# Exploring and Understanding Data

# Stats Starts Here[1]

S tatistics gets no respect. People say things like "You can prove anything with Statistics." People will write off a claim based on data as "just a statistical trick." And Statistics courses don't have the reputation of being students' first choice for a fun elective.

But Statistics *is* fun. That's probably not what you heard on the street, but it's true. Statistics is about how to think clearly with data. A little practice thinking statistically is all it takes to start seeing the world more clearly and accurately.

## So, What Is (Are?) Statistics?

Q: What is Statistics?
A: Statistics is a way of reasoning, along with a collection of tools and methods, designed to help us understand the world.
Q: What are statistics?
A: Statistics (plural) are particular calculations made from data.
Q: So what is data?
A: You mean, "what *are* data?" Data is the plural form. The singular is datum.
Q: OK, OK, so what are data?
A: Data are values along with their context.

It seems every time we turn around, someone is collecting data on us, from every purchase we make in the grocery store, to every click of our mouse as we surf the Web. The United Parcel Service (UPS) tracks every package it ships from one place to another around the world and stores these records in a giant database. You can access part of it if you send or receive a UPS package. The database is about 17 terabytes big—about the same size as a database that contained every book in the Library of Congress would be. (But, we suspect, not *quite* as interesting.) What can anyone hope to do with all these data?

Statistics plays a role in making sense of the complex world in which we live today. Statisticians assess the risk of genetically engineered foods or of a new drug being considered by the Food and Drug Administration (FDA). They predict the number of new cases of AIDS by regions of the country or the number of customers likely to respond to a sale at the mall. And statisticians help scientists and social scientists understand how unemployment is related to environmental controls, whether enriched early education af-

---

[1] This chapter might have been called "Introduction," but nobody reads the introduction, and we wanted you to read this. We feel safe admitting this here, in the footnote, because nobody reads footnotes either.

fects later performance of school children, and whether vitamin C really prevents illness. Whenever there are data and a need for understanding the world, you need Statistics.

So our objectives in this book are to help you develop the insights to think clearly about the questions, use the tools to show what the data are saying, and acquire the skills to tell clearly what it all means.

> The ads say, "Don't drink and drive; you don't want to be a statistic." But you can't be a statistic.
> We say: "Don't be a datum."



*FRAZZ reprinted by permission of United Feature Syndicate, Inc.*

## Statistics in a Word

> Statistics is about variation.
> Data vary because we don't see everything and because even what we do see and measure, we measure imperfectly.
> So, in a very basic way, Statistics is about the real, imperfect world in which we live.

It can be fun, and sometimes useful, to summarize a discipline in only a few words. So,

Economics is about . . . *Money (and why it is good)*.

Psychology: *Why we think what we think (we think)*.

Biology: *Life*.

Anthropology: *Who?*

History: *What, where, and when?*

Philosophy: *Why?*

Engineering: *How?*

Accounting: *How much?*

In such a caricature, Statistics is about . . . **Variation.**

Data vary. People are different. We can't see everything, let alone measure it all. And even what we do measure, we measure imperfectly. So the data we wind up looking at and basing our decisions on provide, at best, an imperfect picture of the world. This fact lies at the heart of what Statistics is all about. How to make sense of it is a central challenge of Statistics.

## So, How Will This Book Help?

A fair question. Most likely, this book will not turn out to be quite what you expected.

What's different?

*Close your eyes and open the book to a page at random. Is there a graph or table on that page? Do that again, say, 10 times. We'll bet you saw data displayed in many ways, even near the back of the book and in the exercises.*

We can better understand everything we do with data by making pictures. This book leads you through the entire process of thinking about a problem, finding and showing results, and telling others about what you have discovered. At each of these steps, we display data for better understanding and insight.

You looked at only a few randomly selected pages to get an impression of the entire book. We'll see soon that doing so was sound Statistics practice and reasoning.

*Next, pick a chapter and read the first two sentences. (Go ahead; we'll wait.)*

We'll bet you didn't see anything about Statistics. Why? Because the best way to understand Statistics is to see it at work. In this book, chapters usually start by presenting a story and posing questions. That's when Statistics really gets down to work.

There are three simple steps to doing Statistics right: *think, show,* and *tell:*

**Think** first. Know where you're headed and why. It will save you a lot of work.

**Show** is what most folks think Statistics is about. The *mechanics* of calculating statistics and making displays is important, but not the most important part of Statistics.

**Tell** what you've learned. Until you've explained your results so that someone else can understand your conclusions, the job is not done.

The best way to learn new skills is to take them out for a spin. In **For Example** boxes you'll see brief ways to apply new ideas and methods as you learn them. You'll also find more comprehensive worked examples called **Step-by-Steps.** These show you fully worked solutions side by side with commentary and discussion, modeling the way statisticians attack and solve problems. They illustrate how to think about the problem, what to show, and how to tell what it all means. These step-by-step examples will show you how to produce the kind of solutions instructors hope to see.

Sometimes, in the middle of the chapter, we've put a section called **Just Checking** . . . . There you'll find a few short questions you can answer without much calculation—a quick way to check to see if you've understood the basic ideas in the chapter. You'll find the answers at the end of the chapter's exercises.

---

### MATH BOX

Knowing where the formulas and procedures of Statistics come from and why they work will help you understand the important concepts. We'll provide brief, clear explanations of the mathematics that supports many of the statistical methods in **Math Boxes** like this.

---

**TI Tips**    **Do statistics on your calculator!**

How do I use this thing?

Although we'll show you all the formulas you need to understand the calculations, you will most often use a calculator or computer to perform the mechanics of a statistics problem. Your graphing calculator has a specialized program called a "statistics package." Each chapter contains **TI Tips** that teach you how to use it (and avoid doing most of the messy calculations).

*"Get your facts first, and then you can distort them as much as you please. (Facts are stubborn, but statistics are more pliable.)"*
—Mark Twain

From time to time, you'll see an icon like this in the margin to signal that the *ActivStats* multimedia materials on the available DVD in the back of the book have an activity that you might find helpful at this point. Typically, we've flagged simulations and interactive activities because they're the most fun and will probably help you see how things work best. The chapters in *ActivStats* are the same as those in the text—just look for the named activity in the corresponding chapter.

If you are using TI-Nspire™ technology, these margin icons will alert you to activities and demonstrations that can help you understand important ideas in the text. If you have the DVD that's available with this book, you'll find these there; if not, they're also available on the book's Web site www.aw.com/bock.

One of the interesting challenges of Statistics is that, unlike in some math and science courses, there can be more than one right answer. This is why two statisticians can testify honestly on opposite sides of a court case. And it's why some people think that you can prove anything with statistics. But that's not true. People make mistakes using statistics, sometimes on purpose in order to mislead others. Most of the unintentional mistakes people make, though, are avoidable. We're not talking about arithmetic. More often, the mistakes come from using a method in the wrong situation or misinterpreting the results. Each chapter has a section called **What Can Go Wrong?** to help you avoid some of the most common mistakes.

**Time out.** From time to time, we'll take time out to discuss an interesting or important side issue. We indicate these by setting them apart like this.[2]

You'll find all sorts of stuff in margin notes, such as stories and quotations. For example:

*"Computers are useless. They can only give you answers."*
—Pablo Picasso

While Picasso underestimated the value of good statistics software, he did know that creating a solution requires more than just *Showing* an answer—it means you have to *Think* and *Tell,* too!

There are a number of statistics packages available for computers, and they differ widely in the details of how to use them and in how they present their results. But they all work from the same basic information and find the same results. Rather than adopt one package for this book, we present generic output and point out common features that you should look for. The . . . **on the Computer** section of most chapters (just before the exercises) holds this information. We also give a table of instructions to get you started on any of several commonly used packages, organized by chapters in Appendix B's Guide to Statistical Software.

At the end of each chapter, you'll see a brief summary of the important concepts you've covered in a section called **What Have We Learned?** That section includes a list of the **Terms** and a summary of the important **Skills** you've acquired in the chapter. You won't be able to learn the material from these summaries, but you can use them to check your knowledge of the important ideas in the chapter. If you have the skills, know the terms, and understand the concepts, you should be well prepared for the exam—and ready to use Statistics!

Beware: No one can learn Statistics just by reading or listening. The only way to learn it is to do it. So, of course, at the end of each chapter (except this one) you'll find **Exercises** designed to help you learn to use the Statistics you've just read about.

Some exercises are marked with an orange **T**. You'll find the data for these exercises on the DVD in the back of the book or on the book's Web site at www.aw.com/bock.

---

[2] Or in a footnote.

*"Far too many scientists have only a shaky grasp of the statistical techniques they are using. They employ them as an amateur chef employs a cookbook, believing the recipes will work without understanding why. A more* cordon bleu *attitude . . . might lead to fewer statistical soufflés failing to rise."*

—*The Economist,* June 3, 2004, **"Sloppy stats shame science"**

We've paired up the exercises, putting similar ones together. So, if you're having trouble doing an exercise, you will find a similar one either just before or just after it. You'll find answers to the odd-numbered exercises at the back of the book. But these are only "answers" and not complete "solutions." Huh? What's the difference? The answers are sketches of the complete solutions. For most problems, your solution should follow the model of the Step-By-Step Examples. If your calculations match the numerical parts of the "answer" and your argument contains the elements shown in the answer, you're on the right track. Your complete solution should explain the context, show your reasoning and calculations, and state your conclusions. Don't fret too much if your numbers don't match the printed answers to every decimal place. Statistics is more about getting the reasoning correct—pay more attention to how you interpret a result than what the digit in the third decimal place was.

In the real world, problems don't come with chapters attached. So, in addition to the exercises at the ends of chapters, we've also collected a variety of problems at the end of each part of the text to make it more like the real world. This should help you to see whether you can sort out which methods to use when. If you can do that successfully, then you'll know you understand Statistics.

## Onward!

It's only fair to warn you: You can't get there by just picking out the highlighted sentences and the summaries. This book is different. It's not about memorizing definitions and learning equations. It's deeper than that. And much more fun. But . . .

*You have to read the book!*[3]

---

[3] So, turn the page.