



HZ BOOKS
华章教育

CENGAGE
Learning™



STAT

统计学

(美) 罗伯特 R. 强森 (Robert R. Johnson) 著
帕特里西亚 J. 库比 (Patricia J. Kuby)
夏国风 姜爱萍 等译



机械工业出版社
China Machine Press

乐读
系列教材

STAT

统计学

(美) 罗伯特 R 强森 (Robert R. Johnson) 著
帕特丽西亚 J. 库比 (Patricia J. Kuby)



机械工业出版社
China Machine Press

本书是一本统计学教材,与众多统计学教材不同的是,本教材从简单的理论出发,逐步深入,同时配有大量的专栏、案例,形式新颖,严谨中不失活泼,能够提高学生的阅读兴趣。本书每章末都有大量的练习题,练习题分节设置,有利于学生即学即练,及时掌握和复习相关知识,本书书末设置了乐考卡和上机卡,乐考卡有利于学生在学完全书后梳理全书章节要点,上机卡有利于学生上机操作,为学生上机操作提供了指导。

本书适合统计专业及其他相关专业本科生、高职生阅读,也可用做相关专业人士的参考用书。

Robert R. Johnson, Patricia J. Kuby. STAT.

Copyright © 2010 Brooks/Cole, a part of Cengage Learning.

Original edition published by Cengage Learning. CMP Press is authorized by Cengage Learning to publish and distribute exclusively this simplified Chinese edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan). Unauthorized export of this edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

All rights reserved.

本书原版由圣智学习出版公司出版。本书中文简体字翻译版由圣智学习出版公司授权机械工业出版社独家出版发行。此版本仅限在中华人民共和国境内(不包括中国香港、澳门特别行政区及中国台湾地区)销售。未经授权的本书出口将被视为违反版权法的行为。未经出版者预先书面许可,不得以任何方式复制或发行本书的任何部分。

本书封底贴有 Cengage Learning 防伪标签,无标签者不得销售。

封底无防伪标均为盗版

版权所有,侵权必究

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:01-2009-7327

图书在版编目(CIP)数据

统计学/(美)强森(Johnson, R. R.), (美)库比(Kuby, P. J.)著;夏国风,姜爱萍等译.—北京:机械工业出版社,2011.3

(乐读系列教材)

ISBN 978-7-111-33687-7

I. 统… II. ①强… ②库… ③夏… III. 统计学-教材 IV. C8

中国版本图书馆CIP数据核字(2011)第038471号

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:宁 珊 版式设计:刘永青

中国电影出版社印刷厂印刷

2011年4月第1版第1次印刷

184mm×260mm·25印张

标准书号:ISBN 978-7-111-33687-7

定价:49.00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

购书热线:(010) 68326294; 88379649; 68995259

投稿热线:(010) 88379007

读者信箱:hjzg@hzbook.com

本书中文版中使用的图片均为原版图书中的图片,因其中部分图片无法与其著作权人取得联系,故未能向其支付稿酬。请该等图片的著作权人在看到本书后与本社联系,领取稿酬,本社对您的支持表示衷心的感谢。



致读者

教育是什么？教育意义何在？这些问题一直被历代思想者和教育家所追问。

柏拉图说：“教育是为了以后的生活所进行的训练，它能使人变善，从而高尚地行动。”

巴格莱说：“教育是传递人类积累的知识中具有不朽价值的那部分的过程。”陶行知则说：“生活即教育。”

关于“教育”的定义，也许难以有一个绝对的答案，因为教育是一种实践活动，总是处于不断的实践发展与总结提炼中。

现代教育的目的在于发展学习者的认知结构，培养其创造力和批判力，从而更好地提高其生活技能，使其获得更为幸福的生活。

在教育的过程中，学习是十分重要的一部分，而阅读又是学习活动中不可分割的一部分。教材作为用于向学生传授知识、技能和思想的材料，是教学活动中最为重要的阅读物，但一直以来国内出版界对于教材阅读感受的重视显然不够。

目前国内的教材或者篇幅繁冗、内容庞杂，不能在有效的时间内完成阅读，或者内容过于简单，阅读感差，用“味同嚼蜡”形容一点不为过。

适合轻松愉悦学习的教材颇难寻觅！

生活与学习是一种体验的过程，我们应该让这种体验变得快乐。如何让教育者及学习者从沉闷的教材中体验到快乐，并乐于阅读，这一直是作为教育出版者的我们所思考和不懈为之奋斗的目标。

经过长时间的选题甄选工作，最终有了今日“乐读”系列教材的出版。它是在对美国几百所大学的教师和学生、几十个学科调查研究的基础上，由国外权威出版机构精心打造的一本寓教于乐的全新系列，其一改往日教材的厚重繁复，以内容全面、言简意赅、图文并茂、装帧精美、教辅齐全为主要特点，被奉为快乐阅读的教材榜样，一经推出即获得巨大成功，受到广大师生热捧，迅速成为教材市场的新宠。时至今日，全世界超过1 500所大学、100万的学生曾经或者是正在使用该系列教材。在各方的努力下，中文版得以正式出版，我们相信它们必将成为教师乐教、学生乐学的“乐读”教材。

诚挚祝愿各位读者朋友快乐学习、快乐阅读！

出版者

2010年10月

译者序

写这篇序时，正值学生毕业离校之际，4年的大学生活让他们更加美丽与自信，而他们大二上统计课的情形依然清晰地印在我的脑海里。在上海大学悉尼工商学院从教多年，亲身经历学院的发展壮大，从单一的一个专业发展到现在的国际经济与贸易、金融、会计、信管和工管，每个专业的学生都表现出不同的特性，学生们对知识的理解和掌握也存在差异，教研组的老师经常在一起讨论，到底什么样的教材适合不同性质、不同专业的学生使用。近年来，国内引进了各种英文统计学教材，我们选用了多本，并且对每本教材的特点、难易程度和适用范围都进行了比较，包括MBA和CFA教材。我们发现国外教材非常偏重实际应用，每一部分都引入生活中常见的实例或案例，悄然地把读者引入统计专业知识的殿堂。同时，这些教材几乎都摒弃了烦琐的数学推导，大部分只保留基本公式，少数则采用纯语言的形式。很显然，对本科和高职学生来说，用深奥的理论来武装自己不是明智的选择，学会用统计分析方法解决实际问题才是首要目的。而我们推荐的这本书，除了处处体现“生活中的统计学”的主旨以外，还表现出几个鲜明的特点。首先，本书覆盖了本科和高职教学大纲的全部内容。书中对相关知识的介绍并不粗略，采取了循序渐进、细致入微的方式；其次，本书的结构安排更加合理。一般的统计学教材把估计和假设检验放在两章进行介绍，本书则融合到一章进行对比介绍，这种安排更能让学生明白估计和假设检验是统计推断的两个孪生分支；再次，也是本书最具特色的地方，即乐考卡和上机卡的设置。面对一门抽象的课程，单凭教师每周4小时的课堂讲解远远不够，而课下苦读英文教材又让学生颇感沉重，乐考卡对各章主要内容的提炼无疑点燃了学生学习的兴趣。同时，统计分析对大量数据的处理离不开电脑软件或计算器的使用。上机卡则详细给出了Minitab和Excel上机指引以及TI-83/84计算器的操作说明，学生不必携带厚重的教材，只需带着一张薄薄的卡片就能轻松完成上机任务。

本书的翻译凝结了学院所有统计老师的心血，初稿是按照以下顺序进行分工的。夏国风：第1~9章及对应的乐考卡和附录；姜爱萍：第10~14章及对应的乐考卡和上机卡；林琳：第1~9章上机卡；李晗：第6~8章课后习题；夏国风负责全书的协调与定稿。

感谢机械工业出版社华章公司宁姗、王洪波两位女士，她们热心地促成这部中文版面世。感谢我们的家人，他们给予我们极大的理解和支持。

由于译者水平有限，译稿难免存在疏漏和差错，敬请读者批评指正。

希望本书成为你喜爱并能充分利用的书之一！

夏国风 姜爱萍

2010年6月12日于上海大学嘉定校区文商楼

教学建议

教学目的

本课程教学的目的在于让学生掌握统计学的基本原理并能运用统计分析方法解决实际问题。课程内容主要涉及描述性统计、概率、推断性统计三大部分。本教材不仅介绍了统计学中的基本概念、原则和方法,还给出了统计分析软件的基本操作步骤,使学生一方面对基础理论有所了解,同时能运用软件进行数据分析,从而为其他课程的学习和今后从事相关工作打下坚实基础,并从中受益。

前期需要掌握的知识

高等数学

课时分布建议

教学内容	学习要点	课时安排	
		本科	高职
第 1 章 统 计 学	(1) 了解什么是统计学 (2) 掌握不同的数据收集方法 (3) 明白概率论和统计学之间的关系 (4) 理解统计学和现代科学技术之间的关系	4	4
第 2 章 描述分析和单变量 数据的展示	(1) 了解统计图的制作 (2) 知道如何描述中心趋势 (3) 理解离散程度及标准差的含义 (4) 知道如何测度数据位置	4	4
第 3 章 描述分析和双变量 数据的展示	(1) 理解定量数据间线性相关性的含义并能计算相关系数 (2) 知道如何对双变量数据做回归分析	4	4
第 4 章 概 率	(1) 了解概率的性质 (2) 知道什么是条件概率 (3) 掌握概率运算法则 (4) 理解互斥事件、独立事件的含义	2	2
第 5 章 概率分布 (离散变量)	(1) 理解离散变量和连续变量之间的区别 (2) 学会构建离散概率分布,并能计算其均值和方差 (3) 掌握二项分布及其均值与方差的计算	4	4
第 6 章 正态概率分布	(1) 了解正态分布的特性 (2) 弄清正态分布、标准正态 z 和概率之间的关系 (3) 知道在特定条件下,二项分布可以合理地近似为正态分布	2	2
第 7 章 样本间变异	(1) 理解样本均值的抽样分布的含义 (2) 学会利用样本均值的抽样分布求相应概率	4	4

(续)

教学内容	学习要点	课时安排	
		本科	高职
第 8 章 统计推断概述	(1) 在 σ 已知的情况下, 掌握总体均值 μ 的估计方法 (2) 在 σ 已知的情况下, 学会用临界值法和 p 值法检验关于 μ 的假设	4	4
第 9 章 单总体的推断	(1) 在 σ 未知的情况下, 知道如何用 t 统计量对总体均值 μ 做出估计 (2) 在 σ 未知的情况下, 能够用临界值法和 p 值法对 μ 做 t 检验 (3) 能够对二项比例 P 做出估计和检验 (4) 能够做方差或标准差的假设检验	6	6
第 10 章 双总体的推断	(1) 理解相关样本和独立样本的含义 (2) 掌握相关样本总体均值之差的推断方法 (3) 掌握独立样本总体均值之差和比例之差的推断方法 (4) 掌握独立样本总体方差之比的推断方法	4	4
第 11 章 卡方的应用	(1) 理解卡方统计量的构成 (2) 了解多项试验的特性 (3) 掌握列联表的构建和独立性检验与齐性检验的区别	4	4
第 12 章 方差分析	(1) 理解方差分析的含义和方差分析的基本原理 (2) 学会做单因素的方差分析	4	4
第 13 章 线性相关和回归分析	(1) 理解相关分析和回归分析的含义 (2) 能够对线性相关系数做出推断 (3) 知道如何做线性回归分析并能对回归系数做出推断 (4) 学会构建 y 均值的置信区间	4	4
第 14 章 非参数统计的要素	(1) 理解非参数统计的含义 (2) 知道如何做符号检验、曼-惠特尼 U 检验和游程检验 (3) 掌握秩相关系数的计算及其检验	4	2
课时总计		54	52

说明

- (1) 由于统计课属于专业基础课, 本科和高职教学大纲规定的内容无显著差异, 因此课时安排差别不大, 但任课教师可以根据学生的性质确定授课及考核的难易程度。
- (2) 除第1章外, 每章可额外安排1小时上机课。

Contents

目 录

致 读 者
译 者 序
教学建议

第1章 统计学

- 1.1 什么是统计学.....2
- 1.2 可测量性和变异性.....8
- 1.3 数据收集.....9
- 1.4 概率和统计学比较.....13
- 1.5 统计学和科学技术.....14

第2章 描述分析和单变量数据的展示

- 2.1 图形、排列图和茎叶图.....21
- 2.2 频数分布和直方图.....25
- 2.3 中心趋势度量.....30
- 2.4 离散程度的度量.....34
- 2.5 位置测度.....35
- 2.6 对标准差的解释和理解.....39
- 2.7 统计陷阱.....41

第3章 描述分析和双变量数据的展示

- 3.1 双变量数据.....48
- 3.2 线性相关.....52
- 3.3 线性回归.....56



第4章 概率

- 4.1 事件的概率.....68
- 4.2 事件的条件概率.....73
- 4.3 概率运算法则.....75
- 4.4 互斥事件.....79
- 4.5 独立事件.....82
- 4.6 互斥事件和独立事件有联系吗.....84

第5章 概率分布（离散变量）

- 5.1 随机变量.....93
- 5.2 离散随机变量的概率分布.....94
- 5.3 离散概率分布的均值和方差.....96
- 5.4 二项概率分布.....98
- 5.5 二项分布的均值和标准差.....101

第6章 正态概率分布

- 6.1 正态分布.....110
- 6.2 标准正态分布.....112
- 6.3 正态分布的应用.....115
- 6.4 符号.....118
- 6.5 二项分布的正态近似.....119

第7章 样本间变异

- 7.1 抽样分布.....128
- 7.2 样本均值的抽样分布.....132
- 7.3 样本均值抽样分布的应用.....135

第8章 统计推断概述

- 8.1 估计的性质.....142
- 8.2 均值 μ 的估计 (σ 已知)144
- 8.3 假设检验的性质.....149
- 8.4 均值 μ 的假设检验
(σ 已知): p 值法.....153
- 8.5 均值 μ 的假设检验
(σ 已知): 临界值法.....159

第9章 单总体的推断

- 9.1 均值 μ 的推断 (σ 未知)170
- 9.2 二项成功概率的推断.....177
- 9.3 关于方差和标准差的推断.....183

第10章 双总体的推断

- 10.1 相关样本和独立样本.....198
- 10.2 两个相关样本均值差异的推断.....199
- 10.3 两个独立样本均值差异的推断.....203
- 10.4 两个独立样本总体比例差异
的推断.....208
- 10.5 两个独立样本方差比的推断.....212

第11章 卡方的应用

- 11.1 卡方统计量.....222
- 11.2 多项试验的推论.....225
- 11.3 列联表的推论.....229

第12章 方差分析

- 12.1 方差分析简介.....240
- 12.2 方差分析方法的逻辑.....243
- 12.3 单因子方差分析的应用.....244

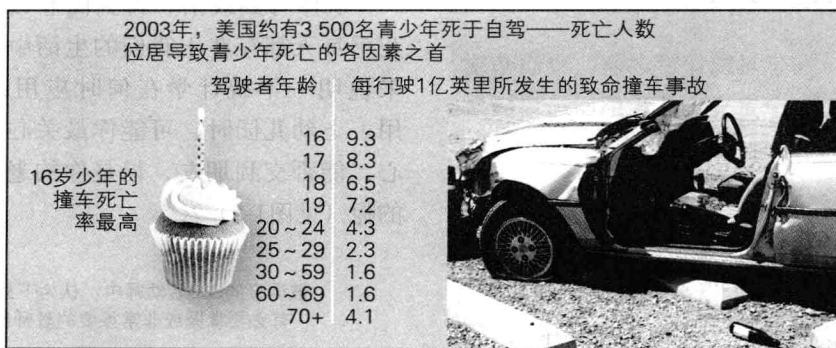
第13章 线性相关和回归分析

- 13.1 线性相关分析.....254
- 13.2 线性相关系数的推论.....256
- 13.3 线性回归分析.....259
- 13.4 回归直线斜率的推论.....263
- 13.5 回归的置信区间.....265
- 13.6 理解相关和回归之间的关系.....269

第14章 非参数统计的要素

- 14.1 非参数统计.....274
- 14.2 统计检验的比较.....275
- 14.3 符号检验.....276
- 14.4 曼-惠特尼 U 检验.....282
- 14.5 游程检验.....287
- 14.6 秩相关.....289

附录A

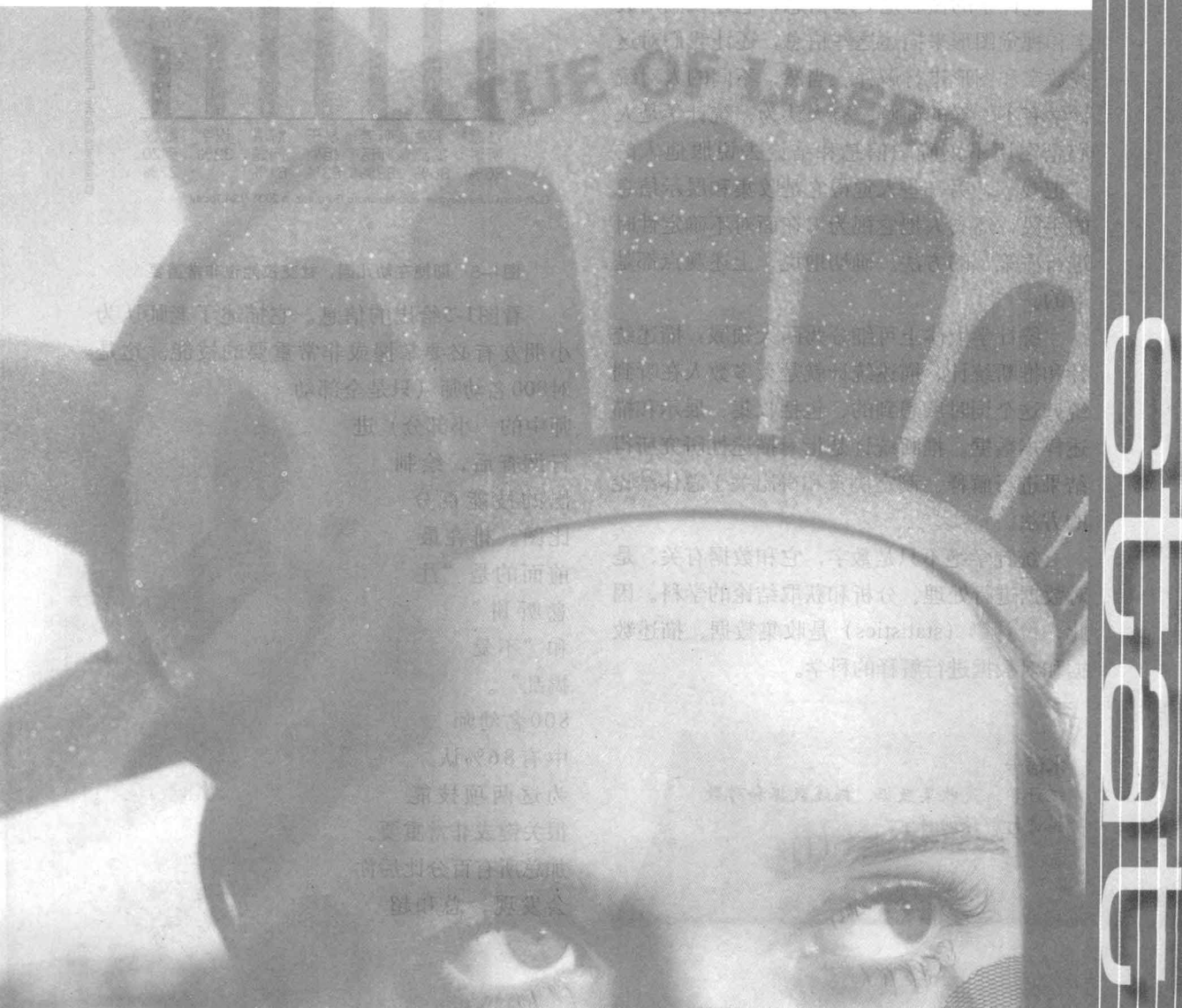


Data from USA Today, © 2003.

© Marie-France Bélanger/istockphoto /
© Tomi/PhotoLink/Getty Images

图1-4 对青少年来说最危险的事就是驾驶

我们常用这样的图表和许多其他指标来描述美国人的生活。



1.1 什么是统计学

在学习统计学之前，我们必须先对统计学下一个定义，并详细阐述该概念的一些细节内容。

统计学已经成为很多学科的通用语言。作为潜在的使用者，我们既要知晓正确使用统计方法所需的“科学”，又要掌握其中的“艺术”。对统计方法的小心使用，可以让我们获取数据中的真实信息。这些方法包括：①仔细界定情境，②收集数据，③精确地描述数据，④推断并得出有意义的结论。

统计学的核心是处理信息，它让我们用数字和视觉图形来描述这些信息，还让我们对这些数字和图形进行解释。当然，不同的人对统计学有不一样的理解。有人认为，统计学是人们企图用不正确的信息和结论去说服他人的“把戏”，另一些人觉得它是收集和展示信息的手段，还有人把它视为“在面对不确定性时进行决策”的方法。确切地说，上述观点都是对的。

统计学大体上可细分为两大领域：描述统计和推断统计。描述统计就是大多数人在听到统计这个词时所想到的，包括收集、展示和描述样本数据。推断统计是指对描述性研究所得结果进行解释、制定决策和得出关于总体结论的方法。

统计学绝不只是数字，它和数据有关，是对数据进行处理、分析和获取结论的学科。因此，统计学（statistics）是收集数据、描述数据和对数据进行解释的科学。

术语卡

统计学：是收集数据、描述数据和对数据进行解释的科学。

1.1.1 生活中的统计学

从体育到政治，再到商业领域，统计学几乎每天都存在于我们的生活中。现在我们来说明一下统计学在何时应用以及如何应用。上幼儿园时，可能你最关心玩得是否开心，能否交到朋友，但是你的老师是怎么想的呢（见图1-5）？

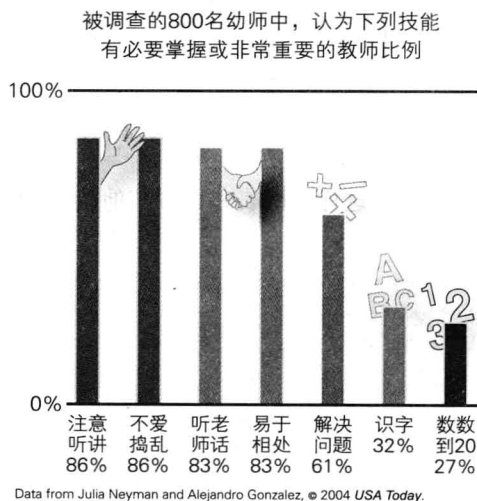


图1-5 即使在幼儿园，社交技能也非常重要

看图1-5给出的信息。它描述了老师认为小朋友有必要掌握或非常重要的技能。这是对800名幼师（只是全部幼师中的一小部分）进行调查后，绘制出的技能百分比图。排在最前面的是“注意听讲”和“不爱捣乱”。800名幼师中有86%认为这两项技能很关键或非常重要。加总所有百分比后你会发现，总和超

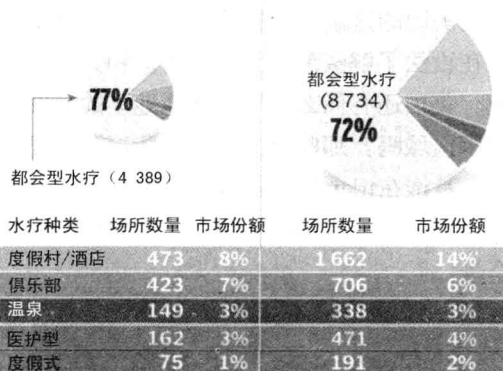
娇养放纵的人也能创造利润

不同类型水疗业的增长率

过去4年间,水疗业激增113个百分点。尽管都会型水疗的市场份额有所下降,但它仍高于其他水疗种类。

2000年水疗业总收入

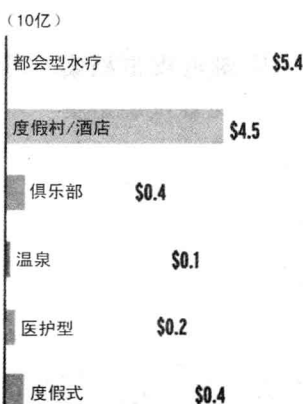
2004年*水疗业总收入



由于四舍五入,百分比之和可能不是100。 * 7月完成该项调查

2003年不同类型水疗业的收入

都会型水疗及度假村和酒店式水疗在该行业占据90%的收入。



有趣的事实

2003年不同性别的消费比例



■ 2003年美国消费群体达1.36亿人

■ 8 102万人选择了都会型水疗

■ 水疗业已经超过游乐园、主题公园和电影院,成为美国第四大娱乐产业

制图: 美术设计师Kevin M. Smith

From Rochester Democrat and Chronicle 12/5/2004. Reprinted with permission.

图 1-6

过100%。显然,这是因为允许老师们同时给出多个答案。

除了描述我们关心的事情外,统计学还可以揭示人们生活方式的转变。比如,蓬勃发展的水疗业。据国际SPA协会的报告,“娇养放纵”的人也能创造利润。在2000~2004年,水疗/美容业的收入增长了113%。实际上,水疗业已经超过游乐园、主题公园和电影院,成为美国第四大娱乐产业。

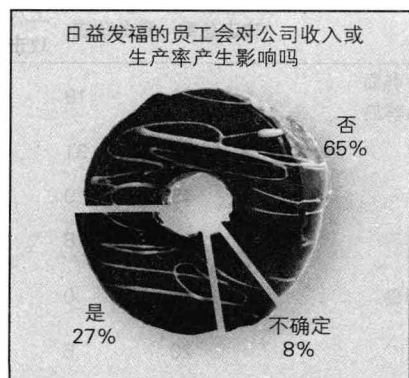
图1-6给出了更多有关水疗业的信息。想一想,需要收集哪些信息来制作图表——除

水疗场所的数目外,还要有水疗的种类和顾客的性别。但是,到哪里去收集这些数据呢? 请注意那些披露出来的统计资料的来源。根据《今日美国》的这则报道,其数据来源于国际SPA协会。该协会是全世界水疗界公认的专业、权威的信息发布机构。

媒体是传递统计信息的一个重要工具。特别是报纸,用各种图表告诉我们各类团体和人们在想些什么。你可曾想到,我们被这些文章直接影响的程度有多大?

图1-7表明,65%的公司都不担心日益发福的员工会对公司收入或生产率造成影响。这一信息是从哪里来的? 请注意资料来源。两位研究人员是怎么收集数据的? 他们是在调查了450名商界和政界人士后,才制作出这个图。给定误差为±5个百分点。根据这一信息,60%~70%的公司并不担心肥胖的员工会影响公司收入或生产率。这好像很让人吃惊,因为新闻里常说肥胖会影响健康,还会浪费人们很多金钱和精力去减肥。

即使是声誉良好的媒体定期发布的统计数



Data from Darryl Haralson and Alejandro Gonzalez, © 2004 USA Today. 给定误差为±5个百分点

© Mike Kemp/Rubberball/Getty Images

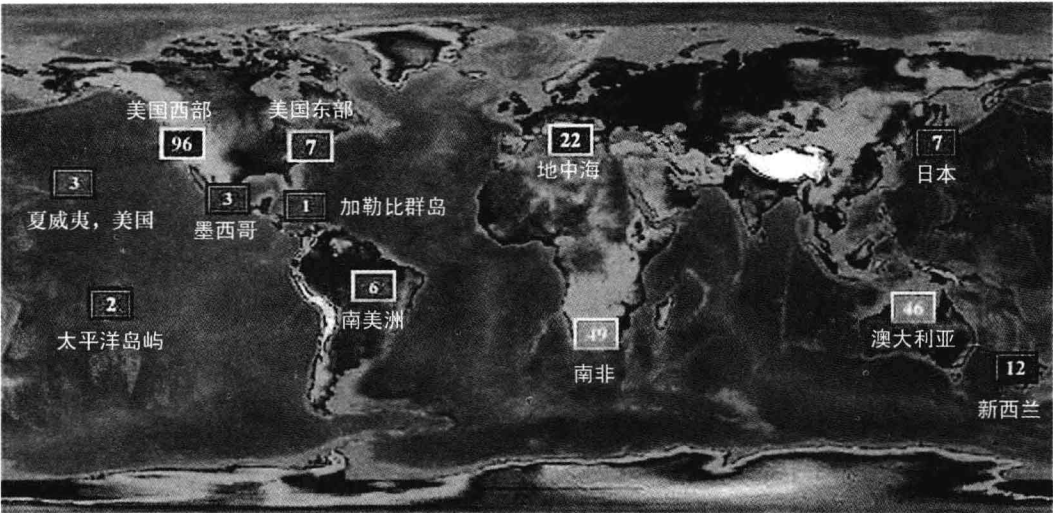
图1-7 公司关心工人的体重吗

字，也不意味着你可以盲目地相信。别忘了看一下统计资料的来源，并确信你看完了全景。此外，一定记住：正确运用统计方法的前提是掌握大量的常识。

由美国软骨鱼类学会和佛罗里达自然历史博物馆共同管理的国际鲨鱼攻击档案（ISAF），收录了所有已知的鲨鱼袭击案例。图1-8和表1-1展示了相关信息。

常识能帮我们做出什么判断呢？利用常识和鲨鱼攻击图，再喜欢大海的人，也要做出远离美国的决定。全球近2/5的鲨鱼袭击发生在美国。美国水域肯定到处是鲨鱼，那些鲨鱼

也一定疯了！常识——记得吗？这幅图是不是略带误导性？还有什么因素会影响这里的统计数字？首先，必须弄清楚一个国家或一个洲的海岸线有多长。其次，是谁在追踪这些攻击？此处表1-1已指明——佛罗里达自然历史博物馆，是美国的这家博物馆在追踪。很明显，美国正在设法了解鲨鱼的无端攻击。和其他地区相比，美国还有什么不同？其他地区的海域属于休闲区域吗？那些地方的经济状况怎么样？还有，是谁在记录那些鲨鱼袭击呢？记住，在看统计报告时看清资料来源，并确信你看到的是全景。



资料来源: http://www.sharkattackphotos.com/Shark_Attack_Map.htm.

图1-8 国际鲨鱼攻击档案（ISAF）

表1-1 佛罗里达自然历史博物馆的鲨鱼攻击数据

地区	攻击总数	致命攻击	上次致命攻击时间	地区	攻击总数	致命攻击	上次致命攻击时间
美国（不包括夏威夷）	761	39	2004	安的列斯群岛和巴哈马群岛	59	19	1972
澳大利亚	294	134	2004	中美洲	58	31	1997
非洲	264	69	2004	新西兰	45	9	1968
亚洲	116	55	2000	欧洲	38	18	1984
太平洋/大洋洲岛屿（不包括夏威夷）	114	47	2003	百慕大群岛	4	0	
夏威夷	100	15	2004	其他地区	20	6	1965
南美洲	96	22	2004	全球	1 969	464	2004

资料来源: <http://gerber.iwarp.com/Attack/GAttack/World.htm>.

1.1.2 统计学的用处

统计学的使用不受任何限制。我们可以找到统计学扮演整个角色的领域，但要想出完全不用统计学的领域却难得多。下面我们给出几个应用实例。

- 教育学频繁地使用描述统计来说明考试结果。
- 在科学领域，研究人员必须收集和分析实验数据。
- 政府随时都在收集各种统计数据。实际上，美国政府可能是全世界最大的数据收集者。

仔细研究统计结果并得出适当的结论是统计分析过程一个非常重要的环节。这些结论必须被精确地传递给读者，如果不把结果奉献出来与他人分享，研究就没有任何意义。我们的生活中到处都有统计报道：报纸、杂志、收音机和电视，所以，我们能够看到或听到各种新的研究成果，特别是与健康有关的很多研究成果。

1.1.3 统计语言

要继续深入学习统计学，我们就要了解它的术语。除描述统计和推断统计外，统计学还包括其他需要被定义和解释的术语。总体就是其中一个最基本的概念。

总体 (population) 是样本收集者感兴趣的所有个人或物体的集合。我们必须认真定义感兴趣的总体，只有在确定了每个个体以后，才能充分定义总体。“在美国上过大学的所有学生”就是一个被明确定义的总体。

一般情况下，我们常把总体看做是人的集合。但是，动物、植物、制成品等的集合也都可以构成统计总体。例如，所有的加州红杉树就是一个总体。

总体可以被分成两类：有限总体和无限总体。当总体中的个体可以（或者可能）被一个一个罗列时，该总体是有限的 (finite)。如果个体有无穷多个，该总体就是无限的

(infinite)。大学图书馆的藏书构成有限总体；在线公共检索目录 (OPAC) 可以列出每一本英文书。美国所有登记在册的选民构成一个非常庞大的有限总体；如果有必要，可以编制美国所有选区全部选民的综合名单。另一方面，可能服用阿司匹林的人和希尔瓦尼亚公司将要生产的40瓦灯泡则构成无限总体。研究庞大的总体会比较困难，因此，我们常常需要抽样，即抽取总体的子集，然后分析样本数据。**样本 (sample)** 是样本收集者从总体中选取的个人、物体或其他观测量的集合。

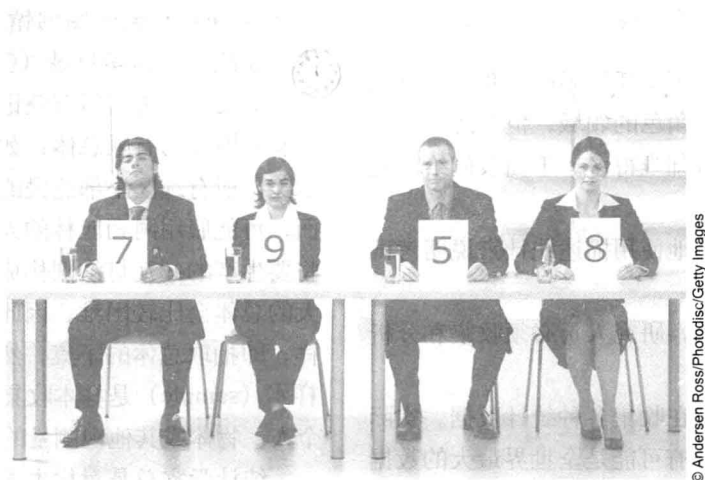
统计学家总是对样本或总体中的特定变量 (variable) 感兴趣。也就是说，他们常常要去研究总体或样本中个体的一个或多个特征。比如，年龄、发色、身高、体重就是变量。总体或样本的每一个体都有相应的变量值。每个个体的变量值被称做**数据值 (data value)**，可能是一个数字、一个词或一个符号。例如，比尔·琼斯“23岁”上大学，当时他长着一头“棕发”，身高“71英寸”，体重“183磅”。这四个值就是琼斯在四个变量上的数据值。

样本中每一个体在某个变量上的全部取值，叫做**数据 (data)**。25名学生的25个身高（或25个体重、年龄和发色）就是一组数据。统计学家通过**实验 (experiment)** 来收集数据，实验是能产生一组数据的经过设计的活动。实验既包括抽取个体的安排，还包括获取数据的活动。

大学生的“平均”入学年龄和入学时超过21岁的“比例”是总体参数的两个例子。**参数 (parameter)** 是描述整个总体某个特征的值。我们常用希腊字母来表示参数。等学到这些参数时我们再介绍这些字母。

每个参数有对应的样本统计量。**统计量 (statistic)** 是描述样本某个特征的数值，参数描述的是总体特征，而统计量则描述样本的特征。

25名学生的“平均”身高就是一个样本统计量，它是描述样本的指标。多数样本统计量



© Andersen Ross/Photodisc/Getty Images

都可以用公式算出，并且用特定的英文字母来表示（比如， $\bar{x}$ ， s 和 r ）。

供您参考

参数描述总体特征，统计量描述样本特征。

每种类型的变量（定量和定性）可再进行细分，如图1-9所示。

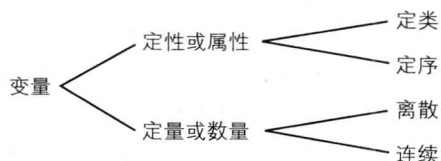


图 1-9

1.1.4 变量的分类

变量大体上可分为两类：①定性变量，对总体中的个体进行定性描述和分类；②定量变量，对个体特征进行量化反映。

对包含4位美发顾客的样本进行调查，询问他们的“发色”、“家乡”和“满意度”。这三个变量都是定性变量（属性变量），它们描述了人的某个特征，具有相同特征的人都属于这一类别。收集到的数据是{金发、棕发、黑发、棕发}，{布莱顿、哥伦布、奥尔巴尼、杰克逊维尔}，{非常满意、满意、尚可}。

本学期学生们购置教材的“总费用”则是定量变量（数值变量）。收集到的样本数据如下：238.87美元、94.57美元、139.24美元。

〔计算“平均费用”，只需将3个数字加总后除以3： $(238.87+94.57+139.24)/3=157.56$ （美元）〕。可见，像加法和求平均值这样的算术运算，对定量数据是有意义的（对定性数据则不适用）。

可以根据定性或属性变量的特征将其划分为定类变量和定序变量。定类变量（nominal variable）是描述（或描绘、陈述）总体中个体类别特征的定性变量。对定类变量做算术运算没有任何意义，即使排序也不具备可操作性。

对4位美发顾客进行调查时，“发色”和“家乡”这两个变量就属于定类变量，它们仅仅刻画了人的某些特征，对这些变量值加总后除以4来求样本平均值毫无意义。比如，（金发+棕发+黑发+棕发）/4是无法被解释的。此外，发色和家乡也不能进行排序。

定序变量（ordinal variable）是体现次序位置或等级的定性变量。对4位美发顾客的调查数据中，“满意度”这个变量就是定序变量，因为它确实体现出了等级次序：“非常满意”排在“满意”的前面，“满意”排在“比较满意”之前。定序变量的另一个例子是，根