PREFACE

This book is an introduction to statistics for students who have completed at least one quarter or semester of calculus. This includes students of mathematics; statistics; computer science; the physical, biological, and social sciences; economics; and management science.

Statistics is the body of knowledge that deals with inductive reasoning. In our everyday life, we use "common sense" to deal with situations where we have to act without having all of the facts needed to make decisions with complete certainty. Statistics is "common sense" raised to the level of a formal study.

A decision on whether a vaccine is effective, or whether an electrical switch will last the life of an appliance, or whether a new teaching method improves reading skills of children cannot be based on wishful thinking or haphazard judgments. Ideally, sources of variability are identified, a controlled experiment is performed, and the plausibility of various conclusions are measured and assessed before a decision is made.

Yet no discipline is misused and misunderstood as much as statistics. We see evidence of this every day in the media and even in scientific papers. Nowadays, running data through statistical software is easy. **But understanding the purposes and the limitations of statistical procedures as well as the correct interpretation of the results is essential to any meaningful application.**

Thus, an understanding of statistical principles is essential for nearly all science and mathematics programs. Elementary statistics courses taught to nonscience majors do not integrate the student's knowledge of calculus and would not satisfy this need. Generally, the only other option is a course in mathematical statistics, which typically consists of one semester of probability followed by one semester of statistics. To include this in the curriculum of a typical science major, could only be done at the expense of some other requirement and would thus distort the balance in the program. Consequently, most mathematics and science programs cannot include statistics in the required part of their curricula.

In teaching our course, we struggled for some time with picking topics from various two-term textbooks. Neither we nor our students found this satisfactory. We decided that a new kind of textbook was needed that was specifically designed for a calculus-based one-semester or one-quarter statistics course. This is the result. The book has been extensively tested by several instructors over several years. Most of the book can be covered in a one-term course intended for science and mathematics students.

KEY FEATURES

Inference before midterm. The traditional format of books at this level is to develop the necessary mathematical probability tools before introducing statistical inference. Unfortunately, postponing this statistical inference to the end of the term may have negative consequences in that often what is learned last is first forgotten. The approach of this book is to introduce statistical inference concepts early and often. These ideas are reinforced by repeatedly reworking old ground at ever-increasing levels of awareness.

Statistical inference is introduced in the first chapter. Then, because students have been introduced to error analysis, precision, and confidence, it is possible to discuss standard error as early as Chapter **??**. This allows students the time to get comfortable working with standard errors and related concepts well before the formal study of sampling experiments in Chapters **??**.

Immediately after binomial and hypergeometric probability models have been studied in Chapter **??**, the formal vocabulary and structure of statistical hypothesis tests are introduced in Chapter **??** in the context of success–failure experiments. We start with single-sample experiments, extend to paired samples, and then to two-sample experiments. Tests such as the **sign test, binomial exact test**, and **Fisher's exact test** are developed here. We think that this approach is the best one for science students for the following reasons:

1. The road from the Introduction to Inference is as short as possible. Students see real statistics using real data without having to endure lengthy and unmotivated theory.

2. Procedures for these tests are intuitive and easily understood. Students thus find the resulting test decisions convincing.

3. Statistical inference at this stage is based only on the elementary binomial and hypergeometric probability models. The Central Limit Theorem is not needed until later (Chapter **??**).

4. All of this can be completed well before midterm. In our approach, the more commonly used z and t tests are introduced in Chapter **??**. By this time, many of the basics of inference are already familiar to the students. This approach also allows for a more complete discussion of confidence interval methods in Chapter **??** than is found in competing texts. For example, we discuss small-sample confidence intervals for proportions and explicitly discuss the coverage curves associated with interval estimation procedures.

5. Finally, in Chapter **??** we revisit inference through the methods of **resampling**.

And all of this can be accomplished in a single term.

Instructors who prefer to use the traditional approach of completing the study of probability models and large sample distributions before beginning the study of formal statistical inference can easily accomplish this by postponing Chapter **??** until after the Central Limit Theorems, normal approximations (Chapter **??**), and sampling distributions of statistics (Chapter **??**) have been completed.

Students who have completed a "short" or reform calculus course are prepared for the calculus used in the book. Multivariate calculus and linear algebra are not required. For example, Chapter **??**, which covers two or more random variables, can be understood without a background in multivariate calculus; the discussion is restricted to discrete random variables with the explanation that all results apply to the continuous case as well. The statistical ideas covered in this book are not too different from those in introductory texts for students who are not mathematically prepared. We do not believe that it is necessary to have a calculus background to understand these statistical ideas. Rather, we believe that students who have a calculus background should have the opportunity to see statistics in this setting. Calculus is used to build standard models. We give our readers the chance to hone their skills by exploring simple probability models used in statistical inference.

Examples are targeted for a science curriculum. The order of presentation in this book is generally to start each new topic with examples that lead to a discussion of new concepts, definitions, and specific results. These are then followed by a more formal treatment of the topic. There are over 150 worked-out examples and over 700 exercises in this book. Almost half of the exercises have answers in the back of the book, and a *Student's Solution Manual* is available. A CD-ROM of all solutions is available to instructors. We try to structure them on science themes. Many are based on laboratory measurements and error analysis. Bayes' Theorem is explained in context of diagnosis of illness, forecasting weather, and predicting the economy. Fisher's exact test and the binomial exact test are presented in the context of public health experiments and drug trials. Tests of hypotheses involve computer programs, physics, chemistry, archaeology, and psychology.

Sections that involve worked-out examples are generally followed by section exercises. At the end of each chapter are review exercises. Nearly half of the exercises have answers in the back of the book. Furthermore, solutions to selected exercises are available on the website

http://www.stat.luc.edu/StatisticsForTheSciences

Exercises are marked by symbols as follows:

$\bullet$ indicates that there is an answer in the back of the book.

$\star$ indicates a thought-provoking question.

C involves programming a graphing calculator or computer.

The separator

* * * * * * * * * *

is followed by exercises that go beyond the basics.

Probability theory and theories of inference have been kept to a minimum. Our goal is to develop statistical intuition and statistical common sense. We have omitted topics such as moment-generating functions and likelihood functions because they would seriously reduce the time one can spend on statistical inference and exploratory data analysis. We explore relatively few probability models so that sufficient time can be spent in their development.

Technology is used, but the presentation is independent of specific software or technology. It is assumed that students have access to a graphing calculator such as the TI-83 Plus[1], statistical software such as Minitab[2], or a spreadsheet such as Microsoft Excel[3]. In Chapter **??** on resampling methods, Maple[4] is also used. However, the treatment in the book is independent of the technology used. Subsections entitled **Notes on Technology** are included next to exercise sets throughout the text to give students instructions on how to use Minitab, Excel, and the TI-83 Plus graphing calculator. This frees the instructor from having to lecture on which buttons to press.

[1] TI-83 Plus is a registered trademark of Texas Instruments Incorporated.
[2] Minitab is a registered trademark of Minitab Incorporated.
[3] Microsoft Excel is a registered trademark of Microsoft Corporation.
[4] Maple is a registered trademark of Waterloo Maple Incorporated.

Graphing-calculator technology or computer software is used to simulate models and to explore the relationships between data and models.

Computer simulations and computer-drawn graphs are used to illustrate sampling distributions. Emphasis is on data analysis with careful attention paid to the assumptions behind statistical inference. Some of the data sets we use are available to the students online at StatLib, hosted by the Statistics Department at Carnegie Mellon University, and in particular at The Data and Story Library, DASL. Other data sets are available online through various federal government agencies. We also use the data in Paul Meier's essay *The Biggest Public Health Experiment Ever: The 1954 Field Trial of the Salk Poliomyelitis Vaccine*, which is available on the Web at http://www.stat.luc.edu, and other data drawn from various journal and newspaper articles.

This book is a gateway to further courses in statistics. We hope that this text will motivate some students to take additional courses in statistics. Students who complete a course from this textbook should be ready for most courses in applied statistics, such as regression analysis, design and analysis of statistical experiments, stochastic processes, and data analysis. Furthermore, this course will complement research methods courses in various disciplines.

The syllabus is flexible. There is flexibility in the syllabus of a course using this textbook. Most of the book can be covered in one semester.

- Some sections can be skipped and others emphasized at the instructor's preference. Usually, the last section of each chapter can be covered or skipped at the option of the instructor. Some interior sections are also marked as optional.

- Chapter **??** introduces testing of hypotheses immediately after the study of binomial and hypergeometric random variables. For those who prefer a more traditional order, these tests—including the sign, binomial exact, and Fisher's exact tests—can be presented at the same time as the z and t tests in Chapter **??**. Also, Chapter **??** (Inference with Categorical Data) can precede Chapter **??** (Correlation and Regression).

- Chapter **??** (**Waiting Time Random Variables**) on probability models involving Poisson and exponential random variables can be postponed to the end of the term.

- Time spent on Chapter **?? Two or More Random Variables** can be reduced by discussing only the boxed results of Section **??** along with some hand waving.

Acknowledgments. A special thanks to E. N. Barron, Anthony Giaquinto, Steven L. Jordan, Richard Maher, Henry Park, Hans-Juergen Petersen, and Alan Saleski as well as their students for classroom testing earlier versions of this text and for providing numerous suggestions. We also wish to thank the reviewers for their numerous insights and suggestions. Finally, many thanks to Carolyn Crockett, Kipp Blackburn, Joseph Rogove, Steven Summerlight, Ann Day, and others on the editorial staff of Duxbury Press for their encouragement and support during the writing of this book.