Preface

In today's complicated world, very few issues are clear-cut and without controversy. In order to understand and form an opinion about an issue, one must usually gather information, or data. To learn from data, one must know something about statistics, which is the art of learning from data.

This introductory statistics text is written for college-level students in any field of study. It can be used in a quarter, semester, or full-year course. Its only prerequisite is high school algebra. Our goal in writing it is to present statistical concepts and techniques in a manner that will teach students not only how and when to utilize the statistical procedures developed, but also to understand why these procedures should be used. As a result we have made a great effort to explain the ideas behind the statistical concepts and techniques presented. Concepts are motivated, illustrated, and explained in a way that attempts to increase one's intuition. It is only when a student develops a feel or intuition for statistics that she or he is really on the path toward making sense of data.

To illustrate the diverse applications of statistics and to offer students different perspectives about the use of statistics, we have provided a wide variety of text examples and problems to be worked by students. Most refer to real-world issues, such as gun control, stock price models, health issues, driving age limits, school admission ages, public policy issues, gender issues, use of helmets, sports, disputed authorship, scientific fraud, and Vitamin C, among many others. Many of them use data that not only are real but are themselves of interest. The examples have been posed in a clear and concise manner and include many thought-provoking problems that emphasize thinking and problem-solving skills. In addition, some of the problems are designed to be open-ended and can be used as starting points for term projects.

SOME SPECIAL FEATURES OF THE TEXT

Introduction   The first numbered section of each chapter is an introduction that poses a realistic statistical situation to help students gain perspective on what they will encounter in the chapter.

Statistics in Perspective   Statistics in Perspective highlights are placed throughout the book to illustrate real-world application of statistical techniques and concepts. These perspectives are designed to help students analyze and interpret data while utilizing proper statistical techniques and methodology.

Real Data   Throughout the text discussions, examples, perspective highlights, and problems, real data sets are used to enhance the students' understanding of the material. These data sets provide information for the study of current issues in a variety of disciplines, such as health, medicine, sports, business, and education.

Historical Perspectives   These enrichment sections profile prominent statisticians and historical events, giving students an understanding of how the discipline of statistics has evolved.

Problems/Review Problems   This text includes hundreds of exercises placed at the end of each section within a chapter, as well as more comprehensive review problems at the end of each chapter. Many of these problems utilize real data and are designed to assess the students' conceptual as well as computational understanding of the material. Selected problems are open-ended and offer excellent opportunity for extended discussion, group activities, or student projects.

Summary/Key Terms   An end-of-chapter summary provides a detailed review of important concepts and formulas covered in the chapter. Key terms and their definitions are listed that serve as a working glossary within each chapter.

Formula Summary   Important tables and formulas that students often refer to and utilize are included on the inside front and back covers of the book. These can serve as a quick reference when doing homework or studying for an exam.

Program CD-ROM   A CD-ROM is provided with each volume that includes programs that can be used to solve basic statistical computation problems. Please refer to Appendix E for a listing of these programs.

THE TEXT

In Chap. 1 we introduce the subject matter of statistics and present its two branches. The first of these, called descriptive statistics, is concerned with the collection, description, and summarization of data. The second branch, called inferential statistics, deals with the drawing of conclusions from data.

Chapters 2 and 3 are concerned with descriptive statistics. In Chap. 2 we discuss tabular and graphical methods of presenting a set of data. We see that an effective presentation of a data set can often reveal certain of its essential features. Chap. 3 shows how to summarize certain features of a data set.

In order to be able to draw conclusions from data it is necessary to have some understanding of what they represent. For instance, it is often assumed that the data constitute a "random sample from some population." In order to understand exactly what this and similar phrases signify, it is necessary to have some understanding of probability, and that is the subject of Chap. 4. The study of probability is often a troublesome issue in an introductory statistics class because many students find it a difficult subject. As a result, certain textbooks have chosen to downplay its importance and present it in a rather cursory style. We have chosen a different approach and attempted to concentrate on its essential features and to present them in a clear and easily understood manner. Thus, we have briefly but carefully dealt with the concept of the events of an experiment, the properties of the probabilities that are assigned to the events, and the idea of conditional probability and independence. Our study of probability is continued in Chap. 5, where discrete random variables are introduced, and in Chap. 6, which deals with the normal and other continuous random variables.

Chapter 7 is concerned with the probability distributions of sampling statistics. In this chapter we learn why the normal distribution is of such importance in statistics.

Chapter 8 deals with the problem of using data to estimate certain parameters of interest. For instance, we might want to estimate the proportion of people who are presently in favor of congressional term limits. Two types of estimators are studied. The first of these estimates the quantity of interest with a single number (for instance, it might estimate that 52 percent of the voting population favors term limits). The second type provides an estimator in the form of an interval of values (for instance, it might estimate that between 49 and 55 percent of the voting population favors term limits).

Chapter 9 introduces the important topic of statistical hypothesis testing, which is concerned with using data to test the plausibility of a specified hypothesis. For instance, such a test might reject the hypothesis that over 60 percent of the voting population favors term limits. The concept of p value, which measures the degree of plausibility of the hypothesis after the data have been observed, is introduced.

Whereas the tests in Chap. 9 deal with a single population, the ones in Chap. 10 relate to two separate populations. For instance, we might be interested in testing whether the proportions of men and of women that favor term limits are the same.

Probably the most widely used statistical inference technique is that of the analysis of variance; this is introduced in Chap. 11. This technique allows us to test inferences about parameters that are affected by many different factors. Both one- and two-factor analysis of variance problems are considered in this chapter.

In Chap. 12 we learn about linear regression and how it can be used to relate the value of one variable (say, the height of a man) to that of another (the height of his father). The concept of regression to the mean is discussed, and the regression fallacy is introduced and carefully explained. We also learn about the relation between regression and correlation. Also, in an optional section, we use regression to the mean along with the central limit theorem to present a simple, original argument to explain why biological data sets often appear to be normally distributed.

In Chap. 13 we present goodness-of-fit tests, which can be used to test whether a proposed model is consistent with data. This chapter also considers populations classified according to two characteristics and shows how to test whether the characteristics of a randomly chosen member of the population are independent.

Chapter 14 deals with nonparametric hypothesis tests, which are tests that can be used in situations where the ones of earlier chapters are inappropriate. Chapter 15 introduces the subject matter of quality control, a key statistical technique in manufacturing and production processes.

NEW TO THIS EDITION

The third edition has many new and updated examples and exercises. In addition, there is a new subsection (12.11.1) on the use of dummy variables in multiple regression models. There is also a new section (14.6) on the use of the Kruskal-Wallis nonparametric test of the equality of multiple probability distributions, with a subsection (14.6.1) giving a discussion of the Freedman test which can be used to test this hypothesis when the data are comparison rankings. There is also a new section (14.7) on the class of nonparametric tests known as permutation tests.