



Statistical Inference

Second Edition

George Casella
Roger L. Berger

0212
C337
E-2

Statistical Inference

Second Edition

George Casella

University of Florida

Roger L. Berger

North Carolina State University

DUXBURY



THOMSON LEARNING

Australia • Canada • Mexico • Singapore • Spain • United Kingdom • United States



Sponsoring Editor: *Carolyn Crockett*
Marketing Representative: *Tom Ziolkowski*
Editorial Assistant: *Jennifer Jenkins*
Production Editor: *Tom Novack*
Assistant Editor: *Ann Day*
Manuscript Editor: *Carol Reitz*
Permissions Editor: *Sue Ewing*

Cover Design: *Jennifer Mackres*
Interior Illustration: *Lori Heckelman*
Print Buyer: *Vena Dyer*
Typesetting: *Integre Technical Publishing Co.*
Cover Printing: *R. R. Donnelley & Sons Co.,
Crawfordsville*
Printing and Binding: *R. R. Donnelley
& Sons Co., Crawfordsville*

All products used herein are used for identification purposes only and may be trademarks or registered trademarks of their respective owners.

COPYRIGHT © 2002 the Wadsworth Group. Duxbury is an imprint of the Wadsworth Group, a division of Thomson Learning Inc.
Thomson Learning™ is a trademark used herein under license.

For more information about this or any other Duxbury products, contact:

DUXBURY
511 Forest Lodge Road
Pacific Grove, CA 93950 USA
www.duxbury.com
1-800-423-0563 (Thomson Learning Academic Resource Center)

All rights reserved. No part of this work may be reproduced, transcribed or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, Web distribution, or information storage and/or retrieval systems—without the prior written permission of the publisher.

For permission to use material from this work, contact us by
www.thomsonrights.com
fax: 1-800-730-2215
phone: 1-800-730-2214

Printed in United States of America

10 9 8 7 6 5 4 3

Library of Congress Cataloging-in-Publication Data

Casella, George.

Statistical inference / George Casella, Roger L. Berger.—2nd ed.
p. cm.

Includes bibliographical references and indexes.

ISBN 0-534-24312-6

1. Mathematical statistics. 2. Probabilities. I. Berger, Roger L.

II. Title.

QA276.C37 2001

519.5—dc21

2001025794



To Anne and Vicki

Duxbury titles of related interest

Daniel, *Applied Nonparametric Statistics* 2nd

Derr, *Statistical Consulting: A Guide to Effective Communication*

Durrett, *Probability: Theory and Examples* 2nd

Graybill, *Theory and Application of the Linear Model*

Johnson, *Applied Multivariate Methods for Data Analysts*

Kuehl, *Design of Experiments: Statistical Principles of Research Design and Analysis* 2nd

Larsen, Marx, & Cooil, *Statistics for Applied Problem Solving and Decision Making*

Lohr, *Sampling: Design and Analysis*

Lunneborg, *Data Analysis by Resampling: Concepts and Applications*

Minh, *Applied Probability Models*

Minitab Inc., *MINITAB™ Student Version 12 for Windows*

Myers, *Classical and Modern Regression with Applications* 2nd

Newton & Harvill, *StatConcepts: A Visual Tour of Statistical Ideas*

Ramsey & Schafer, *The Statistical Sleuth* 2nd

SAS Institute Inc., *JMP-IN: Statistical Discovery Software*

Savage, *INSIGHT: Business Analysis Software for Microsoft® Excel*

Scheaffer, Mendenhall, & Ott, *Elementary Survey Sampling* 5th

Shapiro, *Modeling the Supply Chain*

Winston, *Simulation Modeling Using @RISK*

To order copies contact your local bookstore or call 1-800-354-9706. For more information contact Duxbury Press at 511 Forest Lodge Road, Pacific Grove, CA 93950, or go to: www.duxbury.com

Preface to the Second Edition

Although Sir Arthur Conan Doyle is responsible for most of the quotes in this book, perhaps the best description of the life of this book can be attributed to the Grateful Dead sentiment, “What a long, strange trip it’s been.”

Plans for the second edition started about six years ago, and for a long time we struggled with questions about what to add and what to delete. Thankfully, as time passed, the answers became clearer as the flow of the discipline of statistics became clearer. We see the trend moving away from elegant proofs of special cases to algorithmic solutions of more complex and practical cases. This does not undermine the importance of mathematics and rigor; indeed, we have found that these have become more important. But the manner in which they are applied is changing.

For those familiar with the first edition, we can summarize the changes succinctly as follows. Discussion of asymptotic methods has been greatly expanded into its own chapter. There is more emphasis on computing and simulation (see Section 5.5 and the computer algebra Appendix); coverage of the more applicable techniques has been expanded or added (for example, bootstrapping, the EM algorithm, p-values, logistic and robust regression); and there are many new Miscellanea and Exercises. We have de-emphasized the more specialized theoretical topics, such as equivariance and decision theory, and have restructured some material in Chapters 3–11 for clarity.

There are two things that we want to note. First, with respect to computer algebra programs, although we believe that they are becoming increasingly valuable tools, we did not want to force them on the instructor who does not share that belief. Thus, the treatment is “unobtrusive” in that it appears only in an appendix, with some hints throughout the book where it may be useful. Second, we have changed the numbering system to one that facilitates finding things. Now theorems, lemmas, examples, and definitions are numbered together; for example, Definition 7.2.4 is followed by Example 7.2.5 and Theorem 10.1.3 precedes Example 10.1.4.

The first four chapters have received only minor changes. We reordered some material (in particular, the inequalities and identities have been split), added some new examples and exercises, and did some general updating. Chapter 5 has also been reordered, with the convergence section being moved further back, and a new section on generating random variables added. The previous coverage of invariance, which was in Chapters 7–9 of the first edition, has been greatly reduced and incorporated into Chapter 6, which otherwise has received only minor editing (mostly the addition of new exercises). Chapter 7 has been expanded and updated, and includes a new section on the EM algorithm. Chapter 8 has also received minor editing and updating, and

has a new section on p-values. In Chapter 9 we now put more emphasis on pivoting (having realized that “guaranteeing an interval” was merely “pivoting the cdf”). Also, the material that was in Chapter 10 of the first edition (decision theory) has been reduced, and small sections on loss function optimality of point estimation, hypothesis testing, and interval estimation have been added to the appropriate chapters.

Chapter 10 is entirely new and attempts to lay out the fundamentals of large sample inference, including the delta method, consistency and asymptotic normality, bootstrapping, robust estimators, score tests, etc. Chapter 11 is classic oneway ANOVA and linear regression (which was covered in two different chapters in the first edition). Unfortunately, coverage of randomized block designs has been eliminated for space reasons. Chapter 12 covers regression with errors-in-variables and contains new material on robust and logistic regression.

After teaching from the first edition for a number of years, we know (approximately) what can be covered in a one-year course. From the second edition, it should be possible to cover the following in one year:

Chapter 1: Sections 1–7	Chapter 6: Sections 1–3
Chapter 2: Sections 1–3	Chapter 7: Sections 1–3
Chapter 3: Sections 1–6	Chapter 8: Sections 1–3
Chapter 4: Sections 1–7	Chapter 9: Sections 1–3
Chapter 5: Sections 1–6	Chapter 10: Sections 1, 3, 4

Classes that begin the course with some probability background can cover more material from the later chapters.

Finally, it is almost impossible to thank all of the people who have contributed in some way to making the second edition a reality (and help us correct the mistakes in the first edition). To all of our students, friends, and colleagues who took the time to send us a note or an e-mail, we thank you. A number of people made key suggestions that led to substantial changes in presentation. Sometimes these suggestions were just short notes or comments, and some were longer reviews. Some were so long ago that their authors may have forgotten, but we haven’t. So thanks to Arthur Cohen, Sir David Cox, Steve Samuels, Rob Strawderman and Tom Wehrly. We also owe much to Jay Beder, who has sent us numerous comments and suggestions over the years and possibly knows the first edition better than we do, and to Michael Perlman and his class, who are sending comments and corrections even as we write this.

This book has seen a number of editors. We thank Alex Kugashev, who in the mid-1990s first suggested doing a second edition, and our editor, Carolyn Crockett, who constantly encouraged us. Perhaps the one person (other than us) who is most responsible for this book is our first editor, John Kimmel, who encouraged, published, and marketed the first edition. Thanks, John.

*George Casella
Roger L. Berger*

Preface to the First Edition

When someone discovers that you are writing a textbook, one (or both) of two questions will be asked. The first is “Why are you writing a book?” and the second is “How is your book different from what’s out there?” The first question is fairly easy to answer. You are writing a book because you are not entirely satisfied with the available texts. The second question is harder to answer. The answer can’t be put in a few sentences so, in order not to bore your audience (who may be asking the question only out of politeness), you try to say something quick and witty. It usually doesn’t work.

The purpose of this book is to build theoretical statistics (as different from mathematical statistics) from the first principles of probability theory. Logical development, proofs, ideas, themes, etc., evolve through statistical arguments. Thus, starting from the basics of probability, we develop the theory of statistical inference using techniques, definitions, and concepts that are statistical and are natural extensions and consequences of previous concepts. When this endeavor was started, we were not sure how well it would work. The final judgment of our success is, of course, left to the reader.

The book is intended for first-year graduate students majoring in statistics or in a field where a statistics concentration is desirable. The prerequisite is one year of calculus. (Some familiarity with matrix manipulations would be useful, but is not essential.) The book can be used for a two-semester, or three-quarter, introductory course in statistics.

The first four chapters cover basics of probability theory and introduce many fundamentals that are later necessary. Chapters 5 and 6 are the first statistical chapters. Chapter 5 is transitional (between probability and statistics) and can be the starting point for a course in statistical theory for students with some probability background. Chapter 6 is somewhat unique, detailing three statistical principles (sufficiency, likelihood, and invariance) and showing how these principles are important in modeling data. Not all instructors will cover this chapter in detail, although we strongly recommend spending some time here. In particular, the likelihood and invariance principles are treated in detail. Along with the sufficiency principle, these principles, and the thinking behind them, are fundamental to total statistical understanding.

Chapters 7–9 represent the central core of statistical inference, estimation (point and interval) and hypothesis testing. A major feature of these chapters is the division into methods of *finding* appropriate statistical techniques and methods of *evaluating* these techniques. Finding and evaluating are of interest to both the theorist and the

practitioner, but we feel that it is important to separate these endeavors. Different concerns are important, and different rules are invoked. Of further interest may be the sections of these chapters titled Other Considerations. Here, we indicate how the rules of statistical inference may be relaxed (as is done every day) and still produce meaningful inferences. Many of the techniques covered in these sections are ones that are used in consulting and are helpful in analyzing and inferring from actual problems.

The final three chapters can be thought of as special topics, although we feel that some familiarity with the material is important in anyone's statistical education. Chapter 10 is a thorough introduction to decision theory and contains the most modern material we could include. Chapter 11 deals with the analysis of variance (oneway and randomized block), building the theory of the complete analysis from the more simple theory of treatment contrasts. Our experience has been that experimenters are most interested in inferences from contrasts, and using principles developed earlier, most tests and intervals can be derived from contrasts. Finally, Chapter 12 treats the theory of regression, dealing first with simple linear regression and then covering regression with "errors in variables." This latter topic is quite important, not only to show its own usefulness and inherent difficulties, but also to illustrate the limitations of inferences from ordinary regression.

As more concrete guidelines for basing a one-year course on this book, we offer the following suggestions. There can be two distinct types of courses taught from this book. One kind we might label "more mathematical," being a course appropriate for students majoring in statistics and having a solid mathematics background (at least $1\frac{1}{2}$ years of calculus, some matrix algebra, and perhaps a real analysis course). For such students we recommend covering Chapters 1–9 in their entirety (which should take approximately 22 weeks) and spend the remaining time customizing the course with selected topics from Chapters 10–12. Once the first nine chapters are covered, the material in each of the last three chapters is self-contained, and can be covered in any order.

Another type of course is "more practical." Such a course may also be a first course for mathematically sophisticated students, but is aimed at students with one year of calculus who may not be majoring in statistics. It stresses the more practical uses of statistical theory, being more concerned with understanding basic statistical concepts and deriving reasonable statistical procedures for a variety of situations, and less concerned with formal optimality investigations. Such a course will necessarily omit a certain amount of material, but the following list of sections can be covered in a one-year course:

Chapter	Sections
1	All
2	2.1, 2.2, 2.3
3	3.1, 3.2
4	4.1, 4.2, 4.3, 4.5
5	5.1, 5.2, 5.3.1, 5.4
6	6.1.1, 6.2.1
7	7.1, 7.2.1, 7.2.2, 7.2.3, 7.3.1, 7.3.3, 7.4
8	8.1, 8.2.1, 8.2.3, 8.2.4, 8.3.1, 8.3.2, 8.4

9	9.1, 9.2.1, 9.2.2, 9.2.4, 9.3.1, 9.4
11	11.1, 11.2
12	12.1, 12.2

If time permits, there can be some discussion (with little emphasis on details) of the material in Sections 4.4, 5.5, and 6.1.2, 6.1.3, 6.1.4. The material in Sections 11.3 and 12.3 may also be considered.

The exercises have been gathered from many sources and are quite plentiful. We feel that, perhaps, the only way to master this material is through practice, and thus we have included much opportunity to do so. The exercises are as varied as we could make them, and many of them illustrate points that are either new or complementary to the material in the text. Some exercises are even taken from research papers. (It makes you feel old when you can include exercises based on papers that were new research during your own student days!) Although the exercises are not subdivided like the chapters, their ordering roughly follows that of the chapter. (Subdivisions often give too many hints.) Furthermore, the exercises become (again, roughly) more challenging as their numbers become higher.

As this is an introductory book with a relatively broad scope, the topics are not covered in great depth. However, we felt some obligation to guide the reader one step further in the topics that may be of interest. Thus, we have included many references, pointing to the path to deeper understanding of any particular topic. (The *Encyclopedia of Statistical Sciences*, edited by Kotz, Johnson, and Read, provides a fine introduction to many topics.)

To write this book, we have drawn on both our past teachings and current work. We have also drawn on many people, to whom we are extremely grateful. We thank our colleagues at Cornell, North Carolina State, and Purdue—in particular, Jim Berger, Larry Brown, Sir David Cox, Ziding Feng, Janet Johnson, Leon Gleser, Costas Goutis, Dave Lansky, George McCabe, Chuck McCulloch, Myra Samuels, Steve Schwager, and Shayle Searle, who have given their time and expertise in reading parts of this manuscript, offered assistance, and taken part in many conversations leading to constructive suggestions. We also thank Shanti Gupta for his hospitality, and the library at Purdue, which was essential. We are grateful for the detailed reading and helpful suggestions of Shayle Searle and of our reviewers, both anonymous and non-anonymous (Jim Albert, Dan Coster, and Tom Wehrly). We also thank David Moore and George McCabe for allowing us to use their tables, and Steve Hirdt for supplying us with data. Since this book was written by two people who, for most of the time, were at least 600 miles apart, we lastly thank Bitnet for making this entire thing possible.

George Casella
Roger L. Berger

*“We have got to the deductions and the inferences,” said Lestrade, winking at me.
“I find it hard enough to tackle facts, Holmes, without flying away
after theories and fancies.”*

Inspector Lestrade to Sherlock Holmes

The Boscombe Valley Mystery

Contents

1	Probability Theory	1
1.1	Set Theory	1
1.2	Basics of Probability Theory	5
1.2.1	Axiomatic Foundations	5
1.2.2	The Calculus of Probabilities	9
1.2.3	Counting	13
1.2.4	Enumerating Outcomes	16
1.3	Conditional Probability and Independence	20
1.4	Random Variables	27
1.5	Distribution Functions	29
1.6	Density and Mass Functions	34
1.7	Exercises	37
1.8	Miscellanea	44
2	Transformations and Expectations	47
2.1	Distributions of Functions of a Random Variable	47
2.2	Expected Values	55
2.3	Moments and Moment Generating Functions	59
2.4	Differentiating Under an Integral Sign	68
2.5	Exercises	76
2.6	Miscellanea	82
3	Common Families of Distributions	85
3.1	Introduction	85
3.2	Discrete Distributions	85
3.3	Continuous Distributions	98
3.4	Exponential Families	111
3.5	Location and Scale Families	116

3.6	Inequalities and Identities	121
3.6.1	Probability Inequalities	122
3.6.2	Identities	123
3.7	Exercises	127
3.8	Miscellanea	135
4	Multiple Random Variables	139
4.1	Joint and Marginal Distributions	139
4.2	Conditional Distributions and Independence	147
4.3	Bivariate Transformations	156
4.4	Hierarchical Models and Mixture Distributions	162
4.5	Covariance and Correlation	169
4.6	Multivariate Distributions	177
4.7	Inequalities	186
4.7.1	Numerical Inequalities	186
4.7.2	Functional Inequalities	189
4.8	Exercises	192
4.9	Miscellanea	203
5	Properties of a Random Sample	207
5.1	Basic Concepts of Random Samples	207
5.2	Sums of Random Variables from a Random Sample	211
5.3	Sampling from the Normal Distribution	218
5.3.1	Properties of the Sample Mean and Variance	218
5.3.2	The Derived Distributions: Student's t and Snedecor's F	222
5.4	Order Statistics	226
5.5	Convergence Concepts	232
5.5.1	Convergence in Probability	232
5.5.2	Almost Sure Convergence	234
5.5.3	Convergence in Distribution	235
5.5.4	The Delta Method	240
5.6	Generating a Random Sample	245
5.6.1	Direct Methods	247
5.6.2	Indirect Methods	251
5.6.3	The Accept/Reject Algorithm	253
5.7	Exercises	255
5.8	Miscellanea	267
6	Principles of Data Reduction	271
6.1	Introduction	271
6.2	The Sufficiency Principle	272
6.2.1	Sufficient Statistics	272
6.2.2	Minimal Sufficient Statistics	279
6.2.3	Ancillary Statistics	282
6.2.4	Sufficient, Ancillary, and Complete Statistics	284

6.3	The Likelihood Principle	290
6.3.1	The Likelihood Function	290
6.3.2	The Formal Likelihood Principle	292
6.4	The Equivariance Principle	296
6.5	Exercises	300
6.6	Miscellanea	307
7	Point Estimation	311
7.1	Introduction	311
7.2	Methods of Finding Estimators	312
7.2.1	Method of Moments	312
7.2.2	Maximum Likelihood Estimators	315
7.2.3	Bayes Estimators	324
7.2.4	The EM Algorithm	326
7.3	Methods of Evaluating Estimators	330
7.3.1	Mean Squared Error	330
7.3.2	Best Unbiased Estimators	334
7.3.3	Sufficiency and Unbiasedness	342
7.3.4	Loss Function Optimality	348
7.4	Exercises	355
7.5	Miscellanea	367
8	Hypothesis Testing	373
8.1	Introduction	373
8.2	Methods of Finding Tests	374
8.2.1	Likelihood Ratio Tests	374
8.2.2	Bayesian Tests	379
8.2.3	Union–Intersection and Intersection–Union Tests	380
8.3	Methods of Evaluating Tests	382
8.3.1	Error Probabilities and the Power Function	382
8.3.2	Most Powerful Tests	387
8.3.3	Sizes of Union–Intersection and Intersection–Union Tests	394
8.3.4	p-Values	397
8.3.5	Loss Function Optimality	400
8.4	Exercises	402
8.5	Miscellanea	413
9	Interval Estimation	417
9.1	Introduction	417
9.2	Methods of Finding Interval Estimators	420
9.2.1	Inverting a Test Statistic	420
9.2.2	Pivotal Quantities	427
9.2.3	Pivoting the CDF	430
9.2.4	Bayesian Intervals	435

9.3	Methods of Evaluating Interval Estimators	440
9.3.1	Size and Coverage Probability	440
9.3.2	Test-Related Optimality	444
9.3.3	Bayesian Optimality	447
9.3.4	Loss Function Optimality	449
9.4	Exercises	451
9.5	Miscellanea	463
10	Asymptotic Evaluations	467
10.1	Point Estimation	467
10.1.1	Consistency	467
10.1.2	Efficiency	470
10.1.3	Calculations and Comparisons	473
10.1.4	Bootstrap Standard Errors	478
10.2	Robustness	481
10.2.1	The Mean and the Median	482
10.2.2	M-Estimators	484
10.3	Hypothesis Testing	488
10.3.1	Asymptotic Distribution of LRTs	488
10.3.2	Other Large-Sample Tests	492
10.4	Interval Estimation	496
10.4.1	Approximate Maximum Likelihood Intervals	496
10.4.2	Other Large-Sample Intervals	499
10.5	Exercises	504
10.6	Miscellanea	515
11	Analysis of Variance and Regression	521
11.1	Introduction	521
11.2	Oneway Analysis of Variance	522
11.2.1	Model and Distribution Assumptions	524
11.2.2	The Classic ANOVA Hypothesis	525
11.2.3	Inferences Regarding Linear Combinations of Means	527
11.2.4	The ANOVA F Test	530
11.2.5	Simultaneous Estimation of Contrasts	534
11.2.6	Partitioning Sums of Squares	536
11.3	Simple Linear Regression	539
11.3.1	Least Squares: A Mathematical Solution	542
11.3.2	Best Linear Unbiased Estimators: A Statistical Solution	544
11.3.3	Models and Distribution Assumptions	548
11.3.4	Estimation and Testing with Normal Errors	550
11.3.5	Estimation and Prediction at a Specified $x = x_0$	557
11.3.6	Simultaneous Estimation and Confidence Bands	559
11.4	Exercises	563
11.5	Miscellanea	572

12 Regression Models	577
12.1 Introduction	577
12.2 Regression with Errors in Variables	577
12.2.1 Functional and Structural Relationships	579
12.2.2 A Least Squares Solution	581
12.2.3 Maximum Likelihood Estimation	583
12.2.4 Confidence Sets	588
12.3 Logistic Regression	591
12.3.1 The Model	591
12.3.2 Estimation	593
12.4 Robust Regression	597
12.5 Exercises	602
12.6 Miscellanea	608
Appendix: Computer Algebra	613
Table of Common Distributions	621
References	629
Author Index	645
Subject Index	649

List of Tables

1.2.1	Number of arrangements	16
4.1.1	Values of the joint pmf $f(x, y)$	141
7.3.1	Three estimators for a binomial p	354
	Counts of leukemia cases	360
8.3.1	Two types of errors in hypothesis testing	383
9.2.1	Location–scale pivots	427
9.2.2	Sterne’s acceptance region and confidence set	431
	Three 90% normal confidence intervals	441
10.1.1	Bootstrap and Delta Method variances	480
	Median/mean asymptotic relative efficiencies	484
10.2.1	Huber estimators	485
	Huber estimator asymptotic relative efficiencies, $k = 1.5$	487
	Poisson LRT statistic	490
10.3.1	Power of robust tests	497
10.4.1	Confidence coefficient for a pivotal interval	500
10.4.2	Confidence coefficients for intervals based on Huber’s M-estimator	504
11.2.1	ANOVA table for oneway classification	538
11.3.1	Data pictured in Figure 11.3.1	542
11.3.2	ANOVA table for simple linear regression	556
12.3.1	Challenger data	594
12.4.1	Potoroo data	598
	Regression M-estimator asymptotic relative efficiencies	601