Introduction   **What This Book Is About**

If you have glanced at the Contents, you might be wondering what the topics in this book could possibly have in common. They include presidential and congressional elections, extramarital affairs, wine quality, college grades, marathon times, baseball performance, college football games, interest rates, and inflation. The answer is that they can all be explained and analyzed using the tools of the social sciences and statistics. The aim of this book is to allow those who are not necessarily well versed in these tools to see how this is done. The widely differing topics have been chosen to show the broad range of these tools and their strengths and weaknesses.

This book does not require that you be social scientists or statisticians or even that you like them. It is also fine if you don't know a Greek letter from a happy face emoticon. The book simply requires some patience in following the movement from a general idea of how something works to a specific prediction of what it will be in the future. The steps involved in this process can be explained without resort to technical material. By the end of the book, you should have a deeper understanding not only of the particular topics discussed but also of the way topics like these can be analyzed.

Knowledge of social science procedures allows a more critical reading of opinions and predictions that we are bombarded with every day. Is there any support for the particular opinion? How might the opinion

be tested? How much confidence can be placed in the prediction? The problem of evaluating views is harder than ever now that we are in the Internet age. Information is available at the click of a mouse, and views can be backed up by vast amounts of information. But is the information any good?

Consider, for example, the topic of U.S. presidential elections. What factors are important in deciding who will win an election? There are many views. Some stress the personalities of the candidates, others stress the amount of money available to the campaigns, still others stress the economy, and so on. Although each view has a story with it, the problem is that there are too many stories. How do we separate the wheat from the chaff? We need some way of deciding which stories have something to them and which do not. One way comprises tools of the social sciences and statistics, which this book discusses. We will see that past election results can be used in a systematic way to decide which stories appear to have merit. But we can do more than this. We can also use the information from past elections to *predict* the outcome of an election that has not yet taken place. We can both explain, in terms of telling a story that seems consistent with past results, and predict.

So, what are these mysterious tools? We begin with some question of interest, such as what determines the outcomes of presidential elections? Why did Richard Nixon beat George McGovern in 1972, and why did Ronald Reagan beat Jimmy Carter in 1980? We usually have some initial ideas about the answer to our question. The economy may play an important role in influencing how people vote, so an obvious preliminary idea is that the economy affects voting behavior. Another obvious initial idea is that an incumbent president running again may have an advantage because he can use the powers of the presidency to try to sway voters.

We call an idea or a set of ideas offered to explain something a *theory*. A theory may or may not be a good explanation. For example, a theory that says that people vote for a candidate solely on the basis of his or her height is not likely to be a good explanation of the way people actually vote. A theory need not be original with you, and it does not really matter where it came from. What is important is that there is some way to test whether the theory is any good.

How would one test the theory that the economy affects voting behavior? This is where past election results come into play. We can collect

data on the economy and on past election outcomes and see how closely the theory explains the data. Was it generally the case that the incumbent party won when the economy was good and lost when the economy was bad? This movement from proposing a theory to testing the theory using data is what most of this book is about. We will see that an important feature about testing theories is that once the tests have been performed, it is usually possible to move fairly easily to prediction.

The topics in this book have been chosen to appeal to a wide range of people. Political junkies and others interested in voting behavior should find the results on presidential and congressional elections helpful in understanding how people vote. Other dimensions of human behavior are covered in the chapters on extramarital affairs and college grades. In the chapter on extramarital affairs, we examine the factors that increase or decrease the chances that someone will have an affair. In the chapter on college grades, we examine how class attendance affects grades. As those of us who teach college students know, not every student shows up for every class. Does this matter in terms of grades and, if so, how much?

The chapter on wine quality shows that the quality a new wine eventually achieves after proper aging can be predicted fairly well simply by knowing the weather in the harvest year. This knowledge can help one decide whether a new wine is under- or overpriced and thus whether one should purchase it.

More serious investment issues are the concern of the chapters on interest rates and inflation. These chapters discuss how interest rates and inflation can be explained and predicted. This is the macroeconomic part of the book.

The chapter on marathon times should be of special interest to people older than age 35. It gives estimates of how fast people slow down as they age. If you are a runner older than 35 and have noticed that you are not quite as fast as you used to be, you can see if you are slowing down faster than you should be. The nice thing about this chapter is that it shows you need not feel bad just because you are slowing down. You only need to feel bad if you are slowing down too fast!

There are also chapters on baseball and football. The aim in the baseball chapter is to estimate aging effects. At what age do players peak, how fast do they improve up to the peak age, and how fast do they decline after the peak age? We will see that once we have estimated these effects,

we have a better measure than simple lifetime averages to rank players. We can also use the results to examine unusual, possibly drug-aided, performances at the older ages.

The aim in the football chapter is to see if there is independent information in the various college football computer rankings in predicting point spreads. Can we combine the various computer rankings to produce a better ranking than any one individual ranking? The answer is yes. We will see, however, that the Las Vegas betting spread completely dominates all the rankings, even the best combination. All useful information is in the betting spread. The betting markets are efficient in this sense.

This book may also appeal to college students taking introductory courses in the social sciences and statistics. It provides an intuitive discussion of the tools used in these courses and contains a number of examples of their use. It is often said that the three main interests of college students are sex, drinking, and sports, with perhaps grades ranked fourth. This book has all four!

The most difficult chapter in the book is Chapter 2, which discusses the tools. The chapter is divided into seven lessons, one for each day of the week. Lesson 4 on Thursday is the hardest. It explains a key test statistic, the *t-statistic*, showing why a *t*-statistic greater than about 2 is supportive of the theory being tested, whereas smaller values are not. If you are willing to accept this result—that large *t*-statistics support a theory but small ones do not—on faith, you can skip Lesson 4. You will miss an explanation of why this is true, but it should not hinder your reading the rest of the book. Also, if the material in Chapter 2 is completely new to you, you may want to read the chapter quickly the first time through and then come back to it as you go through the examples in Chapters 3 and beyond.

The tools in Chapter 2 are best explained with an example, and we use voting behavior in presidential elections as the example. Chapter 1 introduces this topic. It presents a theory of what determines votes for president and discusses the data that are available to test the theory. This sets the stage for Chapter 2, which explains the tools. Chapter 3 continues the discussion of presidential elections using the tools from Chapter 2. Each of the remaining chapters is a separate topic, each using the tools from Chapter 2. These chapters need not be read in order, although Chapter 11 on interest rates should be read before Chapter 12 on inflation. There is a

glossary at the end of the book of key words and concepts. The Chapter Notes at the end of the book give references.

In his advice to Harvard students in *Under Which Lyre*, W. H. Auden wrote:

> Thou shalt not answer questionnaires
> Or quizzes upon World-Affairs,
>   Nor with compliance
> Take any test. Thou shalt not sit
> With statisticians nor commit
>   A social science.

Alas, I am giving the opposite advice. Come sit with statisticians and social scientists for a while and see what they can do.