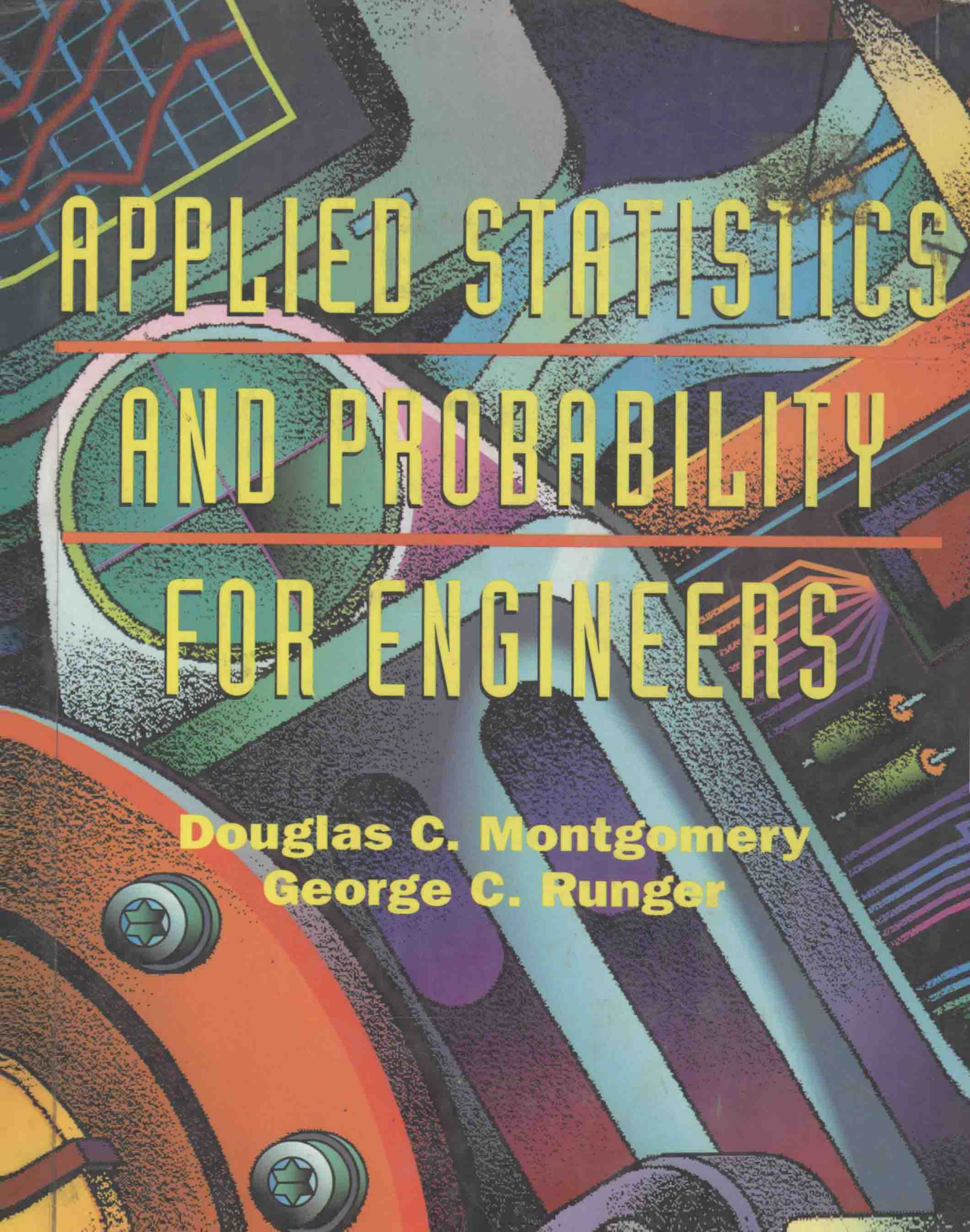

APPLIED STATISTICS AND PROBABILITY FOR ENGINEERS

Douglas C. Montgomery
George C. Runger

Applied Statistics and Probability for Engineers

Douglas C. Montgomery
Arizona State University

George C. Runger
Rensselaer Polytechnic Institute

John Wiley & Sons, Inc.
New York • Chichester • Brisbane • Toronto • Singapore

Acquisitions Editor	Charity Robey
Marketing Manager	Susan Elbe
Senior Production Editor	Nancy Prinz
Designer	Lynn Rogan
Manufacturing Manager	Andrea Price
Illustration Coordinators	Anna Melhorn, Rosa Bryant

This book was set in Times Roman by Monotype Composition Company and printed and bound by R. R. Donnelley & Sons Company, Crawfordsville. The cover was printed by Phoenix Color Corp.

Recognizing the importance of preserving what has been written, it is a policy of John Wiley & Sons, Inc. to have books of enduring value published in the United States printed on acid-free paper, and we exert our best efforts to that end.

Library of Congress Cataloging-in-Publication Data

Montgomery, Douglas C.

Applied statistics and probability for engineers / Douglas C. Montgomery, George C. Runger.

p. cm.

Includes bibliographical references and index.

ISBN 0-471-54041-2

1. Statistics. 2. Probabilities. I. Runger, George C. II. Title.

QA276.12.M645 1994 93-29954

519.5—dc20 CIP

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Preface

American industry must continue to improve the quality of its products and services if it is to continue to compete effectively in both domestic and world markets. A significant portion of this quality improvement effort will be driven by engineers and scientists, because these are the individuals who design and develop new products, and new manufacturing systems and processes, and who improve existing systems. Statistical methods are an important tool in these activities because they provide the engineer with both descriptive and analytical methods for dealing with the variability in observed data.

This is an introductory textbook for a first course in applied statistics and probability for undergraduate students in engineering and the physical or chemical sciences. While many of the methods we present are fundamental to statistical analysis in other disciplines, such as business or management, the life sciences, and the social sciences, we have elected to focus on an engineering-oriented audience. We believe that this approach will best serve engineering students and allow them to concentrate on the many applications of statistics in these disciplines. We have worked hard to ensure that all of our examples and exercises are engineering-based, and in almost all cases we have used examples of real data—either taken from a published source or based on our consulting experiences.

We believe that engineers in all disciplines should take at least one course in statistics. Unfortunately, because of other requirements, most engineers will only take one statistics course. We have written this book for that single course, although we have provided enough material for two courses in the hope that more engineering students will see the important applications of statistics in their everyday work and elect a second course. We believe that the book will also serve as a useful reference.

ORGANIZATION OF THE BOOK

The book has a very modest mathematical level; engineering students who have completed one or two semesters of calculus should have no difficulty reading almost all of the text. It is our intent to give the reader an understanding of the methodology and how to apply it, not the mathematical theory.

Chapter 1 is an introduction to the field of statistics and a presentation of basic methods of descriptive statistics. In addition to the relatively standard topics (stem-and-leaf displays, box plots, histograms, etc.), we have also included time sequence plots, digidot plots, and the control chart. Much engineering data are time oriented, and the student should gain some basic appreciation of that from this chapter.

Chapters 2, 3, 4, and 5 cover the basic concepts of probability, discrete and continuous random variables, expected values, joint probability distributions, and independence. We have given a reasonably complete treatment of these topics but have avoided many of the mathematical or more theoretical details.

Chapters 6, 7, and 8 present the basic tools of statistical inference; point estimation, interval estimation, and hypothesis testing. Our presentation is distinctly applications-oriented. We want the student to become interested in how these methods can be used to solve real-world engineering problems and to get some understanding of the concepts behind them. We give a logical, heuristic development of these concepts rather than a formal mathematical one.

Chapters 9 and 10 present simple and multiple linear regression. We use matrix algebra throughout the multiple regression material (Chapter 10), because, quite frankly, it is the only easy way to understand the concepts presented. Scalar arithmetic presentations of multiple regression are awkward at best, and we have found that undergraduate engineers are exposed to enough matrix algebra to understand the presentation of this material.

Chapters 11 and 12 deal with single-factor and multifactor experiments, respectively. The notions of randomization, blocking, factorial designs, interactions, graphical data analysis, and fractional factorials are emphasized. Chapter 13 gives a brief introduction to the methods and applications of nonparametric statistics, whereas Chapter 14 introduces the reader to statistical quality control, emphasizing the control chart and the fundamentals of statistical process control.

In addition to the usual collection of statistical tables and charts, we have also provided some supplemental technical material in the Appendix. This material includes an introduction to moment-generating functions, the change of variable technique, permutations and counting methods, development of the t- and F-distributions, Bayes estimation, and the likelihood ratio principle. This material may be of interest to some instructors and students, and we have provided it as a reference.

Each chapter has an extensive collection of exercises, including end-of-section exercises that emphasize the material in that section; supplemental exercises at the end of the chapter that cover the scope of chapter topics; and mind-expanding exercises that often require the student to extend the text material somewhat or to apply it in a novel situation.

USING THE BOOK

This is a very flexible textbook, because instructors' ideas about what should be in a first course on statistics for engineers vary widely, as do the abilities of different groups of students. Therefore, we hesitate to give too much advice but will explain how we use the book.

We believe that a first course in statistics for engineers should be primarily *an applied statistics course,* not a probability course. In our one-semester course we cover all of Chapter 1 (in three to four lectures), overview the material on probability, putting most of the emphasis on the normal distribution (six to eight lectures), discuss most of Chapters 7 and 8 on confidence intervals and tests (ten lectures), introduce regression models in Chapter 9 (four lectures), give an introduction to the design of experiments from Chapters 11 and 12 (six lectures), and present the basic concepts of statistical process control, including the Shewhart control chart from Chapter 14 (six lectures). This leaves about three to four periods for exams and review. Let us emphasize that the purpose of this course is to introduce engineers to how statistics can be used to solve real-world engineering problems, not to weed out the less mathematically gifted students. This course is not the "baby math-stat" course that is all too often given to engineers.

If a second semester is available, then it is possible to cover the entire book, including some of the appendix material, if appropriate for the audience. It would also be possible to assign and work many of the homework problems in class to reinforce understanding of the concepts. Obviously, multiple regression and more design of experiments would be major topics in a second course.

USING THE COMPUTER

In practice, engineers use computers to apply statistical methods to solve problems. Therefore, we strongly recommend that the computer be integrated into the class. Throughout the book we have presented output from Statgraphics and SAS as typical examples of what can be done with modern statistical software. In teaching, we have used not only these packages but others as well, such as EXECU-STAT, MINITAB, DESIGN-EASE, and SPSS. We did not clutter up the book with examples from many different packages, because how the instructor integrates the software into the class is ultimately more important than which package is used. All text data is available on computer disk.

In our own classrooms, we bring a notebook PC and liquid-crystal display pad to almost every lecture and show how the technique is implemented in the computer as soon as it is discussed in lecture. Student versions of many statistical software packages are available at low cost, and students can either purchase their own copy or use the products available on the PC local area networks. We have found that this greatly improves both the pace of the course and student understanding of the material.

ACKNOWLEDGMENTS

We would like to express our grateful appreciation to the many organizations and individuals who have contributed to this book. Several reviewers provided many helpful suggestions, including Dr. Dwayne Rollier, Arizona State University; Dr. Smiley Cheng, University of Manitoba; Dr. John Wilkinson, Rensselaer Polytechnic Institute; and Dr. J. E. Mann, Virginia Polytechnic Institute. We thank Mr. Dale Kennedy for his editorial assistance. We are also indebted to Dr. Smiley Cheng for permission to adapt many of the statistical tables from his excellent book (with Dr. James Fu), *Statistical Tables for Classroom and Exam Room.* John Wiley and Sons, Prentice Hall, the Biometrika Trustees, The American Statistical Association, the Institute of Mathematical Statistics, and the editors of *Biometrics* allowed us to use copyrighted material, for which we are grateful.

Contents

1 Introduction and Descriptive Statistics

1-1 AN OVERVIEW OF STATISTICS AND PROBABILITY

1-1.1 What Is Statistics?

The field of statistics deals with the collection, presentation, analysis, and use of data to make decisions and solve problems. Everyone, both in professional careers and in everyday life through contact with newspapers, television, and other media, is presented with information in the form of data. We often need to draw some conclusion from the information in the data, and so some understanding of statistics would be helpful to anyone. Since engineers and scientists routinely engage in obtaining and analyzing data, knowledge of statistics is especially important in these fields. Specifically, knowledge of statistics and probability can be a powerful tool to help engineers and scientists in designing new products and systems, improving existing designs, and designing, developing, and improving production processes. This book seeks to equip engineers and scientists with the basic statistical tools to successfully practice these aspects of their professions.

1-1.2 The Role of Statistics in Engineering and Science

The importance of statistics in engineering, science, and management has been underscored by the involvement of American industry in quality improvement. Many American companies have realized that poor product quality in the form of manufacturing defects and/or unsatisfactory product reliability and field performance dramatically affects their overall productivity, their market share and competitive position, and ultimately their profitability. Improving these aspects of quality can eliminate waste; reduce scrap and rework, the requirements for inspection and test, and warranty losses; enhance customer satisfaction; and enable the company to become the high-quality, low-cost producer in its market. Statistics is a critical skill in quality improvement, because statistical techniques can be used to describe and understand **variability.**

Virtually all real-world processes and systems exhibit variability. For example, consider a situation where we select several aircraft engine components from a machining process and measure the vane height (a critical dimension) on each part. If the measuring instrument has sufficient resolution, the vane heights will be different on each part; that is, there will be variability in the dimension. Alternatively, if we count the number of defects in personal computer cabinets, we will find variability in the counts, for some cabinets will have few defects and others will have many. This notion of variability extends to all environments. There is variability in the thickness of oxide coatings on silicon wafers, the hourly yield of a chemical process, the number of errors on engineering drawings, and the flow time required to assemble an automobile engine.

Why does variability occur? Generally, variability is the result of changes in the conditions under which observations are made. In a manufacturing context, these changes may be differences in material properties, differences in the way people do the work, differences in process variables, such as temperature, pressure, or holding time, and differences in environmental factors, such as relative humidity. Variability also occurs because of the measurement system. As an illustration, the measurement obtained from a scale may depend on where the test item is placed on the pan. Sampling may also cause variability. For example, suppose that a lot of 5000 integrated circuit devices has exactly 50 defective chips. If we inspected all 5000 devices, and if our inspection process was perfect (containing no inspection or measurement error), we would find all 50 defective devices. However, suppose that we select a sample of 100 devices. Now some of the devices in the sample will likely be defective. In fact, we would expect the sample to be about 1 percent defective (because the lot contains $50/5000 \times 100 = 1\%$ defective), but it could be zero percent or 2 percent or 5 percent defective, depending on the specific devices selected in the sample. Thus, the process of sampling has produced variability in the observed results in the sense that the observed proportion of defective units may vary from the actual proportion of defective units.

The field of statistics and probability consists of methods for describing and modeling variability, and for making decisions when variability is present. In **inferential statistics,** we usually want to make a decision about a particular population. The term *population* refers to the collection of measurements on all elements of a universe about which we wish to draw conclusions or make decisions. For example, the population may consist of the lot of 5000 integrated circuit devices referred to previously. Suppose that the manufacturer is interested in the transistor gain for a particular circuit in each device. We may think of the transistor gain levels in the devices as the population of interest. In this case, each population value is a numerical measurement, such as 5.10 or 5.24; consequently, the data are referred to as *variables* or *numerical data.* On the other hand, the manufacturer may be interested in whether or not each device produces a gain that conforms to the requirement. Then we think of the population as consisting of *attribute data,* in which each device is assigned a value of one if the unit is nonconforming and a value of zero if it conforms to requirements. In this book we will present techniques for dealing with both variables and attribute data.

In most applications of statistics, the available data consist of a **sample** from the population of interest. This sample is just a subset of observations selected from the population. In the integrated circuit example, suppose that the sample consists of five devices, selected from the lot of 5000 devices. The transistor gains observed in these devices are 5.10, 5.24, 5.13, 5.19, and 5.08. We might be interested in questions such as "Does the information in this sample conclusively demonstrate that transistor gain is less than 5.50?", or "How confident can we be that transistor gain is in the interval between 5.00 and 5.50?". The methods of **inferential statistics** are used to answer questions such as these.

The field of inferential statistics has developed primarily since the early 1900s. It is an outgrowth of methods for organizing and summarizing data whose origins are much earlier, going back hundreds of years. These methods for organizing and summarizing data are called **descriptive statistics.** Most of the modern use of statistics, particularly in engineering and the sciences, focuses on inference rather than description. For example, an engineer who designs a new computer chip will manufacture a sample or prototypes and will want to draw conclusions about how all of these devices will work once they are in full-scale production.

The primary focus of this book is on the inferential statistical techniques that are most useful to engineers and scientists. The rest of this chapter presents some useful descriptive statistical techniques. Chapters 2–5 present the basic concepts of **probability.** Knowledge of probability builds the base that enables us to understand how statistical inference and decision-making techniques are developed, why they work, and how the conclusions from these procedures can be presented and interpreted correctly. Thus, probability is the language and mathematical foundation of inferential statistics in the same sense that the rules of English grammar provide the foundation for organizing thoughts from the words of the language. Chapters 6–8 define and illustrate three important inferential procedures; **point estimation of parameters, confidence interval estimation,** and **hypothesis testing.** Many engineering problems can be effectively formulated and solved using these methods. In chapters 9–14 we will see how these procedures form the