


# A COURSE IN MODERN BUSINESS STATISTICS

SECOND EDITION

TERRY SINCICH



BOOK NOT RETURNABLE IF DISK  
ENVELOPE SEAL IS BROKEN  
© 1998 WILEY-INTERSCIENCE, A JOHN WILEY & SONS, INC. PUBLICATION  
WILEY COLLEGE PUBLISHING COMPANY

# A COURSE IN MODERN BUSINESS STATISTICS

**SECOND EDITION**

---

**TERRY SINCICH**

University of South Florida

**DELLEN**

an imprint of

**MACMILLAN COLLEGE  
PUBLISHING COMPANY**

New York

**MAXWELL MACMILLAN  
CANADA**

Toronto

**MAXWELL MACMILLAN  
INTERNATIONAL**

New York Oxford Singapore Sydney

**On the cover:** "Morning Mirror" is a bold acrylic on panel painting, 11 × 14 inches, by New York artist Gary Lang. His "mirror" is a complex composition of interwoven opaque, transparent, and fluorescent colors that vibrates with optical intensity.

Lang received a Master of Fine Arts degree from Yale University in 1975 and has been the focus of numerous one-man exhibitions throughout the United States and Europe. His work is included in the permanent collections of the Whitney Museum of American Art, New York City; the Museum of Contemporary Art, Los Angeles; the Madison Art Center, Madison, Wisconsin; and the Menil Collection, Houston, Texas. He is represented by Brian Gross Fine Art, San Francisco, and Michael Klein, Inc., New York City.

Editor: Don Dellen  
Production supervisor: Phyllis Niklas  
Text designer: Janet Bollow  
Cover designer: John Williams  
Cartoons: Tom Barnett  
Technical illustrations: Folium

Copyright 1994 by Macmillan College Publishing Company, Inc.  
Dellen is an imprint of Macmillan College Publishing Company.

Printed in the United States of America

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the Publisher.

Macmillan College Publishing Company  
866 Third Avenue  
New York, NY 10022

Macmillan College Publishing Company is part of the  
Maxwell Communication Group of Companies.

Maxwell Macmillan Canada, Inc.  
1200 Eglinton Avenue East, Suite 200  
Don Mills, Ontario M3C 3N1

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

Sincich, Terry.

A course in modern business statistics / Terry Sincich. —2nd ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-02-410481-7

1. Commercial statistics. 2. Social sciences—Statistical methods. I. Title.

HF1017.S533 1994

93-39589

CIP

Printing: 1 2 3 4 5 6 7 8 9      Year: 4 5 6 7 8

# PREFACE

---

## **PEDAGOGY**

*A Course in Modern Business Statistics*, Second Edition, is designed so that all major topics in introductory applied statistics can be covered in a single course. The prerequisite is only a high school background in mathematics.

The text incorporates American Statistical Association (ASA) guidelines developed from a series of conferences titled “Making Statistics More Effective in Schools of Business.” At these annual conferences, there is a substantial consensus among the participants on the following points:

1. Students are most effectively motivated by seeing statistics at work in real applications, problems, cases, and projects.
2. Students should be provided with the opportunity to work with real data and to make serious use of statistical computations.
3. Formal training in probability needs to be downplayed in favor of intuitive concepts of probability.
4. We need to reduce our emphasis on the formal theory of statistics and increase emphasis on applications.

These ideas have culminated in a text that focuses on practical applications and interpretation of results, with less emphasis on calculations and probability.

---

## **FEATURES**

Specifically, the main features of the text are:

1. An **early introduction to confidence intervals and hypothesis testing** (Chapter 5) familiarizes students with the basics of inferential statistics. This enables the instructor to cover all the fundamental statistical ideas and methods (eg, descriptive statistics, sampling distributions, interval estimation, hypothesis testing, control charts, and regression) in a single semester, if desired.
2. **Examples and problem sets are based on real data.** The data sets and problem ideas are extracted from academic journals, newspaper articles, and magazines. This practical orientation helps the student relate statistics to problems encountered in the real world and, hopefully, will develop a pattern of thought that will persist after the student begins his/her career.
3. **Case studies** detail interesting current problems in business. Each chapter ends with a case study that poses a series of questions for the student. The student is expected to solve the problem using the statistical concepts and methodology

presented in the chapter. Like the examples and problems, the cases are extracted from academic journals, newspaper articles, and magazines.

4. **Computer printouts** generated from three popular statistical software packages (**SAS, SPSS, MINITAB**) are presented throughout the text. This enables the instructor to emphasize interpretation of the statistical results rather than the computation of those results.
5. **Computer labs** appear throughout the text. These sections provide the SAS, SPSS, and MINITAB commands used to generate the computer printouts displayed in the chapter.
6. **Nonparametric and parametric statistical methods are integrated** throughout. Whenever a parametric technique is presented, a corresponding distribution-free (nonparametric) method is also discussed. This provides the student with an alternative method to solving a problem when, as often occurs in the real world, the parametric assumptions are violated.
7. **Chapter notes**, summarizing key ideas, definitions, formulas, and interpretations, are provided for the student at the end of each chapter.
8. **Key concepts are highlighted.** Throughout the text, definitions, formulas, steps to follow in performing a statistical procedure, and warnings (indicating a specific situation where a method is likely to be misused) are boxed and highlighted.
9. **Sample exam questions** are provided for logically grouped chapters. These give the student a feel for the types of questions that might be expected on an exam, and provides the instructor with sample questions that cover the material presented.
10. **Short answers to all problems** are provided at the back of the book. Answers to the case studies and sample exam questions are excluded so that they may be used for homework assignments, quizzes, etc.

---

## REVISIONS

The second edition contains several substantial modifications, additions, and enhancements:

1. The material on **exploring and describing data is expanded and presented in two chapters:** Chapter 2 (qualitative data) and Chapter 3 (quantitative data). New sections in Chapter 2 cover summary frequency tables (Section 2.1) and cross-classification tables (Section 2.5). New sections added to Chapter 3 cover dot plots (Section 3.1) and sequence plots (Section 3.4).
2. The original chapter on **probability and statistical inference is split into two chapters:** Chapter 4 (basic concepts of probability) and Chapter 5 (introduction to statistical inference).
3. Ideas on **statistical process and quality control have been consolidated into a single chapter** (Chapter 9). New sections on total quality management (Section 9.1) and tolerance limits (Section 9.7) have been added.
4. Material on **regression analysis is expanded and presented in two parts:** Chapter 10 (part I) provides the basics of performing a regression analysis and presents several different types of linear models. Chapter 11 (part II) focuses on model building and model diagnostics. New sections in Chapter 11 cover variable transformations (Section 11.3) and influential observations (Section 11.7).

5. **Two new sections have been added to the time series chapter** (Chapter 12), one on index numbers (Section 12.3) and one on forecasting using lagged values of the dependent variable (Section 12.8).
6. A **new chapter on experimental design and analysis of variance** has been written (Chapter 13).
7. Several **new case studies** have been added.
8. **More problems** have been added to the problem sets in each chapter.

---

## **COURSE COVERAGE**

Depending on the amount of material covered, the text can be used at either a 2-year junior college or 4-year institution, and at the undergraduate or graduate (ie, master's) level. For example, the text can be used in the following courses:

1. One-semester introductory statistics course at a 2-year junior college (suggested coverage: Chapters 1–6, 8).
2. One-semester introductory statistics course at a 4-year college or university (suggested coverage: Chapters 1–6, 8–10).
3. Two-semester introductory statistics course at a 4-year college or university (suggested coverage: Chapters 1–8 in semester 1; Chapters 9–13 in semester 2).
4. One-semester course in applied statistics for MBA students (suggested coverage: Chapters 1–10, 12).

At first glance, these multiple uses for the book may seem inconsistent. For example, how can a text be appropriate for both undergraduate and graduate business students? The answer lies in the content. In contrast to a course in statistical theory, the level of mathematical knowledge required for an applied statistics course is minimal. Consequently, the difficulty encountered in learning the mechanics is much the same for both undergraduate and graduate students. The challenge is in the application—diagnosing practical problems, deciding on the appropriate statistical method for a given situation, and knowing which inferential technique will answer a manager's practical question. This takes *experience*, and it explains why a student can take an undergraduate course in applied statistics and still benefit from covering some of the same ground in an MBA course.

---

## **SUPPLEMENTS**

This text is accompanied by the following supplementary material:

1. **Student's Solutions Manual** (by Mark Dummeldinger). The manual contains the full solutions for the odd-numbered problems contained in the text.
2. **Instructor's Solutions Manual** (by Mark Dummeldinger). The manual contains the complete solutions to the even-numbered problems, case studies, and sample exam questions contained in the text.
3. **Data sets on diskette**. Several of the large data sets analyzed in examples and case studies are available (in ASCII format) on either a  $3\frac{1}{2}$  or  $5\frac{1}{4}$  inch IBM PC diskette. The data sets include the starting salaries of 2,000 college graduates, sale prices for 8,000 residential properties, costs of 300 HMO physicians, and low-bid prices for 300 bread contracts. The disk also includes data for all problem sets containing 20 or more observations.

4. **ASP statistical software diskette.** New to this edition, the text includes inside the back cover a 3½ inch micro disk containing the ASP program, *A Statistical Package for Business, Economics, and the Social Sciences*. ASP, from DMC Software, Inc., is a user-friendly, totally menu-driven program that contains all the major statistical applications covered in the text, plus many more. ASP runs on any IBM-compatible PC with at least 512K of memory and two disk drives. With ASP, students with no knowledge of computer programming can create and analyze data sets easily and quickly. Appendix C contains start-up procedures and a short tutorial on the use of ASP. Full documentation is provided complimentary to adopters of the text.
5. **ASP Tutorial and Student Guide** (by George Blackford). Most students have little trouble learning to use ASP without documentation. Some, however, may want to purchase the *ASP Tutorial and Student Guide*. Book stores can order the tutorial from DMC Software, Inc., 6169 Pebbleshire Drive, Grand Blanc, MI, 48439.
6. **Test Bank Manual** (by Mark Dummeldinger). This manual provides a large number of test items utilizing real data.
7. **DellenTest software.** This unique computer-generated random test system is complimentary to adopters. Utilizing an IBM (or compatible) PC and printer, the system will generate an almost unlimited number of quizzes, chapter tests, final examinations, and drill exercises. At the same time, the system produces an answer key and student worksheet with an answer column that exactly matches the column on the answer key.

## **ACKNOWLEDGMENTS**

Several individuals deserve special recognition upon completion of the second edition of this text. Don Dellen, my publisher, encouraged me to write the text and then provided all the necessary support to ensure the quality of the finished product. Joyce Curry-Daly (California Polytechnic Institute) provided a comprehensive and constructive review of the first edition; many of the changes you see in the text are due to her suggestions. Phyllis Niklas did an excellent job of supervising the production of the second edition. My wife, Faith Sincich, served in several capacities on both editions of the text, including word-processing specialist, proofreader, acquisition of permissions, moral supporter, and my toughest critic. Without Faith, this text could not have been completed. Finally, I thank my daughters, Kara and Kelly, for their love and understanding during those times when they failed to receive the attention they so richly deserved.

# CONTENTS

Preface ix

- 1 INTRODUCTION: STATISTICS AND DATA 1**
  - 1.1** What Is Statistics? 2
  - 1.2** Types of Data 3
    - Computer Lab 1** Entering and Listing Data 6
  - 1.3** Fundamental Elements of a Statistical Analysis 12
  - 1.4** Collecting Data 17
    - Computer Lab 2** Generating a Random Sample 20
  - 1.5** Overview of This Text 22
    - Chapter Notes** 23
    - Looking Ahead** 23
    - Case Study** The Latest Hite Report—Controversy Over the Numbers 24
    - References** 26
  
- 2 EXPLORING AND DESCRIBING QUALITATIVE DATA 27**
  - 2.1** Summary Frequency Tables 28
  - 2.2** Bar Graphs 30
  - 2.3** Pie Charts 36
  - 2.4** Pareto Diagrams 40
  - 2.5** Cross-Classification (Two-Way) Tables 43
    - Computer Lab** Describing Qualitative Data Sets 47
    - Chapter Notes** 53
    - Looking Ahead** 54
    - Case Study** Characteristics of HMO Physicians 54
    - References** 58
  
- 3 EXPLORING AND DESCRIBING QUANTITATIVE DATA 59**
  - 3.1** Dot Plots 60
  - 3.2** Stem-and-Leaf Displays 64



<b>3.3</b>	Histograms	69
<b>3.4</b>	Sequence Plots	79
	<b>Computer Lab 1</b> Describing Quantitative Data Graphically	84
<b>3.5</b>	Descriptive Statistics: Central Tendency	92
<b>3.6</b>	Descriptive Statistics: Spread	99
<b>3.7</b>	Some Rules for Describing Quantitative Data	103
<b>3.8</b>	Descriptive Statistics: Relative Standing	110
<b>3.9</b>	Detecting Unusual Observations: Outliers	116
<b>3.10</b>	Detecting Bivariate Relationships	123
	<b>Computer Lab 2</b> Describing Quantitative Data Numerically	134
	<b>Chapter Notes</b>	140
	<b>Looking Ahead</b>	143
	<b>Case Study</b> Fudging the Data	143
	<b>References</b>	145
	<b>Sample Exam Questions:</b> Chapters 1–3	145

## **4 PROBABILITY AND PROBABILITY DISTRIBUTIONS** 149

<b>4.1</b>	Basic Concepts of Probability	150
<b>4.2</b>	Random Variables	162
<b>4.3</b>	Probability Models	164
<b>4.4</b>	The Binomial Distribution for Discrete Data	166
<b>4.5</b>	The Normal Distribution for Continuous Data	174
	<b>Computer Lab</b> Normal Probability Plots	185
	<b>Chapter Notes</b>	187
	<b>Looking Ahead</b>	189
	<b>Case Study</b> Break-Even Analysis—When to Market a New Product	189
	<b>References</b>	190

## **5 INTRODUCTION TO STATISTICAL INFERENCE** 191

<b>5.1</b>	Sampling Distributions	192
<b>5.2</b>	Large-Sample Confidence Intervals	201
<b>5.3</b>	Large-Sample Tests of Hypothesis	210
<b>5.4</b>	Inferences Based on Small Samples	224
<b>5.5</b>	Reporting Test Results: $p$ -Values	229
<b>5.6</b>	Distribution-Free Tests: Nonparametrics	232
	<b>Chapter Notes</b>	235
	<b>Looking Ahead</b>	237
	<b>Case Study</b> A Decision Problem for Financial Managers: When to Investigate Cost Variances	237
	<b>References</b>	240
	<b>Sample Exam Questions:</b> Chapters 4–5	240

<b>6</b>	<b>ONE-SAMPLE AND TWO-SAMPLE INFERENCE METHODS FOR QUANTITATIVE DATA: MEANS</b>	243
6.1	General Form of Confidence Intervals and Test Statistics	244
6.2	Inferences About a Single Population Mean	245
6.3	Comparing Two Population Means: Independent Samples	254
6.4	Comparing Two Population Means: Paired Samples	268
6.5	Determining the Sample Size	279
	<b>Computer Lab</b> Tests and Confidence Intervals for Means	286
	<b>Chapter Notes</b>	295
	<b>Looking Ahead</b>	296
	<b>Case Study</b> The SAT Pill	296
	<b>References</b>	297
	<b>Sample Exam Questions:</b> Chapter 6	298
<b>7</b>	<b>ONE-SAMPLE AND TWO-SAMPLE INFERENCE METHODS FOR QUANTITATIVE DATA: VARIANCES</b>	303
7.1	Inferences About a Single Population Variance	304
7.2	Comparing Two Population Variances	312
	<b>Computer Lab</b> Testing Variances	326
	<b>Chapter Notes</b>	329
	<b>Looking Ahead</b>	330
	<b>Case Study</b> Bid-Rigging in the Florida Bread Market	330
	<b>References</b>	331
	<b>Sample Exam Questions:</b> Chapter 7	332
<b>8</b>	<b>INFERENCE FOR QUALITATIVE DATA: PROPORTIONS</b>	335
8.1	Inferences About a Single Population Proportion	336
8.2	Comparing Two Population Proportions	341
8.3	Determining the Sample Size	347
8.4	Comparing Multinomial Proportions: One-Way Table	352
8.5	Comparing Multinomial Proportions: Two-Way Table	358
	<b>Computer Lab</b> Testing Multinomial Proportions	367
	<b>Chapter Notes</b>	372
	<b>Looking Ahead</b>	374
	<b>Case Study</b> Does an Aspirin a Day Really Keep the Heart Doctor Away?	374
	<b>References</b>	376
	<b>Sample Exam Questions:</b> Chapter 8	376
<b>9</b>	<b>STATISTICAL PROCESS AND QUALITY CONTROL</b>	379
9.1	Total Quality Management	380
9.2	Variable Control Charts	380

9.3	Control Chart for Means	386
9.4	Control Chart for Process Variation	394
9.5	Detecting Trends in a Control Chart: Runs Analysis	398
9.6	Control Chart for Percent Defectives	401
9.7	Tolerance Limits	405
	<b>Computer Lab</b> Control Charts	410
	<b>Chapter Notes</b>	413
	<b>Looking Ahead</b>	415
	<b>Case Study</b> Monitoring the Manufacture of Steel Rods	415
	<b>References</b>	417
	<b>Sample Exam Questions:</b> Chapter 9	417

## 10 MODELING RELATIONSHIPS BETWEEN VARIABLES: REGRESSION (PART I) 421

10.1	The Concept of a Statistical Model	422
10.2	Collecting the Data for Regression Analysis	425
10.3	Simple Linear Regression	428
	<b>Computer Lab 1</b> Simple Linear Regression	454
10.4	Multiple Regression Models	459
10.5	First-Order Models with Quantitative Variables	461
10.6	Interaction Models with Quantitative Variables	480
10.7	Second-Order (Quadratic) Models with Quantitative Variables	489
10.8	Models with Qualitative (Dummy) Variables	498
	<b>Computer Lab 2</b> Multiple Regression	505
10.9	Nonparametric Regression	511
	<b>Chapter Notes</b>	518
	<b>Looking Ahead</b>	521
	<b>Case Study</b> The Salary Race: Males Versus Females	522
	<b>References</b>	523

## 11 MODEL BUILDING AND MODEL DIAGNOSTICS: REGRESSION (PART II) 525

11.1	Model Building	526
11.2	Stepwise Regression	528
11.3	Variable Transformations	538
11.4	Multicollinearity	546
11.5	Tests for Comparing Nested Models	555
11.6	Residual Analysis	564
11.7	Influential Observations	581
	<b>Computer Lab</b> Stepwise Regression, Residuals, and Regression Diagnostics	591
	<b>Chapter Notes</b>	607
	<b>Looking Ahead</b>	610
	<b>Case Study</b> Building a Model for Road Construction Costs	610

**References** 617

**Sample Exam Questions:** Chapters 10–11 617

## **12 TIME SERIES ANALYSIS AND FORECASTING** 623

**12.1** Time Series Components 624

**12.2** Collecting Time Series Data 626

**12.3** Index Numbers 627

**12.4** Forecasting Using Smoothing Techniques 634

**12.5** Forecasting Using Regression 647

**12.6** Testing for Autocorrelation 653

**12.7** Forecasting Using Autoregressive Models 659

**12.8** Forecasting Using Lagged Values of the Dependent Variable 668

**Computer Lab** Time Series Methods 669

**Chapter Notes** 673

**Looking Ahead** 674

**Case Study** Forecasting and Back-Casting Peak Electricity Demand 675

**References** 681

**Sample Exam Questions:** Chapter 12 682

## **13 EXPERIMENTAL DESIGN: ANALYSIS OF VARIANCE** 685

**13.1** Introduction 686

**13.2** Experimental Design: Terminology 686

**13.3** The Logic Behind an Analysis of Variance 689

**13.4** Completely Randomized Designs 691

**13.5** Randomized Block Designs 700

**13.6** Factorial Experiments 712

**13.7** Follow-Up Analysis: Multiple Comparisons of Means 725

**13.8** The Relationship Between ANOVA and Regression 732

**13.9** Checking ANOVA Assumptions 737

**13.10** Selecting the Sample Size for an Experiment 743

**13.11** The Importance of Randomization 745

**Computer Lab** Analysis of Variance 745

**Chapter Notes** 756

**Case Study** Reluctance to Transmit Bad News: The MUM Effect 758

**References** 760

**Sample Exam Questions:** Chapter 13 761

## **APPENDIX A: STATISTICAL TABLES** A1

**Table 1** Cumulative Binomial Probabilities A3

**Table 2** Normal Curve Areas A6

<b>Table 3</b>	Critical Values for Student's $t$	A7
<b>Table 4</b>	Critical Values for the Wilcoxon Rank Sum Test	A8
<b>Table 5</b>	Critical Values for the Wilcoxon Signed Ranks Test	A9
<b>Table 6</b>	Critical Values for the $\chi^2$ Statistic	A10
<b>Table 7</b>	Critical Values for the $F$ Statistic: $F_{.10}$	A12
<b>Table 8</b>	Critical Values for the $F$ Statistic: $F_{.05}$	A14
<b>Table 9</b>	Critical Values for the $F$ Statistic: $F_{.025}$	A16
<b>Table 10</b>	Critical Values for the $F$ Statistic: $F_{.01}$	A18
<b>Table 11</b>	Critical Values for the Ansari–Bradley Test	A20
<b>Table 12</b>	Factors Used When Constructing Control Charts	A25
<b>Table 13</b>	Values of $K$ for Tolerance Limits for Normal Distributions	A26
<b>Table 14</b>	Sample Size $n$ for Nonparametric Tolerance Limits	A27
<b>Table 15</b>	Critical Values of Spearman's Rank Correlation Coefficient	A28
<b>Table 16</b>	Critical Values for the Theil Zero-Slope Test	A29
<b>Table 17</b>	Critical Values for the Durbin–Watson $d$ Statistic ( $\alpha = .05$ )	A31
<b>Table 18</b>	Critical Values for the Durbin–Watson $d$ Statistic ( $\alpha = .01$ )	A32

## **APPENDIX B: CALCULATION FORMULAS FOR PARAMETRIC ANALYSIS OF VARIANCE**

Completely Randomized Design	A33
Randomized Block Design	A33
Two-Factor Factorial Experiment	A34

## **APPENDIX C: ASP TUTORIAL**

<b>SHORT ANSWERS TO PROBLEMS</b>	A41
<b>INDEX</b>	11

# CHAPTER 1

# INTRODUCTION: STATISTICS AND DATA

## CONTENTS

1.1 What Is Statistics?

1.2 Types of Data

### Computer Lab 1

Entering and Listing Data

1.3 Fundamental Elements of a Statistical Analysis

1.4 Collecting Data

### Computer Lab 2

Generating a Random Sample

1.5 Overview of This Text

## Chapter Notes

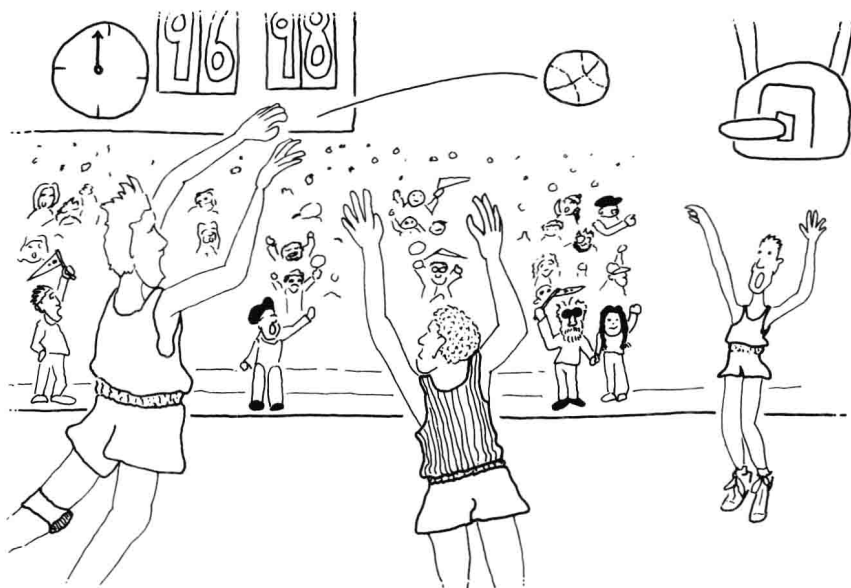
### Looking Ahead

### Case Study

The Latest Hite Report—Controversy Over the Numbers

## References

**C**an taking drugs improve your score on the Scholastic Aptitude Test (SAT)? Are the final outcomes of professional basketball games determined solely by what happens in the fourth quarter? How many job offers, on average, do graduating MBAs receive? Questions such as these can be answered if relevant data are collected. This chapter presents a general overview of the science of statistics and its value in extracting useful and reliable information from data.



# 1.1

 Consider the following recent items from the news media:

*The Wall Street Journal*, July 14, 1992

Video card games (eg, video poker), once only legal in casinos, are now available for play in laundromats and grocery stores of states that have a lottery. These video card games are beginning to create adult addicts, many of them women, who gamble until they're broke and in debt. In a study of 52 women in Gamblers Anonymous, 90% were video poker players. After they became hooked on video poker, 75% exhausted their family savings, 33% embezzled from their employers, 25% declared bankruptcy, and 10% turned to prostitution to raise money.

*Chance*, Vol. 5, No. 3–4, 1992

Folk wisdom suggests that the final outcome of a professional basketball game is heavily dependent on what happens in the last quarter (final 12 minutes), if not the final 2 minutes, of playing time. Information for 189 National Basketball Association games was collected from the box scores reported in the *St. Louis Post-Dispatch*. In these games, the team leading after 3 quarters of play won 150 times and lost only 39 times—a “leader-win” percentage of almost 80%.

*U.S. News & World Report*, Mar. 22, 1993

According to *U.S. News'* fourth annual survey of “America’s Best Graduate Schools,” an MBA degree candidate received an average of only one job offer prior to graduation, compared to 3.8 job offers a decade ago. On some campuses last year, one in five MBA candidates had not received a single job offer by graduation.

*Tampa Tribune*, Jan. 10, 1990

A computer study conducted at Columbia University demonstrated that shuffling a deck of cards seven times prevents bettors from accurately predicting the next card in blackjack. Most Atlantic City casinos will continue with their customary two shuffles of the deck, however, since the extra 5 minutes it takes to perform the additional shuffles could result in a substantial profit loss.

*Time*, Jan. 22, 1990

Nissan’s ad campaign for its Infiniti luxury cars was originally renowned for a novel gimmick: The autos were nowhere in sight. The Infiniti ads, which depicted lushly photographed trees, boulders, lightning bolts and ocean waves (but no cars), were found by a Gallup poll to be the best-recalled commercial on television. Unfortunately, Infiniti dealers in the United States sold only 32 cars a day during the last 2 months of the year, compared to the 134 cars per day sold by its archrival, the Toyota Lexus.

Every day we are inundated with bits of information—data—like those in the examples above, whether we are in the classroom, on the job, or at home. Many of you taking this course are studying to be (or may already be) managers of a business or firm. Some of you will be **data producers**, but most of you will be **data users**. As such, you will need to be able to make sense out of the mass of data that others produce for you. What specialized tools will enable you to become effective data users? The answer is **statistics**.

A common misconception is that a statistician is simply a “number cruncher,” or a person who calculates and summarizes numbers, like baseball batting averages or unemployment rates. Statistics involves numbers, but there is much more to it than that.

According to *The Random House College Dictionary* (1988 ed.), statistics is “the science that deals with the collection, classification, analysis, and interpretation of numerical facts or data.” In short, statistics is the **science of data**—a science that will enable you to be proficient data producers and efficient data users.

#### DEFINITION 1.1

**Statistics** is the science of data. This involves collecting, classifying, summarizing, organizing, analyzing, and interpreting data.

In this chapter we explore the different types of data that you will encounter in business and introduce you to some ideas on methods for collecting data. The various statistical methods for summarizing, analyzing, and interpreting data are presented in the chapters that follow.

## TYPES OF DATA

**1.2** Data are obtained by measuring some characteristic or property of the objects (usually people or things) of interest to us. These objects upon which the measurements (or observations) are made are called **experimental units**, and the properties being measured are called **variables** (since, in virtually all studies of interest, the property varies from one observation to another).

#### DEFINITION 1.2

An **experimental unit** is an object (person or thing) upon which we collect data.

#### DEFINITION 1.3

A **variable** is a characteristic (property) that differs, or varies, from one observation to the next.

All data (and consequently, the variables we measure) are either **quantitative** or **qualitative** in nature.\* Quantitative data are data that can be measured on a numerical scale. In general, qualitative data take values that are nonnumerical; they can only be classified. The statistical tools that we use to analyze data depend on whether

\* A finer breakdown of data types into nominal, ordinal, interval, and ratio data is possible. **Nominal** data are qualitative data with categories that cannot be meaningfully ordered. **Ordinal** data are also qualitative data, but a distinct ranking of the groups from high to low exists. **Interval** and **ratio** data are two different types of quantitative data. For most statistical applications (and all the methods presented in this introductory text), it is sufficient to classify data as either quantitative or qualitative.



the data are quantitative or qualitative. Thus, it is important to be able to distinguish between the two types of data.

**DEFINITION 1.4**

**Quantitative data** are observations measured on a numerical scale.

**DEFINITION 1.5**

Nonnumerical data that can only be classified into one of a group of categories are **qualitative data**.

To illustrate, consider the data in Table 1.1. This data set, obtained from *Business Week's* 1993 "Executive Compensation Scoreboard," contains information on 10 of the 365 corporate executives who participated in the survey. In this example, the experimental units are the 10 corporate executives. For each executive (ie, each observation) five variables are recorded: (1) company, (2) industry group, (3) total 1992 pay (in thousands of dollars), (4) return to shareholders (in dollars) on a \$100 investment made 3 years earlier, and (5) pay-for-performance rating measured on a scale of 1 (excellent) to 5 (poor).

The first two variables (company and industry group) are qualitative since the data they produce are values that are nonnumerical values; they can only be classified into categories or groups. The next two variables (total pay and shareholder return) are quantitative since they are measured on a numerical scale. The fifth variable (performance rating), although coded as a number (1–5), is really qualitative in nature. The performance categories are "excellent," "above average," "average," "be-

**TABLE 1.1**  
Data on 10 CEOs in the 1993 Executive Compensation Scoreboard

CEO	(1) COMPANY	(2) INDUSTRY GROUP	(3) TOTAL PAY	(4) SHAREHOLDER RETURN	(5) PERFORMANCE RATING
C. Lazarus	Toys 'R' Us	Retailing	64,231	168	5
A. O'Reilly	Heinz	Food processing	36,918	135	5
W. Anders	General Dynamics	Aerospace	29,015	239	5
L. Bantle	UST Inc.	Tobacco	24,602	222	4
R. Allen	Delta	Airlines	2,740	80	2
H. Poling	Ford Motor	Automotive	6,167	113	3
D. Calloway	Pepsico	Beverages	15,984	201	3
A. Hirsig	Arco	Chemicals	1,950	141	1
P. Freiman	Syntex	Drugs	3,170	103	2
J. Marriott	Marriott	Food & lodging	3,421	65	4

Source: "Executive Pay: The Party Ain't Over Yet." *Business Week*, Apr. 26, 1993. Reprinted from Apr. 26, 1993 issue of *Business Week* by special permission, copyright © 1993 by McGraw-Hill, Inc. Used with permission.