



Subject-Based English

高等学校专业英语教材

统计学

专业英语教程

☆ 王忠玉 宋要武 编著 ☆



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

高等学校专业英语教材

统计学专业英语教程

王忠玉 宋要武 编著

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

内 容 简 介

本书内容分为三部分：第一部分是描述统计学，共有7个单元，包括统计学初步、单变量数据的描述分析、两个变量数据的描述分析、概率初步、离散概率模型、连续概率模型、抽样分布和中心极限定理；第二部分是推断统计学（数理统计学），共有4个单元，包括统计推导初步、一个总体的统计推断、两个总体的统计推断、简单回归、回归的统计分析；第三部分是统计学与数据科学专题，只有1个单元。

和同类书籍相比，本书具有如下特点：（1）比较系统地阐述基础统计学的知识，即以初阶统计学的基本内容为主体，又适当地加入并介绍中阶统计学的部分内容；（2）在大多数章后，我们提供课外进一步阅读和学习的补充知识，有统计学家简介等；（3）紧跟当今时代发展，给出“统计学与数据科学”的阅读学习内容。另外，在每一章前面，我们精心选取了一些著名统计学家或教授的名言或警句，同时，特别绘制了有趣的漫画。本书提供部分习题参考答案、教学PPT、音频资料、部分课文参考译文及其他辅助资料，读者可从华信教育资源网 www.hxedu.com.cn 免费下载，也可扫描二维码获取。

本书适合于各个专业对统计学专业英语感兴趣的大学生，学习双语统计学的统计学专业大学生，希望学习和掌握中阶统计学的相关专业的低年级研究生，以及有关的科研人员等。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目 (CIP) 数据

统计学专业英语教程 / 王忠玉，宋要武编著. —北京：电子工业出版社，2016.8

高等学校专业英语教材

ISBN 978-7-121-28928-6

I. ①统… II. ①王… ②宋… III. ①统计学—英语—高等学校—教材 IV. ①H31

中国版本图书馆 CIP 数据核字 (2016) 第 117241 号

策划编辑：秦淑灵

责任编辑：秦淑灵

印 刷：三河市双峰印刷装订有限公司

装 订：三河市双峰印刷装订有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编：100036

开 本：787×1 092 1/16 印张：24.75 字数：824 千字

版 次：2016 年 8 月第 1 版

印 次：2016 年 8 月第 1 次印刷

印 数：3000 册 定价：49.80 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010)88254888，88258888。

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

本书咨询联系方式：(010)88254531。

序 言

当前,就我国大学英语教学的目标或模式而言,通常本科基础英语或传统的综合英语模式,不论是理、工、农林类,还是管理、财经类等大学生都要学习基础外语,最终以通过四级或六级英语考试为评价阶段性英语教学任务的标准。然后,各高校因专业不同而开设各自专业的英语,其目标是为大学生有机会接触用英语讲授的专业讲座和专业课程提供查阅、搜索和研究某个专题文献综述的一个良好开端。这样的教学模式可用图 1 来表示,也就是英语教学的实用目标和专业英语、基础英语三者之间的关系。

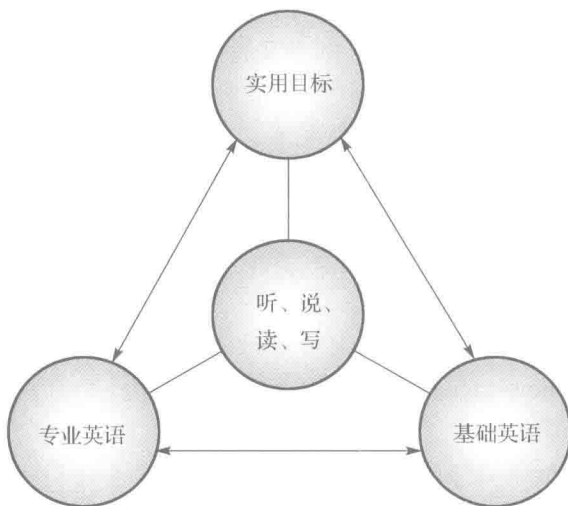


图 1 专业英语、基础英语和实用目标三者关系

2012年4月,上海市大学英语教学指导委员会宣布了以学术英语为导向的指导性文件《上海市大学英语教学参考框架(试行)》。2014年4月,由上海交通大学出版社出版《新核心综合学术英语教程》第四册,从而完成第一套大学生学术英语教材(共4册),这表明大学英语的教学方式向以“专门用途英语”为导向的转型之路走出了重要一步。出现这样的变化,可以说反映出一种新的趋势和发展事态,即随着大学教育的日益普及、大学生的外语水平普遍提升,国际交往和交流越来越多,各高校的不同专业大学生接触和联系外国专家、学者的机会也不断增多,试图通过外文直接获取、学习本专业知识的途径非常便利。同时,由于互联网的发展和移动互联网的普及,使得以往制约人们寻找、发现和获取新信息的瓶颈不复存在,进而出现了新的发展趋势。

那么,究竟什么是学术英语呢?学术英语的含义是 *English for Academic Purposes*, 记为 EAP, 一般可以分成两大类:第一类是通用学术英语 (*English for General Academic Purposes*, EGAP);第二类是专门学术英语 (*English for Specific Academic Purposes*, ESAP)。前者是一个跨越多学科的语言教学,目的是为各专业学生提供所需的通用共性的基本技能,包括学术口语交流能力和学术书面交流能力。具体地说,比如如何听讲座、做笔记、搜索和阅读文献、

撰写课程小论文、参加学术讨论等。后者则是某个特定学科领域（如数学、统计学、生物学、经济学、物理学）的英语教学。

实际上，如果从更广阔的视角来考察，许多高校的博士研究生英语教学早在多年前就已经执行了学术英语的教学。当今，随着英语逐渐成为世界上各个学科交流科研成果、各个学术团体及组织、会议和期刊的通用语言，学术英语迅速扩展到全世界。由此可见，学术英语的目标是培养大学生对本专业文献信息的查阅、搜集、评价、组织及表达的能力。尤其是，开展以问题或项目为指针的教学，使学生具有独立思维、独立学习的研究能力，这是每一名大学生所必备的学术素养（见图1）。

编写这本书的主要目的是，尝试提供一本针对统计学领域的专业英语，比较系统地阐述基础统计学的知识。作为统计学导论的书籍，本书深入浅出地讲解和阐述什么是统计学，特别是初阶统计学的基本内容。同时，紧跟当今时代发展，整理出“统计学与数据科学”可供选学和课外阅读的单元。

实际上，从数据科学（Data Science）的交叉属性来看，可将数据科学看成计算机科学、数值计算、现代数据分析等的交叉融合而形成的新兴学科，目的是从数据中获得知识，获得有价值的信息，服务于社会。如果从应用视角看，数据科学应具备三个条件：第一个条件是底层构架开发或使用能力，如 Spark, MapReduce, Hadoop 等；第二个条件是程序开发能力；第三个条件是数据建模和解决问题能力。

美国加州大学伯克利分校统计系的郁彬（Bin Yu）教授提出，一个合格的数据科学家应具备的基本素质和技能，可概括为 SDC³：

- Statistics (S) 统计学；
- Domain (science) knowledge (D) 深厚的（科学）知识；
- Computing (C) 计算技术；
- Collaboration (“team work”) (C) 团队的合作能力；
- Communication (to outsiders) (C) 与外界的沟通能力。

并认为

$$\text{Data Science} = \text{SDC}^3$$

美国统计学家吴建民教授（C.F. Jeff Wu）早在 1998 年的一个学术会议上就曾建议：

Statistics → Data Science

Statisticians → Data Scientists

Several good names have been taken up: computer, information science, material science, cognitive science. “Data Science” is likely the remaining good name reserved for us.

为了适应这一技术变革趋势与新兴的社会需求，伊利诺伊大学香槟分校从 2011 年起举办“数据科学暑期研究班”；哥伦比亚大学从 2013 年起开设《应用数据科学》课程，并从 2013 年起开设相关培训项目，从 2014 年起设立硕士学位，2015 年设立博士学位；纽约大学从 2013 年秋季起设立“数据科学”硕士学位。在英国，邓迪大学从 2013 年起设立“数据科学”硕士学位。

特别要提及的是，美国的得克萨斯大学奥斯汀分校（The University of Texas Austin）的自然科学学院（College of Natural Sciences）索性将统计系改名为统计及数据科学系（Department of Statistics and Data Sciences），而其他大学（如美国的西弗吉尼亚大学（West

Virginia University)) 统计系硕士研究生设有数据科学方向 (Master of Data Science)。另外, 斯坦福大学统计学系研究生层面教育也有数据科学方向。由此可见, 统计学是数据科学中最重要的组成部分之一。

作为数据科学三大支柱之一的计算机科学, 迄今为止的发展经历了三个阶段。早期阶段, 让计算机可以工作, 发展重点在于程序语言、编译原理、操作系统以及支撑它们的数学理论; 中期阶段, 让计算机变得有用, 发展重点在于算法和数据结构; 当前阶段, 让计算机具有更多的应用, 发展重点从离散类数学转到概率与统计。

如果从计算机科学处理数据的核心技术看, 机器学习就是当前最核心的技术之一, 而且发展势头非常强劲, 那么就不能不提到统计 (或统计方法)、数据、计算和机器学习这四者的关系, 如图 2 所示。因此, 在“统计学与数据科学”中对机器学习、统计计算等都有所涉及。另外, 有一种观点认为, 机器学习等价于“数据矩阵+统计学+最优化+算法”。由此可见统计学作为数据科学的另一个支柱的重要性。

2014 年 6 月 25 日, 全国科学技术名词审定委员会发布试用 204 条科技新词, 其中包括“大数据”、“云计算”、“物联网”、“三维打印”等 42 条热点名词以及“暗能量”、“宏基因组”等 162 条专业新词。具体地说, 大数据 (big data) 是指具有数量巨大 (无统一标准, 一般认为在 T 级或 P 级以上, 即 10^{12} 或 10^{15} 以上), 类型多样 (既包括数值型数据, 也包括文字、图形、图像、音频、视频等非数值型数据), 处理时效紧, 数据源可靠性保证度低等综合属性的数据集合。再比如, 物联网 (internet of things) 是指综合采用计算机、网络、传感器、控制设备等, 让能够被独立寻址的相关物理对象互联互通, 实现对其识别、监控和管理的智能化网络。(摘自 <http://tech.gmw.cn/> 中的 http://tech.gmw.cn/2014-06/25/content_11727262.htm; http://tech.gmw.cn/2014-06/25/content_11727659.htm.)

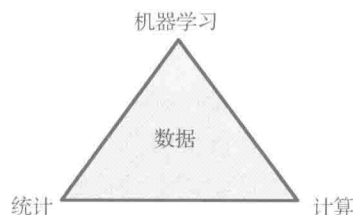


图 2 统计、数据、计算和机器学习四者的关系

本书的“统计学与数据科学”单元就包括了大数据、物联网这样的科技新词。

作者曾经出版过《统计学专业英语 (第 3 版)》(哈尔滨工业大学出版社, 2015 年 4 月), 但是这两本书在内容素材选取、难易程度等方面, 各自有不同的特点, 这两本书的关系可以说是相互补充, 没有替代性。具体而言, 《统计学专业英语 (第 3 版)》是针对学习过统计学或数理统计学的大学生, 提供了不同专题的单元内容以学习和掌握统计学专业英语。而本书则是针对以前没有学习过统计学, 打算了解和掌握利用英语阐明统计学基础知识的大学生, 内容定位为初阶统计学, 内容素材和前者相比, 完全不同。

另外, 本书在内容选取上, 充分兼顾“特性群体的大学生”, 也就是懂一点统计学但又想学习统计学专业英语的各类专业大学生或研究生, 各行各业有这方面需求的工作者。

本书尝试在下述几方面进行探索。

(1) 定位: 这是初阶、中阶统计学领域的专业英语, 目的是使学生初步认识、了解和掌握统计学专业领域的常用术语, 掌握统计学的基本内容, 学会运用基本统计分析方法。

(2) 教学内容: 以单元形式提供相关的统计内容, 给出有关的英文术语及词汇表。另外, 为方便教学, 提供有关用于教学的 PPT 等。

(3) 习题解答: 对某些较难的计算习题, 给出参考答案。另外, 我们为使用本书的教师

提供一些额外的教学资料,可以直接联系编辑或作者(编辑 E-mail: qinshl@phei.com.cn)。

部分习题参考答案、教学 PPT、音频资料、部分课文参考译文及其他辅助资料,可登录华信教育资源网 www.hxedu.com.cn 免费下载,也可扫描二维码获取。

全书的内容安排以 Unit(单元)为独立形式,前面有一个名言或语录、漫画,然后是单元的详细目录。随后,进入统计知识的阐述和讲解。最后有些单元后面还提供“补充阅读内容”,这包括两类知识:一类是正文内容的补充,另一类是历史人物、重要知识点等。音标所注为英式发音。

与此同时,为了扩展知识面,紧密联系当代统计学的新应用,本书特别编写“统计学和数据科学”的内容,这是一个选学单元。因此,这是一本初阶、中阶统计学专业英语教程,其中有些小节带有星号,表示这样的内容更适合于中阶。

全书整体设计有别于其他教材(不论是国内统计学教材,还是外文统计学教材),目的是试图编写出一本既有专业性,又有趣味性、可读性的教材或参考书。这本书就是在这种理念下构思编写而成的。

另外,本书的附录提供了几个有用的内容,包括统计学领域专业术语的标准翻译,参考了国家标准化委员会发布和出版的关于统计学方面的术语及英文翻译,也就是《GB/T 3358.1-2009 统计学词汇及符号 第 1 部分:一般统计术语与概率的术语》。实际上,如果读者需要学习和了解更多的统计学专业术语,可以参看《GB/T 3358.2-2009 统计学词汇及符号 第 2 部分:应用统计》;《GB/T 3358.3-2009 统计学词汇及符号 第 3 部分:实验设计》。

全书内容安排如下:王忠玉(哈尔滨工业大学经济与管理学院)编写第 1、6、7、8、9、10、12 单元,宋要武(黑龙江科技大学、哈尔滨股权投资协会)编写第 2、3、4、5 单元。哈尔滨理工大学经济学院的张莹老师编写 11 单元和书后附录的整理工作(包括统计学常用术语翻译、附录)。漫画制作人员有黑龙江农垦职业学院李辰光、戈娇老师,哈尔滨师范大学传媒学院曹文龙老师,研究生朱砚。本书采用漫画形式,以此概括或揭示本章内容的某种浓缩特色。这些漫画为书增添了趣味性和可读性。

另外,英国帝国理工大学的数学系学生朱烜燊提供了有益的建议。其他同学,如哈尔滨工业大学的王初旭、邢喆、周子涵、温雅欣、仇派、于娜、陈悦竹、夏晴、牟思涵、郑天慧、范晓菲,另外 2013 级金融学的武杰、满达,2014 级金融学的董赫,黑龙江大学的王天元等,也提供了许多有益的帮助。

书中难免存在纰漏和错误,欢迎广大读者、教师批评指正。E-mail: h20061111@126.com。

王忠玉 宋要武



部分习题参考答案



部分课文参考译文



附录 D-H



音频

目 录

Part I Descriptive Statistics

Unit 1 Statistics	3
1.1 What is Statistics?.....	4
1.1.1 Meanings of Statistics	4
1.1.2 Definition of Statistics	5
1.1.3 Types of Statistics	6
1.1.4 Applications of Statistics	6
1.2 The language of Statistics	9
1.2.1 Population and Sample	9
1.2.2 Kinds of Variables	11
1.3 Measurability and Variability.....	14
1.4 Data Collection.....	16
1.4.1 The Data Collection Process.....	17
1.4.2 Sampling Frame and Elements.....	18
1.5* Single-Stage Methods.....	21
1.5.1 Simple Random Sample	21
1.5.2 Systematic Sample	22
1.6* Multistage Methods.....	25
1.7* Types of Statistical Study	27
1.8 The Process of a Statistical Study	31
Glossary	34
Reading English Materials.....	35
Passage 1. What is Statistics?	35
Passage 2. From Data to Foresight.....	35
Problems	36
Unit 2 Descriptive Analysis of Single-Variable Data	40
2.1 Graphs, Pareto Diagrams, and Stem-and-Leaf Displays	41
2.1.1 Qualitative Data	41
2.1.2 Quantitative Data	43
2.2 Frequency Distributions and Histograms	47
2.2.1 Frequency Distribution.....	47

2.2.2	Histograms	51
2.2.3	Cumulative Frequency Distribution and Ogives	53
2.3	Measures of Central Tendency	55
2.3.1	Finding the Mean	55
2.3.2	Finding the Median	56
2.3.3	Finding the Mode	57
2.3.4	Finding the Midrange	58
2.4	Measures of Dispersion	60
2.4.1	Sample Standard Deviation	62
2.5	Measures of Position	64
2.5.1	Quartiles	64
2.5.2	Percentiles	64
2.5.3	Other Measures of Position	66
2.6	Interpreting and Understanding Standard Deviation	70
2.6.1	The Empirical Rule and Testing for Normality	70
2.6.2	Chebyshev's Theorem	72
	Glossary	74
	Problems	75
Unit 3	Descriptive Analysis of Bivariate Data	79
3.1	Bivariate Data	80
3.1.1	Two Qualitative Variables	80
3.1.2	One Qualitative and One Quantitative Variable	82
3.1.3	Two Quantitative Variables	83
3.2	Linear Correlation	85
3.2.1	Calculating the Linear Correlation Coefficient, r	86
*3.2.2	Causation and Lurking Variables	89
3.3	Linear Regression	91
3.3.1	Line of Best Fit	92
3.3.2	Making Predictions	97
	Reading English Materials	99
	Passage 1. The First Regression	99
	Passage 2. Simpson's Paradox	99
	Problems	100
Unit 4	Introduction to Probability	104
4.1	Sample Spaces, Events and Sets	105
4.1.1	Introduction	105
4.1.2	Sample Spaces	105
4.1.3	Events	106

4.1.4	Set Theory	108
4.2	Probability Axioms and Simple Counting Problems	109
4.2.1	Probability Axioms and Simple Properties	109
4.2.2	Interpretations of Probability	111
4.2.3	Classical Probability	112
4.2.4	The Multiplication Principle	113
4.3	Permutations and Combinations	115
4.3.1	Introduction	115
4.3.2	Permutations	116
4.3.3	Combinations	118
4.3.4	The Difference Between Permutations and Combinations	120
4.4	Conditional Probability and the Multiplication Rule	122
4.4.1	Conditional Probability	122
4.4.2	The Multiplication Rule	123
4.5	Independent Events, Partitions and Bayes Theorem	124
4.5.1	Independence	124
4.5.2	Partitions	125
4.5.3	Law of Total Probability	126
4.5.4	Bayes Theorem	126
4.5.5	Bayes Theorem for Partitions	127
	Reading English Materials	130
	Passage 1. Probability and Odds	130
	Passage 2. The Relationship between Odds and Probability	130
	Passage 3. How the Odds Change across the Range of the Probability	131
	Problems	132
Unit 5	Discrete Probability Models	134
5.1	Introduction, Mass Functions and Distribution Functions	135
5.1.1	Introduction	135
5.1.2	Probability Mass Functions (PMFs)	136
5.1.3	Cumulative Distribution Functions (CDFs)	137
5.2	Expectation and Variance for Discrete Random Quantities	138
5.2.1	Expectation	138
5.2.2	Variance	139
5.3	Properties of Expectation and Variance	140
5.3.1	Expectation of a Function of a Random Quantity	140
5.3.2	Expectation of a Linear Transformation	140
5.3.3	Expectation of the Sum of Two Random Quantities	141
5.3.4	Expectation of an Independent Product	141

5.3.5	Variance of an Independent Sum	142
5.4	The Binomial Distribution	142
5.4.1	Introduction	142
5.4.2	Bernoulli Random Quantities	143
5.4.3	The Binomial Distribution	143
5.4.4	Expectation and Variance of a Binomial Random Quantity	145
5.5	The Geometric Distribution	146
5.5.1	PMF	146
5.5.2	CDF	147
5.5.3	Useful Series in Probability	148
5.5.4	Expectation and Variance of Geometric Random Quantities	148
5.6	The Poisson Distribution	149
5.6.1	Poisson as the Limit of a Binomial	149
5.6.2	PMF	150
5.6.3	Expectation and Variance of Poisson	151
5.6.4	Sum of Poisson Random Quantities	152
5.6.5	The Poisson Process	152
	Reading English Materials	154
	Passage 1. The Founder of Modern Statistics—Karl Pearson	154
	Passage 2. The Relations of Several Discrete Probability Models	154
	Problems	155

Unit 6	Discrete Probability Models	158
6.1	Introduction, PDF and CDF	159
6.1.1	Introduction	159
6.1.2	The Probability Density Function	159
6.1.3	The Distribution Function	160
6.1.4	Median and Quartiles	161
6.2	Properties of Continuous Random Quantities	161
6.2.1	Expectation and variance of continuous random quantities	161
6.2.2	PDF and CDF of a Linear Transformation	162
6.3	The Uniform Distribution	163
6.4	The Exponential Distribution	165
6.4.1	Definition and Properties	165
6.4.2	Relationship with the Poisson Process	166
6.4.3	The Memoryless Property	167
6.5	The Normal Distribution	168
6.5.1	Definition	168

6.5.2	Properties	168
6.6	The Standard Normal Distribution	169
6.6.1	Properties of the Standard Normal Distribution	170
6.6.2	Finding Area to The Right of $z = 0$	171
6.6.3	Finding Area in The Right Tail of a Normal Curve	171
6.6.4	Finding Area to the Left of a Positive z Value	172
6.6.5	Finding Area from a Negative z to $z = 0$	172
6.6.6	Finding Area in the Left Tail of a Normal Curve	172
6.6.7	Finding Area from A Negative z to a Positive z	172
6.6.8	Finding Area Between two z Values of the Same Sign	173
6.6.9	Finding z -Scores Associated with a Percentile	173
6.6.10	Finding z -scores that Bound an Area	174
6.7	Applications of Normal Distributions	175
6.7.1	Probabilities and Normal Curves	175
6.7.2	Using the Normal Curve and z	176
6.8	Specific z -score	178
6.8.1	Visual Interpretation of $z(a)$	179
6.8.2	Determining Corresponding z Values for $z(a)$	179
6.8.3	Determining z -scores for Bounded Areas	180
6.9	Normal Approximation of Binomial and Poisson	181
6.9.1	Normal Approximation of the Binomial	181
6.9.2	Normal Approximation of the Poisson	182
	Problems	182
Unit 7	Sampling Distributions and CLT	187
7.1	Sampling Distributions	188
7.1.1	Forming a Sampling Distribution of Means	188
7.1.2	Creating a Sampling Distribution of Sample Means	189
7.2	The Sampling Distribution of Sample Means	192
7.2.1	Central Limit Theorem	193
7.2.2	Constructing a Sampling Distribution of Sample Means	194
7.3	Application of the Sampling Distribution of Sample Means	199
7.3.1	Converting $\bar{x}$ Information into z -scores	199
7.3.2	Distribution of $\bar{x}$ and Increasing Individual Sample Size	200
7.4	Advanced Central Limit Theorem	202
7.4.1	Central Limit Theorem (Sample Mean)	203
7.4.2	Central Limit Theorem (Sample Sum)	203
	Problems	207

Part II Inferential Statistics

Unit 8 Introduction to Statistical Inferences	210
8.1 Point Estimation and Interval Estimation	211
8.1.1 Point Estimate	211
8.1.2 Interval Estimate	212
8.2 Estimation of Mean μ (σ Known)	214
8.2.1 The Principle of Constructing a Confidence Interval	214
8.2.2 Applications	216
8.2.3 Sample Size and Confidence Interval	217
8.3 Introduction to Hypothesis Testing	220
8.3.1 Null Hypothesis and Alternative Hypothesis	220
8.3.2 Four Possible Outcomes in a Hypothesis Test	222
8.4 Formulating the Statistical Null and Alternative Hypotheses	226
8.4.1 Writing Null and Alternative Hypothesis in One-Tailed Situation	226
8.4.2 Writing Null and Alternative Hypothesis in Two-Tailed Situation	227
8.5 Hypothesis Test of Mean μ (σ Known): A Probability-Value Approach	228
8.5.1 One-Tailed Hypothesis Test Using the p -Value Approach	229
8.5.2 Two-Tailed Hypothesis Test Using the p -Value Approach	233
8.5.3 Evaluating the p -Value Approach	234
8.6 Hypothesis Test of Mean μ (σ Known): A Classical Approach	235
8.6.1 One-Tailed Hypothesis Test Using the Classical Approach	236
8.6.2 Two-Tailed Hypothesis Test Using the Classical Approach	239
Problems	241
Unit 9 Inferences Involving One Population	246
9.1 Inferences about the Mean μ (σ Unknown)	247
9.1.1 Using the t-Distribution Table	249
9.1.2 Confidence Interval Procedure	251
9.1.3 Hypothesis-Testing Procedure	252
9.2 Inferences about the Binomial Probability of Success	258
9.2.1 Confidence Interval Procedure	259
9.2.2 Determining Sample Size	261
9.2.3 Hypothesis-Testing Procedure	263
9.3 Inferences about the Variance and Standard Deviation	268
9.3.1 Critical Values of Chi-Square	269
9.3.2 Hypothesis-Testing Procedure	270
Problems	279

Unit 10	Inferences Involving Two Populations	284
10.1	Dependent and Independent Samples	285
10.2	Inferences Concerning the Mean Difference Using Two Dependent Samples	287
10.2.1	Procedures and Assumptions for Inferences Involving Paired Data	287
10.2.2	Confidence Interval Procedure	288
10.2.3	Hypothesis-Testing Procedure	290
10.3	Inferences Concerning the Difference between Means Using Two Independent Samples	294
10.3.1	Confidence Interval Procedure	295
10.3.2	Hypothesis-Testing Procedure	297
10.4	Inferences Concerning the Difference between Proportions	301
10.4.1	Confidence Interval Procedure	303
10.4.2	Hypothesis-Testing Procedure	304
10.5	Inferences Concerning the Ratio of Variances Using Two Independent Samples	308
10.5.1	Writing for the Equality of Variances	308
10.5.2	Using the F -Distribution	309
10.5.3	One-Tailed Hypothesis Test for the Equality of Variances	310
10.5.4	Critical F -Values for One- and Two-Tailed Tests	313
	Problems	315
Unit 11	An Introduction to Simple Regression	321
11.1	Regression as a Best Fitting Line	322
11.1.1	Regression as a Best Fitting Line	322
11.1.2	Errors and Residuals	324
11.2	Interpreting OLS Estimates	326
11.3	Fitted Values and R^2 : Measuring the Fit of a Regression Model	328
11.4	Nonlinearity in Regression	331
	Reading English Materials	335
	Problems	336

Part III Statistical Methods and Data Science

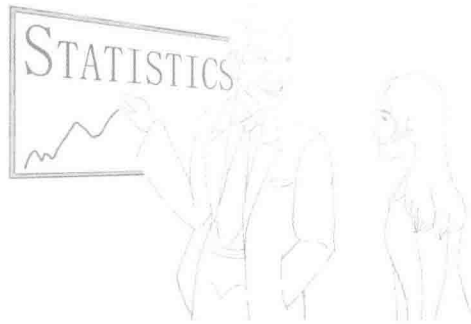
Unit 12	Statistics and Data Science	339
12.1	Statistics and Data Science (I)	340
12.1.1	What is Data Science	340
12.1.2	Statistics and Data Science	340
12.2	Statistics and Data Science (II)	343
12.2.1	Statistics as Part of Data Science	343
12.2.2	The Modern Statistical Analysis Process	344

12.2.3	Statistician and Data Scientist	345
12.3	Statistical Thinking	348
12.3.1	What is Statistical Thinking	348
12.3.2	The Two Cultures of Statistical Modeling	348
12.3.3	A New Research Community	350
12.4	Distinguishing Analytics, Business Intelligence, Data Science	352
12.4.1	Analytics	352
12.4.2	Business Intelligence	355
12.4.3	Data Science	356
	Reading English Materials	359
	Problems	361
	Commonly Used Statistical Terms	362
	Appendix A Commonly Used Statistical Tables	367
	Appendix B Summary of Univariate Descriptive Statistics and Graphs for the Four Level of Measurement	379
	Appendix C Order of Magnitude of Data	380
	References	381

Part I Descriptive Statistics

Most statistical work is concerned directly with the provision and implementation of methods for study design and for the analysis and interpretation of data.

—— D.R. Cox (He is one of the world's preeminent statisticians. His work on the proportional hazards regression model is one of the most-cited and most influential papers in modern statistics. In 2010 he won the Copley Medal of the Royal Society 'for his seminal contributions to the theory and application of statistics'. He is currently an Honorary Fellow at Nuffield College, Oxford.)



Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.

— H. G. Wells (1866—1946)

