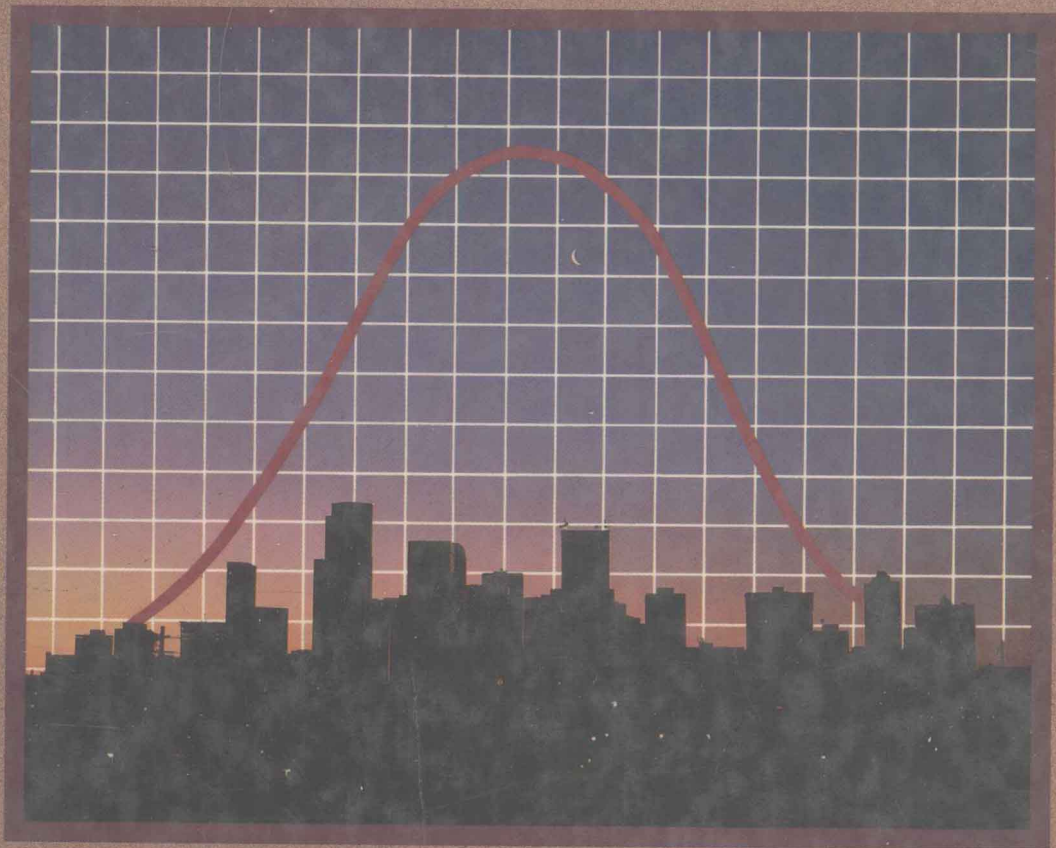


STATISTICS TODAY

A Comprehensive Introduction



Donald R. Byrkit

STATISTICS TODAY

A Comprehensive Introduction

Donald R. Byrkit

The University of West Florida

With contributions from Robert L. Schaefer & John H. Skillings,
Miami University of Ohio



The Benjamin/Cummings Publishing Company, Inc.
Menlo Park, California • Reading, Massachusetts
Don Mills, Ontario • Wokingham, U.K. • Amsterdam • Sydney
Singapore • Tokyo • Madrid • Bogota • Santiago • San Juan

To Steven and Linda

Sponsoring Editor: Craig Bartholomew
Developmental Editor: Martine Westermann
Production Editor: Richard Mason, Bookman Productions
Production Supervisor: Mary Picklum
Copy Editor: Don Yoder
Interior and Consulting Designer: Hal Lockwood
Illustrations: Carl Brown
Composition: Graphic Typesetting Services

The basic text of this book was designed using the Modular Design System, as developed by Wendy Earl and Design Office Bruce Kortebein.

Cover photo © Andy Caulfield/The Image Bank.

Copyright © 1987 by The Benjamin/Cummings Publishing Company, Inc.

Output examples from SAS procedures, Version 5 and later releases, Copyright © 1985, SAS Institute, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher. Printed in the United States of America. Published simultaneously in Canada.

The programs presented in this book have been included for their instructional value. They have been tested with care but are not guaranteed for any particular purpose. The publisher does not offer any warranties or representations, nor does it accept any liabilities with respect to the programs.

Library of Congress Cataloging-in-Publication Data

Byrkit, Donald R.

Statistics today.

Includes index.

1. Statistics. I. Title.

QA276.12.B96 1987 519.5 86-20710

ISBN 0-8053-0740-0

DEFGHIJ-DO-89

The Benjamin/Cummings Publishing Company, Inc.
2727 Sand Hill Road
Menlo Park, California 94025

STATISTICS TODAY

A Comprehensive Introduction

Preface

Statistics is a dynamic field of study today. The computer has augmented the ability of statisticians to analyze large data sets thoroughly. New statistical techniques such as robust methods and exploratory data analysis are making it easier to evaluate data sets that were once considered “messy.” The student of statistics is now entering a field that is on the cutting edge of our information-oriented economy. To keep up with the current developments in this exciting field, I have written *Statistics Today*. Rather than a patched-together book from another era, this is a new text that reflects the position of statistics. . . *today*.

What’s Different About This Book?

Statistics Today goes beyond previous books on elementary statistics. A few key elements of this book set it apart:

- Coverage of Minitab and the SAS software system—Minitab and SAS procedures are thoroughly covered in the book. Each chapter concludes with an optional section that presents the appropriate commands for both Minitab and the SAS software system. Later in the book SAS software systems’ usage is integrated throughout each chapter. This is a step beyond the convention of merely inserting computer output as an afterthought.
- Modern Methods—Many contemporary statistical methods are included in this book. If these methods cannot be covered in class due to time

constraints, it is hoped that the student will refer to them for further statistical study. Such topics include stem and leaf plots, box plots, normal probability plots, the pseudo-standard deviation, using trimmed and Winsorized data sets, and the Fisher-Behrens t' .

- Thinking “statistically”—An effort has been made in this book to encourage the student to approach problems like a real statistician. In addition, a number of historical comments and notes are presented in the margins in the interest of building an appreciation for statistics as a continually developing science.
- Motivation—A common misconception amongst uninitiates is that statistics is a boring field of study. This is not true! Every effort has been made in this book to make the text enjoyable reading, by showing the utility and real-life application of statistical methods. Each chapter is motivated by an intriguing chapter-opening case that should pique the reader’s interest. This case is solved at the end of the chapter by using the methods presented in that chapter. Real-data problems and examples are also sprinkled throughout each chapter.
- Nonparametric Methods—Most books “tack on” nonparametric tests as an apparent afterthought. In this book some nonparametric tests are introduced in Chapter 10. Appropriate nonparametric techniques are then presented side-by-side with their counterparts in Chapters 11, 12, and 13.
- Foundations for further study—Rather than presenting *only* a brief overview of statistics suitable for a general education survey course, this book also lays a complete foundation for further courses requiring statistics. Such courses could include experimental psychology, biometrics, econometrics, decision theory, marketing research, and many others. This text also provides alternative procedures to fit most situations encountered in real life.

Courses That Might Use This Book

Statistics Today provides a complete first course in general statistics at the algebra-based level. It is suitable for either a one-term or two-term course, and for students with a wide range of mathematical maturity. A first course in high-school algebra and some familiarity with the use of a calculator is recommended. Examples and problems have been chosen from a variety of disciplines including business, economics, accounting, management, biology, agriculture, psychology, sociology, anthropology, and education.

Flexibility and Organization

The book has been organized in a logical sequence to allow for great flexibility in a course. The first eight chapters contain the core of a beginning statistics course, although much additional material has been included. The accompanying dependency chart (pp. xxiv–xxv) shows which topics are considered essential to the logical development of the course. A few options are:

- For a two-course sequence, the first course might include the “core” of Chapters 1–8 (see dependency chart), along with additional topics from these chapters, or with some of Chapter 9. The second course could then cover most of the remainder of the material from Chapters 9–14.
- A one-term introductory course should include all of those topics listed as the “core” of the course, together with a selection of additional topics chosen by the instructor.
- For a course minimizing the coverage of probability, much of Chapter 3 can be omitted. The fundamental ideas of probability can be covered in Section 3.1 along with the idea of independence (in Section 3.2), which may be presented intuitively.
- An introduction to categorical data analysis, extension of two-sample methods to more than two populations, and an introduction to prediction via regression analysis might include Chapter 9, Sections 11.1 and 11.2, and Sections 13.1 and 13.2.
- An investigation of the methods for checking assumptions and using alternate approaches might include Sections 2.4, 6.4.3–6.4.5, 7.2.3–7.2.4, 8.2.3–8.2.5, and Chapter 10.

Pedagogical Aids

As mentioned previously, a concerted attempt has been made to motivate the student to study and enjoy the material in this book. The chapter opening Cases should stimulate the student’s interest in each chapter. In addition to these Cases the following study aids are woven into the text:

- **Proficiency Checks**—Most sections contain proficiency checks. These are short exercises designed to test to what extent students have mastered the concepts in that section, and to reinforce the retention of those concepts. Complete solutions of the proficiency checks are provided for easy reference at the conclusion of each chapter.

- Subsections—Each section has been divided into smaller “bite-size” subsections. Each subsection presents a single topic for students to master. It is hoped that these subsections will be helpful for students when reviewing the material for tests or for reference.
- Examples—Nearly every new idea is illustrated by at least one example showing its application. These examples are meant to amplify and clarify the theory behind an idea.
- Problems—The book contains a large number of both end-of-section and chapter-review problems. All problems are divided into two categories. Those problems labeled *Practice* are designed to provide exercise in using the mechanics covered in that section or chapter. Those problems marked *Applications* are realistic scenarios in which students can apply these methods. Many of these application problems are based on studies or situations that have appeared in newspapers, magazines, or journals.
- Symbols—Symbols that appear for the first time in each chapter are listed at the end of the chapter. All symbols that appear in the book are listed with their meaning on the book’s endpapers.
- Key Terms—Key terms are printed in boldface type where they first occur. An index to key terms has been provided at both the end of each chapter and at the back of the book.
- References—References for additional reading are provided where appropriate in each chapter. These are repeated in a list at the end of the chapter along with additional references.

Use of Computer Packages

Minitab and the SAS software system have been chosen for extensive use in this book. These systems were selected both because they are widely used and available, and because they are sufficiently different to be of interest to different users. Both packages are now available for mainframes and microcomputers.

To give the student something more useful than the standard “printout followed by interpretation” approach, emphasis has been placed on hands-on use of the computer. Each chapter includes basic explanations of how to use Minitab and the SAS software system, with examples and illustrations chosen from that chapter. Thus, very little explanation or supervision should be required by the instructor in order for the student to work problems using either a SAS software system or Minitab installation, once the local system has been mastered.

Minitab is introduced in Chapter 1, together with the data-entry, storage, and retrieval methods used throughout the book. A separate section in each chapter is devoted to using Minitab to perform some or all of the procedures introduced in that chapter. In some chapters the computer coverage is used to amplify or illustrate some of the concepts presented in that chapter. We suggest that these sections be read even if a computer is not used in the course. A complete listing of Minitab commands is presented in Appendix A.1.

Coverage of the SAS software system parallels that of Minitab in Chapters 1–10. Beginning in Chapter 11, SAS software system coverage is integrated within each section, since the computer is an important adjunct to the material presented in this chapter and succeeding chapters. A selected listing of SAS statements is presented in Appendix A.2.

Appendix B contains two data sets for use with the computer. Data set B.1 contains data used in a study of 100 female students at Miami University, Oxford, Ohio; data set B.2 reproduces a survey of 166 elderly people conducted by the U.S. Department of Agriculture. Additional data sets are included in the Instructor's Guide.

Testing of Assumptions

One of the most important aspects of statistical inference is understanding the meaning of the statistical procedures being used. To enhance student understanding, all procedures in this book are carefully examined, yet without using intimidating mathematical proofs or notation. A knowledge of the underlying assumptions for each procedure is essential if these procedures are not to be misused. Therefore, those assumptions are frequently emphasized. Unlike most other books, methods are given to test the assumptions. For example, several methods are given to test the ubiquitous assumption of normality, including normal probability plots, a goodness-of-fit test, and some modern, exploratory data-analysis procedures. Hartley's test is applied to the assumption of homogeneity of variance in ANOVA.

Supplements

A number of supplements to this book are provided to assist the instructor and the student in presenting and learning the material:

- *Instructor's Guide*—The Instructor's Guide contains additional optional material not included in the book. This includes derivations, the use of

Chebyshev's theorem on statistical inference, additional data sets, and additional SAS statements and procedures. Transparency masters and the answers to even-numbered problems are also included here. All data sets are available on disk or tape.

- *Student Solutions Manual*—Step-by-step worked-out solutions to the odd-numbered problems (answers to most of which are in the back of the text) are available in a separate Solutions Manual. This supplement also includes comprehensive chapter summaries.
- *Testbank*—A Testbank for instructors includes a large number of problems for use in tests.
- *STATDISK*—STATDISK is a set of statistical programs available for the IBM-PC or the Apple II. It is available from the publisher for adopters of this book. The accompanying *Statdisk Manual*, a tutorial self-study student manual, is available at a reasonable cost for student purchase.

Acknowledgments

Thanks are due to many people who helped in the preparation of this book, especially to Craig Bartholomew, who believed in the project, and Martine Westermann, who made many, many valuable suggestions for improvements. Particular thanks are due to John Skillings and Robert Schaefer of Miami University of Ohio for supplying a draft of the Minitab material. Special thanks also go to my wife, Marnette, who encouraged me through many 80- and 90-hour work weeks over the past year and a half, and gave up our weekends to work on the book. Thanks also to the many reviewers who made valid criticisms and valuable suggestions: Barnard Bissinger, Penn State University; John W. Dirksey, California State University, Bakersfield; Shirley Dowdy, West Virginia University; Richard G. Driskell, Ball State University; Judith M. Ekstrand, San Francisco State University; Dale Everson, University of Idaho; Iris Ibrahim, Clemson University; William Koellner, Montclair State College; Purushottam Laud, Northern Illinois University; Bruce Lind, University of Puget Sound; Michael Shing, Stanford University; John Skillings, Miami University; Leonard Sweet, University of Akron; John Wasik, North Carolina State University; and Ann Watkins, Los Angeles Pierce College.

I am indebted to the SAS Institute and Minitab, Inc. for permission to document their packages, and to Barbara F. Ryan of Minitab who checked all Minitab commands and printouts. Output from SAS procedures are printed with permission of SAS Institute Inc., Cary, NC. 27511-8000, Copyright © 1985. I am also indebted to the Biometrika Trustees for permission to use Tables 8, 12, 18, 29, and 31 from *Biometrika Tables for Statisticians*,

Volume I, Third Edition, by Pearson and Hartley; to American Cyanamid for permission to use Tables 1 and 2 from *Some Rapid Approximate Statistical Procedures* by Wilcoxon and Wilcox; to the authors and publishers of *Statistical Tables* by R. R. Sokal and F. J. Rohlf (W. H. Freeman and Company) for permission to reprint part of Table O; and to the staff of the Institute for Statistical and Mathematical Modeling of the University of West Florida for compiling Tables 2, 3, and 4.

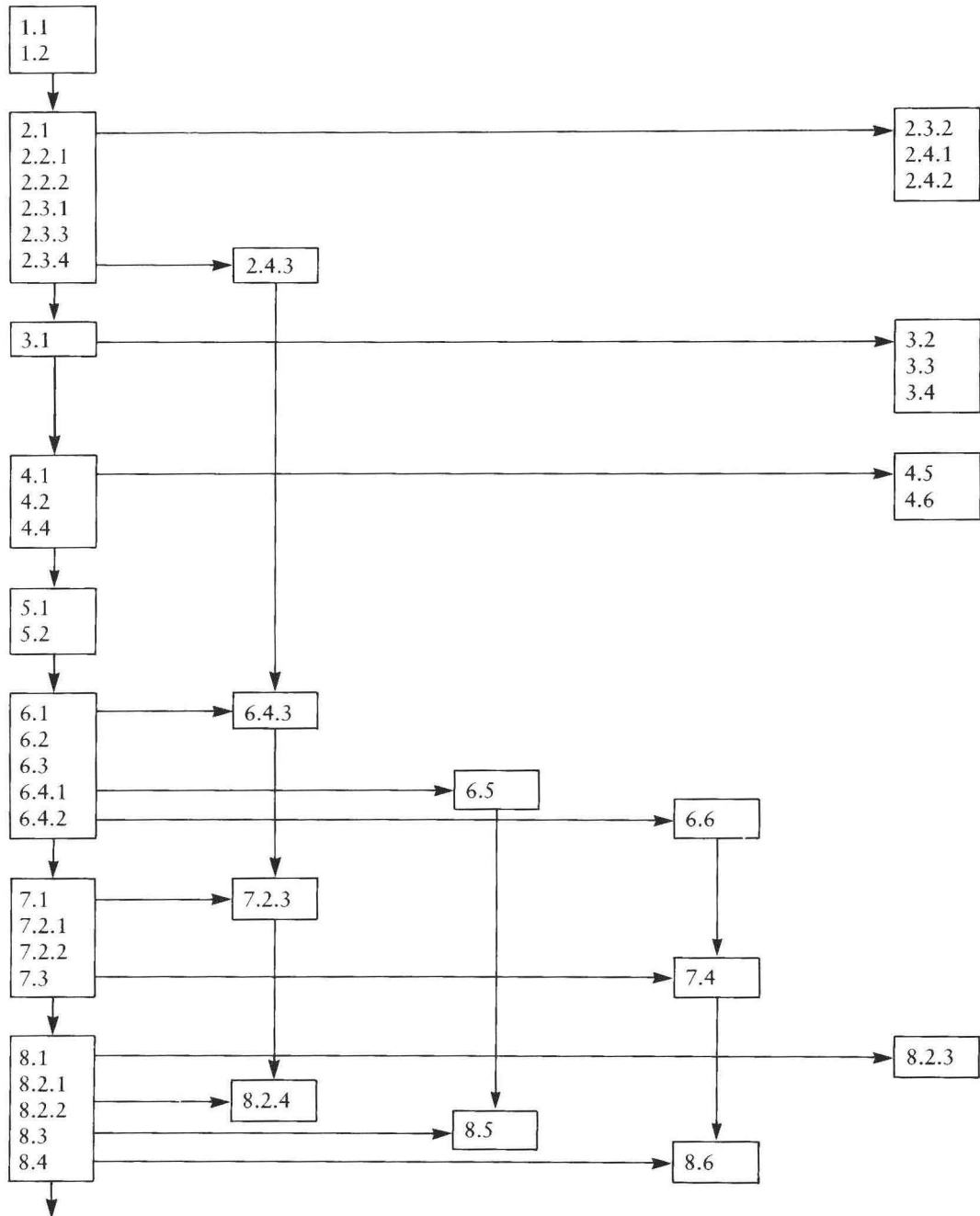
Donald R. Byrkit
Pensacola, Florida

Dependency Chart

Core of
the Course

Non-core, but
needed for later use

Additional
topics

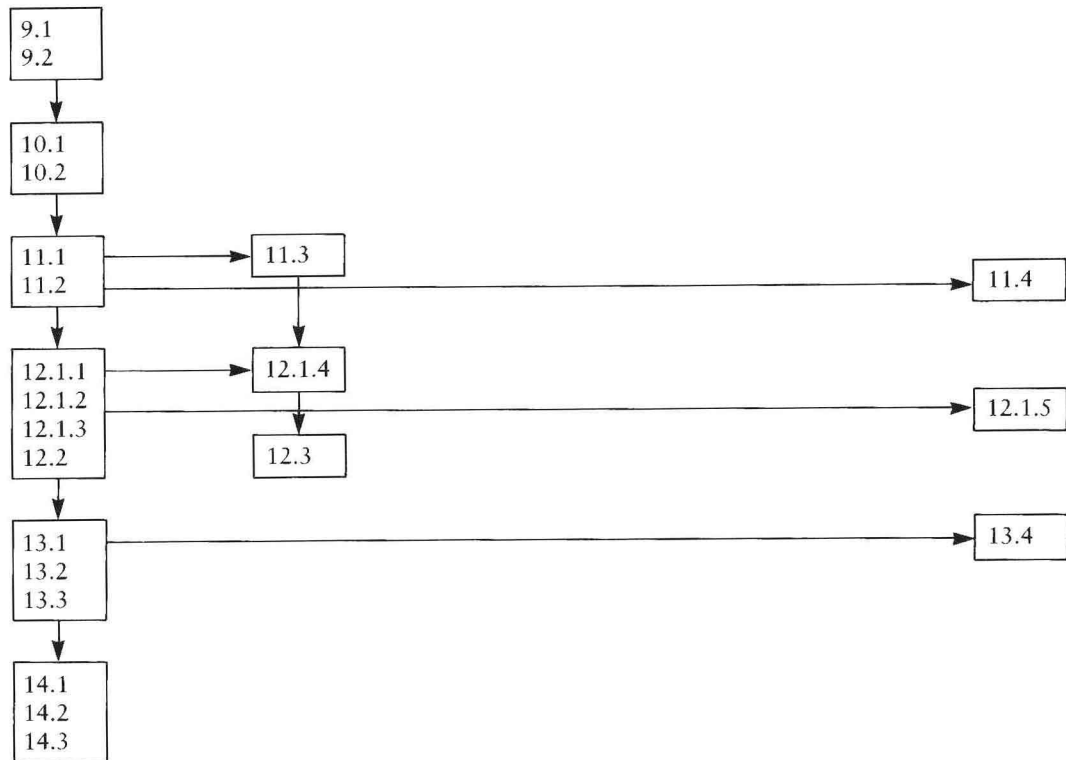


Dependency Chart

Core of
the Course

Non-core, but
needed for later use

Additional
topics



Brief Contents

Chapter 1	Organization and Presentation of Data	1
Chapter 2	Summary Descriptive Measures	54
Chapter 3	Probability	110
Chapter 4	Discrete Probability Distributions	172
Chapter 5	The Normal Distribution	238
Chapter 6	Estimating Population Parameters	279
Chapter 7	Hypothesis Testing	356
Chapter 8	Statistical Inferences from Two Samples	428
Chapter 9	Chi-Square Analysis	498
Chapter 10	Nonparametric Methods	541
Chapter 11	Single-Factor Analysis of Variance	588
Chapter 12	Analysis of Variance—Two Factors	646
Chapter 13	Linear Regression Analysis	711
Chapter 14	Multiple Regression	791
Appendix A.1	List of Minitab Commands	A-1
Appendix A.2	List of SAS Statements	A-7
Appendix B.1	Miami Female Study Data	B-1
Appendix B.2	Department of Agriculture Survey	B-5
Appendix C	Answers to Odd-numbered Problems	C-1
Appendix D	Statistical Tables	D-1
Index		I-1

Detailed Contents

1	Organization and Presentation of Data	1
1.0	Case 1: What Grapes to Grow?	2
1.1	Frequency Distributions	2
1.1.1	What Is Statistics?	2
1.1.2	The Basics	4
1.1.3	Frequency Tables	5
1.1.4	Grouping Data	8
1.1.5	Cumulative Frequency Tables	13
1.2	Graphical Methods	16
1.2.1	Bar Graphs	17
1.2.2	Histograms	18
1.2.3	Frequency Polygons	21
1.2.4	Stem and Leaf Plots	22
1.2.5	Other Graphical Methods	23
1.2.6	A Word About Distortion	24
1.3	Computer Usage	27
1.3.1	Computer Software	28
1.3.2	Introduction to Minitab	29
1.3.3	Minitab Commands for This Chapter	32
1.3.4	Introduction to the SAS System	35
1.3.5	SAS Statements for This Chapter	38
1.4	Case 1: What Grapes to Grow?—Solution	44
1.5	Summary	45
	Index of Terms	49

Answers to Proficiency Checks	49
References	53

2	Summary Descriptive Measures	54
2.0	Case 2: The Problem Promotion	55
2.1	Measures of Central Tendency	56
2.1.1	The Mean	56
2.1.2	The Median	60
2.1.3	The Mode	62
2.2	Measures of Variability	66
2.2.1	The Variance and Standard Deviation	67
2.2.2	Chebyshev's Theorem and the Empirical Rule	71
2.2.3	Pearson's Index of Skewness	73
2.3	Other Descriptive Measures	78
2.3.1	The z -Score	78
2.3.2	The Coefficient of Variation	80
2.3.3	Quartiles and the Interquartile Range	81
2.3.4	The Box Plot	83
2.4	Outliers and the Trimmed Mean	86
2.4.1	Identifying Outliers	86
2.4.2	The Pseudo-Standard Deviation	88
2.4.3	Trimmed and Winsorized Data Sets	89
2.5	Computer Usage	92
2.5.1	Minitab Commands for This Chapter	92
2.5.2	SAS Statements for This Chapter	96
2.6	Case 2: The Problem Promotion—Solution	101
2.7	Summary	102
	Index of Terms	105
	Glossary of Symbols	106
	Answers to Proficiency Checks	106
	Some Notes on Summation Notation	108
	References	109

3	Probability	110
3.0	Case 3: Where Did the Sponge Come From?	111
3.1	Probability	112
3.1.1	Probabilities for Equally Likely Outcomes	112
3.1.2	Sample Spaces	114

3.1.3	Probability Rules	118
3.1.4	Compound Events	120
3.2	Conditional Probability and Independent Events	126
3.2.1	Conditional Probability	127
3.2.2	Independent Events	134
3.2.3	Bayes' Rule	136
3.3	Random Sampling	143
3.3.1	Random Samples	144
3.3.2	Using a Random Number Table	145
3.4	Counting Rules	147
3.4.1	Permutations	148
3.4.2	Partitions	150
3.4.3	Combinations	151
3.5	Computer Usage	155
3.5.1	Minitab Commands for This Chapter	156
3.5.2	SAS Statements for This Chapter	158
3.6	Case 3: Where Did the Sponge Come From?—Solution	162
3.7	Summary	163
	Index of Terms	169
	Glossary of Symbols	169
	Answers to Proficiency Checks	170
	References	171
4	Discrete Probability Distributions	172
4.0	Case 4: The Parapsychological Puzzle	173
4.1	Random Variables and Probability Distributions	174
4.1.1	Random Variables	174
4.1.2	Discrete Probability Distributions	175
4.2	Mean and Variance of a Random Variable	179
4.2.1	Mean of a Random Variable	179
4.2.2	Variance of a Random Variable	182
4.3	Case Study: Repair, Replace, or Test?	187
4.4	Binomial Probability	188
4.4.1	The Binomial Random Variable	189
4.4.2	Using Binomial Tables	193
4.4.3	Mean and Variance of a Binomial Random Variable	197
4.5	The Poisson Random Variable	202
4.5.1	The Poisson Distribution	202
4.5.2	Approximating the Binomial Distribution	205