平上的单侧假设检验的统计决断。

表 9-1 0.05 和 0.01 显著性水平的假设检验

检验类型	备择假设的形式	0.05 水平上 拒绝虚无假设 的标准分数	0.01 水平上 拒绝虚无假设 的标准分数
左侧检验	$H_a: \mu < \text{声称值}$	$z \leq -1.645$	$z \leq -2.33$
右侧检验	$H_a: \mu > \text{声称值}$	$z \geq 1.645$	$z \geq 2.33$

思考题

假定一个右侧检验，一个研究计算出样本平均数的标准分数是 $z=3$ ，另一个研究计算出样本平均数的标准分数 $z=10$ 。两个标准分数都在 0.01 水平显著，对于拒绝虚无假设来说，那一个结果提供的证据更为强有力？请加以解释。

例 9-6 哥伦比亚大学毕业生假设检验的统计显著性

假定虚无假设为真的前提下，哥伦比亚大学毕业

生的平均起薪是 39 000 美元，计算出一个样本平均数只有 37 000 美元，这是否具有统计显著性？依据你的回答，应该拒绝还是不应该拒绝虚无假设？

解答 回想一下，这个假设检验属于左侧检验（因为备择假设是“小于”的形式， $\mu < 39\,000$ 美元），我们已经知道样本平均数 = 37 000 美元，所对应的标准分数是 -3.25 。这个结果在 0.05 水平上具有统计显著性，因为这个标准分数 $<$ 临界值 $z = -1.645$ 。实际上，这个结果在 0.01 水平上也具有统计显著性，因为标准分数 $<$ 临界值 $z = -2.33$ 。因此，我们有强有力的理由拒绝虚无假设，并且可以得出结论，哥伦比亚大学官方声明确实夸大了他们大学毕业生的平均起薪。

4. 计算 P 值

正如我们在 9.1 节中所讨论的，我们可以使用 P 值来说明统计显著性会更精确。回顾一下，在假定虚无假设为真的前提下，P 值用来计算样本平均数作为一个观察到的特例的概率。这一章我们讨论了假设检验，一般情况下，我们通过计算样本平均数的标准分数计算对应的 P 值。图 9-4 呈现了这个观念。样本平均数分布曲线以下的全部面积被定义为 1，因此我们用曲线下的面积来解释概率。对于一个左侧检验来说（图 9-4a），计算一个样本平均数少于或等于某个特定值的概率，就是简单地找出曲线下样本平均数左侧部分的面积。对于一个右侧检验来说（图 9-4b），概率就是曲线下样本平均数右侧的面积。

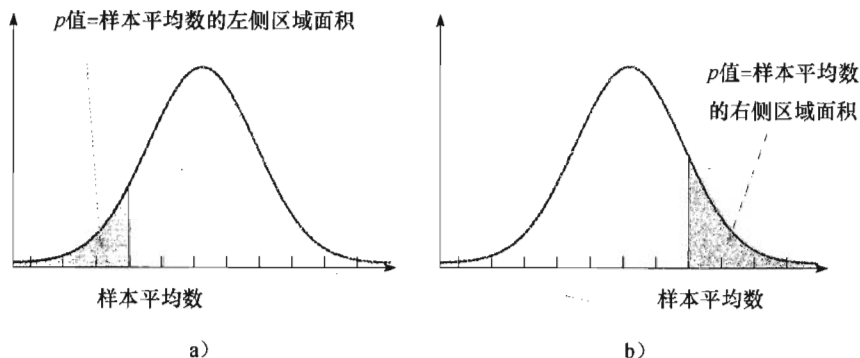


图9-4 单侧假设检验的 P 值对应抽样分布曲线下所涵盖的某一面积。当我们计算某一特定样本平均数的标准分数时，我们就可以根据表 A-1 中的标准分数表来计算对应的面积。（注意：表 A-1 列出了从左侧开始的每一个标准分数所对应的面积，因此，对于右侧检验而言，计算样本平均数右侧的面积需要 1 减去查表值）

a) 左侧检验 b) 右侧检验

作为一个例子，让我们来计算哥伦比亚大学毕业生假设检验的 P 值。我们已经知道样本平均数（=37 000 美元）所对应的标准分数是 $z = -3.25$ 。如果在表 A-1 中查找，你可以看到这个标准分数所对应的“从左侧开始起的累计面积”是 0.000 6，这就是我们所要计算的 P 值。这个很小的 P 值提供了一个强有力的否定虚无假设的证据。毫无疑问，这个 P 值小于 0.01，这就是为什么这个结果在 0.01 水平上具有统计显著性。它小于 0.01 的事实告诉我们，我们有强有力的证据拒绝虚无假设。

5. 总体平均数单侧检验的概括

在这一节开始，目标是学习如何实现假设检验四步骤中的第三步和第四步。我们现在已经成功完成了平均数的单侧假设检验。现在概括如下。

- 因为我们想知道总体平均数，虚无假设就是这样的形式， $\mu = \text{声称值}$ 。为了做出统计决断，是否拒绝虚无假设，在虚无假设为真的前提下，我们必须确定，在假设检验中作为一个特例的我们得到这个样本是有可能发生还是很少有可能发生。
- 我们通过样本平均数所对应的标准分数来确定这个可能性，具体计算通过如下公式完成：

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

式中， n 是样本容量， μ 是虚无假设所声称的总体平均数， σ 是总体标准差（在这一节中，我们通常用样本标准差 s 作为总体标准差的近似代表值）。

- 我们可以通过以下两种方式使用标准分数：①我们可以通过比较计算得到的标准分数和查表 9-1 所得到的临界值，来得到统计显著性水平。②我们可以通过查像表 A-1 这样的标准分数表来确定 P 值。对于左侧检验而言， P 值是正态分布曲线下标准分数左侧的面积，对于右侧检验而言，是正态分布曲线下标准分数右侧的面积。
- 如果在这些选定水平上（通常是 0.05 或者 0.01 显著性水平）结果具有统计显著性，我们就拒绝虚无假设。如果没有统计显著性，我们就不拒绝虚无假设。

例 9-7 租赁汽车公司的平均公里数

回顾一下租赁汽车公司老板的例子（9.1 节的例 5），他怀疑公司汽车实际的年公里数大于全国 12 000 公里的平均水平。他从公司中随机选取了 $n = 225$ 辆汽车，并且测量的结果平均公里数是 12 375 公里，样本标准差是 $s = 2415$ 公里。确定假设检验的统计显著性和 P 值，并且解释计算结果。

解答 为了确定假设检验的统计显著性和 P 值，我们首先要计算样本平均数 = 12 375 公里所对应的标准分数，已知所声称的总体平均数（ $\mu = 12\,000$ 公里）和样本容量（ $n = 225$ ），我们不知道总体标准差，但是我们可以假定它与样本标准差一样，这样总体标准

差就是 $\sigma = 2\,415$ 公里。据这些已知条件，我们可以计算得出：

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{12\,375 - 12\,000}{2415 / \sqrt{225}} = 2.33$$

计算出来的标准分数大于 0.05 显著性水平的临界值 1.645，并且等于 0.01 显著性水平的临界值 2.33。这样我们就有一个强有力的理由去拒绝虚无假设，并且得出结论说，租赁汽车公司的平均年里程数实际上大于全国平均水平。我们可以从表 A-1 中查出 P 值，表中显示出标准分数 $z = 2.33$ 左侧的面积数是 0.990 1，因为这是一个右侧检验，所以概率是右侧的面积（参考图 9-4b），也就是 $1 - 0.990\,1 = 0.009\,9$ 。因此 P 值是 0.009 9，非常接近于 0.01。

9.2.2 双侧检验

同样的基本原理可以应用于双侧假设检验，备择假设是“不等于”的形式， $\mu \neq$ 声称值。但是，双侧假设检验的临界值和 P 值的计算过程稍微有别于单侧检验。

通常情况，如果作为一个特例所抽出的样本实际上的概率小于或等于 0.05，那么假设检验在 0.05 水平具有统计显著性。对于单侧检验而言，0.05 的概率对应左侧检验的第 5 百分等级和右侧检验第 95 百分等级（见图 9-3）。对于双侧检验来说，“一个实际抽样出来的特定样本”取值可能位于抽样分布的左侧，也可能位于右侧。5%（或者 0.05）的概率对应的标准分数既包括抽样分布左侧的 2.5%，也包括抽样分布右侧的 2.5%。从表 A-1 中可以查出，第 2.5 百分等级对应的标准分数是 -1.96，第 97.5 百分等级对应的标准分数是 1.96。这些标准分数就成为双侧检验 0.05 显著性水平的临界值。同样的，双侧检验 0.01 显著性水平的临界值是第 0.5 百分等级和第 99.5 百分等级，从表 A-1 中可以查出，所对应的标准分数值是 -2.575 和 2.575。

同样思考一下双侧检验中 P 值的应用。回想一下单侧检验，左侧检验的 P 值是曲线之下样本平均数左侧部分的面积，右侧检验的 P 值是曲线之下样本平均数右侧部分的面积。因为双侧检验要求我们考虑所声称的平均数两侧的极端值，双侧检验的 P 值应该相当于两个单侧检验的 P 值。

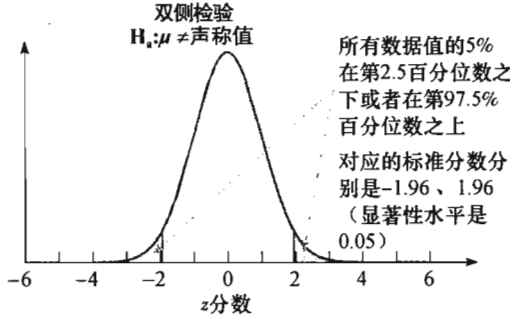


图 9-5 对于双侧检验，0.05 显著性水平的临界值是第 2.5 百分等级和第 97.5 百分等级（与单侧检验的第 5 百分等级和第 95 百分等级相对应）

双侧检验 ($H_a: \mu \neq$ 声称值)

统计显著性 (statistical significance)：如果样本平均数的标准分数在临界值 -1.96 以下或者临界值 1.96 以上，双侧检验就在 0.05 水平具有统计显著性，对于 0.01 的显著性水平来说，临界值是 -2.575 和 2.575。

P 值 (P -Value)：为了计算双侧检验样本平均数的 P 值，首先假定检验是单侧的，使用样本平均数的标准分数计算 P 值，然后将计算出的 P 值乘以 2 倍就是双侧检验的 P 值。

下述的一个例子可以帮助我们将这些观念分类。比如，一个制药公司想要确认它生产的“500 毫克”阿司匹林片剂是否真的包含 500 毫克阿司匹林的成分。如果片剂包含的阿司匹林成分少于 500 毫克，消费者将得不到广告所声称的剂量。如果片剂包含的阿司匹林成分多于 500 毫克，消费者将可能服药过量。虚无假设说的是阿司匹林成分的总体平均数是 500 毫克：

$$H_0: \mu = 500 \text{ 毫克}$$

制药公司对阿司匹林成分的平均重量多于或少于 500 毫克的可能性均感兴趣。由于制药公司对所声称的 500 毫克的平均数两侧的可能性都给予关注，我们采用双侧检验，备择假设如下。

$$H_a: \mu \neq 500 \text{ 毫克}$$

假如制药公司随机选择了一个样本， $n = 100$ 片，测量得到它们的平均重量是 501.5 毫克，再假定总体标准差 = 7.0 毫克，那么，样本平均数的标准分数是：

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{501.5 - 500}{7 / \sqrt{100}} = 2.14$$

这个标准分数超过双侧检验的临界值 1.96，因此在 0.05 水平上具有统计显著性（在 0.01 水平

上不具有统计显著性，因为它没有超出 0.01 水平的临界值 2.575)。这个结果给了我们一个很好的理由拒绝虚无假设，并且得出结论说，“500 毫克”的阿司匹林片剂平均重量不等于 500 毫克。

我们通过查表 A-1 来得出 P 值，结果显示标准分数 $z = 2.14$ 左侧的面积是 0.9838。因而，如果这是一个单侧检验， P 值应该是右侧的面积，即 $1 - 0.9838 = 0.0162$ 。因为这是一个双侧检验，我们将这个数字乘以 2 就可以得到 P 值是 $0.0162 \times 2 = 0.0324$ 。换句话说，在虚无假设为真的前提下，作为所发现的极端值，抽出这种样本的概率是 0.0324。

例 9-8 人类平均体温是多少

再次回顾马里兰大学的研究者所测量的人体温度的研究，样本包括 $n = 106$ 位健康成人（参考 8.2 节的例 8-7）。测得样本量平均体温 = 98.20 °F，样本标准差 = 0.62。假定总体标准差和样本标准差是一样的。样本是否提供了证据来拒绝下述的人们关于体温一般观念，即人体的平均体温是 98.6 °F。

解答 虚无假设是关于人体平均温度是 98.6 °F 这个人们一般的观念，或者用公式表示为 $H_0: \mu = 98.6$ °F。我们想检验的备择假设是人体平均温度不是 98.6 °F，可以用公式表示为： $H_1: \mu \neq 98.6$ °F。“不等于”形式的备择假设告诉我们需要双侧检验。已知样本容量 $n = 106$ ，样本平均数 = 98.2，假定总体标准差 = 0.62。那么样本平均数的标准分数就是：

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{98.20 - 98.60}{0.62 / \sqrt{106}} = -6.64$$

这个标准分数 < 0.05 显著性水平的临界值为 -1.96，并且小于 0.01 显著性水平的临界值 -2.575（见图 9-6）。

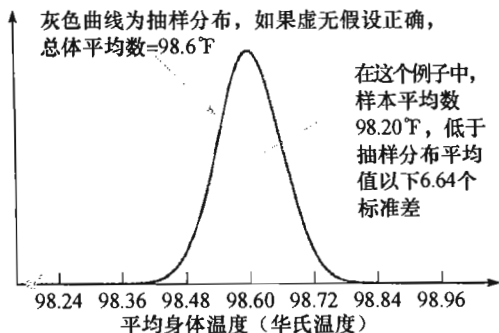


图 9-6 零假设表明人类的平均体温为华氏 98.6 °F，如果这是正确的话，样本平均数分布的平均数就是 98.6，基于样本容量和标准差，样本平均数为华氏 98.2 °F，低于所声称的平均体温 6 个标准差，这样的话说明这个声称很可能是错误的，人类的平均体温并不等于华氏 98.6 °F

因为这个结果在 0.01 水平上具有统计显著性，它给我们提供了一个否定虚无假设的强有力的证据。因此我们拒绝虚无假设，并且得出结论，人类的平均体温不等于 98.6 °F。（在这个例子中，标准分数 -6.64 偏离很远，我们不能在表 A-1 中找到其所对应的精确 P 值，采用计算器可以计算出这个 P 值大约是 3×10^{-11} 。■

9.2.3 假设检验的两类错误

我们现在已经涉猎了假设检验的全部理论基础，以及用于样本平均数假设检验特定的方法。然而，即便假设检验是正确进行的，常见的两种类型的错误还是会影响到统计结论。为了理解这些错误，回顾一下 9.1 节结尾部分法学上的类比，在这种类比中，虚无假设是：被告是无罪的。如果我们得出结论被告是有罪的，在事实中，他是无罪的，这是一种类型的错误。在任何情况下，虚无假设（无罪）被错误拒绝了。另外一种类型的错误是这样的，如果我们判决被告无罪而事实上他是真的有罪的。在这种情况下，我们不拒绝虚无假设就犯了错误。

用一个统计案例来说，再次考虑制药公司检验他们所声称的阿司匹林片剂平均含量为 500 毫克 ($H_0: \mu = 500$ 毫克)。在通过检验做出一个统计结论以后，公司可能会犯以下两种类型的错误。

- 制药公司可能拒绝虚无假设，得出结论——阿司匹林片剂平均含量不是 500 毫克，事实上，阿司匹林片剂平均含量就是 500 毫克。这类错误可能导致浪费时间和金钱来完成一个持续的修复过程。这种类型的错误，也就是虚无假设被错误地拒绝了，叫做 I 型错误 (type I error)。
- 公司可能不拒绝虚无假设，事实上，阿司匹林片剂平均含量不是 500 毫克。在这种情况下，公司可能在分配片剂时导致阿司匹林分量或多或少，消费者可能提出诉讼请求。这种类型的错误，也就是说没有拒绝虚无假设是错误的，这个称做 II 型错误 (type II error)。

表 9-2 概括了四种可能的情况。

表 9-2 针对 H_0 和 H_1 统计决断表

		真实情况	
		H_0 真	H_1 真
统计决断	拒绝 H_0	I 型错误	正确决断
	不拒绝 H_0	正确决断	II 型错误

如果认真思考，你就会认识到在假设检验的显著性水平和 I 型错误（错误的拒绝 H_0 ）之间存在重要的联系：显著性水平就是犯 I 型错误的概率。例如，如果我们使用 0.05 的显著性水平建立一个假设检验，当虚无假设为真的时候，我们拒绝虚无假设就会犯错误，这个概率是 0.05。如果我们使用 0.01 显著性水平，错误地拒绝虚无假设的概率就是 0.01（II 型错误的概率也可以量化，但它超出了本书的范围）。

例 9-9 身体温度检验的错误

考虑例 9-8 中的虚无假设，它是指身体平均温度为 98.6 °F。

(1) 在虚无假设的前提下，正确的统计决断可能是什么样的？

(2) 说明在这种情况下，I 型错误和 II 型错误的意思是什么？

(3) 在例 9-8 中，我们拒绝虚无假设，做出这样的统计决断犯错误的概率是多少？

解答 (1) 任何一个假设检验都有两种可能的正确决断。在这个例子里，第一种正确的统计决断是指，如果平均体温确实是 98.6 °F，并且我们不拒绝虚无假设；第二种正确的统计决断是，如果平均体温事实上不是 98.6 °F，并且我们拒绝虚无假设。

(2) 当我们拒绝虚无假设但事实上它是真的，这时候会发生 I 型错误。在这个例子里，如果平均体温真的是 98.6 °F，但是我们得出结论却说不是 98.6 °F，这种情况下发生的错误是 I 型错误；如果我们不拒绝虚无假设但事实上它是假的，在这种情况下发生的错误是 II 型错误。在这个例子里，如果平均体温不是 98.6 °F，但是我们却没有拒绝虚无假设，这时所发生的错误就是 II 型错误。

(3) 我们在 0.01 显著性差异水平上拒绝虚无假设。这样，拒绝虚无假设而实际上它是真的，犯下这种 I 型错误的概率是 0.01。

选择假设的偏差

如我们所看到的，一个设计良好的假设检验会遵循一个相当严格的过程，它尽力使破坏检验的偏差降低到最小的概率。尽管如此，偏差还是会以很多方式存在。尤其可能发生在假设最初的选择上。考虑下述的一个情况，调查一家工厂排放污染物进入附近一条河流的水平是否超出了政府所允许排放的最高水平。虚无假设的合理选择是实际上的污染物排放水平等于政府所允许排污的最高水平。

H_0 : 平均污染量 = 政府允许的最高水平

然而，这让我们放弃了备择假设的两个合理选择，每一个都会导致一些偏差，影响最终的结论。

如果厂方实施这个检验，他们会声称他们的污染水平低于政府允许的最高水平。也就是他们将选择：

H_a : 平均污染量 < 政府允许的最高水平

在这种备择假设下，两个可能的结果是：

(1) 拒绝虚无假设，在这种情况下我们将得出结论，这家工厂的平均污染水平低于所允许的最高水平。这一结果将有利于工厂官方。

(2) 不拒绝虚无假设，在这种情况下，检验不能做出结论。

换句话说，通过选择左侧检验（“少于”），工厂官方能够确保最坏的情况下，我们也只能承认没有证据说明工厂所排放的污染量超过了所允许的最高污染水平，可能这样的结论工厂是乐于接受的。

现在假定一个环保组织进行这个检验。因为环保组织怀疑这个工厂违反了政府标准，他所选择的备择假设可能是所排放的平均污染量大于允许的最高水平，用公式表述为：

平均污染量 > 允许的最高水平

由于选择这个备择假设，拒绝虚无假设意味着这家工厂违反了标准，当不能拒绝虚无假设是不能得出结论的。换句话说，选择右侧检验（“大于”）会产生这样一种情况，工厂可能会被发现违反国家标准，但是不能证明污染量没有超标。

例 9-10 黄金矿区

珍贵金属矿山的成功开采有赖于所采集的矿石纯净度和矿石的市场价格。假定金矿石的纯净度是至少每吨矿石包含 0.5 盎司黄金，以便保持一个特定的矿山开发。金矿石样本通常被用来估计矿山中全部矿石的纯净度。讨论 I 型错误和 II 型错误对于两种可能的备择假设的影响。

H_0 : 纯净度 < 0.5 盎司 / 吨

H_a : 纯净度 > 0.5 盎司 / 吨

解答 对于右侧检验的情况，虚无假设和备择假设如下：

H_0 : 纯净度 = 0.5 盎司 / 吨

H_a : 纯净度 < 0.5 盎司 / 吨

在这种情况下两种可能的统计决断如下：

(1) 拒绝虚无假设, 得出结论说, 金矿石的纯净度小于开发荒山所需要的水平, 因此矿山不能开发。

(2) 不拒绝虚无假设, 这意味着没有足够的证据得出结论, 金矿石的纯净度小于开发荒山所需要的水平, 所以矿山可以开发。

I 型错误 (错误拒绝一个真的虚无假设) 意思是指矿山封闭, 但事实上, 金矿石的纯净度足够开发矿山。对于矿山的开发者来说, 这意味着矿山的潜在价值被遗漏了。II 型错误 (没有拒绝一个假的虚无假设) 意思是指矿山可以继续开发, 但是实际上它却是没有价值的。

现在假定, 如果矿石的纯净度大于 0.5 盎司/吨, 矿山就持续开发。这是一个右侧检验的情况, 虚无假设和备择假设分别是:

H_0 : 纯净度 = 0.5 盎司/吨

H_a : 纯净度 > 0.5 盎司/吨

两种可能的统计决断分别如下:

(1) 拒绝虚无假设, 得出结论说, 金矿石的纯净度大于开发荒山所需要的水平, 因此矿山继续开发。

(2) 不拒绝虚无假设, 这意味着没有足够的证据得出结论, 金矿石的纯净度大于开发荒山所需要的水平, 所以矿山可以封闭不再开发。

I 型错误 (错误拒绝一个真的虚无假设) 意思是指矿山继续开发, 但事实上, 金矿石的纯净度不够, 是没有价值的。II 型错误 (没有拒绝一个假的虚无假设) 意思是指封闭矿山, 但是实际上它却是有价值的矿石, 让雇员停止工作是没有理由的。■

9.2 练习题

统计素养和批判性思维

1. 符号 要想建立总体平均数的假设检验, 理解这些符号 n 、 $\bar{x}$ 、 s 、 σ 和 μ 所代表的意思是非常重要的。描述每一个符号所代表的变量。
2. 心率 一个大学要设计一个检验以验证大学指导教师的平均心率大于每分钟 70 次。能够测量大学中 225 位指导教师的心率, 并且计算出有足够的证据实施这个声称。这个计划错误在哪里?
3. 错误类型 I 型错误是什么? II 型错误是什么?
4. 统计显著性 在对阿特金斯健康饮食法的检验中, 当检验采用此饮食法体重会减轻的声称, 计算得出 P 值是 0.004, 在 12 个月以后平均减去的体重是 2.1 磅。这个平均减去的 2.1 磅体重是否具有统计显著性? 平均减重 2.1 磅是否具有实践价值? 为什么?

判断正误: 判断练习题 5 ~ 12 中的陈述是有意义的 (表述正确) 还是无意义的 (表述错误), 并加以解释。并非所有的陈述都有明确的答案, 所以你的解释比判断更重要。

5. 假设检验 在一项假设检验中, P 值和显著性水平是相等的。
6. P 值 在检验有关总体平均数声称的时候, 检验统计值 $z=0.20$ 比检验统计值 $z=2.00$ 所对应的 P 值要大。
7. P 值 在一项假设检验中, P 值是 0.999 说明你应该支持备择假设。
8. 备择假设 在检验监狱犯人的平均智商分数小

于 100 这样的声称中, 备择假设表达为 $\bar{x} < 100$ 。

9. 统计显著性 假设检验的显著性水平就是犯 I 型错误的概率。
10. I 型错误 当一个消费者组织检验阿司匹林片剂的含量是 350 毫克的声称时, 不能错误地拒绝一个真的虚无假设是非常重要的。这样, 最好选择 0.01 水平而不是 0.05 的统计显著性水平。
11. 显著性水平 因为统计显著性水平就是犯 I 型错误的概率, 选择一个 0 显著性水平是非常明智的, 这样就不存在犯错误的概率。
12. 记忆术 一种解释假设检验中 P 值的方便记忆是: “如果 P 值很低, 那么就会接近零”。

概念与应用

备择假设能否得到支持? 在练习题 13 ~ 20 中, 计算标准分数值 z , 并且确定在 0.05 显著性水平上备择假设能否得到支持?

13. $H_a: \mu > 100$, $n=100$, $\bar{x}=102$, $\sigma=15$ 。
14. $H_a: \mu > 69.0$, $n=400$, $\bar{x}=75.2$, $\sigma=12.0$ 。
15. $H_a: \mu < 98.6$, $n=64$, $\bar{x}=98.2$, $\sigma=0.6$ 。
16. $H_a: \mu < 1518$, $n=81$, $\bar{x}=1516$, $\sigma=14$ 。
17. $H_a: \mu \neq 37.40$, $n=400$, $\bar{x}=37.30$, $\sigma=0.80$ 。
18. $H_a: \mu \neq 24.7$, $n=100$, $\bar{x}=24.0$, $\sigma=1.9$ 。
19. $H_a: \mu \neq 1050$, $n=49$, $\bar{x}=1058$, $\sigma=20$ 。
20. $H_a: \mu \neq 777$, $n=900$, $\bar{x}=779$, $\sigma=42$ 。

计算 P 值: 在练习题 21 ~ 34 中, 使用表 5-1 计算标准分数所对应的 P 值, 并且确定在 0.05 显著性水平上备择假设是否得到支持。

21. 备择假设 $H_a: \mu < 25$, $z = -0.4$ 。

22. 备择假设 $H_a: \mu < 727, z = -2.4$ 。
 23. 备择假设 $H_a: \mu > 36.35, z = 1.9$ 。
 24. 备择假设 $H_a: \mu > 0.227, z = 1.5$ 。
 25. 备择假设 $H_a: \mu \neq 0.389, z = -1.6$ 。
 26. 备择假设 $H_a: \mu \neq 172, z = -2.0$ 。
 27. 备择假设 $H_a: \mu \neq 75, z = 1.7$ 。
 28. 备择假设 $H_a: \mu \neq 25.7, z = 1.9$ 。
 29. 备择假设 $H_a: \mu > 19.4, z = 2.7$ 。
 30. 备择假设 $H_a: \mu > 75, z = 3.5$ 。
 31. 备择假设 $H_a: \mu < 1007, z = -2.1$ 。
 32. 备择假设 $H_a: \mu < 149.6, z = -1.1$ 。
 33. 备择假设 $H_a: \mu \neq 90.3, z = 0.15$ 。
 34. 备择假设 $H_a: \mu \neq 1022, z = 3.5$ 。
 35. **P 值的解释** 假定你进行一项检验, 备择假设形式是 $H_a: \mu >$ 声称值。如果样本平均数对应的标准分数 $z = -1.0$, 你将做出什么样的统计结论? 为什么不必在实际中建立正式的假设检验?
 36. **P 值的解释** 假定你进行一项检验, 备择假设形式是 $H_a: \mu <$ 声称值。如果样本平均数对应的标准分数 $z = 0.5$, 你将做出什么样的统计结论? 为什么不必在实际中建立正式的假设检验?
- 平均数的假设检验:** 对于练习题 37 ~ 50, 采用 0.05 的显著性水平, 根据本章所描述的四步骤建立一个完整的假设检验。并且明确陈述你的研究结论。
37. **洛普民意调查** 一项洛普民意调查从拥有汽车的人群中随机选出 100 个人的样本。在样本内, 这些人使用一辆汽车的平均时间是 7.01 年, 标准差是 3.74 年, 请检验一个汽车大代理商所声称的下面这个说法, 所有汽车的平均使用寿命少于 7.5 年。
 38. **医院停留时间** 按照疾病控制中心的研究, 全美国婴儿出生后在医院的平均停留时间是 2.0 天。参考自己所在医院的记录, 一位医院院长计算出 81 位女性样本在孩子出生后在医院的平均停留时间是 2.2 天, 标准差是 1.2 天。假定这代表了总体的一个随机样本, 检验如下声称“这所医院妈妈在孩子出生后的住院停留时间要长于全国的平均水平”, 如果样本标准差是 0.8 天, 这个结果又会是怎样的?
 39. **包装** 一种非常有名的意大利面品牌的制造者需要确认, 包装上注明是“16 盎司”的意大利

面, 其实际重量既不多 (这导致产品的成本过高) 也不少 (这会导致消费者的不满)。随机抽取了 $n = 144$ 包意大利面作为一个样本, 测得其平均重量是 15.8 盎司, 标准差是 1.6 盎司。检验下面的声称, 包装上注明是“16 盎司”的意大利面其实际平均重量就是 16 盎司。

40. **摩托车的生产制造** 一家摩托车的生产厂商必须生产一种车轴以适合特殊的规格。特别是这种车轴的直径必须是 8.5 厘米, 随机抽取了包括 $n = 64$ 个车轴的一个样本, 测得其平均直径是 8.56 厘米, 标准差是 0.24 厘米, 检验这家厂商所生产的车轴实际上符合这种特殊规格。
41. **药物的数量** 一种治疗感冒的药物“Dozon-al”标明, 每盎司药液包含 600 毫克对乙酰氨基酚作为药物的活性成分。食品和药物管理委员会检验了 65 份 1 盎司药液样品, 测得样品中的平均对乙酰氨基酚成分是 589 毫克, 标准差是 21 毫克, 检验食品和药物委员会声称, 这种药物没有达到所需要的对乙酰氨基酚成分含量标准。
42. **收入** 犹他州瓦沙奇近期的平均家庭收入是 41 045 美元, 标准差是 1 605 美元。假如抽取 13 267 户居民作为犹他州的一个随机样本。依据这个样本, 检验犹他州的平均年收入与全联邦的平均年收入 35 100 美元是有差异的。
43. **燃料消费** 按照能源信息管理机构的数据, 所有机动车的平均公里数是每加仑 21.4 公里, 一个由 40 辆运动越野车组成的随机样本, 平均公里数是每加仑 19.8 公里, 标准差是每加仑 3.5 公里。检验如下声称, 所有运动越野车的公里数低于每加仑 21.4 公里。
44. **棒球** 选取了由 40 个新棒球组成的随机样本。每一个棒球都被扔到水泥地面上, 弹起来的平均高度是 92.67 英尺, 标准差是 1.79 英尺 (数据来自《今日美国》)。检验下面的说法, 新棒球弹起的平均高度低于旧棒球弹起的 92.84 英尺的平均高度。
45. **强迫购物狂** 研究者们编制一个问卷来识别强迫购物狂。由 32 名确诊为强迫购物狂的被试组成的一个随机样本, 他们在问卷上的平均分数是 0.83, 标准差是 0.24 (数据来自《强迫购物狂的临床筛选》, 法伯尔和吉恩, 《消费者研究杂志》第 19 卷)。检验下面的声称, 自我识别的强迫购物狂在问卷中

的平均分数大于一般人群 0.21 的分数。

46. 硬币重量 根据美国财政部规定, 25 美分的平均重量是 5.670 克。随机抽取 50 枚硬币测得其平均重量是 5.622 克, 标准差是 0.068, 该实验表明, 发行的 25 美分硬币的平均重量是 5.670。
 47. 出生时的体重 母亲在怀孕时服用维生素制剂, 这样产下的 121 个男婴在出生时的平均体重是 3.67 千克, 标准差是 0.66 千克 (数据来自纽约州卫生部门)。检验下面的说法是否成立, 母亲在怀孕期间服用维生素制剂所产下男婴的平均重量等于全部男性婴儿的平均体重——3.39 千克。
 48. 汽车的里程数 按照能源信息委员会的说法 (数据来自联邦高速数据委员会), 美国的机动车每年的平均里程数是 11 725 公里, 一家汽车租赁公司的所有者从他全部的汽车中随机选取了 225 辆汽车作为样本。对于这个样本来说, 平均年里程数是 12 145 公里, 标准差是 3 000 公里。检验下面的说法, 这家租赁公司的平均年里程数要大于美国的平均水平。
 49. 对贪污犯的控告 随机选择 70 名被控告贪污的犯人, 他们被宣判服刑的平均时间长度是 22.1 个月, 标准差是 8.6 个月 (数据来自美国司法部)。杰尼·弗莱明服务于一家警务机关, 从事被控告有罪的人的干预治疗, 检验他的这种说法, 因为被控贪污而入狱服刑的平均刑期少于 24 个月。
 50. 收入 按照美国人口普查局的最近一次调查, 获得学士学位的男性平均年收入是 46 700 美元, 假定随机抽取具有学士学位的 400 名女性作为一个样本, 她们的平均年收入是 28 700 美元 (这也是全美国女性的平均年收入)。标准差是 16 500 美元, 检验下面的声称: 女性的平均年收入少于男性的平均年收入。
- I 型错误和 II 型错误:** 在练习题 51~54 中, 已知虚无假设和备择假设, 不用虚无假设和备择假设这两个术语来说明 I 型错误和 II 型错误。
51. 虚无假设: 患者所患的某种特定疾病已经治愈, 备择假设: 患者具有某种疾病。
 52. 虚无假设: 被告无罪。备择假设: 被告有罪。

53. 虚无假设: 彩票是公平的。备择假设: 彩票里面有欺诈。
54. 虚无假设: 一种新型奥迪汽车的悬浮系统中, 一种螺丝平均长度是 3.456 厘米。备择假设: 一种新型奥迪汽车的悬浮系统中, 一种螺丝平均长度不等于 3.456 厘米。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 9 章的有用链接 “links for internet projects”。

55. 与全国平均水平的比较, 选择在班级里或一个学生样本中相对容易测量的几个变量。这些变量应该包括能够被计算平均数的量化数值 (例如, 身高、体重、家庭人口数、血压、心率、反应时等), 使用互联网或者其他参考工具得到这些变量的全国平均数。搜集这些变量的数据, 选取至少包括 50 名个体的随机样本。对此进行假设检验, 并且确立样本平均数和总体平均数是否有显著差异。
56. 乡村的数据 《美国统计学概要》和最近的人口调查, 提供了一个有关社会科学、经济学和人口统计的县水平、州水平和统计水平的大量数据。应用互联网, 以下述方式比较各州和全美的数据: (1) 选择你所关注的一个变量, 包含某一个特定州的平均数, 例如, 伊利诺伊州的家庭平均人口数; (2) 计算这个变量的最近的全国数据, 例如, 全美的家庭平均人口数; (3) 选择这个州的某一特定的县, 并且得到这一县的对应数据作为样本数据, 例如, 伊利诺伊州库克县的家庭平均人口数; (4) 假定这一个县是这一个州的随机样本, 检验下面的声称, 这个州在这个变量上的平均水平, 高于或者低于全国平均数; (5) 讨论和解释你的结果, 假设检验依赖样本是总体的一个随机样本。讨论你的结论中这个因素的影响。
57. 假设检验的程序 以“假设检验和程序”为关键词, 使用谷歌等搜索引擎, 使用这样一个程序并且运行它, 描述这个程序是怎样工作的, 以及它能说明什么。
58. 检验力 使用搜索引擎谷歌, 比如寻找“统计检验力”, 描述统计检验力是什么意思。



9.3 总体比率的假设检验

对于比率的假设检验，前面一节已经说明了所有的基础知识，现在我们只需要一个不同的方法来计算抽样分布的标准差，让我们用一个例子来说明这个过程。假定一名政党候选人，在即将到来的选举前，委托进行一项民意测验，随机抽取 400 个可能的投票者作为样本，调查结果发现，其中的 204 人支持这名候选人。这名候选人应该有获胜的信心吗？在第 8 章，我们讨论了怎样确立这种类型的误差幅度和置信间距，现在我们将这一问题放入假设检验中，对于一项假设检验来说，我们问民意测验的结果（也就是指样本统计量）是否支持如下的假设。

候选人拥有超过 50% 的投票者，通常我们用 p 来代表投票者总体中支持这名候选人的比例，我们用字母 $\hat{p}$ 代表样本中的人群支持这名候选人的比例，因为样本中支持候选人的人数是 204 人，所以样本比率是：

$$\hat{p} = 204/400 = 0.510$$

现在我们用公式来表达虚无假设和备择假设，通常我们将虚无假设表示成为一个等式， $H_0: p = 0.5$ （50% 的投票的支持候选人），候选人想知道他是否获得大多数人的支持，所以备择假设是右侧形式。 $H_a: p > 0.5$ （超过 50% 的投票者支持候选人）。

9.3.1 总体比率假设检验的计算

我们怎样才能确定样本中是否有足够的证据拒绝虚无假设，回顾假设检验的四个步骤，我们已经完成了前两步（用公式表示假设，以及搜集样本数据）。现在，如同我们在 9.2 节中处理样本平均数那样，我们必须确定在虚无假设为真的前提下，样本结果偶然发生的可能性是多少。

方法注解

样本比率分布的标准差的公式，只有当 $np \geq 5$ 和 $n(1-p) \geq 5$ 时才准确，在本书中所讲的一些问题会涉及这个条件。

如同我们处理样本平均数那样的过程，我们设想选择了许多样本容量为 $n = 400$ 的样本，

对于每一个样本，我们计算其中支持候选人的比率。接下来，因为我们有足够大的样本容量，因此，样本比率的分布应该是非常接近正态分布的。在虚无假设为真的前提下（即总体中的人们支持候选人的比率是 0.5），这个分布的峰值就应该是虚无假设所声称的总体比率 $p = 0.5$ 。这个抽样分布的标准差可以根据下面的公式计算得来：

$$\text{样本比率分布的标准差} = \sqrt{\frac{p(1-p)}{n}} \quad (9-4)$$

对于这个例子而言，样本标准差是：

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.5(1-0.5)}{400}} = 0.025$$

图 9-7 呈现了样本比率的分布。我们这样考虑，这次民意测验的样本比率（ $\hat{p} = 0.510$ ）作为抽样分布的一个单独的点，像平均数假设检验那样（见图 9-1），我们也可以观察到以下规律。

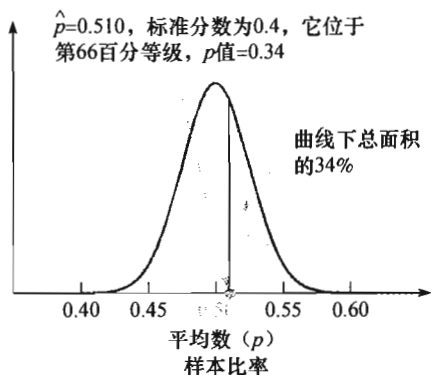


图 9-7 一次选举民意调查的样本比率分布（总体平均数 p 是 0.50，标准差是 0.025；样本比率是 0.510，标准分数是 0.4， P 值是 0.34。）

(1) 如果样本结果接近抽样分布的峰值，那么我们就没有理由认为虚无假设是假的，我们不能拒绝虚无假设。

(2) 如果样本结果远离抽样分布的峰值，可能的理由是抽样分布的分值并不是虚无假设所声称的，在这种情况下，我们拒绝虚无假设。

和样本平均数的假设检验一样，通过用标准分数量化距离的方法，我们确定样本比率是接近还是远离抽样分布的峰值（假定虚无假设是真的前提下），标准分数的公式总是同样的形式，对于样本比率的分布，公式就变成：

$$z = (\text{样本比率} - \text{总体比率}) / \text{抽样分布的标准差} \\ = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \quad (9-5)$$

现在我们就能够计算并解释上面例子中的标准分数了, 样本容量 $n = 400$, 样本比率是 51%, 总体比率是 0.5%, 这样样本比率的标准分数就是:

$$z = (\text{样本比率} - \text{总体比率}) / \text{抽样分布的标准差} \\ = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.510 - 0.5}{0.025} = 0.4$$

换句话说, 这次民意测验中的样本比率是高于样本比率分布峰值的 0.4 个标准差, 如你在图 9-7 中所看到的, 这个样本结果并不很极端, 这说明我们并不应该拒绝虚无假设, 也就是说候选人并没有获得大多数投票者的支持。

假设检验中的样本比率的标准分数:

已知样本容量 (n), 样本比率 ($\hat{p}$), 以及声称的总体比率 (p), 样本比率的标准分数就是:

$$z = (\text{样本比率} - \text{总体比率}) / \text{抽样分布的标准差} \\ = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

相关知识

在 1948 年美国总统选举中, 大多数调查机构和报纸预测共和党候选人杜威 (Dewey) 战胜民主党候选人杜鲁门 (Truman), 当选总统, 事实上, 杜鲁门赢得了大选, 而不是盖洛普预测的杜威 49.5% 对杜鲁门的 45.1%。

9.3.2 显著性水平和 P 值

我们可以通过使用显著性水平和 P 值进一步量化。先使用显著性水平, 因为选举权的民意测验这个例子使用了右侧检验, 我们可以通过查表 9-1 来得到显著性水平的临界值, 标准分 $z = 0.4$ 并不大于 0.05 显著性水平下的临界值 $z = 1.645$, 这证实这个结果不应该使我们拒绝虚无假设。下面计算 P 值, 我们使用表 A-1, 这个表显示从左边开始到标准分数 $z = 0.4$ 这部分的面积是 0.6554, 因为这是一个右侧检验, 我们用 1 减去这个值, 得到右侧的面积 (见图

9-7), 这也就是这个假设检验的 P 值。 $P = 1 - 0.6554 = 0.3446$, 这个 P 值告诉我们, 在虚无假设为真的前提下, 随机抽取一个样本, 就如这次民意测验中所得到的作为极端值的概率大于 0.34, 随机抽取这样一个样本的概率如此之高, 我们没有理由拒绝虚无假设。所以候选人没办法假定他获得 50% 以上投票者的支持。

9.3.3 总体比率假设检验的概括

现在我们可以概括一下, 对总体比率进行假设检验的步骤, 必须注意我们仍然会沿用前面所说的假设检验的常用四个步骤, 我们仅仅关注如何处理比率检验的特殊性。

(1) 因为我们要解决总体比率的问题, 虚无假设的形式是 p 等于声称值。为了做出统计决断, 拒绝还是不拒绝虚无假设, 我们必须确定, 如果虚无假设为真的前提下, 在假设检验中, 样本作为一个所发现的特定例子, 是更可能还是更不可能发生。

(2) 我们根据样本比率的标准分数来确定这个比率的可能性。标准分数的计算公式如下:

$$z = (\text{样本比率} - \text{总体比率}) / \text{抽样分布的标准差} \\ = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

公式中, n 是样本容量, $\hat{p}$ 是样本比率, p 是虚无假设所声称的总体比率。

(3) 然后, 我们使用标准分数, 完成以下两个步骤, 这和平均数的假设检验过程是一样的。

A: 我们通过比较计算得来的标准分数, 和查表 9-1 所得来的临界值, 来确定统计显著性水平。

B: 我们通过查表 A-1 来计算 P 值, 对左侧检验来说, P 值就是从左开始到标准分数之间, 这一部分曲线下的面积。对右侧检验来说, 它就是从标准分数到右侧这一部分曲线下的面积。对于双侧检验来说, 它就是单侧检验所计算出的 P 值的两倍。

(4) 如果在所选择的水平上具有统计显著性 (通常是 0.05 或 0.01 显著性水平), 我们拒绝虚无假设, 如果不具有统计显著性, 我们就不拒绝虚无假设。

例 9-11 当地的失业率

假定全美的失业率为 31.5%，在威斯康星州一个县里，调查了 450 名被试，结果发现 22 人处于失业状态，基于前述的调查数据，本地官方上报到州当地的失业率上升到全国的平均水平。请在 0.05 水平上检验这个声称。

解答 让我们采用四个基本步骤来解决这个问题。第一步，全美失业率是一个总体比率（与总体平均数相对应）。虚无假设就是，假设当地的失业率等于全美的失业率，或者用公式表达为 $H_0: p = 0.135$ 。备择假设就是县政府所声称的失业率高于全国平均水平，或者用公式表达为 $H_a: p > 0.135$ 。注意这是一个右侧检验。第二步，样本统计量就是样本容量 $n = 450$ 和样本比率，根据已知数据，样本比率等于：

$$\hat{p} = 22/450 = 0.0489$$

第三步，我们现在确定，在虚无假设为真的前提下，抽出一个特定样本，样本比率至少是 0.0489。这种可能性是多少，我们计算样本比率的标准分数：

$$z = (\text{样本比率} - \text{总体比率}) / \text{抽样分布的标准差} \\ = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.0489 - 0.035}{\sqrt{\frac{0.035(1-0.035)}{450}}} = 1.60$$

这个标准分数非常接近右侧检验的临界值 $Z = 1.645$ 。但是并不大于这个值。所以这个结果在 0.05 水平上并不具有统计显著性。从表 A-1 中查出，标准分数 1.60 对应的 P 值是 $1 - 0.9452 = 0.0548$ ，这说明如果虚无假设是为真的前提下，随机抽取这样一个样本的概率大于 0.05，图 9-8 显示了抽样分布的这些基本观念。

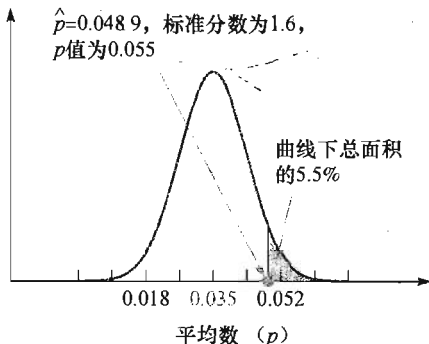


图 9-8 样本失业率

注：例 9-11 中样本比率的分布，样本比率 $\hat{p} = 0.0489$ 没有达到 0.05 显著性水平的临界值右侧。

第四步，因为这个检验并没有超出 0.05 显著性水平的临界值，我们不能拒绝虚无假设，换句话说，这个样本不能提供足够的样本来支持县政府所声称的，本地失业率大于全国平均水平。■

例 9-12 左利手的总体

随机抽取了 750 个人的样本，发现其中 92 人是左利手，使用这个样本数据检验“10% 的人口是左利手”这个说法。

解答 我们再次按照假设检验的四个步骤来完成上述问题。

第一步，这个声称与人口中左利手的比率有关，所以这是一个总体比率的假设检验。虚无假设就是所声称的人口 10% 为左利手，或者用公式表达为 $H_0: p = 0.1$ 。为了检验这个声称，我们需要计算实际的人口比率，小于或大于 10%，这样被择假设检验就是 $H_a: p \neq 0.1$ ，这是一个双侧检验。

第二步，样本统计量是样本容量 $n = 750$ 以及样本中左利手的比率：

$$\hat{p} = 92/750 = 0.123$$

第三步，样本比率的标准分数是：

$$z = (\text{样本比率} - \text{总体比率}) / \text{抽样分布的标准差} \\ = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.123 - 0.1}{\sqrt{\frac{0.1(1-0.1)}{750}}} = 21.09$$

0.05 显著性水平的双侧检验的标准分数的临界值是小于 -1.96 或者大于 1.96，这个检验的标准分数 2.09，大于 1.96，所以我们得出结论这个检验在 0.05 水平显著。通过查表 A-1，从 2.09 到右侧的曲线下的面积，是 $1 - 0.9817 = 0.0183$ ，这就是单侧检验的 P 值。因为这是一个双侧检验，所以我们将这个 P 值乘以 2 等于 0.0366，图 9-9 呈现了这个分布的意义。

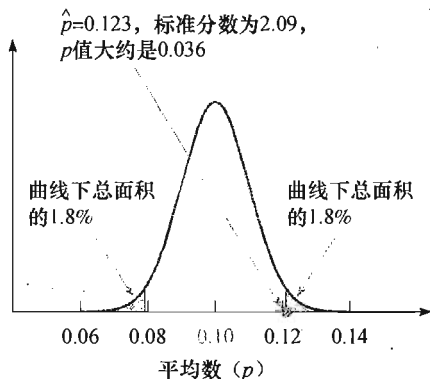


图 9-9 左利手人的样本比例

注：例 9-12 的样本比率分布中的样本比率 $\hat{p} = 0.123$ 所在位置。由于是双侧检验， P 值所对应的是在峰两侧超出 2.09 个标准差的位置以外曲线下的面积。

第四步，因为这个检验在 0.05 水平显著，我们拒绝这个虚无假设，并且得出结论说，左利手的总体

比率不等于10%，记住这个显著性水平就是犯I型错误的概率，我们就知道当虚无假设实际上是真的，而

我们拒绝它，从而犯下I型错误的概率实际上就是0.05。

9.3 练习题

统计素养和批判性思维

1. 符号 符号 p 、 $\hat{p}$ 和 P 值分别代表什么意思？
2. 分布 在建立这一节所描述的一项假设检验时，会用到哪一种特定的分布？为什么？
3. 抽样 美国在线进行了一项调查，在调查中互联网的使用者被问到一个问题。在 96 772 名被调查者中，76 885 名被调查者回答说是。使用这个样本结果检验如下的声称是否有效，这个说法是总体中的大多数对这个问题通常回答“是”。为什么？
4. P 值 当使用样本数据检验“大多数汽车事故发生于离家 5 公里以内”这个说法时，计算得到的 P 值是 0.000 01。这个 P 值会告诉我们什么？

判断正误 判断练习题 5 ~ 12 中的陈述是有意义的（表述正确）还是无意义的（表述错误），并加以解释。并非所有的陈述都有明确的答案，所以你的解释比判断更重要。

5. 虚无假设 检验大多数美国人同意对所有持枪者进行注册，虚无假设是： $p = 0.5$ 。
6. 备择假设 检验大多数美国人是否同意的对所有持枪者进行注册，备择假设是： $p > 0.5$ 。
7. 备择假设 在对一个总体比率的声称进行双侧检验时， P 值是标准分数右侧的面积。
8. 备择假设 如果样本比率小于 0.5，所声称的总体比率 $p > 0.5$ 一定不能被支持。

概念与应用

假设检验：对于练习题 9 ~ 18，采用 0.05 的显著性水平，并且使用文章中所描述的统计检验四步骤来建立一个假设检验，并应明确陈述你的结论。

9. 投票者民意测验 在一项选举前的民意测验中，一位地方检察官的候选人得到了 400 个投票者中 205 人的支持。假定这些人的民意测验代表了投票总体的一个随机样本，检验大多数投票支持这位候选人的这个说法。
10. 大学中的女性 按照美国人口普查局的数据，25 岁及以上的大学生中 58% 是女性。假设在克莱瑞恩大学抽取 2 100 名 25 岁及以

上的大学生作为一个随机样本，其中 1 234 名大学生是女性。检验这所大学的女性的百分数高于全国平均水平的说法。

11. 成绩的压力 美国教育部所进行的一项研究，数据来源于 1 015 名随机选择的青少年，结果发现 44% 的青少年回答说学习成绩是他们最大的压力来源。检验“全部青少年中将近半数感觉学习成绩是他们最大的压力源”这个说法。
12. 结婚的成年人 在近年，125 800 000 名成年人，或者说全部美国成年人中的 58.6%，是已婚的。在一个新英格兰城镇，一个包括 1 445 名成年人的随机样本，其中 56.0% 的人是已婚的。检验这个声称：来自总体样本数的结婚率低于 58.6%。
13. 药物滥用 美国卫生和人类服务部的一项有关 12 ~ 17 岁青少年的非法药物使用的研究，报告说药物滥用从 1997 年的 11. % 下降到现在 9.9%。假定在一所高中重复进行这个调查，随机选取 1 050 名中学生作为样本，结果发现其中有 98 名使用非法药物。检验该中学校长所说的在他的学校里的非法使用药物人数要低于现在的全国平均水平。
14. 贫困 按照最近的估计，艾奥瓦州的卡斯特县 4 342 人中有 12.1% 生活在贫困线以下。假定这个县的人代表了艾奥瓦州一个随机样本，通过这个样本，检验艾奥瓦州个贫困率低于全国 13.3% 的平均水平。
15. 堕胎调查 加州大学洛杉矶分校（UCLA）的高等教育研究所进行的对于一年级大学生的年度调查，针对他们对各种事物的态度，调查了大约 276 000 名大学生。按照最近调查结果，51% 的一年级大学生认为堕胎应该是合法的（这个比例在 1990 年是 65%）。检验下述说法是成立的，“超过半数的一年级大学生认为堕胎应该是合法的”。
16. 吸烟 全部美国成年人（超过 21 岁）的吸烟率是 32%（数据来自美国药物滥用学会），假如随机抽取了 75 名超过 21 岁的工

艺美术专业大学生作为样本，吸烟率是35%。采用0.05的显著性水平检验所有工艺美术专业大学生吸烟率高于全国平均水平。

17. 天然气使用 按着能源信息委员会的数据，1997年全国范围内53.0%的家庭使用天然气。最近由3600个家庭组成的随机样本显示出54.0%使用天然气。采用0.05的显著性水平检验全国使用天然气53.0%的平均水平是否改变。
18. 临床试验 在对药物“立普妥”进行的一项临床试验中，863名患者采用这种药物进行治疗，其中19名出现流感的症状（数据来自Parke Davis）。检验采用这种药物治疗出现流感症状的人数比率大于没有服用药物的罹患流感者的1.9%的比率。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第9章的有用链接“links for internet projects”。

19. 左利手 已知10%的美国人是左利手，在你所在的大学里随机选择不少于50名同学并确定他们是否为左利手。采用假设检验公式来验证上述的说法。
20. 吸烟 使用互联网或学校图书馆的参考资料，确定美国人吸烟的比率。检验你所在大学的大学生中吸烟比率是否与全部美国人吸烟的总体比率有差异。选择样本数据为至少50名所在学校的大学生。
21. 女性大学生 使用互联网或学校图书馆的参

考资料，美国大学里女性大学生的比率。检验你所在大学女性比例不同于全部美国大学生的比率。选择样本数据不少于100名大学生。

22. 县市的数据 《美国统计概览》和最近的美国人口普查提供一个全面的有关社会、经济和生命统计的数据以及当地的数据资料。使用网络信息，以下面的方式对州和全国的数据进行比较。

- (1) 选择你所关注的一个变量，包含某一个特定州的比率（例如，亚利桑那州生活在贫困线以下的人口比率）。
- (2) 计算这个变量的最近的全国数值（例如，全美的生活贫困率）。
- (3) 选择这个州的某一特定的县，并且得到这一县的对应数据作为样本数据（例如，亚利桑那州皮马县的生活贫困率）。
- (4) 假定这个县是这—个州的随机样本，检验下面的声称：这个州在这个变量上的比率，高于或者低于全国的比率。
- (5) 讨论和解释你的结果，假设检验依赖样本是总体的一个随机样本。讨论你的结论中这个因素的影响。

新闻中的统计学

23. 新闻中的假设检验 寻找近期的一篇新闻报道文章，其中描述了（可能不是非常明确的）一个有关总体比率的假设检验，剪贴这篇报道并概括其所使用的方法。

第9章复习题

1. 随机选择36听可乐，并且测量可乐的容量（盎司）。下面就是所测的样本数据，平均数是12.19盎司，标准差是0.11盎司。假如我们采用0.05水平显著性，检验一听可乐的平均容量大于12盎司。假定总体标准差已知， $\sigma = 0.115$ 盎司。

12.3 12.1 12.2 12.3 12.2 12.3
12.0 12.1 12.2 12.1 12.3 12.3
11.8 12.3 12.1 12.1 12.0 12.2
12.2 12.2 12.2 12.2 12.2 12.4
12.2 12.2 12.3 12.2 12.2 12.3
12.2 12.2 12.1 12.4 12.2 12.2

- (1) 虚无假设是什么？
- (2) 备择假设是什么？
- (3) 样本平均数12.19盎司对应的标准分数值是多少？
- (4) 临界值是多少？
- (5) P 值是多少？
- (6) 统计结论是什么？（要明确原始的声称是平均数大于12盎司）
- (7) 描述这个检验的I型错误。
- (8) 描述这个检验的II型错误。
- (9) 如果所检验的声称变成平均数不是12盎司的话（代替平均数大于12盎司），

计算检验的 P 值。

2. 在一项通过采用可替宁贴治疗方法戒烟的研究中, 39 位烟民在治疗一年后, 其中 32 位戒烟时间已经达到一年 (数据来自《高剂量可替宁贴治疗方法》, 戴尔等, 《美国医学联合会杂志》, 274 卷第 17 期)。我们想要采用 0.05 的显著性水平检验如下的声称: “采用可替宁贴治疗方法戒烟一年后, 大多数吸烟者仍在吸烟”。

- (1) 虚无假设是什么?
- (2) 备择假设是什么?
- (3) 样本比率所对应的标准分数是多少?
- (4) 临界值是多少?
- (5) P 值是多少?
- (6) 统计结论是什么? (要明确原始的声称是“采用可替宁贴治疗方法戒烟一年后, 大多数吸烟者仍在吸烟”。)

(7) 描述这个检验的 I 型错误。

(8) 描述这个检验的 II 型错误。

(9) 如果所检验的声称变成“总体比率等于 0.5”, 计算 P 值是多少。

3. 我们选择检验如下的声称: “克拉克性别选择方法对于增加女孩出生的概率是有效的。”在由 80 多位采用了克拉克性别选择方法的夫妇所组成的随机样本中, 结果发现在 80 个新出生的婴儿中其中 35 位是女婴。对如上的声称我们将做出怎样的统计结论的? 为什么不必再经过统计假设检验正式的全部步骤?
4. 迈克·珊丽想要检验如下这个假设, “成年男性的平均体重大于 150 磅”。她调查了家庭中的成年男性, 并且获得了 40 个样本数据, 这个数据支持她的备择假设“成年男性的总体平均体重大于 150 磅”。她的这个研究过程最基本的缺陷在哪里?

第9章小测验

1. 声称“经常运动的男性比率大于 0.2”的备择假设是什么?
2. 检验如下声称“经常运动的男性比率大于 0.2”, 所使用的虚无假设是什么?
3. 声称“因为贪污受贿而被判入狱服刑的重型犯的比率等于 0.6”, 其备择假设是什么?
4. 声称“ $P > 0.05$ ”是左侧检验、右侧检验还是双侧检验?

在练习题 5~10 中, 我们想要采用 0.05 水平来检验如下的声称: “职业喜剧的演员的平均智商分数大于 110”。

5. 虚无假设是什么?
6. 备择假设是什么?
7. 如果检验结果所计算的 P 值是 0.2500, 对于上述的声称将做出什么样的统计结论?
8. 关于虚无假设两种可能的统计结论分别是什么?
9. 关于被检验的声称两种可能的统计结论分别是什么?
10. 如果你错误地得出结论, 职业喜剧演员的平均智商分数大于 110, 但事实上他们的平均智商等于 100, 你所犯的误差是 I 型误差还是 II 型误差?

技术应用

为了描述总体平均数的假设检验, 这一章只考虑了正态分布应用的例子。第 10 章会介绍使用 t 分布的一些方法。在第 10 章的结尾处可以看到这些其他方法的使用过程。

SPSS

总体平均数的假设检验: SPSS 并没有一个程序适用于总体标准差已知的前提下, 建立针对一个所声称的总体平均数的假设检验。在第 10 章的结尾部分可以看到这样的程序, 它适用于总体标准差未知的前提下, 对一个总体平均数声称的假设检验。

总体比率的假设检验: SPSS 并没有一个程序适用于针对一个总体比率声称的假设检验。

Excel

检验所声称的一个总体平均数: 使用数据工作平台 Data Desk/XL, 将它安装于 Excel。Data Desk/XL 是一个统计插件, 它完善了 Excel 的标准统计功能。它可以有效地用于分析实际数据。在子目录 DDXL 下找到文件 ReadMe.txt, 打开这个文件可找到安装指南。

首先进入样本数据列 A。如果使用 Excel 2003, 点击 [DDXL]。如果你使用 Excel

2007, 请点击 [Add-Ins], 然后再点击 [DDXL]。选择 [Hypothesis Tests] (“假设检验”)。在功能键中, 选择 [1 Var z Test]。然后点击铅笔图标, 输入数据值的范围, 如果在 A 数据列有 45 个数据值, 你就输入 A1:A45。然后点击 [OK]。然后顺序完成对话框中所列出的四个步骤, 在第四个步骤成以后点击 [Compute], 这样就会得到 P 值、检验统计量和统计结论。

检验所声称的一个总体比率: 使用数据工作台 Data Desk/XL, 将它安装于 Excel。

首先在单元格 “A1” 输入成功的次数, 然后在单元格 “B1” 输入实验的总次数。如果使用 Excel 2003, 点击 [DDXL]。如果你使用 Excel 2007, 请点击 [Add-Ins], 然后再点击 [DDXL]。选择 [Hypothesis Tests] (“假设检验”)。在功能键中, 选择 [Summ 1 Var

Prop Test]。(“使用一种变量的概括数据来检验所声称的总体比率”)。然后点击内容为 “Num Successes” (“成功次数”) 并且输入 “A1”, 然后点击内容为 “Num Trials” (“试验总次数”) 并且输入 “B1”, 然后点击 [OK]。在对话框中, 然后顺序完成对话框中所列出的四个步骤, 在第四个步骤成以后点击 [Compute], 这样就会得到 P 值、检验统计量和统计结论。

STATDISK

选择 [Analysis] 然后再选择 [Hypothesis Tests]。对于这一章所讨论的方法, 选择 [Mean- One Sample] 或者 [Proportion- One Sample]。在出现的对话框中, 点击左上角的框, 选择与所检验的相匹配的项目, 然后点击 [Evaluate], 结果就会出现检验统计量和 P 值。

聚焦健康和教育

你所受的教育是否有利于你长寿

在这一章里, 我们关注的是假设检验最简单的形式, 我们想检验有关总体的一个说法, 看是否能通过一个单独的样本搜集的证据来支持我们的假设。毫无疑问, 很多有趣的统计问题需要更复杂的分析。想想看, 比如这样的问题, 哪些因素使得人们更加长寿。

我们经常听到一些缩短我们寿命的因素。例如, 吸烟、过量饮酒、肥胖这些因素, 都有可能使人们早逝。知道这些可以帮助人们避免不良生活方式。但是有哪些因素能够提高人们的寿命呢?

尽管这个问题比我们在这一章里遇到的其他问题都更难回答, 但是这个问题也有它遵循的基本规律。例如, 假定一个研究者认为吃燕麦能使一个人更长寿。为了检验这个猜想, 这个研究者建立了一次假设检验。首先建立虚无假设: 吃燕麦与长寿没有关系。其次对数据进行检测寻找证据来支持应该拒绝虚无假设, 从而得出结论说食用燕麦确实能够促进长寿。这项研究最困难的地方在于搜集数据并举例说明, 找到吃燕麦和不吃燕麦的人进行比较, 以及找到一种方法, 能够把食用燕麦对寿命的影响和其他大量变量也对长寿的影响作用分离开来。

在过去几十年里, 研究者们已经发现了很多能够促进长寿的因素。例如, 富有和长寿呈正相关, 长跑对于长寿也有重要的促进作用, 有节制地饮食对于健康也存在积极的影响, 运动通常对于促进长寿会起到积极作用。令人惊奇的是, 一个因素比其他因素更能促进长寿, 那就是受教育的年限。这意味着你在学校呆得时间越长, 你的寿命也就越长, 最起码在平均水平上看来是这样。图 9-10 和图 9-11 呈现了支持这个结论的一些数据。



是生病了不去上学，还是因为不能上学而生病了？

什么因素使一个特定群体的人比其他人群更能健康长寿，在大量的研究之后，世界各国的研究者一致认为教育是教育，新的研究发现，相关事实上反映了原因和效应。

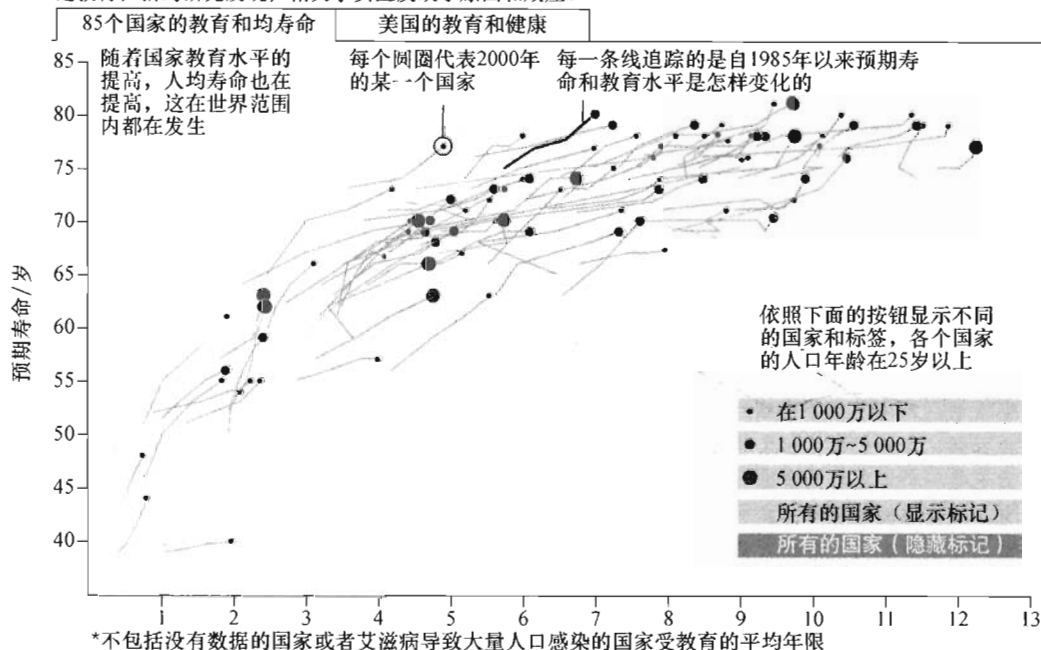


图 9-10 受教育年限与寿命

注：图中的点对应的是不同国家人们的预期寿命与受教育年限。每一个国家的线开始于1985年的预期寿命，结束于2000年的预期寿命（点）。实际上所有的线几乎都是向上的，显示了随着受教育年限的增长，预期寿命也会随之增长。（数据来源：《纽约时报》）

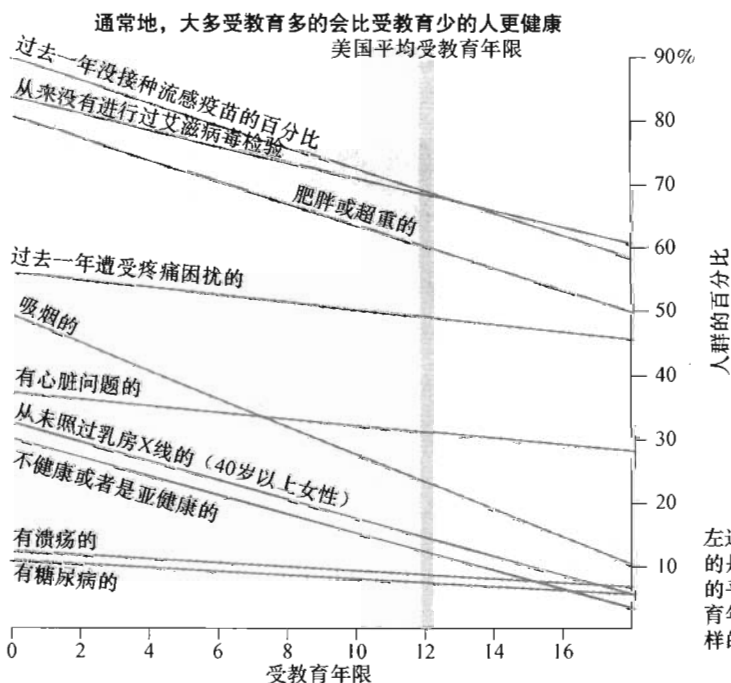


图 9-11 受其他因素影响的寿命

注：图中呈现了人数百分比，各种健康风险因素趋向于减少在校年限。例如，最上面的线显示了90%没有接受正规教育的人在过去的这一年里没有接种过任何疫苗，同时60%的人只接受了16年的教育，许多风险因素随着受教育年限增长而下降，这说明教育是导致健康生活方式的可能原因。（数据来源：《纽约时报》）

为什么受教育年限越长，人的寿命会越长呢？研究者们已经发现了很多可能的原因。例如，受教育年限越长的人，可能更少地从事缩短寿命的行为，例如吸烟，过度饮酒，而是会更多地参加一些积极的、有利于长寿的活动，如运动。另一种有意思的假设是，教育使人们学会为了得到长远的回报先做出短暂的牺牲（例如，现在为你所受的教育做出投资，可以期望以后会有个高收入的工作）。这种延迟满足能力使人们能够做出更有利于长寿的决定。当然，没有人知道确切的答案，也许还有更多的假设检验要去完成。

尽管如此，根据现有的数据，我们至少能够得到一个和你阅读此书有关的非常有意思的结论：因为需要，你来参加统计课程的学习，但是它同时还能促使你更加长寿，尤其当你在学校里学习的时候。

问题讨论

1. 图 9-10 和 9-11 都呈现了大量的数据。详细地解释每个图。确认你理解了每一条直线（或者曲线）所表达的意思。每个图是怎样支持受教育年限越长会越长寿的假设。请说明原因。
2. 为什么你认为教育能够促进长寿？例如，你认为是受教育年限越长导致了人们的长寿，还是受教育年限影响其他因素，进而再影响寿命？怎样检验你的假设？
3. 吸烟被公认为是导致疾病和早逝的最危险因素。因此，政府和关爱健康的机构经常组织耗费巨大的活动来劝阻人们吸烟（或者从来都不吸烟）。但是图 9-11 显示随着受教育年限的增加，吸烟的人数急剧下降。那么，为了禁止吸烟，与其把钱花在反对吸烟的活动上，倒不如将钱花在增长人们的受教育年限上更有意义。你是否认为这是一个好主意？请说明你的意见。
4. 根据教育年限能够促进长寿的发现，还有事实说明长寿使得对于健康的投资减少，并且有利于提高生产效率。一些人提议政府应该增加教育资金投入，使更多的人接受更长时间的教育。你是否支持这项建议，并阐明你的理由。

建议阅读

1. Kolata, G. "A Surprising Secret to a Long Life; Stay in School." *New York Times*, January 3, 2007.
2. Lleras-Muney, Adriana. "The Relationship Between Education and Adult Mortality in the United States." *Review of Economic Studies*, Vol. 72, No. 1, 2005.

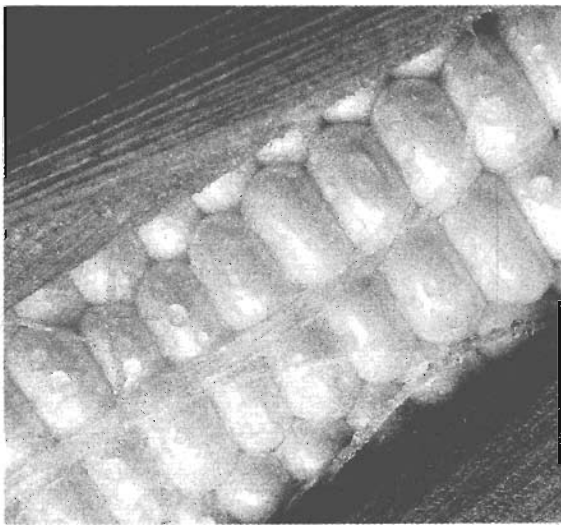
| 聚焦农业 |

转基因食品是否安全

英国的报纸称它们为基因改造食物（“Frankenfood”），它来源于单词“Frankenstein”（意思为毁灭创造者自身之物）。大多数欧洲人根本就不吃这些食物。但是今天，美国销售的很多食物，包括大约 36% 的美国玉米和 55% 的美国黄豆，都属于这一类。我们都在说“转基因食品”，或者 GM 食品。GM 食品是开始于 20 世纪 90 年代中期的农业变革。尽管如此，它们还是引起了农业历史上最大的争论。

转基因生物体是指科学家人为地在某种生物体中插入一个基因，此基因在自然条件下不存在于该生物体的种类或其近支亲本中。各种目的不同的转基因生物体被检验并被投入使用。例如，科学家发现了一种细菌，它含有产生胰岛素的基因。

在农业生产中，转基因技术使科学家们能够培育一种新的农作物，采用传统的种植技术很难或者根本不可能获得这种农作物。我们用一种被广泛种植的转基因农作物——“Bt 玉米”来解释转基因技术的含义。玉米经常被各种害虫破坏，结果，农民经常在庄稼上喷洒农药。可不幸的是，很多农药还会对除了害虫以外的其他动物有害，因此会造成环境污染（也不利于人类食用）。



Bt 是一种细菌，准确名称是“苏云金芽孢杆菌”（“bacillus thuringiensis”）。Bt 在自然条件下生活在土壤

中。早在1911年,科学家发现了Bt能产生能杀死某种害虫的毒素。不同类型的Bt能杀死不同的害虫,但是对其他的动物和人类一般是无害的。直到1960年,这些特点使得Bt作为一种环境友好型农药喷洒在农作物上。不幸的是,Bt农药非常贵而且效率低,因为只有当害虫吃到它的时候才发生作用(更有效的农药只要接触便能发生作用)。另一个原因是它在阳光下很快被分解,下雨的时候也会被冲刷掉,只有在恰当的时间它才能发挥作用。

转基因技术解决了Bt农药的问题。Bt细菌产生一种特殊的蛋白质,这种蛋白质便是这种农药中发生作用的东西。科学家找到了产生这种蛋白质的基因,然后把这种基因转移到植物,例如玉米中。一旦玉米中包含了这些基因,它便能够产生像Bt细菌产生的能杀死害虫的蛋白质。喷洒农药就没有必要了,因为玉米本身就能杀死害虫。也就不必担心农药分解或被雨水冲刷掉而不起作用。

相关知识

超过50种新转基因农产品获准在美国销售,包括杀虫基因的玉米和大豆,以及存放更长时间的西红柿。发展中的转基因农产品包括抗擦伤的马铃薯、口味更佳的大豆以及包含更多维生素和其他更多营养的粮食,以提高贫穷国家的营养水平。

Bt玉米相对于传统玉米的优势非常明显,但是,它是否也会存在一些不利因素呢?这就是关于GM食品引起争论的方面。一方面,许多科学家说GM食品非常安全,它们有助于提高全世界人民的营养水平。然而,另外一些人,也包括一些科学家,把GM食品称做科学怪食物(“Frankenfoods”),还声称这是至今发明的最危险的技术之一。

从广义上说,关于GM食品是否安全的问题主要涉及以下三个方面:

1. GM食品对人类是否有毒害作用?
2. GM食品中包含的新蛋白质对一些人会不会产生过敏反应?
3. GM农作物是否会引起一些看不见的环境污染,例如把它们的基因移植到种子中(从而制作“超级种子”),或者杀死除了害虫以外的其他动物?

这些问题可以通过假设检验来验证。针对每一个问题,我们都会首先建立一个虚无假设,具体表述为:传统食品和GM食品在安全性上没有差别。例如,针对问题一所建立的虚无假设是:GM食品不比传统食品里(通常包含极少量的有毒物质)所包含的毒害物质含量更多。备择假设:GM食品与传统食品有毒物质的含量是有差异的。科学家们进行实验检验这个假设。如果实验提供的证据提示两种食品之间的毒害物质有显著差异(超过了预计的随机因素所产生的差异),就有理由拒绝虚无假设。

不幸的是,到今天为止,实验仍然不能解决这个争议。例如,一些转基因农作物确实包含对人类有害的物质——它们因此不能获准进入市场销售。对于GM食品的支持者来说,这个事实也提供了支持他们的证据,因为它显示当前的监管程序(例如,GM食品需要通过美国食品和医药管理会的检验)充分地确保了只有安全食品才会进入市场。同样道理,尽管GM食品中的一些蛋白质无疑会引起一些过敏反应,但是,很多科学家认为他们知道了这些反应,会更有利于他们确保只有安全的食品才会通过检验。

反对GM食品的一方使用同样的实验结果来支持他们的观点。他们认为尽管科学家已经在实验中证明了一些明显的有害物质存在,但是还有可能会忽视了长期的影响,这些影响可能会在很多年以后才显现出来。同样道理,他们声称我们肯定不可能了解所有的过敏反应,因此通过检验的GM食品有可能会对一部分人造成严重的过敏反应。

环境问题就更难研究清楚了。例如,一项研究表明Bt玉米对“美洲王蝶”(一种蝴蝶,又叫大斑蝶,分布于美洲)有害,这些蝴蝶不是害虫,本来也没打算伤害它们。但实验是在实验室进行的,不能准确地代表在真实的环境中所发生的一切。同样,存在“基因跳跃”的可能性——Bt基因会从玉米扩散到其他植物中,现在还不能对其机制形成很好地理解,也很难对其进行研究。

在未来的一些年里,关于GM食品的争论很可能还在继续,甚至会更加激烈。毕竟,当涉及食品时,大家都会非常关注的。

相关知识

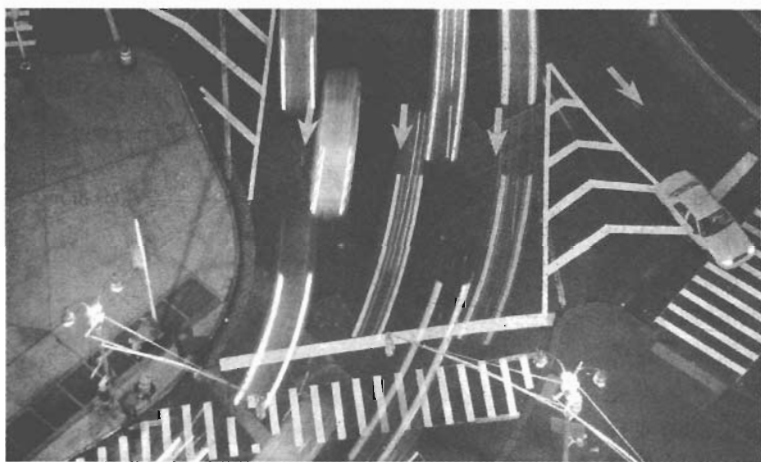
转基因食品的反对者经常指出,发现中毒和环境污染的长期效应非常艰难,杀虫剂DDT就是一个典型例子。在发现它的有害效应时,DDT已经使用了60年。

问题讨论

1. 设计一个实验，用来检验 GM 食品的安全性，例如玉米。详细描述你的实验，并讨论在实验过程中和结果解释时会遇到的实际困难。
2. 忽略 GM 食品的安全问题，列举 GM 食品会对人类有益的方面。然后，忽略 GM 食品的益处，列举 GM 食品可能会对人类产生危害的方面。通过以上分析，你认为 GM 食品的进一步应用是应该提倡还是禁止？说明你的理由。
3. 一些人建议让消费者自行选择 GM 食品，只需要在包含转基因成分的食品上贴上标签予以注明就可以。你是否认为这是一个好办法？为什么？
4. 调查关于 GM 食品争论的最近发展情况，新的信息有没有改变争论中的焦点？请加以解释。

建议阅读

1. 关于转基因食品的争论频繁产生的头条新闻，为发现最近的研究，请以“转基因食品”为关键词进行网络搜索。
2. Belsie, Laurent. "Superior Crops or 'Frankenfood'?" *Christian Science Monitor*, March 1, 2000.
3. Barrionuevo, Alexei. "Can Gene-Altered Rice Rescue the Farm Belt?" *New York Times*, August 16, 2005.
4. Rosenthal, Elisabeth. "Biotech Food Tears Rifts in Europe." *New York Times*, June 6, 2006.
5. Rosenthal, Elisabeth. "Questions on Biotech Crops with No Clear Answers." *New York Times*, June 6, 2006.
6. Yoon, Carol K. "Pollen from Genetically Altered Corn Threatens Monarch Butterfly, Study Finds." *New York Times*, May 20, 1999.



这个世界是必然性和偶然性交织在一起的，已经习惯于必然却发现某些事物反复无常的人会非常苦恼，而有人却把这些归因于机遇。

——歌德

第 10 章 t 检验、双向表和方差分析

学习目标

■ 10.1 平均数的 t 检验

理解在构建一个置信区间或者建立总体平均数的假设检验时， t 分布比正态分布有着更广泛的适用性，理解如何正确运用 t 分布。

■ 10.2 双向表的假设检验

理解和掌握独立变量数据被编制成双向表时的假设检验。

■ 10.3 方差分析（一元方差分析）

理解和掌握使用一元方差分析的方法完成假设检验。

我们在第 1~9 章探索了统计学的核心理念和应用，如果继续你的统计学习，你会遇到更多的统计应用。在最后一章，我们将探讨三个特别常用的统计学方法，你很可能在以后的课程中会遇到。它们也有助于你更好地理解统计学在日常生活中的应用。这三个应用方面都建立在第 9 章所学习的重要统计假设检验技术的基础上，我们以 t 检验开始，它既用于置信区间也用于假设检验，然后探讨两个变量间的假设检验，最后是众所周知的方差分析（ANOVA）。



10.1 平均数的 t 检验

8.2 节讨论了总体平均数置信区间的估计。当假设样本平均数分布是正态分布时，我们估计误差幅度 E 为：

$$E \approx \frac{2s}{\sqrt{n}}$$

注意，误差幅度的精确公式使用 1.96 而不是 2。1.96 是右侧 0.025 的面积所对应的临界标准分数 z （表 A-1）。

9.2 节讨论关于所声称的某一特定总体平均数的假设检验，同样也假设抽样分布符合正态分布。我们使用下述公式来计算样本平均数的标准分数：

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

9.2 节也提供了拒绝虚无假设的临界值（在 0.05 显著性水平上）：左侧检验 $z \leq -1.645$ ，右侧检验 $z \geq 1.645$ ，双侧检验 $z \leq -1.96$ 或者 $z \geq 1.96$ 。-1.645、1.645、-1.96、1.96 这些值都来自标准正态分布。

在实际应用中，统计学专家对置信区间和

假设检验不一定都假定符合正态分布。（事实上，一些专业学术期刊上的文章显示正态分布实际上很少应用。）其主要原因是，当使用正态分布的标准分数公式，我们需要已知总体标准差。因为我们通常并不能知道总体标准差，我们必须用样本标准差来估计总体标准差。这样统计学家们就提出了一种不需要知道总体标准差的方法。这样的一种情况下就会用到学生 t 分布（student t distribution），或者简称 t 分布（ t distribution）。这是在我们不知道总体标准差的情况下也可以应用的一种分布，它只需要样本容量大于 30 或者总体分布是正态分布就可以。

相关知识

学生 t 分布是威廉·戈塞特（William Gosset）发现的，他当时是吉尼斯啤酒厂的雇员，需要一个可用于相对较小样本的分布，他工作的爱尔兰啤酒厂不允许公开发表研究成果，所以戈塞特以笔名“学生”发表了他的研究。

总体平均数的推断：在 t 分布和正态分布之间进行选择

t 分布：总体标准差未知并且符合正态分布或者当总体标准差未知并且样本容量大于 30。

正态分布：总体标准差已知并且总体符合正态分布或者总体标准差已知并且样本容量大于 30。

在本书中我们不会详细讲解 t 分布的性质。但是它的基本原理非常容易理解： t 分布形态和对称性与正态分布非常相似，但是对于小样本来说预计两者之间会有很大的变异性。图 10-1 将不同样本容量 $n=3$ 和 $n=12$ 两种情况下的 t 分布和标准正态分布进行比较。请注意，对于较大的样本容量， t 分布比较接近于正态分布。实际上，样本容量越大， t 分布就越接近于正态分布。

t 分布的真实值，是我们可以将置信区间或者假设检验的应用扩展到更多我们不能使用正态分布的情况中。因为使用正态分布要求总体标准差已知。但是要记住，尽管如此， t 分布也并非适用所有情况。例如，如果我们有一个样本容量在 30 以下的小样本，或者样本数据提示，总体分布完全有别于正态分布。这些情况下， t 分布和正态分布都不能应用的。这需要用

到本书中没有讨论过的其他方法来解决。

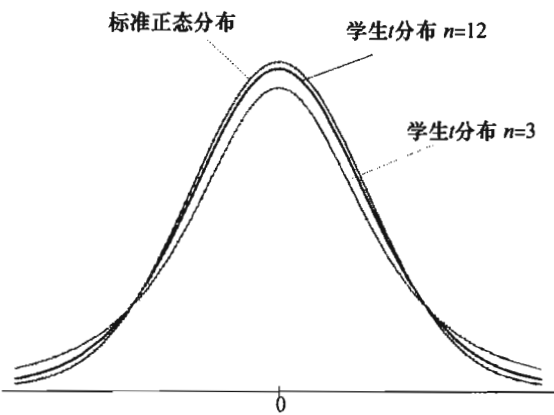


图 10-1 两种不同样本容量的 t 分布和标准正态分布的比较

注：请注意，随着样本容量的增大， t 分布就越来越接近正态分布。

10.1.1 使用 t 分布来估计置信区间

如同我们在 8.2 节和 9.2 节中所提到的对于正态分布的假定。我们也可以使用 t 分布来建构总体平均数的置信区间或者建立一个总体平均数的假设检验。然而，替代依据标准 z 分数（比如 -1.645 , 1.645 , -1.96 , 1.96 ）来确定显著性水平，我们使用 t 值，常用的 t 值呈现在表 10-1 中（统计软件也能计算 t 值）。

表 10-1 t 值标准表

自由度 ($n-1$)	单侧面积	
	0.025	0.05
	双侧面积	
	0.05	0.10
1	12.706	6.314
2	4.303	2.920
3	3.182	2.353
4	2.776	2.132
5	2.571	2.015
6	2.447	1.943
7	2.365	1.895
8	2.306	1.860
9	2.262	1.833
10	2.228	1.812
11	2.201	1.796
12	2.179	1.782
13	2.160	1.771
14	2.145	1.761

(续)

自由度 (n-1)	单侧面积	
	0. 025	0. 05
	双侧面积	
	0. 05	0. 10
15	2. 131	1. 753
16	2. 120	1. 746
17	2. 110	1. 740
18	2. 101	1. 734
19	2. 093	1. 729
20	2. 086	1. 725
21	2. 080	1. 721
22	2. 074	1. 717
23	2. 069	1. 714
24	2. 064	1. 711
25	2. 060	1. 708
26	2. 056	1. 706
27	2. 052	1. 703
28	2. 048	1. 701
29	2. 045	1. 699
30	2. 042	1. 697
31	2. 040	1. 696
32	2. 037	1. 694
34	2. 032	1. 691
36	2. 028	1. 688
38	2. 024	1. 686
40	2. 021	1. 684
50	2. 009	1. 676
100	1. 984	1. 660
无限大	1. 960	1. 645

我们从置信区间开始介绍。为了确定置信区间，我们首先必须计算误差幅度 E ，在 t 分布中，所用的公式是：

$$E \approx t \cdot \frac{s}{\sqrt{n}} \tag{10-1}$$

式中， n 是样本容量， s 是样本标准差， t 是我们在表 10-1 中所得到的值。唯一的“新方法”是从表中找到正确的 t 值。我们按照以下步骤来进行。

(1) 首先，确定样本数据的自由度 (degrees of freedom)，自由度定义为样本容量 - 1：

$$t \text{ 分布的自由度} = n - 1 \tag{10-2}$$

(2) 其次，查表 10-1，表的第一列是自由度，找到与你的样本数据相对应的那一行的自由度。并且从这一行中找到适合的 t 值。对于总体平均数的置信区间来说，95% 置信度的 t 值

对应于第二列的数字，90% 置信度的 t 值对应于第三列的数字。(我们使用表中的“双侧面积”是因为误差幅度是在平均数上下两边；例如，95% 置信水平意思是指 t 分布左右两端加起来 0.05 的面积，如图 10-1 所示。)

一旦确定了数据的 t 值和置信水平，你就可以如同在 8.2 节中那样确定置信区间，这就要使用误差幅度 E 的新公式。

应用 t 分布估计总体平均数的置信区间

如果已知条件需要我们使用 t 分布 (总体标准差未知并且样本容量 $n > 30$ ，或者总体呈正态分布)，总体平均数真值的置信区间就可以通过样本平均数加上和减去误差幅度估计得到。也就是说，总体平均数的置信区间是：

$$\bar{x} - E < \mu < \bar{x} + E \text{ (或者表示为 } \bar{x} \pm E \text{)}$$

在公式中的误差幅度是，

$$E = t \cdot \frac{s}{\sqrt{n}}$$

t 值通过查表 10-1 得出。

例 10-1 舒张压的置信区间

随机选择 5 位成年男性，测得其舒张压分别是：78，54，81，68，66。这 5 个数据结果的样本统计量是： $n = 5$ ，样本平均数 $\bar{x} = 69.4$ ，样本标准差 $s = 10.7$ 。使用上述样本，建构所有成年男性总体平均舒张压 95% 水平的置信区间的估计值。

解答 因为总体标准差是未知的，并且有理由假定成年男性的舒张压符合正态分布，在置信区间的估计中，我们使用 t 分布来代替正态分布。样本容量 $n = 5$ ，所以样本的自由度是：

$$t \text{ 分布的自由度} = n - 1 = 5 - 1 = 4$$

在 95% 的置信水平上，我们使用表 10-1 的第二列找到 $t = 2.776$ 。现在我们使用已知的样本容量 ($n = 5$) 和样本标准差 ($s = 10.7$) 来计算误差幅度 E ：

$$E = t \cdot \frac{s}{\sqrt{n}} = 2.776 \times \frac{10.7}{\sqrt{5}} = 13.3$$

最后，我们使用误差幅度和样本平均数来计算 95% 的置信区间：

$$\begin{aligned} \bar{x} - E &< \mu < \bar{x} + E \\ 69.4 - 13.3 &< \mu < 69.4 + 13.3 \\ 56.1 &< \mu < 82.7 \end{aligned}$$

依据这 5 个样本测量值，我们在 95% 的置信水平上估计所有成年男性的总体平均舒张压在 56.1 ~ 82.7。

10.1.2 应用 t 分布进行假设检验 (t 检验)

当 t 分布用于有关总体平均数的假设检验时 ($H_0: \mu = \text{特定值}$), t 值如同我们在 9.2 节学习的假设检验中标准分数的作用是一样的。回顾一下, 我们通过将样本平均数的标准分数与临界值或者与计算出的 P 值进行比较, 然后确定统计显著性。在 t 分布中, 我们使用下列公式计算 t 值来代替计算标准分数 z 。

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \quad (10-3)$$

式中, n 是样本容量, $\bar{x}$ 是样本平均数, s 是样本标准差, μ 是虚无假设所声称的总体平均数。

当我们计算出 t 值以后, 我们就可以通过将计算出的 t 值与查表 10-1 的临界值作比较, 然后做出统计决断是否拒绝虚无假设。临界值有赖于下面不同的检验类型。

(1) **左侧检验**: 如果计算出的检验统计量 t 大于或等于表 10-1 中标注着“单侧面积”一列的临界值, 我们就可以拒绝虚无假设。值得注意的是, 对于单侧检验来说, 第二列所给出的临界值是 0.025 显著性水平, 第三列所给出的临界值才是 0.05 显著性水平。

(2) **右侧检验**: 如果计算出的检验统计量 t 小于或等于表 10-1 中标注着“单侧面积”一列的临界值, 我们就可以拒绝虚无假设。同样, 因为这也是一个单侧检验, 第二列所给出的临界值是 0.025 显著性水平, 第三列所给出的临界值才是 0.05 显著性水平。

(3) **双侧检验**: 如果计算出的检验统计量 t 的绝对值大于或者等于表 10-1 中标注着“双侧面积”一列的临界值, 我们就可以拒绝虚无假设。在这种情况下, 第二列所给出的临界值就是 0.05 显著性水平, 而第三列所给出的临界值则是 0.10 显著性水平。

所计算出的检验统计量 t 也能用来计算 P 值, 然而, 通常使用统计软件而不是使用统计表来计算 P 值。

例 10-2 平均数的右侧假设检验

下面列出的是随机选择的 10 个学习统计课程的学生们的智商分数:

111 115 118 100 106 108 110 105 113 109

使用第 4 章所学习的方法, 你可以计算出这些数据的样本统计量: $n = 10$, $\bar{x} = 109.5$, $s = 5.2$ 。采用 0.05 显著性水平检验如下声称: “学习统计课程的学生们的平均智商分数高于 100, 也就是高于一般人群的平均智商分数。”

解答 基于所声称的“学习统计课程的学生们的平均智商分数高于 100”。我们建立虚无假设 $H_0: \mu = 100$, 备择假设 $H_a: \mu > 100$, 因为所有学习统计课程的学生智力测验分数的总体标准差是未知的, 并且有理由假定所有学习统计课程的学生智力测验分数符合正态分布, 因此样本平均数分布符合 t 分布。 t 检验的统计量的值计算过程如下:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{109.5 - 100}{\frac{5.2}{\sqrt{10}}} = 5.777$$

我们现在需要将这个检验统计量的值和查表 10-1 找到的合适的临界值相比较:

- 这组数据的自由度是 $n - 1 = 10 - 1 = 9$, 根据这个我们就可以从表 10-1 中找到对应的行。
- 因为这是单侧检验, 我们要求的显著性水平是 0.05, 它所对应的是第 3 列。
- 查表 10-1 找到自由度为 9 的行和第 3 列, 这样我们就可以得到 0.05 显著性水平的临界值是 $t = 1.833$ 。

因为样本检验统计量 $t = 5.777$ 大于临界值 $t = 1.833$, 所以我们拒绝虚无假设。我们可以得出结论说, 有充分的证据支持所声称的“学习统计课程的学生们的平均智商分数高于 100”。

我们可以通过使用软件更精确地计算假设检验的 P 值, 本例结果是 0.000 135。值得注意的是 P 值远小于 0.05, 所以我们可以更高的显著性水平上拒绝虚无假设, 支持所声称的“学习统计课程的学生们的平均智商分数高于 100”。

例 10-3 平均数的双侧假设检验

使用例 10-2 中的样本数据和同样的显著性水平 0.05, 检验如下的声称“学习统计课程的学生们的平均智商分数不等于 100”。

解答 基于如下的声称: “学习统计课程的学生们的平均智商分数不等于 100”。我们建立虚无假设 $H_0: \mu = 100$ 和备择假设 $H_a: \mu \neq 100$ 。“不等于”的备择假设形式意味着采用的是双侧检验。

样本检验统计量的值仍然是例 10-2 中的结果 ($t = 5.777$), 但是双侧检验的临界值不同于单侧检验的临界值。因为我们使用的是同样的数据, 自由度仍然是 9, 这告诉我们查表 10-1 的哪一行。但是对于单侧检验来说, 查找 0.05 显著性水平的临界值是在第 2 列而不是第 3 列。表 10-1 中第 9 行和第 2 列的临界值是 $t = 2.262$ 。

要注意, 检验统计量的绝对值是 $t=5.777$, 大于查表所得临界值 $t=2.262$, 所以我们还是要拒绝虚无假设。我们可以得出结论说, 有充分的证据支持所声称“学习统计课程的学生们的平均智商分数不等于100”。(如果使用统计软件, 我们可以计算出检验统计量对应的 P 值是 0.000 269。)

例 10-4 平均数的右侧假设检验

因为花费过高的原因, 汽车碰撞试验经常使用较少的样本。在一项研究中, 5 辆宝马汽车在标准条件下进行碰撞试验, 使用修车的费用 (以美元计) 来检验如下的声称: “所有宝马汽车平均修车费用少于 3 000 美元。” 5 辆样本宝马汽车的平均修车费用是 2 835 美元, 标准差是 883 美元。使用 0.05 显著性水平检验如下的声称: “所有宝马汽车总体平均修车费用少于 3 000 美元。”

解答 基于声称“宝马汽车的平均修车费用少于 3 000 美元”, 建立虚无假设 $H_0: \mu = 3\,000$ 美元, 和备择假设 $H_a: \mu < 3\,000$ 美元。总体标准差不知。我们假定修车费用符合正态分布 (对于稍微偏离正态分

布的情况 t 检验不是很敏感, 因此这是一个合理的假定)。 t 检验的统计量计算值就是:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{2\,835 - 3\,000}{883/\sqrt{5}} = -0.418$$

对于这个例子来说, 样本容量是 $n=5$, 因此自由度是 $n-1=5-1=4$ 。因为这是一个单侧检验, 因此我们在表 10-1 的第 3 列查找 0.05 显著性水平的临界值。这样, 如果你查到第 4 行的第 3 列, 你会发现临界值 $t = -2.132$ 。对于左侧检验, 我们要寻找的值会小于或等于这个值, 或 $t = -2.132$ 。

因为样本统计量 $t = -0.418$, 它不小于或者等于临界值 $t = -2.132$, 我们不能拒绝虚无假设。这就是没有充分证据支持这个声称: “宝马汽车的平均修车费用少于 3 000 美元。” 如果使用统计软件进行计算, 你会得到检验统计量的 P 值是 0.348 8。换句话说, 如果虚无假设为真, 把所得到的样本看做一个界限值, 选择这样一个样本的概率大约是 35%。这是一个相对较高的概率, 我们没有理由认为虚无假设是错误的。

10.1 练习题

统计素养和批判思维

1. t 分布 假定你想对一个总体平均数建构一个置信区间, 或者建立一个统计假设检验, 描述 t 分布适用的条件。
2. t 分布 在建构总体平均数置信区间的估计或者对总体平均数的假设检验, 为什么 t 分布要比正态分布用得更广泛?
3. 学生 t 分布和 t 分布之间有什么不同?
4. 抽样方法 假定你想估计美国人平均携带的现金数量, 如果你调查 12 个最好的朋友, 你能使用 t 分布估计一个置信区间吗? 这个结果对总体平均数的估计是否恰当?

判断正误: 判断练习题 5~8 中的陈述是有意义的 (表述正确) 还是无意义的 (表述错误), 并加以解释。并非所有的陈述都有明确的答案, 所以你的解释比判断更重要。

5. t 检验 使用一个少于 30 个数值的样本, 对总体平均数进行检验, t 分布总是适用的。
6. 抽样方法 使用一个从正态分布总体中抽出来的样本, 对总体平均数进行检验, 总体标准差未知, 不管采集样本数据的方法是什么, 总是可以使用 t 分布。
7. t 检验 采用 t 分布检验如下这个声称“如果样本平均数是 103, 学生们的平均智商大于

105”, 这个假设检验是不必进行的, 我们可以得出结论说, 没有充分的证据支持这个声称。

8. t 分布与正态分布 因为 t 分布不需要知道总体标准差的值, 并且因为总体标准差的值很少有机会知道, 所以 t 检验在实际情况下要比使用正态分布的检验多。

概念与应用

置信区间: 在练习题 9~16 中, 使用 t 分布对总体平均数建立置信区间的估计。

9. 智商分数 从正态分布的总体中抽取一个随机智商分数的样本, 样本统计量是: 样本容量 $n=25$, 样本平均数 $\bar{x}=102.5$, 标准差 $S=12.8$, 建构总体平均数估计值的 95% 的置信区间。
10. NBA 篮球运动员的身高 从 NBA 篮球运动员中获得一个随机的身高样本, 总体接近于正态分布, 样本统计量分别是: 样本容量 $n=16$, 样本平均数 $\bar{x}=77.9$ 英寸, 样本标准差 $s=3.5$ 英寸, 建立总体平均数估计值的 95% 的置信区间。
11. 男性的肘部到手指尖的长度 选取一个男性的随机样本, 并且测量每个人肘部到手指尖的长度, 这个长度的总体符合正态分布, 样

本统计量是： $n=35$ ，样本平均数 $\bar{x}=14.5$ 英寸，标准差 $s=0.7$ 英寸，建立总体平均数估计值的95%的置信区间。

12. SAT 测验分数 随机抽取一些人的 SAT 测验分数，总体接近于正态分布，样本统计量分别是： $n=41$ ，样本平均数 $\bar{x}=1503$ ，标准差 $s=352$ ，建立总体平均数估计值的95%的置信区间。

13. 交通事故的医疗花费 一项研究用来估计使用安全带发生交通事故以后在医院的花费，随机选取的20个案例接近于钟形分布，平均数是9004美元，标准差是5629美元（数据来源于美国交通部）。

(1) 建立这个花费的平均数95%的置信区间。

(2) 如果你是一家保险公司的经营管理人员，对于使用安全带的司机给予了较低的费率，你想保守地估计最坏的情况，那么你对使用安全带所发生的交通事故在医院的花费可能的估计值是多少？

14. 天气预报和实际温度 本书的一位作者编制了实际最高气温的一份清单，并且对应于3天前天气预报的一份清单，用3天前预报的高温减去实际的高温就是每天的差值，列出31个差值的结果，平均数是-0.419度，标准差是3.704度。

(1) 建立所有实际高温和天气预报的高温之间的平均数差异估计值的95%的置信区间。

(2) 这个置信区间是否包括零度？如果这个置信区间不包括零度，我们能否得出结论说3天前的天气预报是不准确的？

15. 估计汽车污染 在7辆汽车的一个样本中测量每一辆车的氮氧排放物（克/公里），得到如下的结果：0.06, 0.11, 0.16, 0.15, 0.14, 0.08, 0.15（数据来自环境保护署）。

(1) 假定这个样本是使用中的汽车的代表性样本，建立所有汽车氮氧排放物平均数估计值的95%的置信区间。

(2) 环境保护署要求氮氧排放物要少于0.165克/公里，有人声称氮氧排放物的平均数是0.165克/公里，这个置信

区间说明这个声称无效吗？为什么？

16. 监控空气中的铅 下面列出了所测得的空气中的铅含量（微克/立方米）：

5.40, 1.10, 0.42, 0.73, 0.48, 1.10

环境保护署提供了一个空气中铅含量的量化标准1.5微克/立方米，上述的测量是世贸中心大厦5号建筑物旁不同日子的即时测量记录（世贸大厦在2001年9月11日被恐怖分子袭击）。在世贸大厦两个建筑物倒塌之后，有关空气质量量化的考虑增多了。使用上述已知的数据建立空气中铅含量的95%的置信区间。有什么事情提示这个置信区间可能不是很理想？请加以解释。

假设检验：在练习题17~24中，检验给出的声称。

17. 谷类食品的含糖量 随机抽取16种不同的谷类食品作为一个样本，测量每种谷类食品的含糖量（每克谷类食品中糖含量的克数），样本平均数是0.295克，标准差是0.168克。采用0.05的显著性水平检验如下的声称：“所有谷类食品的平均含糖量少于0.3克。”

18. 检验手表的准确度 同学们随机选取了30个人，并且测量他们手表的准确性，正的误差代表手表比实际时间快，负的误差代表手表比实际时间慢，30个值的平均数是117.3秒，标准差是185.0秒。采用0.05的显著性水平检验如下的声称：“所有手表误差的总体平均数等于0秒。”关于人们手表的准确性，我们可以得出什么样的结论？

19. 棒球 在前面的检验中，棒球从24米高掉落到坚硬的地面上，它们弹起的平均高度是92.84英寸，在对20个新球样本的测试中，弹起的平均高度是92.67英寸，标准差是1.79英寸（数据来自美国国家图书馆和《今日美国》）。使用0.05的显著性水平来确定是否有充分的证据来支持如下的声称：“新球弹起的平均高度不等于92.84英寸。”看起来新球是有差异的吗？

20. 航空无线电的可靠性 Telekronic 公司用于小型飞行器的航空无线电平均使用寿命是420小时，15个改进后的航空无线电试图提

高其可靠性，建立一个假设验证其平均寿命。这15个改进后的航空无线电平均使用寿命为442小时，标准差44.2小时，采用0.05的显著性水平检验改进后的声称无线电平均使用寿命大于420小时。这样的改进是否提高了航空无线电的可靠性呢？

21. 服用维生素对于婴儿体重的影响 一个随机抽取的由16名男婴组成的样本记录其出生时的体重，这些婴儿的母亲在怀孕期间服用特定的维生素。这个样本的平均体重是3.675千克，标准差是0.657千克（数据来源于纽约州的卫生部门）。采用0.05的显著性水平检验如下的声称：“母亲在怀孕期间服用维生素的这些男婴与所有男婴3.39千克的平均体重是有显著差异的。”基于这个结果，能否说母亲怀孕期间服用维生素对于婴儿的体重会产生影响？

22. 心率 本书的一位作者说，她的心率低于上统计课程学生的平均心率。作者的心率测量值结果是60次/分钟。并且测量了他班级里20名学生的心率。这20名学生的平均心率是74.5次/分钟，标准差是10.0次/分钟。是否有充分的证据支持本书作者的说法“上统计课程学生的平均心率高于60次/分钟”，检验采用0.05的显著性水平。

23. 奥林匹克运动的获奖者 下面列出的是在连续数届夏季奥林匹克运动会上100米短跑比赛获奖者的时间。按照奥林匹克运动会的比赛届次顺序排列。

12.0	11.0	11.0	11.2	10.8	10.8
10.8	10.6	10.8	10.3	10.3	10.3
10.4	10.5	10.2	10.0	9.95	10.14
10.06	10.25	9.99	9.92	9.96	

假定这些结果是从过去和未来的奥林匹克运动会百米短跑比赛成绩中随机抽取的样本，采用0.05的显著性水平来检验“所有的平均成绩低于10.5秒是否正确”。你观察到的这些数字的精确度是多少？在假设检验中，这个数据组的什么重要特征没有被考虑到？

24. 香烟中的尼古丁 卡罗来纳州烟草公司广告说他们公司销售最好的不带滤嘴卷烟所含的尼古丁是40毫克以下，但是《消费者权益倡导者》杂志进行了一项检验，从中随机抽取10支香烟，测量发现其中尼古丁的含量如下：

47.3	39.3	40.3	38.3	46.3
43.3	42.3	49.3	40.3	46.3

采用0.05的显著性水平检验这家杂志的声称“香烟的平均尼古丁含量大于40毫克”。

网络和远程项目

请在网站 www.aw.com/bbt 上打开第10章的有用链接“links for internet projects”。

25. 著名人士 《世界大事年鉴》包括一部分内容叫做“著名人士”，包括如下各个组成部分，建筑师、艺术家、商业领袖、漫画家以及其他几个类别。从一个类别的人群中选择一个样本然后计算他们的平均寿命。然后检验如下说法：“这个类别人群的平均寿命与一般大众的77岁平均寿命是有差异的”。

26. 戈塞特 t 分布最初是由威廉·希利·戈塞特（William Sealey Gosset）发现的。使用互联网查找“William Sealey Gosset”或者“William S. Gosset”。请写一段关于威廉·希利·戈塞特及其贡献的短评。

系。基本过程和其他假设检验是一样的。确定虚无假设和备择假设，采用样本数据检验虚无假设，然后通过样本数据是否有证据拒绝或者是不拒绝虚无假设。如果拒绝虚无假设，也就是意味着接受备择假设。

10.2.1 确定双向表的假设

假如一所大学的管理员关注在不同系中男生和女生的学位授予是否存在偏差。他们搜集



10.2 双向表的假设检验

到目前为止，我们已经学习的所有假设检验虚无假设都是声称总体平均数 μ 或者总体比率 p 等于某个值。但是也有很多情况，虚无假设也会采用其他不同形式。在这一节中，我们所设计的假设检验是要寻找两个变量之间的关

了不同系男生和女生所授予的学位人数。这些数据涉及两个变量：专业 and 性别。变量“专业”的取值较多，比如生物学、商学、数学和音乐。变量“性别”只有两个取值，男性和女性。为了检验在学位授予中是否存在偏差，管理员思考了如下问题：

“这两个变量间的数据是否存在关系？”

如果这两个变量之间有关系，那么大的男性和女性在选择专业方面会有不同的比率，这说明一个人的性别可能会影响到他的专业选择（可能由于选择也可能由于不同专业存在的某些偏见）。如果没有关系，它意味着没有证据说明性别会影响一个人选择的专业。

上述观念说明存在下述两种可供选择的假设，虚无假设 H_0 是：两个变量是相互独立的（它们之间没有关系），在我们现在的样本中，它可以表述为性别和专业之间没有关联。备择假设 H_a 的陈述是相反的，两个变量之间存在关系，在此情况下，意味着性别影响一个人的专业选择。

两个变量的虚无假设和备择假设

虚无假设（null hypothesis）， H_0 ：是指两个变量是相互独立的（它们之间没有关系）。

备择假设（alternative hypothesis）， H_a ：是指两个变量之间存在关系。

10.2.2 双向表中的数据呈现

如果假设确定了，假设检验的下一步对数据进行分析，看它是支持拒绝虚无假设还是不拒绝虚无假设。我们现在的例子所搜集的数据意思是指要找出不同专业男生和女生授予学位的人数。搜集了这样的数据，就需要用一种恰当的方法呈现这些数据。因为这些数据涉及两个变量，我们可以用双向表（two-way table）来有效地呈现这些数据，也叫做列联表（contingency table），它的命名是因为它呈现了两个变量。

表 10-2 呈现了双向表在专业 and 性别两个变量可能得到的数据。每个单元格呈现了两个变量的一种组合次数。例如，横行“女生”和纵列“生物学”这个组合的单元格，呈现了获得生物学学士学位的女生为 32 人，同样的

道理，横行“男生”和纵列“商学”这个组合的单元格，呈现获得商学学士学位的男生为 87 人。

表 10-2 由专业和性别两个变量构成的双向表

变量 1 专业 →	生物学	商学	数学	心理学	...
女生	32	110	18	75	...
男生	21	87	15	70	...

↑
变量 2 性别

注：一个变量呈现在横行，另外一个变量则呈现在纵列。因为性别包括男性和女性，所以只有两行，而专业则有很多列，我们只呈现了其中的几列。

双向表

双向表呈现了两个变量之间的关系，一列变量呈现在横行，而另外一列变量则呈现在纵列。表中的每个单元格都是次数（或者是计数数据）。

如果我们探讨专业和性别之间的关系，我们需要所有专业的完整数据。这就要表 10-2 有 12 列。此外，如同我们稍后所理解的内容，进行假设检验的计算要求我们得出所有横行和纵列的总次数。这样，为了简化计算，我们只关注两个专业，生物学和商学，也就是说，不再考虑性别和所有专业的关系，只是考虑学生的性别和生物学以及商学这两个专业的关系。我们就只关注下面这样简单的问题：一个学生的性别是否影响他选择生物学或者是商学？表 10-3 呈现了从表 10-2 中摘录的生物学和商学的数据，并且分别列出了横行和纵列的次数合计。

表 10-3 获得生物学和商学学位的双向表

	生物学	商学	合计
女生	32	110	142
男生	21	87	108
合计	53	197	250

思考题

使用表 10-3 来回答。如下的问题：
①授予商学学位的男生是多少人？②总共授予多少人商学学位？③将男生和女生授予学位的总人数与商学及生物学授予学位的总人数进行比较，这两个总人数是相同的还是不同的？为什么？

相关知识

在美国的所有学院和综合大学里,男生人数超出女生人数的专业有工程技术、计算机技术和农业等,女生人数超出男生人数的专业主要有心理学、艺术、生物学和基础教育等。

例 10-5 一项调查的双向表

表 10-4 呈现了一次选举前的有关控制持枪的调查。使用下面的双向表回答下述的问题。

表 10-4 关于控制持枪调查的双向表

	支持严格 控制持枪 的法律	反对严格 控制持枪 的法律	说不清	合计
民主党人	456	123	43	622
共和党人	332	446	21	799
合计	788	569	64	1 421

(1) 确定表中所呈现的两个变量。

(2) 民主党人支持严格控制持枪法律的百分比是多少?

(3) 所有投票人支持严格控制持枪法律的百分比是多少?

(4) 反对严格控制持枪法律的人中共和党人所占百分比是多少?

解答 注意横行的总人数和纵列的总人数是相等的。

(1) 第一行呈现的是变量“调查的态度”,包括3类分别是:“支持严格控制持枪的法律”、“反对严格控制持枪的法律”、“说不清楚”;第一列呈现的是变量“党派关系”,包括两类分别是“民主党”“共和党”。

(2) 调查了共有622名民主党人,其中456人支持严格控制持枪的法律,支持严格法律的百分比是 $456/622=0.733$,占73.3%。

(3) 总共调查了1421人,其中788人支持严格控制持枪的法律,支持严格法律的百分比是 $788/1421=0.555$,占55.5%。

(4) 共有569人反对严格控制持枪的法律,其中446人是共和党人。因为 $446/569=0.783$,所以严格控制持枪的法律当中78.3%是共和党人。 ■

10.2.3 假设检验的实施

当假设确定以后,并且数据也以双向表的形式呈现出来,现在我们开始进行假设检验。

假设检验的基本理论总是一样的,确定样本数据能否提供充分的证据来拒绝虚无假设。对于双向表假设检验的例子,比较特殊的步骤如下所述:

- 和前面所的假设检验一样,首先我们假设虚无假设为真,这意味着两个变量之间没有关系。在这种情况下,我们预期双向表中的次数(每个单元格中的数字)是随机的,我们的第一步是寻找一种计算预期偶然发生的次数。
- 接下来我们将预期的次数和所观察到的样本次数进行比较,样本次数呈现在双向表中。我们通过计算样本的卡方统计量完成以上步骤。卡方值和在第9章正态分布假设检验中的 z 值或者10.1节中 t 检验中的 t 值起到一样的作用。
- 再次回顾第9章的假设检验,通过比较样本标准分数的计算值和查表所得的临界值,我们可以做出统计决断,拒绝还是不拒绝虚无假设。同样的道理,在10.1节中,通过将样本统计量的 t 值和查表所得的 t 值临界值进行比较,也可以做出统计决断。在这里,我们完成同样的任务,只是不再使用标准分数或者 t 分布的临界值,我们使用卡方统计量的临界值。

下面让我们以表10-3的数据为例,来完成上述的检验过程。

1. 计算随机的理论预期次数

第一步是计算表10-3中随机的理论预期次数,如果两个变量间没有关系,理论上的预期次数是相等的。让我们首先开始计算商学专业男生的理论预期次数。为了完成这个计算,首先要计算样本中所有学生得到商学学士学位的比率:

$$\frac{\text{获得商学学士学位的总人数}}{\text{所有获得学位的总人数}} = \frac{197}{250}$$

如同在第6章所讨论的,我们可以将这个结果解释为相对次数概率。也就是说,如果我们随机选择样本中的一位学生,他获得商学学士学位的概率是 $197/250$ 。使用概率符号,可以描述成:

$$P(\text{商学学士}) = \frac{197}{250}$$

同样的道理，如果我们从样本中随机选择一位学生，这位学生是一位男生的概率是：

$$P(\text{男生}) = \frac{\text{男生的总人数}}{\text{男生和女生的总人数}} = \frac{108}{250}$$

现在我们已经得到全部信息，可以计算商学男生的理论预期次数，回顾 6.5 节中两个独立事件（一个事件的发生并不影响另外一个事件发生的概率） A 和 B 同时发生的概率，我们可以得到：

$$P(A \text{ 和 } B) = P(A) \times P(B)$$

我们可以使用上述概率的乘法定理来计算一个学生既是男生又是商学专业的概率（假定虚无假设是性别和专业的相互独立）：

$$\begin{aligned} P(\text{男生并且是商学专业}) \\ &= P(\text{男生}) \times P(\text{商学专业}) \\ &= (108/250) \times (197/250) \approx 0.3404 \end{aligned}$$

如果性别和专业这两个变量没有关系的话，这个概率就等于理论预期的全部商学专业男生所占的比率。这样，我们将这个概率乘以样本中全部学生人数（250）就是理论预期的全部商学专业男生的人数（或者是次数）：

$$(108/250) \times (197/250) \times 250 \approx 85.104$$

我们将这个值称做商学专业男生人数的理论期望次数（expected frequency）。（注意在 6.3 节中所讨论的期望值和这里所讨论的理论期望次数之间的相似之处。）

定义

◆ 理论期望次数是指在双向表中，如果横行和纵列的两个变量之间没有关系的话，我们理论预期的随机发生的次数。

例 10-6 表 10-3 中的理论期望次数

计算表 10-3 其余的三个单元格的理论期望次数，然后绘制一个包括实际观察次数和理论期望次数组成的双向表。

解答 我们采用上述商学专业男生人数的过程中来计算另外三个单元格的理论期望次数。

我们已经知道 $P(\text{男生})$ 和 $P(\text{商学})$ ，我们还需要知道 $P(\text{女生})$ 和 $P(\text{生物学})$ ：

$$\begin{aligned} P(\text{女生}) &= \frac{\text{女生的总人数}}{\text{男生和女生的总人数}} \\ &= 142/250 = 0.5680 \end{aligned}$$

$$\begin{aligned} P(\text{生物学}) &= \frac{\text{获得生物学学士学位的总人数}}{\text{所有获得学位的总人数}} \\ &= 53/250 = 0.2120 \end{aligned}$$

我们将每个单独的概率组合在一起来计算 3 种剩余单元格的概率：

$$\begin{aligned} P(\text{女生并且是商学专业}) \\ &= P(\text{女生}) \times P(\text{商学专业}) \\ &= \left(\frac{142}{250}\right) \times \left(\frac{197}{250}\right) \approx 0.4476 \\ P(\text{男生并且是生物学专业}) \\ &= P(\text{男生}) \times P(\text{生物学专业}) \\ &= \left(\frac{108}{250}\right) \times \left(\frac{53}{250}\right) \approx 0.0916 \\ P(\text{女生并且是生物学专业}) \\ &= P(\text{女生}) \times P(\text{生物学专业}) \\ &= \left(\frac{142}{250}\right) \times \left(\frac{53}{250}\right) \approx 0.1204 \end{aligned}$$

请注意：正如我们所预想的，所有四个单元格概率之和是 $0.3404 + 0.4476 + 0.0916 + 0.1204 = 1.0000$ 。

现在我们通过将每一个单元格的概率乘于所有学生的总次数（250），就可以计算出理论期望次数：

$$\begin{aligned} \text{商学专业女生的理论期望次数} &= \frac{142}{250} \times \frac{197}{250} \times 250 \\ &\approx 111.896 \\ \text{生物学专业男生的理论期望次数} &= \frac{108}{250} \times \frac{53}{250} \times 250 \\ &\approx 22.896 \\ \text{生物学专业女生的理论期望次数} &= \frac{142}{250} \times \frac{53}{250} \times 250 \\ &\approx 30.104 \end{aligned}$$

表 10-5 是表 10-3 中数据的重复，这是在表中呈现了每一个单元格的理论期望次数（括号中的）。为了检验我们所做工作是否正确，我们将所有的理论期望次数的总和和样本中所有学生人数进行比较：

$$85.104 + 111.896 + 22.896 + 30.104 = 250.000$$

请注意：横行的“合计”和纵列的“合计”是相同的，同时也是实际观察的总次数和理论期望的总次数。它们也总是相同的，这是从另外一方面对你所做的计算进行检查的一种方法。

表 10-5 表 10-3 中实际观察次数和理论期望次数（括号中）

	生物学	商学	总数
女生	32(30.104)	110(111.896)	142(142.000)
男生	21(22.896)	87(85.104)	108(108.000)
总数	53(53.000)	197(197.000)	250(250.000)

2. 检验统计量卡方的计算

请注意，表 10-5 的理论期望次数和实际观察次数是比较接近的。比如，商学专业的

男生理论期望次数是 85.1, 实际观察次数是 87, 这两者是非常接近的。我们可以更进一步猜测这些数据没有给出我们理由来拒绝虚无假设“性别和专业这两个变量没有关系”。尽管如此, 我们还需要找出特定的方法来对实际观察的次数和理论期望次数之间的差异进行量化。

让我们用符号 O 代表实际观察次数, 用符号 E 代表理论期望次数, 用符号 $O - E$ 来代表每个单元格里实际观察次数和理论期望次数之间的差值。我们现在要计算整表的总差异, 但是我们不能简单地将每个单元格的差值 ($O - E$) 加起来, 因为它们的总和是 0。我们可以考虑用每一个单元格中差值的平方 ($O - E$)², 我们可以进一步将 ($O - E$)² 除以它所对应的理论期望次数作为一种相对差异, 这样我们就可以通过 ($O - E$)²/ E 来量化每一个单元格中理论次数和实际次数的差异。将每一个单元格中的 ($O - E$)²/ E 累加起来就是检验统计量卡方 (chi-square statistic), 符号为 χ^2 (χ 是希腊字母)。

检验统计量卡方的计算

第一步, 对于双向表的每一个单元格, 在虚无假设为真的前提下 (两个变量间没有关系), 确定其实际观察次数 O 和理论期望次数 E 。

第二步, 计算每一个单元格的 ($O - E$)²/ E 取值。

第三步, 将第二步所计算的每一个单元格的 ($O - E$)²/ E 值累加起来得到检验统计量卡方:

$$\chi^2 = \text{所有单元格 } (O - E)^2 / E \text{ 值的累加}$$

χ^2 的值越大, 各个单元格的实际观察次数和理论预期次数的平均差异就越大。

为了能够进行系统的计算, 最好还是编制如表 10-6 的表格, 它呈现出了卡方值的完整计算过程, 如表中最后一个单元格所呈现的, 性别和专业两个变量的卡方最后结果是 $\chi^2 = 0.350$ 。

表 10-6 表 10-5 中数据的卡方值计算

结果	O	E	$O - E$	$(O - E)^2$	$(O - E)^2 / E$
女性/商学	110	111.896	-1.896	3.595	0.032
女性/生物学	32	30.104	1.896	3.595	0.119
男性/商学	87	85.104	1.896	3.595	0.042
男性/生物学	21	22.896	-1.896	3.595	0.157
总和	250	250.000	0.000	14.380	$\chi^2 = 0.350$

思考题

为什么 $O - E$ 的累加和总是等于零?

3. 做出统计推断

χ^2 的值提供给我们一个检验虚无假设“两个变量之间没有关系”的方式。如果 χ^2 的值小, 那么实际观测次数和理论期望次数之间的差异就小, 我们就不应该拒绝虚无假设。如果 χ^2 的值大, 那么实际观测次数和理论期望次数之间的差异就大, 我们就有理由拒绝两个变量之间独立的虚无假设。为了量化我们所说的“大”和“小”, 我们将计算出的样本 χ^2 值和临界值进行比较:

- 如果计算出的样本 χ^2 值小于临界值, 那么实际观测次数和理论期望次数之间的差异很小, 并没有充分的证据拒绝虚无假设。
- 如果计算出的样本 χ^2 值大于临界值, 那么样本就提供了充分的证据拒绝虚无假设 (在一定的显著性水平上)。

表 10-7 列出了两个显著性水平的 χ^2 临界值 (0.05 水平和 0.01 水平)。请注意大小不同的表临界值是有差别的, 所以你必须确定所查找的临界值适合你样本的表格。对于前面所讨论的表 10-3 和表 10-5 的性别和专业的数据, 表中由两行两列组成 (不考虑行和列的合计), 这个表的大小是 2×2 , 在表 10-7 中的第一行中, 我们可以看到这种大小的表 0.05 显著性水平的 χ^2 临界值是 3.841, 我们所计算得到的性别和专业的检验统计量 $\chi^2 = 0.350$, 这比临界值 3.841 要小, 所以我们不能拒绝虚无假设。当然, 不能拒绝虚无假设并不是证明性别和专业队是相互独立的。它只是表明我们没有充分证据来拒绝两个变量相互独立的虚无假设。

表 10-7 χ^2 的标准值: 如果 χ^2 大于标准值时拒绝 H_0 假设

表格大小 (行 $\times$ 列)	显著性水平	
	0.05	0.01
2×2	3.841	6.635
2×2 或 3×2	5.991	9.210
3×3	9.488	13.277
2×4 或 4×2	7.815	11.345
2×5 或 5×2	9.488	13.277

方法注解

检验统计量 χ^2 是个离散变量，然而 χ^2 分布却是连续的。只要每一个单元格的理论预期次数不小于5，这种矛盾并不会导致实质性的问题，我们假定本书中的所有例子都是符合这个条件的。

例 10-7 维生素 C 的检验

一项研究（虚拟的）试图确定维生素 C 对预防感冒的影响。在由 220 人组成的样本中，随机抽取其中的 105 人每日服用维生素 C 药丸持续 10 周的时间，其余的 115 人则服用安慰剂持续 10 周时间。在 10 周结束以后，记录罹患感冒的人数，表 10-8 概括了研究的结果。请确定服用维生素 C 和罹患感冒之间是否存在关联。

表 10-8 各个类别观测次数的双向表

	感冒	没有感冒	总数
维生素 C	45	60	105
安慰剂	75	40	115
总数	120	100	220

解答 我们从建立虚无假设和备择假设开始，

H_0 （虚无假设）：服用维生素 C 和罹患感冒之间没有关系，也就是说，维生素 C 在预防感冒上不比安慰剂更有效。

H_a （备择假设）：服用维生素 C 和罹患感冒之间存在一定的关系，也就是说，两组之间罹患感冒的次数是不同的，我们预期服用维生素 C 的一组 and 服用安慰剂的一组效果是不等同的。

和通常情况一样，我们假定虚无假设为真，然后计算表中每一个单元格的理论期望次数。要注意样本容量是 220，和例 10-6 中的过程一样，我们对理论期望次数进行如下的计算。

服用维生素和罹患感冒：

$$220 \times \frac{105}{220} \times \frac{120}{220} \approx 57.273$$

服用维生素和没有罹患感冒：

$$220 \times \frac{105}{220} \times \frac{100}{220} \approx 47.727$$

安慰剂和罹患感冒：

$$220 \times \frac{115}{220} \times \frac{120}{220} \approx 62.727$$

安慰剂和没有罹患感冒：

$$220 \times \frac{115}{220} \times \frac{100}{220} \approx 52.273$$

表 10-9 呈现了双向表，括号中是理论期望次数。

表 10-9 维生素研究中的实际观测次数和理论期望次数

	感冒	没有感冒	总数
维生素 C	45(57.273)	60(47.727)	105(105.000)
安慰剂	75(62.727)	40(52.273)	115(115.000)
总数	120(120.000)	100(100.000)	220(220.000)

现在我们来计算样本数据检验统计量卡方值。表 10-10 呈现了我们如何系统地完成这个工作，你应该能够确定所有计算过程。

表 10-10 维生素研究中的检验统计量卡方的计算过程

结果	O	E	O-E	(O-E) ²	(O-E) ² /E
维生素 C/ 感冒	45	57.273	-12.273	150.627	2.630
维生素 C/ 没有感冒	60	47.727	12.273	150.627	3.156
安慰剂/ 感冒	75	62.727	12.273	150.627	2.401
安慰剂/ 没有感冒	40	52.273	-12.273	150.627	2.882
总数	220	220.000	0.000	602.508	$\chi^2 = 11.069$

为了做出统计决断，是否拒绝虚无假设，我们将样本数据的卡方计算值， $\chi^2 = 11.069$ 和从表 10-7 中所查出的临界值进行比较。我们要查的表大小 2×2 ，因为表 10-8 中的原始数据有两行两列（不考虑“合计”）。我们得到 0.01 显著性水平的 χ^2 临界值是 6.635，因为样本检验统计量 $\chi^2 = 11.069$ 大于这个临界值，我们拒绝虚无假设，并且得出结论说服用维生素 C 和感冒之间存在一定关联。也就是说，基于样本数据，我们有理由相信服用维生素 C 的确要比服用安慰剂有效。

相关知识

关于维生素 C 和感冒的关系，已经进行了 12 项精细的研究，一些研究发现维生素 C 的效应有很高的置信度，但是另一些研究却没有这种发现。由于这些研究冲突的结果，维生素是否可以预防感冒的问题仍然存在争议。

例 10-8 有罪辩护还是无罪辩护

表 10-11 中的双向表呈现了旧金山地区的随机选择的 1 028 件盗窃案件，有罪辩护或者无罪辩护对判决的影响。检验如下的声称：“判决（入狱服刑或者无罪释放）和辩护之间是相互独立的。”

表 10-11 判决和辩护的实际观测数据

	被判决入狱服刑	被判决无罪释放	合计
有罪辩护	392	564	956
无罪辩护	58	14	72
合计	450	578	1 028

解答 这个问题的虚无假设和备择假设分别是
 H_0 (虚无假设): 在盗窃案件中宣判和辩护是相互独立的。

H_1 (备择假设): 在盗窃案件中的宣判受到辩护的影响。

我们对每一类的理论期望次数进行如下的计算, 假设行变量和列变量是相互独立的:

有罪辩护和宣判入狱服刑:

$$1\,028 \times \frac{956}{1\,028} \times \frac{450}{1\,028} = 418.482$$

有罪辩护和宣判无罪释放:

$$1\,028 \times \frac{956}{1\,028} \times \frac{578}{1\,028} = 537.518$$

无罪辩护和宣判入狱服刑:

$$1\,028 \times \frac{72}{1\,028} \times \frac{450}{1\,028} = 31.518$$

无罪辩护和宣判无罪释放:

$$1\,028 \times \frac{72}{1\,028} \times \frac{578}{1\,028} = 40.482$$

表 10-12 判决和辩护的实际观测数据和理论期望次数 (括号中的数字)

	被判决入狱服刑	被判决无罪释放	合计
有罪辩护	392 (418.482)	564 (537.518)	956
无罪辩护	58 (31.518)	14 (40.482)	72
合计	450	578	1 028

和通常情况一样, 表 10-13 系统地呈现了计算 χ^2 值的过程, O 代表实际观测次数, E 代表理论期望次数. 如同表中最后一个单元格所呈现的, 检验统计量 $\chi^2 = 41.556$, 这个值要大于 2×2 表在 0.05 显著性水平的临界值 ($\chi^2 = 3.841$), 也大于 0.01 显著性水平的临界值 ($\chi^2 = 6.635$). 这样, 样本数据支持我们拒绝虚无假设, 接受备择假设. 基于样本数据, 我们有理由相信盗窃案件的判决和辩护是有关联的. 尤其在于, 做有罪辩护的被告, 实际上被判决入狱服刑的人数要比理论期望人数少, 而被判决无罪释放的人要比理论期望人数要多. 那些做无罪辩护的人们, 实际上被判决入狱服刑的人数要比理论期望人数多, 而被判决无罪释放的人要比理论期望的人数要少. 但是要记住这个检验并不能证明辩护和判决之间存在因果关系。

表 10-13 χ^2 值的计算

	O	E	$O - E$	$(O - E)^2$	$(O - E)^2 / E$
有罪辩护/ 入狱服刑	392	418.482	-26.482	701.296	1.676
有罪辩护/ 无罪释放	564	537.518	26.482	701.296	1.305
无罪辩护/ 入狱服刑	58	31.518	26.482	701.296	22.251
无罪辩护/ 无罪释放	14	40.482	-26.482	701.296	17.324
总数	1 028	1 028.000	0.000	2 805.184	$\chi^2 = 42.556$

思考题

如果你是一个盗窃案件犯罪嫌疑人的辩护律师, 前面这个例子的结果对你的当事人的辩护策略有什么样的影响? 请加以解释。

10.2 练习题

统计素养和批判性思维

1. 双向表 什么是双向表? 双向表中应该输入什么样的数据?
2. 变量之间的关系 假定我们已经分析过双向表中, 性别 (男/女) 次数对应于横行, 利手 (左利手/右利手) 对应于纵列. 现在假定我们拒绝性别和利手相互独立. 我们能否得出结论, 一个人的性别对他是左利手还是右利手会产生影响? 为什么?
3. 理论期望次数 假定我们知道由性别 (男/女) 作为横行的变量, 超速驾驶罚单 (是/否) 作为纵列变量. 已知实际观测变量一定是整数, 理论期望次数也一定是整数吗? 请

加以解释。

4. 符号 在双向表这一部分内容中, 描述下列符号的含义: O 、 E 、 χ^2 。

判断正误: 判断练习题 5~8 中的陈述是有意义的 (表述正确) 还是没有意义的 (表述错误), 并加以解释. 并非所有的陈述都有明确的答案, 所以你的解释比判断更重要。

5. 调查 在一次快速调查中, 被试被询问他们是否知道美国宣布独立是在哪一年. 对于这个问题的结果可以概括成双向表。
6. χ^2 检验统计量 一个双向表可以用来计算检验统计量, 所得到的 $\chi^2 = -2.500$ 。
7. χ^2 检验统计量 在一个双向表中, 所有实际

观测次数非常接近于理论期望次数，因此检验统计量 χ^2 非常小，我们不能拒绝“横行和纵列的变量是相互独立的”虚无假设。

8. 虚无假设 假定双向表被设计成性别（男/女）作为横行的变量，对一个调查问题的回答（是/否）作为纵列变量。虚无假设可以陈述为性别和对问题的反应是相互独立的。

概念与应用

调查结果：在练习题 9 ~ 12 中，假定一个调查问题的备选答案是“是/否”，然后随机抽取了包括男性和女性的一个样本。结果被概括成如下形式的双向表。使用下面已知的检验统计量 χ^2 和显著性水平来检验性别和对问题的回答之间是否相互独立。

- 9. 检验统计量： $\chi^2 = 0.051$ ；显著性水平：0.01。
- 10. 检验统计量： $\chi^2 = 10.785$ ；显著性水平：0.01。
- 11. 检验统计量： $\chi^2 = 2.924$ ；显著性水平：0.05。
- 12. 检验统计量： $\chi^2 = 15.238$ ；显著性水平：0.05。

完成假设检验：在练习题 13 ~ 20 中，进行如下步骤完成假设检验。

- (1) 陈述虚无假设和备择假设。
- (2) 假定两个变量之间是相互独立的，计算表中每一个单元格的理论期望次数。
- (3) 计算检验统计量 χ^2 的值。
- (4) 使用给出的显著性水平查找 χ^2 的临界值。
- (5) 使用给出的显著性水平，完成对声称“两个变量相互独立的”的检验。

13. 大学生的人口统计特征 下表呈现了 148 名全日制和业余的大学生以及他们性别的分布情况（与联邦数据相一致）。采用 0.05 显著性水平检验如下的声称“大学生的性别和学生身份（全日制和业余）是相互独立的”。

	全日制	业余
女大学生	47	37
男大学生	38	26

14. 参加投票的投票者 右上表呈现了在上次的总统选举中参加投票的一个公民样本的人数，按照性别和是否参加投票分别列出对应

的人数（与联邦数据相一致）。采用 0.05 显著性水平检验如下的声称“性别和是否投票没有关联”。

	参加投票	没有参加投票
女性	140	120
男性	130	110

15. 电子邮件和个人隐私 对工人和上级管理阶层的老板进行调查，询问他们“如果监控雇员的电子邮件是否是**非常不道德的”。调查的结果概括成如下表（基于盖洛普调查公司的数据）。采用 0.05 显著性水平检验如下的声称“人们对问题的回答与他们是工人还是老板无关”。

	是	否
工人	192	244
老板	40	81

16. 自行车头盔的有效性 进行的一项研究是关于在自行车碰撞中受伤害 531 个人的；这个随机选择的样本结果概括在下表中（based on data from “A Case-Control Study of the Effectiveness of Bicycle Safety Helmets in Preventing Facial Injury,” by Thompson, Thompson, Rivara, and Wolf, *American Journal of Public Health*, Vol. 80, No. 12），采用 0.01 显著性水平检验如下的声称“带不带安全头盔与脸部是否受到伤害无关”。

	带安全头盔	不带安全头盔
脸部受到伤害	30	182
脸部没有受到伤害	83	236

17. 关节炎的治疗 进行的一项临床实验中，作为实验组，给予 98 个关节炎罹患患者一项新的物理治疗，56 位症状有所改善。另外 92 位实验参与者给予安慰剂治疗，其中 49 位症状有所改善。针对这些数据建立一个双向表，并采用 0.05 显著性水平如下的声称“关节炎罹患者的症状改善与采用物理治疗还是给予安慰剂无关”。

18. 饮酒和怀孕 随机抽取了由 1252 名 25 岁以下的孕妇组成的样本，其中包括 13 名在怀孕期间饮酒的孕妇。另外随机抽取了 2029

名 25 岁及以上的孕妇组成的样本，其中包括 37 名在怀孕期间饮酒的孕妇（数据来源于美国健康统计中心的研究结果），并采用 0.05 显著性水平如下的声称“孕妇的年龄（25 岁以下和 25 岁及以上）和怀孕期间是否饮酒没有关系”。

19. 犯罪和陌生人 下表列出了由不同的犯罪受害者组成的随机样本的调查结果（数据来源于美国司法部）。采用 0.01 显著性水平如下的声称“不同类型的犯罪和罪犯与是否是陌生人无关”。

	杀人	抢劫	伤害
罪犯为陌生人	12	379	727
罪犯为熟人或亲属	39	106	642

20. 中国的烟民情况 下表概括了生活在上海闵行地区 15 岁及以上的男性的一个调查结果（based on data from “Cigarette Smoking in China” by Gong, Koplan, Feng, et al., *Journal of the American Medical Association*, Vol. 274, No. 15）。这些男性按照他们现在的受教育水平以及是否吸烟进行分类，采用 0.05 显著性水平检验如下的声称“吸烟与所受教育水平没有关系”。

	小学	中学	大学
吸烟	606	1 234	100
不吸烟	205	505	137

网络和远程项目

请在网站 www.aw.com/bbt 上打开第 10 章的有用链接“links for internet projects”。

21. 编制双向表 选择可能有关系并且值得研究的两个变量。其中一个变量应该至少分为两类，例如，两类不同年龄、不同的种族或者不同的地域，另外一个变量社会、经济或者健康因素，也应该至少分为两类，例如，两种不同的收入分类，是否饮酒，接受的教育水平的高低。所需要的总体或者样本数据填入由两个变

量组成的一个双向表。讨论这两个变量看起来是否相关。一个好的开始方法是搜索美国人口普查局的“美国统计概要”这样的网页。

22. 双向表的分析 现在看起来可能会有关系并且值得研究的两个变量。其中一个变量应该至少分为两类，例如，两类不同年龄、不同的种族或者不同的地域。另外一个变量社会、经济或者健康因素应该至少也分为两类，例如，两种不同的收入分类，是否饮酒，接受的教育水平的高低。找到次数数据填入两个变量组成的双向表中，进行假设检验这两个变量之间是否有关联。一个非常好的资源是美国人口普查局的“美国统计概要”这样的网页。

新闻中的统计学

23. 新闻中的双向表 新闻文章中的一个双向表并不常见（但并非不可能出现），但一则新闻故事却经常会提供可以用双向表表达的信息。寻找一篇讨论可以用双向表表达的两个变量间关系的文章，将表绘制出来。
24. 新闻中的假设检验 新闻报道经常描述统计研究的结果，其结论来自于包含双向表的假设检验。尽管如此，报道中很少给出实际的表或者描述假设检验的细节。寻找一篇近期的新闻报道，你认为其中的结论可能基于双向表的假设检验。假如你是正确的，用语言描述假设检验是如何进行的。也就是描述零假设和备择假设，以及研究者可能如何进行检验的过程。
25. 你自己的假设检验 设想一个你很想知道是怎么回事并且能用双向表进行假设检验的例子。不必实际搜集数据进行运算，仅仅描述你是怎样设计你的研究的。也就是说，你打算如何收集数据，解释怎样将这些数据编制成一个双向表，并且陈述虚无假设、备择假设，描述怎样进行这个假设检验并且最后得出统计结论。



10.3 方差分析（一元方差分析）

迄今为止，我们已经学习过虚无假设采用

三种不同形式进行声称的假设检验：声称一个总体平均数等于某个值，这种情况我们在 9.2 节中采用正态分布进行检验，在 10.1 节采用 *t* 分布进行检验；声称一个总体比率等于某个值，这种情况我们在 9.3 节中采用正态分布进行检

验；声称两个变量是相互独立的，这种情况我们在 10.2 节中进行了讨论。统计学家进一步发展了统计技术来考虑其他类型的虚无假设，使它有可能应用于更广泛的领域中。为了让你知道统计学可能适用的范围，也是为了鼓励更多地学习统计学，在本书的最后一节，我们简要地介绍一下更多的假设检验类型。

10.3.1 变异性的假设检验

在三种不同的书中，我们各自随机抽取了 12 页作为一个样本：汤姆·克兰西的《熊与龙》，J. K. 罗琳的《哈利波特和魔法石》，列夫·托尔斯泰的《战争与和平》，对这些样本中的每一页都采用弗莱什阅读难度指数进行评分，结果呈现在表 10-4 中。弗莱什阅读难度指数的分数意思是指，文本的分数越高，就越容易阅读；分数越低，著作就越难读懂。我们在这一节的目标是使用每一本书的 12 页文本样本来推断每一本书所有页的总体易读性。

我们可以通过集中量数、变异性、分布和异常值等，非正式地探讨样本数据。表 10-15 呈现了我们这个例子的重要样本统计量，如果你将表 10-14 的原始数据和表 10-15 的样本平均数进行比较，你会注意到比起大多数其他分数来，尽管有极少的分数远离平均数（比如汤姆·克兰西的《熊与龙》最低的 43.9 和 J. K. 罗琳《哈利波特和魔法石》的 70.9），但是并没有分数看起来非常极端，因此我们不能认为这些值是异常值。此外，详细地研究说明样本来自的总体比较接近于正态分布。

表 10-14 弗莱什阅读难度指数		
汤姆·克兰西的 《熊与龙》	J. K. 罗琳的 《哈利波特和魔法石》	列夫·托尔斯泰 的《战争与和平》
58.2	85.3	69.4
73.4	84.3	64.2
73.1	79.5	71.4
64.4	82.5	71.6
72.7	80.2	68.5
89.2	84.6	51.9
43.9	79.2	72.2
76.3	70.9	74.4
76.4	78.6	52.8
78.9	86.2	58.4
69.4	74.0	65.4
72.9	83.7	73.6

表 10-15 弗莱什阅读难度指数			
	汤姆·克兰西 《熊与龙》	J. K. 罗琳的 《哈利波特和魔法石》	列夫·托尔斯泰 的《战争与和平》
样本容量 n	12	12	12
样本平均数 $\bar{x}$	70.73	80.75	66.15
样本标准差 s	11.33	4.68	7.86

其实在研究这些数据之前，我们也会预计到 J. K. 罗琳的《哈利波特和魔法石》是这三本书中最容易读懂的，因为它就是为儿童写的作品。我们会预计列夫·托尔斯泰的《战争与和平》是最难读懂的，因为它是翻译的俄罗斯经典著作。现在我们来看下表 10-15 中的平均阅读难度指数。再次回顾弗莱什阅读难度指数分数越高，阅读水平越容易。数据看起来正是我们的设想：J. K. 罗琳的《哈利波特和魔法石》分数最高，列夫·托尔斯泰的《战争与和平》的分数最低。然而，这三个样本平均数的差别并不是特别大，值域范围从 J. K. 罗琳的 66.15 到列夫·托尔斯泰的 80.75，而且每一个样本容量相对较小 $n=12$ 。现在再回到这一节中的关键统计问题：这些数据能够给我们提供充分的证据，从而得出结论说汤姆·克兰西的《熊与龙》、J. K. 罗琳《哈利波特和魔法石》、列夫·托尔斯泰的《战争与和平》这三本著作的弗莱什阅读难度指数是有显著差异的吗？

为了回答这个问题，我们再来看一下 9.1 节中的假设检验的基本原理。因为我们想知道这三本书的弗莱什阅读难度指数是否真的有差异，让我们从它们没有差异这个假设开始。换句话说，我们的虚无假设就是，这三本书的弗莱什阅读难度指数是相等的。备择假设就是，这三本书的弗莱什阅读难度指数是不同的。假设检验应该告诉我们拒绝还是不拒绝虚无假设。拒绝虚无假设可以让我们得出这三本书的弗莱什阅读难度指数有显著差异的结论，正如我们所预期的。不拒绝虚无假设将告诉我们这些数据不能提供充分的证据，让我们得出如下结论“这三本书的弗莱什阅读难度指数有显著差异”。

请注意这个例子中的每一个总体平均数，所代表的是如果我们对每一本书的全部页数进行测量后计算出的弗莱什阅读难度指数平均数。使用更规范的符号，我们可以将虚无假设写作：

$$H_0: \mu_{\text{克兰西}} = \mu_{\text{罗琳}} = \mu_{\text{托尔斯泰}}$$

正如你所看到的, 我们需要一个假设检验来告诉我们对三个不同总体的平均数是否相同。这种方法我们叫做方差分析 (analysis of variance), 通常缩写成 ANOVA。这个名字来自我们已经知道的, 代表一组样本数据变异性的正式统计量; 前面在 4.3 节中简要讨论过的, 方差被定义成样本标准差的平方, 或者是 s^2 。(例如, 一个身高样本标准差是 $s = 3.0\text{cm}$, 它的方差就是 $s^2 = 9.0\text{cm}^2$ 。)

定义

- ◆ 方差分析是一种假设检验方法, 通过分析样本方差来检验三个或更多的总体平均数是否相等。

方法注解

方差分析的方法也可以用于两个平均数, 但是它相当于对两个样本方差进行 t 检验, 这是另一种可以用于两个独立样本的 t 检验, 并且效率更高, 但是本书并不包括这些形式的 t 检验。

更为特别的是, 这种方法经常用来分析像表 10-14 这样的数据, 叫做一元方差分析, 因为样本数据仅仅按照一种属性 (或者说是一种因素) 被分成不同的类别。在这个例子中, 属性就是作者 (汤姆·克兰西、J.K. 罗琳、列夫·托尔斯泰)。此外, 也可以进行二元方差分析, 但这是按照两个属性 (或者两个因素) 各自分成不同的类别然后比较它们之间的总体平均数。例如, 我们将人们的身高使用如下的两个属性进行分类: ①性别 (男性/女性), ②左右利手。在本书中我们不再讨论二元方差分析。

10.3.2 建立假设检验

方差分析是建立在下面的基本理论上的: 我们假定总体平均数都有相同的方差, 然后比较样本之间的方差和样本之内的方差。此外, 医院方差分析的检验统计量 (通常叫做 F 值) 是两个方差之间的比值:

$$\text{检验统计量 } F(\text{一元方差分析}) = \frac{\text{样本之间的方差}}{\text{样本之内的方差}}$$

这些样本统计量实际计算过程是非常繁琐的, 因此现在一般用统计软件来完成这个计算

(参考本章的技术应用)。尽管如此, 我们对这个统计量做如下的解释, 所使用的例子就是前面提到的三本书的易读性。

- 样本之间的方差是指三个样本平均数 (见表 10-15) 之间彼此差异的指标。
- 样本之内的方差是指每一个样本中各自的页面弗莱什阅读难度指数彼此之间的差异。
- 如果这三个平均数是真的相等, 那么就如同虚无假设所声称的那样, 我们可以预期单独的任何一个样本的样本平均数会落入另外任何一个样本的变异范围之内。检验统计量 ($F = \text{样本间方差} / \text{样本内方差}$) 将告诉我们下面的情况:

A. 计算出的检验统计量 F 值大, 这告诉我们样本平均数与样本内的个体数据有显著差异, 如果总体平均数真的是相等的话 (如同虚无假设所声称的), 这是不大可能的。也就是说, 大的检验统计量 F 值提供给我们证据来拒绝“总体平均数是相等的”这个虚无假设。

B. 计算出的检验统计量 F 值小, 这告诉我们样本平均数与样本内的个体数据没有显著差异, 这说明样本平均数之间的差异可能只是随机抽样的结果。这样, 小的检验统计量 F 值不能提供给我们证据来拒绝“总体平均数是相等的”这个虚无假设。

请注意, 方差分析中的检验统计量 F 值, 和我们前面所学习的假设检验中的标准分数 z 以及 t 检验中的 t 值所起的作用是相同的。这样, 如同我们在前面的例子中所做的, 我们也可以通过计算 P 值来量化地解释这个检验统计量 F , 而 P 值是指作为一种临界值, 我们得到这样一个样本结果的概率。假定虚无假设为真 (所有总体平均数全部是相等的)。一个较小的 P 值显示出我们通过随机抽样得到这样的几个样本是不大可能的。而较大的 P 值是指在总体平均数相等的前提下, 得到这样的样本结果还是容易的。

和检验统计量相似, P 值计算通常要借助软件。当我们计算出了 P 值, 像我们在其他例子中一样, 我们采用它作出最后的统计决断: 如果 P 值很小 (比如 0.05 或者更小), 就拒绝总体平均数相等的虚无假设; 如果 P 值较大 (比如大于 0.05), 就不能拒绝总体平均数相等的虚无假设。下面概括了一元方差分析的要求

和这一节中所提到的软件使用过程。

检验一元方差分析 $H_0: \mu_1 = \mu_2 = \mu_3 = \dots$

第一步，将样本数据输入一个统计软件包，然后使用软件确定检验统计量（ $F = \text{样本间方差} / \text{样本内方差}$ ）和检验统计量的 P 值。

第二步，基于所计算的检验统计量的 P 值，做出统计决断是拒绝还是不拒绝虚无假设：

- 如果 P 值小于或等于显著性水平，拒绝总体平均数相等的虚无假设，可以得出结论，至少有一个总体平均数不同于其他总体平均数。
- 如果 P 值大于显著性水平，不能拒绝总体平均数相等的虚无假设。

只有符合下面的要求时这个方法才是有效的：总体平均数的分布接近于正态分布，并且方差齐性（相等），来自每一个总体中的样本是相互独立的随机样本。

例 10-9 汤姆·克兰西、J. K. 罗琳、列夫·托尔斯泰三位作家的易读性

根据表 10-14 给出的三位作家的易读性分数，在 0.05 的显著性水平上检验如下的虚无假设：“三个样本各自所代表的总体平均数是相等的”。

解答 我们现在开始检查一下是否符合一元方差分析的要求。如同前面内容所提示的，对数据非常仔细的检查说明样本所来自总体接近于正态分布。样本标准差异不是很明显，所以有理由假定三个总体平

均数的方差是相等的。样本是随机抽取的并且相互之间是独立的。符合进行一元方差分析的要求。

现在我们检验虚无假设“三个样本各自所代表的总体平均数是相等的”（ $H_0: \mu_1 = \mu_2 = \mu_3$ ）。技术应用部分描述了应用不同的统计软件包怎样计算统计量 F 值和对应的 P 值。下表呈现了采用 Excel 进行处理的结果，其他的统计软件包也会给出相似的表格。

请注意表格中包括 F 值和 P 值的两列。这是我们两个比较关注的项目，我们解释如下：

- F 值是一元方差分析的检验统计量（ $F = \text{样本间方差} / \text{样本内方差}$ ）。注意这个值大于 1 很多的话，说明样本平均数和我们所预期的总体平均数相等是有差异的。
- P 值告诉我们如果虚无假设为真的前提下，作为一个极端值获得这样一个样本的概率。请注意本例中的 P 值非常小，远小于 0.05，这必然在 0.05 显著性水平上拒绝虚无假设（也远小于 0.01，必定也在 0.01 显著性水平上拒绝虚无假设）。

我们得出结论说有充分的证据拒绝虚无假设。这意味着样本数据支持如下的声称：“三个总体平均数是不相同的”。基于对汤姆·克兰西的《熊与龙》、J. K. 罗琳的《哈利波特和魔法石》、列夫·托尔斯泰的《战争与和平》这三本书的随机选择的页码，我们可以得出结论说这三本书的易读性是不相同的。值得注意的是，我们不能得出我们所预计的关于这三本书的易读性顺序的结论，J. K. 罗琳的《哈利波特和魔法石》易读性最高，列夫·托尔斯泰的《战争与和平》易读性最低，因为假设检验知识显示这三本书的易读性是不同的。不过，我们的预期看起来是合理的，因为表 10-15 中的样本平均数的结果和我们预期的顺序是一样的。

	平方和	自由度	均方	F	P 值	F 临界值
组间	1 338. 002 222	2	669. 001 111	9. 469 487 401	0. 000 562 13	3. 284 924 333
组内	2 331. 386 667	33	70. 648 081			
合计	3 669. 388 889	35				

10.3 练习题

统计素养和批判性思维

1. 方差分析 ANOVA 代表什么方法？这个方法的目标是什么？
2. 方差 样本的方差是什么？
3. 专业的比较 纽波特大学的一位学生管理员在他的大学里随机选择英语、数学和科学三个专业的学生进行抽象推理测验。然后他使用一元方差分析的方法得出结论：这些平均

- 分数是不相等的。他能否得出这样的结论：在美国，英语、数学和科学这三个专业的学生在抽象推理测验上的平均分数是不相同的。为什么？
4. 一元方差分析 为什么这一部分所学习的方法叫做一元方差分析？
- 判断正误：**判断练习题 5~8 中的陈述是有意义的（表述正确）还是没有意义的（表述

错误), 并加以解释。并非所有的陈述都有明确的答案, 所以你的解释比判断更重要。

5. *P* 值 在一个测试来自美国、挪威和中国的人们的平均胆固醇水平是否相等的检验中, 样本数据 *P* 值的结果是 0.99, 所以我们拒绝三个总体平均数相等的假设。

6. *P* 值 对动画片、喜剧片和恐怖片这三种不同类型电影的观众平均年龄是否相等的检验中, *P* 值的结果是 0.009, 因此我们拒绝三个总体平均数相等的假设。

7. 一元方差分析 在一个有关儿童智商测验分数的研究中, 随机选择一些有三个孩子的家庭, 第一个样本包括所有的排行老大的智商分数, 第二个样本包括所有的排行老二的智商分数, 第三个样本包括所有的排行老三的智商分数。一元方差分析的方法可以用来检验三个总体平均数是否相等。

8. 一元方差分析 一家自动减震器的制造商检验三种不同的机器, 采用一元方差分析比较这三种不同机器变异性的结果。

概念和应用

9. 作者的易读性 这一节所使用的汤姆·克兰西、J. K. 罗琳、列夫·托尔斯泰三位作家的弗莱什阅读难度指数的例子。当使用 lesch-kincaid 难度系数来代替弗莱什阅读难度指数时, 采用 STATDISK 进行一元方差分析的结果显示在图 10-2 中。假定我们在

0.05 显著性水平检验如下的虚无假设“三位作家的 lesch-kincaid 难度系数平均数是相同的”。

- (1) 什么是虚无假设?
- (2) 什么是备择假设?
- (3) 请确定 *P* 值。
- (4) 依据下述的结果, 你可以得到三个总体平均数相等的结论吗?

	自由度	平方和	均方	检验统计量	<i>F</i> 临界值	<i>P</i> 值
组间	2	68.187 222	34.093 611	8.978 506	3.284 914	0.000 77
组内	33	125.309 167	3.797 247			
合计	35	193.496 389	5.528 468			

图 10-2

10. 不同实验室的织物耐燃性检测 应用于儿童服装的纺织物经常要进行垂线半抑制检测。在控制的条件下点燃一小片纺织物。当燃烧停止的时候, 烧焦部分的长度被测量并记录。同样的纺织物样本在不同的实验室进行检测。下面是采用 EXCEL 进行一元方差分析的结果。

- (1) 什么是虚无假设?
- (2) 什么是备择假设?
- (3) 请确定 *P* 值。
- (4) 是否有充分的证据支持这样的声称“不同实验室所检测的平均数并不是全部相同的”? 检验采用 0.05 显著性水平。

	平方和	自由度	均方	<i>F</i>	<i>P</i> 值	<i>F</i> 临界值
组间	2.087 194 264	4	0.521 798 566	2.949 333 035	0.030 665 893	2.588 834 036
组内	7.607 597 403	43	0.176 920 287			
合计	9.694 791 667	47				

11. 马拉松比赛的完成时间 从完成纽约马拉松比赛的男运动员中抽取一个随机样本, 应根据年龄将人们分成 21~29 岁、30~39 岁和 40 岁以上三类, 记录样本中的运动员比赛完成的时间(以秒计)。下面是采用 EXCEL 进行一元方差分析的结果。

- (1) 什么是虚无假设?
- (2) 什么是备择假设?
- (3) 请确定 *P* 值。
- (4) 是否有充分的证据支持这样的声称“不同年龄组的运动员平均比赛完成时间是不相同的”?

	平方和	自由度	均方	<i>F</i>	<i>P</i> 值	<i>F</i> 临界值
组间	3 532 063. 284	2	1 766 031. 642	0.188 679 406	0.828 324 293	3.080 387 501
组内	1 010 875 649	108	9 359 959. 71			
合计	1 014 407 712	110				

12. 不同年龄组的心脏收缩压 随机抽取的 40 位女性样本按照 20 岁以下、20~40 岁和 40 岁以上分成 3 组。下面是采用 SPSS 进行一元方差分析的结果。
- (1) 什么是虚无假设？
- (2) 什么是备择假设？
- (3) 请确定 P 值。
- (4) 是否有充分的证据支持这样的声称“不同年龄组女性心脏收缩压的平均水平是不相同的”？

	平方和	自由度	均方	F	Sig.
组间	937.930	2	468.965	1.655	0.205
组内	10484.470	37	283.364		
合计	11422.400	39			

13. 汽车事故中的头部伤害 美国交通安全委员会设计的汽车碰撞试验中，新购买的汽车以每小时 35 公里的速度与固定的障碍物进行碰撞。这些微型汽车分别是福特护卫者、本田思域、现代雅绅特、尼桑日产、土星 SL4；小型汽车分别是雪佛兰骑士、道奇霓虹、马自达 626dx 陵、旁蒂克太阳火、斯巴鲁力狮；中型汽车分别是雪佛兰大黄蜂、道奇无畏、福特野马、本田雅阁、沃尔沃 S70；大型汽车包括奥迪 A8、凯迪拉克帝威、福特皇冠维多利亚、奥兹莫比尔曙光、庞蒂克·博纳维尔。在驾驶员座位的仿真人的头部伤害数据列于下表。使用 0.05 显著性水平检验如下的虚无假设：“不同重量分类的汽车头部伤害平均数相等。”样本数据说明大型汽车更安全吗？
- 微型汽车：681 428 917 898 420
- 小型汽车：643 655 442 514 525
- 中型汽车：469 727 525 454 259
- 大型汽车：384 656 602 687 360
14. 汽车碰撞中的胸部减速 下表列出了练习题 13 中所描述的试验中的胸部减速的数据，请采用 0.05 的显著性水平，检验“不同重量类别的汽车有着相同的平均数水平”这个虚无假设。这个数据说明大型汽车更安全些吗？
- 微型汽车：55 47 59 49 42
- 小型汽车：57 57 46 54 51
- 中型汽车：45 53 49 51 46
- 大型汽车：44 45 39 58 44

15. 考古学：来自不同时期的头盖骨宽度 下表的数值是所测量到的，来自不同时期的埃及男性头盖骨的最大宽度（以毫米计），(based on data from *Ancient Races of the Thebaid*, by Thomson and Randall-Maciver)。不同时期头形大小的变化说明了随着移民的进入，发生了混血交配现象。请使用 0.05 显著性水平检验如下的声称“不同时期的人们有着相同的头盖骨宽度的平均数”。

公元前 4000 年	公元前 1850 年	公元 150 年
131	129	128
138	134	138
125	136	136
129	137	139
132	137	141
135	129	142
132	136	137
134	138	145
138	134	137

16. 不同天气状况下的太阳能 本书作者的一位学生家里有一个电子太阳能系统。在每天的不同时间，他收集了连接这个系统的电压表读数。下表呈现了所收集的结果。请采用 0.05 显著性水平检验如下的声称“在三种不同类型的天气下，电压表读数的平均数是相同的”。我们预计晴朗的天气比多云或者雨天里一个太阳能系统能提供给我们更多的电能。有足够的证据支持总体平均数的不同声称吗？

晴天	多云天气	雨天
13.5	12.7	12.1
13.0	12.5	12.2
13.2	12.6	12.3
13.9	12.7	11.9
13.8	13.0	11.6
14.0	13.0	12.2

网络和远程项目

- 请在网站 www.aw.com/bbt 上打开第 10 章的有用链接“links for internet projects”。
17. 著名人物 《世界大事年鉴》其中包括一部分叫做“著名人物”的内容，并且按照建筑家、艺术家、商业领袖、漫画家和其他类别进行分类。设计并实施一项研究，从中选择几类然后在每一类中选择一些人物作为样本，然后对不同类别人们的平均寿命进行比较。是否有一类人物看起来要比其他不同

类别的人物平均寿命有差异。

- 一元方差分析 寻找一篇期刊上的文章，其中使用到了一元方差分析。请确定所用的检验并且描述所得到的结论。对检验结果是否拒绝了平均数相等的假设？它的 P 值是多少？一元方差分析在其中起到了什么作用？

新闻中的统计学

- 运动 寻找近期的一篇新闻报道文章，内容是讨论不同运动队（比如一个棒球队）的运动员的薪水。使用一元方差分析检验平均数相等的虚无假设。概括你的计算结果，写一篇简要报告。

第10章复习题

- 下表列出了由 AT&T 和美国汽车联合会赞助的一项研究包括的一种样本数据。

	过去一年中 发生交通事故	过去一年中 没发生交通事故
开车使用手机	23	282
开车不使用手机	46	407

- 比较发生交通事故中使用手机者的百分比和没有使用手机者所占的百分比。基于这个结果，使用手机看起来很危险吗？
 - 确定虚无假设和备择假设，检验如下声称“使用手机和发生交通是相互独立的”。
 - 假定使用手机和发生交通事故是相互独立的，计算表中每一个单元格的理论期望值。
 - 计算检验如下声称“使用手机和发生交通是相互独立的”的检验统计量卡方值。
 - 根据第 4 小题计算的结果以及表的大小，查表 10-7 确定对应的 P 值是多少。
 - 基于前面的结果，我们能根据“两个变量（手机使用和发生交通事故）相互独立”的假设检验得出什么样的结论。
 - 第 6 小题的结论和我们现在关于开车和使用手机的观念一致吗？
- 一个可乐罐的轴向负荷是指它所能支持的最大力量，一定大于 165 磅，因为这个力量是当它的顶盖被按压到位时所承受的最大压力。下面列出了随机选取的 12 盎司铝可乐罐的轴向负荷。

270 273 258 204 254 228 282
采用 0.05 的显著性水平检验这些样本来自一个总体平均数大于 165 磅的总体。

- 使用复习题 2 中的样本数据，建立一个所有铝可乐罐平均轴向负荷估计值的 95% 的置信区间。
- 下表列出的是随机选择的 3 个不同年龄组被试的体温。再下面的表格则呈现了采用 STATDISK 检验统计图 10-3 的结果值。假定我们采用 0.05 的显著性水平检验如下的声称“3 个年龄组有着相同的身体温度”。

	自由度	平方和	均方	检验统计量 F	F 临界值	P 值
组间	2	1.729333	0.864667	1.87971	3.88529	0.194915
组内	12	5.52	0.46			
合计	14	7.249333	0.51781			

图 10-3

18 ~ 20 岁	21 ~ 29 岁	30 岁及以上
98.0	99.6	98.6
98.4	98.2	98.6
97.7	99.0	97.0
98.5	98.2	97.5
97.1	97.9	97.3

- 虚无假设是什么？
- 备择假设是什么？
- 请确定 P 值。
- 是否有充分的证据拒绝“三个年龄组有着相同的身体温度”这个虚无假设？

第10章小测验

- 从一个正态分布总体中随机抽取样本包括 15 个值，总体标准差未知，下面哪一种分布是最适合这种情况下关于总体平均数的假设检验？

- 正态分布
- 卡方分布
- t 分布
- 均匀分布

2. 从一个正态分布总体中随机抽取 45 个值作为样本，总体标准差未知，下面哪一种分布是最适合这种情况下关于总体平均数的假设检验？
 - (1) 正态分布
 - (2) 卡方分布
 - (3) t 分布
 - (4) 均匀分布
3. 从一个正态分布总体中随机抽取 45 个值作为样本，总体标准差已知，下面哪一种分布是最适合这种情况下关于总体平均数的假设检验？
 - (1) 正态分布
 - (2) 卡方分布
 - (3) t 分布
 - (4) 均匀分布
4. “SAT 测验的平均分数大于 1 500” 这个声称的虚无假设是什么？
5. “SAT 测验的平均分数大于 1 500” 这个声称的备择假设是什么？
6. 说明下面的陈述是对还是错：“一个 t 检验经常用来检验的声称是，在一个双向表中，同行的变量和纵列的变量以某种方式存在关联。”
7. 假定你想检验如下声称：“专业分别是科学、文学和商学的学生平均智商分数是相同的。”采用哪一种检验方法会适合？
8. 如果对测验题 7 进行假设检验的 P 值结果是 0.350 0，关于虚无假设将做出什么样的统计结论？
9. 调查的结果被编制成双向表，包括两个横行代表性别（男/女），两个纵列代表对问题的回答（是/否）。确定性别和对问题的回答是否存在某种关联，虚无假设是什么？
10. 如果对测验题 9 进行假设检验的 P 值结果是 0.001，关于虚无假设将作出什么样的统计结论？

技术应用

使用 t 分布确定置信区间

SPSS: 进入 SPSS 的数据编辑窗口，点击 [Analyze]，然后选择 [Descriptive Statistics]，然后选择菜单项 [Explore]。在对话框的右侧点击要处理的变量，然后点击中间的按钮，将变量选入 [Dependent List] 框中。在同一个对话框中，点击 [Statistics] 按钮。请确定“Descriptives”核对无误，并且置信水平是预期的。点击 [Continue]，然后点击 [OK]。置信区间的“Lower Bound”（下限）和 [Upper Bound]（上限）就会呈现在标注为 [Descriptives] 的方框中。

Excel: 使用数据工作台 Data Desk/XL，将它安装于 Excel。Data Desk/XL 是一个统计插件，它完善了 Excel 的标准统计功能。它可以有效地用于分析实际数据。在子目录 DDXL 下找到文件 ReadMe.txt，打开这个文件可找到安装指南。

首先进入样本数据列 A。如果使用 Excel 2003，点击 [DDXL]。如果你使用 Excel 2007，请点击 [Add-Ins]，然后再点击 [DDXL]。选择 [Confidence Intervals]（“置信区间”）。在功能键中，如果 σ 未知，选择 [1 Var t Interval]。（如果 σ 已知，选择 [1 Var z Inter-

val]。）然后点击铅笔图标，比如 A 数据列有 12 个数值，就输入数据列 A1: A12。然后点击 [OK]。在对话框中，选择置信水平（如果你使用 [1 Var z Interval]，就要输入已知总体标准差 σ 的数值），然后点击 [Compute Interval]，这样置信区间就会呈现出来。

STATDISK: 选择 [Analysis] 然后再选择 [Confidence Intervals]。然后选择 [Mean-One Sample]。在出现的对话框中，首先输入一位小数作为显著性水平。输入 0.95 或者 95% 作为置信水平。接下来输入其他所需要的项目。然后点击 [Evaluate]，置信区间就会呈现出来。

使用 t 分布进行假设检验

SPSS: 进入 SPSS 的数据编辑窗口，点击 [Analyze]，然后选择菜单项 [Compare Means]，然后选择 [One-Sample T Test]，选择对话框左侧的变量，然后点击中间的按钮，将选中的变量选入“Test Variable”框中。将假定的总体平均数的值输入“Test Value”框中（这与虚无假设使用的值是相同的），然后点击 [OK]。呈现的视窗中包括的主要组成部分是标题为“One-Sample Test”的一个表格。表格中包含一行最上面标明“ t ”的检验

统计量 t 值,还包括一行注明“Sig. (2-tailed)”的 P 值。要注意的是,列出的 P 值适用于双侧检验,但可以调整成单侧检验。对于单侧检验,一般只需要将这个 P 值除以2就可以了,但是在备择假设和样本平均数不一致时,不可以这样应用,比如以下两种情况:①在左侧检验中,样本平均数大于虚无假设假定的平均数,②右侧检验中,样本平均数小于虚无假设假定的平均数。

Excel: Excel 本身没有进行 t 检验的功能,使用数据工作台 Data Desk/XL,将它安装于 Excel。Data Desk/XL 是一个统计插件,它完善了 Excel 的标准统计功能。

首先进入样本数据列 A。如果使用 Excel 2003,点击 [DDXL]。如果你使用 Excel 2007,请点击 [Add-Ins],然后再点击 [DDXL]。选择 [Hypothesis Tests],在功能类型选项中,选择 [1 Var t Test],然后点击铅笔图标,输入数据范围,比如 A 数据列有 12 个数值,就输入 A1:A12。然后点击 [OK]。接下来完成在对话框中所列出的四个步骤,在第四步完成以后点击 [Compute],就会呈现出 P 值、检验统计量和统计结论。

STATDISK: 选择 [Analysis],然后再选择 [Hypothesis Testing],对于 10.1 中所讨论的方法,选择 [Mean-One Sample]。将会出现一个对话框,点击对话框左上角,然后选择被检验的项目,继续输入对话框中的其他项目,完成以后点击 [Evaluate],结果包括检验统计量和 P 值。

双向表的假设检验

SPSS: 在 SPSS 的数据窗中,在第一列中输入次数,在第二列输入对应的列名称,在第三列中输入对应的行名称。所计的次数必须按如下步骤进行加权处理:点击 [Data],然后选择菜单项 [Weight Cases],点击标明“Weight Case by”的按钮,将代表次数的变量选入标注着“Frequency Variable”字样的框中。然后再点击 [Analyze],选择 [Descriptive Statistics],接下来选择 [Crosstabs],将代表行变量名称的变量选入标注着“Row (s)”的框中,将代表列变量名称的变量选入标注着“Column (s)”的框中。现在点击 [Statistics] 按钮,然后在打开的对话框中

选择标注“Chi-Square”的小方框,然后点击 [Continue]。接下来点击 [Cells] 按钮,然后在打开的对话框中核对标明“Observed”和“Expected”的小方框,然后点击 [Continue]。最后,点击 [OK]。呈现出来的一个表格将显示实际观测次数和理论期望次数,标注为“Chi-Square Tests”的表格则显示检验统计量卡方的值(上面标注“Pearson Chi-Square”),同一行中还显示 P 值。

Excel: Excel 本身没有进行检验的功能,使用数据工作台 Data Desk/XL,将它安装于 Excel。Data Desk/XL 是一个统计插件,它完善了 Excel 的标准统计功能。

首先在列 A 中输入列变量名称,其次在列 B 中输入行变量名称,最后,在列 C 中输入对应的实际观测次数。如果使用 Excel 2003,点击 [DDXL]。如果你使用 Excel 2007,请点击 [Add-Ins],然后再点击 [DDXL]。

现在选择 [Tables],然后选择 [Indep. Test for Summ Data]。然后点击标注 [Variable One Names] 的铅笔图标,输入数据范围,包括列变量的名称,比如 A1:A6。再点击标注 [Variable Two Names] 的铅笔图标,输入数据范围,包括行变量的名称,比如 B1:B6。接下来点击标注 [Counts] 的铅笔图标,输入对应各类观测次数,比如 C1:C6。然后点击 [OK] 在对话框中得到检验结果,包括 P 值和检验统计量卡方(理论期望次数也可呈现出来)。

STATDISK: 首先在数据窗口的列中输入实际观测次数,从主菜单条中选择 [Analysis] 然后再选择 [Contingency Tables],继续确定次数列,完成以后点击 [Evaluate],结果包括检验统计量、临界值、 P 值和统计结论。

一元方差分析

SPSS: 在 SPSS 数据窗口第一列输入所有数据,第二列输入对应的分类编码数字,比如输入 1 对应第一种类别,输入 2 对应第二种类别,依次类推。点击 [Analyze],选择 [Compare Means] [One-Way ANOVA]。在出现的对话框中,点击第一列的变量名称,然后点击中间的按钮将它移入标注着“Dependent List”的框中,然后点击第二列的分类变量,然后点击中

间的按钮将它移入标注着“Factor”的框中，点击 [OK] 就可以得到结果，其中包括 *P* 值。

Excel: Excel 本身没有进行检验的功能，使用数据工作台 Data Desk/XL，将它安装于 Excel。Data Desk/XL 是一个统计插件，它完善了 Excel 的标准统计功能。

首先在列 A, B, C... 中输入数据。如果你使用 Excel 2003，点击主菜单项中的 [Tools]，然后选择 [Data Analysis]。如果你使用 Excel 2007，点击 [Data]，然后点击 [Data A-

nalysis]，然后选择 [Anova: Single Factor]。在对话框中，输入样本数据的范围。（比如，如果第一个数据在 A 列第一行，最后一个数据在 C 列第 12 行，就输入“A1: C12。”结果包括一元方差的 *P* 值。

STATDISK: 在数据窗口的列里输入数据，从主菜单条中选择 [Analysis] 然后再选择 [One Way Analysis of Variance]，然后选择样本数据列，完成以后点击 [Evaluate]，结果包括一元方差分析的 *P* 值。

| 聚焦犯罪学 |

当你看到一个事物时能否辨别真假

假设你的导师给你布置了一项家庭作业，让你抛掷一枚硬币 200 次，并按照实际的顺序记录下结果。下面两组数据是两个学生所得到的结果。现在假如你知道其中一个学生真的做了这个实验，而另一个学生的数据则完全是伪造的。你能分辨出来哪一个是假的吗？

第一组数据 (H 代表正面向上, T 代表反面向上):

H T H T H H T T T T T H T H T T T T H
H H T T T T H T T H T T H H H T T H H T
H T H T H H H H T T H T H T H H H H T H
T T H H T T H H H T T T T T H H T H T
T T H T H T H H T T T H H T H H H H H T
H T H H T H T T T T H T T T H T H T H H
H T H H T H H H T H T H H H T T T H T H
T H T T T H T H H T H H H H T T T H H T
T H T H H T T H T H H H T H H T H T T H

第二组数据 (H 代表正面向上, T 代表反面向上):

T H H T T H H T T H T H H T T H H T H T
H T T H T T H H T T H T T T H H T H T H
H H T T H T H H T H T T H T T H H T T H
H T H H T T H T H T H H T H T H T H H T
H T H H T H T T H H T T H H T H T H H T
T H T H H T T H T H T T H T T H T H H T
H T H T T H T H H T H T H T H T H H T H
T T H T H T H H T T H T T H T H H T H H
H T H H T T H T H T H T H T H H T H T T
T H H T H T H H T H H T T H H T H T H T



为了让这个猜测工作容易些，下表概括了这两组数据的一些特征，可以帮助你判断哪一组数据是假的。

第一组数据的特征	第二组数据的特征
总计 97H, 103T	总计 101H, 99T
一行中有连续 6T 有两例	在一行中没有超过 3H 或 3T 的
一行中有连续 4H 有 5 例	
一行中有连续 4T 有 3 例	

如果和大多数人一样的话，你很可能会猜测第一组数据是伪造的。因为它正面朝上和反面朝上的次数与人们期望的正反面各占100次相差很远，而且出现了两次在一行里连续6个反面朝上，还有数次在一排里连续4个正面或反面朝上。

但是，做如下考虑：在一行里出现6次正面朝上的概率是 $(1/2)^6$ ，即 $1/64$ 。出现6次连续反面朝上的概率也是 $1/64$ 。这样的话，掷硬币200次，出现至少一次在一排里连续6次正面朝上或反面朝上是可能的，所以在第一组数据中出现连续的正面朝上或反面朝上不应该感到奇怪。相反，在第二组数据中，在一排里连续4个正面朝上或反面朝上的都没有，尽管这样连续的可能性只有 $(1/2)^4$ ，即 $1/16$ ；我们得出结论第二组数据基本可以确定是伪造的。

这是一个揭示了统计学在犯罪学领域中重要应用的简单例子：它经常可能用于发现人们是否伪造各种数据。例如，统计学帮助银行管理者和财政审计员查出造假的财务报表，也帮助美国国税局查出造假的税收申报，帮助科学家发现其他科学家伪造的数据。

一个最有效的发现数据造假的工具是物理学家弗兰克·本福德（Frank Benford）发现的。在20世纪30年代，弗兰克·本福德注意到一个对数表（科学家和工程师在计算机发明之前经常使用的工具）从数字1开始的前几页比起后面的页面来磨损得更厉害。根据这个观察，不久他发现了在我们的日常生活中存在很多的数字集合，例如股票价格、棒球计分、湖水的面积，而且他发现以1开头的数字比起以2开头的数字多，以2开头的数字比以3开头的数字多，依此类推。他最终发明了一个公式来说明数据如何从不同的数字开始，这个公式被称做本福德法则。图10-4显示根据他的这个法则所预测的很多数据的第一个数字，以及来自几个数据集合中实际的数字结果。注意到本福德法则能够很好地解释结果。[有趣的是，本福德法则第一次被发现是在50年以前，是天文学家和数学家西蒙·纽科姆（Simon Newcomb）在1881年发表的。然而，直到本福德完成了他的工作之前纽科姆的文章却被忽视了。]

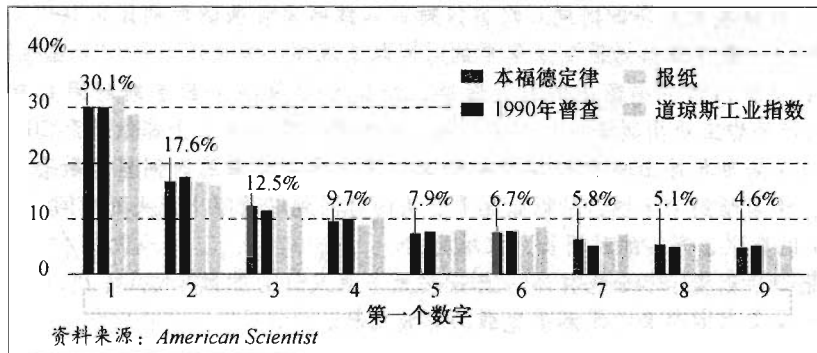


图10-4 非预期的模型

注：不同类型的数据遵从本福德定律，误差幅度在2%内，包括出现在报纸前页的数据，1990年人口普查中的3141个县城的人口，从1990到1993年的道琼斯工业指数

资料来源：Malcolm W. Browne, "Following Benford's Law, or Looking Out for No. 1," *New York Times*, August 4, 1998, p. B10.

本福德法则令人感到吃惊，因为大多数人认为从1~9每一个数字作为开头的概率相等。确实，作为一组随机数据，比如抽奖数字，以1开头的数字并不比其他数字开头的更多（所以本福德法则不能被用来选取抽奖数字）。但是，因为图10-5真正的税务表格中的数据（左二）遵循本福德法则（左一）。相反，1995年一项研究中的财务报表数据（右二）不遵循本福德法则。根据这个事实，地方检察官怀疑是造假，而且最终被证明确实是造假了。“随机猜测数据”（右一）来自美国佐治亚理工学院的西奥多P. 希尔（Theodore P. Hill）教授的学生（本聚焦篇很多信息的来源）。这个猜测一点都不符合本福德法则，这就解释了为什么人们伪造的数据经常会被发现。

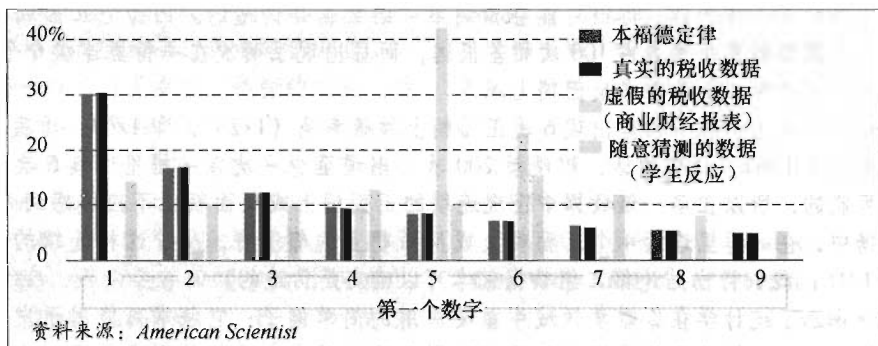


图 10-5 调查谎报数据

注: 来自 169 662 行税收表单的真实税收数据的第一个数字遵循本福德定律, 来自 1995 的布兰克林商业研究中现金支出和工资单的谎报数据并不遵循本福德定律, 同样, 来自要求 743 名大学一年级新生随意写下的六位随机数字, 也不遵循本福德定律。

资料来源: Adapted from Malcom W. Browne, "Following Benford's Law, or Looking Out for No. 1," *New York Times*, August 4, 1998, p. B10.

本福德法则在过去几十年里让科学家和数学家感到很困惑, 今天它看起来能够被理解了, 但仍然很难解释。这里有一个解释, 为什么本福德法则能够用于道琼斯平均工业指数, 这得益于南方卫理公会大学的马克 J. 尼格瑞尼博士 (Mark J. Nigrini) 在《纽约时报》上的报道: 假设道琼斯平均工业指数是 1 000, 那么第一个数字是 1, 每年按 20% 的比率增加。按这个速度增长至少 4 年时间, 这一段时间里道琼斯平均工业指数仍会保持在 1 000 多, 仍然以数字 1 开头, 大约 4 年以后, 到指数到达了 2 000, 它会保持以数字 2 开头直至指数达到 3 000。然而, 从 2 000 到达 3 000 只需要增长 50%, 只需要两年多的时间。这样, 指数以数字 2 开头的时间比以数字 1 开头的的时间的一半要多些。第一个数字连续地改变所需要的时间越来越短。当道琼斯平均工业指数达到 9 000 的时候, 它只需要增加 11%, 花费几个月的时间, 所以 9 当第一个数字的时间只有几个月。到了 10 000 点, 道琼斯平均工业指数又以数字 1 开头, 直到道琼斯平均工业指数达到 20 000 点的时候才会发生改变, 意味着每年按 20% 的速度增长的话, 又一个 4 年多的时间里指数会一直以数字 1 开头。这样的话, 如果绘制一个图, 从以数字 1 开头的的时间到以数字 9 开头的的时间, 你会发现以数字 1 开头的的时间比以数字 2 开头的的时间长, 依次类推。

总之, 本福德法则揭示了数据并不一直是按大多数人认为的频率显示。它不仅解释了关于数字的一些神秘的地方, 它还是一个调查犯罪的有效工具。

问题讨论

1. 尝试着和你的朋友一起做以下这个实验: 让你的一个朋友记录真实的 200 次投掷硬币的实验结果, 另一个朋友则编造一个 200 次掷硬币的实验结果。把他们的结果匿名地呈现给你, 保证你不知道数据是来自哪个朋友的。用这一节所学习的观念来判断哪一份数据是伪造的。在你做出猜测之后, 让你的朋友来告诉你猜测是否正确。讨论你识别伪造的能力。
2. 简要叙述怎样用本福德法则来查出伪造的税务报表。
3. 只采用本福德法则就能证明数据是伪造的吗? 还是只能够说明数据需要进一步调查? 请说明理由。
4. 找一组类似于图 10-4 的数据, 做一个描述以数字 1 开头和以数字 9 开头的次数的条形图。这组数据服从本福德法则吗? 请说明理由。

建议阅读

1. Browne, Malcolm W. "Following Benford's Law, or Looking Out for No. 1." *New York Times*, August 4, 1998.
2. Hill, Theodore P. "The First Digit Phenomenon." *American Scientist*, Vol. 86, No. 358, July-

August 1998.

3. Nigrini, Mark J. *Digital Analysis Using Benford's Law*. Global Audit Publications, Vancouver, 2000.

| 聚焦教育 |

四年级的学生能用统计学做什么

9岁的艾米丽·罗莎是一个四年级的学生，正尝试着决定做一个什么样的学校科学展览项目。她考虑要做一项关于M&M公司糖果颜色的项目，这时她注意到妈妈，一个护士正在观看一个叫做“接触治疗”或“TT”的练习录像。TT是一项首选的医学治疗方法，世界各地都在使用，在一些护士学校里甚至作为课程讲授。但是它的有效性从来没有人证明。艾米丽告诉妈妈她想做一个检验TT有效性的实验，并且把它作为她的科学展览项目。

尽管叫做接触治疗，实际上TT治疗师并不接触他们的罹患者，而是在罹患者身体上方几英寸的地方移动他们的手。TT的支持者声称受过训练的治疗师移动他们的手就能够感觉到并且操纵所谓的“人体能量场”。通过正确地进行这些操作，治疗师坚信他们能够治愈很多的疾病。艾米丽·罗莎的科学展览项目力求发现TT治疗师是否真的能感觉到人们的能量场。

为了完成这个项目，艾米丽找了21个TT治疗师来参加她的简单实验。每一个治疗师都坐在艾米丽的对面，把他或她的胳膊在一个小洞里伸出来，手掌向上。艾米丽在治疗师和她之间立起有两个分隔的洞的硬纸板。这使得治疗师和她都不能看到对方的脸，但是艾米丽能看到治疗师的手。

艾米丽将她的手放在一个治疗师双手其中一支的上面几英寸的位置，如果治疗师真的能感受到她的“身体能量场”，治疗师应该能够说出是自己的左手还是右手更接近艾米丽的手。这样，每一次试验结束后艾米丽都记录治疗师判断她手的位置是正确的还是错误的。

艾米丽采取了几项措施以保证她实验的统计有效性。例如，为了随机地选择一只手，在每一次试验时艾米丽投掷一枚硬币来决定自己是把手放在治疗师的左手上还是右手上。为了有足够的数​​据使实验结果符合统计显著性的要求，21个治疗师里面有14个人每人进行10次试验，另外7个人每人进行了20次试验。

这些结果对TT治疗师来说是严重的失败。因为在每次试验里只有两个可能的答案，左手或右手，如果纯属猜测的话，治疗师猜对的可能性应该有50%。但是这次实验全部数据显示TT治疗师猜对的次数只占44%。此外，没有一位治疗师操作的结果能够显著地好于随机猜测。艾米丽还检验了经验丰富的治疗师是否比经验较少的治疗师表现更好，结果也否定了这一假设。艾米丽根据实验结果得出结论：如果存在“人类能量场”（她怀疑这是不存在的），TT治疗师也不能感觉到，所以即使“人类能量场”存在的话，如果TT治疗师不能感受到它的存在，就更别提通过操纵它来治疗疾病了。

这项研究最有意思的一个方面是艾米丽能够完成这个实验。在过去一些对TT有效性的怀疑论者希望进行相似的研究，但是TT治疗师拒绝参与。一个著名的怀疑者魔术师詹姆斯·兰迪（James Randi）甚至提出对通过类似艾米丽实验的每一个治疗师提供110万美元。但只有一个治疗师参加了他的挑战，而且20次试验中她只通过了11次，和随机猜测得到的结果差不多。为什么艾米丽能够成功，而很多具有丰富经验的研究者都失败了？显然，这些治疗师同意参加艾米丽的实验，是因为他们认为这个四年级的小学生对他们不会造成威胁。

艾米丽的实验引起了媒体的注意，不久就传到了退休的宾夕法尼亚州精神病学家史蒂芬·巴雷特（Stephen Barrett）那里。巴雷特博士专门从事于揭穿“假冒医生”的治疗，他说服艾米丽和她的妈妈将她们的研究结果发表在医学研究论文上。这篇论文发表在《美国医学联合会杂志》上，这时的艾米丽才11岁，就成了在非常权威的杂志上发表文章的年龄最小的作者。



问题讨论

1. 在艾米丽的研究结果发表之后，很多 TT 的支持者声称她的实验无效，因为她和她的妈妈都有反对 TT 的偏见。根据她的实验设计的方式，你认为她的个人偏见是否会影响结果，为什么？
2. 另一个反对艾米丽实验的说法是，这是一个单盲实验，而不是双盲实验。就是治疗师看不到艾米丽在干什么，但是艾米丽能看到治疗师在干什么。你认为在这个例子中的反对是否有效？你是否能想到一种方法设计一个双盲实验对艾米丽的实验进行重复？
3. 艾米丽的实验并没有直接测验 TT 治疗方法是否真的有效，因为它没有去检验通过 TT 治疗病人是否真的好转。设计一种统计上有意义的方法检验 TT 治疗是否会比安慰剂更有效。
4. 根据艾米丽研究的结果，怀疑论者认为 TT 既然无效，就不应该被继续应用。你同意这种看法吗，为什么？

建议阅读

1. Ball, T. S., and Alexander, D. D. "Catching Up with Eighteenth Century Science in the Evaluation of Therapeutic Touch." *Skeptical Inquirer*, July/August 1998.
2. Kolata, Gina. "4th Grader Challenges Alternative Therapy." *New York Times*, April 2, 1998.
3. Rosa, E. "TT and Me." *Skeptic Magazine*, September 1998.
4. Rosa, L., Rosa, E., Sarnier, L., and Barrett, S. "A Close Look at Therapeutic Touch." *Journal of the American Medical Association*, Vol. 279, No. 1005, April 1, 1998.

结语：统计学展望

一个统计学的入门课程是不能使你成为一个统计学专家的。学习了本书，你可能会感到自己仍然没有达到在实际中自信地应用统计的程度。不管怎样，你现在可以充分理解统计学，可以批判性地理解新闻中的统计研究，也可以做到在需要更多信息时与统计专家进行交流。并且如果你继续进行更深入的统计课程学习，你应该可以更好地理解超过本书范畴的更重要主题。

更重要的是，写这本书的初衷并不是使你成为统计专家，而是旨在促进你的工作能力，使你成为一个受过良好教育的人。你应该理解概率和机遇的基本概念；你应该知道怎样获得数据，它对研究集中量数（比如平均数和中位数）、差异量数（比如全距和标准差）、分布形态（以图和表的方式呈现）以及确定异常值存在是非常重要的；你应该知道和理解总体参数估计的重要性（比如总体平均数和总体比率），以及总体参数的检验；你应该理解良好抽样的重要性；你应该知道即使一个样本容量看起来相对小，但许多调查和民意测验仍然可以获得一个良好的结果。尽管许多人拒绝相信它，但是如果抽样是精心设计并且可操作的，全国范围内一个只有 1 700 人的调查也可以提供一个非常好的结果。

过去，如果一个人能够阅读本书，就会被看做受过良好的教育，但是我们处在一个有更多要求的新世纪，今天一个受过良好教育的人必须能够阅读和书写、能够理解文艺复兴的价值，会操作计算机，并且能进行统计推理。统计课程的学习，可以帮助我们看清有时被错误方法歪曲或者被数据呈现方式隐藏的真相。对于雇主和被雇用者乃至所有公民而言，理解统计学都是必不可少的。赫伯特·乔治·威尔斯曾经说过“和读写能力一样，终有一天统计思维会成为每个合格市民的必备素质。”这一天现在已经到来了。

附录 A 标准分数表

表 A-1 是表 5-1 的更详细的版本，注意曲线下的面积对应百分比，阅读表 A-1，先在右侧一栏找出标准分数的前两位数，然后在第一行找到第三位数字。

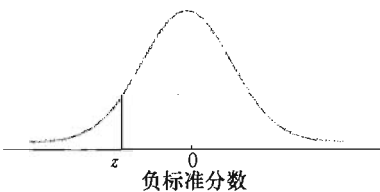


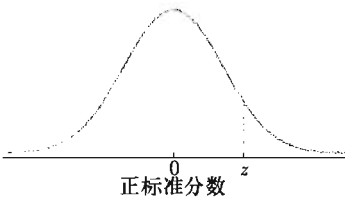
表 A-1 标准正态分布：从左侧开始累计的区域

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.50 及以下	0.000 1									
-3.4	0.000 3	0.000 3	0.000 3	0.000 3	0.000 3	0.000 3	0.000 3	0.000 3	0.000 3	0.000 2
-3.3	0.000 5	0.000 5	0.000 5	0.000 4	0.000 4	0.000 4	0.000 4	0.000 4	0.000 4	0.000 3
-3.2	0.000 7	0.000 7	0.000 6	0.000 6	0.000 6	0.000 6	0.000 6	0.000 5	0.000 5	0.000 5
-3.1	0.001 0	0.000 9	0.000 9	0.000 9	0.000 8	0.000 8	0.000 8	0.000 8	0.000 7	0.000 7
-3.0	0.001 3	0.001 3	0.001 3	0.001 2	0.001 2	0.001 1	0.001 1	0.001 1	0.001 0	0.001 0
-2.9	0.001 9	0.001 8	0.001 8	0.001 7	0.001 6	0.001 6	0.001 5	0.001 5	0.001 4	0.001 4
-2.8	0.002 6	0.002 5	0.002 4	0.002 3	0.002 3	0.002 2	0.002 1	0.002 1	0.002 0	0.001 9
-2.7	0.003 5	0.003 4	0.003 3	0.003 2	0.003 1	0.003 0	0.002 9	0.002 8	0.002 7	0.002 6
-2.6	0.004 7	0.004 5	0.004 4	0.004 3	0.004 1	0.004 0	0.003 9	0.003 8	0.003 7	0.003 6
-2.5	0.006 2	0.006 0	0.005 9	0.005 7	0.005 5	0.005 4	0.005 2	0.005 1	0.004 9	0.004 8
-2.4	0.008 2	0.008 0	0.007 8	0.007 5	0.007 3	0.007 1	0.006 9	0.006 8	0.006 6	0.006 4
-2.3	0.010 7	0.010 4	0.010 2	0.009 9	0.009 6	0.009 4	0.009 1	0.008 9	0.008 7	0.008 4
-2.2	0.013 9	0.013 6	0.013 2	0.012 9	0.012 5	0.012 2	0.011 9	0.011 6	0.011 3	0.011 0
-2.1	0.017 9	0.017 4	0.017 0	0.016 6	0.016 2	0.015 8	0.015 4	0.015 0	0.014 6	0.014 3
-2.0	0.022 8	0.022 2	0.021 7	0.021 2	0.020 7	0.020 2	0.019 7	0.019 2	0.018 8	0.018 3
-1.9	0.028 7	0.028 1	0.027 4	0.026 8	0.026 2	0.025 6	0.025 0	0.024 4	0.023 9	0.023 3
-1.8	0.035 9	0.035 1	0.034 4	0.033 6	0.032 9	0.032 2	0.031 4	0.030 7	0.030 1	0.029 4
-1.7	0.044 6	0.043 6	0.042 7	0.041 8	0.040 9	0.040 1	0.039 2	0.038 4	0.037 5	0.036 7
-1.6	0.054 8	0.053 7	0.052 6	0.051 6	0.050 5	0.049 5	0.048 5	0.047 5	0.046 5	0.045 5
-1.5	0.066 8	0.065 5	0.064 3	0.063 0	0.061 8	0.060 6	0.059 4	0.058 2	0.057 1	0.055 9
-1.4	0.080 8	0.079 3	0.077 8	0.076 4	0.074 9	0.073 5	0.072 1	0.070 8	0.069 4	0.068 1
-1.3	0.096 8	0.095 1	0.093 4	0.091 8	0.090 1	0.088 5	0.086 9	0.085 3	0.083 8	0.082 3
-1.2	0.115 1	0.113 1	0.111 2	0.109 3	0.107 5	0.105 6	0.103 8	0.102 0	0.100 3	0.098 5
-1.1	0.135 7	0.133 5	0.131 4	0.129 2	0.127 1	0.125 1	0.123 0	0.121 0	0.119 0	0.117 0
-1.0	0.158 7	0.156 2	0.153 9	0.151 5	0.149 2	0.146 9	0.144 6	0.142 3	0.140 1	0.137 9
-0.9	0.184 1	0.181 4	0.178 8	0.176 2	0.173 6	0.171 1	0.168 5	0.166 0	0.163 5	0.161 1
-0.8	0.211 9	0.209 0	0.206 1	0.203 3	0.200 5	0.197 7	0.194 9	0.192 2	0.189 4	0.186 7
-0.7	0.242 0	0.238 9	0.235 8	0.232 7	0.229 6	0.226 6	0.223 6	0.220 6	0.217 7	0.214 8

注：对于标准分数在 -3.49 以下，对应于 0.000 1。

①使用内插法得到的常用数值。

标准分数	区域
-1.645	0.050 0
-2.575	0.005 0



(续)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0.6	0.274 3	0.270 9	0.267 6	0.264 3	0.261 1	0.257 8	0.254 6	0.251 4	0.248 3	0.245 1
-0.5	0.308 5	0.305 0	0.301 5	0.298 1	0.294 6	0.291 2	0.287 7	0.284 3	0.281 0	0.277 6
-0.4	0.344 6	0.340 9	0.337 2	0.333 6	0.330 0	0.326 4	0.322 8	0.319 2	0.315 6	0.312 1
-0.3	0.382 1	0.378 3	0.374 5	0.370 7	0.366 9	0.363 2	0.359 4	0.355 7	0.352 0	0.348 3
-0.2	0.420 7	0.416 8	0.412 9	0.409 0	0.405 2	0.401 3	0.397 4	0.393 6	0.389 7	0.385 9
-0.1	0.460 2	0.456 2	0.452 2	0.448 3	0.444 3	0.440 4	0.436 4	0.432 5	0.428 6	0.424 7
-0.0	0.500 0	0.496 0	0.492 0	0.488 0	0.484 0	0.480 1	0.476 1	0.472 1	0.468 1	0.464 1
0.0	0.500 0	0.504 0	0.508 0	0.512 0	0.516 0	0.519 9	0.523 9	0.527 9	0.531 9	0.535 9
0.1	0.539 8	0.543 8	0.547 8	0.551 7	0.555 7	0.559 6	0.563 6	0.567 5	0.571 4	0.575 3
0.2	0.579 3	0.583 2	0.587 1	0.591 0	0.594 8	0.598 7	0.602 6	0.606 4	0.610 3	0.614 1
0.3	0.617 9	0.621 7	0.625 5	0.629 3	0.633 1	0.636 8	0.640 6	0.644 3	0.648 0	0.651 7
0.4	0.655 4	0.659 1	0.662 8	0.666 4	0.670 0	0.673 6	0.677 2	0.680 8	0.684 4	0.687 9
0.5	0.691 5	0.695 0	0.698 5	0.701 9	0.705 4	0.708 8	0.712 3	0.715 7	0.719 0	0.722 4
0.6	0.725 7	0.729 1	0.732 4	0.735 7	0.738 9	0.742 2	0.745 4	0.748 6	0.751 7	0.754 9
0.7	0.758 0	0.761 1	0.764 2	0.767 3	0.770 4	0.773 4	0.776 4	0.779 4	0.782 3	0.785 2
0.8	0.788 1	0.791 0	0.793 9	0.796 7	0.799 5	0.802 3	0.805 1	0.807 8	0.810 6	0.813 3
0.9	0.815 9	0.818 6	0.821 2	0.823 8	0.826 4	0.828 9	0.831 5	0.834 0	0.836 5	0.838 9
1.0	0.841 3	0.843 8	0.846 1	0.848 5	0.850 8	0.853 1	0.855 4	0.857 7	0.859 9	0.862 1
1.1	0.864 3	0.866 5	0.868 6	0.870 8	0.872 9	0.874 9	0.877 0	0.879 0	0.881 0	0.883 0
1.2	0.884 9	0.886 9	0.888 8	0.890 7	0.892 5	0.894 4	0.896 2	0.898 0	0.899 7	0.901 5
1.3	0.903 2	0.904 9	0.906 6	0.908 2	0.909 9	0.911 5	0.913 1	0.914 7	0.916 2	0.917 7
1.4	0.919 2	0.920 7	0.922 2	0.923 6	0.925 1	0.926 5	0.927 9	0.929 2	0.930 6	0.931 9
1.5	0.933 2	0.934 5	0.935 7	0.937 0	0.938 2	0.939 4	0.940 6	0.941 8	0.942 9	0.944 1
1.6	0.945 2	0.946 3	0.947 4	0.948 4	0.949 5 ^①	0.950 5	0.951 5	0.952 5	0.953 5	0.954 5
1.7	0.955 4	0.956 4	0.957 3	0.958 2	0.959 1	0.959 9	0.960 8	0.961 6	0.962 5	0.963 3
1.8	0.964 1	0.964 9	0.965 6	0.966 4	0.967 1	0.967 8	0.968 6	0.969 3	0.969 9	0.970 6
1.9	0.971 3	0.971 9	0.972 6	0.973 2	0.973 8	0.974 4	0.975 0	0.975 6	0.976 1	0.976 7
2.0	0.977 2	0.977 8	0.978 3	0.978 8	0.979 3	0.979 8	0.980 3	0.980 8	0.981 2	0.981 7
2.1	0.982 1	0.982 6	0.983 0	0.983 4	0.983 8	0.984 2	0.984 6	0.985 0	0.985 4	0.985 7
2.2	0.986 1	0.986 4	0.986 8	0.987 1	0.987 5	0.987 8	0.988 1	0.988 4	0.988 7	0.989 0
2.3	0.989 3	0.989 6	0.989 8	0.990 1	0.990 4	0.990 6	0.990 9	0.991 1	0.991 3	0.991 6
2.4	0.991 8	0.992 0	0.992 2	0.992 5	0.992 7	0.992 9	0.993 1	0.993 2	0.993 4	0.993 6
2.5	0.993 8	0.994 0	0.994 1	0.994 3	0.994 5	0.994 6	0.994 8	0.994 9 ^①	0.995 1	0.995 2
2.6	0.995 3	0.995 5	0.995 6	0.995 7	0.995 9	0.996 0	0.996 1	0.996 2	0.996 3	0.996 4
2.7	0.996 5	0.996 6	0.996 7	0.996 8	0.996 9	0.997 0	0.997 1	0.997 2	0.997 3	0.997 4
2.8	0.997 4	0.997 5	0.997 6	0.997 7	0.997 7	0.997 8	0.997 9	0.997 9	0.998 0	0.998 1
2.9	0.998 1	0.998 2	0.998 2	0.998 3	0.998 4	0.998 4	0.998 5	0.998 5	0.998 6	0.998 6
3.0	0.998 7	0.998 7	0.998 7	0.998 8	0.998 8	0.998 9	0.998 9	0.998 9	0.999 0	0.999 0
3.1	0.999 0	0.999 1	0.999 1	0.999 1	0.999 2	0.999 2	0.999 2	0.999 2	0.999 3	0.999 3
3.2	0.999 3	0.999 3	0.999 4	0.999 4	0.999 4	0.999 4	0.999 4	0.999 5	0.999 5	0.999 5
3.3	0.999 5	0.999 5	0.999 5	0.999 6	0.999 6	0.999 6	0.999 6	0.999 6	0.999 6	0.999 7
3.4	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 7	0.999 8
3.50	0.999 9									
及以上										

注：对于标准分数在-3.49 以下，对应于 0.000 1。

①使用内插法得到的常用数值。

标准分数	区域
-1.645	0.950 0
-2.575	0.995 0

表 A-2 选择标准分数的临界值

	左侧检验	右侧检验	双侧检验
0.05 显著性水平	-1.645	1.645	-1.96 或者 1.96
0.01 显著性水平	-2.33	2.33	-2.576 或 2.576

附录 B 随机数字表

生成一组随机选择的数字对于许多统计应用是有用的，你可以使用计算器或者计算机生成随机数字，但有时使用下面的随机数字表是非常方便的。使用如下这个表可以生成随机数字，可以从表中的任意一个位置机选择数字 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 中的一个，每一个数字出现在任何一个顺序的机会都是相等的。这样，你可以从表的任意一个位置开始，按照它们所呈现的顺序得到一个简单的随机数字结果。更大的随机数字可以利用网络资源 (<http://www.aw.com/bbt>)。

例 B-1 随机生成一个回答问题做“是”“否”的一个清单

解答 从表中任意选取一个点，如果选取的是数字 0, 1, 2, 3, 4 就记做答案为“是”；如果选取的数字是 5, 6, 7, 8, 9 就记做答案为“否”，从你开始的点继续下去，使用每一个数字以确定答案为“是”“否”，最后形成一个完整的清单。

例 B-2 生成一个字母形式成绩的随机清单

解答 让 0 或者 1 代表 A, 2 或者 3 代表 B, 4 或者 5 代表 C, 6 或者 7 代表 D, 8 或者 9 代表 E, 从表中任意选取一个点开始，然后使用所列出的数字来确定一个成绩的列表。

9 9 3 2 7	5 6 0 8 1	6 0 2 3 2	8 3 3 1 2	4 7 6 3 4
9 7 1 8 1	6 6 7 6 6	5 4 4 7 7	6 8 1 7 1	0 8 4 9 9
8 1 7 5 0	7 8 5 2 0	9 4 3 9 0	7 6 1 9 1	0 8 7 3 4
0 5 1 0 2	8 7 4 0 3	9 2 6 2 5	8 4 2 2 5	1 9 8 3 4
2 7 8 0 8	1 8 5 6 6	4 4 5 5 4	9 3 5 2 8	6 5 5 4 3
4 8 8 3 3	8 4 6 9 1	8 2 5 7 6	9 7 1 2 3	6 5 1 8 2
5 4 5 8 7	3 8 4 5 7	4 5 2 2 2	7 7 0 2 3	0 2 4 8 6
4 1 9 4 3	0 7 1 9 0	7 3 1 4 0	8 3 2 8 0	5 0 1 0 1
2 8 7 4 6	5 7 7 6 0	0 8 9 5 5	4 0 7 3 9	1 6 3 3 2
5 8 6 8 6	9 6 7 7 5	1 3 5 2 9	7 6 6 3 5	9 4 6 0 5
8 0 9 4 8	5 0 5 6 9	1 0 6 9 5	9 0 7 8 9	9 4 8 3 7
6 1 0 4 1	7 4 0 0 3	5 6 4 2 1	5 1 1 9 0	0 2 5 0 7
3 4 0 9 1	9 3 9 5 1	0 7 4 8 1	1 9 7 0 7	1 4 5 2 6
9 9 6 5 0	7 8 6 1 0	4 9 8 7 7	4 4 7 4 0	7 8 6 4 9
1 5 0 7 8	1 6 3 6 9	5 4 9 5 4	2 4 6 0 4	3 2 6 8 4
5 7 2 8 5	8 3 1 6 4	4 2 2 3 7	0 6 6 3 2	3 5 0 4 6
1 2 3 5 7	1 9 7 7 6	5 5 3 1 9	6 0 5 1 6	3 8 0 3 4
3 4 9 3 5	0 3 7 4 6	2 1 8 5 1	4 1 7 0 2	1 4 9 4 9
2 4 2 6 6	9 9 7 2 1	7 7 3 2 0	6 7 0 7 3	9 3 3 7 5
9 9 1 5 2	8 3 9 9 4	8 9 6 0 6	7 4 6 3 1	3 6 9 8 3
2 6 3 3 9	3 4 1 9 3	0 6 9 1 2	4 1 9 9 8	7 3 2 6 9
1 3 7 8 0	1 0 6 1 6	5 4 9 3 0	7 0 7 2 3	1 7 8 9 2
8 8 4 8 4	0 2 8 5 5	7 3 7 1 2	1 8 3 5 2	0 1 5 3 2
4 2 9 2 8	8 8 9 1 2	4 6 7 1 7	5 1 5 1 9	3 2 2 8 0
1 3 7 2 1	0 6 4 7 6	1 0 8 4 8	9 7 6 3 5	5 1 2 2 9

参 考 答 案

1.1 练习题答案

1. 总体就是要研究的全部人和事物，样本就是总体的一个子集。这二者之间的不同在于样本仅仅是全部总体的一部分。
3. 样本统计量就是通过对原始数据进行合并或者总结而得到的样本特征。总体参数就是这个总体的特征。因为对于庞大的总体来说，一般情况下直接测量总体的参数是不现实的，所以我们通常通过测量样本统计量来推断总体参数的值。
5. 无意义
7. 无意义
9. 无意义
11. 样本：选择的1002个成年人。总体：全美国所有成年人。样本统计量：48%。总体参数的值不知道，但它是指所有美国成年人中如果被问及这个问题就会说喜欢的人在全部美国成年人中所占的百分比。
13. 样本：所选取的恒星之间的距离。总体：银河系中所有恒星间的距离。样本统计量：所选取的恒星间距离的平均数。总体参数的值不知道，但如果所有银河系中的恒星都知道的话，那么作为总体参数的这个平均距离就可以计算出来了。
15. 45% ~ 51%
17. 22.6 ~ 29.4
19. 尽管没有保证，但是调查结果说明共和党人更可能赢得大多数人的支持，因为他非常有可能获得全部选票的55% ~ 61%。
21. 基于调查结果，估计实际的投票者的百分比67% ~ 73%，而根据实际结果61%的投票来看，估计范围并没有包括这个数值。如果这个调查设计得理想的话，也就不会出现与实际结果这么大的差别，这意味着不是被试故意说谎来讨好民意测验的实施者，就是他们的记忆不准确。
23. (1) 目标：调查经营管理人员对职业选择的满意程度。总体：所有的经营管理人员。

总体参数：经营管理人员被问到，如果他们可以再次重新选择他们的职业，其中会选择一个新领域的人数所占的百分比。

(2) 样本：此项调查所选取的1733名经营管理人员。原始数据：每个被试独立的对问题的回答。样本统计量：51%。

(3) 48% ~ 54%
25. (1) 目标：调查成人中失业人数的百分比。总体：全部成年人。总体参数：全部成人中失业人数所占的百分比。
- (2) 样本：调查随机选取的成人样本。原始数据：每个被试独立的对问题的回答。样本统计量：4.6%。
- (3) 4.4% ~ 4.8%
27. 第一步，目标：确定那些拥有驾照的司机，上周在开车的时候至少使用过一次手机的百分比。
- 第二步，在有驾照的司机中选择一个样本。
- 第三步，对这个有驾照的司机样本进行调查，来确定他们当中有多少上周在开车的时候至少使用一次手机。针对这个样本，计算拥有驾照的司机上周在开车的时候至少使用一次手机的百分比。
- 第四步，使用科学统计技术做一个统计分析，推断关于持有驾照的司机上周在开车的时候至少使用一次手机的百分比。
- 第五步，基于这个总体参数的可能取值，得出一个结论，内容是关于拥有驾照的司机上周在开车的时候至少使用一次手机的百分比。
29. 第一步，目标：确定所有商务航空旅客的平均重量。
- 第二步，选择一个航空公司的乘客作为样本。
- 第三步，给每个被选择的航空乘客称重，然后计算他们的平均重量。
- 第四步，使用科学统计技术做一个统计分析，估计所有商务航空旅客的平均重量。

第五步，基于这个总体参数的可能取值，得出关于所有商务航空旅客平均重量的结论。

1.2 练习题答案

1. 普查是指对总体中每一个成员数据的收集，样本则是对总体中一部分个体数据的收集。
3. 进行分组抽样，是指我们选择了随机抽取的小组（或群集）的所有成员；分层抽样，则是从每一不同的小组（或层中）中来选取样本。
5. 无意义
7. 有意义
9. 普查具有实际意义。总体包括洛杉矶湖人队中的一小部分球员，测得球员的身高是容易做到的。
11. 普查是不可行的。统计学指导教师的数量十分庞大，让他们所有人都参加一次智商测试是极其困难的。
13. 样本是选取的四位参议员的任职时间，总体是全部 100 位参议员的任职时间。这是一个随机抽样的实例（或者叫做简单随机抽样）。然而，由于选取的样本容量太小，所以结果对总体来说不具有代表性。
15. 样本是随机选择的 1 012 名成年人，总体是全体美国成年人。因为选择的样本容量足够大，并且是从一家信誉良好的调查公司得到的数据，因此抽样结果对于总体具有较强的代表性。
17. 样本③最具有代表性，因为清单上非常可能代表了大多数人，而且没有理由认为前 1 000 人和其他人比较会存在什么特殊之处；样本①的结果有偏差，因为这项调查只选取了豪华车辆的拥有者作为样本。样本②的结果也有偏差，因为这项调查只对一个地区的人进行了抽样。样本④的结果也有偏差，因为它包括一个自选择的样本，在这个样本中，参加调查的人们可能对信用卡债务这件事情有较强烈的主观情绪。
19. 因为那位影评人是间接为迪斯尼公司工作的，所以她很可能写出有利于迪斯尼公司的一份评论。
21. 这些大学里的科学家从孟山都公司领取薪酬，为了将来能得到更多的研究项目，他们很可能会提供一些有利于孟山都公司的调查

结果来取悦这家公司。存在一种提供有利证据的倾向。

23. 这个样本是一个简单随机抽样，可能比较具有代表性，因为在抽样过程中并不存在偏差。
25. 这个样本是一个分组抽样，可能比较有代表性，尽管选取投票站的确切方法可能对抽样结果是否有偏差会有影响。
27. 这个样本是一个方便取样，结果很可能会产生偏差，因为抽样选取的家庭成员很可能具有相同的体型特征，而这种特征可能并不具有适用于一般大众的普遍性。
29. 这个样本是一个分层抽样，结果很可能会产生偏差。因为被测试者的年龄段在整个人口群体中很可能并不是均匀分配的。然而，这个结果可以通过调整来反映总体的年龄分布。
31. 这个样本是系统抽样。它可能比较具有代表性，因为没有理由可以说，如果按照字母表的顺序抽样会导致偏差。
33. 这个样本是分层抽样。结果很可能会产生偏差，因为三种不同类别中的人口不一定是相等的。然而，这个结果可以通过加权的方法反映现实人口的分布。
35. 这个样本是一个方便样本。结果很可能会有偏差，因为这是一个自选择的样本，被试包括那些对议题本身有强烈主观情绪的人们。
37. 这个样本为一个简单随机抽样，它可能比较具有代表性。
39. 对学生的体征进行简单随机抽样，只有测量足够数量的不同体型的数据，得到的结果才有意义。
41. 死亡记录的分组抽样应该可以提供很好的数据。分组时应该根据性别或者地区或者二者兼顾。

1.3 练习题答案

1. 安慰剂在外形上和治疗药物相同，但它缺少有效的活性成分，因此安慰剂本身并不能起到任何作用。安慰剂是非常重要的，只有将被试接受真正药物治疗和接受安慰剂的结果进行比较，结果才有意义。
3. 混淆是指不同因素混合在一起产生的效应。因此并不能确定某些特定结果就是由某些特定因素导致的。例如，让男性接受真正的治

疗，给予女性的是安慰剂，这样我们就不能确定实验效应是由于治疗产生的作用，还是由于被试性别的差异所导致的。

5. 进行双盲实验是很有意义的，但这种情况下的实验设计稍微有些棘手，因为被测者能够清楚地看到他们身上所穿衣服的颜色，而这对于实验评测者来说，这些衣服的颜色也是清晰可见的。只有不告知被试实验目的情况下，“盲试”才会发挥作用，因为他们关于衣服颜色的知识并不会影响实验结果。因为实验评测者进行“盲试”非常困难，所以要确保实验结果是基于客观测量而不是实验评测者的主观判断，这一点是尤其重要的。
7. 当一名研究者或者是实验者通过某些方式，诸如面部表情、声音声调或者态度在某种程度上影响了被试，就会产生实验者效应。可以通过双盲实验来避免实验者效应，这样可以使评估实验结果的实验者不知道哪些被试给予了真正的治疗，哪些被试给予了安慰剂。
9. 这是一项观察研究，因为要研究的被试只接受了测试，并没有给予任何实验处理。关注变量是每次试验正确的或者错误的结果。
11. 这是一项实验研究，因为实验处理组里的被试给予磁性手镯，而控制组里的被试手镯则没有磁性。这个研究也许不可能采用“盲试”，因为乘客可以轻易地检测出他所带的手镯是否带有磁性（通过将手镯靠近金属）。本实验的关注变量是乘客是否出现晕船症状。
13. 这是一项观察研究，同时也是一项回溯研究，研究旨在检测双胞胎（同卵双生或异卵双生）在出生前就已经确定的素质，如何影响他们后来心智技能的发展。
15. 这是一项元分析研究。关注变量是被试是否罹患了前列腺癌。
17. 这是一项实验研究。实验处理组包括转基因玉米，控制组包括非转基因玉米。关注变量是根部释放的杀虫剂含量。
19. 这是一项实验研究。实验组包括给予磁疗的被试。控制组则是给予非磁疗设施的被试。关注变量是对背部疼痛的测量。
21. 有可能发生混淆。如果两组所发生的实验效应不一样，没有办法确定这些差别是归因于实验处理（肥料或者灌溉），还是由于不同

的地区类型（潮湿或者干燥）。这个混淆变量可以通过以下方法得以避免：在潮湿和干燥的地区都使用肥料，并且在潮湿和干燥的地区也都进行灌溉。

23. 有可能发生混淆的。如果在燃料消耗中存在差异，没有办法确定这些差异是由燃料中的辛烷值所导致的还是由不同交通工具类型所导致的。通过以下方法可以避免混淆变量：一半大篷车和一半越野车使用 87 号辛烷值汽油，另外一半大篷车和另外一半越野车使用 91 号辛烷值汽油。
25. 混淆变量可能来自安慰剂效应或者实验者效应。可以通过双盲实验得以避免混淆变量。
27. 可能存在混淆变量，因为有可能存在安慰剂效应或者实验者效应。网球起到了安慰剂的作用。最好在不同时间里要求相同的被试分别举重和举网球，一半被试先举网球后举重，另一半被试先举重后举网球。
29. 控制组包括那些不听贝多芬音乐被试，实验组包括那些收听贝多芬音乐的被试。盲试可以通过对被试进行随机编码，让测试这些被试智力的实验者不知道被试分别是哪一组的。
31. 控制组包括使用的燃料不包含乙醇添加剂的汽车，实验组包括使用的燃料包含乙醇添加剂的汽车。对于汽车来说盲试是不必要，对研究者来说盲试也是不必要的，因为公里数的测试是通过客观仪器来完成的，但是，应该使用同样的汽车来使用或者不使用添加剂，这样，无关因素所引起的差异就会被排除掉。

1.4 练习题答案

1. 同行评议是同领域的专家在研究者的报告没有公开发表之前的评估研究者报告的过程。这对于增加研究的可信度是有用的，因为它意味着其他专家认为这个研究的实施过程是合适的。
3. 当实验的参与者自己选择被调查时，这些被试可能对于调查的主题持有强烈的态度，那么这样的被试组对于一般人群来说不具有代表性。
5. 无意义
7. 无意义

9. 因为研究者来自公共关系部门, 所以研究者可能是有偏差的, 指导二可能与这种偏差最有关联。
13. 这是一个自选择样本, 所以被试偏差是一个严重问题, 指导三可能与这种偏差最有关联。
15. 这个问题的语言表述是有偏差的, 它倾向于引出否定回答, 指导六可能与这种偏差最有关联。
17. 因为资金的大部分是由糖果公司和巧克力制造联合会提供的, 所以研究者会倾向于得出有利于他们的结论。如果研究者没有接收巧克力制造商的资助, 这种偏差就可能避免。如果这是研究做下去的唯一方法, 研究人员应该制定研究过程以确保所有研究结果包括负面的结果应该公开发表。
19. 因为对问题的回答是自选择的, 所以他们对问题的回答会带有强烈的情感色彩。需要一种好的抽样方法, 比如大多数调查公司所采用的简单随机抽样。
21. 第一个问题中的词语“错误”可能会误解。一些人可能认为堕胎是错误的但还是选择赞成。第二个问题可能会导致混淆, 一些人们可能认为“医生建议”意思是指女性的生命处于危险之中, 这会影响他们对于堕胎的态度。反对堕胎的人更可能引证第一个问题的结果, 而支持堕胎的人更可能会引证第二个问题的结果。
23. 第一个问题需要网络约会的研究。第二个问题包括一个确定已婚的人的第一次约会是否为网络约会的研究。第二个研究要比第一个研究有更多限制。
25. 第一个问题包括对一般大学生的研究; 第二个问题包括对那些经常酗酒的人的研究。第一个问题可以通过调查大学生来完成, 第二个问题通过调查酗酒者完成, 第二个问题要比第一个问题更难以调查。
27. 标题是吸毒, 然而故事涉及吸毒、饮酒和吸烟, 因为通常认为吸毒只涉及毒品, 而不包括香烟和酒精, 标题非常容易引起误解。
29. 没有关于平均数置信度的已知信息, 也没有提供样本容量和误差幅度。
31. 没有已知信息证明“很多公司尝试信任天气预报的结果”这个陈述。如果引用的只

有4家公司是新的, 这个增长是相当不显著的。

第1章复习题答案

1. (1) 35% ~ 41% (2) 所有美国成年人
- (3) 这是一项观察研究。因为所有被试不以任何方式加以控制和改变。关注变量是家庭中拥有的枪支, 这一研究可以有两种取值: 是或者否。
- (4) 这个值是一个样本统计量, 因为它基于1 012名成年人的样本, 不是所有成年人的总体。
- (5) 不是有效的。因为它是一个自选择样本, 很可能有被试偏差。
- (6) 随机选择一个美国成年人的样本会有许多不同的方法。但是任何一种方法应该被设计成能够确保所有可能的由1 012名成年人组成的样本被选中的机会是等同的。一种简单方法是编制一个所有美国成年人的数字编号清单, 然后用计算机随机生成1 012个不同的数字, 然后选择那些和选取的数字相对应的被试。
- (7) 在每个州选取一个家庭样本。
- (8) 在随机抽取的几个选区中选取这些选区中的所有家庭。
- (9) 根据所在城市街道的地址, 每隔10个家庭选择一个。
- (10) 选择你同学的家庭。
2. (1) 一个随机样本是这样选取样本的方法, 每个同样大小的样本都有相同的被选中的机会。
- (2) 这不是一个简单随机样本。因为不是每个由2 007个人组成的样本都有同样的被抽中的机会。例如, 在同一个基本抽样单元的2 007个人被抽取作为样本是不可能的。
- (3) 重复这个随机抽取一个基本抽样单元的过程并且从中随机抽取每一个成员。如果任何人被重复抽取, 则放弃这个样本再继续抽取。
3. (1) 不会的。因为没有关于不使用辛伐他汀的人们中发生头痛的信息。
- (2) 因为使用辛伐他汀的人们中发生头痛的比率很小, 看起来头痛不是辛伐他汀的

副作用。

- (3) 采用双盲实验，实验参与者和评估实验结果的实验者都不知道被试的所服用的是辛伐他汀还是安慰剂。
 - (4) 这是一项试验研究，因为所有被试给予了实验处理。
 - (5) 如果实验者通过某种方式，比如面部表情、声调或者态度影响到了被试，就会产生实验者效应。可以通过使用双盲实验来避免。
4. (1) 第二个问题，因为使用词汇“福利”产生了负面影响。
 - (2) 第一个问题，因为它更可能消除否定回答。
 - (3) 这里面存在大量的主观判断。一些专业调查员持反对态度，因为所有问题都会有主观偏见。但是另外一些人认为可以采用。一个重要的考虑是调查问题都能够改变人们的思考，但如果人们没有意识到或者同意的话，这样的改变不应该发生。

第1章小测验答案

1. (1) 2. (3) 3. (1) 4. (2) 5. (3)
6. (1) 7. (3) 8. (3) 9. (2) 10. (2)
11. (2) 12. (2) 13. (2) 14. (3) 15. (2)

2.1 练习题答案

1. 定性数据包括各种能够被分进不同非数字类别的取值，然而，定量数据则包括表示计数或者测量的数值。
3. 是的，可以这样说。数据既包含属性又可以量化（数字），因此所有数据既可以定性又可以量化。
5. 血型是定性数据，因为不能测量也不能计数。
7. 身高是定量数据，因为是由测量值构成。
9. 电影的长度是定量数据，因为是由测量值构成。
11. 电视节目是定性数据，因为不能测量也不能计数。
13. 人数是定量数据，因为它是计数而来的。
15. 薪资是定量数据，因为它是对金钱的计数。
17. 人数是离散变量，因为只能计整数，并且两个整数字之间不可能有别的取值。

19. 硬币数量是离散变量；因为只能计整数，并且两个整数字之间的取值无意义。
21. 时间是连续变量，因为它在一定值域范围内可以取任意值。
23. 这个选择题的测验分数是间断变量，因为只能计数。
25. 速度是连续变量，因为它在一定值域范围内可以取任意值。
27. 恒星数量是间断变量，因为只能计数。
29. 比率变量 31. 称名变量 33. 顺序变量
35. 顺序变量 37. 称名变量 39. 比率变量
41. 不能应用比率水平。比率是无意义的，因为星级不是测量或者计数得来的，星级之间的差值是无意义的。
43. 可以应用比率水平。450 公里/小时是 150 公里/小时的 3 倍。
45. 可以应用比率水平。“2 倍”比率是有意义的。
47. 可以应用比率水平。薪资 150 000 美元是 75 000 美元的两倍，“两倍”比率是有意义的。
49. 数据是定量数据，是比率水平的测量值。数据是连续变量。时间有个绝对零点，而且可以是某取值范围当中的任意值。
51. 数据是定性数据，是称名水平的变量。不同的数字是表示不同人名的方式，既不能计算也不能计数。
53. 数据是定量数据，是等距水平的测量值。数据是离散的，因为只包括整数。年份是从人为的零点开始的（年份 0），不是自然意义的 0 开始点，年份之间的差值是有意义的，但比率是无意义的。
55. 数据是定性数据，是称名水平的变量。等级包括一个顺序，但是并不表示可以计数或者测量。

2.2 练习题答案

1. 因为这是测量系统中一个记录问题的结果，这是一个系统误差。这不是一个随机误差，因为在测量过程中这不是一个随机的结果，也不是一个本质上不可预测的事件。
3. 因为这个记录的身高有很多位小数，它是非常精确的，但是因为这个记录的身高并不接近于真正的身高，所以它不是很准确。

5. 无意义 7. 有意义
9. 影响结果的错误往往来自随机误差,但是有利于纳税人的不诚实报告会导致系统误差。
11. 因为大约一半的螺丝比 25 厘米长,一半的螺丝比 25 厘米短,这看起来是一个随机误差。
13. 当收入报告不正确时或者调查被试确实不知道他们的准确收入时会产生随机误差。当人们为了让自己显得成功而过高地报告他们的收入时会产生系统误差。
15. 当量具的读数不一致或者在读刻度以及记录数据时发生错误的时候会产生随机误差。当一个量具一直都读数过高或者过低时会产生系统误差。
17. 当使用的雷达枪不稳定或者警官不诚实地记录速度时,会产生随机误差。当使用的雷达枪没有校准而导致读数一致性地过高或过低时,会产生系统误差。
19. 计算发生错误会产生随机误差。当少报香烟的数量以逃避印花税时,会产生系统误差。
21. 绝对误差: 1 750 美元; 相对误差, 141% (使用误差值 1 750 美元和正确的账单 1 245 美元计算得到)。
23. 绝对误差: -2 英里/加仑; 相对误差: -8.3%。
25. (1) 这些误差是随机误差。
(2) 要使随机误差最小化最好的选择是取平均数。
(3) 可能会产生系统误差,例如,测量工具有问题或者定义房间“宽度”的问题。
27. 交通部门的测量结果是非常精确的,因为它所测得的重量 3 298.2 磅要比另外所测得的重量 3 250 磅更精细一些。制造商的测量是相对准确的,因为他所测量的重量 3 250 磅更接近于真值(测得的 3 250 磅的重量误差是 23 磅,交通部门所测得的 3 298.2 磅的重量误差是 25.2 磅)。
29. 健身房的数字秤更精确,但是它并不比健康秤更准确(假定你真实的体重如你所知道的)。
31. 给出的数字是非常精确的,但是它可能并不准确。我们今天也不能精确地测量人口总数,1860 年,这种不确定性会更高。

33. 给出的数字是非常精确的,但是它可能并不准确。任何一个国家的人口普查都有可能产生大量的误差。因为建立一个国家的人口普查是非常困难。在这一年中,中国的人口很可能会发生变化。
35. 一座建筑物的高度在一个合理的精确度水平上是非常容易准确测量的,比如精度为 1/10 英尺,但是给出的这个数字过于精确(它所提供的这个高度数字最小单位要比一个原子还小),这个说法是不太可信的。
37. 大学生的数量随着入学和辍学经常会发生改变,所以这个给出的数字一定是个估计值。这个给出的数字说明它只精确到百万,看起来像是一个非常好的估计值。所以它是可信的,当然如果我们要确定它是否真准确就需要更多信息。

2.3 练习题答案

1. 对于十年级的学生来说,现在的吸烟率是 18.3%,增长的百分比是 45%。这样,现在吸烟率是过去吸烟率的 1.45 倍,因此十年级学生过去的吸烟率是 12.6%。对于八年级学生来说,现在的吸烟率是 10.4%,增长的百分比是 44%。这样,现在吸烟率是过去吸烟率的 1.44 倍,因此八年级学生过去的吸烟率是 7.2%。
3. 不正确的陈述意味着误差幅度是 5% 的 1.2%, 因为 5% 的 1.2% 是 0.06%, 这个不正确的陈述意思是指误差幅度为 0.06% (从 4.94% 提高到 5.06%); 实际上的误差幅度是以 5% 为中心上下浮动 1.2 个百分点 (3.8% ~ 6.2%)。
5. 无意义 7. 有意义
9. (1) $2/5$; 0.4; 40% (2) $150/100$, 或者 $3/2$; 1.50; 150%
(3) $1/4$; 0.25; 25% (4) $30/100$, 或者 $3/10$; 0.30; 30%
11. (1) 93% (2) 44% (3) 41%
(4) 81%
13. -35% (从 1990 年开始下降了 35%)。
15. 115% (从 1980 年开始增加了 115%)。
17. 61%。《华尔街日报》比《纽约时报》的发行量高出 61%。
19. -19%。俄亥俄州芝加哥飞机场要比亚特兰

- 大哈兹菲尔德飞机场接待的旅客要少 19%。
21. 66 23. 834
25. 140%。大象的重量相当于一辆汽车重量的基础上再加 40%。
27. 80%。蒙大拿州的人口是新罕布什尔的人口减去 20%。
29. 是的，两者的意思是不一样的。百分点对应的范围是 86% ~ 92%，3% 的误差幅度指的是 89% 的 3%，也就是 $0.03 \times 0.89 = 0.0267$ ，这不是我们所指的意思。
31. -15.5 个百分点；-22.7%。
33. 22 个百分点；56.4%。

2.4 练习题答案

1. 指数是一个比率，没有单位。呈现的数字是 2007 年实际的石油价格，不是一个指数。
3. 是的。消费价格指数是基于食品、服务和住房等的价格，所以这些方面价格的增长最终会导致消费价格指数的增长。
5. 573.2
7. 1.12 美元
9. 93.3 100, 181.7, 383.3 386.2, 496.8, 740.4
11. 21.69 美元
13. 2004 年的学费是 1980 年学费的 466.4%，2004 年的 CPI 是 1980 年 CPI 的 229.2%，所以学费要比食品、服务和住房等的典型价格增长得多。
15. 2004 年的房屋价格是 1990 年房屋价格的 206.5%，2004 年的 CPI 是 1990 年 CPI 的 144.5%，所以房屋价格要比食品、服务和住房等的典型价格增长得多。
17. 582 000 美元；180 000 美元。
19. 1 591 667 美元；1 491 667 美元

第 2 章复习题答案

1. (1) 706 (2) 离散变量；一个人不可能部分存活。
- (3) 23.89% (4) 64 (5) 比率变量
- (6) 称名变量
2. (1) 860 (2) 37% (3) 顺序变量，因为存在一个顺序关系。
- (4) 因为调查所采用的被试是自愿参与的，样本是一个自选择样本，它不大可能准确地反映总体的态度。

3. 2004 年健康护理的花费是 1973 年消费价格指数的 2 150%，但 2004 年的消费价格指数是 1973 年消费价格指数的 325.5%，所以健康护理花费的增长要远大于食品、服务和住房的一般价格指数。
4. (1) 2.78 美元 (2) 5.77 美元 (3) 最低工资的增长赶不上通货膨胀的增长，所以 2006 年的实际最低工资 5.15 美元相当于 1996 年 4.14 美元的购买力。1996 年工人所得到的最低工资 4.75 美元要比 2006 年所得到的最低工资 5.15 美元收入更高些。

第 2 章小测验答案

1. 称名变量 2. 连续变量 3. 比率变量
4. -10cm 5. -18.1% 6. 52% 7. 52
8. 178 厘米，179.18 厘米 9. 123.7
10. 52 972 美元

3.1 练习题答案

1. 一个次数分布表由二列组成，一列是分类，另外一列是次数。分类是指一个变量有不同的取值（例如，不同的眼睛颜色或者喜欢的冰激凌口味）。次数是指每一类中数据点（计数）的个数。
3. 2, 11, 25, 37, 40 5. 无意义 7. 无意义
- 9.

成绩	次数	相对次数	累加次数
A	4	16.7%	4
B	7	29.2%	11
C	8	33.3%	19
D	3	12.5%	22
E	2	8.3%	24
合计	24	100%	24

11.

重量 (磅)	次数	相对次数	累加次数
0.790 0 ~ 0.794 9	1	1/36	1
0.795 0 ~ 0.799 9	0	0	1
0.800 0 ~ 0.804 9	1	1/36	2
0.805 0 ~ 0.809 9	3	3/36	5
0.810 0 ~ 0.814 9	4	4/36	9
0.815 0 ~ 0.819 9	17	17/36	26
0.820 0 ~ 0.824 9	6	6/36	32
0.825 0 ~ 0.829 9	4	4/36	36
合计	36	100%	36

13.

年龄 (岁)	获奖演员的数量
20 ~ 29	0
30 ~ 39	12
40 ~ 49	13
50 ~ 59	5
60 ~ 69	3
70 ~ 79	1

15.

分类	次数	相对次数
A	13	24%
B	9	18%
C	12	24%
D	11	22%
E	6	12%
合计	50	100%

17. (1) 200 (2) 142 (3) 16%
(4) 0.135, 0.155, 0.210, 0.200, 0.140, 0.160
(5) 27, 58, 100, 140, 168, 200

19. (1)

等级	次数	相对次数
0 ~ 2	20	38.5%
3 ~ 5	14	26.9%
6 ~ 8	15	28.8%
9 ~ 11	2	3.8%
12 ~ 14	1	1.9%
合计	52	100%

(2)

单词指向率	次数	相对次数
0 ~ 2	33	63.5%
3 ~ 5	19	36.5%
6 ~ 8	0	0%
9 ~ 11	0	0%
12 ~ 14	0	0%
合计	52	100%

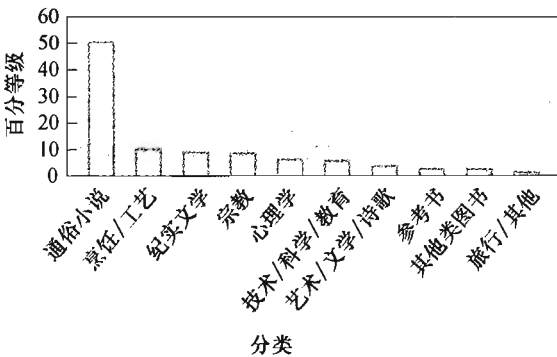
- (3) 德沃夏克键盘看起来好像更有效, 因为它低单词指向率的次数更多, 而高单词指向率的次数更少。

3.2 练习题答案

1. 数据的分布是指在所有可能的数据取值范围中数据的散落情况。直方图提供了一种图示化的分布形态。与数据值的陈列比较来说, 直方图帮助我们更容易理解分布的形态。
3. 数据必须是按照某种时间顺序来排列的, 时间序列图可以揭示随着时间延续呈现出来的事物发展的趋势。
5. 无意义 7. 有意义
9. 一个直方图能够很好地呈现不同分数类别的次数分布。
11. 一个帕累托图或者一个饼图可以很好地呈现出人们的死亡原因这样的数据。但是帕累托图在呈现人们最常见的死亡原因时会表现更直观准确。
13. (1)

阅读书籍种类	相对次数
通俗小说	50.4%
烹调/工艺	10.2%
散文	8.9%
宗教	8.6%
心理学/保健	6.3%
技术/科学/教育	5.6%
艺术/文学/诗	3.7%
参考资料	2.6%
其他类别	2.5%
旅行/地区	1.3%

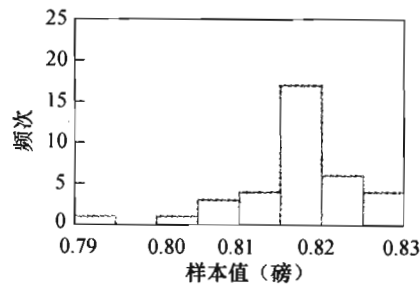
(2)



人们阅读的图书

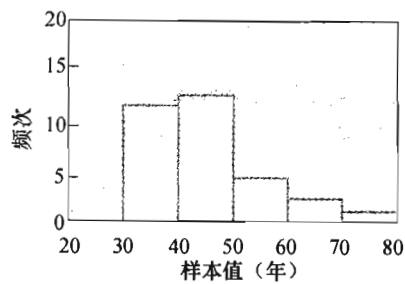
- (3) 在帕累托图上更容易看出哪种类型最受欢迎。

15.



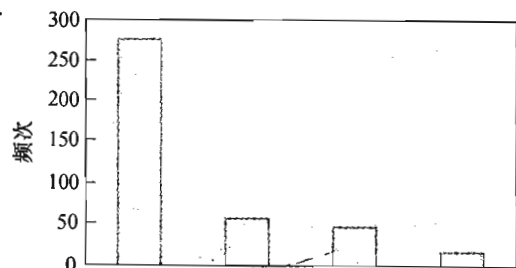
可乐的重量

17.



奥斯卡获奖演员

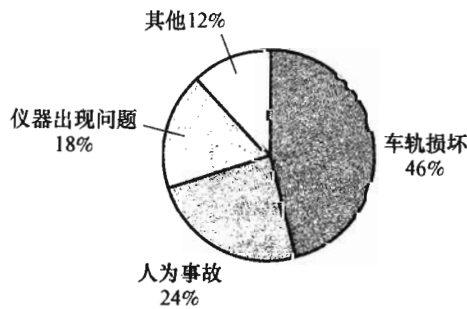
19.



网络 和 广播电视 招聘广告 猎头公司 群发邮件

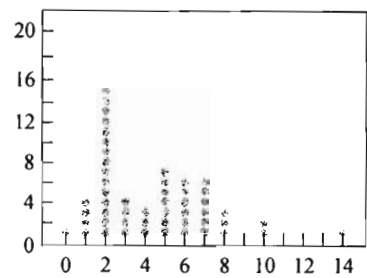
人们如何获得工作

21.



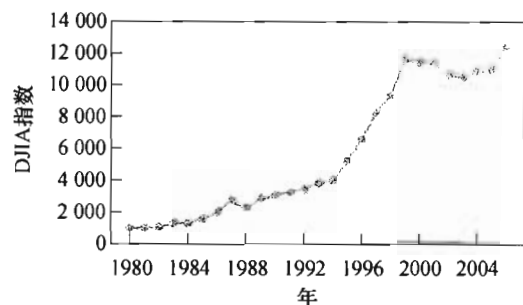
列车出轨原因

23.



打字机式键盘

25.



有股票价格上升趋势的年份是好的投资时机。

27.

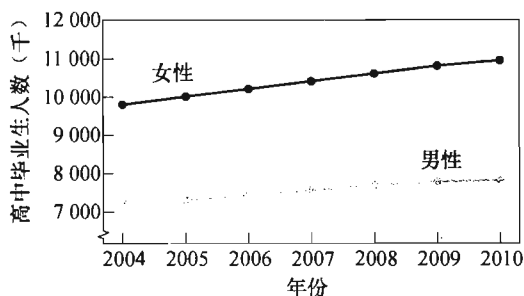
茎	叶
6	7
7	25
8	5 899
9	09
10	0

列的宽度与直方图的条形长度是相同的，较长的数据对应于较高的次数。

3.3 练习题答案

- 石油桶的使用是不合适的，因为数据不是三维的。使用一个简单的条图就能够很好地说明这些数据。
- 地理数据是对应于不同地理区域的原始数据。呈现地理数据的两个例子是彩色编码图和等高图。
- 无意义 7. 无意义
- (1) 女性人数持续超过男性人数。男性和女性的人数都随着时间而持续增长。

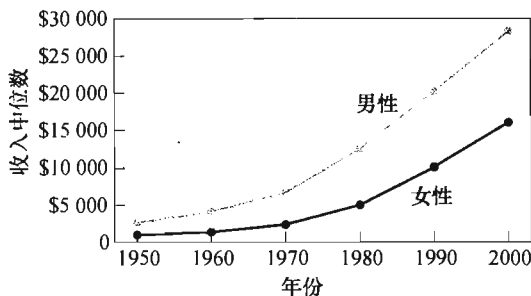
(2)



不同年份的高中毕业生男女生人数线图

11. (1) 男性收入的中位数持续高于女性，男性和女性的收入都随时间而持续稳定。从左到右比较条形的高度，我们可以看到男性与女性收入的比率呈现出下降的趋势。最左边的条形显示出男性的收入超过女性收入的两倍，但是在最右边的条形看起来男性的收入达不到女性收入的两倍。

(2) 线图更容易看出随着时间发展的趋势。但是，由于线图不是聚集在一起的，所以图中的信息更容易理解和解释。

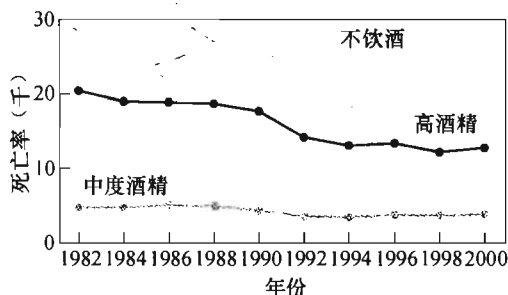


不同年代的男性和女性的收入中位数线图

13. 这个堆栈图显示了总体预算的百分数，这样，总的高度一定是 100%，我们不能直接确定政府所花费的实际钱数。然而，从图中我们更容易看到政府如何花钱的长期发展趋势。例如，联邦预算中用于国防开支的比率持续下降，从 1960 年占到联邦预算的 50% 下降到 2005 年占联邦预算的 13%。在同一时期，联邦预算用于支付个体费用（比如社会保障、医疗保险和社会福利）比率却在急剧增加，从占联邦预算的 30% 上升到 70%。花费在纯利息上的费用比率也在成倍增加。
15. 看起来在南方诸州和西部诸州，较高的黑色素瘤的死亡率是一个一般趋势。这可能是由

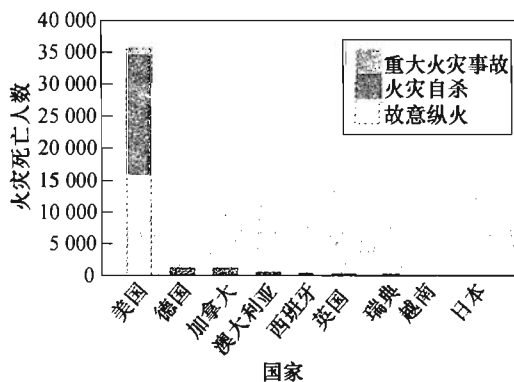
于在这些州的人们花更多时间在户外，暴露在太阳光下的时间更多。作为一个研究者，你可以特别关注具有背离一般趋势的某一个特定的地区。例如，在东部华盛顿州的一个县呈现出了非常高的黑色素瘤死亡率。首先，你想核实一下这个数据点是否准确，是否存在某种类型的误差。如果数据是准确的，你可能想寻找为什么这个地方有这么高的黑色素瘤死亡率。

17.



涉及酒精的交通死亡事故

19.



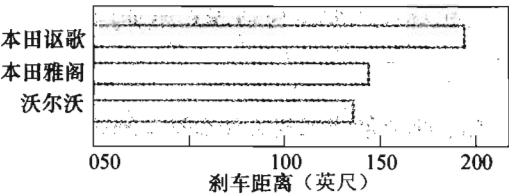
不同国家因火灾而导致的死亡人数

3.4 练习题答案

- 纵坐标轴的刻度值不是从零开始时，就会歪曲读者的知觉，导致读者只关注图形的一部分而不是关注整体图形。只有整体图形才能对所有数据形成更好的直觉。这种改变会导致夸大。
- 替代相同的等距刻度，在某些数字（最常见的是 10）以次方的形式发生变化时，使用一个指数刻度。例如，如果用 10 的幂的话，等距的刻度代表的是 1, 10, 100 等。依次类推，指数方式在呈现变化范围巨大的数据时

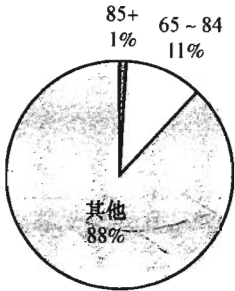
是非常有帮助的。

5. 因为横坐标不是开始于零点，所以导致图形的误解。本田讴歌的刹车距离大于沃尔沃42%。

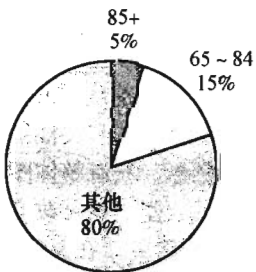


7. 每个国家石油使用的数量看起来是和图中石油桶的容量是相关的，但实际上它只是和石油桶的高度相关，美国的石油消费是日本的4倍，而不是石油桶的容量所暗示的64倍。

9. (1) 由于饼图的三维外观形式，楔形的大小没办法比较百分比，相反，整个图呈现了如果从某个角度看整个的饼图楔形所呈现的外观。这种歪曲使看清真实的关系变得非常困难。
- (2) 由于饼图的三维外观形式，楔形的大小没办法比较百分比，这种歪曲使看清真实的关系变得非常困难。
- (3)



1990 年不同年龄的人口分布

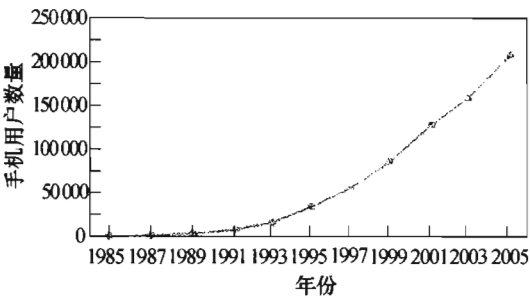


2050 年不同年龄的人口分布

(4) 2050 年与 1990 年相比，美国人口中老年人数越来越多，青年人越来越少。值

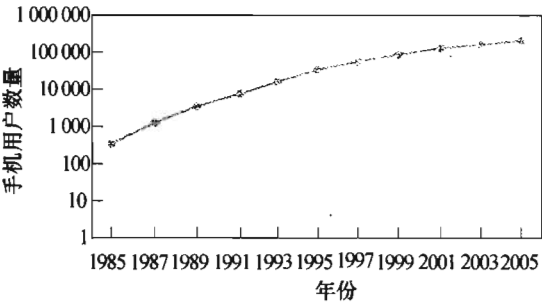
得注意的是，因为人口总体是增长的，所以在 2050 年，所有年龄组的预计人口要比 1990 年多。

11. (1)



手机用户的时间序列图

(2)



手机用户的指数图

- (3) 第 1 小题中的图形给我们手机用户变化全部的真实情况一个最好的直观展示。第 2 小题中的图形则让我们更容易看清最早年份的变化情况，指数刻度使图形更容易适合全部的数据。
13. 1990 年的消费价格指数变化百分比要小于 1998 年或者 2002 年的变化，后面两个年份的变化看起来是相同的。实际上，1991 年的价格要比 1990 年的价格高出 4.2%。价格每年都在增长，但是这种增长在结尾的年份比起在开始年份要低一些。
15. 自从 1955 年以来，实际上的最低工资在没有经过调整前，保持不变或者说稳定增长，但是自从 1980 年以来，购买力已经下降，2006 年的购买力小于 1955 年的购买力。

第 3 章复习题答案

1. (1) 次数分别是 5，12，12，5，0，2。

重量	次数
0.8130~0.8179	5
0.8180~0.8229	12
0.8230~0.8279	12
0.8280~0.8329	5
0.8330~0.8379	0
0.8380~0.8429	2

(2) 次数分别是 5, 5, 16, 8, 2。

重量	次数
0.7740~0.7779	5
0.7780~0.7819	5
0.7820~0.7859	16
0.7860~0.7899	8
0.7900~0.7939	2

(3) 一罐普通百事可乐的重量一般要大于一罐无糖百事可乐的重量, 因为无糖百事可乐里面不包括糖的重量。

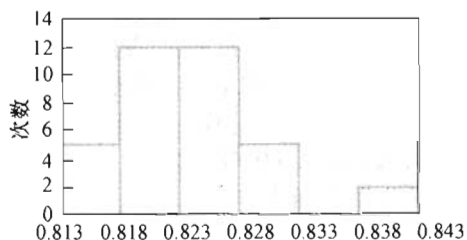
2. (1) 相对次数是 0.139, 0.333, 0.333, 0.139, 0, 0.056。

重量	相对次数
0.8130~0.8179	0.139
0.8180~0.8229	0.333
0.8230~0.8279	0.333
0.8280~0.8329	0.139
0.8330~0.8379	0
0.8380~0.8429	0.056

(2) 累加次数是 5, 17, 29, 34, 34, 36。

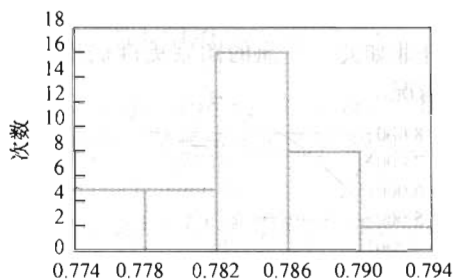
重量	累加次数
0.8130~0.8179	5
0.8180~0.8229	17
0.8230~0.8279	29
0.8280~0.8329	34
0.8330~0.8379	34
0.8380~0.8429	36

3. (1)



普通百事可乐的重量/磅

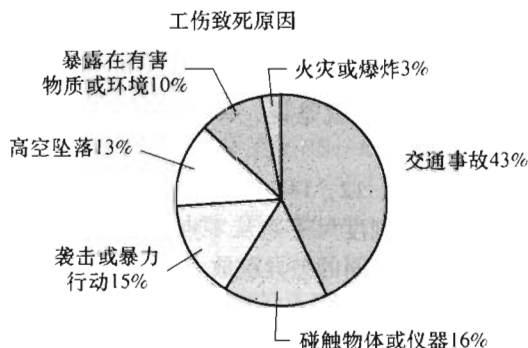
(2)



无糖百事可乐的重量/磅

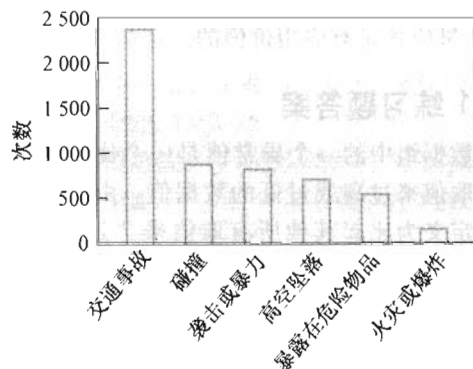
(3) 两个直方图的形态没有明显的差别, 说明可乐重量的分布是相同的, 然而, 取值范围是明显不同的, 说明一罐普通百事可乐的重量要明显大于一罐无糖百事可乐的重量。

4.



工作时死亡原因的饼图

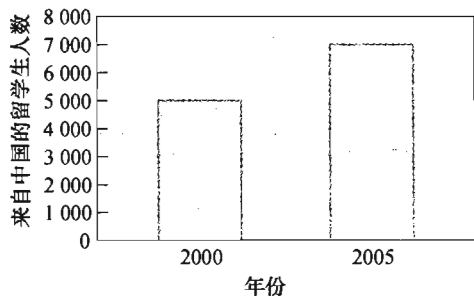
5. 如果关注最严重的死亡原因, 帕累托图是非常有效的。帕累托图在呈现不同死亡原因的相对重要性上更直观。



工作时死亡原因的帕累托图

6. 图形的纵坐标不是从零点开始, 所以两个次

数分布的差异被放大了。图形看起来 2005 年接收留学生数是 2000 年的两倍多，但实际情况并非如此。下面的图则更准确地反映了实际情况。



2000 年和 2005 年去美国留学的中国学生数量

第 3 章小测验答案

1. 条形图 2. 复合条形图 3. 直方图
4. 在 50 和 59 之间有 7 个值。
5. 在所有取值中，取值在 50 ~ 59 之间的所占比率是 0.2。也就是说，20% 的取值在 50 ~ 59 之间。
6. 5, 10, 12, 12, 14
7. 纵坐标的刻度尺不是从零点开始，使得两个次数分布之间的差异被放大了。
8. 最大的数据值是 9，最多的次数是 3（数据值 6）。
9. 1982 年美国酒后驾驶机动车致死人数是 25 000，2000 年美国酒后驾驶机动车致死人数是 16 652。
10. 与价格或者收入不同，每年的汽油消耗量不受通货膨胀的影响，所以实际没有调整的数量应该是有应用价值的。

4.1 练习题答案

1. 数据组中的一个异常值是一个比起其他所有取值来过高或过低的数据值。由于异常值被定义为比起其他所有取值来“过高”、“过低”，并没有对它们进行明确和客观的定义。确定一些异常值需要一些判断。
3. 不是一个好的估计值。结果可能会被一些通勤时间较少的小州所歪曲。由于不同州的人口数不同，平均数应该给以加权。
5. 无意义 7. 有意义
9. 收入数据经常会得到一些高端的异常值，所

以中位数是适合的。

11. 分布可能是相当对称的，所以平均数或者中位数都可以应用。
13. 平均数：58.3 秒；中位数：55.5 秒；众数：49 秒。
15. 平均数：0.188；中位数：0.165；众数：0.16。
17. 平均数：0.807mm；中位数：0.840mm；众数：0.84mm。
19. 平均数：0.919 4g；中位数：0.920 0g；众数：没有。
21. (1) 平均数：157 586；中位数：104 100。
(2) 阿拉斯加州是一个高端的异常值，如果没有阿拉斯加州的数据，平均数是 81 317，中位数是 78 650。
(3) 康涅狄格州是一个低端的异常值，如果没有康涅狄格州的数据（但是有阿拉斯加州的数据），平均数是 182 933 中位数是 109 050。
23. (1) 73.75 分 (2) 80 分 (3) 不可能，5 次小测验以后平均分达到 80 分需要总共 400 分，你不可能在第 5 次测验得分超过 100 分。
25. 如果在下一次小测验中得到 90 分，新的平均分就是 81.4。你能得到的最高分数是 100 分，这样会使得你分数的新平均数是 82.9 分，能得到的最坏分数是 0 分，这样会使得你分数的新平均数是 68.6。
27. 你学生的平均分数是 74.7。中位数是 70。这些都是平均水平的含义，如果使用平均数，你学生在平均水平以上，如果使用中位数，你学生在平均水平以下。
29. 0.39 磅。
31. 每个学生参加登记人数是 20 人的三门课程和登记人数是 100 人的一门课程，所以每个学生的班容量是 $160/4 = 40$ 。有三门课程人数是 100 名学生，有 45 门课程人数是 20 名学生，所有登记的学生人数是 1 200 人，一共有 48 门课程。这样，每门课程的平均登记人数就是 $1\,200/48 = 25$ 。
33. 0.417，它是每击打一次的成功率。
35. 不可以这样计算，除非检验者在两家农场检验相同数目的鸡蛋。
37. 投票的最后结果是反对这个方案。600 票反

对对 400 票赞成。

39. 平均数是 1.9；中位数是 2；众数是 1。众数是 1 说明皮光滑且是黄色的性状发生的次数超过其他性状，但是平均数和中数没有意义，因为数据是称名水平的变量。

4.2 练习题答案

1. 对称性是指分布的形态是这样的，左半边分布是右半边分布的镜像。
3. 因为上统计课的学生必须具备一些先决条件并且正在学习一门大学课程，他们智商分数的变异性可能小于随机选择的成年人智商分数的变异性。统计课学生较低的智商分数变异性这个结果，比起随机选择的成年人的智商分数，在图中的分布就会呈现得比较狭窄。
5. 无意义 7. 无意义
9. 两个众数，左偏态，较大的变异性。
11. 单峰分布，接近对称分布，中等水平变异性。
13. (1) 右偏态 (2) 150 (300 的一半)
(3) 不能，它依赖于分布。
15. (1) 选择众数 0 美元
(2) 右偏态，因为许多学生没有收入或者有很少的一些收入。
17. (1) 一个众数 (2) 接近对称分布
19. (1) 2 个众数
(2) 右偏态分布（假定每组中滑冰者的人数相等）
21. (1) 一个众数 (2) 右偏态分布
23. (1) 一个众数 (2) 右偏态分布
25. (1) 一个众数 (2) 对称分布
27. (1) 一个众数 (2) 右偏态分布
29. (1) 一个众数 (2) 右偏态分布

4.3 练习题答案

1. 标准差是一个偏离平均数多少的量化指标。
3. 陈述是不正确的，因为它定义标准差是一个依赖于最大值和最小值的值，但是标准差的计算会用到每个数据值。
5. 无意义 7. 有意义
9. 全距：26.0 秒，标准差：9.5 秒。
11. 全距：0.170，标准差：0.057。
13. 全距：0.280mm，标准差：0.094mm。

15. 全距：0.1160g，标准差：0.0336g

17. 《热铁皮屋顶上的猫》：全距 = 10.0，标准差 = 2.6。

《戴帽子的猫》：全距 = 3.0，标准差 = 0.9。
在词的长度上《戴帽子的猫》所用词的变异性较小。

19. (实际最高温) - (1 天前预测的最高温)：全距 = 11.0，标准差 = 2.6
(实际最高温) - (5 天前预测的最高温)：全距 = 15.0，标准差 = 4.5
看起来 5 天前预测的最高温的误差的变异性要更大一些。
21. (1) 第 5 百分等级
(2) 第 69 百分等级
(3) 第 48 百分等级
23. 只回答第一组数据的结果：
(1) 直方图中只有 7 个数据值得 9，没有其他数据值。
(2) 最小值 = 9；第一四分位数 = 9；中位数 = 9；第三四分位数 = 9；最大值 = 9。
(3) 0。
25. (1) 教师停车场：平均数 = 2；中位数 = 2；全距 = 4。学生停车场：平均数 = 6.18；中位数 = 6；全距 = 9。
(2) 教师停车场：最小值 = 0；第一四分位数 = 1；中位数 = 2；第三四分位数 = 3；最大值 = 4。学生停车场：最小值 = 1；第一四分位数 = 4；中位数 = 6；第三四分位数 = 9；最大值 = 10。
(3) 教师停车场：标准差 = 1.18。学生停车场：标准差 = 3.03。
(4) 教师停车场：全距/4 = 1.0。学生停车场：全距/4 = 2.25。
27. (1) 前 7 位总统：平均数 = 58.3；中位数 = 57；全距 = 4。后 7 位总统：平均数 = 57.6；中位数 = 56；全距 = 23。
(2) 前 7 位总统：最小值 = 57；第一四分位数 = 57；中位数 = 57；第三四分位数 = 61；最大值 = 61。后 7 位总统：最小值 = 46；第一四分位数 = 52；中位数 = 56；第三四分位数 = 64；最大值 = 69。
(3) 前 7 位总统：标准差 = 1.9。后 7 位总统：标准差 = 7.7。

(4) 前7位总统：全距/4 = 1。后7位总统：
全距/4 = 5.75。

29. 从第一家比萨店订货（标准差是3分钟）。

31. 对于长期投资来说，高平均增长率值的长期投资，对于短期投资来说，较小的标准差风险会更少。

4.4 练习题答案

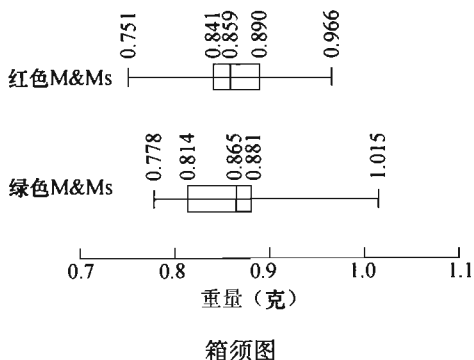
1. 假阳性是指检测结果说明某个人使用了药物，而事实上这个人并没有使用药物；假阴性是指检测结果说明某个人没有使用药物，而事实上这个人却使用了药物。
3. 说谎者的比率通常非常小，这意味着说真话的人的比率是非常大的。即使说真话的人们只有很小的百分比得出不正确的结论，这也可能涉及相当多的人数。
5. 有意义 7. 无意义
9. 约什、约什、尤德。
11. (1) 新泽西州在两个族裔中都有较高得分，内布拉斯加州在所有族裔中有较高的总体平均分。(2) 略。(3) 略。
13. (1) 白种人：0.18%；非白种人：0.54%；总体：0.19%。
(2) 白种人：0.16%；非白种人：0.34%；总体：0.23%。
(3) 纽约白种人和非白种人肺结核死亡人数的比率都高于弗吉尼亚州里士满市，但是弗吉尼亚州里士满市总体的肺结核死亡人数的比率却高于纽约。
15. (1) 就主客场分别来看，斯佩尔曼学院在主场（34.5%对32.1%）和客场（75.0%对73.3%）的成绩都要高于莫哈斯学院。
(2) 就全场来看，莫哈斯学院比斯佩尔曼学院平均成绩更高。
(3) 莫哈斯学院的篮球队水平要高于斯佩尔曼学院篮球队，因为篮球比赛通常是整体计分的。
17. (1) 略。
(2) 216名员工被指控为说谎。在其中，只有18名是真正说谎的，而另外198名说的是真话。
(3) 按照测谎仪的检测结果，1784名员工被认为是诚实的。在其中，1782名员工说的确实是真话，另外两名实际说的是谎话。被检测为说真话的人中99.9%确实说的是真话。
19. 分别看白领和蓝领职位，被雇用女性的百分比都要高于男性，这说明了对于女性的雇用偏好。就总体来看，200名女性中的20%（40人）被雇用从事白领职位，100名女性中的85%（85人）被雇用从事蓝领职位。这样， $200 + 100 = 300$ 中的 $40 + 85 = 125$ 名女性被雇用，百分数为41.7%。这总体来看，200名男性中的15%（30人）被雇用从事白领职位，400名女性中的75%（300人）被雇用从事蓝领职位。这样， $400 + 100 = 500$ 中的 $30 + 300 = 330$ 名男性被雇用，百分数为55.0%。
21. (1) 在一般人群中，样本中的20000人被感染的人数是 $57 + 3 = 60$ 。感染率是 $60/20000 = 0.3\%$ 。在高风险人群中，样本中的5000人被感染的人数是 $475 + 25 = 500$ ，感染率是 $500/5000 = 10.0\%$ 。
(2) 在高风险人群中，检测阳性的500人中确实感染HIV的有475名，占95%。 $475 + 225 = 700$ 人检测出是阳性，其中确实感染HIV的有475名，百分比是 $475/700 = 67.9\%$ 。这两个百分数有所不同是因为它们测量的事物是不同的。700名检测阳性的人包括225名是假阳性。检验能够正确地确定感染HIV人们中的95%，同时，它也会对那些实际上并没有感染HIV的人们做出不正确的判断。
(3) 在高风险人群中，检验阳性的罹患者大约有68%的机会确实罹患这种疾病。这是高风险人群中感染率（10%）的7倍，这样，这个检验对于诊断是否感染了HIV还是非常有用的。
(4) 在一般人群中，检测阳性的60人中确实感染HIV的是57名，占95%。 $57 + 997 = 1054$ 人检测出是阳性，其中确实感染HIV的有57名，百分比是 $57/1054 = 5.4\%$ 。这两个百分数有所不同是因为它们测量的事物是不同的。

1 054 名检测阳性的人包括 997 名是假阳性。检验能够正确地确定感染 HIV 人中的 95%，同时，它也会对那些实际上并没有感染 HIV 的人做出不正确的判断。这样，检验阳性的人中只有 5.4% 的人确实感染了 HIV。

- (5) 在一般人群中，检验阳性的罹患者只有 5.4% 的机会确实罹患这种疾病。这是一般人群中感染率 (0.3%) 的 18 倍，这样，这个检验对于诊断是否感染了 HIV 还是非常有用的。

第 4 章复习题答案

- (1) 红色 M&Ms 巧克力豆平均数: 0.863 5 克, 绿色 M&Ms 巧克力豆平均数: 0.863 5 克; 红色 M&Ms 巧克力豆中位数: 0.859 0 克, 绿色 M&Ms 巧克力豆中位数: 0.865 0 克。
- (2) 红色 M&Ms 巧克力豆全距: 0.215 0 克, 绿色 M&Ms 巧克力豆全距: 0.237 0 克。红色 M&Ms 巧克力豆标准差: 0.057 6 克, 绿色 M&Ms 巧克力豆标准差: 0.057 0 克。
- (3) 红色 M&Ms 巧克力豆五数概括法描述为最小值: 0.751 0 克, 第一四分位数: 0.825 0 克, 中位数: 0.859 0 克, 第三四分位数: 0.897 5 克, 最大值: 0.996 0 克。
绿色 M&Ms 巧克力豆五数概括法描述为最小值: 0.778 0 克, 第一四分位数: 0.814 0 克, 中位数: 0.865 0 克, 第三四分位数: 0.881 0 克, 最大值: 1.015 0 克。



- (4) 红色 M&Ms 巧克力豆标准差的估计值是 $0.215\ 0/4 = 0.0537\ 5$ 克, 实际标准差是 0.057 6 克。绿色 M&Ms 巧克力豆标准

差的估计值是 $0.237\ 0/4 = 0.0592\ 5$ 克, 实际标准差是 0.057 0 克。使用全距的拇指法则对标准差进行估计在两种情况下都是比较合理的。

- (5) 两个数据组结果的比较说明它们之间差异是不明显的, 看起来红色 M&Ms 巧克力豆和绿色 M&Ms 巧克力豆在重量上基本是相同的。
- (1) 第 31 百分等级 (2) 0.865 克
 - (1) 0 (2) 虽然两种汽车电池的平均使用寿命相同, 但是标准差更小的电池更好一些, 因为它们的使用寿命都接近平均数, 这样会使驾驶员很少产生超乎预料的抛锚现象。
 - (3) 异常值会使平均数偏高或偏低, 这取决于这个异常值是远高于平均数还是远低于平均数。 (4) 这个异常值对中位数没有什么影响。 (5) 异常值会使全距变大。
 - (6) 异常值会使标准差增大。

第 4 章小测验答案

- 平均数 2. 标准差
- 不是标准差的真值。它只是这个真值最接近的估计值。
- 这 3 个陈述都可能是正确的。
- 其中的第二个、第三个、第四个陈述是正确的, 第一个和第五个陈述是错误的。
- 最小值: 30; 最大值: 70。 7.2 8.0 9.7.0
- 五数概括法: 最小值, 第一四分位数, 中位数 (第二四分位数), 第三四分位数, 最大值。

5. 1 练习答案

- 术语“正态”在统计学上有特定的含义。它指的是一种特定的分布类型, 呈钟形分布。
- 不是正态分布。10 个不同的数字出现的可能性是相等的, 因此分布的形态是平直的, 而不是钟形分布。
- 无意义 7. 无意义
- b 分布不是正态分布, c 分布的标准差比较大。
- 符合正态分布。对于制造出的一种产品来说, 比如 CD, 通常都符合正态分布。在平均重量上下波动的产品变化情况是相同的, 因此分布是个单峰对称分布。

13. 非正态分布。1 枚均匀的骰子抛掷的结果 1, 2, 3, 4, 5, 6 点出现的机会是相同的, 所以是个均匀分布, 而不是正态分布。分布的图形也倾向于是一条直线, 而不是钟形分布。
15. 正态分布。这样的生理测量通常都符合正态分布。很少一部分男性具有较高的握力, 另外很少一部分男士具有较弱的握力。倾向于在平均数处形成分布的峰值。
17. 非正态分布。等候时间倾向于是均匀分布。
19. 非正态分布。电影时长有最小时长 (0), 但没有最大时长。
21. 接近正态分布。每个硬币重量的分布, 会围绕着平均数上下波动。
23. (1) 1 (2) 0.20 (3) 0.80
(4) 0.35 (5) 0.45
25. (1) 115 (2) 15%
(3) 45 (4) 15%

5.2 练习题答案

1. 0
3. 不能够得出这个结论。68 - 95 - 99.7 法则适用于正态分布。抛掷 1 枚骰子结果的分布是均匀分布, 而不是正态分布。
5. 无意义 7. 有意义
9. (1) 50% (2) 0.84 (3) 97.5%
(4) 16% (5) 0.025%
(6) 16% (7) 2.5% (8) 0.84
(9) 68% (10) 81.55
11. (1) 68% (2) 95%
(3) 99.7% (4) 47.5%
13. 50% 15. 15.87% 17. 2.28% 19. 1.39%
21. 68.26% 23. 80.64% 25. 50.00%
27. 84.13% 29. 99.38% 31. 0.13%
33. 68.26% 35. 89.25%
37. 在所有情况下, 5% 的硬币被拒绝。
每一种的硬币自动售货机接受的重量范围分别是:

硬币 (美分)	自动售货机接受的重量范围 (克)
1	2.44 ~ 2.56
5	4.88 ~ 5.12
10	2.208 ~ 2.328
25	5.530 ~ 5.810
50	11.060 ~ 11.620

39. (1) 6.68% (2) 48.01% (3) 36.54%
41. (1) 0.47% (2) 接近 4% (3) 30.055 ~ 30.745 英寸汞柱; (4) 这些读数的平均数就是大气压真值的最好估计值, 也就是 30.4 英寸汞柱。
43. 接近 96.33%

5.3 练习题答案

1. 不能够使用中心极限定理分析样本中的平均身高。因为样本是一个便利样本, 会导致一些偏差。不能当做随机样本对其进行预期。
3. 不符合正态分布。中心极限定理说的是, 对大样本容量来说, 样本平均数的分布将接近于正态分布。但是样本容量只有两个显然是不够的, 除非原始总体自身是正态分布。
5. (1) 平均数是 100, 标准差是 2。(2) 平均数是 100, 标准差是 1.6。(3) 随着样本容量的增大 (比如第 2 小题), 样本平均数越来越接近, 因此它们之间的变异性越来越小, 结果就是标准差越来越小。
7. (1) 平均数是 6.5, 标准差是 0.384。
(2) 平均数是 6.5, 标准差是 0.345。
(3) 随着样本容量的增大 (比如第二小题), 样本平均数越来越接近, 因此它们之间的变异性越来越小, 结果就是标准差越来越小。
9. 40%, 4%
11. 29.6%, 99.96%
13. (1) 0.35% (2) 不是的。看起来平均数大于 12.00 盎司, 但并不是欺骗了消费者。因为 1 听装满后的可乐容量大于标准而不是低于标准。
15. (1) 57.93% 的机会 (2) 97.72% 的机会
(3) 虽然 100 个随机挑选的人头围平均数非常有可能小于 6.2 英寸, 但是仍然会有许多男士不能使用这个头盔, 因为他们的头围大于 6.2 英寸。基于第 1 小题的结果, 这种大小的头盔适合 42% 的男性。
17. (1) 52.91% 的机会
(2) 接近于 73% 的机会
(3) 第 1 小题的结果更加适合女性飞行员, 因为弹射座椅要适合女性个人, 而不是适合女性的群体。

19. (1) 16 枚 (280 枚中的 5.74%)。
- (2) 标准分数超过了表 5-1 中的范围, 但表 5-1 显示出平均数落在 5.550 ~ 5.790g 是非常不可能的 (小于 0.04% 的机会)。
- (3) 第 1 小题的结果更有关系。因为自动售货机拒绝硬币会导致销售量减少或造成较少的收益。
21. (1) 样本平均数分布的标准差是 $\sigma/\sqrt{n} = \sigma/10$ 。
- (2) 用平均数分布的标准差接近于 $\sigma/\sqrt{n} = \sigma/\sqrt{1000} = \sigma/32$, 大约是第 1 小题中标准差的 1/3。
- (3) 随着样本容量增加, 样本平均数分布的标准差在减小。

第 5 章复习题答案

1. (1) 因为 38 个结果出现的可能性是完全相等的, 所以分布是均匀分布, 而不是正态分布。
- (2) 同一物种体重的总体, 比如成年金毛猎犬的体重, 是典型的正态分布。
- (3) 渐近正态分布, 人类的生理和心理特征倾向于正态分布 (然而, 这些分布有些轻微的偏态, 因为反应时不可能是负数, 但是可能有少数人会有非常高的反应时)。
2. (1) 95%
- (2) 99.7%
- (3) 是很少出现, 因为这样一个高分发生的机会大约是 0.3%。
3. (1) 第 90 百分等级 (2) 1.29 (3) 身体温度达到 99.00°F 并不是非常罕见, 因为这个数据值在平均数两个标准差以内。
- (4) 0.0065 (5) 这个身体温度值是非常罕见的, 它在平均数两个标准差以外。
- (6) 99.22 (7) 97.18 (8) 小于 0.01%
- (9) 样本平均数在平均数 6.6 个标准差以下, 抽出这样一个样本的机会是非常之小的, 假定的平均数 (98.60) 可能是不正确的。
5. 2 6. -1 7. 50% 8. 2.28% 9. 2.28%
10. 第 3 小题中的分布接近正态分布。

6.1 练习题答案

1. 这个声称是无效的, 因为在自然情况下, 在 20 个刚出生的婴儿中得到 11 个女孩儿是非常有可能的。
3. 不一定。在 0.05 水平上具有统计显著性, 是指这个结果偶然发生的概率小于 0.05, 但是偶然发生的概率小于 0.05 并不意味着它偶然发生的概率一定就小于 0.01。
5. 无意义 7. 有意义
9. 不具有统计显著性。这个结果非常接近于 250 次试验所预期的结果, 这样的结果是很容易随机产生的。
11. 具有统计显著性。在可能的结果是 6 种的情况下, 我们预期 60 个试验结果中 10 次的结果会是点数 3, 而实际上没有得到一个 3 点, 在自然情况下是非常不可能发生的。
13. 具有统计显著性。当随机挑选 20 个成年人时他们全部是女性, 这是非常不可能发生的。
15. 不具有统计显著性。这个结果很容易随机产生。
17. 虽然样本的容量较小, 但是公里数 21% 的提高还是非常显著的。
19. 在 325 名婴儿中, 女孩人数通常在 162 个左右, 295 个女孩这样的结果实质上是背离了预期随机发生的结果。这个结果看起来是具有统计显著性的。
21. (1) 如果随机选取 100 个样本, 其中 5 个样本或更少的样本, 平均体温是 98.20 或者更小时, 具有统计显著性。
- (2) 选择这样一个平均数很小的样本几乎是不可能的, 不大可能期望它是偶然发生的。
23. 这个结果在 0.05 水平上不具有统计显著性, 因为如果预备课程没有真的提高的话, 它随机发生的概率大于 0.05。

6.2 练习题答案

1. $P(A)$ 代表事件发生的概率, $P(\text{非 } A)$, 或者 $P(\bar{A})$ 代表事件不发生时的概率。
3. 这个推理是错误的, 因为它假定两个可能的

第 5 章小测验答案

1. 第 2 小题和第 5 小题是正确的 2. 95%
3. 50 4. 0.5

结果——下雨和不下雨发生的可能性是相等的，但它们不可能是相等的。

5. 有意义 7. 有意义 9. 无意义
11. (1) 结果 (2) 事件，因为它以 3 种不同的结果出现。 (3) 结果 (4) 结果
 (5) 事件，因为它以 3 种不同的结果出现。
 (6) 事件，因为它以 6 种不同的结果出现。
 (7) 事件，因为它以 6 种不同的结果出现。
13. 1/3，假定这类骰子是公平的，所出现的结果可能性是相等的。
15. 2/38，假定所有的结果可能性是相等的。
17. 1/365，或者 1/19，假定出现所有结果的可能性是相等的。
19. 1/2 或者 0.5 21. 4/6 或者 2/3 23. 36/38 或者 18/19
25. 0.45 27. 0.720 29. 0.22, 0.33, 0.44, 0.56
31. (1) $1/8 = 0.125$ (GGG) (2) $3/8 = 0.375$ (BBG、BGB、GBB) (3) $1/8 = 0.125$ (GBB) (4) $7/8 = 0.875$ (GGG、BGG、GBG、GGB、BBG、BGB、GBB) (5) $4/8 = 0.5$ (BBG、BGB、GBB、BBB)
33. 0.40 35. $P(\text{成功}) = 0.86$
37. 在 2000 年随机遇到一个人是超过 65 岁的老年人的概率是 $34\,700\,000/281\,000\,000 = 0.123$ ，在 2050 年随机遇到一个人是超过 65 岁的老年人的概率可能是 $78.9/394 = 0.200$ 。这样，在 2050 年随机遇到一个老年人的概率要大于 2000 年。
39. (1) 抛掷 4 枚硬币所能呈现的所有结果如下表所示：

硬币 1	硬币 2	硬币 3	硬币 4	结果	概率
正	正	正	正	正正正正	1/16
正	正	正	反	正正正反	1/16
正	正	反	正	正正反正	1/16
正	正	反	反	正正反反	1/16
正	反	正	正	正反正正	1/16
正	反	正	反	正反反正	1/16
正	反	反	正	正反反正	1/16
正	反	反	反	正反反反	1/16
反	正	正	正	反正正正	1/16
反	正	正	反	反正正反	1/16
反	正	反	正	反正反正	1/16
反	正	反	反	反正反反	1/16
反	反	正	正	反反正正	1/16
反	反	正	反	反反正反	1/16
反	反	反	正	反反反正	1/16
反	反	反	反	反反反反	1/16

(2) 抛掷 4 枚硬币正面向上次数的概率分布如下表所示：

结果	概率
4 枚正面向上 (0 枚反面向上)	1/16
3 枚正面向上 (1 枚反面向上)	4/16
2 枚正面向上 (2 枚反面向上)	6/16
1 枚正面向上 (3 枚反面向上)	4/16
0 枚正面向上 (4 枚反面向上)	1/16
总计	1

- (3) $6/16 = 0.375$ (4) $15/16 = 0.9375$
(5) 2 枚硬币正面向上 (概率是 0.375)

6.3 练习题答案

1. 如果某些过程可以在很多试验中重复，某些事件在试验中发生的比率将会接近事件的概率。随着试验次数的增多，某些事件在试验中发生的比率就越来越接近于事件的概率。
3. 他的赌博策略是不明智的，因为赌场是要赢利的。他以前的操作并不会影响将来的事件。他的推理是不正确的。虽然他获胜的比率可能会提高，但是当他继续赌下去，他输掉的钱数也会越来越多。这个有缺陷的推理可能会使很多赌徒输掉更多钱。
5. 有意义 7. 有意义
9. 不会的。你不可能预期刚好得到 5 000 次正面向上的结果。因为某种特定结果的概率是非常小的，随着投掷次数的增多，正面向上的比率越来越接近于 0.5。
11. 你所参与的这个游戏预期的结果是 $(5 \text{ 美元} \times 0.25) + (-1 \text{ 美元} \times 0.75) = 1.25 \text{ 美元} - 0.75 \text{ 美元} = 0.50 \text{ 美元}$ 。因为你在第一次游戏中或者赢得 5 美元或者输掉 1 美元，你不可能赢得预期的 0.50 美元。如果继续 100 次游戏，预期你会赢得 $100 \times 0.50 \text{ 美元} = 50 \text{ 美元}$ 。
13. 0.94, 0.74 15. 15 分钟 17. -0.78 美元, -285 美元 19. -7.07 美分, -1.4 美分
21. (1) 决策一：选项 1，预期值 = 100 万美元；选项 2，预期值 = 114 万美元。决策二：选项 1，预期值 = 110 万美元；选项 2，预期值 = 250 万美元；
(2) 决策一中的回答和预期值是不一样的，但是决策二的回答和预期值是一样的。
(3) 看起来人们会在决策一中选择一个确定的结果。

23. (1) 0.5, 0.5 (2) 输掉 10 美元
 (3) 输掉 16 美元 (4) 输掉 20 美元
 (5) 45%, 46%, 48%; 虽然百分比数字越来越接近于 50%, 但是偶数次数和奇数次数的差异也在增加。
 (6) 60 个偶数。

6.4 练习题答案

1. 因为将出生人口表达为总体每 1 000 人中所出生人数, 这样出生率就可以直接比较了。实际的人口出生数要考虑不同国家的总体人数, 因此这样的比较需要知道总体的出生人数, 即使如此, 由于所包含的人数不同, 比较起来也不是很容易。
 3. 预期寿命是一个已知年龄的人能够继续生存的平均年限。一个 30 岁人的预期寿命要比一个 20 岁人的预期寿命要短, 也就是说, 一个 30 岁所预期的继续生存年限要小于一个 20 岁人的。
 5. 无意义 7. 有意义
 9. 1995: 0.0208; 2000: 0.0102; 2004: 0.0013。2004 年是最安全的。因为他在每 1 000 名乘客中的死亡率是最低的。
 13. 58.8 年 15. 每 10 000 人中死亡 8.3 人。
 17. (1) 犹他州: 137; 迈阿密州: 38
 (2) 犹他州: 20.8; 迈阿密州: 10.7
 19. (1) 4 230 000 (2) 2 520 000
 (3) 1 710 000 (4) 0.0057; 0.57%

6.5 练习题答案

1. 两个事件是独立事件的意思是指一个事件的发生并不影响另一个事件发生的概率。
 3. 放回取样。第二次抽样的结果和第一次抽样的结果是相互独立的。
 5. 无意义 7. 无意义
 9. 1/8, 或者 0.125
 11. 1/260; 密码太短不是一个有效的保护措施。
 13. (1) 0.0039 (2) 0.00098
 (3) 0.125 (4) 0.0625
 (5) 对于每一次选择都有 60 种可供利用的歌曲。不管第一次选择的是哪首歌, 第二次选中同一首歌的概率是 $1/60 = 0.0167$ 。
 15. $P(\text{无罪释放或者说入狱服刑}) = 1\ 014/1\ 028 = 0.986$ 。
 17. 0.865 19. 0.381 21. 0.410 23. 0.920

25. 0.0195
 27. (1) 0.733 (2) 1 (3) 0.643
 (4) 0.22
 29. (1) 0.5625 (2) 0.375 (3) 0.0625
 (4)

事件	概率
AA	0.5625
Aa	0.1875
aA	0.1875
aa	0.0625

- (5) 0.9375
 31. (1) 0.518 (2) 0.491

第6章复习题答案

1. 85/99, 或者 0.859 2. 83/99, 或者 0.838
 3. 88/99, 或者 0.889 4. 96/99, 或者 0.970
 5. 0.702 6. 0.736
 7. (1) 0.73 (2) 0.073 (3) 1.35
 (4) 0.0014; 不能相信成比率是 27%, 应该怀疑所陈述的是否准确。
 8. (1) 0.10; (2) 这不是一个“不寻常”事件, 这个事件的概率大于 0.05。
 9. (1) $1/38 = 0.026$ (2) 这是一个“不寻常”事件, 这个事件的概率小于 0.05。
 10. (1) 0.14 (2) 这不是一个“不寻常”事件, 这个事件的概率大于 0.05。

第6章小测验答案

1. 1/6 2. 1/216; 或者 0.00463 3. 0.55
 4. 7.5 5. 对 6. 0.502 7. 0.141 8. 0.613
 9. 0.251 10. 0.123

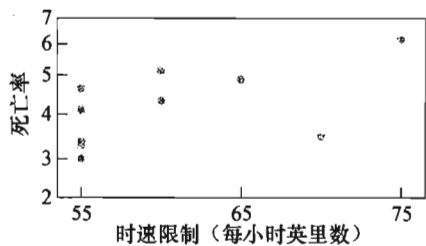
7.1 练习题答案

1. 一个变量较高的取值比较稳定地倾向于和另外一个变量的较高取值相关联, 或者说一个变量较大的取值比较稳定地倾向于和另外一个变量的较小的取值相关联, 我们说这两个变量之间存在相关。统计学上的术语“相关”的意思要比一般的用法更明确一些。
 3. 这些点相当接近一条直线的形状, 位于右边的点的高度更大一些。
 5. 无意义 7. 无意义
 9. 正相关, 因为较高的女士体重会更大一些。
 11. 负相关, 因为小汽车越重倾向于消耗更多的

燃料，所以每加仑燃料消耗量所走的里程数会少一些。

13. 正相关，因为马拉松跑得越快（时间用得越少）比赛名次就越靠前。
15. 变量之间没有相关。
17. 存在一个强的正相关，相关系数接近于0.8。这个相关的大部分是由于大部分粮食生产要用于牲畜饮料。

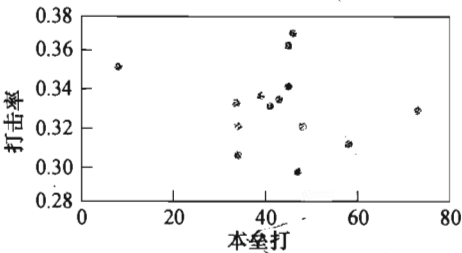
19. (1)



时速限制和死亡率的散点图

- (2) 存在一个中等水平的正相关（确切为 $r = 0.59$ ）。
- (3) 除了英国之外，时速限制通常和高死亡率是相关的，除了时速限制外，死亡率也受其他因素的影响。

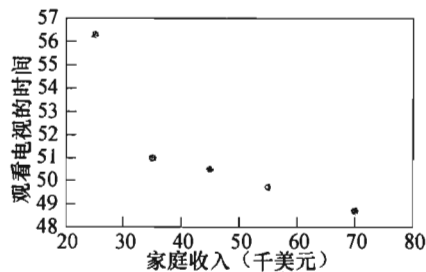
21. (1)



本垒打次数和平均打击率的散点图

- (2) 存在一个中等水平的负相关（确切为 $r = -0.29$ ）。
- (3) 否

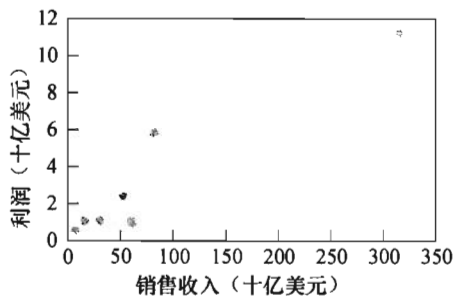
23. (1)



家庭收入和观看电视时间的散点图

- (2) 在家庭收入和每周观看电视时间存在一个强的负相关（确切为 $r = -0.86$ ）。
- (3) 家庭收入越多，就越有机会做其他的事情。

25. (1)



销售收入和利润的散点图

- (2) 在销售收入和利润之间存在一个强的正相关（具体为 $r = 0.92$ ）。这个强的正相关受一个异常值的影响较大。
- (3) 较高的销售收入并不一定能够转化为利润，一些公司会有较高的支出，导致利润下降。

27. 变量 x 和 y 在公式中是对等的（将 x 和 y 互换并不影响公式的使用）。

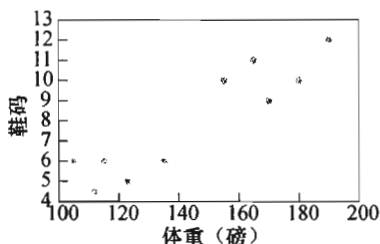
7.2 练习题答案

1. 如果在两个变量间存在显著的相关，这只是统计学意义上的关系，这种关联并不意味着其中一个变量是另外一个变量某种方式的原因或者对其产生直接影响。
3. 异常值是在数据组中远离大多数值的一个数据值，异常值可以使得数据看起来存在明显的相关，而实际上这个相关并不存在或者可能掩盖了一个真正的相关。
5. 无意义
7. 无意义
9. 存在一个正相关，可能会是一个潜在的原因。许多犯罪是由注册的枪支造成的。
11. 存在一个正相关，应该是直接原因，因为马萨诸塞州收费高速公路是根据行驶的距离来收费的。
13. 存在一个正相关，可能会是一个通常的原因，交通信号灯的增加，汽车之间的碰撞事故也在增加。
15. 存在一个正相关，可能是直接原因，当汽油价格增长时，人们不可避免要开车，但是可以通过减少开车次数来削减开支。

17. (1) 异常值是位于右上角的一个点 (0.4, 1.0)。如果没有异常值的话, 相关系数就是 0.0。

(2) 如果包括异常值, 相关系数就是 -0.58。

19. (1) 实际的相关系数是 $r = 0.92$, 它在 0.01 水平上显著, 因此这是体重和鞋码之间非常强的相关。



体重和鞋码的散点图

- (2) 分别在两组数据中计算各自的相关, 要比总体数据计算的相关要弱。

21. (1) 实际的相关系数是 $r = 0.77$, 它在 0.01 水平上显著, 说明这是一个强相关。

(2) 右侧的 16 个点对应相对贫穷的国家, 比如乌干达。余下的点对应于比较富裕的国家, 比如瑞典。

(3) 看起来在贫穷国家里这是一个负相关, 而在比较富裕的国家里, 这是一个正相关。

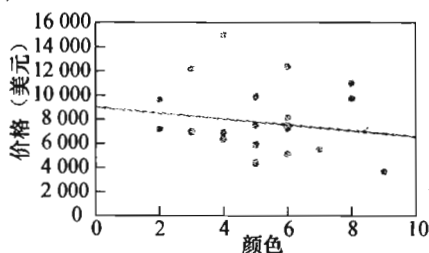
7.3 练习题答案

1. 最佳拟合线 (或者说回归线) 是散点图上的一条线, 这条线对数据的拟合要好于其他任何一条可能的线 (按照统计学标准的拟合程度)。一条最佳拟合线对于根据一个变量已知的数据预测另外一个变量未知的数据是非常有用的。

3. r^2 是相关系数的平方, 它是能用最优拟合线解释的一个变量的变异性比例。

5. 无意义 7. 无意义

9. (1)

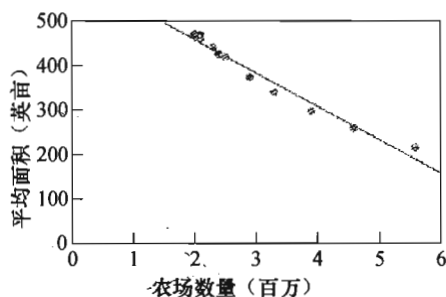


颜色和价格的最佳拟合线

- (2) 实际的相关系数 $r = -0.16$; $r^2 = 0.026$, 价格中的变异性大约 3% 可以用最佳拟合线解释。

(3) 这个最佳拟合线不应该用于预测。

11. (1)

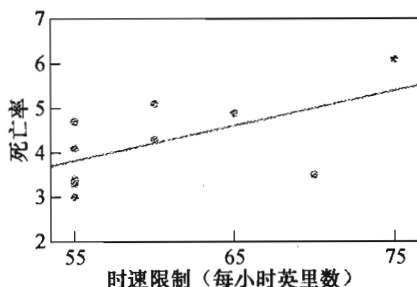


农场数量和农场大小的最佳拟合线

- (2) 实际的相关系数 $r = -0.99$; $r^2 = 0.97$, 大约 97% 的变异性可以用最佳拟合线解释。

(3) 这个最佳拟合线在农场数量这个数据的取值范围内可以用于预测。因为对数据点来说看起来有一些轻微的弯曲, 因此预测不应该超出数据范围。

13. (1)



时速限制和死亡率的最佳拟合线

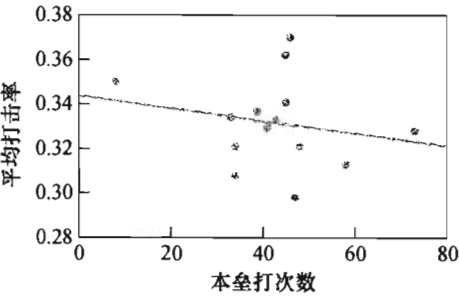
- (2) 实际的相关系数 $r = 0.59$; $r^2 = 0.35$, 大约 35% 的变异性可以用最佳拟合线解释。

(3) (75, 6.1) 和 (70, 3.5) 这两个点都可能是异常值。因为这两个点中的前者是远大于大多数点, 后者这个点要按照其余点的情况进行预计的话, 实际的死亡率比预期的要少得多。因为一个点在最优拟合线上方, 而另外一个点在拟合线下方, 两个点的综合效

应可能就相互抵消了。

- (4) 如果考虑到最优拟合线的可靠性的话，对于预测来说 r 的值就显得很小了。

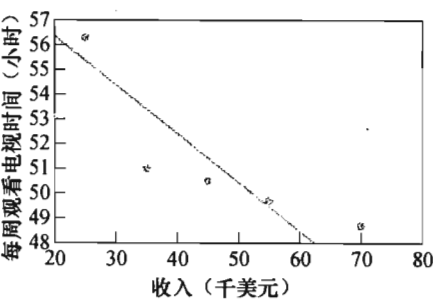
15. (1)



本垒打次数和平均打击率的最佳拟合线

- (2) 实际的相关系数 $r = -0.29$; $r^2 = 0.08$, 大约 8% 的变异性可以用最佳拟合线解释。
- (3) (0.366, 49) 是一个异常值，因为它远离其他数点。
- (4) 基于最佳拟合线的预测是不可靠的。

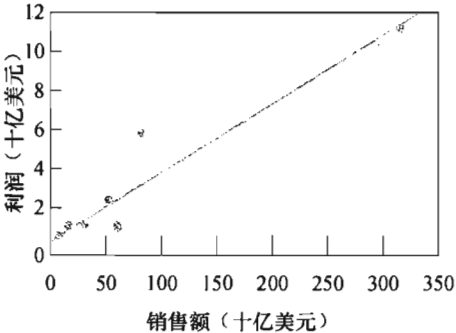
17. (1)



收入和观看电视时间的最佳拟合线

- (2) 实际的相关系数 $r = -0.86$; $r^2 = 0.74$, 大约 74% 的变异性可以用最佳拟合线解释。
- (3) (25 000, 56.3) 点是异常值。如果这个点去除掉的话，最佳拟合线交将呈现出更平缓的斜率。
- (4) 根据最佳拟合线的预测可能是可信的，但是由于这个异常值的存在，会导致对最佳拟合线的可靠性产生疑问。

19. (1) 实际的相关系数 $r = 0.92$; $r^2 = 0.84$, 大约 84% 的变异性可以用最佳拟合线解释。



销售额和利润的最佳拟合线

- (2) (81.5, 5.8) 是一个异常值。
- (3) 根据最佳拟合线的预测是可信的，但是由于这个异常值的强烈作用，会导致对最佳拟合线的可靠性产生疑问。需要有更多数据来解决这个问题。

7.4 练习题答案

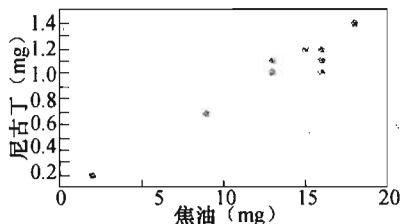
1. (1) 因此是一个巧合的结果
- (2) 相关可能是一个常见的潜在原因
- (3) 相关是一个变量对另外一个变量的直接影响。
3. 一个混淆变量是指没有包含在研究分析之内的变量，但是它对研究结果却产生了影响。如果没有排除或者考虑到混淆变量，可能会使研究者在研究中会遗漏一个可能的潜在原因。
5. 无意义 7. 无意义
9. 因果关系是有效的。当球杆头的速度加快时，作用于高尔夫球上的力量也会增加，所以高尔夫球会飞得更远些。
11. 因果关系是有效的。酒精对中枢神经系统会有抑制作用，它会产生一些副作用包括反应时减慢。这是酒后驾驶非常危险的原因。
13. 指导一；指导二和五；指导三和五。头疼和工作天数有某种形式的联系。但是头痛与可乐或者可乐中的咖啡因没有关联。头痛很可能是建筑物中通风不良的结果。
15. 吸烟只是增加了已经存在的风险。
17. 这是一个观察研究。生育较晚只是反映了一种可能的潜在原因。这个结论有可能是正确的，但也存在其他可能的解释。例如，可能

是,生活在某一时期的年轻女性选择在 40 岁以后生孩子可能性非常小,但是她们活到 100 岁仍然有可能的。

19. 高凶杀率不是仅有持有枪支这一原因。社会的、经济的或者个人的原因都可能会使个体使用枪支犯罪。

第 7 章复习题答案

1. 是强的正相关, 92.5%
2. 是强的正相关, 83.6%
3. 是强的正相关, 95.8%
4. 因为这些点的形态非常接近一条直线, 散点图说明了在两个变量之间存在一个相关。



焦油和尼古丁的散点图

5. 数据应该包括住所临近高压线人们的死亡率和住所距高压线的距离。在高压线和白血病的死亡率之间可能会存在一个相关,但是很难建立一个因果关系。
6. 散点图上的点位于一条直线上并且有一个负值斜率(向右下降)。
7. 不能仅凭相关就说明因果关系,在这个例子中,可以确定看牙医的次数增多并不会导致家庭收入增加。家庭可支配收入越多就能支付更多的看牙医的费用,或者能支付包括牙医在内的医疗保险费用。
8. 营销房屋价格的因素包括房屋所在的位置、大小、已经使用年限、房屋的环境和小区规模。房屋所在的位置经常是应用的最重要因素,要考虑的方面包括离学校、商店和医院的远近。前一位房主的年龄是很少与房屋价格有关联的。
9. 所收集的数据是不相关的。但数据所代表的变量之间仍然有可能以某种非线性的方式存在关联。换句话说,代替直线相关可能是一个曲线的散点图形式。
10. 相关系数是 -0.056 , 任何一个接近 0 的正值或负值的估计值都是合理的。

第 7 章小测验答案

1. $-1, 1$
2. (3), (4) 及 (5)
3. (3)
4. 实际的相关系数 $r = -0.934$, 但是任何一个介于 $-0.5 \sim -0.99$ 之间的数值都是这个相关系数的合理估计值。
5. 正确
6. 错误
7. 错误
8. 正确
9. 错误
10. 正确

8.1 练习题答案

1. 均匀分布; 渐近正态分布。
3. $\bar{x}$ 代表样本平均数, μ 代表总体平均数。
5. 无意义
7. 无意义
9. 东北部总体比率的最佳估计值是样本比率: 0.19, 已知的样本比率不是美国儿童总体比率的一个最佳估计值, 地区因素会对总体比率存在一种强烈的影响, 不同地区的儿童比率是不一样的。
11. (1) 平均数以上 9.5 个标准差 ($z = 9.5$)。
(2) 概率非常小, 小于 0.01。
(3) 消费者没有被欺骗, 一听可乐充满后的实际容量大于 12.00 盎司。
13. (1) 0.528 (2) 0.49 (3) 样本比率非常小。
15. 783 个人, 大样本可能会提供更为可信的总体比率的估计值。
17. (1) 平均数以下 0.67 个标准差 ($z = -0.67$)
(2) 0.253
19. (1) 1.0, 1.5, 3.0, 1.5, 2.0, 3.5, 3.0, 3.5, 5.0 (2) 2.7
(3) 第 2 小题中样本平均数分布的平均数等于总体所列出的 3 个数值的平均数, 这两个平均数总是相等的。
- 21.

样本容量 $n = 1$

样本	平均数
A	52
C	5
G	60
M	33
O	97
平均数	49.4

样本容量 $n=2$	
样本	平均数
AC	28.5
AG	56.0
AM	42.5
AO	74.5
CG	32.5
CM	19.0
CO	51.0
GM	46.5
GO	78.5
MO	65.0
平均数	49.4
样本容量 $n=3$	
样本	平均数
ACG	39.00
ACM	30.00
ACO	51.33
AGM	48.33
AGO	69.67
AMO	60.67
CGM	32.67
CGO	54.00
CMO	45.00
GMO	63.33
平均数	49.4
样本容量 $n=4$	
样本	平均数
ACGM	37.50
ACGO	53.50
ACMO	46.75
ACMO	60.50
CGMO	48.75
平均数	49.4
样本容量 $n=5$	
样本	平均数
ACGMO	49.4
平均数	49.4

8.2 练习题答案

- 1. 女性胆固醇含量总体平均水平的真值位于 183.3 ~ 298.5g，我们可以说有 95% 的置信度。预计依据类似的这种样本所建构的置信区间其中 95% 会包含总体平均数的真值。
- 3. 媒体报道经常遗漏置信水平参考值，经常采

- 用的置信水平是 95%。术语“平均数”经常被用于替代术语“平均水平”。
- 5. 无意义 7. 无意义
 - 9. 误差幅度：2.0 $73.0 < \mu < 77.0$ 。
 - 11. 误差幅度：1 469 美元；44 736 美元 $< \mu < 47\,674$ 美元。
 - 13. 49 15. 484 17. 185 19. 225
 - 21. 5. 639g；5. 619g $< \mu < 5. 659$ g。
 - 23. 5. 15yr；5. 10yr $< \mu < 5. 20$ yr。
 - 25. 146 磅 $< \mu < 221$ 磅
 - 27. (1) 3.1 (2) 1.4 (3) 3.1
(4) $2.6 < \mu < 3.6$
(5) 样本较小并且集中于你的邻居中选取的，不大可能是美国家庭的代表性样本。

8.3 练习题答案

- 1. 总体比率的真值介于 0.457 ~ 0.551，我们可以说这个范围有 95% 的置信度。我们可以预计到基于类似的样本所建构的置信区间结果其中 95% 会包含总体比率真值。
- 3. 媒体报道经常遗漏置信水平的参考值，经常采用的置信水平是 95%。
- 5. 无意义 7. 无意义
- 9. 误差幅度：0.086 6； $0.163 < p < 0.337$ 。
- 11. 误差幅度：0.025 7； $0.202 < p < 0.254$ 。
- 13. 10 000
- 15. 278
- 17. $0.140 < p < 0.160$
- 19. $E = 0.021$ ； $0.779 < p < 0.821$
- 21. (1) $0.153 < p < 0.191$
(2) $0.118 < p < 0.146$
(3) $0.215 < p < 0.243$
- 23. (1) 0.98
(2) $0.966 < p < 0.994$
(3) 由于样本是一个随机样本，所有电影被选取的机会是相等的。
- 25. 95% 的置信区间是 $0.538 < p < 0.587$ ，预估选举获胜是合理的。
- 27. 调查一： $0.494 < p < 0.546$ ；调查二： $0.494 < p < 0.534$ ；调查三： $0.498 < p < 0.532$ ；由于所有置信区间包括范围都有小于 0.5 的值，所以马丁内斯获得大多数人的支持是不可信的。

29. (1) $0.50 < p < 0.58$; (2) 621

第8章复习题答案

- (1) $0.008 < p < 0.056$;
(2) 总体比率的真值介于 $0.008 \sim 0.056$, 我们可以说这个范围有 95% 的置信度。如果随机选择容量为 221 的许多样本, 基于每个样本所建立的置信区间其中 95% 会包含总体比率的真值。
(3) 0.0237 (4) 分布接近于正态分布, 或者说是钟形分布。
(5) 这个范围将变得更狭窄。
(6) 总体比率是确定值, 不是一个随机变量, 只能说它是位于根据样本所建立的置信区间之内还是不在, 所以这里不存在概率问题。
- (1) 114
(2) 如果样本容量大于所需要的数量, 就所需要的置信区间范围的意义上而言, 样本容量越大, 置信区间越要符合条件。如果样本容量小于必要的数量, 置信区间的范围就越宽泛, 就会不符合预定的要求。
(3) 因为比起全部人群来, 大学生是高度同质的人群, 大学生的标准差可能小于 16。如果我们使用小于 16 的真值, 所需要的样本容量就少于 114 人。
- (1) $6.4 < \mu < 7.8$ (2) 我们可以说有 95% 的置信水平总体平均数的真值位于 $6.4 \sim 7.8$ 。随机抽取样本容量为 50 的样本多次, 依据这些样本所建构的置信区间, 其中 95% 会包含总体平均数的真值。
(3) 0.707 (4) 400
- (1) 2 500 (2) 这样的调查结果不恰当, 你的样本容量虽然已经足够大了, 但其中有很大可能存在偏差。那些拒绝回答问题的人们很可能是有不同观点的总体中的一部分, 而持这些观点的人将会被不正确地排除在样本之外。

第8章小测验答案

- 最接近正态分布或者是钟形分布。
- 最接近正态分布或者是钟形分布。

- 样本比率 4. $0.52 < p < 0.68$ 5. (2)
- 0.02 7. 20 8. $4.45 < \mu < 4.75$
- $0.400 < p < 0.800$
- 总体比率的置信区间应该在 $0 \sim 1$, 但题目中陈述的却不是。

9.1 练习题答案

- 一个假设检验是检验关于一个总体参数值的声称这样一种标准过程。
- $\mu \neq 98.6, \mu < 98.6, \mu > 98.6$ 。
- 无意义
- 无意义
- 无意义
- 无意义
- (1) 增长相对较小, 这个增长看起来不具有统计显著性。
(2) 这个增长相对较大, 看起来具有统计显著性。
(3) 答案有很多, 但是大约 3 加仑以上的增长就应该具有统计显著性。
- $H_0: \mu = 1L, H_a: \mu < 1L$ 。样本没有提供充分的证据支持如下的声称“每瓶可乐平均容量少于 1L”。样本提供了充分的证据支持如下的声称“每瓶可乐的平均容量少于 1L”。
- $H_0: p = 0.7, H_a: p > 0.7$ 。样本没有提供充分的证据支持如下的声称“女孩的总体比率大于 0.7”。样本提供了充分的证据支持如下的声称“女孩的总体比率大于 0.7”。
- $H_0: \mu = 10$ 盎司, $H_a: \mu \neq 10$ 盎司。样本没有提供充分的证据拒绝如下的声称“咖啡的平均容量是 10 盎司”。样本提供了充分的证据拒绝如下的声称“咖啡的平均容量是 10 盎司”。
- $H_0: p = 0.5, H_a: p > 0.5$ 。样本没有提供充分的证据拒绝如下的声称“总体比率大于 0.5”。样本提供了充分的证据拒绝如下的声称“总体比率大于 0.5”。
- 不具有统计显著性。计算出的 P 值 0.382 大于 0.05 的显著性水平, 这个较高的概率值说明一个 100 名婴儿的随机样本中有 48 名男婴的可能性是相当大的。
- 不具有统计显著性。计算出的 P 值 0.028 大于 0.01 的显著性水平, 说明一个 100 名婴

儿的随机样本中有 40 名男婴的可能性是比较大的。

27. 具有统计显著性。计算出的 P 值 0.000 2 小于 0.01 的显著性水平, 说明一个 100 名男婴的随机样本中有 35 名男婴的可能性非常小。

9.2 练习答案

1. n 代表样本容量, $\bar{x}$ 代表样本平均数, s 代表样本标准差, σ 代表总体标准差, μ 代表总体平均数。
3. I 型错误是当事实上虚无假设为真时拒绝虚无假设所犯的误差; II 型错误是当实际上虚无假设为假时只接受虚无假设犯的误差。
5. 无意义
7. 无意义
9. 有意义
11. 无意义
13. $z = 1.33$, 不支持备择假设。
15. $z = -5.33$, 支持备择假设。
17. $z = -2.50$, 支持备择假设。
19. $z = 2.80$, 支持备择假设。
21. P 值是 0.344 6, 不支持备择假设。
23. P 值是 0.028 7, 支持备择假设。
25. P 值是 0.109 6, 不支持备择假设。
27. P 值是 0.089 2, 不支持备择假设。
29. P 值是 0.003 5, 支持备择假设。
31. P 值是 0.017 9, 支持备择假设。
33. P 值是 0.880 8, 不支持备择假设。
35. 检验为右侧检验, 而已知标准分数是在分布的中心到左边, 因此 P 值一定大于 0.5。不支持备择假设。不必建立正式的假设检验, 是因为如果样本平均数小于某个确定值的话, 没有方法能够支持这样的假设“总体平均数大于某个确定值”。
37. $H_0: \mu = 7.5$ 年, $H_a: \mu < 7.5$ 年。 $n = 100$, $\bar{x} = 7.01$ 年, $s = 3.74$ 年, $\sigma \approx s$, $z = -1.3$, 在 0.05 水平上不具有统计显著性, P 值 = 0.096 8, 没有充分的证据支持如下的声称“自有汽车的平均使用寿命少于 7.5 年”。
39. $H_0: \mu = 16$ 盎司, $H_a: \mu \neq 16$ 盎司。 $n = 144$, $\bar{x} = 15.8$ 盎司, $s = 1.6$ 盎司, $\sigma \approx s$, $z = -1.5$, 在 0.05 水平上不具有统计显著性, P 值 = 0.096 8, 没有充分的证据支持如下的

声称“糖果包的重量与广告宣传的 16 盎司有明显差别”。

41. $H_0: \mu = 600\text{mg}$, $H_a: \mu \neq 600\text{mg}$ 。 $n = 65$, $\bar{x} = 589\text{mg}$, $s = 21\text{mg}$, $\sigma \approx s$, $z = -4.2$, 在 0.05 水平上具有统计显著性, P 值 = 0.000 4, 有充分的证据支持如下的声称“这种药物没有达到所需要的对乙酰氨基酚成分 600mg 的含量标准”。
43. $H_0: \mu = 21.4\text{mpg}$, $H_a: \mu < 21.4\text{mpg}$ 。 $n = 40$, $\bar{x} = 19.8\text{mpg}$, $s = 3.5\text{mpg}$, $\sigma \approx s$, $z = -2.9$, 在 0.05 水平上具有统计显著性, P 值 = 0.001 9, 有充分的证据支持如下的声称“SUV 的平均燃油公里数小于 21.4mpg”。
45. $H_0: \mu = 0.21$, $H_a: \mu > 0.21$ 。 $n = 32$, $\bar{x} = 0.83$, $s = 0.24$, $\sigma \approx s$, $z = 14.6$, 在 0.05 水平上具有统计显著性, P 值 < 0.000 2, 有充分的证据支持如下的声称“自我识别的强迫购物狂在问卷中的平均分数大于一般人群 0.21 的分数”。
47. $H_0: \mu = 3.39\text{kg}$, $H_a: \mu \neq 3.39\text{kg}$ 。 $n = 121$, $\bar{x} = 3.67\text{kg}$, $s = 0.66\text{kg}$, $\sigma \approx s$, $z = 4.7$, 在 0.05 水平上具有统计显著性, P 值 < 0.000 4, 有充分的证据支持如下的声称“母亲在怀孕期间服用维生素制剂所产下男婴的平均重量显著不同于全部男性婴儿 3.39 千克的平均体重”。
49. $H_0: \mu = 24\text{mo}$; $H_a: \mu < 24\text{mo}$; $n = 70$, $\bar{x} = 22.1\text{mo}$, $s = 8.6\text{mo}$, $\sigma \approx s$, $z = -1.8$; 显著性水平在 0.05 水平, P 值 < 0.035 9, 有充分的证据支持如下声称“平均值小于 24mo”。
51. I 型错误: 当罹患者的某种疾病实际上已经治愈却拒绝如下声称“罹患者所罹患的某种疾病已经治愈”; II 型错误: 当罹患者的某种疾病实际上还没有治愈却没有拒绝“罹患者的某种疾病已经治愈”的声称。
53. I 型错误: 当彩票实际上是公平时却拒绝“彩票是公平的”的声称; II 型错误: 当彩票实际上是不公平时却没有拒绝“彩票是公平的”的声称。

9.3 练习题答案

1. 符号 p 代表总体比率, 符号 $\hat{p}$ 代表样本比率, P 值代表当至少作为一个样本比率的例子时抽取这样的样本比率的概率。

3. 这个样本是无效的, 这个样本是一个自反应样本, 不是一个简单随机样本。回答这个问题的群体并不能够代表一般总体。
5. 有意义
7. 无意义
9. $H_0: p = 0.5$, $H_a: p > 0.5$ 。 $n = 400$, $\hat{p} = 0.51$, 样本比率分布的标准差是 0.025, $z = 0.5$, 在 0.05 水平上不具有统计显著性, P 值 = 0.344 6。没有充分的证据支持如下的声称“大多数投票支持这位候选人”。
11. $H_0: p = 0.50$, $H_a: p < 0.50$, $n = 1\ 015$, $\hat{p} = 0.44$, 样本比率分布的标准差是 0.015 7, $z = -3.8$, 在 0.05 水平上具有统计显著性, P 值 $< 0.000\ 2$ 。有充分的证据支持如下的声称“全部青少年中将近半数感觉学习成绩是他们最大的压力源”。
13. $H_0: p = 0.099$, $H_a: p < 0.099$, $n = 1\ 050$, $\hat{p} = 0.093\ 3$, 样本比率分布的标准差是 0.009 2, $z = -0.6$, 在 0.05 水平上不具有统计显著性, P 值 = 0.274 3。没有充分的证据支持如下的声称“这所学校里的非法使用药物要低于现在的全国平均水平”。
15. $H_0: p = 0.5$, $H_a: p > 0.5$; $n = 276\ 000$, $\hat{p} = 0.51$, 样本比率分布的标准差是 0.000 952, $z = 10.5$, 在 0.05 水平上具有统计显著性, P 值 $< 0.000\ 2$ 。有充分的证据支持如下的声称“超过半数的一年级大学生认为堕胎应该是合法的”。
17. $H_0: p = 0.53$, $H_a: p \neq 0.53$; $n = 3\ 600$, $\hat{p} = 0.54$, 样本比率分布的标准差是 0.008 3, $z = 1.2$, 在 0.05 水平上不具有统计显著性, P 值 = 0.230 2。没有充分的证据支持如下的声称“全国使用天然气 53.0% 的平均水平有所改变”。

第 9 章复习题答案

1. (1) $H_0: \mu = 12$ 盎司 (2) $H_a: \mu \neq 12$ 盎司
(3) $z = 10.4$ (4) $z = 1.645$
(5) 小于 0.000 2 (6) 拒绝虚无假设, 并且声称“每听可口可乐的容量大于 12 盎司”。
- (7) I 型错误是这样的结果, 如果我们得出结论总体平均数大于 12 盎司, 而实际上却并不如此。

- (8) II 型错误是这样的结果, 如果我们没有得出结论总体平均数大于 12 盎司, 实际上, 总体平均数确实大于 12 盎司。
- (9) P 值是 z 大于 10.4 和小于 -10.4 的概率。因为 z 大于 10.4 或者小于 -10.4 的概率小于 0.000 2, 所以 P 值小于 $2 \times 0.000\ 2 = 0.000\ 4$ 。
2. (1) $H_0: p = 0.5$ (2) $H_a: p > 0.5$ (3) $z = 0.8$ (4) $z = 1.645$ (5) 0.211 9
(6) 没有足够的证据支持如下的声称“采用可替宁贴治疗方法试图戒烟一年后, 大多数吸烟者仍在吸烟”。
- (7) I 型错误是这样的错误, 如果我们得出结论戒烟一年后大多数吸烟者仍在吸烟, 而实际上一年后吸烟者的比率等于或小于 0.5。
- (8) II 型错误是这样的错误, 如果我们没有得出如下结论, 戒烟一年后大多数吸烟者仍在吸烟, 而实际上, 戒烟一年后大多数吸烟者仍在吸烟。
- (9) P 值小于 $2 \times 0.211\ 9 = 0.423\ 8$ 。
3. 不支持女孩出生概率增长的声称, 样本比例小于 0.5, 没有办法来支持总体比率 $p > 0.5$ 的声称。
4. 样本不是随机的。在很大可能性上, 他的家庭成员不是总体的代表性样本, 所以结果是有偏差的。

第 9 章小测验答案

1. $p > 0.2$ 2. $p = 0.2$ 3. $p \neq 0.6$ 4. 右侧检验
5. $\mu = 110$ 6. $\mu > 110$
7. 不支持平均数大于 110 的假设。
8. 拒绝虚无假设。不拒绝虚无假设。
9. 支持平均数大于 110 的声称, 不支持平均数大于 110 的声称。
10. I 型错误。

10.1 练习题答案

1. 总体标准差未知, 总体符合正态分布; 或者总体标准差未知; 并且样本容量大于 30。
3. 没有什么差别。术语“ t 分布”是学生 t 分布的通常说法, 这两个术语是相同的。
5. 无意义 7. 有意义
9. $97.2 < \mu < 107.8$ 11. 14.3 英寸 $< \mu < 14.7$ 英寸

13. (1) 6370 美元 $< \mu < 11\,638$ 美元
(2) 11 638 美元
15. (1) $0.085 < \mu < 0.157$
(2) 这个声称看起来是无效的。因为 0.165 克/公里并不包括在置信区间内, 这个值不大可能是平均数。
17. $H_0: \mu = 0.3$, $H_a: \mu < 0.3$ 。检验统计量 $t = -0.119$ 小于查表 10-1 所得临界值 $t = -1.753$, 所以不能拒绝虚无假设 (P 值 = 0.4534)。没有充分证据支持如下的声称“所有谷类食品的平均含糖量少于 0.3 克。”
19. $H_0: \mu = 92.84$, $H_a: \mu \neq 92.84$ 。检验统计量 $t = -0.425$ 小于查表 10-1 所得临界值 $t = -2.093$, 所以不能拒绝虚无假设 (P 值 = 0.6758)。没有充分的证据支持如下的声称“新球弹起的平均高度不等于 92.84 英寸”。新球看起来并没有什么不同。
21. $H_0: \mu = 3.39$, $H_a: \mu \neq 3.39$ 。检验统计量 $t = 1.735$ 小于查表 10-1 所得临界值 $t = 2.131$, 所以不能拒绝虚无假设 (P 值 = 0.1302)。没有充分证据支持如下的声称“母亲在怀孕期间服用维生素的这些男婴与所有男婴 3.39 千克的平均体重是有显著差异的”。基于上述结果, 母亲怀孕期间服用维生素对于婴儿的体重不会产生影响。
23. $H_0: \mu = 10.5$, $H_a: \mu < 10.5$ 。检验统计量 $t = -0.0095$ 小于查表 10-1 所得临界值 $t = -1.717$, 所以不能拒绝虚无假设 (P 值 = 0.4627)。没有充分证据支持如下的声称“所有的平均成绩低于 10.5 秒”。在后来的年份里时间的测量精确度越来越高, 假设检验没有考虑时间减少的模式。
- 关系。
11. 临界值: $\chi^2 = 3.841$ 。不能拒绝两个变量相互独立的虚无假设。性别和反应时看起来没有关系。
13. (1) H_0 : 性别和学生类型 (全日制还是业余) 是相互独立的, H_a : 性别和学生类型 (全日制还是业余) 是有关联的。
(2) 48.24, 35.76, 36.76, 27.24
(3) 0.174 (4) 3.841
(5) 不能拒绝两个变量相互独立的虚无假设, 看起来性别和学生类型 (全日制还是业余) 是没有关联的。
15. (1) H_0 : 对问题的回答与被试是工人还是老板没有关系, H_a : 对问题的回答与被试是工人还是老板有关联。
(2) 181.60, 254.40, 50.40, 70.60
(3) 4.698 (4) 3.841
(5) 拒绝两个变量相互独立的虚无假设, 对问题的回答与被试是工人还是老板有一定的关联。
17. (1) H_0 : 关节炎罹患者的症状改善与采用物理治疗还是给予安慰剂无关。 H_a : 关节炎罹患者的症状改善与采用物理治疗还是给予安慰剂有关系。
(2) 54.16, 43.84, 50.84, 41.16
(3) 0.289 (4) 3.841
(5) 不能拒绝两个变量相互独立的虚无假设, 关节炎罹患者的症状改善与采用物理治疗还是给予安慰剂有关系。
19. (1) H_0 : 不同类型的犯罪和罪犯是否为陌生人无关。 H_a : 不同类型的犯罪和罪犯是否为陌生人有关系。
(2) 29.93, 284.64, 803.43, 21.07, 200.36, 565.57
(3) 119.330 (4) 9.210 (5) 拒绝两个变量相互独立的虚无假设, 不同类型的犯罪和罪犯是否为陌生人有关系。

10.2 练习题答案

- 一个双向表包括对于两个不同变量的各个不同类别的全部次数。一个变量作为横行的分类, 另外一个变量作为纵列的分类。
- 理论期望次数不一定是整数。通常计算出来的理论期望次数不是整数。它们是理论预期的平均次数, 并不是在特定情况下观察到的次数。
- 无意义 7. 有意义
- 临界值: $\chi^2 = 6.635$ 。不能拒绝两个变量相互独立的虚无假设。性别和反应时看起来没有

10.3 练习题答案

- ANOVA 代表一元方差分析, 它是检验三个以上或更多平均数是否相等的方法。
- 不能得出下面这样的结论。在美国, 英语、数学和科学这三个专业的学生在抽象推理测验上的平均分数是不相同的。纽波特大学的

学生不一定是美国所有大学生的代表性样本。

5. 无意义 7. 无意义

9. (1) H_0 : 三位作家的 Flesch-Kincaid 难度系数平均数是相同的。

(2) H_a : 三位作家的 Flesch-Kincaid 难度系数平均数是不同的。

(3) 0.000 77

(4) 可以得出这样的结论: 三位作家的 Flesch-Kincaid 难度系数平均数是有显著差异的。

11. (1) H_0 : 不同年龄组的运动员平均比赛完成时间是相同的。

(2) H_a : 不同年龄组的运动员平均比赛完成时间是不一样的。

(3) 0.828 324 293

(4) 不能。 P 值说明不能拒绝平均数相等的虚无假设, 不同年龄组的运动员平均比赛完成时间看起来并没有差异。

13. P 值是 0.421 6, 不能拒绝平均数相等的虚无假设, 没有足够的证据拒绝如下的声称“不同重量类别的汽车有着相同的平均数水平”。

15. P 值是 0.030 5, 拒绝平均数相等的虚无假设, 有足够的证据拒绝如下的声称“不同时期的人们有着相同的头盖骨宽度的平均数”。

故是相互独立的, H_a : 去年一年中使用手机和发生交通事故是有关联的。

(3) 27.76, 277.24, 41.24, 411.76

(4) 检验统计量 1.505 低于 3.841, P 值大于 0.05。

(5) 没有充分的证据拒绝如下的声称“去年一年中使用手机和发生交通事故是相互独立的”。基于这些数据, 使用手机并不是一个导致交通事故的因素。

(6) 更多的数据说明手机使用和驾驶是有关系的, 使用手机导致注意力分散是非常危险的。一些州立法禁止开车使用手机。

2. $H_0: \mu = 165$, $H_a: \mu > 165$ 。检验统计量 $t = 8.398$ 大于查表 10-1 的临界值, 所以拒绝虚无假设 (P 值 = 0.000 08), 有充分的证据支持如下的声称“可乐罐的轴向负荷大于 165 磅”。

3. 227.2 磅 $< \mu < 278.3$ 磅。

4. (1) 三个不同的年龄组有着相同的平均体温。

(2) 三个不同的年龄组平均体温是不同的。

(3) 0.194 915

(4) P 值大于 0.05, 这说明我们拒绝平均数相等的虚无假设。

第 10 章复习题答案

1. (1) 发生交通事故中使用手机者的百分比是 7.5%, 发生交通事故中没有使用手机者所占的百分比为 10.2%, 基于以上数据, 使用手机看起来并不危险。

(2) H_0 : 去年一年中使用手机和发生交通事

第 10 章小测验答案

1. (2) 2. (2) 3. (1) 4. $H_0: \mu = 1500$

5. $H_0: \mu > 1500$ 6. 错误 7. 一元方差分析

8. 不拒绝 3 个平均数相等的虚无假设。

9. H_0 : 性别和反应时是相互独立的。

10. 拒绝虚无假设。

推荐阅读

◎ 心理学与人类世界：无处不在的心理学

作者：（美）罗伯特 S. 费尔德曼 ISBN: 978-7-111-34800-9 定价：65.00元

◎ 人生的心理智慧（原书第9版）

作者：（美）韦恩·韦登 ISBN: 978-7-111-33058-5 定价：65.00元

◎ 左看右看心理学

作者：（美）戴维 N. 萨特勒 ISBN: 978-7-111-31163-8 定价：45.00元

◎ 心理学的50大奥秘

作者：（美）斯科特 O. 利林费尔德 等 ISBN: 978-7-111-37788-7 定价：39.00元

◎ 发展心理学——孩子的成长

作者：（美）丹尼斯·布伊德 等 ISBN: 978-7-111-34108-6 定价：65.00元

◎ 发展心理学：探索人生发展的轨迹

作者：（美）罗伯特 S. 费尔德曼 ISBN: 978-7-111-34764-4 定价：70.00元

◎ 你一生的故事——走近发展心理学

作者：（美）斯潘塞 A. 拉瑟斯 ISBN: 978-7-111-33699-0 定价：59.00元

◎ 打开“积极心理学”之门

作者：（美）克里斯托弗·彼得森 ISBN: 978-7-111-31422-6 定价：65.00元

◎ 社会心理学（原书第12版）

作者：（美）罗伯特 A. 巴伦 ISBN: 978-7-111-35320-1 定价：65.00元

◎ 教育心理学

作者：（美）莉萨·博林 等 ISBN: 978-7-111-36547-1 定价：69.00元

◎ 人格心理学：经典理论和当代研究（原书第4版）

作者：（美）霍华德 S. 弗里德曼 ISBN: 978-7-111-33921-2 定价：70.00元

◎ 工业与组织心理学（原书第4版）

作者：（美）保罗 E. 斯佩克特 ISBN: 978-7-111-31391-5 定价：70.00元

◎ 心理学：主题曲和变奏曲（英文版·原书第8版）

作者：（美）韦恩·韦登 ISBN: 978-7-111-32965-7 定价：92.00元

◎ 认知心理学

作者：（美）罗伯特·索尔所 等 ISBN: 978-7-111-31144-7 定价：70.00元

◎ 实验心理学

作者：（美）巴里 H. 坎特威茨 等 ISBN: 978-7-111-31168-3 定价：62.00元

◎ 心理测验：原理、应用和问题

作者：（美）罗伯特 M. 卡普兰 等 ISBN: 978-7-111-31169-0 定价：70.00元