

Statistics in Action

UNDERSTANDING A WORLD OF DATA

Ann E. Watkins
Richard L. Scheaffer
George W. Cobb



Key Curriculum Press
Innovators in Mathematics Education

Project Editor	Cindy Clements
Editors	Anna Werner, Mary Jo Cittadino
Project Administrator	Michael Hyett
Editorial Assistants	Kristin Burke, Carrye De Mers, Heather Dever, Sigi Nacson, Amy Vidali
Mathematics Reviewer	Mary Parker, Austin Community College, Austin, Texas
AP Teacher Reviewers	Angelo DeMattia, Columbia High School, Maplewood, New Jersey Beth Fox-McManus, formerly of Alan C. Pope High School, Marietta, Georgia Dan Johnson, Silver Creek High School, San Jose, California
Multicultural Reviewers	Gil Cuevas, University of Miami, Coral Gables, Florida Genevieve Lau, Skyline College, San Bruno, California Beatrice Lumpkin, Malcolm X College (retired), Chicago, Illinois
Accuracy Checkers	Dudley Brooks, Monica Johnston
Editorial Production Manager	Deborah Cogan
Production Editor	Jacqueline Gamble
Production Director	Diana Jean Parks
Production Coordinator	Charice Silverman
Text Designer	Kathy Cunningham
Compositor	TSI Graphics
Art Editor	Laura Murray
Illustrator	Wendy Wray
Technical Artist	Matt Perry
Technical Art Consultant	Brett Garrett
Photo Researcher	Margee Robinson
Art and Design Coordinators	Caroline Ayres, Kavitha Becker
Cover Designer	Greg Dundis
Cover Photo Credits	<i>top:</i> Strauss-Curtis/Corbis; <i>center left:</i> Hulton Deutsch Collection/Corbis; <i>center right:</i> Tom Nebbia/Corbis; <i>bottom:</i> Bettman/Corbis
Prepress	TSI Graphics
Printer	Von Hoffmann Press
Executive Editor	Casey FitzSimons
Publisher	Steven Rasmussen

©2004 by Key Curriculum Press. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, photocopying, recording, or otherwise, without the prior written permission of the publisher.

®Key Curriculum Press is a registered trademark of Key Curriculum Press. ™Fathom Dynamic Statistics is a trademark of KCP Technologies. All other registered trademarks and trademarks in this book are the property of their respective holders.

Key Curriculum Press
1150 65th Street
Emeryville, CA 94608
editorial@keypress.com
<http://www.keypress.com>

Printed in the United States of America
10 9 8 7 6 5 4 3 08 07 06 05
ISBN: 1-55953-313-7

About the Authors

Ann E. Watkins is Professor of Mathematics at California State University, Northridge. She received her Ph.D. in education from the University of California, Los Angeles. She is a former president of the Mathematical Association of America and a Fellow of the American Statistical Association. Dr. Watkins has served as co-editor of the *College Mathematics Journal*, as a member of the Board of Editors of the *American Mathematical Monthly*, and as Chair of the Advanced Placement Statistics Development Committee. She was selected as the 1994–1995 CSUN Outstanding Professor and won the 1997 CSUN Award for the Advancement of Teaching Effectiveness. Before moving to CSUN in 1990, she taught for the Los Angeles Unified School District and at Pierce College in Los Angeles. In addition to numerous journal articles, she is the co-author of books based on work produced by the Activity-Based Statistics Project (co-authored by Watkins and Scheaffer and now published by Key College Publishing), the Quantitative Literacy Project, and the Core-Plus Mathematics Project.

Richard L. Scheaffer is Professor Emeritus of Statistics at the University of Florida. He received his Ph.D. in statistics from Florida State University. He then joined the faculty of the University of Florida and served as chairman of the Department of Statistics for 12 years. Dr. Scheaffer's research interests are in the areas of sampling and applied probability, especially in their applications to industrial processes. He has published numerous papers and is co-author of four college-level textbooks. In recent years, much of his effort has been directed toward statistics education throughout the school and college curriculum. He was one of the developers of the Quantitative Literacy Project in the United States. As a developer on this project, Dr. Scheaffer helped form the basis of the data analysis emphasis in mathematics curriculum standards recommended by the National Council of Teachers of Mathematics. Dr. Scheaffer also directed the task force that developed the Advanced Placement Statistics Program and served as its first Chief Faculty Consultant. He continues to work on educational projects at the elementary, secondary, and college levels. Dr. Scheaffer is Fellow and past president of the American Statistical Association, from whom he received a Founders Award.

George W. Cobb is the Robert L. Rooke Professor of Statistics at Mt. Holyoke College, where he served a three-year term as Dean of Studies. He received his Ph.D. in statistics from Harvard University. In addition to his fundamental contributions to the emerging science of confectionery ballistics (the statistics of firing gummy bears from a launcher), he is an expert in statistics education with a significant publication record in this field. He chaired the joint committee on undergraduate statistics of the Mathematical Association of America and the American Statistical Association. He also led the STATS project of the Mathematical Association of America, which helped professors of mathematics learn to teach statistics. He is the author of *Introduction to Design and Analysis of Experiments*, published by Key College Publishing. Dr. Cobb served on the National Research Council's Committee on Applied and Theoretical Statistics. Over the past two decades, he has frequently served as an expert witness in lawsuits involving alleged employment discrimination. Dr. Cobb is a Fellow of the American Statistical Association.

Contents

	A Note to Students from the Authors	vii
1	A Case Study of Statistics in Action	2
	1.1 Discrimination in the Workplace: Data Exploration	4
	1.2 Discrimination in the Workplace: Inference	11
2	Exploring Distributions	22
	2.1 The Shapes of Things: Visualizing Distributions	24
	2.2 Graphical Displays for Distributions	38
	2.3 Measures of Center and Spread	53
	2.4 The Normal Distribution	79
3	Relationships Between Two Quantitative Variables	100
	3.1 Scatterplots	102
	3.2 Getting a Line on the Pattern	113
	3.3 Correlation: The Strength of a Linear Trend	135
	3.4 Diagnostics: Looking for Features That the Summaries Miss	154
	3.5 Shape-Changing Transformations	171
4	Sample Surveys and Experiments	208
	4.1 Why Take Samples, and How Not To	211
	4.2 Randomizing: Playing It Safe by Taking Chances	223
	4.3 Experiments and Inference About Cause	234
	4.4 Designing Experiments to Reduce Variability	249
5	Sampling Distributions	266
	5.1 Sampling from a Population	268
	5.2 Generating Sampling Distributions	274
	5.3 Sampling Distribution of the Sample Mean	284
	5.4 Sampling Distribution of the Sample Proportion	300
	5.5 Sampling Distribution of the Sum and Difference	311
6	Probability Models	326
	6.1 Sample Spaces with Equally Likely Outcomes	328
	6.2 The Addition Rule and Disjoint Events	339
	6.3 Conditional Probability	348
	6.4 Independent Events	364
7	Probability Distributions	378
	7.1 Random Variables and Expected Value	380
	7.2 The Binomial Distribution	391
	7.3 The Geometric Distribution	400

8	Inference for Proportions	412
	8.1 Estimating a Proportion with Confidence	414
	8.2 Testing a Proportion	432
	8.3 A Confidence Interval for the Difference of Two Proportions	453
	8.4 A Significance Test for the Difference of Two Proportions	463
9	Inference for Means	478
	9.1 Toward a Confidence Interval for a Mean	480
	9.2 Toward a Significance Test for a Mean	492
	9.3 When You Estimate σ : The t -Distribution	503
	9.4 The Effect of Long Tails and Outliers	517
	9.5 Inference for the Difference Between Two Means	528
	9.6 Paired Comparisons	547
10	Chi-Square Tests	576
	10.1 Testing a Probability Model: The Chi-Square Goodness-of-Fit Test	578
	10.2 The Chi-Square Test of Homogeneity	592
	10.3 The Chi-Square Test of Independence	607
11	Inference for Regression	628
	11.1 Variation in the Estimated Slope	631
	11.2 Making Inferences About Slopes	643
	11.3 Transforming for a Better Fit	658
12	Statistics in Action: Case Studies	678
	12.1 Mum's the Word!	680
	12.2 Baseball: Does Money Buy Success?	683
	12.3 <i>Martin v. Westvaco</i> Revisited: Testing for Possible Employment Discrimination	690
	Appendix: Statistical Tables	699
	Table A: Standard Normal Probabilities	699
	Table B: t -Distribution Critical Values	701
	Table C: Chi-Square Distribution Critical Values	702
	Table D: Random Digits	703
	Glossary	704
	Brief Answers to Selected Problems	712
	Index	743
	Photo Credits	752

A Note to Students from the Authors

Whether you talk about income, prices of goods and services, sports, health, politics, or the weather, data enter the conversation. In fact, in this age of information technology, data come at you at such a rapid rate that you can catch only a glimpse of the masses of numbers. The only way to cope intelligently with this quantitative world and make informed decisions is to gain an understanding of the basic concepts of statistics and practice what you have learned with real data.

What's in This Book

This book is designed for an introductory statistics course—either an introductory college course or its high school equivalent, Advanced Placement Statistics—and includes all of the standard topics for that course. Beginning in Chapter 1 with a court case about age discrimination, you will be immersed in real problems that can be solved only with statistical methods. You will learn to

- explore, summarize, and display data
- design surveys and experiments
- use probability to understand random behavior
- make inferences about populations by looking at samples from those populations
- make inferences about the effect of treatments from designed experiments

How This Book Is Different

Statistical work is more active than it was a generation ago. Computers and graphing calculators have automated the graphical exploration of data, and in the process have made the practice of statistics a more visual enterprise. Statistical techniques are also changing as simulations allow statisticians (and you) to shift the emphasis from following recipes for calculations to paying more attention to statistical concepts. Your instructor has selected this book for you because he or she

- wants you to learn this modern, data-analytic approach to statistics
- encourages you to be an active participant in the classroom
- wants you to see real data (If you have only pretend data, you can only pretend to analyze it.)
- believes that statistical analyses must be tailored to the data
- uses graphing calculators or statistical software for data analysis and for simulations

Throughout this textbook you will see many graphical displays, lots of real data, activities that introduce each major topic, computer printouts, questions for you to discuss with your class, and practice problems so you can be sure you understand the basics before you move on. These features grow out of the vigorous changes that

have been reshaping the practice of statistics and the teaching of statistics over the last quarter century.

The most basic question to ask about any data set is, “Where did the data come from?” Good data for statistical analysis must come from a good plan for data collection. Thus, *Statistics in Action* gives an honest and thorough treatment to the design and analysis of both experiments and surveys.

What You Should Know Before You Start

You will be using this book in an introductory statistics course; thus, you aren’t required to know anything yet about statistics. You may find that your perseverance in trying to understand what you read will contribute more to your success in statistics than your skill with algebra. However, basic topics from algebra, such as the equation for a line, slope, exponential equations, and the idea of a logarithm, will come up throughout the book. Be prepared to review those as you go along, if the need arises.

Acknowledgments

This book is a product of what we have learned from the statisticians and teachers who have been actively involved in helping the introductory statistics course evolve into one that emphasizes activity-based learning of statistical concepts while reflecting modern statistical practice. This book is written in the spirit of the recommendations from the MAA’s STATS project and Focus Group on Statistics, the ASA’s Quantitative Literacy projects, and the College Board’s AP Statistics course. We hope that it adequately reflects the wisdom and experience of those with whom we have worked and who have inspired and taught us.

It has been an awesome experience to work with the Key Curriculum staff and field-test teachers, who always put the interests of students and teachers first. Their commitment to excellence has motivated us to do better than we ever could have done on our own. Steve, Casey, Mary Jo, Anna, Dudley, Bill, and the rest of the staff have been professional and astute throughout. Our deepest gratitude goes to Cindy Clements, our editor, who has been a joy to work with. (Not all authors say that—and mean it—about their editors.) Cindy was an outstanding statistics and calculus teacher before coming to Key. She brings an extraordinary intellectual curiosity and talent for teaching and for statistics to her current position. Her organizational skills, experience in the classroom, and insight have improved every chapter of this text.

Ann Watkins

Dick Scheaffer

George Cobb

Statistics in Action with R. The purpose of this course is to show how statistics may be efficiently used in practice. The course presents both statistical theory and practical analysis on real data sets. The R statistical software and several R packages are used for implementing methods presented in the course and analyzing real data. Topics covered in the current version of the course are: hypothesis testing (single and multiple comparisons). regression models (linear and nonlinear models). Statistics with R, and open source stuff (software, data, community). Menu. Home.Â Until Aug 21, 2013, you can buy the book: R in Action, Second Edition with a 44% discount, using the code: "œmlria2bl". K-means clustering. The most common partitioning method is the K-means cluster analysis. Bayesian statistics in action. Downloaded on 01 June 2012 Published on 01 June 2012 on <http://pubs.rsc.org> | doi:10.1039/C2AY90023H. Analytical Methods. Cite this: DOI: 10.1039/c2ay90023h www.rsc.org/methods. Bayesian statistics in action. Analytical Methods Committee, AMCTB No 52. Received 23rd May 2012 DOI: 10.1039/c2ay90023h.