时代教育·国外高校优秀教材精选

（英文版·原书第2版）

统计思想

Mind on Statistics

（美）Jessica M. Utts
Robert F. Heckard 著

机械工业出版社
CHINA MACHINE PRESS

附赠光盘

时代教育·国外高校优秀教材精选

统计思想

（英文版·原书第2版）

Mind on Statistics

（美） Jessica M. Utts
Robert F. Heckard 著

机械工业出版社

Jessica M. Utts, Robert F. Heckard

Mind on Statistics,2e

EISBN: 0-534-39305-5

981-254-712-6

北京市版权局著作权合同登记号：图字 01-2004-5364 号

图书在版编目（CIP）数据

统计思想：英文版·原书第 2 版/（美）尤茨（Utts,J.M.），（美）赫卡德（Heckard,R.F.）著.—北京：机械工业出版社，2005.1

（时代教育·国外高校优秀教材精选）

ISBN 7-111-10946-5

Ⅰ.统… Ⅱ.①尤…②赫… Ⅲ.统计—高等学校—教材—英文… Ⅳ.C8

中国版本图书馆 CIP 数据核字（2004）第 114372 号

机械工业出版社(北京市百万庄大街 22 号　邮政编码 100037)

责任编辑：郑　玫

封面设计：饶　薇　　责任印制：施　红

北京铭成印刷有限公司印刷·新华书店北京发行所发行

2005 年 1 月第 1 版第 1 次印刷

787mm × 1092mm 1/16·42.25 印张·1050 千字

定价：66.00 元（含 1CD）

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

本社购书热线电话（010）68993821、88379646

封面无防伪标均为盗版

国外高校优秀教材审定委员会

出 版 说 明

随着我国加入 WTO，国际间的竞争越来越激烈，而国际间的竞争实际上也就是人才的竞争、教育的竞争。为了加快培养具有国际竞争力的高水平技术人才，加快我国教育改革的步伐，国家教育部近来出台了一系列倡导高校开展双语教学、引进原版教材的政策。以此为契机，机械工业出版社陆续推出了一系列国外影印版教材，其内容涉及高等学校公共基础课，以及机、电、信息领域的专业基础课和专业课。

引进国外优秀原版教材，在有条件的学校推动开展英语授课或双语教学，自然也引进了先进的教学思想和教学方法，这对提高我国自编教材的水平，加强学生的英语实际应用能力，使我国的高等教育尽快与国际接轨，必将起到积极的推动作用。

为了做好教材的引进工作，机械工业出版社特别成立了由著名专家组成的国外高校优秀教材审定委员会。这些专家对实施双语教学做了深入细致的调查研究，对引进原版教材提出了许多建设性意见，并慎重地对每一本将要引进的原版教材一审再审，精选再精选，确认教材本身的质量水平，以及权威性和先进性，以期所引进的原版教材能适应我国学生的外语水平和学习特点。在引进工作中，审定委员会还结合我国高校教学课程体系的设置和要求，对原版教材的教学思想和方法的先进性、科学性严格把关。同时尽量考虑原版教材的系统性和经济性。

这套教材出版后，我们将根据各高校的双语教学计划，举办原版教材的教师培训，及时地将其推荐给各高校选用。希望高校师生在使用教材后及时反馈意见和建议，使我们更好地为教学改革服务。

机械工业出版社

序

本书保持了第1版的特点：作为统计学的入门教材，对数学基础知识的要求较低；用丰富、新鲜的例子说明统计学的概念、原理和运算程序，使全书极具趣味性和引人入胜；同时计算机、互联网以及多媒体教学手段的合理充分的使用，改变了统计学课程教学中原有的面貌，使学习和掌握统计学变得生动、活泼和十分容易。

全书内容包括数据的采集、整理与概括，变量之间的相关关系与分析，概率与随机变量，随机变量的数字特征，点估计和区间估计，回归分析与方差分析等内容。本书的特点是以讲统计学为主，概率论的内容服从和服务于统计学，占很小部分，而且两者穿插结合得如此恰当、完美，令人佩服。本书前四章讲抽样调查，把问题讲得很清楚、很好懂，是国内教材中不多见的。本书有两章讲分类变量之间的关系与分析，也是国内教材中不安排或很少安排的，而实际上这部分内容是很有用的。

全书自成体系，对统计学的原则和概念叙述得相当清楚和准确。书中的框图、小结等也使本书增色不少。

本书第2版比第1版在各章节中适当增加了一些内容和例子。各章节的习题有所增加，某些章节中增加了“打开计算机”的环节，为读者提供通过计算机掌握统计概念的手段。书末的光盘提供的内容要比第1版的光盘内容丰富许多。对全书的主要内容已作出PowerPoint，可直接用于双语教学。同时光盘中有数据和用各种统计软件解决实例的程序。

北京工商大学

章栋恩

Case Studies and Examples* by Chapter and Supplemental Topics

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

* Examples marked by an asterisk are revisited for further discussion later in the chapter.

Chapter 8

Chapter 9

Chapter 10

Chapter 11

Chapter 12

Chapter 13

Chapter 14

Chapter 15

Chapter 16

Chapter 17

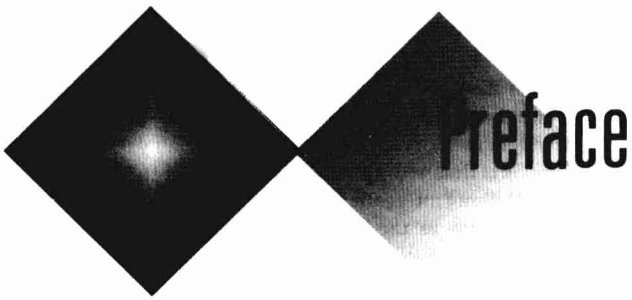
Supplemental Topic 1

Supplemental Topic 2

Supplemental Topic 3

Supplemental Topic 4

Supplemental Topic 5



Preface

A CHALLENGE

Before you continue, think about how you would answer the question in the first bullet, and read the statement in the second bullet. We will return to them a little later in this Preface.

- What do you *really know* is true, and how do you know it?
- The diameter of the moon is about 2160 miles.

WHAT IS STATISTICS AND WHO SHOULD CARE?

Because people are curious about many things, chances are that your interests include topics to which statistics has made a useful contribution. As written in Chapter 17, "information developed through the use of statistics has enhanced our understanding of how life works, helped us learn about each other, allowed control over some societal issues, and helped individuals make informed decisions. There is almost no area of knowledge that has not been advanced by statistical studies."

Statistical methods have contributed to our understanding of health, psychology, ecology, politics, music, lifestyle choices, and dozens of other topics. A quick look through this book, especially Chapters 1 and 17, should convince you of this. Watch for the influences of statistics in your daily life as you learn this material.

Although statistics courses are often offered through a mathematics department, statistics is not a branch of mathematics. Mathematics is to statistics as wood, hammer, and nails are to building a house: a partial set of materials and tools. Statistics also draws materials and tools from philosophy, graphics, computing, psychology, and language.

HOW IS THIS BOOK DIFFERENT? TWO BASIC PREMISES OF LEARNING

We wrote this book because we were tired of being told that what statisticians do is boring and difficult. We think statistics is useful and not difficult to learn, and yet the majority of college graduates we've met seemed to have had a negative experience taking a statistics class in college. We hope this book will help to overcome these misguided stereotypes.

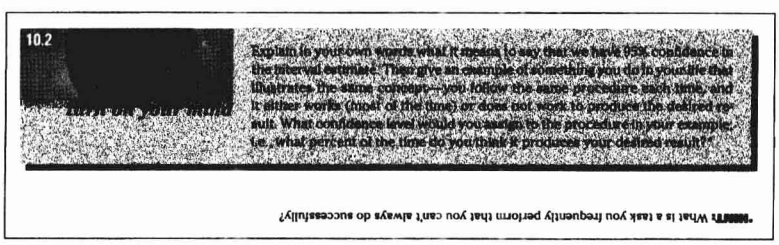
Let's return to the two bullets at the beginning of this Preface. Without looking, do you remember the diameter of the moon? Unless you already had a pretty good idea, or have an excellent memory for numbers, you probably don't remember. One premise of this book is that new material is much easier to learn and remember if it is related to something interesting or previously known. The diameter of the moon is about the same as the air distance between Atlanta and Los Angeles, San Francisco and Chicago, London and Cairo, or Moscow and Madrid. Picture the moon sitting between any of those pairs of cities, and you are not likely to forget the size of the moon again. Throughout this book, new material is presented in the context of interesting and useful examples. The first and last chapters (1 and 17) are exclusively devoted to examples and case studies, which illustrate the wisdom that can be generated through statistical studies.

Now answer the question asked in the first bullet: What do you really know is true and how do you know it? If you are like most people, you know because it's something you have experienced or verified for yourself. It is not likely to be something you were told or heard in a lecture. The second premise of this book is that new material is easier to learn if you actively ask questions and answer them for yourself. *Mind on Statistics* is designed to help you learn statistical ideas by actively thinking about them. Throughout most of the chapters there are boxes entitled *Turn On Your Mind.* Thinking about the questions in those boxes will help you to discover and verify important ideas for yourself. We encourage you to think and question, rather than simply read and listen.

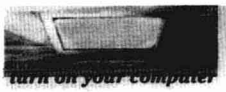
TOOLS FOR EXPANDED LEARNING

There are a number of tools provided in this book and beyond to enhance your learning of statistics.

Turn On Your Mind boxes appear throughout each chapter to help you develop your statistical reasoning and intuition. *Hints* are now provided to help you develop these skills.



New to this edition: *Turn On Your Computer* features provide additional opportunities for independent, hands-on exploration of key statistical concepts.



Random Sampling in Action

It should be obvious to you by now that each time you take a random sample from a population you are likely to get a different group of individuals, even though the population remains fixed. The **Sampling** applet on the CD for this text allows us to watch what happens when we repeatedly take random samples from the same population. Figure 4.1 illustrates how the applet appears when it starts. Each stick figure represents one member of a population of 100 college students: 55 females (blue) and 45 males (red). Their heights have been recorded as well, and if you look carefully you will see that the little stick figures are all of slightly different heights. The mean height for the population is 68 inches.

FIGURE 4.1 The sampling applet starting point

Examples throughout the chapters tie statistical concepts to everyday life occurrences.

EXAMPLE 10.2
If I Won the Lottery I'd Consider Quitting

What would you do for the rest of your life if you won a lot of money in a lottery? In a 1997 poll conducted by the Gallup Organization, one of the questions was "If you won 10 million dollars in the lottery would you continue to work, or would you stop working?" The results were reported at the Gallup Organization's website (www.gallup.com).

Surprisingly, 59% of the 616 employed respondents answered that they would continue working, 40% said they would stop working, and 1% had no opinion. The website article also gave this information about the poll:

> The current results are based on telephone interviews with a randomly selected sample of 1,014 adults, conducted August 22–25, 1997. Among this group, 616 are employed full-time or part-time. For results based on this sample of "workers," one can say with 95% confidence that the error attributable to sampling could be plus or minus 4 percentage points.

Gallup describes the margin of error with the phrase "could be plus or minus 4 percentage points," but the phrase "could be *as large as* plus or minus 4 percentage points" would have been more informative. The margin of error is a likely upper limit on the sampling error. Here, there is a 95% chance that the sampling error is actually smaller than 4 percentage points. ◆

Key Terms at the end of each chapter, organized by section, can be used as a "quick-finder" and as a review tool.

KEY TERMS

Section 9.1
statistic, 294, 296
parameter, 294, 296
sampling distribution, 294–295, 296

Section 9.2
sample proportion, 296
normal curve approximation rule for sample proportions, 297, 298
sampling distribution of $\hat{p}$, 298
standard deviation of $\hat{p}$, 300
standard error of $\hat{p}$, 300

Section 9.3
normal curve approximation rule for sample means, 302–303, 304
sampling distribution of $\bar{x}$, 304
sampling distribution of the mean, 304
standard deviation of the sample mean, 304
standard error of the mean, 305
law of large numbers, 306

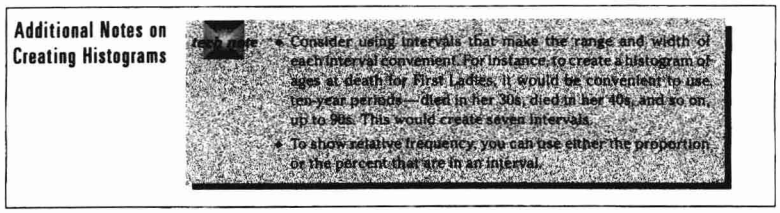
Section 9.4
central limit theorem, 309

Section 9.6
standardized statistic, 313–314
standardized *z*-statistic, 313–314, 315

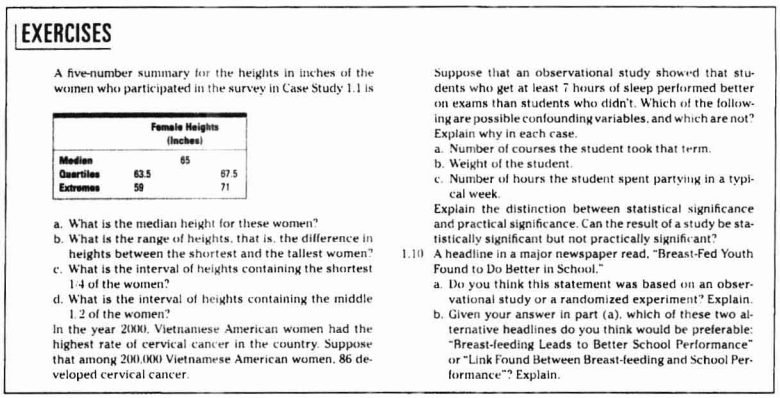
Section 9.7
Student's *t*-distribution, 314
t-distribution, 314
degrees of freedom, (df), 315
standardized *t*-statistic, 315

Section 9.8
statistical inference, 316
confidence interval, 317
hypothesis testing, 317
significance testing, 317
statistical significance, 317

Special *Tech Note* boxes provide additional technical discussion as well as details about using MINITAB® statistical software* and Microsoft® Excel®.



New to this edition: *Mind on the Basics* Exercises focus on practice and review; these exercises, color-coded in blue and appearing at the beginning of each exercise section, complement the conceptual and data-analysis exercises.



EXERCISES

A five-number summary for the heights in inches of the women who participated in the survey in Case Study 1.1 is

	Female Heights (Inches)		
Median		65	
Quartiles	63.5		67.5
Extremes	59		71

a. What is the median height for these women?
b. What is the range of heights, that is, the difference in heights between the shortest and the tallest women?
c. What is the interval of heights containing the shortest 1/4 of the women?
d. What is the interval of heights containing the middle 1/2 of the women?

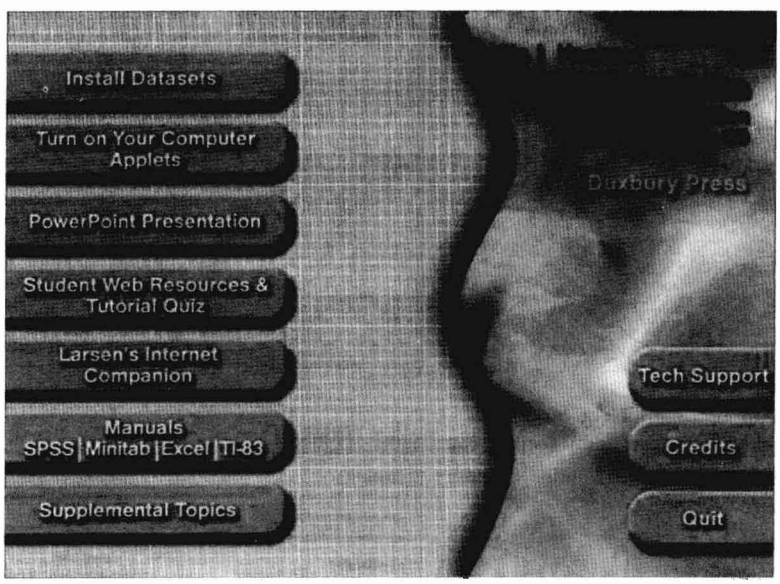
In the year 2000, Vietnamese American women had the highest rate of cervical cancer in the country. Suppose that among 200,000 Vietnamese American women, 86 developed cervical cancer.

Suppose that an observational study showed that students who get at least 7 hours of sleep performed better on exams than students who didn't. Which of the following are possible confounding variables, and which are not? Explain why in each case.
a. Number of courses the student took that term.
b. Weight of the student.
c. Number of hours the student spent partying in a typical week.

Explain the distinction between statistical significance and practical significance. Can the result of a study be statistically significant but not practically significant?

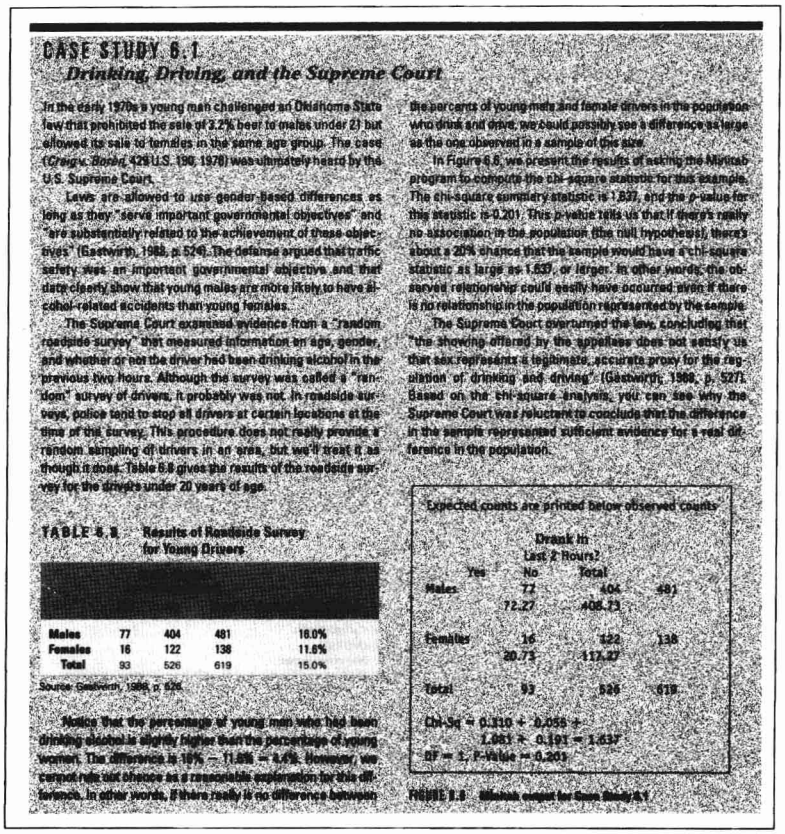
1.110 A headline in a major newspaper read, "Breast-Fed Youth Found to Do Better in School."
a. Do you think this statement was based on an observational study or a randomized experiment? Explain.
b. Given your answer in part (a), which of these two alternative headlines do you think would be preferable: "Breast-feeding Leads to Better School Performance" or "Link Found Between Breast-feeding and School Performance"? Explain.

A Student's Suite CD provided with the book includes the *Turn On Your Computer* applets, tutorial quizzes, InfoTrac® College Edition and Internet exercises, Microsoft PowerPoint® presentation slides, links to both the Book Companion Web Site and *Internet Companion for Statistics* by Michael Larsen, Supplemental Topics identified in the Table of Contents, datasets formatted for MINITAB, Microsoft Excel, SPSS®, SAS®, JMP®, and ASCII, and technology manuals designed for use with MINITAB, SPSS, Excel, and the TI-83.



*MINITAB is a trademark of Minitab, Inc. and is used herein with the owner's permission (www.minitab.com).

Case Studies apply statistical ideas to intriguing news stories. As the *Case Studies* are developed, they model the statistical reasoning process.



CASE STUDY 6.1

Drinking, Driving, and the Supreme Court

In the early 1970s a young man challenged an Oklahoma State law that prohibited the sale of 3.2% beer to males under 21 but allowed its sale to females in the same age group. The case (*Craig v. Boren*, 429 U.S. 190, 1976) was ultimately heard by the U.S. Supreme Court.

Laws are allowed to use gender-based differences as long as they "serve important governmental objectives" and "are substantially related to the achievement of these objectives" (Gastwirth, 1988, p. 524). The defense argued that traffic safety was an important governmental objective and that data clearly show that young males are more likely to have alcohol-related accidents than young females.

The Supreme Court examined evidence from a "random roadside survey" that measured information on age, gender, and whether or not the driver had been drinking alcohol in the previous two hours. Although the survey was called a "random" survey of drivers, it probably was not. In roadside surveys, police tend to stop all drivers at certain locations at the time of the survey. This procedure does not really provide a random sampling of drivers in an area, but we'll treat it as though it does. Table 6.8 gives the results of the roadside survey for the drivers under 20 years of age.

TABLE 6.8 Results of Roadside Survey for Young Drivers

Males	77	404	481	16.0%
Females	16	122	138	11.6%
Total	93	526	619	15.0%

Source: Gastwirth, 1988, p. 524.

Notice that the percentage of young men who had been drinking alcohol is slightly higher than the percentage of young women. The difference is 16% − 11.6% = 4.4%. However, we cannot rule out chance as a reasonable explanation for this difference. In other words, if there really is no difference between the percents of young males and female drivers in the population who drink and drive, we could possibly see a difference as large as the one observed in a sample of this size.

In Figure 6.8, we present the results of asking the Minitab program to compute the chi-square statistic for this example. The chi-square summary statistic is 1.637, and the *p*-value for this statistic is 0.201. This *p*-value tells us that if there's really no association in the population (the null hypothesis), there's about a 20% chance that the sample would have a chi-square statistic as large as 1.637, or larger. In other words, the observed relationship could easily have occurred even if there is no relationship in the population represented by the sample.

The Supreme Court overturned the law, concluding that "the showing offered by the appellees does not satisfy us that sex represents a legitimate, accurate proxy for the regulation of drinking and driving" (Gastwirth, 1988, p. 527). Based on the chi-square analysis, you can see why the Supreme Court was reluctant to conclude that the difference in the sample represented sufficient evidence for a real difference in the population.

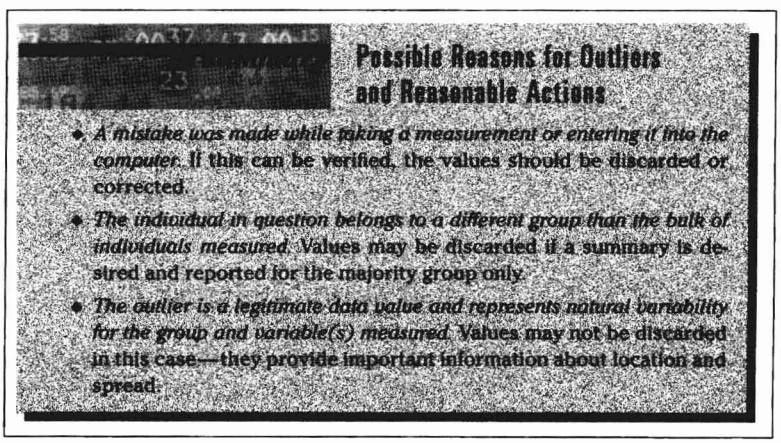
Expected counts are printed below observed counts

	Drank in Last 2 Hours?		
	Yes	No	Total
Males	77	404	481
	72.27	408.73	
Females	16	122	138
	20.73	117.27	
Total	93	526	619

Chi-Sq = 0.310 + 0.055 +
1.061 + 0.191 = 1.637
DF = 1, P-Value = 0.201

FIGURE 6.8 Minitab output for Case Study 6.1

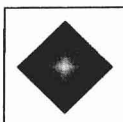
In Summary boxes appear at appropriate points to enhance key concepts and calculations; many are new in this edition.



Possible Reasons for Outliers and Reasonable Actions

- *A mistake was made while taking a measurement or entering it into the computer.* If this can be verified, the values should be discarded or corrected.
- *The individual in question belongs to a different group than the bulk of individuals measured.* Values may be discarded if a summary is desired and reported for the majority group only.
- *The outlier is a legitimate data value and represents natural variability for the group and variable(s) measured.* Values may not be discarded in this case—they provide important information about location and spread.

Many *Exercises* (indicated by bold numbers) have complete or partial solutions, found in the *Answers to Selected Exercises* at the back of the book, to check your answers on those exercises and guide your thinking on similar exercises.



Answers to Selected Exercises

Chapter 1

1.2 a. .00043
1.5 a. 400
1.7 c. Randomized experiment.
d. Observational study.
1.11 189/11,034, or about 17/1000, based on placebo group.
1.15 a. 150 mph.
b. 55 mph.
c. 80 mph.
d. 1/2
e. 51
1.19 No.
1.22 The base rate for that type of cancer.
1.26 a. 212/1525 = .139.
b. $1/\sqrt{1525}$ = .026.
c. .113 to .165.
1.28 a. Self-selected or volunteer.
b. No. Readers with strong opinions will respond.

Chapter 2

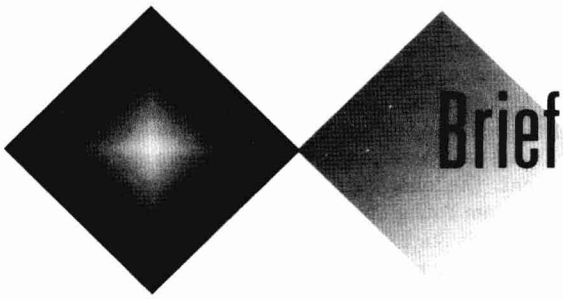
2.1 a. Categorical.
b. Quantitative.
c. Quantitative.
d. Categorical.
2.2 a. Categorical.
b. Ordinal.
2.4 a. Not continuous.
2.5 a. Explanatory is amount of walking or running; response is lung function.
2.6 a. Population.
b. Sample.
2.8 a. Support ban or not; categorical.
b. Gain on verbal and math SATs after program; quantitative.
c. Smoker or not and Alzheimer sufferer or not; both categorical.
2.9 a. Question 1a.
2.11 Example: Letter grades (A, B, etc.) converted to GPA.
2.14 a. 1427/2530 = .564, or 56.4%.
b. 100% − 56.4% = 43.6%.
2.15 a. 1700/2470 = .688, or 67.8%.
b. 1056/1700 = .621, or 62.1%.
2.16 c. Overweight = 26.57%; about right = 69.23%; underweight = 4.20%.
2.18 a. Explanatory is smoked or not; response is developed Alzheimer's disease or not.
2.21 Either could be justified as being more informative.
2.23 a. Skewed to the right.
b. 13 ear pierces may be an outlier.
c. 2 ear pierces; about 45 women had this number.
d. About 32 or so.
2.25 a. Roughly symmetric.
b. Highest = 92.
c. Lowest = 64.
d. 5/20 = .25, or 25%.
2.29 Example: The age of a person who is 80 years old would be an outlier at a traditional college, but not at a retirement home.
2.31 Whether it's the male author (then not an outlier) or the female author (then an outlier).
2.34 Yes. Values inconsistent with the bulk of the data will be obvious.
2.36 a. This is personal preference; some may prefer a very large family.
b. An outlier (in the high direction).
2.37 a. Median = (72 + 76)/2 = 74; mean = 74.33.
b. Median = 7; mean = 25.
2.39 a. Range = 225 − 123 = 102.
b. IQR = 35.
c. 50%

A Note to Instructors

The Instructor's Suite CD includes everything featured on the Student's Suite CD, plus a Test Bank in Microsoft Word® format, sample course outlines and syllabi, Web resources for examples and case studies, complete solutions to the exercises found in the text, suggested answers for the *Turn On Your Mind* questions, recommended in-class noncomputer-oriented projects, and information on how to correlate text material with the AP Statistics curriculum. Additional Instructor's Resources are available through your Duxbury/Thomson representative.

ACKNOWLEDGMENTS

We thank William Harkness, Professor of Statistics at Penn State University, for continued support and feedback throughout our careers and during the writing of this book, and for his remarkable dedication to undergraduate statistics education. Preliminary editions of *Mind on Statistics* were used at Penn State, the University of California at Davis, and Texas A & M University, and we thank the many students who provided comments and suggestions on those and on the first edition. Thanks to Dr. Melvin Morse (Valley Children's Clinic and University



Brief Contents

of Washington) for suggesting the title for Chapter 17 and to Deb Niemeier, University of California, Davis, for suggesting that we add a supplemental chapter on Ethics (on the CD). At Penn State, Dave Hunter, Steve Arnold, and Tom Hettmansperger have provided many helpful insights. At the University of California at Davis, Rodney Wong has provided insights as well as material for some exercises and the Test Bank. The following reviewers offered valuable suggestions: Patti B. Collings, Brigham Young University; James Curl, Modesto Junior College; Donald Harden, Georgia State University; Rosemary Hirschfelder, University Sound; Sue Holt, Cabrillo Community College; Tom Johnson, North Carolina University; Andre Mack, Austin Community College; D'Arcy Mays, Virginia Commonwealth; Mary Murphy, Texas A & M University; N. Thomas Rogness, Grand Valley State University; Heather Sasinouska, Clemson University; Robert Alan Wolf, University of San Francisco; Elizabeth Clarkson, Wichita State University; Megan Meece, University of Florida; Thomas Nygren, Ohio State University; Wade Ellis, West Valley College; David Robinson, St. Cloud University; Jay Gregg, Colorado State University; Mark Johnson, University of Central Florida; and Nancy Pfenning, University of Pittsburgh.

Our sincere appreciation and gratitude goes to Carolyn Crockett and the staff of Duxbury, without whom this book could not have been written, and to Martha Emry who kept us on track throughout the editing and production of both editions of the book. Finally, for their support, patience, and numerous prepared dinners, we thank our families and friends, especially Candace Heckard, Molly Heckard, Wes Johnson, Claudia Utts-Smith, and Dennis Smith.

Jessica Utts
Robert Heckard