



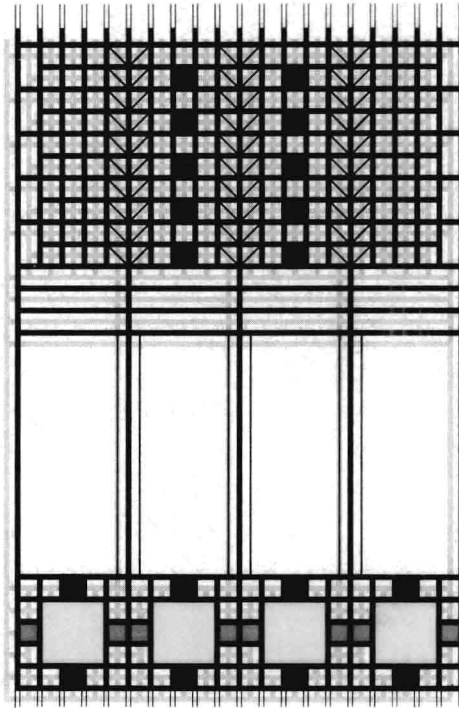
Sheldon M. Ross

Macintosh
high-density
disk enclosed

P/N 843887-X

INTRODUCTORY STATISTICS

INTRODUCTORY STATISTICS



Sheldon M. Ross

University of California, Berkeley



The McGraw-Hill Companies, Inc.
New York St. Louis San Francisco Auckland Bogotá Caracas
Lisbon London Madrid Mexico City Milan Montreal
New Delhi San Juan Singapore Sydney Tokyo Toronto

TO MY FAMILY

This book was set in Times Roman by GTS Graphics.
The editors were Jack Shira, Maggie Lanzillo, and David A. Damstra;
the designer was Robin Hoffmann;
cover photo by Patrick Barta;
the production supervisor was Friederich W. Schulte.
The photo editor was Kathy Bendo;
the photo researcher was Debra P. Hershkowitz.
Von Hoffmann Press, Inc., was printer and binder.

Introductory Statistics

Copyright ©1996 by The McGraw-Hill Companies, 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.

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 VN H VN H 9 0 9 8 7 6 5

P/N 053932-4

Library of Congress Cataloging-in-Publication Data

Ross, Sheldon M.

Introductory statistics / Sheldon M. Ross.

p. cm. — (McGraw-Hill series in probability and statistics)

Includes index.

ISBN 0-07-912244-2 (IBM set). — ISBN 0-07-912245-0 (MAC set)

I. Mathematical statistics. I. Title. II. Series.

QA276.R684 1996

519.5—dc20

95-19670

INTRODUCTORY STATISTICS

McGRAW-HILL SERIES IN PROBABILITY AND STATISTICS

Brillinger: Time Series: Data Analysis
and Theory
De Groot: Optimal Statistical Decisions
Dixon and Massey: Introduction to Sta-
tistical Analysis
Milton and Arnold: Introduction to Prob-
ability and Statistics: Principles and

Applications for Engineering and the
Computing Sciences
Milton: Statistical Methods in the Bio-
logical and Health Sciences
Mood, Graybill, and Boes: Introduction
to the Theory of Statistics
Morrison: Multivariate Statistical

Methods
Ross: Introductory Statistics
Snell: Introduction to Probability
Steel and Torrie: Principles and Proce-
dures of Statistics

Also available from McGraw-Hill

SCHAUM'S OUTLINE SERIES IN MATHEMATICS & STATISTICS

Most outlines include basic theory, definitions and hundreds of example problems solved in step-by-step detail, and supplementary problems with answers.

Related titles on the current list include:

Advanced Calculus
Advanced Mathematics for Engineers &
Scientists
Analytic Geometry
Basic Mathematics for Electricity &
Electronics
Basic Mathematics with Applications to
Science & Technology
Beginning Calculus
Boolean Algebra & Switching Circuits
Calculus
Calculus for Business, Economics, & the
Social Sciences
College Algebra
College Mathematics
Combinatorics
Complex Variables

Descriptive Geometry
Differential Equations
Differential Geometry
Discrete Mathematics
Elementary Algebra
Finite Differences & Difference
Equations
Finite Element Analysis
Finite Mathematics
Fourier Analysis
General Topology
Geometry
Group Theory
Laplace Transforms
Linear Algebra
Mathematical Handbook of Formulas &
Tables

Mathematical Methods for Business &
Economics
Mathematics for Nurses
Matrix Operations
Modern Abstract Algebra
Numerical Analysis
Partial Differential Equations
Probability
Probability & Statistics
Real Variables
Review of Elementary Mathematics
Set Theory & Related Topics
Statistics
Technical Mathematics
Tensor Calculus
Trigonometry
Vector Analysis

SCHAUM'S SOLVED PROBLEMS SERIES

Each title in this series is a complete and expert source of solved problems with solutions worked out in step-by-step detail.

Titles on the current list include:

3000 Solved Problems in Calculus
2500 Solved Problems in College
Algebra and Trigonometry

2500 Solved Problems in Differential
Equations
2000 Solved Problems in Discrete
Mathematics

3000 Solved Problems in Linear Algebra
2000 Solved Problems in Numerical
Analysis
3000 Solved Problems in Precalculus

Available at most college bookstores, or for a complete list of titles and prices, write to:

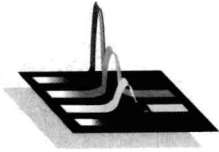
Schaum Division
McGraw-Hill, Inc.
1221 Avenue of the Americas
New York, NY 10020

ABOUT THE AUTHOR

Sheldon M. Ross is a professor in the Department of Industrial Engineering and Operations Research at the University of California at Berkeley. He received his Ph.D. in statistics at Stanford University in 1968 and has been at Berkeley ever since. He has published many technical articles and textbooks in the areas of statistics and applied probability. Among his texts are *A First Course in Probability* (4th edition), *Introduction to Probability Models* (5th edition), and *Introduction to Probability and Statistics for Engineers and Scientists*.

Professor Ross is the founding and continuing editor of the journal *Probability in the Engineering and Informational Sciences*. He is a fellow of the Institute of Mathematical Statistics, and a recipient of the Humboldt U.S. Senior Scientist award.

P R E F A C E



*Statistical thinking will one day be as necessary
for efficient citizenship as the ability to read and write.*
H. G. Wells (1866–1946)

In today's complicated world, very few issues are clear cut and without controversy. In order to understand and form an opinion about an issue, one must usually gather information, or data. To learn from data, one must know something about statistics, which is the art of learning from data.

This introductory statistics text is written for college level students in any field of study. It can be used in a quarter, semester, or full year course. Its only prerequisite is high school algebra. Our goal in writing it is to present statistical concepts and techniques in a manner that will teach students not only how and when to utilize the statistical procedures developed, but also to understand why these procedures should be used. As a result we have made a great effort to explain the ideas behind the statistical concepts and techniques presented. Concepts are motivated, illustrated, and explained in a way that attempts to increase one's intuition. It is only when a student develops a feel or intuition for statistics that she or he is really on the path toward making sense of data.

To illustrate the diverse applications of statistics, and to offer students different perspectives about the use of statistics, we have provided a wide variety of text examples and problems to be worked by students. Most refer to real world issues, such as gun control, stock price models, health issues, driving age limits, school admission ages, public policy issues, gender issues, use of helmets, sports, disputed authorship, scientific fraud, Vitamin C, and many others. Many of them use data that are not only real but are themselves of interest. The exercises have been posed in a clear and concise manner, and include many thought provok-

ing problems that emphasize thinking and problem solving skills. In addition, some of the problems are designed to be open ended and can be used as starting points for term projects.

SOME SPECIAL FEATURES OF THE TEXT

Introduction The first numbered section of each chapter is an introduction which poses a realistic statistical situation to help students gain perspective on what they will encounter in the chapter.

Statistics in Perspective Statistics in Perspective highlights are placed throughout the book to illustrate real-world application of statistical techniques and concepts. These perspectives are designed to help students analyze and interpret data, as well as to utilize proper statistical techniques and methodology.

Real Data Throughout the text discussions, examples, perspective highlights, and exercises, real data sets are used to enhance the students' understanding of the material. These data sets provide information for the study of current issues in a variety of disciplines, such as health, medicine, sports, business, and education.

Historical Perspectives These enrichment sections profile prominent statisticians and historical events, giving students an understanding of how the discipline of statistics has evolved.

Problems/Review Problems This text includes hundreds of exercises which are placed at the end of each section within a chapter, as well as more comprehensive review exercises at the end of each chapter. Many of these problems utilize real data and are designed to assess the students' conceptual as well as computational understanding of the material. Selected problems are open-ended and offer excellent opportunity for extended discussion, group activities or student projects.

Summary/Key Terms An end-of-chapter summary provides a detailed review of important concepts and formulas covered in the chapter. Key terms and their definitions are listed which serve as a working glossary within each chapter.

Formula Card Important tables and formulas which students refer to and utilize often are included in a detachable table/formula card. This card can be used as a quick reference when doing homework or studying for an exam.

Minitab Laboratories/Computer Applications Extended laboratory sessions for Minitab are provided at the end of most chapters. These laboratories, written by Lloyd R. Jaisingh of Morehead State University, include detailed instructions and extended examples for using Minitab and allow students to explore the topics covered

in the corresponding chapter. In addition, exercises are placed at the end of each laboratory which can be used with Minitab or any other statistical software package.

Program Disk A disk is provided within each text which includes programs that can be used to solve basic statistical computation problems. Please refer to Appendix E for a listing of these programs.

SUPPLEMENTS

Instructor's Resource Manual The Instructor's Resource Manual contains detailed solutions to the even-numbered end-of-section and end-of-chapter problems and a data disk that contains data sets from the text and exercise material.

Print Test Bank/Computerized Test Bank The Print Test Bank provides instructors with over 1,000 test questions to choose from. A computerized version of this print test bank is also available in both IBM and Macintosh formats.

Student Solutions Manual The Student Solutions Manual contains detailed solutions to the odd-numbered end-of-section and end-of-chapter exercises.

McGraw-Hill's Statistics Discovery Series: A Guide to Learning Statistics This supplement is intended to help students enhance their understanding of introductory statistics. Each section of this study guide contains study objectives, an overview of the topics covered, key terms and definitions, worked-out examples, helpful hints to the student, and new exercises and their solutions.

McGraw-Hill's Statistics Discovery Series: A Guide to Minitab This supplement helps the student gain a better understanding of statistics through the use of the statistics software Minitab. Worked-out examples and new exercises for use with Minitab are presented, along with a data disk containing data sets ready for use with Minitab. The supplement contains command information for DOS, Windows, and Macintosh platforms, and is packaged with either an IBM or a Macintosh disk.

McGraw-Hill's Statistics Discovery Series: A Guide to TI Graphing Calculators for Statistics This supplement contains instructions for the student using a graphic calculator in an introductory statistics course through worked-out examples and new exercises. Appendixes for Texas Instruments 81, 82, and 85 graphic calculators are also included.

Against All Odds Videotapes Videotapes depicting the use of statistics in our world, produced by the Annenberg/CPB Project, are available.

Mystat This student version of the statistical software program Systat is available through McGraw-Hill.

THE TEXT

In chapter 1 we introduce the subject matter of statistics and present its two branches. The first of these, called descriptive statistics, is concerned with the collection, description, and summarization of data. The second branch, called inferential statistics, deals with the drawing of conclusions from data.

Chapters 2 and 3 are concerned with descriptive statistics. In Chapter 2 we discuss tabular and graphical methods of presenting a set of data. We see that an effective presentation of a data set can often reveal certain of its essential features. Chapter 3 shows how to summarize certain of the features of a data set.

In order to be able to draw conclusions from data it is necessary to have some understanding of what they represent. For instance, it is often assumed that the data constitute a “random sample from some population.” In order to understand exactly what this and similar phrases signify, it is necessary to have some understanding of probability, and that is the subject of Chapter 4. The study of probability is often a troublesome issue in an introductory statistics class because many students find it a difficult subject. As a result, certain textbooks have chosen to downplay its importance and present it in a rather cursory style. We have chosen a different approach, and have attempted to concentrate on its essential features, and present them in a clear and easily understood manner. Thus, while avoiding the topic of counting principles, we have briefly but carefully dealt with the concepts of the events of an experiment, the properties of the probabilities that are assigned to the events, and the idea of conditional probability and independence. Our study of probability is continued in chapter 5 where discrete random variables are introduced and in chapter 6, which deals with the normal and other continuous random variables.

Chapter 7 is concerned with the probability distributions of sampling statistics. In this chapter we learn why the normal distribution is of such importance in statistics.

Chapter 8 deals with the problem of using data to estimate certain parameters of interest. For instance, we might want to estimate the proportion of people that are presently in favor of congressional term limits. Two types of estimators are studied. The first of these estimates the quantity of interest with a single number (for instance, it might estimate that 52 percent of the voting population favors term limits). The second type provides an estimator in the form of an interval of values (for instance, it might estimate that between 49 and 55 percent of the voting population favors term limits).

Chapter 9 introduces the important topic of statistical hypothesis testing, which is concerned with using data to test the plausibility of a specified hypothesis. For instance, such a test might reject the hypothesis that over 60 percent of the voting population favors term limits. The concept of p -value, which measures the degree of plausibility of the hypothesis after the data has been observed, is introduced.

Whereas the tests in chapter 9 deal with a single population, the ones in chapter 10 relate to two separate populations. For instance, we might be interested in testing whether the proportions of men and of women that favor term limits are the same.

Probably the most widely used statistical inference technique is that of the analysis of variance, and this is introduced in chapter 11. This technique allows us to test infer-

ences about parameters that are affected by many different factors. Both one- and two-factor analysis of variance problems are considered in this chapter.

In Chapter 12 we learn about linear regression and how it can be used to relate the value of one variable (say the height of a man) to that of another (the height of his father). The concept of regression to the mean is discussed and the regression fallacy is introduced and carefully explained. We also learn about the relation between regression and correlation. Also, in an optional section, we use regression to the mean along with the central limit theorem to present a simple, original argument to explain why biological data sets often appear to be normally distributed.

In Chapter 13 we present goodness of fit tests, which can be used to test whether a proposed model is consistent with data. This chapter also considers populations classified according to two characteristics and shows how to test whether the characteristics of a randomly chosen member of the population are independent.

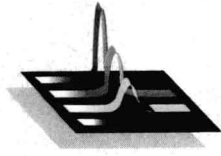
Chapter 14 deals with nonparametric hypothesis tests, which are tests that can be used in situations where the ones of earlier chapters are inappropriate. Chapter 15 introduces the subject matter of quality control, a key statistical technique in manufacturing and production processes.

ACKNOWLEDGMENTS

We would like to thank Margaret Lin, Erol Pekoz, and the following reviewers of this text for their many helpful comments: William H. Beyer, University of Akron; Patricia Buchanan, Pennsylvania State University; Michael Eurgubian, Santa Rosa Junior College; Larry Griffey, Florida Community College, Jacksonville; James E. Holstein, University of Missouri; James Householder, Humboldt State University; Robert Lacher, South Dakota State University; Jacinta Mann, Seton Hill College; C. J. Park, San Diego State University; Ronald Pierce, Eastern Kentucky University; Lawrence Riddle, Agnes Scott College; Gaspard T. Rizzuto, University of Southwestern Louisiana; Jim Robison-Cox, Montana State University; Walter Rosenkrantz, University of Massachusetts, Amherst; Bruce Sisko, Belleville Area College; Glen Swindle, University of California, Santa Barbara; Paul Vetrano, Santa Rose Junior College; Joseph J. Walker, Georgia State University; Deborah White, College of the Redwoods; and Cathleen Zucco, LeMoyne College.

Sheldon M. Ross

C O N T E N T S



Preface

xvii

CHAPTER 1 INTRODUCTION TO STATISTICS

2

1.1 Introduction

4

1.2 The Nature of Statistics

5

1.3 Populations and Samples

8

1.4 A Brief History of Statistics

10

Problems

13

The Changing Definition of Statistics

16

Key Terms

17

Minitab Lab: Introduction to the Minitab Software

18

CHAPTER 2 DESCRIBING DATA SETS

22

2.1 Introduction

24

2.2 Frequency Tables and Graphs

24

2.3 Grouped Data and Histograms

36

2.4 Stem-and-Leaf Plots

48

2.5 Sets of Paired Data

56

2.6 Some Historical Comments	63
<i>Key Terms</i>	63
<i>Summary</i>	64
<i>Review Problems</i>	67
Minitab Lab: Describing Data Sets	75

CHAPTER 3

USING STATISTICS TO SUMMARIZE DATA SETS

84

3.1 Introduction	86
3.2 Sample Mean	87
3.3 Sample Median	97
3.4 Sample Mode	112
3.5 Sample Variance and Sample Standard Deviation	114
3.6 Normal Data Sets and the Empirical Rule	125
3.7 Sample Correlation Coefficient	136
<i>Key Terms</i>	150
<i>Summary</i>	151
<i>Review Problems</i>	153
Minitab Lab: Using Statistics to Summarize Data	157

CHAPTER 4

PROBABILITY

166

4.1 Introduction	168
4.2 Sample Space and Events of an Experiment	168
4.3 Properties of Probability	175
4.4 Experiments Having Equally Likely Outcomes	183
4.5 Conditional Probability and Independence	191
*4.6 Bayes' Theorem	208
<i>Key Terms</i>	213
<i>Summary</i>	214
<i>Review Problems</i>	216

CHAPTER 5

DISCRETE RANDOM VARIABLES

222

5.1 Introduction	224
5.2 Random Variables	224

5.3 Expected Value	231
5.4 Variance of Random Variables	244
5.5 Binomial Random Variables	251
*5.6 Hypergeometric Random Variables	260
<i>Key Terms</i>	262
<i>Summary</i>	263
<i>Review Problems</i>	264
Minitab Lab: Discrete Random Variables	268

CHAPTER 6

NORMAL RANDOM VARIABLES

276

6.1 Introduction	278
6.2 Continuous Random Variables	278
6.3 Normal Random Variables	282
6.4 Probabilities Associated with a Standard Normal Random Variable	287
6.5 Finding Normal Probabilities: Conversion to the Standard Normal	292
6.6 Additive Property of Normal Random Variables	295
6.7 Percentiles of Normal Random Variables	299
<i>Key Terms</i>	305
<i>Summary</i>	306
<i>Review Problems</i>	308
Minitab Lab: Normal Random Variables	311

CHAPTER 7

DISTRIBUTIONS OF SAMPLING STATISTICS

316

7.1 A Preview	318
7.2 Introduction	318
7.3 Sample Mean	319
7.4 Central Limit Theorem	324
7.5 Sampling Proportions from a Finite Population	333
7.6 Distribution of the Sample Variance of a Normal Population	343
<i>Key Terms</i>	345
<i>Summary</i>	346
<i>Review Problems</i>	347
Minitab Lab: Distributions of Sampling Statistics	350

CHAPTER 8	356
ESTIMATION	
8.1 Introduction	358
8.2 Point Estimator of a Population Mean	359
8.3 Point Estimator of a Population Proportion	362
8.4 Estimating a Population Variance	368
8.5 Interval Estimator of the Mean of a Normal Population with Known Population Variance	373
8.6 Interval Estimators of the Mean of a Normal Population with Unknown Population Variance	385
8.7 Interval Estimators of a Population Proportion	396
<i>Key Terms</i>	406
<i>Summary</i>	406
<i>Review Problems</i>	409
 CHAPTER 9	 414
TESTING STATISTICAL HYPOTHESES	
9.1 Introduction	416
9.2 Hypothesis Tests and Significance Levels	416
9.3 Tests Concerning the Mean of a Normal Population: Case of Known Variance	422
9.4 The t Test for the Mean of a Normal Population: Case of Unknown Variance	437
9.5 Hypothesis Tests Concerning Population Proportions	449
<i>Key Terms</i>	459
<i>Summary</i>	459
<i>Review Problems and Proposed Case Studies</i>	463
Minitab Lab: Estimation and Testing Statistical Hypotheses for a Single Population	467
 CHAPTER 10	 474
HYPOTHESIS TESTS CONCERNING TWO POPULATIONS	
10.1 Introduction	476
10.2 Testing Equality of Means of Two Normal Populations: Case of Known Variances	477
10.3 Testing Equality of Means: Unknown Variances and Large Sample Sizes	484

10.4 Testing Equality of Means: Small-Sample Tests When the Unknown Population Variances Are Equal	493
10.5 Paired-Sample t Test	502
10.6 Testing Equality of Population Proportions	511
<i>Key Terms</i>	521
<i>Summary</i>	521
<i>Review Problems</i>	526
Minitab Lab: Hypothesis Tests Concerning Two Populations	530

CHAPTER 11

ANALYSIS OF VARIANCE

534

11.1 Introduction	536
11.2 One-Factor Analysis of Variance	537
11.3 Two-Factor Analysis of Variance: Introduction and Parameter Estimation	545
11.4 Two-Factor Analysis of Variance: Testing Hypotheses	551
11.5 Final Comments	561
<i>Key Terms</i>	561
<i>Summary</i>	562
<i>Review Problems</i>	565
Minitab Lab: Analysis of Variance	568

CHAPTER 12

LINEAR REGRESSION

574

12.1 Introduction	576
12.2 Simple Linear Regression Model	577
12.3 Estimating the Regression Parameters	581
12.4 Error Random Variable	591
12.5 Testing the Hypothesis that $\beta = 0$	594
12.6 Regression to the Mean	601
12.7 Prediction Intervals for Future Responses	610
12.8 Coefficient of Determination	615
12.9 Sample Correlation Coefficient	619
12.10 Multiple Linear Regression Model	621
<i>Key Terms</i>	627
<i>Summary</i>	627
<i>Review Problems</i>	631
Minitab Lab: Linear Regression	636

CHAPTER 13	
CHI-SQUARED GOODNESS OF FIT TESTS	642
13.1 Introduction	644
13.2 Chi-Squared Goodness of Fit Tests	646
13.3 Testing for Independence in Populations Classified According to Two Characteristics	657
13.4 Testing for Independence in Contingency Tables with Fixed Marginal Totals	667
<i>Key Terms</i>	673
<i>Summary</i>	674
<i>Review Problems</i>	676
Minitab Lab: Chi-Square Goodness of Fit Tests	682
CHAPTER 14	
NONPARAMETRIC HYPOTHESES TESTS	686
14.1 Introduction	688
14.2 Sign Test	688
14.3 Signed-Rank Test	697
14.4 Rank-Sum Test for Comparing Two Populations	706
14.5 Runs Test for Randomness	715
<i>Key Terms</i>	722
<i>Summary</i>	722
<i>Review Problems</i>	725
Minitab Lab: Nonparametric Hypotheses Tests	727
CHAPTER 15	
QUALITY CONTROL	734
15.1 Introduction	736
15.2 The \bar{X} Control Chart for Detecting a Shift in the Mean	736
15.3 Control Charts for Fraction Defective	742
<i>Key Terms</i>	745
<i>Summary</i>	745
<i>Review Problems</i>	745
Minitab Lab: Quality Control	747