

HZ BOOKS
华章教育

Pearson

统计学精品译丛

商务统计学

Business Statistics



[美] 罗伯特 A. 唐纳利 著
Robert A. Donnelly, Jr.

高璞 徐园植 林艳 译

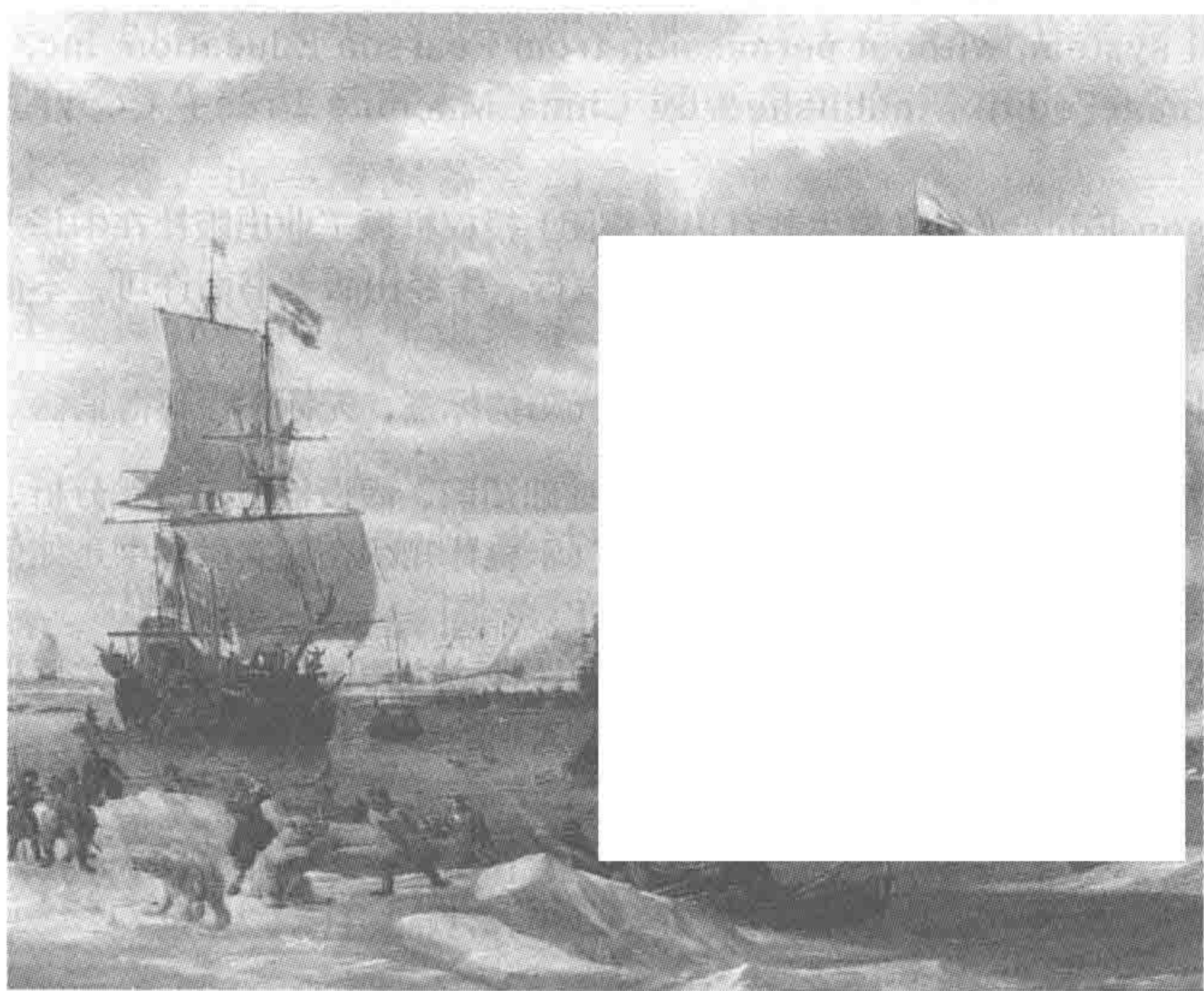


机械工业出版社
China Machine Press

统计学精品译丛

商务统计学

Business Statistics



[美] 罗伯特 A. 唐纳利
Robert A. Donnelly

高璞 徐园植 译



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

商务统计学 / (美) 罗伯特 A. 唐纳利 (Robert A. Donnelly, Jr.) 著; 高璞, 徐园植, 林艳译.
—北京: 机械工业出版社, 2018.2

(统计学精品译丛)

书名原文: Business Statistics

ISBN 978-7-111-58992-1

I. 商… II. ①罗… ②高… ③徐… ④林… III. 商业统计学—高等学校—教材 IV. F712.3

中国版本图书馆 CIP 数据核字 (2018) 第 013963 号

本书版权登记号: 图字 01-2013-1649

Authorized translation from the English language edition, entitled Business Statistics, ISBN: 9780132145398 by Robert A. Donnelly, Jr., published by Pearson Education, Inc., Copyright © 2013 by Pearson Education, Inc.

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

Chinese simplified language edition published by China Machine Press, Copyright © 2018.

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 (不包括香港、澳门特别行政区及台湾地区) 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。

本书是一本统计学入门教材, 内容涉及描述统计学、概率论基础知识、假设检验、方差分析、卡方检验、回归分析和预测等。作者以对话式风格来讲述主要内容, 展示统计知识在现实世界中是如何应用的。“思考题”贯穿全书, 鼓励学生利用所学知识求解问题。

本书适合作为经济管理类本科生的统计学教材以及继续教育机构、培训机构、MBA 相关课程的教材, 也适合作为从事会计、市场营销、银行、金融服务等行业专业人士的参考书。

出版发行: 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: 和 静

责任校对: 李秋荣

印 刷: 北京市兆成印刷有限责任公司

版 次: 2018 年 3 月第 1 版第 1 次印刷

开 本: 186mm × 240mm 1/16

印 张: 42.75

书 号: ISBN 978-7-111-58992-1

定 价: 139.00 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) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

译者序

“统计”是个啥？不就是一门数学类的课程嘛！反正就是一堆概念定理加公式呗！我最头疼的就是这个了，太枯燥！别急别急，这本书可是《商务统计学》，理论联系实际，绝对可以颠覆你对统计学刻板枯燥的认知。具体有啥看头，请听我娓娓道来。

罗伯特 A. 唐纳利真不简单，他这本书包含一般统计学应该涉及的知识点，最最关键的是，你在读他有趣故事的同时已经把这些知识点理解了（相信离应用就不远了）。我们会追随罗伯特来到他的生活中，看看他如何用统计学知识来解决他的家庭生活（妻子黛比出镜率很高）以及社会生活（他工作学校的学生、具有时代特征的美国焦点问题也在其中）中的实际问题。当我们漫步在罗伯特的生活场景中，欣赏他呈现的一个个统计故事时，常常会因他内心独白中的幽默而会心一笑（哈哈大笑也是有可能的）。一本统计学读完犹如欣赏了一部美式大片，还学会了使用统计学这个工具，真是文理兼修。

此外，罗伯特在书中还非常用心地设计了很多细节来帮助读者学习（请看前言）。至于学习中需要的学生资源和教师资源，那更是配备得相当齐全。同时，对于解决统计学问题所需的技术支持，也一应配备妥当。

一个《商务统计学》入门级读者所需的万事已经具备，就差您捧起一读这个东风啦！

不知不觉间这篇译者序也秉承了本书的文风（看到本书的代入感威力了吧）。好吧，下面回归正常理科教师风。

作为一名统计学教师同行，深知学生在学习中存在学不懂只会机械做题的焦虑，这本书恰恰可以让大家在轻松愉快的过程中学会统计学原理，并掌握统计学与商务问题等实际问题的联系，从而可以灵活地学以致用。希望它能给广大被统计学折磨的读者带来曙光，从而爱上统计这门学科，并能利用它解决工作中的问题。

全书共 16 章，由高璞、徐园植和林艳 3 人合作完成。感恩我们能遇到这样一本出色的书并有幸成为它的译者，我们尽量保持原书风格进行翻译，但很多地方可能仍难以准确表达作者描述的意境。限于水平，译文不可避免会有不完善的地方，欢迎广大读者提出宝贵意见，不胜感激！

感谢家人及朋友对我们在翻译过程中的帮助和支持！

高璞于北京西山

前 言

本书是一到两个学期的教材，以一种对话的风格写作，以期减少商科学生在学习统计课程时的焦虑。

现在有太多商科学生被他们的统计教材吓到了。这些学生常常把他们的教材看作是要跨越的障碍，而不是帮他们成功的工具。针对这个问题，我用一种直白的、对话的风格编写本书。作为作者和教师，经验告诉我当学生感觉他们与老师有个人联系时，学习效率更高。许多传统的教材总是“高高在上”，让很多学生感觉很难理解。我更喜欢教材与学生“对话”，就像我在课堂上一样，向学生提供一种我与他们在一起的场景，在每一个学习过程中鼓励他们。

我深信学生在学习新概念时解决统计问题比起之后再解决（常常是下一次考试前），学习效率会更高。为了奉行这种理念，我给出了标为“思考题”的平行问题，这样学生就可以对照相应章节的例题完成这些题目。我试图用一种轻松的语气促使学生完成这些题目，但这并不意味着我祈求学生完成。我在每章末都给出了思考题的完整答案[⊖]，这样学生就可以快速地检验所得结果的正误。我称此为“学习，练习，检验”循环，在这个循环中，学生通过读例题学习，自行完成类似的习题，最后检查结果来验证是否真的学会了。实际上，本书也起着工具书的作用，使学生能积极地参与到学习过程中来。学生常常跳过书中过程完整的例题，感觉自己已经理解了概念，直到接下来的考试中遇到类似的问题而不会解时才发现自己没有学会。我的方法鼓励学生仔细阅读例题并在他们进入下一个知识主题时能确保已经掌握了学过的概念。

我在每章都会插入一些注释或提示，这些内容给出了有用的深刻见解。这种特色与老师上课时为帮助学生理解而准备的讲义类似。我发现这是在学生困惑或难以应对时帮助他们集中注意力的有效方法。

本书特色

- **当前的商务例题，能引起学生的兴趣**——学生能够了解统计过程在产品和服务上的应用，例如：
 - 英国石油公司在墨西哥湾的漏油事件；
 - 金宝保险公司对享受优秀学生折扣和不享受优秀学生折扣客户车祸比例的比较；
 - 对不同品牌智能手机满意度分数的比较；
 - 使用原来的 IPv4 造成的 IP 地址短缺。
- **每个主要小节之后的思考题**——这些问题被战略性地放在各章中以加强对关键概念的理解。这些问题的完整答案可在每章末找到。我发现学生主动地解题与他们跳过解答完整的例题相比，学习效率更高。

⊖ 限于篇幅，思考题答案放到了华章网站(www.hzbook.com)上。——编辑注

- **注释或提示**——在图注、表注、脚注及正文中，你会发现帮助阐明特定主题的注释或提示。这些内容常常指向文章中某个位置，与老师在课堂上为学生准备的深入见解类似。
- **实践中的统计学**——在本书中，关于统计学如何应用在当今商业环境中的例子被放在章首或正文穿插的线框中。例如：
 - 政府使用置信区间以及相应结果解释对失业数字的报道；
 - 对赫尔食品公司不同零食的顾客反馈的比较；
 - 医疗保险公司调查诊疗室反常计费使用的统计方法；
 - 对奥运会运动员在不同体育赛事上表现的比较。
- **复杂统计过程的分步方法**——很多学生在面对像假设检验、方差分析(ANOVA)以及回归分析这样的复杂过程时都会“迷失在森林中”。我的办法是将这些过程分成可以接受的重复性的步骤，这些步骤可以用来解决各种问题。这样，学生在用某种方法解题时就有一个可以遵循的路径图来决定怎么做。
- **贯穿全书的微软 Excel**——我用 Excel 说明商务统计学中各方法的应用，但该软件的使用不影响对概念的理解。我对那些告诉我会我在电脑上执行像 ANOVA 这样的过程而不会解释结果的学生如是说：如果让 Excel 完成所有的工作，你就会失去理解相应方法背后概念的机会。本书的理念是避免这种情况发生。

学生资源

学生解题手册——由 Goldey-Beacom 学院的 Bob Donnelly 和 Roman Erenshteyn 完成，该手册给出了所有偶数习题的详细解答。学生可以通过访问 www.mypearsonstore.com 搜索 ISBN 0-13-213568-X 购买。

在线资源——在线资源可从 www.pearsonhighered.com/donnelly 免费下载，包含下列内容。

- **数据文件**——这些是各章习题中的 Excel 数据文件。
- **在线案例文件**——由 Bob Donnelly 创建，这些案例可与各章配合使用，作为进一步的课堂讨论或课程项目、课后作业。
- **PHStat2**——适用于 Windows Excel 2003、2007、2010 的培生统计插件。该版本删除了 Excel 分析工具包插件，从而简化了安装和设置。

教师资源[⊖]

教师资源中心——网址为 www.pearsonhighered.com/donnelly，其中包含完整的教师解题手册、考题文件以及教学幻灯片。

- **教师解题手册**——由 Goldey-Beacom 学院的 Bob Donnelly 和 Roman Erenshteyn 完成，该手册给出了各节习题以及各章复习题的详细解答。电子答案在教师资源中心

⊖ 关于本书教辅资源，只有使用本书作为教材的教师才可以申请，需要的教师请联系机械工业出版社华章公司，电话 136 0115 6823，邮箱 wangguang@hzbook.com。——编辑注

致 学 生

信息过剩是我们的时代标签，日常的数据海啸并不会很快消退。毫无疑问，你的整个职业生涯都将面临空前庞大的数据。当周围的很多人叫嚷着“被淹没了！”甚至“有人落水了！”时，你有一条不同的、更好的路可走。学习组织和解释重要信息的技能是成功的关键，商业尤其如此。

本书将提供给你利用数据得出不错的商业决策的无价工具。比如，如果你是美国电话电报公司(AT&T)或威瑞森无线通讯公司的经理，你能根据每家公司的样本得出两家公司的掉线率是否显著不同吗？仅凭该数据就能影响你的商业管理吗？或者，你会想到进一步地用其他因素(比如掉线发生的地点)来分析吗？在学完第10章后，你就能分析这样的数据并能用较高的信度来回答这些问题了。

在你的整个商务职业生涯中，你将会发现不同水平的人做出的具有真实结果的决策会影响商业利润和人们的工作。专业的数据分析会给你一个令人惊异的视角来了解商务，而不做此分析，这种视角就不明显。

我为你们——亲爱的学生们——写了这本教材，主要目的就是帮助你们成功地学习这门课并能在今后的职业生涯中获得成功。我汇集了多年的教学方法，在这些经验基础上，我给出如下的建议来帮你们达到各自的目标：

- 要去上课。如果你不去上课，将会错过帮你掌握这门课的细节。不管教材写得再好，没有任何的教材能代替你的老师和课堂上的交流。说真的，一定要去上课！
- 充分利用每一章中的“思考题”。对这些问题的解答会加深对关键概念的理解，并能使自己知道是否完全掌握了书中的知识。每章末给出的答案会给你及时的反馈，所以我鼓励你在做完题目后再看答案，不要在没完成或不知道怎么做时偷看答案！
- 在考试前要尽可能多地做章节习题。商务统计学不是一门临阵磨枪管用的学科。“练习出成绩”的说法尤其适合这门课，做一下各种类型的题目有助于你建立学习的信心。我已将所有偶数题目的答案放在附录 B[⊖]中。

我希望你们能够感受到我所感受的商务统计学的价值，你在本课程中的所学会对你今后在商务世界中的成功有所帮助。

罗伯特 A. 唐纳利

⊖ 附录 B 可从华章网站(www.hzbook.com)下载。——编辑注

目 录

译者序	
前言	
关于作者	
致学生	
第 1 章 商务统计学简介	1
1.1 商务统计学及其应用	1
1.1.1 市场调研	1
1.1.2 广告宣传	2
1.1.3 商业运营	2
1.1.4 财政金融	2
1.2 数据	2
1.2.1 数据来源	3
1.2.2 两个主要数据类型	5
1.2.3 通过计量水平对数据分类	6
1.2.4 时间序列与截面数据	8
1.3 描述统计学与推断统计学	9
1.4 道德与统计学——数据的危险	
世界	11
复习题	12
第 2 章 描述统计学	14
2.1 技术在统计学中的作用	15
2.1.1 用 Excel 2010 执行统计	
分析	16
2.1.2 安装 PHStat2	17
2.2 描述定量数据	17
2.2.1 定量数据的频率分布	17
2.2.2 相对频率分布	18
2.2.3 累积相对频率分布	19
2.2.4 Excel 的 FREQUENCY	
函数	19
2.2.5 用直方图表示频率分布	19
2.2.6 直方图的形状	22
2.2.7 用分组定量数据构造频率	
分布	23
2.2.8 构造分组定量数据直方图	25
2.2.9 分组太多和太少的后果	26
2.2.10 离散数据还是连续数据	27
2.2.11 折线图	29
2.3 描述定性数据	34
2.3.1 定性数据的频率分布	34
2.3.2 条形图	35
2.3.3 帕累托图	40
2.3.4 饼图	42
2.4 列联表	46
2.5 茎叶显示	50
2.6 散点图	52
复习题	57
第 3 章 计算描述统计学	61
3.1 集中趋势的度量	61
3.1.1 均值	61
3.1.2 加权平均值	63
3.1.3 用均值归纳数据的优缺点	64
3.1.4 中位数	65
3.1.5 众数	67
3.1.6 频率分布的形状	68
3.1.7 用 Excel 计算均值、中位数和	
众数	69
3.1.8 选哪种集中趋势度量方法:	
均值、中位数还是众数	72
3.2 变异性的度量	73
3.2.1 极差	74
3.2.2 方差和标准差	74
3.3 共同使用均值和标准差	82

3.3.1 变异系数	83	第 5 章 离散型概率分布	144
3.3.2 z 值	84	5.1 离散型概率分布简介	145
3.3.3 经验法则	87	5.1.1 离散型随机变量	145
3.3.4 切比雪夫定理	88	5.1.2 离散型概率分布的规则	145
3.4 处理分组数据	91	5.1.3 离散型概率分布的均值	146
3.4.1 分组数据的均值	91	5.1.4 离散型概率分布的方差和 标准差	147
3.4.2 分组数据的方差和标准差	92	5.1.5 预期货币价值	152
3.5 相对位置的度量	94	5.2 二项分布	154
3.5.1 百分位数	94	5.2.1 二项试验的特征	154
3.5.2 四分位数	98	5.2.2 二项分布的应用	155
3.5.3 盒须图	100	5.2.3 二项分布的均值和标准差	159
3.5.4 五数概括法	102	5.2.4 二项概率表	160
本章主要公式	104	5.2.5 用 Excel 和 PHStat2 计算 二项概率	161
复习题	105	5.3 泊松分布	166
第 4 章 概率	109	5.3.1 泊松过程的特征	166
4.1 概率简介	109	5.3.2 泊松概率分布的应用	167
4.1.1 传统概率	110	5.3.3 用泊松分布计算到达 概率	169
4.1.2 经验概率	111	5.3.4 泊松概率表	170
4.1.3 主观概率	112	5.3.5 用 Excel 和 PHStat2 计算 泊松概率	171
4.1.4 概率的基本属性	113	5.3.6 用泊松分布近似二项分布	174
4.2 多事件的概率规则	116	5.4 超几何分布	177
4.2.1 事件的交集	116	5.4.1 计算超几何分布的概率	177
4.2.2 事件的并集	117	5.4.2 用 Excel 和 PHStat2 计算超 几何概率	180
4.2.3 加法法则	118	本章主要公式	183
4.2.4 条件概率	120	复习题	183
4.2.5 相互独立事件与相依事件	122	第 6 章 连续型概率分布	188
4.2.6 乘法法则	123	6.1 连续型随机变量	188
4.2.7 概率列联表	126	6.2 正态概率分布	190
4.2.8 互斥事件和相互独立事件	127	6.2.1 正态概率分布的特征	190
4.2.9 贝叶斯定理	128	6.2.2 用标准正态概率表计算正态 分布的概率	191
4.3 计数原理	133	6.2.3 重温经验法则	197
4.3.1 基本计数原理	134		
4.3.2 排列	136		
4.3.3 组合	137		
本章主要公式	140		
复习题	140		

6.2.4 其他正态概率区间	197	8.3 总体标准差 σ 未知时均值的置信	
6.2.5 用 Excel 计算正态概率	200	区间计算	265
6.2.6 用正态分布近似二项分布	203	8.3.1 学生 t 分布的使用	265
6.3 指数概率分布	208	8.3.2 用 Excel 和 PHStat2 确定 σ 未知	
6.3.1 计算指数概率	209	时均值的置信区间	269
6.3.2 用 Excel 计算指数概率	211	8.4 比例的置信区间计算	272
6.4 均匀概率分布	213	8.5 样本容量的确定	276
本章主要公式	217	8.5.1 估计总体均值时样本容量的	
复习题	217	计算	276
第 7 章 抽样和抽样分布	222	8.5.2 用 PHStat2 计算估计总体均值时	
7.1 为什么抽样	222	的样本容量	277
7.2 抽样类型	223	8.5.3 估计总体比例时样本容量的	
7.2.1 概率抽样	223	计算	279
7.2.2 非概率抽样	229	8.5.4 用 PHStat2 计算估计总体比例时	
7.3 抽样误差和非抽样误差	230	的样本容量	280
7.4 中心极限定理	233	8.6 有限总体置信区间的计算	282
7.4.1 中心极限定理的应用	237	本章主要公式	285
7.4.2 中心极限定理在检验断言中		复习题	286
的应用	238	第 9 章 单个总体的假设检验	289
7.4.3 样本容量对抽样分布的		9.1 假设检验简介	289
作用	240	9.1.1 假设的设定	289
7.4.4 有限总体均值的抽样分布	241	9.1.2 双尾假设检验	290
7.5 比例的抽样分布	246	9.1.3 单尾假设检验	291
本章主要公式	250	9.1.4 假设检验的逻辑	291
复习题	251	9.1.5 假设表述的设定总结	292
第 8 章 置信区间	253	9.1.6 I 型错误和 II 型错误的	
8.1 点估计	253	区别	293
8.2 总体标准差 σ 已知时均值的		9.2 σ 已知时总体均值的假设	
置信区间计算	254	检验	294
8.2.1 边际误差的计算	256	9.2.1 一个对 σ 已知时总体均值单尾	
8.2.2 置信区间的解释	257	假设检验的例子	295
8.2.3 置信水平的改变	259	9.2.2 假设检验的 p 值法: 单尾	
8.2.4 置信区间在商务中的应用	261	检验	298
8.2.5 用 Excel 和 PHStat2 确定 σ 已知		9.2.3 一个对 σ 已知时总体均值双尾	
时均值的置信区间	261	假设检验的例子	299
8.2.6 总体标准差 σ 已知时小样本均值		9.2.4 假设检验的 p 值法: 双尾	
的置信区间计算	263	检验	301

9.2.5	假设检验中 α 的作用	302	10.2.2	用置信区间比较两均值之差： 独立样本， σ_1 和 σ_2 未知，总体 方差相等	342
9.2.6	用 PHStat2 进行 σ 已知时总体 均值的假设检验	303	10.2.3	用 PHStat2 进行比较两均值之差的 假设检验：独立样本， σ_1 和 σ_2 未知，总体方差相等	343
9.3	σ 未知时总体均值的假设检验	306	10.2.4	比较两均值之差的假设检验：独立 样本， σ_1 和 σ_2 未知，总体方差 不相等	345
9.3.1	一个对 σ 未知时总体均值单尾 假设检验的例子	306	10.2.5	用 PHStat2 进行比较两均值之差的 假设检验：独立样本， σ_1 和 σ_2 未知，总体方差不相等	347
9.3.2	用学生 t 分布估计 p 值	309	10.2.6	用 Excel 进行比较两均值之差的 假设检验：独立样本， σ_1 和 σ_2 未知，总体方差不相等	349
9.3.3	一个对 σ 未知时总体均值双尾 假设检验的例子	310	10.2.7	对总体均值之差不为 0 的 检验	351
9.3.4	用 PHStat2 进行 σ 未知时总体 均值的假设检验	311	10.3	相关样本的假设检验	354
9.4	总体比例的假设检验	314	10.3.1	比较相关样本两均值之差的 假设检验	354
9.4.1	一个对比例单尾假设检验的 例子	315	10.3.2	用 Excel 进行比较相关样本 两均值之差的假设检验	357
9.4.2	对比例假设检验的 p 值法	316	10.3.3	用置信区间比较相关样本的 两均值之差	360
9.4.3	用 PHStat2 进行比例的假设 检验	317	10.4	独立样本的两总体比例 比较	362
9.5	II 型错误	319	10.4.1	用置信区间比较两比例之差	365
9.5.1	计算 II 型错误发生的概率	319	10.4.2	比较两比例之差的假设 检验	365
9.5.2	α 对 β 的作用	323	10.4.3	用 PHStat2 比较两总体 比例	367
	本章主要公式	326		本章主要公式	371
	复习题	326		复习题	373
第 10 章	比较两总体的假设检验	330	第 11 章	方差分析过程	378
10.1	比较两总体均值：独立样本， 总体标准差 σ_1 和 σ_2 已知	330	11.1	单因素 ANOVA：检验单因素 对总体均值的作用	379
10.1.1	比较两均值之差的假设检验： 独立样本， σ_1 和 σ_2 已知	332	11.1.1	用 Excel 进行单因素 方差分析	389
10.1.2	用 PHStat2 进行比较两均值 之差的假设检验：独立 样本， σ_1 和 σ_2 已知	335			
10.1.3	用置信区间比较两均值之差： 独立样本， σ_1 和 σ_2 已知	336			
10.2	比较两总体均值：独立样本， 总体标准差 σ_1 和 σ_2 未知	339			
10.2.1	比较两均值之差的假设检验： 独立样本， σ_1 和 σ_2 未知，总体 方差相等	339			

11.1.2 多重比较: 比较成对总体均值 (单因素 ANOVA)	391	第 13 章 关于总体方差的假设检验 ...	471
11.1.3 用 PHStat2 进行总体均值的 多重比较	394	13.1 单总体方差的检验	471
11.2 随机化区组 ANOVA: 通过对 第二因子分区组检验单因素的 作用	397	13.1.1 单总体方差的单尾假设检验 ...	471
11.2.1 用 Excel 进行随机化区组方差 分析	403	13.1.2 用 PHStat2 进行总体方差的 检验	474
11.2.2 要不要分区组: 这是个 问题	405	13.1.3 单总体方差的双尾假设 检验	475
11.2.3 多重比较: 比较成对总体均值 (随机化区组 ANOVA)	410	13.2 比较两总体方差	479
11.3 双因素 ANOVA: 检验双因素 对总体均值的作用	414	13.2.1 两总体方差的单尾检验	479
11.3.1 用 Excel 进行双因素方差 分析	415	13.2.2 用 Excel 进行两总体方差的 比较	481
11.3.2 双因素 ANOVA 的均方和 ...	417	13.2.3 两总体方差的双尾检验	483
11.3.3 解释双因素 ANOVA 的输出 结果	419	13.2.4 用 PHStat2 进行两总体方差的 比较	485
11.3.4 再论交互作用	422	本章主要公式	487
11.3.5 多重比较: 比较成对总体均值 (双因素 ANOVA)	425	复习题	488
本章主要公式	430	第 14 章 相关性与简单回归分析	491
复习题	431	14.1 因变量和自变量	492
第 12 章 卡方检验	434	14.2 相关分析	492
12.1 两个及以上总体比例的比较 ...	435	14.2.1 相关系数	493
12.2 判定观测频率是否服从已知 概率分布	443	14.2.2 用 Excel 计算相关系数	495
12.2.1 关于离散型概率分布的 检验	443	14.2.3 相关系数显著性的假设 检验	495
12.2.2 关于泊松分布的检验	445	14.3 简单回归分析	497
12.2.3 关于二项分布的检验	449	14.3.1 最小二乘法	498
12.2.4 关于正态分布的检验	452	14.3.2 斜率和 y 轴截距的计算	499
12.3 两个变量独立性的检验	459	14.3.3 用 Excel 计算斜率和 y 轴 截距	500
本章主要公式	466	14.3.4 分解平方和	502
复习题	466	14.3.5 决定系数的计算	504
		14.3.6 决定系数显著性的假设 检验	505
		14.4 利用回归进行预测	508
		14.4.1 值 x 已知时关于 y 的平均值的 置信区间	508
		14.4.2 值 x 已知时关于特定 y 值的 预测区间	510

14.4.3	用 PHStat2 计算置信区间和 预测区间	511	15.4	定性自变量的运用	552
14.5	回归方程斜率的显著性 检验	513	15.5	建立模型	558
14.6	关于回归分析的假设	517	15.5.1	多重共线性	558
14.7	一个相关系数为负时简单线性 回归的例子	520	15.5.2	一般逐步回归分析法	563
14.7.1	相关系数	521	15.5.3	最小子集回归分析法	565
14.7.2	斜率和 y 轴截距的计算	522	15.5.4	其他的选择法	569
14.7.3	相关系数为负时分解 平方和	523	15.5.5	残差分析	569
14.7.4	决定系数的计算	524	15.5.6	建模小结	572
14.7.5	相关系数为负时计算置信 区间和预测区间	524	本章主要公式		574
14.7.6	相关系数为负时回归斜率 显著性的检验	526	复习题		575
14.8	一些非常重要的最终想法	528	第 16 章 预测		578
	本章主要公式	529	16.1	预测简介	578
	复习题	529	16.2	平滑预测法	580
第 15 章	多元回归与建模	533	16.2.1	简单移动平均法预测	580
15.1	多元回归模型的构造	533	16.2.2	加权移动平均法预测	583
15.1.1	确定回归系数	534	16.2.3	指数平滑法预测	585
15.1.2	利用回归模型进行 预测	536	16.2.4	指数平滑法的趋势调整 预测	587
15.2	解释因变量的变异	539	16.3	回归分析预测	592
15.2.1	多重决定系数	539	16.3.1	趋势外推法	592
15.2.2	回归模型显著性的检验	541	16.3.2	用 PHStat2 进行趋势外推	596
15.2.3	修正的多重决定系数	543	16.3.3	自相关的检验	597
15.3	关于自变量的推断	547	16.4	季节性预测法	605
15.3.1	回归系数的显著性检验	547	16.4.1	时间序列的乘法分解法	605
15.3.2	回归系数的置信区间	549	16.4.2	用虚拟变量代表季节性	614
			本章主要公式		621
			复习题		621
			附录 A 常用表格		624
			单位换算表		668
			思考题答案 [⊖]		
			附录 B 偶数复习题答案 [⊖]		

⊖ 思考题答案和附录 B 可从华章网站(www.hzbook.com)下载。——编辑注

第1章 商务统计学简介

欢迎大家来到统计学的世界！尽管你们当中可能有一部分人听说要学习统计学会很激动，另一部分人可能就没那么兴奋了。可能你来上“统计”课只是出于专业培养要求而已。不过，在你把这次学习机会的价值抹掉之前，我们先来讨论一下统计学在你生活中发挥的作用。在今天的世界上，每个人都是统计学的使用者。我的意思是，在你周围充斥着各种各样的数据和报表，它们影响着你的购物意愿、表决意向，或者对某个事件的看法。比如下面的例子：

- 福克斯体育台宣布有 1.11 亿观众收看了 2011 年超级碗。你知道算出这个数字的方法吗？（福克斯体育台如何知道你、我或者其他谁看了这个节目？）
- 电视上说总统奥巴马的支持率是 56%，旁边有小字 $\pm 4\%$ ，你知道这个百分率的含义吗？
- 当你看到一则新产品广告说五个医生当中有四个都推荐它，你会不会质疑这个声明的有效性？（比如，医生是否收取代言费了？）
- 某在线调查结果显示，佳能品牌的数码相机比尼康更受欢迎。你是否会有这样的顾虑，该调查的大部分受访者均为佳能支持者且重复投票，从而导致了结果的偏差？或者该调查是由佳能用户的论坛发起？


现在存在于你周围的数据和信息量是人类历史上前所未有的。我们的感受、想法以及在个人生活和职业生涯中做出的决策都受到统计学的强烈影响。因此，我们所报道的统计数据必须准确、客观和公正，以确保它们被正确地使用，这一点非常重要。

1.1 商务统计学及其应用

统计学是一门数学科学，对数据进行收集、分析和演示，之后这些结果将作为推理和归纳的基础。商务统计学是应用在商业世界中的统计学，协助改善人们在不同领域中制定的决策，例如，市场、运营、金融和人力资源等。下面我们就来看几个商务统计学是如何帮助到企业中的决策制定者的。

1.1.1 市场调研

企业会用市场调研的方法来判断消费者的喜好，这在很大程度上依赖于商务统计学的知识。例如，Kellogg 公司想通过口味测试来判断消费者对于奶酪小饼干的喜好，是喜欢 Kellogg 的 Cheez-It，还是喜欢 Nabisco 的 Cheese Nips。（作为一个忠实的 Cheez-It 爱好者，我知道我的选票花落谁家。）收集参与口味测试的消费者信息，Kellogg 便可以判断出对这两种具有相似特征的产品来说消费者的喜好在哪里。调研结果可以为 Kellogg 以后的市场推广提供非常有用的信息。

 20 世纪 80 年代，万豪酒店曾经就酒店现有产品在潜在客户中进行了一次广泛的调查。根据调查结果，万豪酒店设计出一种新型连锁酒店——“万豪庭院酒店”，并且获得了巨大的成功。

1.1.2 广告宣传

电视网络根据收看观众的多少来制定商业广告费率。电视网络获取数据的来源是尼尔森媒体研究，他们在全美大约收集了 25 000 份用户信息。因为样本用户经过了精心的挑选，所以结果信息可以反映全美收视习惯。在这里，正确地使用统计学知识保证了样本的正确选择以及数据能够转化为对电视网络有用的信息。有了这样的信息，福克斯向 2011 年超级碗 30 秒的广告收了 300 万美元的广告费。

1.1.3 商业运营

统计学能帮助商业运营得更好。我们用所有成功企业都极为关注的质量控制来举例说明。商务统计学和质量控制是天生一对。正确实施统计学可以帮制造业和服务业企业监控整个流程并判断出现质量问题的时间。这样 Kellogg 就可以监测出我的 Cheez-It 什么时间可能烤过头或者盐放多了。（据我正在采样的这个包装盒，我可以说这家公司的统计质量管理做得非常棒。）

1.1.4 财政金融

当我开始写这本书的时候，美国正经历着严重的经济衰退。导致经济衰退的部分原因，是银行的不良贷款，特别是抵押贷款。如果能够正确地使用统计学，它将会成为极好的工具来帮助银行区分具备和不具备良好信贷风险的消费者，判断依据是收入水平、教育水平和住房所有权等。

1.2 数据

数据是统计学领域的基础，可被定义为指定的观察值和测量值。如果我想收集我太太打鼾习惯的数据，我可以选择多种方法。我可以计算黛比在 10 分钟之内打鼾的次数。我也可以记录每次打鼾的时间长度。我还可以用短语来形容每次打鼾的声音强度，比如“听起来像刚从冬眠中醒来的熊”或者“哇！听起来像呼叫孩子的海狮”。（一个穿 2 号牛仔裤的人如何发出这样的声音是我无法理解的。）

在每种情况下，我对同一事件用不同的形式来记录。第一种情况，我记录的是频率，或者发生次数。第二种情况，我记录的是持续时间，或者时间长度。在最后一种情况，我记录事件的方法是用文字而不是数字。

然而，单独的数据没有特别的用途。根据定义，数据是关于测量值的原始事实和数字。而**信息**，是为了制定决策而从事实中推导出来的。我们使用统计学的一个主要原因是要把数据转换为信息。例如，表 1-1 是两个月以来我的高尔夫球得分。（供不打高尔夫球者参考，分数越低成绩越好。）

每个高尔夫球分数都可以作为一个数据点。单独来看的话，这些数据点除了可以表明我不应该退出我的日常工作之

表 1-1 高尔夫球得分数据

日期	得分
6 月 13 日	94
6 月 20 日	96
6 月 27 日	93
7 月 10 日	89
7 月 16 日	86
7 月 24 日	89

外，价值有限。（如果要放弃日常工作转做职业高尔夫球手，我的得分必须在 60~70 分范围

内。)让我再加上一点,就是我迫切地希望能通过买新设备来提高我的高尔夫球成绩。我在7月1日的时候悄悄地买了一套全新的球杆,我认为太太不会发现我的球包里面多了新的高尔夫球杆(她发现了)。通过这些数据,我们可以用统计学推断,购买新球杆确实让我的成绩提高了(想象一下黛比脸上布满的怀疑)。如果要回答这个问题,我们要用到本书后面章节涉及的统计分析知识。请继续关注我是如何用统计学来说服我太太我是真的需要这套球杆。

1.2.1 数据来源

我们把数据分成两大类:初级源数据和次级源数据。次级源数据是由他人收集,并供人使用的数据。美国政府收集并且发布了各类数据,可以在线查看。比如,美国劳工部收集了关于消费价格、通货膨胀、失业率和生产力等各类海量的数据。该部门的网站首页如图1-1所示。

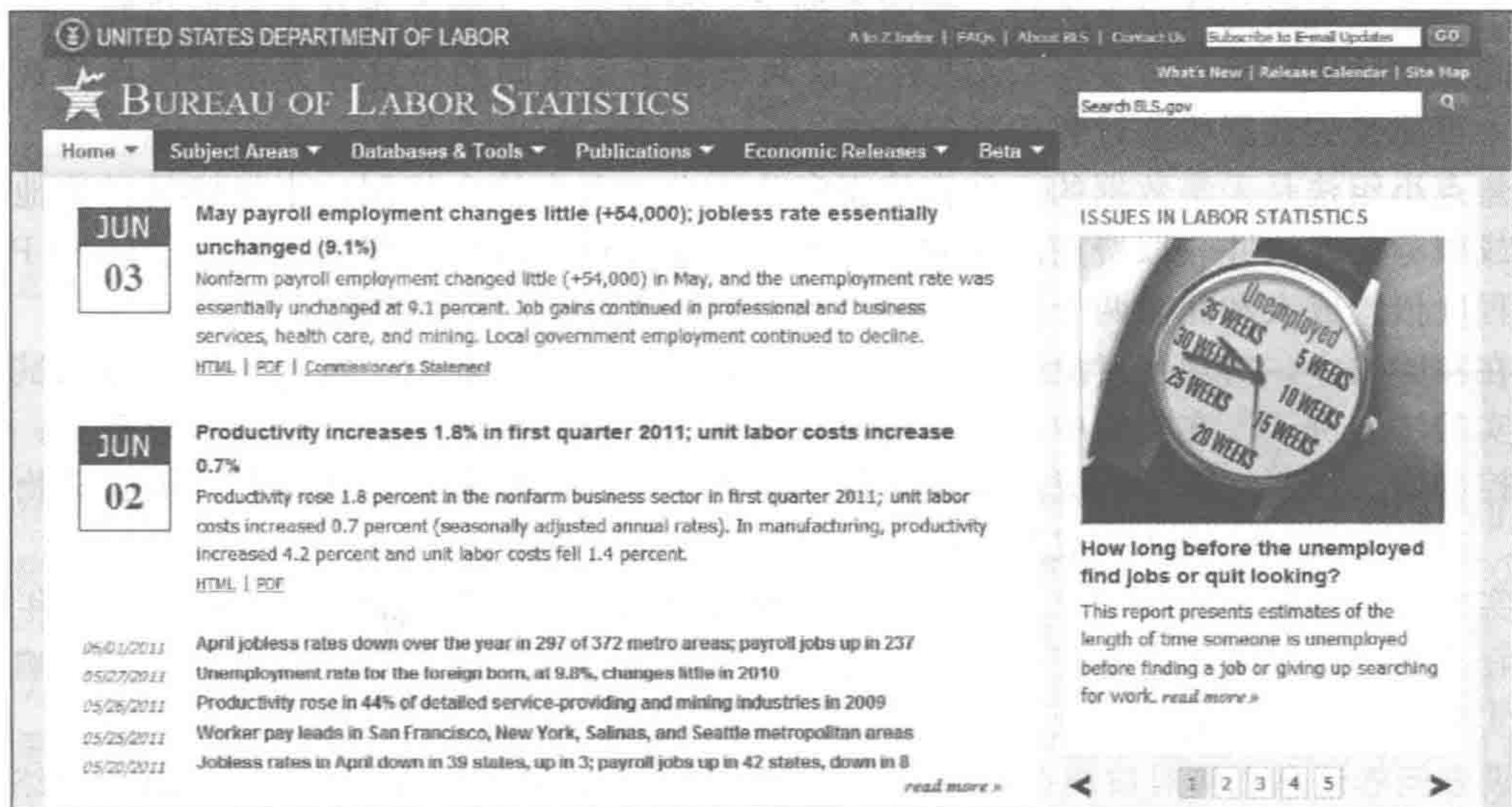


图 1-1 美国劳工统计局主页

每10年美国商务部进行一次全国性普查,收集全国人口的各种相关数据。国会使用数据结果来决定全美各州社区服务经费。同时,这些数据也用来调整各州在国会中的代表人数。普查数据结果在商务部网站上公布。

👉 各行各业在互联网发布的信息已经成为统计数据的丰富来源。你在一个典型互联网搜索引擎上显示的99 999个站点中糊里糊涂地找到一个网站,上面都有可能存在有用的东西。有一次,我就发现了一个日本作者研究氟对蟾蜍胚胎的影响。在这个发现以前,我根本不知道蟾蜍有牙齿,更不要说腔内问题了。

美国地质调查局提供各种各样琳琅满目的科学信息,令人叹为观止,这些信息可以用来管理整个地球上的水、能源、生物和矿产资源。比如说,你知道世界上的松鼠有250种吗?如果你不相信,可以到美国地质调查局的主页上去查查,那么你就成为你所在区域的

松鼠“专家”了。

使用次级源数据的最大问题在于你无法控制所收集到的数据。人们倾向于相信打印出来的东西，即使那不是真的。（你相信我，对吗？）这里面有一些是错的，而且，在本章的后面你会发现，有一些是故意弄错的。次级源数据的优势就是它们便宜（有时是免费的），而且它们立刻可用。如果有人想快速找到数据，那么次级源数据能马上满足他们的意愿（当然假设数据精确）。

初级源数据，从另一方面，是由个人或者组织收集并最终使用的数据。获取初级源数据的代价不菲，最大的优势在于这是你自己的数据。如果你把这些数据弄得乱七八糟，那么除了你自己之外没有别人可以责备。获取初级源数据的方法有很多种，可以通过直接观察法、实验法或者调查问卷法。


直接观察法是采集数据的一种方法，在观察对象（兴趣主体）处于自然环境中，通常在它们不知道被观察的情况下进行。举两个例子，观察野生动物在森林中跟踪猎物，观察星期五晚上购物中心里的青少年。（或者，这两个例子是相同的？）这种方法的优点是观察对象不太可能受到数据收集过程的任何影响。

焦点小组法是采集数据的一种直接观察法，由一名主持人控制，小组成员讨论他们对产品或服务的看法并收取费用。比如，为了获得对新玩具的反馈意见，费雪（Fisher Price）很大程度依赖成人和儿童两个焦点小组。参与者知道自己正在被观察。

在**实验法**中，对观察对象做不同的处理，记录感兴趣的数据。新的医疗药品测试就是一个实验法的例子。在实验中，把参与者分成两组：第一组拿到的是新的医疗药品；第二组是控制组。控制组里面的成员被告知他们拿到了新的医疗药品，实际上是不含药物的安慰剂。通过测量和比较两组成员的反应，判断新药是否有效。

实验的好处是让统计学家来控制可能影响结果的各项因素，像参与者的性别、年龄和教育程度等。由于受访者清楚他们正参与着一项研究，所以他们的反应可能受到影响，这是在实验中收集数据的主要顾虑。

调查问卷法是直接对被调查人询问一系列问题，可以采用的方式有电子邮件、网络、邮政信件、面对面或者电话调查。（我最“喜欢”的方式是电话调查，特别是我接到电话的时候是我刚坐下来要吃饭的时候，或者要去洗澡的时候，或者手头这章的工作终于有了进展的时候。）问卷问题必须要仔细设计，避免任何可能影响到参与者反馈或者让参与者困惑的偏见问题。有的问卷问题设计的方式鼓励或者引导参与者回答某一个特定答案，这就是偏见。例如，“你会不会赞同所有的司机都系安全带？”这就是一个有偏见的问题。调查问卷本身对受访者回答的影响也影响着数据收集的质量。一些受访者会根据他们感觉到调查问卷期望的答案来反馈。图 1-2 是我为特拉华州克莱蒙德社区中心的用户做的调查问卷的部分问题。为了鼓励受访者积极参与，实际的调查问卷需要在问卷头阐述调查目的，提问方式清晰简明，并且把涉及个人的人口统计问题放到最后——这时受访者对整个过程中会感觉更舒服。

 研究表明，调查问题提问的方式能够影响人们对调查问卷的回答。用正面语气提出的问题易于得到更为正面的回答；用负面语气提出的问题易于得到更为负面的回答。一个好的策略就是在把调查问卷发给实际参与人群之前，在小范围内先进行测试。