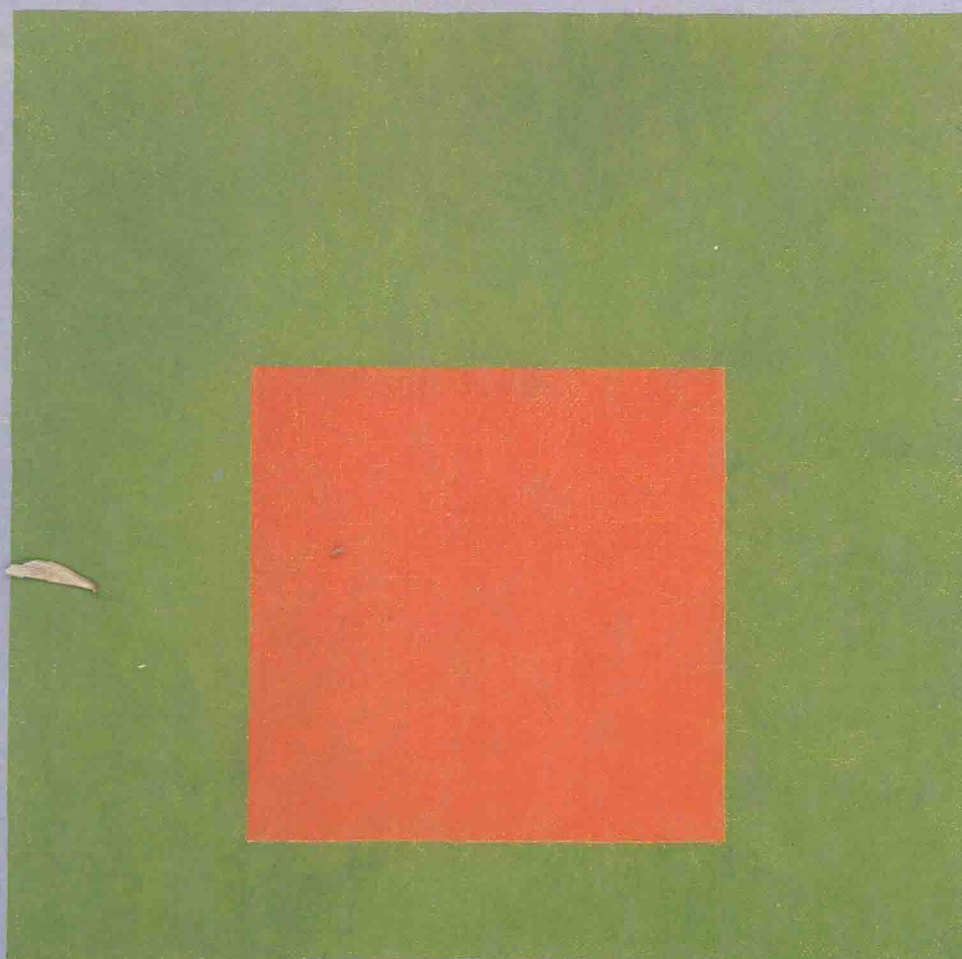


# STATISTICS

FOR BUSINESS AND ECONOMICS



International  
Edition  
Special Price!

ROBERT SANDY

# **STATISTICS FOR BUSINESS AND ECONOMICS**

**ROBERT SANDY**

INDIANA UNIVERSITY—PURDUE UNIVERSITY AT INDIANAPOLIS

**McGRAW-HILL PUBLISHING COMPANY**

New York St. Louis San Francisco Auckland  
Bogotá Caracas Hamburg Lisbon London Madrid  
Mexico Milan Montreal New Delhi Oklahoma City  
Paris San Juan São Paulo Singapore Sydney  
Tokyo Toronto

## STATISTICS FOR BUSINESS AND ECONOMICS

Copyright © 1990 by McGraw-Hill, Inc. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the publisher.

2 3 4 5 6 7 8 9 0 DOH DOH 9 4 3 2 1 0

ISBN 0-07-557119-6

This book was set in Bembo by York Graphic Services, Inc.  
The editors were Bonnie Binkert and Judith Kromm;  
the production supervisor was Stacey Alexander.  
Design was done by Binns & Lubin.  
R. R. Donnelley & Sons Company was printer and binder.

Cover painting: Josef Albers, *Homage to the Square: Young*. 1951.  
Oil on masonite. 23 $\frac{3}{8}$  inches  $\times$  23 $\frac{1}{2}$  inches. The Metropolitan Museum of Art,  
George A. Hearn Fund, 1953.

### Library of Congress Cataloging in Publication Data

Sandy, Robert.

[1st ed]

Statistics for business and economics / Robert Sandy.

p. cm.

Includes index.

ISBN 0-07-557119-6

1. Social sciences—Statistical methods. 2. Commercial statistics. 3. Economics—Statistical methods. 4. Statistics.

I. Title.

HA29.S254 1990

519.5'024658—dc19

88-15672

*To Elaine, Steven, and Rachel who have borne most of the cost*

# Preface

*Historically, we must admit that the classical version of our key introductory college course has been an abysmal failure, semester after semester, year after year. It deserves a grade of F—yet we persist in teaching it more or less the same way we have been for forty years.*

This disheartening quotation is from an article by Brian Joiner, one of the authors of MINITAB (“Let’s Change How We Teach Statistics,” *Chance*, 1988, no. 1). Joiner advocates two changes in the way introductory Statistics is taught: First, give students real data sets to manipulate throughout the course, and second, eliminate most of the standard material on parametric tests. This text reflects the first recommendation, but I chose not to perform major surgery on the course topics, as the second suggestion would require. I present the traditional topics, but I’ve done everything I can to make them intelligible and interesting.

My main objectives in writing this text were to:

- Explain the material in a highly intuitive and conversational manner with well-written and lively examples,
- Systematize the problem-solving process,
- Emphasize graphical methods to describe models and analyze data, and
- Provide real computer applications for more topics than any other text of this kind.

Wherever possible, explanations take the place of perfunctory statements such as “it can be shown that . . .” or “in advanced statistics texts it is shown that . . .” Proofs are relegated to an appendix. The discussion of why the formula for the sample standard deviation has  $n - 1$  in the denominator (Chapter 3) illustrates this intuitive approach. Other examples are the explanations of why the distribution of sample means approaches a normal distribu-

tion as the sample size increases (Chapter 7) and why the chi-square test statistic is distributed as the sum of a series of independent, squared standard normal distributions (Chapter 11).

This text emphasizes the development of the skills needed to recognize which distribution applies in a given situation and to translate a problem from words into equations. The charts inside the front and back covers present clear-cut rules that lead students to the correct formula for almost every hypothesis test or confidence interval discussed in the text.

Coverage is extensive, yet some familiar topics have been omitted. The choice was guided by two criteria. First, each topic covered should have a realistic business application. The discrete uniform distribution and kurtosis are two of the topics that were excluded because they lacked obvious applications. On the other hand, the exponential distribution, ARIMA, and probit regression, which are generally not covered in other texts, are discussed because they have good business applications. The second criterion for deciding which topics to cover was my conviction that the text should reflect the widespread use of computers in the practice of statistics in business. I believe that virtually all business-related statistical computations will soon be done with computers and that there is no benefit to covering techniques aimed solely at reducing computation time. Assuming that students using this text have access to either a computer with a statistical package or a hand-held calculator that has functions such as summation, mean and standard deviation, and a memory register, there is no need to discuss short-cut formulas and coding.

The growing use of computers also led to the inclusion of some topics that usually are *not* covered in an introductory text. For example, most statistical packages present the pooled and the unpooled variance versions of a difference-in-means test, and many computer packages have graphing routines that simplify the creation of charts such as box-and-whisker plots. These and other topics are addressed so that students can handle the options presented by a computer package or take advantage of some easy-to-understand but computationally tedious techniques.

Although the text is designed to be used in courses with or without computers, it offers many advantages if students have access to a computer. The computer applications start with Chapter 2 and continue through Chapter 17. The sections of the chapters that have computer applications are marked with either one or two computer disks in the table of contents and in the text. It is recommended that the double-disk sections be covered only in classes that have access to a computer. The rest of the text does not require use of a computer.

A computer manual is bound into the text as Appendix A. This manual contains printouts and instructions for four widely used computer packages: DATA DESK, MINITAB, SAS, and SPSSx. MYSTAT, a student version of SYSTAT, is available from McGraw-Hill either separately or in a package with the text. The documentation and examples are packaged with the MYSTAT

disk rather than bound into the text. This extensive coverage is intended to help instructors integrate computers in the introductory statistics course by accommodating a wide variety of computer resources. DATA DESK works on Macintosh computers. MINITAB, SAS, and SPSSx are usually available on university mainframe computers. MYSTAT works on IBM PCs and compatibles, as well as Macintosh computers. The five packages were chosen to reflect current mainframe usage in statistics courses as well as the emerging transition to microcomputers.

Other significant features of the text include the following:

- Over 800 exercises. Within each chapter the exercises increase in difficulty up to the highly challenging starred exercises at the end. Each section includes several drill exercises that review the mechanics of applying a formula. Exercises at the end of each section and at the end of the chapter provide a realistic context for the business use of a technique. Adding interest are exercises relating to student life and exercises having a humorous twist.
- Solutions to the odd-numbered exercises show the intermediate steps. Alternative solutions to a problem, given the computational resources available and the validity of parametric assumptions, are detailed.
- Optional sections within most chapters which allow instructors to tailor the course to special interests and to students' varied abilities.
- Non-calculus proofs of most theorems in the text (Appendix C).
- Carefully refined statistical notation. For example, all parameters are symbolized by Greek letters and all statistics are symbolized by italic Latin letters.

A *Study Guide* to accompany this text is available from the publisher. Other supplements, which are available to instructors, include a sophisticated test bank with class-tested questions. The computerized test generator prints graphs, Greek letters, and statistical notation and randomizes the order of multiple choice questions. The test bank is also available in printed form. In addition, there are over 30 data sets on disk for use with the computer applications, transparency masters, and an instructor's manual. The transparency masters include the figures in the text as well as additional worked-out examples. The instructor's manual consists of solutions to all exercises in the text plus suggested outlines for one- and two-semester courses with and without access to a computer.

## ACKNOWLEDGMENTS

To a textbook author, reviewers retained by the publisher are invisible antagonists who have to be mentally wrestled into compliance with the author's goals and limitations of the publication process. I struggled with most of the reviewers' suggestions, repeatedly asking whether a given change fit my concept of the text. In the end, I accepted most of their ideas. Nevertheless, the reviewers

bear no responsibility for any errors that remain. My appreciation goes to the following reviewers: William Cooke, University of Wyoming; Larry Cornwell, Bradley University; Robert Kowalczyk, Southeastern Massachusetts University; Ralph Miller, California State University—Pomona; Amitava Mitra, Auburn University; Stephen K. Pollard, California State University—Los Angeles; Ruby Ramirez, Loyola University; Michael Sklar, University of Georgia; and Carol Stamm, Western Michigan University. Robert Kowalczyk and Lee Ing-Tong, both at Southeastern Massachusetts University, reviewed the computer appendixes. Susan Reiland did an accuracy check of the manuscript. Her reputation as a meticulous and expert reviewer is entirely deserved. The McGraw-Hill Business Series Advisor, David Dannenbring of Bernard Baruch College, reviewed the manuscript at all stages, reviewed the computer appendixes and checked the solutions to the exercises. I thank him for all of his help.

The first thirteen chapters of this text have been used at IUPUI for three years. I thank my colleagues who class-tested the manuscript: Robert Calhoun, Martin Spechler, Chris Starkey, Shah Towfighi, and Gang Yi. I am also grateful to the IUPUI students who offered suggestions. I thank David Bivin, Gang Yi, and Monte Juillerat for advice on specific topics. Martin Spechler's insightful advice and unflagging encouragement were well beyond any reasonable expectation.

In conclusion, I think that my treatment as an author by my publisher has been exemplary. Courtesy, patience, and consistency were the hallmarks of all of their actions. My contacts were with Seib Adams, editorial director of the college division; June Smith, director of editing and development; Johanna Schmid, acquisitions editor; Judith Kromm, project manager; Dan Alpert, senior development editor; Anne Mahony, development editor; and Stacey Alexander, production manager. Betty Binns of Binns & Lubin and Associates and Karl Illg of IUPUI provided invaluable input on design and graphics problems. Finally, my most earnest appreciation goes to Susan Badger who worked closely with me throughout the project.

ROBERT SANDY

*Indianapolis, Indiana*



# Contents

- means that a computer application is provided for that topic in Appendix A, but it is not necessary to have the use of a computer in order to cover the topic.
- ■ means that a computer application is provided at the end of the chapter for that topic and that it is recommended that the topic be covered only if a computer is available for students' use.

## Preface vii

## Chapter 1 Introduction 1

---

- 1.1 Defining statistics 1
- 1.2 Why many students have difficulty with statistics 4
- 1.3 Outline of the topics covered in the text 5

## Chapter 2 Types of data and describing data sets with graphs 8

---

- 2.1 Data sources and types of data 10
- 2.2 Graphic presentation of data 14

## Chapter 3 Numerical descriptive statistics 36

---

- 3.1 Describing the center of the data 37
- 3.2 Describing the spread of the data, the relative spread, and the proportions in the tails 51
- 3.3 Exploratory data analysis 61

## Chapter 4 Probability 80

---

- 4.1 Introduction to probability and definitions of probability 81
- 4.2 Addition and multiplication rules 84

- 4.3 Joint probability distributions 98
- 4.4 Permutations and combinations 103
- 4.5 Bayes' theorem 109

## Chapter 5 Discrete random variables 128

---

- 5.1 Introduction to random variables 129
- 5.2 The expected value, standard deviation, and variance of a random variable 132
- 5.3 Arithmetic properties of the expected value and variance of a random variable 142
- 5.4 The binomial distribution 146
- 5.5 The Poisson distribution 155
- 5.6 The Poisson approximation to the binomial distribution 160
- 5.7 The hypergeometric distribution 164

## Chapter 6 Continuous random variables and the normal and exponential distributions 178

---

- 6.1 Continuous random variables 180
- 6.2 The normal distribution 186
- 6.3 The normal approximation to the binomial distribution 197
- 6.4 The normal approximation to the Poisson distribution (optional) 204
- 6.5 The exponential distribution (optional) 208

## Chapter 7 Sampling methods, sampling distributions, and the properties of estimators 218

---

- 7.1 Types of samples 219
- 7.2 Sampling distributions and the central limit theorem 231
- 7.3 The sampling distribution of the proportion 243
- 7.4 The properties of point estimators (optional) 250

---

**Chapter 8 Confidence intervals 266**

---

- 8.1 Confidence intervals for the population mean when the population is normal and the population standard deviation is known 268
- 8.2 Confidence intervals for the population mean when the population standard deviation is not known 277
- 8.3 Confidence intervals for the population proportion when the normal approximation is used 284
- 8.4 Minimum sample sizes for a given confidence interval 291

---

**Chapter 9 Hypothesis tests with one population 306**

---

- 9.1 Formulation of a hypothesis test, null and alternative hypotheses, and Type I and Type II errors 308
- 9.2 Plotting the probability of a Type II error (optional) 320
- 9.3 Hypothesis tests on the population mean: population standard deviation not known 331
- 9.4 Hypothesis tests on the population proportion with the normal approximation 337
- 9.5 Hypothesis tests with  $p$ -values (optional) 343

---

**Chapter 10 Confidence intervals and hypothesis tests with two populations 354**

---

- 10.1 Confidence intervals and hypothesis tests on the difference between two population means: standard deviations known and independent samples 356
- 10.2 Confidence intervals and hypothesis tests on the difference between two population means: standard deviations not known and independent samples 366
- 10.3 Confidence intervals and hypothesis tests on the difference between two population proportions 383
- 10.4 Hypothesis tests on the difference between two population means with paired samples 392

## Chapter 11 The chi-square distribution and tests for independence and goodness of fit 410

---

- 11.1 Tests for goodness of fit 412
- 11.2 Tests for equality among proportions, independence, and homogeneity 432

## Chapter 12 Analysis of variance 450

---

- 12.1 One-way analysis of variance and the  $F$  distribution 453
- 12.2 Confidence intervals and hypothesis tests in one-way analysis of variance 467
- ■ 12.3 Two-way analysis of variance (optional) 473

## Chapter 13 Single regression and correlation 502

---

- 13.1 Introduction to regression and correlation 504
- 13.2 Confidence intervals and hypothesis tests on the population slope and intercept 530
- 13.3 Confidence intervals and hypothesis tests on the mean of  $Y$ , the predicted values of  $Y$ , and the coefficient of correlation 539
- ■ 13.4 Logarithmic transformations for establishing linearity and weighted least squares (optional) 555

## Chapter 14 Multiple regression 578

---

- ■ 14.1 Introduction to multiple regression 580
- ■ 14.2 Confidence intervals and hypothesis tests in multiple regression 603
- ■ 14.3 Dummy variables, interaction terms, and transformations to establish linearity (optional) 613
- ■ 14.4 Probit regression (optional) 629

## Chapter 15 Forecasting time series 650

---

- 15.1 Introduction to forecasting 653
- 15.2 Data smoothing 669

■ ■ 15.3 Autoregressive models 679

■ ■ 15.4 ARIMA models (optional) 689

---

## Chapter 16 Index numbers 712

■ 16.1 Introduction to index numbers 714

---

## Chapter 17 Nonparametric tests 734

■ 17.1 Nonparametric tests for the population median 737

■ 17.2 Nonparametric tests that use rankings 749

■ 17.3 Goodness-of-fit tests 776

---

## Chapter 18 Decision theory 796

■ 18.1 Introduction to decision making 798

■ 18.2 Decision making with Bayes' theorem 809

# Appendixes

---

## Appendix A Computer applications 830

■ ■ A2 Computer-generated graphs for descriptive statistics 832

■ ■ A3 Numerical descriptors and exploratory data analysis 836

■ ■ A4 Probability calculations and joint probability tables 846

■ ■ A5 Determining probabilities and drawing random samples 850

■ ■ A6 Determining probabilities and drawing random samples: uniform, normal, and exponential distributions 857

■ ■ A7 Simulating the sampling distribution of the mean and drawing random samples from a list 864

■ ■ A8 Calculating confidence intervals 869

■ ■ A9 Hypothesis tests with one population 873

■ ■ A10 Two-population confidence intervals and hypothesis tests 876

■ ■ A11 Using the chi-square distribution for tests of homogeneity, independence, and goodness of fit 882

- • A12 Analysis of variance 889
- • A13 Simple regression and correlation 901
- • A14 Multiple regression analysis 911
- • A15 Forecasting time series 918
- • A16 Index numbers 929
- • A17 Nonparametric tests 938

## Appendix B Confidence intervals and hypothesis tests for the population variance 950

---

## Appendix C Proofs 962

---

## Appendix D Statistical tables 973

---

- D.1 Binomial probabilities 973
- D.2 Poisson probabilities 978
- D.3 The hypergeometric probability distribution 980
- D.4 Areas under the standard normal probability distribution 988
- D.5 Areas under the exponential distribution 989
- D.6 Table of random digits 990
- D.7 Percentiles of student's  $t$  distribution 991
- D.8 Percentiles of  $\chi^2$  distribution 992
- D.9 Percentiles of  $F$  distribution 994
- D.10 Critical values of the Wilcoxon-Mann-Whitney  $T$  distribution 997
- D.11 Critical values of the Durbin-Watson test statistic 998
- D.12 Critical values of the Kolmogorov-Smirnov test statistic 1000
- D.13 Bias correction factors for estimating  $\sigma$  when the population is normal 1000

- D.14 Lilliefors test statistic for normality 1001
- D.15 Critical values of the Spearman rank correlation test statistic 1001
- D.16 Critical values of the Wilcoxon signed rank test statistic 1002
- D.17 Percentage points of the studentized range 1003

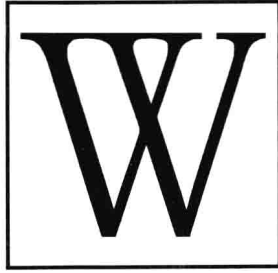
Appendix E Data sets—definitions and descriptions 1005

---

Appendix F Answers to odd-numbered exercises 1011

---

Index 1103



When the first American Indians visited London, England, in the early seventeenth century, they came with instructions to count the men in the city and report their findings back to their tribe. They took sticks along, intending to keep count by their usual method—making a scratch on a stick for every man they saw. Given the number of men in London at that time, the method was hopeless. Besides quickly running out of sticks, they had no idea if they were counting the same men twice. Clearly the methods of recording and analyzing data that are adequate at one time or in one environment can be totally inadequate at another time or in another place.

---



## 1.1 DEFINING STATISTICS

Today the scope of the information required to make sound decisions to resolve problems in finance, marketing, production, and other aspects of business demands the use of modern Statistics. **Statistics**—with a capital “S”—is *the branch of mathematics devoted to the collection, display, and analysis of data to aid in decision making*. We could estimate the population of a large city, for example, by mapping the city’s blocks and sampling houses in blocks selected at random to find the number of residents. We could then project these sample results upward to estimate the population of the entire city. The word “data” (plural of “datum,” something given), which was used in the definition of Statistics, also requires a definition. **Data** consist of *a set of numbers that represent the measurements of a numerical characteristic or characteristics of some group*. The number of men, women, and children in each sampled block are data. Their heights, weights, and incomes are also data. A data set *can also contain terms that describe nonnumerical characteristics associated with some group*. The race of each person interviewed is a nonnumerical characteristic that could be designated by the word “white,” “black,” “Oriental,” or “other.”

The following examples will give you a better idea of why the scope of information needed for sound business decisions requires the use of modern Statistics.

Consider the problem of determining relative prices in competing supermarkets. Fifty years ago a grocery store owner could stroll through a competitor’s store and get a good impression of relative prices in a few minutes. Supermarkets today carry between 15,000 and 25,000 items and in a typical week the prices of



1,000 of them will be changed. Supermarkets do occasionally check the price of every item in their competitors' stores, but this procedure is too expensive and time-consuming to do as frequently as it is needed. A monthly or even weekly pulse of relative prices is essential. Clearly, a sample of items must be taken to represent the average level of prices in a supermarket. The science of Statistics deals with such questions as how the particular items that make up the sample should be selected, how many items need to be in the sample, and how accurately the average price of the sample data will reflect the storewide average level of prices.

Ten years ago the process of brazing or joining the tubes that make up a bicycle frame was done by hand. The alignment of the tubes is critical because a bicycle with a misaligned frame will pull to the right or left when it is ridden. A skilled craftsman would check the alignment after each tube was brazed. The whole process took several hours. Today most bicycle frames are brazed by robots and the process takes a few minutes. When the frames are measured and deviations from a perfect alignment are observed, one of two factors may be responsible: the settings on the brazing robots may have drifted away from their specified values or there was an unavoidable small random variation. The tolerances are so close, in thousandths of an inch, that when a deviation is found, it is uncertain whether the robots have to be reset or if settings are correct and the deviation is unavoidable. Statistical process control is used to track the alignments of each frame as it is completed, and when the trend of measurements from frame to frame is analyzed, it is possible to identify with a high degree of certainty when the settings have drifted even slightly, so that the robots can be stopped and recalibrated.

The marketing of a new toy is much more expensive today than it was a few years ago. Instead of just relying on television commercials to make children aware of a new toy, many toy manufacturers subsidize the creation of a Saturday-morning cartoon show based on the toy. Sometimes even these efforts fail: the kids simply don't like the toy. To reduce the uncertainty inherent in marketing a new toy, manufacturers sample children to gauge their reactions. There are many millions of children in the potential market for a new toy. The science of Statistics is used to estimate the proportion of this market that would purchase the toy on the basis of the reactions of a small sample of children.

In all of these examples, samples were taken from a population and the sample was used to obtain an estimate of a parameter of the population. A **parameter** is *an overall summary measure applied to a population*, such as the average price of all items in a supermarket, the average alignment of all frames made under the current calibration of the brazing robots, or the proportion of all children who will buy the toy. A **population** is *every element in a group that is the subject of analysis*. The three populations in the above examples are the prices of all items in a supermarket, the dimensions of all future bicycle frames brazed by the robot under the current settings, and the decision to buy or not to buy a particular toy by every child in a certain age range. The word "popu-