

Introductory Statistics

Introductory Statistics

Third Edition

Sheldon M. Ross

University of Southern California



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
Elsevier, The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

Copyright © 2010, Elsevier Inc. All rights reserved.

No part of this publication 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. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

Ross, Sheldon M.

Introductory statistics / Sheldon M. Ross. – 3rd ed.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-12-374388-6 (hardcover : alk. paper) 1. Mathematical statistics. I. Title.

QA276.R684 2010

519.5–dc22

2009050832

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN: 978-0-12-374388-6

For information on all Academic Press publications
visit our Web site at www.elsevierdirect.com

Typeset by: diacriTech, India

Printed in Canada

09 10 11 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

About the Author

Sheldon M. Ross

Sheldon M. Ross received his Ph.D. in Statistics at Stanford University in 1968 and then joined the Department of Industrial Engineering and Operations Research at the University of California at Berkeley. He remained at Berkeley until Fall 2004, when he became the Daniel J. Epstein Professor of Industrial and Systems Engineering in the Daniel J. Epstein Department of Industrial and Systems Engineering at the University of Southern California. 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* (eighth edition), *Introduction to Probability Models* (tenth edition), *Simulation* (fourth edition), and *Introduction to Probability and Statistics for Engineers and Scientists* (fourth edition).

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.

For Rebecca and Elise

Contents

ABOUT THE AUTHOR.....	v
PREFACE.....	xvii
ACKNOWLEDGMENTS	xxi
CHAPTER 1 Introduction to Statistics	1
1.1 Introduction	1
1.2 The Nature of Statistics	3
1.2.1 Data Collection.....	3
1.2.2 Inferential Statistics and Probability Models ...	4
1.3 Populations and Samples	5
*1.3.1 Stratified Random Sampling.....	6
1.4 A Brief History of Statistics	7
Key Terms.....	10
The Changing Definition of Statistics.....	11
Review Problems	11
CHAPTER 2 Describing Data Sets.....	17
2.1 Introduction	18
2.2 Frequency Tables and Graphs.....	18
2.2.1 Line Graphs, Bar Graphs, and Frequency Polygons.....	19
2.2.2 Relative Frequency Graphs.....	21
2.2.3 Pie Charts.....	24
Problems	25
2.3 Grouped Data and Histograms	32
Problems	39
2.4 Stem-and-Leaf Plots	44
Problems	47

* The asterisk signifies optional material not used in the sequel.

2.5	Sets of Paired Data	51
	Problems	54
2.6	Some Historical Comments	58
	Key Terms	59
	Summary	60
	Review Problems	63
CHAPTER 3	Using Statistics to Summarize Data Sets	71
3.1	Introduction	72
3.2	Sample Mean	73
	3.2.1 Deviations	78
	Problems	79
3.3	Sample Median	83
	Problems	86
	3.3.1 Sample Percentiles	90
3.4	Sample Mode	97
	Problems	98
3.5	Sample Variance and Sample Standard Deviation	99
	Problems	105
3.6	Normal Data Sets and the Empirical Rule	109
	Problems	114
3.7	Sample Correlation Coefficient	120
	Problems	128
	Key Terms	134
	Summary	136
	Review Problems	138
CHAPTER 4	Probability	145
4.1	Introduction	146
4.2	Sample Space and Events of an Experiment	146
	Problems	150
4.3	Properties of Probability	153
	Problems	156
4.4	Experiments Having Equally Likely Outcomes	161
	Problems	164
4.5	Conditional Probability and Independence	167
	Problems	177
*4.6	Bayes' Theorem	185
	Problems	187

*4.7	Counting Principles	189
	Problems	195
	Key Terms	198
	Summary	200
	Review Problems	201

CHAPTER 5	Discrete Random Variables	209
	5.1 Introduction	210
	5.2 Random Variables	211
	Problems	215
	5.3 Expected Value	218
	5.3.1 Properties of Expected Values	221
	Problems	225
	5.4 Variance of Random Variables	231
	5.4.1 Properties of Variances	233
	Problems	236
	5.5 Binomial Random Variables	238
	5.5.1 Expected Value and Variance of a Binomial Random Variable	243
	Problems	244
	*5.6 Hypergeometric Random Variables	248
	Problems	249
	*5.7 Poisson Random Variables	250
	Problems	253
	Key Terms	254
	Summary	254
	Review Problems	256
CHAPTER 6	Normal Random Variables	261
	6.1 Introduction	262
	6.2 Continuous Random Variables	262
	Problems	264
	6.3 Normal Random Variables	266
	Problems	269
	6.4 Probabilities Associated with a Standard Normal Random Variable	271
	Problems	276
	6.5 Finding Normal Probabilities: Conversion to the Standard Normal	277

6.6	Additive Property of Normal Random Variables	279
	Problems	281
6.7	Percentiles of Normal Random Variables	284
	Problems	289
	Key Terms.....	290
	Summary	290
	Review Problems	293
CHAPTER 7	Distributions of Sampling Statistics	297
7.1	A Preview.....	298
7.2	Introduction.....	298
7.3	Sample Mean	299
	Problems	303
7.4	Central Limit Theorem	304
7.4.1	Distribution of the Sample Mean.....	306
7.4.2	How Large a Sample Is Needed?	310
	Problems	311
7.5	Sampling Proportions from a Finite Population	313
7.5.1	Probabilities Associated with Sample Proportions: The Normal Approximation to the Binomial Distribution.....	317
	Problems	319
7.6	Distribution of the Sample Variance of a Normal Population	323
	Problems	325
	Key Terms.....	325
	Summary	326
	Review Problems	327
CHAPTER 8	Estimation	331
8.1	Introduction.....	332
8.2	Point Estimator of a Population Mean.....	333
	Problems	334
8.3	Point Estimator of a Population Proportion.....	336
	Problems	338
*8.3.1	Estimating the Probability of a Sensitive Event	341
	Problems	342
8.4	Estimating a Population Variance	342
	Problems	344

- 8.5 Interval Estimators of the Mean of a Normal Population with Known Population Variance 347
 - 8.5.1 Lower and Upper Confidence Bounds 355
 - Problems 357
- 8.6 Interval Estimators of the Mean of a Normal Population with Unknown Population Variance 359
 - 8.6.1 Lower and Upper Confidence Bounds 364
 - Problems 366
- 8.7 Interval Estimators of a Population Proportion 371
 - 8.7.1 Length of the Confidence Interval 373
 - 8.7.2 Lower and Upper Confidence Bounds 375
 - Problems 377
- Key Terms 380
- Summary 381
- Review Problems 383

CHAPTER 9 Testing Statistical Hypotheses 387

- 9.1 Introduction 388
- 9.2 Hypothesis Tests and Significance Levels 388
 - Problems 392
- 9.3 Tests Concerning the Mean of a Normal Population: Case of Known Variance 394
 - Problems 400
 - 9.3.1 One-Sided Tests 403
- 9.4 The t Test for the Mean of a Normal Population: Case of Unknown Variance 409
 - Problems 417
- 9.5 Hypothesis Tests Concerning Population Proportions 421
 - 9.5.1 Two-Sided Tests of p 425
 - Problems 429
- Key Terms 433
- Summary 433
- Review Problems and Proposed Case Studies 437

CHAPTER 10 Hypothesis Tests Concerning Two Populations 443

- 10.1 Introduction 444
- 10.2 Testing Equality of Means of Two Normal Populations: Case of Known Variances 446
 - Problems 450

10.3	Testing Equality of Means: Unknown Variances and Large Sample Sizes	453
	Problems	459
10.4	Testing Equality of Means: Small-Sample Tests when the Unknown Population Variances Are Equal...	463
	Problems	468
10.5	Paired-Sample t Test	471
	Problems	476
10.6	Testing Equality of Population Proportions	481
	Problems	490
	Key Terms	493
	Summary	493
	Review Problems	498
CHAPTER 11	Analysis of Variance	503
11.1	Introduction	504
11.2	One-Factor Analysis of Variance	505
	A Remark on the Degrees of Freedom	507
	Problems	510
11.3	Two-Factor Analysis of Variance: Introduction and Parameter Estimation	514
	Problems	518
11.4	Two-Factor Analysis of Variance: Testing Hypotheses	520
	Problems	527
11.5	Final Comments	529
	Key Terms	530
	Summary	530
	Review Problems	533
CHAPTER 12	Linear Regression	537
12.1	Introduction	539
12.2	Simple Linear Regression Model	540
	Problems	542
12.3	Estimating the Regression Parameters	544
	Problems	548
12.4	Error Random Variable	553
	Problems	556
12.5	Testing the Hypothesis that $\beta = 0$	557
	Problems	560

12.6	Regression to the Mean	564
*12.6.1	Why Biological Data Sets Are Often Normally Distributed	569
	Problems	570
12.7	Prediction Intervals for Future Responses	573
	Problems	575
12.8	Coefficient of Determination	578
	Problems	580
12.9	Sample Correlation Coefficient	582
	Problems	583
12.10	Analysis of Residuals: Assessing the Model	584
	Problems	586
12.11	Multiple Linear Regression Model	586
12.11.1	Dummy Variables for Categorical Data	590
	Problems	592
	Key Terms	595
	Summary	595
	Review Problems	599
CHAPTER 13	Chi-Squared Goodness-of-Fit Tests	605
13.1	Introduction	606
13.2	Chi-Squared Goodness-of-Fit Tests	609
	Problems	615
13.3	Testing for Independence in Populations Classified According to Two Characteristics	620
	Problems	626
13.4	Testing for Independence in Contingency Tables with Fixed Marginal Totals	631
	Problems	634
	Key Terms	637
	Summary	638
	Review Problems	640
CHAPTER 14	Nonparametric Hypotheses Tests	647
14.1	Introduction	648
14.2	Sign Test	648
14.2.1	Testing the Equality of Population Distributions when Samples Are Paired	652
14.2.2	One-Sided Tests	653
	Problems	655

14.3	Signed-Rank Test	657
14.3.1	Zero Differences and Ties	662
	Problems	664
14.4	Rank-Sum Test for Comparing Two Populations	667
14.4.1	Comparing Nonparametric Tests with Tests that Assume Normal Distributions	672
	Problems	673
14.5	Runs Test for Randomness	676
	Problems	681
14.6	Testing the Equality of Multiple Probability Distributions	683
14.6.1	When the Data Are a Set of Comparison Rankings	685
	Problems	688
14.7	Permutation Tests	689
	Problems	692
	Key Terms	693
	Summary	693
	Review Problems	696
CHAPTER 15	Quality Control	699
15.1	Introduction	700
15.2	The \bar{X} Control Chart for Detecting a Shift in the Mean ..	700
	Problems	705
15.2.1	When the Mean and Variance Are Unknown...	707
15.2.2	S Control Charts	710
	Problems	713
15.3	Control Charts for Fraction Defective	715
	Problems	717
15.4	Exponentially Weighted Moving-Average Control Charts	717
	Problems	721
15.5	Cumulative-Sum Control Charts	722
	Problems	725
	Key Terms	725
	Summary	725
	Review Problems	726

APPENDICES	727
APPENDIX A A Data Set	729
APPENDIX B Mathematical Preliminaries	733
B.1 Summation	733
B.2 Absolute Value	733
B.3 Set Notation	734
APPENDIX C How to Choose a Random Sample	735
APPENDIX D Tables	739
Table D.1 Standard Normal Probabilities	739
Table D.2 Percentiles $t_{n,\alpha}$ of t Distributions	740
Table D.3 Percentiles $\chi_{n,\alpha}^2$ of the Chi-Squared Distributions	741
Table D.4 Percentiles of F Distributions	743
Table D.5 Binomial Distribution Function	749
APPENDIX E Programs	755
ANSWERS TO ODD-NUMBERED PROBLEMS	757
INDEX	807

Preface

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, and Vitamin C, among many others. Many of them use data that not only are real but are themselves of interest. The examples have been posed in a clear and concise manner and include many thought-provoking 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 that 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 while utilizing proper statistical techniques and methodology.

Real Data Throughout the text discussions, examples, perspective highlights, and problems, 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 placed at the end of each section within a chapter, as well as more comprehensive review problems 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 that serve as a working glossary within each chapter.

Formula Summary Important tables and formulas that students often refer to and utilize are included on the inside front and back covers of the book. These can serve as a quick reference when doing homework or studying for an exam.

Program CD-ROM A CD-ROM is provided with each volume that includes programs that can be used to solve basic statistical computation problems. Please refer to Appendix E for a listing of these programs.

THE TEXT

In Chap. 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 Chap. 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. Chap. 3 shows how to summarize certain 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 Chap. 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 attempted to concentrate on its essential features and to present them in a clear and easily understood manner. Thus, we have briefly but carefully dealt with the concept 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 Chap. 5, where discrete random variables are introduced, and in Chap. 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 who 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 have been observed, is introduced.

Whereas the tests in Chap. 9 deal with a single population, the ones in Chap. 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; this is introduced in Chap. 11. This technique allows us to test inferences about parameters that are affected by many different factors. Both one- and two-factor analysis of variance problems are considered in this chapter.

In Chap. 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 Chap. 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.

NEW TO THIS EDITION

The third edition has many new and updated examples and exercises. In addition, there is a new subsection (12.11.1) on the use of dummy variables in multiple regression models. There is also a new section (14.6) on the use of the Kruskal-Wallis nonparametric test of the equality of multiple probability distributions, with a subsection (14.6.1) giving a discussion of the Freedman test which can be used to test this hypothesis when the data are comparison rankings. There is also a new section (14.7) on the class of nonparametric tests known as permutation tests.

Acknowledgments

We would like to thank the following reviewers of the third edition:

Katherine T. Halvorsen, Smith College
Liam O'Brien, Colby College

In addition we wish to thank Margaret Lin, Erol Pekoz, and the following reviewers of the first edition 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