

The background is a complex collage of geometric shapes and textures. It includes a large blue circle with a white speckled pattern on the left, a pink rectangle, a pink triangle, and a pink square with a white speckled pattern on the right. There are also several thin, dark lines and a gold-colored spiral in the bottom left corner. The overall color palette is dominated by blue, pink, and gold.

*fifth edition*

Basic  
Business Statistics  
*Concepts and Applications*

---

MARK L. BERENSON  
DAVID M. LEVINE



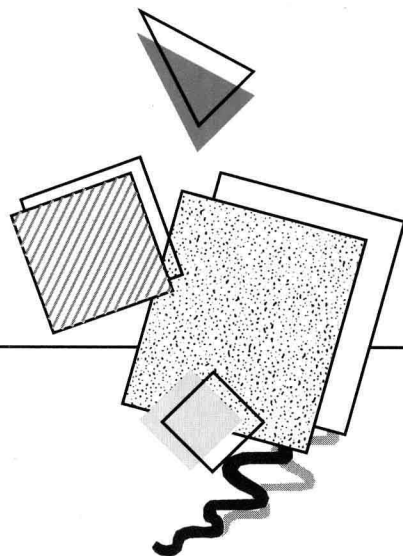
*fifth edition*

# *Basic Business Statistics*

*Concepts and Applications*  
.....

MARK L. BERENSON  
DAVID M. LEVINE

*Department of Statistics and Computer Information Systems  
Baruch College, City University of New York*



PRENTICE HALL Englewood Cliffs, New Jersey

**Library of Congress Cataloging-in-Publication Data**

Berenson, Mark L.  
Basic business statistics : concepts and applications / Mark L.  
Berenson, David M. Levine.—5th ed.  
p. cm.  
Includes bibliographical references and index.  
ISBN 0-13-099540-1  
1. Commercial statistics. 2. Statistics. I. Levine, David M.,  
II. Title.  
HF1017.B38 1992 91-37674  
519.5—dc20 CIP

Editor in Chief: JOSEPH HEIDER  
Acquisition Editor: VALERIE ASHTON  
Managing Editor: FRAN RUSSELLO  
Production Editor: RACHEL J. WITTY, Letter Perfect, Inc.  
Design Supervisor: CHRISTINE GEHRING-WOLF  
Designer: MAUREEN EIDE  
Cover Designer: MAUREEN EIDE  
Cover Art: CLINT EAGER, ORION; uncropped, 36 x 48 inches.  
Front Line Graphics, Inc., © 1988.  
Copy Editor: ADELE GORELICK  
Photo Editor: LORI MORRIS NANTZ  
Prepress Buyer: TRUDY PISCIOTTI  
Manufacturing Buyer: ROBERT ANDERSON  
Marketing Manager: SANDRA STEINER  
Editorial Assistants: ANNMARIE DUNN, BRIAN HATCH  
Design Assistant: THOMAS NERY



© 1992 by Prentice Hall, Inc.  
A Division of Simon & Schuster  
Englewood Cliffs, New Jersey 07632

*All rights reserved. No part of this book may be  
reproduced, in any form or by any means,  
without permission in writing from the publisher.*

Printed in the United States of America

10 9 8 7 6 5 4 3 2

**ISBN 0-13-099540-1**

Prentice-Hall International (UK) Limited, *London*  
Prentice-Hall of Australia Pty. Limited, *Sydney*  
Prentice-Hall Canada Inc., *Toronto*  
Prentice-Hall Hispanoamericana, S.A., *Mexico*  
Prentice-Hall of India Private Limited, *New Delhi*  
Prentice-Hall of Japan, Inc. *Tokyo*  
Simon & Schuster Asia Pte. Ltd., *Singapore*  
Editora Prentice-Hall do Brasil, Ltda, *Rio de Janeiro*

*To our wives,  
Rhoda Berenson and Marilyn Levine  
and to our children,  
Kathy Balto, Lori Berenson, and Sharyn Levine*

# Preface

.....

When planning or revising a textbook, the authors must decide how the text will differ from those already available and what contribution it will make to the field of study. These issues are resolved in several ways in this fifth edition of *Basic Business Statistics*.

## Survey Database

As in previous editions, a set of data based upon the results of a survey is included. Specifically, the data set constitutes the characteristics of 362 sampled single-family houses in two suburban New York counties. This survey, used throughout the text, serves as a means of integrating the various topics, permitting a cohesive study of descriptive statistics, probability, statistical inference, and regression analysis. The use of an actual survey, examined from beginning to end, will aid the student in conducting basic research in future courses, theses, or occupational situations.

## Problems, Special Data Set Problems, Database Exercises, and Case Studies

“Learning” results from “doing.” This text provides the student with the opportunity to select and solve from among the 1,520 problems presented at the end of sections as well as at the end of chapters. Most of these problems apply to realistic situations (using real data whenever possible) in various fields including accounting, economics, finance, health care administration, information systems, management, marketing, and public administration. The answers to selected problems (indicated by the symbol ■) appear at the end of the text. In addition, a series of special data set problems, pertaining to six large data sets displayed in Chapter 3, are presented throughout the text. Moreover, 282 database exercises pertaining to the real estate survey are presented at the end of most chapters. Furthermore, detailed case studies are included at the end of each of ten chapters.

## ACTION (ACTION) and Lightning (⚡) Problems

Statistics is a living, breathing subject. It is not mere numbers crunching! Emphasis here is placed on understanding and interpretation, since it is essential that students be able to express what they have learned. To these ends, action (ACTION) problems enhance literacy by asking the student to write letters, memos, and reports, and prepare talks; lightning (⚡) problems are particularly thought provoking or have no “exact” answer. Together, action and lightning problems allow the students to think and better enable them to understand the utility of statistical analysis as an aid to the solution of real problems in an organizational setting.

## Computer Packages

A major feature of this text is the discussion of the use of such computer packages as GBSTAT, Minitab, MYSTAT, SAS, SPSS<sup>x</sup>, STATGRAPHICS, and STATISTIX. Not only is output from these packages illustrated throughout the text, but the use of the computer as a tool for assisting in the decision-making process is interwoven in the various chapters. Extensive coverage is given to plotting data, interpreting computer output, and evaluating the assumptions of the particular statistical techniques, thereby simulating the approach used by a statistician in conducting an actual statistical analysis. Moreover, using the results from the real estate survey, an entire chapter (Chapter 5) is included to highlight the use of computer packages in a descriptive statistical analysis.

## The Deming Philosophy For Quality and Productivity

It is widely recognized that the field of business has entered a new economic age, one in which competition from all over the world must be faced. This text provides detailed coverage (particularly in Chapter 19) of the use of statistical methods for total quality management with a discussion of the fourteen points of the “*management-by-process*” philosophy of W. Edwards Deming. The topics presented include operational definitions, statistical thinking, enumerative versus analytical studies. Pareto diagrams, digidot plots, process control charts, fishbone diagrams, process flow diagrams, and the parable of the red beads, an experiment intended to demonstrate the concepts of common versus special cause variation.

## Modern Methods

A sixth important feature of this edition is the inclusion of methodology that, over the past few years, has gained widespread usage. As examples, exploratory data analysis (EDA) techniques, dot charts, and supertables are discussed (Chapters 3 and 4); normal probability plots are described (Chapter 8); the subject of meta-analysis is introduced (Chapter 11); a  $p$ -value approach to hypothesis testing is used (Chapters 11–13); residual and influence analysis and model-building in regression are covered (Chapters 16 and 17); and various business forecasting methods are considered (Chapter 18).

It is our hope and anticipation that the pedagogical features along with the unique approaches taken in this textbook will make the study of basic statistics more meaningful, rewarding, and comprehensible for all readers.

We are extremely grateful to the many organizations and companies that generously allowed us to use their actual data for developing problems and examples throughout our text. In particular, we would like to cite Time Inc. (publisher of *Fortune* and *Money*), CBS Inc. (publisher of *Road & Track*), The Condé Nast Publications, Inc. (publisher of *Street and Smith's*), Consumers Union (publisher of *Consumer Reports*), McGraw Hill, Inc. (publisher of *Business Week*), Standard & Poor's Corporation (publisher of *Standard N.Y.S.E. Stock Reports*), Moody's Investors Service (publisher of *Moody's Handbook of Common Stocks*), American Hospital Publishing, Inc. (publisher of *Hospitals*), Los Angeles Times Syndicate International (publisher of *New York Newsday*), CMP Publications Inc. (publishers of *Health Week*), the American Automobile Association, Zagat Survey, Brian Joiner (Joiner Associates, Inc.), M.I.T. Center for Advanced Engineering Study, CEEPress Books, The Quantum Company, Goal/QPC, Preview Publishing, The American Association of University Professors (publisher of *Academe*), Gale Research, Inc., and the Association of Research Libraries.

Moreover, we would like to thank the *Biometrika* Trustees, American Cyanamid Company, The Rand Corporation, The Chemical Rubber Company, the Institute of Mathematical Statistics, and the American Society for Testing and Materials for their kind permission to publish various tables in Appendix E and the American Statistical Association for its permission to publish diagrams from *The American Statistician*. Furthermore, we would like to acknowledge GBSTAT Inc., Minitab Inc., the SAS Institute, SPSS Inc., STATISTIX Inc, STSC Inc. (i.e., STATGRAPHICS), and SYSTAT Inc. (i.e., MYSTAT) for their permission to present computer output throughout the text.

In addition, we wish to thank some of our colleagues at Baruch College: Professors Ann Brandwein, Pasquale DiPillo, Shulamith Gross, Theodore Joyce, and Manus Rabinowitz, as well as Professors Mai Chou, University of Kentucky, Robert Curry, DeVry, Kansas City, Mark Eakin, University of Texas at Arlington, Daniel Gordon, Salem State College, Jacqueline Hoell, Virginia Polytechnic Institute and State University, John Neufeld, University of North Carolina at Greensboro, John McKenzie, Babson College, Alan Olinsky, Bryant College, Elena Pernas, Florida International University, Patricia Ramsey, Fordham University, Donald R. Robinson, Illinois State University, Barbara Russell, St. Bonaventure University, Lillian Russell, University of Delaware, Ernest Scheuer, California State at Northridge, Richard Smith, Bryant College, and David Sylvester, University of Tennessee for their constructive comments during the revision of this textbook.

We would like to close by expressing our thanks to Valerie Ashton, Ann Marie Dunn, Brian Hatch, Dennis Hogan, P. J. McCue, Kate Moore, Sandy Steiner, and Garrett White of the editorial staff at Prentice-Hall and production editor, Rachel J. Witty of Letter Perfect, Inc., for their continued encouragement. We also wish to thank Marie Anne Downey, Uri Peyser, Susan Stanton, and Beth Zuckerman for their assistance in data collection and file development. Finally, we would like to thank our wives and children for their patience, understanding, love, and assistance in making this book a reality. It is to them that we dedicate this book.

MARK L. BERENSON  
DAVID M. LEVINE

# Contents

.....

## PREFACE xiv

## 1 INTRODUCTION 1

- 1.1 What Is Modern Statistics? 2
- 1.2 The Growth and Development of Modern Statistics 2
- 1.3 Enumerative Versus Analytical Studies 3
- 1.4 The Role of Computer Packages in Statistics 5
- 1.5 Statistical Thinking and Modern Management 6

## 2 DATA COLLECTION 9

- 2.1 Introduction: The Need for Research 10
- 2.2 Sources of Data for Research 10
- 2.3 Obtaining Data through Survey Research 11
- 2.4 Designing the Questionnaire Instrument 16
- 2.5 Choosing the Sample Size for the Survey 19
- 2.6 Types of Samples 20
- 2.7 Drawing the Simple Random Sample 21
- 2.8 Obtaining the Responses 24
- 2.9 Data Preparation: Editing, Coding, and Transcribing 26
- 2.10 Data Collection: A Review and a Preview 35

*Case Study A* 37

## 3 DESCRIBING AND SUMMARIZING DATA 39

- 3.1 Introduction: What's Ahead 40
- 3.2 Exploring the Data 41
- 3.3 Properties of Quantitative Data 41
- 3.4 Measures of Central Tendency 42
- 3.5 Measures of Dispersion 50
- 3.6 Shape 57
- 3.7 Dealing with Larger Batches of Quantitative Data:  
"Exploratory Data Analysis" Techniques 58
- 3.8 The Stem-and-Leaf Display 60
- 3.9 Quantiles 63



- 3.10** Descriptive Summary Measures Using Quartiles 65
- 3.11** Using the Five-Number Summary for the Box-and-Whisker Plot: A Graphical Technique 66
- 3.12** Calculating Descriptive Summary Measures from a Population 69

## 4

### DATA PRESENTATION 99

- 4.1** Introduction 100
- 4.2** Tabulating Quantitative Data: The Frequency Distribution 100
- 4.3** Obtaining Descriptive Summary Measures from a Frequency Distribution 106
- 4.4** Tabulating Quantitative Data: The Relative Frequency Distribution and Percentage Distribution 113
- 4.5** Graphing Quantitative Data: The Histogram and Polygon 117
- 4.6** Cumulative Distributions and Cumulative Polygons 125
- 4.7** Graphing Quantitative Data in Sequence: The Digidot Plot 132
- 4.8** Tabulating Qualitative Data: The Summary Table 135
- 4.9** Graphing Qualitative Data: Bar, Pie, and Dot Charts 136
- 4.10** Graphing Qualitative Data: The Pareto Diagram 141
- 4.11** Tabularizing Qualitative Data: Cross-Classification Tables and Supertables 144
- 4.12** Data Presentation: An Overview 147

## 5

### USING THE COMPUTER FOR DESCRIPTIVE STATISTICAL ANALYSIS 165

- 5.1** Introduction and Overview 166
- 5.2** Using Statistical Packages for Quantitative Data 166
- 5.3** Using Statistical Packages for Qualitative Data 174
- 5.4** Using Statistical Packages for Cross-Tabulations 175
- 5.5** The Computer and the Researcher: Interacting Roles 176
- Case Study B 181*

## 6

### BASIC PROBABILITY 187

- 6.1** Introduction 188
- 6.2** Objective and Subjective Probability 188
- 6.3** Basic Probability Concepts 189
- 6.4** Simple (Marginal) Probability 194
- 6.5** Joint Probability 195
- 6.6** Addition Rule 197
- 6.7** Conditional Probability 200
- 6.8** Multiplication Rule 204
- 6.9** Bayes' Theorem 206
- 6.10** Counting Rules 210
- 6.11** Summary 212

## 7

## SOME IMPORTANT DISCRETE PROBABILITY DISTRIBUTIONS 220

- 7.1 Introduction: What's Ahead 221
- 7.2 The Probability Distribution for a Discrete Random Variable 221
- 7.3 Mathematical Expectation and Expected Monetary Value 222
- 7.4 Discrete Probability Distribution Functions 229
- 7.5 Uniform Distribution 230
- 7.6 Binomial Distribution 231
- 7.7 Hypergeometric Distribution 239
- 7.8 Poisson Distribution 245

## 8

## THE NORMAL DISTRIBUTION 260

- 8.1 Mathematical Models of Continuous Random Variables: The Probability Density Function 261
- 8.2 The Normal Distribution 262
- 8.3 Applications 268
- 8.4 Assessing the Normality Assumption: Evaluating Properties and Constructing Probability Plots 279
- 8.5 The Normal Distribution as an Approximation to Various Discrete Probability Distributions 288
- 8.6 Summary and Overview 297
- Case Study C* 303

## 9

## SAMPLING DISTRIBUTIONS 305

- 9.1 The Need for Sampling Distributions 306
- 9.2 Sampling Distribution of the Mean 306
- 9.3 Sampling Distribution of the Proportion 319
- 9.4 Sampling from Finite Populations 322
- 9.5 Summary and Overview 324

## 10

## ESTIMATION 327

- 10.1 Point and Confidence Interval Estimates 328
- 10.2 Confidence Interval Estimation of the Mean ( $\sigma_x$  Known) 328
- 10.3 Confidence Interval Estimation of the Mean ( $\sigma_x$  Unknown) 333
- 10.4 Prediction Interval for a Future Individual Value 339
- 10.5 Confidence Interval Estimation for the Proportion 341
- 10.6 Sample Size Determination for the Mean 343
- 10.7 Sample Size Determination for a Proportion 346
- 10.8 Estimation and Sample Size Determination for Finite Populations 348

# 11

## HYPOTHESIS TESTING I: INTRODUCTION AND CONCEPTS 359

- 11.1 Introduction 360
- 11.2 The Hypothesis-Testing Procedure 360
- 11.3 Type I and Type II Errors 362
- 11.4 Test of Hypothesis for the Mean ( $\sigma_x$  Known) 364
- 11.5 Summarizing the Steps of Hypothesis Testing 366
- 11.6 The  $p$ -Value Approach to Hypothesis Testing: Two-Tailed Tests 367
- 11.7 Test of Hypothesis for the Mean ( $\sigma_x$  Unknown) 369
- 11.8 One-Tailed Tests 372
- 11.9 The  $p$ -Value Approach to Hypothesis Testing: One-Tailed Tests 374
- 11.10 Test of Hypothesis for a Proportion (One Sample) 376
- 11.11 A Connection between Confidence Intervals and Hypothesis Testing 380
- 11.12 Hypothesis Testing and the Real Estate Survey 381
- 11.13 Power of a Test 384
- 11.14 Meta-Analysis: A Controversial Outgrowth of Hypothesis Testing 398

# 12

## HYPOTHESIS TESTING II: DIFFERENCES BETWEEN QUANTITATIVE VARIABLES 411

- 12.1 Introduction 412
- 12.2 Testing for the Difference between the Means of Two Independent Populations Having Equal Variances 412
- 12.3 Testing for the Difference between the Means from Two Related Populations 418
- 12.4 Testing for the Equality of Variances from Two Independent Populations 426
- 12.5 Testing for the Difference between the Means of Two Independent Populations Having Unequal Variances 431
- 12.6 Hypothesis Testing, Computers, and the Real Estate Survey 435
- Case Study D* 447

# 13

## HYPOTHESIS TESTING III: DIFFERENCES BETWEEN PROPORTIONS AND OTHER CHI-SQUARE TESTS 449

- 13.1 Introduction 450
- 13.2 Testing for the Difference between Proportions from Two Independent Populations Using the Normal Approximation 451
- 13.3 Testing for the Difference between Proportions from Two Independent Populations Using the Chi-Square Test 455

- 13.4** Testing the Difference between Proportions from  $C$  Independent Populations 462
- 13.5** Chi-Square Test of Independence in the  $R \times C$  Table 466
- 13.6** The Chi-Square Goodness-of-Fit Test for Probability Distributions 470
- 13.7** Testing for the Difference between Proportions from Two Related Populations: The McNemar Test 475
- 13.8** Testing a Hypothesis About a Population Variance (or Standard Deviation) 479
- 13.9** Chi-Square Tests, Computers, and the Real Estate Survey 484

## 14

### THE ANALYSIS OF VARIANCE 492

- 14.1** Introduction: What's Ahead 493
- 14.2** The One-way Analysis of Variance 493
- 14.3** Assumptions of the Analysis of Variance 505
- 14.4** Hartley's Test for Homogeneity of Variance 506
- 14.5** Multiple Comparisons: The Tukey  $T$  Method 508
- 14.6** The Randomized Block Design 511
- 14.7** Two-Way Analysis of Variance 521
- 14.8** Analysis of Variance, Computers, and the Real Estate Survey 534
- Case Study E 548*

## 15

### NONPARAMETRIC METHODS 550

- 15.1** Introduction 551
- 15.2** Classical Versus Nonparametric Procedures 551
- 15.3** Advantages and Disadvantages of Using Nonparametric Methods 553
- 15.4** Wald-Wolfowitz One-Sample Runs Test for Randomness 554
- 15.5** Wilcoxon Signed-Ranks Test 560
- 15.6** Wilcoxon Rank-Sum Test 569
- 15.7** Kruskal-Wallis Test for  $c$  Independent Samples 574
- 15.8** Friedman Rank Test for  $c$  Related Samples 579
- 15.9** Spearman's Rank-Correlation Procedure 584
- 15.10** Nonparametric Procedures, Computers, and the Real Estate Survey 590
- 15.11** Summary 591
- Case Study F 603*

## 16

### THE SIMPLE LINEAR REGRESSION MODEL AND CORRELATION 604

- 16.1** Introduction 605
- 16.2** The Scatter Diagram 605
- 16.3** Types of Regression Models 609
- 16.4** Determining the Simple Linear Regression Equation 611
- 16.5** Standard Error of the Estimate 616



- 16.6** Measures of Variation in Regression and Correlation 618
- 16.7** Correlation—Measuring the Strength of the Association 622
- 16.8** Confidence-Interval Estimate for Predicting  $\mu_{YX}$  626
- 16.9** Prediction Interval for an Individual Response  $\hat{Y}_I$  629
- 16.10** Inferences about the Population Parameters in Regression and Correlation 630
- 16.11** Assumptions of Regression and Correlation 634
- 16.12** Regression Diagnostics: Residual Analysis 635
- 16.13** Regression Diagnostics: Influence Analysis 640
- 16.14** Regression, Computers, and the Real Estate Survey 643
- 16.15** Regression and Correlation—Some Caveats 650
- Case Study G* 666

## 17

### MULTIPLE REGRESSION MODELS 668

- 17.1** Introduction 669
- 17.2** Finding the Regression Coefficients 671
- 17.3** Prediction of the Dependent Variable  $Y$  for Given Values of the Explanatory Variables 676
- 17.4** Testing for the Significance of the Relationship Between the Dependent Variable and the Explanatory Variables 677
- 17.5** Measuring Association in the Multiple Regression Model 679
- 17.6** Testing Portions of a Multiple Regression Model 681
- 17.7** Inferences Concerning the Population Regression Coefficients 685
- 17.8** Confidence-Interval Estimates for Predicting  $\mu_{YX}$  and  $\hat{Y}_I$  688
- 17.9** Coefficient of Partial Determination 689
- 17.10** The Curvilinear Regression Model 690
- 17.11** Testing Portions of the Curvilinear Model 698
- 17.12** Dummy-Variable Models 700
- 17.13** Other Types of Regression Models 704
- 17.14** Multicollinearity 707
- 17.15** Residual Analysis in Multiple Regression 708
- 17.16** Influence Analysis in Multiple Regression 711
- 17.17** An Example of Model Building 714
- 17.18** Computer Packages and Multiple Regression 720
- Case Study H* 736

## 18

### INDEX NUMBERS, TIME SERIES, AND BUSINESS FORECASTING 739

- 18.1** Introduction 740
- 18.2** Index Numbers 740
- 18.3** The Price Index 740
- 18.4** The Consumer Price Index 745
- 18.5** Index Numbers: An Overview 750
- 18.6** The Importance of Business Forecasting 750
- 18.7** Component Factors of the Classical Multiplicative Time-Series Model 751

- 18.8** Smoothing the Annual Time-Series: Moving Averages and Exponential Smoothing 756
- 18.9** Time-Series Analysis of Annual Data: Least-Squares Trend Fitting and Forecasting 766
- 18.10** The Holt-Winters Method for Trend Fitting and Forecasting 777
- 18.11** Autoregressive Modeling for Trend Fitting and Forecasting 781
- 18.12** Choosing an Appropriate Forecasting Model 789
- 18.13** Time-Series Decomposition: Studying the Cyclical Effects 795
- 18.14** Time-Series Analysis of Monthly Data 797
- 18.15** Time-Series Analysis: An Overview 808
- Case Study I* 817

## 19

### STATISTICAL APPLICATIONS IN QUALITY AND PRODUCTIVITY MANAGEMENT 820

- 19.1** Introduction 821
- 19.2** Quality and Productivity—A Historical Perspective 821
- 19.3** The Theory of Control Charts 823
- 19.4** Some Tools for Studying a Process: Fishbone (Ishikawa) and Process Flow Diagrams 825
- 19.5** Deming's Fourteen Points—A Theory of Management by Process 831
- 19.6** Control Charts for a Sequence of Individual  $X$  Values 834
- 19.7** Control Charts for the Mean ( $\bar{X}$ ) and the Range ( $R$ ) 841
- 19.8** Control Charts for the Proportion and Number of Nonconforming Items—The  $p$  and  $np$  Charts 848
- 19.9** The  $c$  Chart—A Control Chart for the Number of Occurrences per Unit 857
- 19.10** Inspection of Incoming Material and Final Product 861
- 19.11** The Red Bead Experiment: Understanding Process Variability 863
- 19.12** Summary and Overview 865
- Case Study J* 868

### ANSWERS TO SELECTED PROBLEMS (■) 871

### APPENDICES

- A.** Review of Arithmetic and Algebra 897
- B.** Summation Notation 899
- C.** Statistical Symbols and Greek Alphabet 904
- D.** The Metric System 905
- E.** Tables 907

### INDEX 945

# CHAPTER

# 1

## *Introduction*

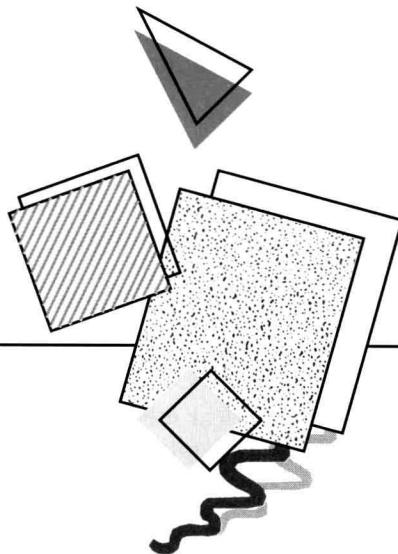


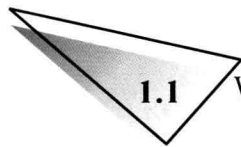
### CHAPTER OBJECTIVE

To present a broad overview of the subject of statistics and its applications

### CONCEPTUAL EMPHASIS

1. Population versus sample
2. Parameter versus statistic
3. Descriptive versus inferential statistics
4. Enumerative versus analytical studies
5. Statistical thinking

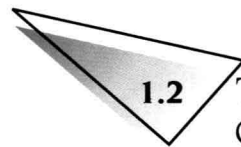




## 1.1 WHAT IS MODERN STATISTICS?

A century ago H. G. Wells commented that “statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.” Each day of our lives we are exposed to a wide assortment of numerical information pertaining to such phenomena as stock market activity, unemployment rates, medical research findings, opinion poll results, weather forecasts, and sports data. Frequently, such information has a profound impact on our lives.

The subject of **modern statistics** encompasses the collection, presentation, and characterization of information to assist in both data analysis and the decision-making process.



## 1.2 THE GROWTH AND DEVELOPMENT OF MODERN STATISTICS

Historically, the growth and development of modern statistics can be traced to two separate phenomena—the needs of government to collect information on its citizenry (see References 4, 5, 16, and 17) and the development of the mathematics of probability theory.

Data have been collected throughout recorded history. During the Egyptian, Greek, and Roman civilizations information was obtained primarily for the purposes of taxation and military conscription. In the Middle Ages, church institutions often kept records concerning births, deaths, and marriages. In America, various records were kept during colonial times (see Reference 17), and beginning in 1790 the Federal Constitution required the taking of a census every ten years. Today these data are used for many purposes, including congressional apportionment and the allocation of federal funds.

### 1.2.1 Descriptive Statistics

These and other needs for data on a nationwide basis were closely intertwined with the development of descriptive statistics.

**Descriptive statistics** can be defined as those methods involving the collection, presentation, and characterization of a set of data in order to properly describe the various features of that set of data.

Although descriptive statistical methods are important for presenting and characterizing information (see Chapters 2 through 5), it has been the development of inferential statistical methods as an outgrowth of probability theory that has led to the wide application of statistics in all fields of research today.



## 1.2.2 Inferential Statistics

The initial impetus for formulation of the mathematics of probability theory came from the investigation of games of chance during the Renaissance. The foundations of the subject of probability can be traced back to the middle of the seventeenth century in the correspondence between the mathematician Pascal and the gambler Chevalier de Mere (see References 5, 8, and 9). These and other developments by such mathematicians as Bernoulli, DeMoivre, and Gauss were the forerunners of the subject of inferential statistics. However, it has only been since the turn of this century that statisticians such as Pearson, Fisher, Gosset, Neyman, Wald, and Tukey pioneered in the development of the methods of inferential statistics that are widely applied in so many fields today.

**Inferential statistics** can be defined as those methods that make possible the estimation of a characteristic of a population or the making of a decision concerning a population based only on sample results.

To clarify this, a few more definitions are necessary.

A **population** (or **universe**) is the totality of items or things under consideration.

A **sample** is the portion of the population that is selected for analysis.

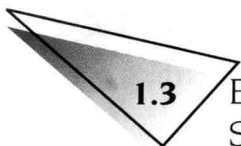
A **parameter** is a summary measure that is computed to describe a characteristic of an entire population.

A **statistic** is a summary measure that is computed to describe a characteristic from only a sample of the population.

Thus, one major aspect of inferential statistics is the process of using sample statistics to draw conclusions about the true population parameters.

The need for inferential statistical methods derives from the need for sampling. As a population becomes large, it is usually too costly, too time consuming, and too cumbersome to obtain our information from the entire population. Decisions pertaining to the population's characteristics have to be based on the information contained in a sample of that population. Probability theory provides the link by ascertaining the likelihood that the results from the sample reflect the results from the population.

These ideas can be clearly seen in the example of a political poll. If the pollster wishes to estimate the percentage of the votes a candidate will receive in a particular election, he or she will not interview each of the thousands (or even millions) of voters. Instead, a sample of voters will be selected. Based on the outcome from the sample, conclusions will be drawn concerning the entire population of voters. Appended to these conclusions will be a probability statement specifying the likelihood or confidence that the results from the sample reflect the true voting behavior in the population.



## 1.3 ENUMERATIVE VERSUS ANALYTICAL STUDIES

The examples that we have just provided concerning inferential statistical methods allow us to make an important distinction between two types of statistical studies that are undertaken: **enumerative** studies and **analytical** studies.

**Enumerative studies** involve decision-making regarding a population and/or its characteristics.