PART 1

The Basics
As I sat in my favorite coffee shop, latte in hand, wondering how to introduce this book, my mind drifted to the conversations around me. At the table to my right sat a couple of college guys, decked out in sweatshirts touting a nearby university. They were arguing baseball, speculating which team was most likely to win the pennant and make it to the World Series. To my left sat three middle-aged women, speaking in quiet voices I had to strain to hear. They were talking about menopause, comparing their own experiences, trying to sort through the conflicting news about which, if any, treatments actually alleviate the symptoms. Behind me was the liveliest conversation of all. Two men were talking politics. Both men seemed to agree about who should win the next presidential election, but that didn’t keep them from arguing. Many words were exchanged, but it came down to this. One of the men, citing a national poll, insisted his candidate was clearly going to be the winner. The other, citing yet another poll, claimed the outcome was anybody’s guess.

Aside from my tendency to eavesdrop, there’s a common theme to the three conversations. Whether they knew it or not, all of these people were talking statistics.

Most people run across statistics on a daily basis. In fact, in this age of instant information, it’s hard to get away from them. Drug
studies, stock market projections, sales trends, sports, education, crime reports: the list of places you’ll find them goes on and on. Any time somebody takes a large amount of information and reduces it down to a few bullet points, that person is using statistics. And even if you never look at any raw data, when you use those bullet points to make conclusions or decisions, you’re using statistics as well.

Being a statistician has never made me the life of the party. In fact, when I meet a new person, the reaction to my profession is almost universal. Here’s how a typical conversation might go.

**ME:** *How do you know John?*

**NEW FRIEND:** *We work together at XYZ Corp. I’m in sales, and John’s in marketing. How about you?*

**ME:** *Oh, we’ve worked together on a few projects as well. So, are you married?*

**NEW FRIEND:** *Yeah, sure. What projects?*

**ME:** *The Michelson account, the Trends Survey, a few others. How ‘bout kids? Do you have kids?*

**NEW FRIEND:** *Hey, I worked on the Michelson account, too. Ran the sales campaign. But I don’t remember you.*

**ME (glancing around the room):** *I was just a consultant.*

**NEW FRIEND:** *Hey, wait a minute. I do remember you. You told us we couldn’t launch the product in June . . . something about summer and . . .*

**ME (shrugging):** *Seasonal effects on customer demand. Yes, that’s me. I’m the statistician.*

*(There’s a long, awkward pause. New Friend eyes me suspiciously.)*

**NEW FRIEND:** *I always hated statistics.*

It may not be the life of the party, but when it comes to sorting through mounds of information, statistics is the belle of the ball. And it doesn’t take a graduate degree in the subject to know how to use it. If you can apply a few basic statistical tools and a little practical knowledge to a problem, people think you’re genius, and maybe even a little clairvoyant. These qualities may not draw crowds at the neighborhood mixer, but they do tend to result in big raises and big promotions.

Real-world statistics isn’t only about calculating an average and a standard deviation. And it’s not always a highly precise, exact science. Statistics involves gathering data and distilling large amounts
of information down to a reasonable and accurate conclusion. Most statistical analyses begin not with a dataset, but with a question. What will be the impact of our new marketing campaign? Does this drug work? Who’s most likely to win the next presidential election? Answering these questions takes more than a spreadsheet and a few formulas. It’s a process: reducing the question down to a manageable size, collecting data, understanding what the data are telling you, and yes, eventually making some calculations. Often this process is as much an art as it is a science. And it is this art, the art of data analysis, that provides you with the tools you need to understand your data.

There are no proofs in pages that follow. Mathematical formulas are kept to a bare minimum. Instead, this book deals with the practical and very real-world problem of data analysis. Each chapter asks a question and illustrates how it might be answered using techniques taught in any introductory statistics course. Along the way, common issues come up, issues such as:

- How to turn a vaguely worded question into a scientific study
- How different types of statistical analyses are well-suited to different types of questions
- How a well-chosen plot can do most of the data analysis for you
- How to identify the limitations of a study
- What happens if your data aren’t perfect
- How to avoid misleading or completely false conclusions

Every chapter is a case study, complete with a question, a data collection effort, and a statistical analysis. None of these case studies addresses society’s fundamental problems (unless you think the lack of appreciation for superhero sidekicks is one of them). None of them will help you improve your company’s sales (unless those sales are dependent on scientific proof that Bigfoot exists). And none of them will help you pick up women (especially not the one about gender stereotypes). On the other hand, all of them can be used as a template for your own data analysis, whether it be for a classroom project, a work-related problem, or a personal bet you just must win. And all of them illustrate how basic data analysis can be used to answer almost any question you can imagine.

The statistical techniques presented here can be found in most spreadsheet programs and basic data analysis software. I used Microsoft
Excel throughout, and in some cases, the Analysis Add-In pack was required. Here and there, a specific function is mentioned, but this isn’t a book on statistics using Excel. There are plenty of good texts covering that topic. Some of the most popular, written by a man known as Mr. Spreadsheet, are listed in the Bibliography at the end of this chapter.

The outline of this book follows a typical introductory statistics course. Part One gives you the basic tools you need to ask a question and design a study to answer it. Part Two shows what you can do with a solid understanding of these basic tools. Each chapter is self-contained, but like a typical textbook, the concepts build on one another, and the analyses gradually become more sophisticated as the book progresses. If you’re dying of curiosity and you’ve just got to find out when the zombie flu went viral, then go ahead and jump to Chapter 9. But if you can wait, I recommend you turn the page and read through the chapters in order.

I hope you enjoy reading these case studies as much as I enjoyed writing them.

BIBLIOGRAPHY