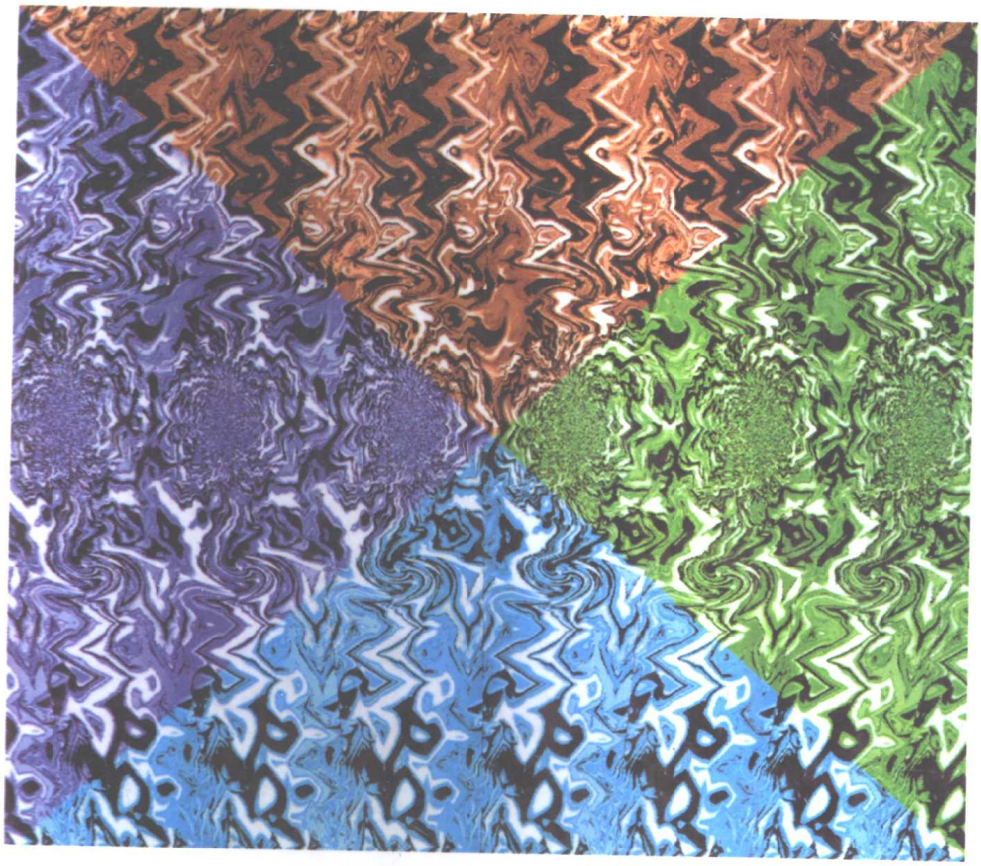
时代教育 • 国外高校优秀教材精选



统计思想（英文版）

Mind on Statistics

（美）杰西卡 M. 尤茨（Jessica M. Utts）
罗伯特 F. 赫卡德（Robert F. Heckard） 著

附赠光盘

机械工业出版社
CHINA MACHINE PRESS

时代教育·国外高校优秀教材精选

统计思想

（英文版）
（附光盘）

Mind on Statistics

（美）杰西卡 M. 尤茨（Jessica M. Utts）
罗伯特 F. 赫卡德（Robert F. Heckard） 著

机械工业出版社

Jessica M. Utts, Robert F. Heckard: Mind on statistics
(ISBN 0-534-35935-3)

Original Edition Copyright © 2002 by the Wadsworth Group. Duxbury is an imprint of the Wadsworth Group, a division of Thomson Learning. All rights reserved.

Reprinted for People's Republic of China by Thomson Learning Asia and China Machine Press under the authorization of Thomson Learning.

No Part of this book may be reproduced in any form without the express written permission of Thomson Learning Asia and China Machine Press.

981—243—399—6

本书英文影印版由汤姆森学习出版集团 THOMSON 授权机械工业出版社独家出版。未经出版者书面许可，不得以任何方式抄袭、复制或节录本书的任何部分。

版权所有，侵权必究。

北京市版权局著作权合同登记号：图字：01-2002-4189

图书在版编目（CIP）数据

统计思想/(美) 尤茨 (Utts, J.M.), (美) 赫卡德 (Heckard, R.F.) 著. -北京：机械工业出版社，2002.9
(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（2002）第 069624 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：刘小慧 郑 玫 封面设计：饶 薇 责任印制：付方敏
北京市密云县印刷厂印刷·新华书店北京发行所发行
2002 年 10 月第 1 版·第 1 次印刷
787mm × 1092mm 1/16· 36.75 印张 · 2 插页 · 906 千字
定价：38·00 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、68326677-2527
封面无防伪标均为盗版

国外高校优秀教材审定委员会

主任委员：杨叔子

委员（按姓氏笔画为序）：

王先逵　王大康　白峰杉　史荣昌　朱孝禄

陆启韶　张润琦　张　策　张三慧　张福润

张延华　吴宗泽　吴　麒　宋心琦　李俊峰

佘远斌　陈文楷　陈立周　俞正光　赵汝嘉

翁海珊　龚光鲁　章栋恩　黄永畅　谭泽光

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，机械工业出版社拟于近期推出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

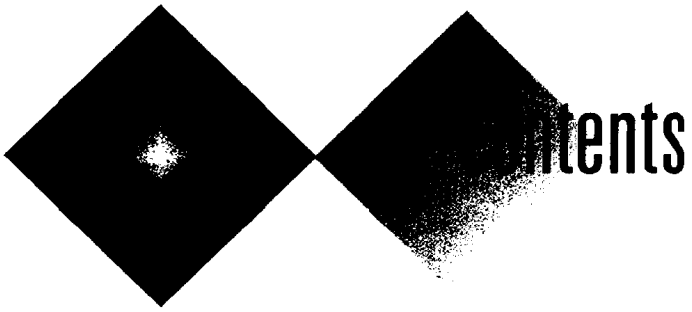
2002 年 3 月

此为试读，需要完整PDF请访问：www.ertongbook.com

序

Utts 和 Heckard 所著“Mind On Statistics”一书是理工科和管理、经济类专业的大学统计学的入门教科书。本书只要求学生具备高中数学知识就能阅读。作者力求使读者认识到统计学非常有用，且容易掌握，因此用丰富、新鲜的例子说明统计学的概念、原则和运算程序，通过例子中的趣味性以及对以前学习知识的联想，使读者能牢记有关的统计学知识。本书内容包括数据的采集、整理、概括（抽样方法和描述性统计）、变量之间的相关关系、概率和随机变量、随机变量数字特征、点估计和区间估计、假设检验、回归分析和方差分析。分类数据的统计分析在本书中也占有重要位置和较大篇幅，因而增强了实用性。从而也使本书成为一本全面、实用的统计学入门教科书。全书篇幅不很大，适合于大专院校理工科各专业、财经和管理类各专业的教学使用。该教材本身自成体系，不需要先学习概率论，也不必先学习微积分，本书对统计学的原则和概念叙述得相当清楚和准确。因此本书的出版将能填补国内这类教材的空白，也能使国内大学教学中所用的统计术语、软件等与国际接轨。

章栋恩
北京工商大学基础部
2002 年 3 月



Contents

出版说明 iv
序 v

CHAPTER 1 Statistics Success Stories and Cautionary Tales 1

1.1 What Is Statistics? 1
1.2 Seven Statistical Stories with Morals 2
- ***Case Study 1.1:*** Who Are Those Speedy Drivers? 2
- ***Case Study 1.2:*** Disaster in the Skies? 3
- ***Case Study 1.3:*** Did Anyone Ask Whom You've Been Dating? 3
- ***Case Study 1.4:*** Who Are Those Angry Women? 4
- ***Case Study 1.5:*** Does Prayer Lower Blood Pressure? 5
- ***Case Study 1.6:*** Does Aspirin Reduce Heart Attack Rates? 5
- ***Case Study 1.7:*** Does the Internet Increase Loneliness and Depression? 6

1.3 The Common Elements in the Seven Stories 7
KEY TERMS 8
EXERCISES 9
REFERENCES 10

CHAPTER 2 Turning Data into Information 12

2.1 Raw Data 13
2.2 Types of Data 14
2.3 Summarizing One or Two Categorical Variables 18
2.4 Finding Information in Quantitative Data 23
2.5 Pictures for Quantitative Data 28
2.6 Numerical Summaries of Quantitative Variables 32
2.7 Bell-Shaped Distributions of Numbers 40

KEY TERMS 46
EXERCISES 46
REFERENCES 51

CHAPTER 3 Gathering Useful Data 52

3.1 Description or Decision? Using Data Wisely 53
3.2 Speaking the Language of Research Studies 55
◆ *Case Study 3.1:* Lead Exposure and Bad Teeth 60
3.3 Designing a Good Experiment 61
◆ *Case Study 3.2:* Kids and Weight Lifting 61
◆ *Case Study 3.3:* Quitting Smoking with Nicotine Patches 66
3.4 Designing a Good Observational Study 67
◆ *Case Study 3.4:* Baldness and Heart Attacks 67
3.5 Difficulties and Disasters in Experiments and Observational Studies 69
KEY TERMS 74
EXERCISES 74
REFERENCES 78

CHAPTER 4 Sampling: Surveys and How To Ask Questions 80

4.1 The Beauty of Sampling 81
4.2 Sampling Methods 84
4.3 Difficulties and Disasters in Sampling 93
◆ *Case Study 4.1:* The Infamous *Literary Digest* Poll of 1936 97
4.4 How to Ask Survey Questions 98
◆ *Case Study 4.2:* No Opinion of Your Own? Let Politics Decide 102
KEY TERMS 102
EXERCISES 103
REFERENCES 107

CHAPTER 5 Relationships Between Quantitative Variables 108

5.1 Looking for Patterns with Scatterplots 110
5.2 Describing Linear Patterns with a Regression Line 115
5.3 Measuring Strength and Direction with Correlation 122
5.4 Why Answers May Not Make Sense 128
5.5 Correlation Does Not Prove Causation 133
◆ *Case Study 5.1:* A Weighty Issue 135
KEY TERMS 137
EXERCISES 137
REFERENCES 143

CHAPTER 6 Relationships Between Categorical Variables 144

6.1 Displaying Relationships Between Categorical Variables 145
6.2 Risk, Relative Risk, Odds Ratio, and Increased Risk 148
6.3 Misleading Statistics about Risk 151
6.4 The Effect of a Third Variable and Simpson's Paradox 154
6.5 Assessing the Statistical Significance of a 2×2 Table 157
- ◆ ***Case Study 6.1:*** Drinking, Driving, and the Supreme Court 163

KEY TERMS 165
EXERCISES 165
REFERENCES 171

CHAPTER 7 Probability 172

7.1 Random Circumstances 173
- ◆ ***Case Study 7.1:*** A Hypothetical Story—Alicia Has a Bad Day 174

7.2 Interpretations of Probability 176
7.3 Probability Definitions and Relationships 181
7.4 Basic Rules for Finding Probabilities 186
7.5 Strategies for Finding Complicated Probabilities 192
7.6 Using Simulation to Estimate Probabilities 199
7.7 Coincidences and Intuitive Judgments about Probability 202

KEY TERMS 209
EXERCISES 209
REFERENCES 214

CHAPTER 8 Random Variables 216

8.1 What is a Random Variable? 217
8.2 Discrete Random Variables 219
8.3 Expectations for Random Variables 223
8.4 Binomial Random Variables 228
8.5 Continuous Random Variables 233
8.6 Normal Random Variables 236
8.7 Approximating Binomial Distribution Probabilities 243
8.8 Sums, Differences, and Combinations of Random Variables 245
- ◆ ***Case Study 8.1:*** Does Caffeine Enhance the Taste of Cola? 251

KEY TERMS 252
EXERCISES 252
REFERENCES 257

CHAPTER 9 Means and Proportions as Random Variables 258

9.1 Understanding Dissimilarity among Samples 259
9.2 Sampling Distributions for Sample Proportions 261
9.3 What to Expect of Sample Means 265
9.4 What to Expect in Other Situations: Central Limit Theorem 270
9.5 Sampling Distribution for Any Statistic 272
9.6 Standardized Statistics 274
9.7 Student's *t*-Distribution: Replacing σ with s 276
9.8 Statistical Inference 278
- ◆ ***Case Study 9.1:*** Do Americans Really Vote When They Say They Do? 279

KEY TERMS 279
EXERCISES 280
REFERENCES 285

CHAPTER 10 Estimating Proportions with Confidence 286

10.1 The Language and Notation of Estimation 288
10.2 Margin of Error 289
10.3 Confidence Intervals 290
10.4 Calculating a Margin of Error for 95% Confidence 292
10.5 General Theory of Confidence Intervals for a Proportion 296
10.6 Choosing a Sample Size for a Survey 301
10.7 Using Confidence Intervals to Guide Decisions 302
- ◆ ***Case Study 10.1:*** Extrasensory Perception Works with Movies 304
- ◆ ***Case Study 10.2:*** Nicotine Patches versus Zyban® 305
- ◆ ***Case Study 10.3:*** What a Great Personality 306

KEY TERMS 306
EXERCISES 307
REFERENCES 311

CHAPTER 11 Testing Hypotheses about Proportions 312

11.1 Formulating Hypothesis Statements 314
11.2 The Logic of Hypothesis Testing: What if the Null Is True? 316
11.3 Reaching a Conclusion about the Two Hypotheses 318
11.4 Testing Hypotheses about a Proportion 320
11.5 The Role of Sample Size in Statistical Significance 331

11.6 Real Importance versus Statistical Significance 334
- ◆ ***Case Study 11.1:*** The Internet and Loneliness: Case Study 1.7 Revisited 335

11.7 What Can Go Wrong: The Two Types of Errors 335
- ◆ ***Case Study 11.2:*** An Interpretation of a *p*-Value Not Fit to Print 339

KEY TERMS 340
EXERCISES 340
REFERENCES 345

CHAPTER 12 More about Confidence Intervals 346

12.1 Examples of Different Estimation Situations 347
12.2 Standard Errors 350
12.3 Approximate 95% Confidence Intervals 354
12.4 General Confidence Intervals for One Mean or Paired Data 357
12.5 General Confidence Intervals for the Difference Between Two Means (Independent Samples) 363
12.6 The Difference Between Two Proportions (Independent Samples) 370
12.7 Understanding Any Confidence Interval 373
- ◆ ***Case Study 12.1:*** Confidence Interval for Relative Risk: Case Study 3.4 Revisited 373
- ◆ ***Case Study 12.2:*** Premenstrual Syndrome? Try Calcium 374

Summary of Formulas for Confidence Intervals 375
KEY TERMS 376
EXERCISES 376
REFERENCES 383

CHAPTER 13 More about Significance Tests 384

13.1 The General Ideas of Significance Testing 386
13.2 Testing Hypotheses about One Mean or Paired Data 387
13.3 Testing the Difference Between Two Means (Independent Samples) 396
13.4 Testing the Difference Between Two Population Proportions 404
13.5 The Relationship Between Significance Tests and Confidence Intervals 409
13.6 The Two Types of Errors and Their Probabilities 411

此为试读,需要完整PDF请访问: www.ertongbook.com

13.7 Evaluating Significance in Research Reports 415
Summary of Chapter 13 Procedures 416
KEY TERMS 416
EXERCISES 416
REFERENCES 423

CHAPTER 14 More about Regression 424

14.1 Sample and Population Regression Models 426
14.2 Estimating the Standard Deviation for Regression 431
14.3 Inference about the Linear Regression Relationship 435
14.4 Predicting the Value y for an Individual 439
14.5 Estimating the Mean y at a Specified x 441
14.6 Checking Conditions for Using Regression Models for Inference 443
◆ ***Case Study 14.1:*** A Contested Election 447
KEY TERMS 449
EXERCISES 450
REFERENCES 455

CHAPTER 15 More about Categorical Variables 456

15.1 The Chi-square Test for Two-Way Tables 457
15.2 Analyzing 2 × 2 Tables 468
15.3 Testing Hypotheses about One Categorical Variable: Goodness of Fit 473
◆ ***Case Study 15.1:*** Do You Mind if I Eat the Blue Ones? 477
KEY TERMS 479
EXERCISES 479
REFERENCES 436

CHAPTER 16 Analysis of Variance 488

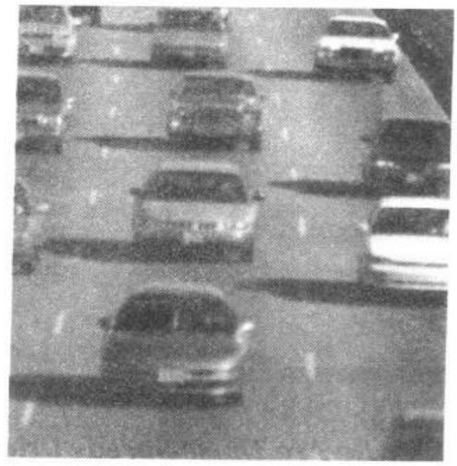
16.1 Comparing Means with an ANOVA F-Test 489
16.2 Details of One-Way Analysis of Variance 497
16.3 Other Methods for Comparing Populations 503
16.4 Two-Way Analysis of Variance 507
KEY TERMS 511
EXERCISES 511
REFERENCES 516

CHAPTER 17 Turning Information into Wisdom 518

17.1 Beyond the Data 519
17.2 Transforming Uncertainty into Wisdom 522
17.3 Making Personal Decisions 522
17.4 Control of Societal Risks 525
17.5 Understanding Our World 528
17.6 Getting to Know You 530
17.7 Words to the Wise 532
EXERCISES 533
REFERENCES 535

Appendix of Tables 537
Answers to Selected Exercises 543
Index 553

The seven stories in this chapter are meant to bring life to the term *statistics*. When you are finished reading these stories, if you still think the subject of statistics is lifeless or gruesome, check your pulse!



Let's face it. You're a busy person. Why should you spend your time learning about a subject that sounds as dull as statistics? In this chapter we give seven examples of situations in which statistics either provided enlightenment or misinformation. With these examples, we hope to convince you that learning about this subject will be interesting and useful.

Each of the stories in this chapter illustrates one or more concepts that will be developed throughout the book. These concepts are given as "the moral of the story" after a case is presented. Definitions of some terms used in the story also are provided following each case. By the time you read all of these stories, you already will have an overview of what statistics is all about. ◆

1.1 WHAT IS STATISTICS?

When you hear the word *statistics* you probably think of lifeless or gruesome numbers, like the population of your state or the number of violent crimes committed in your city last year. The word *statistics,* however, actually is used to mean two different things. The better-known definition is that statistics are numbers measured for some purpose. A more complete definition, and the one that forms the substance of this book, is the following:

Statistics is a collection of procedures and principles for gathering data and analyzing information in order to help people make decisions when faced with uncertainty.

The stories in this chapter are meant to bring life to this definition. When you are finished reading them, if you still think the subject of statistics is lifeless or gruesome, check your pulse!

1.2 SEVEN STATISTICAL STORIES WITH MORALS

The best way to gain an understanding of some of the ideas and methods used in statistical studies is to see them in action. Each of the seven stories presented in this chapter includes interesting lessons about how to gain information from data. The methods and ideas will be expanded throughout the book, but these seven stories will give you an excellent overview of why it is useful to study statistics. To help you understand some basic statistical principles, each case study is accompanied by a "moral of the story" and by some definitions. All of the ideas and definitions will be discussed in greater detail in subsequent chapters.

CASE STUDY 1.1

Who Are Those Speedy Drivers?

A survey taken in a large statistics class at Penn State University contained the question "What's the fastest you have ever driven a car? _____mph." The *data* provided by the 87 males and 102 females who responded are listed here.

Males: 110 109 90 140 105 150 120 110 110 90 115 95 145 140 110 105 85 95 100 115 124 95 100 125 140 85 120 115 105 125 102 85 120 110 120 115 94 125 80 85 140 120 92 130 125 110 90 110 110 95 95 110 105 80 100 110 130 105 105 120 90 100 105 100 120 100 100 80 100 120 105 60 125 120 100 115 95 110 101 80 112 120 110 115 125 55 90

Females: 80 75 83 80 100 100 90 75 95 85 90 85 90 90 120 85 100 120 75 85 80 70 85 110 85 75 105 95 75 70 90 70 82 85 100 90 75 90 110 80 80 110 110 95 75 130 95 110 110 80 90 105 90 110 75 100 90 110 85 90 80 80 85 50 80 100 80 80 80 95 100 90 100 95 80 80 50 88 90 90 85 70 90 30 85 85 87 85 90 85 75 90 102 80 100 95 110 80 95 90 80 90

From these numbers, can you tell which sex tends to have driven faster, and by how much? Notice how difficult it is to make sense of the *data* when you are simply presented with a list. Even if the numbers had been presented in numerical order, it would be difficult to compare the two sexes.

Your first lesson in statistics is how to formulate a simple summary of a long list of numbers. The *dotplot* shown in Figure 1.1 helps us see the pattern in the data. In the plot, each dot represents the response of an individual student. We can see that the men tend to claim a higher "fastest ever driven" speed than do the women.

The graph shows us a lot, and calculating some statistics that summarize the data will provide additional insight. There are a variety of ways to do so, but for this example we examine a *five-number summary* of the data for males and females. The five numbers are the lowest value, the cutoff points for $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ of the data, and the highest value. The three middle values of the summary (the cutoff points for $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ of the data) are called the *lower quartile, median,* and *upper quartile.* Five-number summaries can be represented like this:

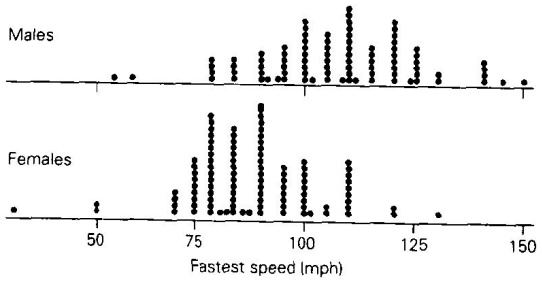


FIGURE 1.1 Responses to "What's the fastest you've ever driven?"

	Males (87 Students)		Females (102 Students)	
Median	110		89	
Quartiles	95	120	80	95
Extremes	55	150	30	130

Some interesting facts become immediately obvious from these summaries. By looking at the medians, you see that half of the men have driven 110 miles per hour or more, while the halfway point for the women is only 89 miles per hour. In fact, $\frac{3}{4}$ of the men have driven 95 miles per hour or more, but only $\frac{1}{4}$ of

the women have done so. These facts were not at all obvious from the original lists of numbers.

Moral of the Story: *Simple summaries of data can tell an interesting story and are easier to digest than long lists.*

Definitions: **Data** is a plural word referring to numbers or nonnumerical labels (such as male/female) collected from a set of entities (people, cities, and so on). The **median** of a numerical list of data is the value in the middle when the numbers are put in order. For an even number of entities, the median is the average of the middle two values. The **lower quartile** and **upper quartile** are (roughly) the medians of the lower and upper halves of the data.

CASE STUDY 1.2

Disaster in the Skies?

If you do a lot of air travel, or know anyone who does, you may have been disturbed by the headline in *USA Today* that read, "Planes get closer in midair as traffic control errors rise" (Levin, 1999). You may have been even more disturbed by the details: "Errors by air traffic controllers climbed from 746 in fiscal 1997 to 878 in fiscal 1998, an 18% increase." Don't cancel your next vacation yet. There is some reassurance when we are told that "most [errors] involve planes passing so far apart that there is no danger of collision."

The news does sound ominous—all those errors and an increase of almost 20%! If things continue at that rate, won't your next flight be quite likely to suffer from air traffic controller error? The answer is a resounding "no," which becomes obvious when we are told the *base rate* or *baseline risk* for errors. "The errors per million flights handled by controllers climbed from 4.8 to 5.5." So, the original *rate* of errors in 1997, from which the 18% increase was calculated, was less than 5 errors per million flights.

Fortunately, the rates for the two years were provided in the story. This is not always the case in news reports of increases in rates. For instance, an article may say that the rate of a certain type of cancer is doubled if you eat a certain unhealthful food. But what good is that information unless you know the actual risk? Doubling your chance of getting cancer from 1 in a million to 2 in a million is trivial, but doubling your chance from 1 in 50 to 2 in 50 is not.

Moral of the Story: *When discussing the* change *in the rate or risk of occurrence of something, make sure you also include the base rate or baseline risk.*

Definitions: The **rate** at which something occurs is simply the number of times it occurs per number of opportunities for it to occur. In fiscal year 1998, the rate of errors was 5.5 per million flights. The **risk** of a bad outcome, if the future is like the past, is estimated using the past rate for that outcome. Based on 1998 data, the risk of an error for any given flight in 1999 would be 5.5/1,000,000, or 0.0000055. The **base rate** or **baseline risk** is the rate or risk at a beginning time period or under specific conditions. For instance, the base rate of air traffic controller errors was 4.8 per million flights in fiscal year 1997.

CASE STUDY 1.3

Did Anyone Ask Whom You've Been Dating?

"According to a new *USA Today*/Gallup Poll of teenagers across the country, 57 percent of teens who go out on dates say they've been out with someone of another race or ethnic group" (Peterson, 1997). That's over half of the dating teenagers, so of course it was natural for the headline in the *Sacramento Bee* to read, "Interracial dates common among today's teenagers." The article contained other information as well, such as: "In most cases, parents aren't a major obstacle. Sixty-four percent of teens say their parents don't mind that they date interracially, or wouldn't mind if they did."

There are millions of teenagers in the United States whose experiences are being reflected in this story. How could the

polltakers manage to ask so many teenagers these questions? The answer is that they didn't. The article states that "the results of the new poll of 602 teens, conducted Oct. 13–20, reflect the ubiquity of interracial dating today. . . ." They asked only 602 teens? Could such a small sample possibly tell us anything about the millions of teenagers in the United States? The answer is "yes" if those teens constituted a *random sample* from the *population* of interest.

The featured statistic of the article is that "57 percent of teens who go out on dates say they've been out with someone of another race or ethnic group." Only 496 of the 602 teens in the poll said they date, so the 57% value is actually a percent based on 496 responses. In other words, the pollsters were using information from only 496 teenagers to estimate something about all teenagers. Figure 1.2 illustrates this situation.

How accurate could this sample possibly be? The answer may surprise you. The results of this poll are accurate to within a *margin of error* of about 4.5%. As surprising as it may seem, the true percentage of all teens in the United States who date interracially is reasonably likely to be within 4.5% of the reported percentage that's based only on the 496 teens asked! We'll be conservative and round the 4.5% margin of error up to 5%. The percent of all teenagers in the United States who date that would say they have dated interracially is likely to be in the range 57% ± 5%, or between 52% and 62%. (The symbol ± is read "plus and minus" and means that the value on the right should be added to and subtracted from the value on the left, to create an interval.)

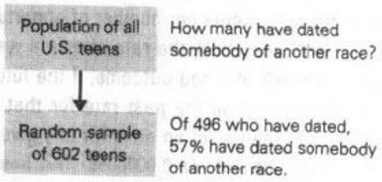


FIGURE 1.2 Population and sample for the survey

Polls and *sample surveys* are frequently used to assess public opinion and to estimate population characteristics like the percentage of teens who have dated interracially. Many sophisticated methods have been developed that allow pollsters to gain the information they need from a very small number of individuals. The trick is to know how to select those individuals. In Chapter 4 we examine a number of other strategies used to ensure that sample surveys provide reliable information about populations.

Moral of the Story: *A representative sample of only a few thousand, or perhaps even a few hundred, can give reasonably accurate information about a population of many millions.*

Definitions: A **population** is a collection of all individuals about which information is desired. The "individuals" are usually people, but could also be schools, cities, pet dogs, agricultural fields, and so on. A **random sample** is a subset of the population selected so that every individual has a specified probability of being part of the sample. In a **sample survey,** the investigators gather opinions or other information from each individual included in the sample. The **margin of error,** for a properly conducted survey, is a number that is added to and subtracted from the sample information to produce an interval that is 95% certain to contain the truth about the population. In the most common types of sample surveys, the margin of error is approximately equal to 1 divided by the square root of the number of individuals in the sample. Hence, a sample of 496 teenagers who have dated produces a margin of error of about $1/\sqrt{496} = 0.045$, or about 4.5%.

CASE STUDY 1.4
Who Are Those Angry Women?

A well-conducted survey can be very informative, but a poorly conducted one can be a complete disaster. As an extreme example, Moore (1997, p. 11) reports that for her highly publicized book *Women and Love,* Shere Hite sent questionnaires to 100,000 women asking about love, sex, and relationships. Only 4.5% of the women responded, and Hite used those responses to write her book. As Moore notes, "The women who responded were fed up with men and eager to fight them. For example, 91% of those who were divorced said that they had initiated the divorce. The anger of women toward men became the theme of the book." Do you think that women who were angry with men would be likely to answer questions about love relationships in the same way as the general population of women?

The Hite sample exemplifies one of the most common problems with surveys—the sample data may not represent the population. Extensive *nonresponse* from a random sample, or the use of a *self-selected* (i.e., *all-volunteer*) *sample,* will probably produce biased results. Those who voluntarily respond to surveys tend to care about the issue and thus

此为试读,需要完整PDF请访问: www.ertongbook.com