

---

# Statistics for Social and Behavioral Sciences

**Series Editor**

Stephen E. Fienberg  
Carnegie Mellon University  
Pittsburgh  
Pennsylvania  
USA

Statistics for Social and Behavioral Sciences (SSBS) includes monographs and advanced textbooks relating to education, psychology, sociology, political science, public policy, and law.

More information about this series at <http://www.springer.com/series/3463>

---

Michael O. Finkelstein • Bruce Levin

# Statistics for Lawyers

Third Edition

With 58 Figures

 Springer

Michael O. Finkelstein  
New York, NY, USA

Bruce Levin  
Department of Biostatistics  
Columbia University Mailman  
School of Public Health  
New York, NY, USA

ISSN 2199-7357                      ISSN 2199-7365 (electronic)  
Statistics for Social and Behavioral Sciences  
ISBN 978-1-4419-5984-3              ISBN 978-1-4419-5985-0 (eBook)  
DOI 10.1007/978-1-4419-5985-0

Library of Congress Control Number: 2015934585

Springer New York Heidelberg Dordrecht London  
© Springer Science+Business Media, LLC 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer Science+Business Media LLC New York is part of Springer Science+Business Media  
([www.springer.com](http://www.springer.com))

*In Memory of  
Herbert Robbins (1915–2001)*

DISSERTATIO INAUGURALIS  
MATHEMATICO JURIDICA  
DE  
USU ARTIS  
CONJECTANDI  
IN JURE,  
Quam  
DIVINA JUVANTE GRATIA  
Auctoritate & Jussu  
*Magnifici & Amplissimi Fictorum Ordinis  
in Academia Patria*  
pro  
GRADU DOCTORATUS  
In Utroque Jure legitime consequendo  
*Ad Diem 14 Junii A. C. MDCCIX*  
L. H. Q. S.  
Publice defendet  
M. NICOLAUS BERNOULLI,  
Basiliensis.



BASILEÆ,

Typis JOHANNIS CONRADI à MECHEL.

The title page of Nicolaus Bernoulli's dissertation *The Use of the Art of Conjecturing in Law*, dated 1709. Nicolaus's thesis appears to be the first extended work applying statistical methods to legal questions.

---

## Preface to the Third Edition

For a reader new to this book, we suggest that you read the prefaces to the first and second editions for an overview; what we wrote there is still relevant as an introduction to our subject and the method of the book. For the teacher who has used prior editions, this edition includes over twenty new sections, including the following: §1.5.4 (SIDS tragedies), §5.5.2 (State trooper literacy exam), §8.1.3 (Rheumatoid arthritis drug), §11.1.4 (Exonerations in death-sentence cases), §13.2.4 (Projecting airline costs), §14.9.1 (New York City police stops), plus new expository material on various statistical techniques. In addition, each data table that appears in the book is available for download as a plaintext file at <http://www.columbia.edu/~BL6/SFLdata.htm>.

As in prior editions, much of the material for the new portions of this edition was collected and explored with help from lawyers and statisticians involved in the cases or studies. We thank them all for their generosity in assisting us. In this connection, we would like to thank in particular Nathan A. Schachtman for directing our attention to a number of new cases that we found useful to include in this edition. We also thank Ms. Dan Bai for helping us to update the references.

New York, NY, USA  
December 2014

Michael O. Finkelstein  
Bruce Levin





---

## Preface to the Second Edition

In the decade that has passed since the first edition of this book appeared, the crest of the wave of interest in statistical evidence has broadened and moved beyond its origins in civil rights law. Significant new developments, reflected in this edition, include, for example, DNA evidence (Sections 2.1.1, 3.1.2, and 3.2.2), epidemiologic studies in toxic substance litigation (Chapter 10), statistical models for adjusting census counts (Section 9.2.1), and vote-dilution cases (Section 13.2.3). It is emblematic of the importance of statistics in the pantheon of scientific evidence that the leading Supreme Court precedent on such evidence—the *Daubert*<sup>1</sup> case—involved toxic substance claims in which epidemiologic studies played a key role. In *Daubert*, the Court rejected the old *Frye* test of general acceptance in the scientific community as the basis for admissibility, and explicitly imposed on federal trial judges a gatekeeping function: they must now assess whether the proffered evidence is both relevant and *reliable*. The new formulation puts judges in the awkward position not only of counting scientific noses, but also of understanding and appraising the scientific basis of what an expert proposes to say, or calculate, on the witness stand. Fortuitously, about a year after *Daubert*, in 1994, the Federal Judicial Center issued and distributed to all federal judges a *Reference Manual on Scientific Evidence*, which is largely a primer on the applications of statistical methods. A new edition of the *Manual*, which unexpectedly turned out to be a best seller, is due to appear this year. Those who find this book heavy going may wish to consult the *Manual* as a useful introduction to at least some subjects.

But new, case-driven applications of statistics are only part of the development. Perhaps even more important, in the long run, is the continuing flow of statistical studies of the legal system itself. Studies of this sort can offer insights that sometimes challenge commonly held views of venerable legal institutions. Section 5.6.3 gives an example of such a study, involving peremptory challenges of prospective jurors, in which the authors analyze data and find that most peremptory challenges are “guesses.” For another example, as this is being written, the media are prominently reporting a large-scale statistical study of the death penalty, undertaken at Columbia Law School, which paints a startling picture of the high

---

<sup>1</sup> *Daubert v. Merrill Dow Pharmaceuticals*, 509 U.S. 579 (1993).

rate of serious errors in criminal trials leading to death sentences. The study will almost certainly influence pending legislation and promises to provide important data in the debate over capital punishment itself. One must note that in both these studies it is the statistical pattern, emerging from the details of individual cases, that tells the most compelling story.

As in the first edition, much of the material for the new portions of this second edition was collected from lawyers, statisticians, or economists who were involved in the cases. We thank them all for their generosity in assisting us. In this connection, we would like to acknowledge in particular Orley Ashenfelter, David Baldus, William Fairley, David Freedman, and Sol Schreiber for their help in furnishing us with their materials and consulting with us on matters of interpretation.

New York, NY, USA  
June 2000

Michael O. Finkelstein  
Bruce Levin

---

## Preface to the First Edition

*For the rational study of the law the black letter man may  
be the man of the present, but the man of the future is the man  
of statistics and the master of economics.*

—Oliver Wendell Holmes  
*The Path of the Law* (1897)

The aim of this book is to introduce lawyers and prospective lawyers to methods of statistical analysis used in legal disputes. The vehicle of this entertainment is a series of case studies interlaced with sections of mathematical exposition. The studies consist of summaries drawn primarily (but not exclusively) from actual cases, which are cast in the form of problems by questions posed to focus discussion. They are designed to illustrate areas of the law in which statistics has played a role (or at least has promised to do so), and to illustrate a variety of ways to reason quantitatively. Also included are some statistical studies of the legal system, and of the impact of proposed legislation or regulation. Wherever possible, excerpts of data are given to expose the reader to the sobering, hands-on experience of calculating statistics and drawing inferences. Judicial opinions are not given because they generally do not elucidate the statistical issues that are our primary concern. On the other hand, some judicial missteps are included so that the reader may exercise critical faculties and enhance self-esteem as a newly minted expert by correcting the bench.

Knowledge of probability or statistics is not required to calculate most of the answers called for by the snippets of data in the case studies. For the uninitiated, the statistical notes supply the technical tools. Some of these notes deal (in rather condensed fashion) with material that is covered in elementary texts; others go beyond that. For a more leisurely, detailed, or expansive discussion of the material, the reader may wish to consult a statistics text; some references are given in the text sections and in the bibliography. Our calculations for the mathematical questions in the case studies are given in Appendix I. The legal issues and the statistical issues not involving calculation are for the most part left to the reader.

Apart from the riveting intellectual interest of the subject, the lawyer or prospective lawyer may fairly question whether one needs to know quite as much about statistics as this book would teach. Of course, not all will. But for increasing

numbers of legal scholars, lawyers, judges, and even legislators, an acquaintance with statistical ideas, to paraphrase Justice Holmes, is not a duty, it is only a necessity. In diverse fields of learning, our knowledge is expressed in data that are appraised statistically. What is true of the general world has filtered into the courtroom. Economists of all stripes, social scientists, geneticists, epidemiologists, and others, testifying in their fields of expertise, make use of statistical tools for description and inference. In economics in particular, and in problems translated into economic terms, the ubiquity of data and computers, and the current fashion, have encouraged the creation of elaborate econometric models that are sufficiently plausible to be accepted in learned journals. But even models with impressive and intimidating technical apparatus may rest on shaky assumptions that, when exposed, undermine their credibility.

Frequently, statistical presentations in litigation are made not by statisticians but by experts from other disciplines, by lawyers who know a little, or by the court itself. This free-wheeling approach distinguishes statistical learning from most other expertise received by the courts and undoubtedly has increased the incidence of models with inappropriate assumptions, or just plain statistical error. The knowledgeable lawyer will be far more effective in proposing useful studies, exposing serious failings in complex models, and making the issues intelligible to a lay decisionmaker than one who is wholly dependent on a consultant for the next question. And although the lawyer usually will not need to make calculations, an appreciation of the principles—which is needed—is best gained from some modest grappling with the data.

Do statistics really matter? This is a question that sometimes vexes statisticians. In the legal setting, the questions are whether statistical models are fairly evaluated in the adversary process and whether statistical findings are given their due in the decisions. Unfortunately, the record here is spotty, even perverse. In some cases the courts have appraised statistical evidence well, but in some important public-issue litigation very good statistical models have been summarily rejected (and very bad ones uncritically embraced) by judges and justices in pursuit of their own agendas. The lawyer of the future predicted by Holmes ninety years ago has not yet come into his or her own.

Despite the trampling of statistical evidence that has occurred in some notable cases, it seems inevitable that studies based on data will continue to be pursued by the scholarly community and presented with increasing frequency in litigated matters involving public issues. A fuller appreciation of the standards for analyzing data and making inferences should at least lead to more accurately focused studies and more discerning treatment of such evidence by the courts. Beyond that, one may hope that the realities exposed by statistical work will in time influence perceptions of justice, even in the courtroom. A knowledgeable lawyer may not dispatch questions of legal policy with statistics, but by knowing more of the subject may hope to contribute to the store of rational and civilized discourse by which insights are gained and new accommodations reached. That, in any event, is the larger purpose of this book.

---

\*\*\*

Much of the material in this book was collected from lawyers and statisticians who were involved in the cases. We thank them all for their generosity in furnishing us with their papers. They are too numerous to list, but we would like to mention in particular David Baldus, Jack Boger, Will Fairley, David Freedman, Elizabeth Holtzman, Jay Kadane, and Jack Weinstein. We would also like to acknowledge Joseph Fleiss, Mervyn Susser, and Zena Stein, and their respective institutions, the Division of Biostatistics, the Sergievsky Center of the Faculty of Medicine, Columbia University, and the New York State Psychiatric Institute, for their liberal support of this project.

We would like especially to thank Margaret Murphy, who steadfastly typed and retyped the manuscript until the error rate was vanishingly small; Marcia Schoen, who typed the calculation notes; Stephen Sullivan, Lynn Cushman, and Matthew Herenstein for checking the citations; and Ann Kinney for editing the case studies and checking the calculations.

We own a debt of gratitude to our families—Claire, Katie, Matthew, Betty, Joby, Laura, and also Julie—for their patience, encouragement, and support during the long evolution of this book.

New York, NY, USA  
August 1989

Michael O. Finkelstein  
Bruce Levin



---

## Acknowledgments

- Figure 1.2.2. Reprinted by permission of the *Journal of Business and Economics Statistics*. Copyright 1985 by American Statistical Association, Alexandria, VA.
- Figure 2.1.1. Reprinted by permission from the National Research Council, *The Evaluation of Forensic DNA Evidence*. Copyright 1996 by the National Academy of Sciences, Washington, DC.
- Figure 9.1.1. Reprinted by permission of *Science*. Copyright 1971 by The American Association for the Advancement of Science, Washington, DC.
- Figure 9.2.1. Reprinted by permission from *Jurimetrics*. Copyright 1999 by the American Bar Association, Chicago, IL.
- Figure 10.2.1. Reprinted by permission from Howard Ball, *Justice Downwind: America's Atomic Testing Program in the 1950s*. Copyright 1986 by Oxford University Press, Inc., New York, NY.
- Figure 10.3.4. Reprinted by permission from the *Evaluation Review*. Copyright 1999 by Sage Publications, Inc., Newbury Park, CA.
- Figure 13.2.3. Reprinted by permission from the *Evaluation Review*. Copyright 1991 by Sage Publications, Inc., Newbury Park, CA.
- Figure 13.3c. Reprinted by permission from F.J. Anscombe, *Graphs in Statistical Analysis*, in 27 *The American Statistician* 17 (1973). Copyright 1973 by the American Statistical Association, Alexandria, VA.
- Figure 14.1. Reprinted by permission from R.J. Wonnacott & T.L.L. Wonnacott, *Econometrics*. Copyright 1970 by John Wiley & Sons, New York, NY.
- Table F and Figures 5.3c and 5.3d. Reprinted by permission from *Biometrika Tables for Statisticians, Volume I* by E.S. Pearson and H.O. Hartley. Copyright 1954 by Cambridge University Press, Cambridge, England.
- Tables G1, G2, H1 and H2. Reprinted by permission from *Handbook Tables for Probability & Statistics 2nd Ed.* Copyright 1968 by CRC Press, Inc., Boca Raton, FL.





---

# Contents

<b>1</b>	<b>Descriptive Statistics</b> . . . . .	1
1.1	Introduction to descriptive statistics . . . . .	1
1.2	Measures of central location . . . . .	2
1.2.1	Parking meter heist . . . . .	7
1.2.2	Taxing railroad property . . . . .	11
1.2.3	Capital infusions for ailing thrifts . . . . .	14
1.2.4	Disproportionate-share hospitals . . . . .	14
1.2.5	Hydroelectric fish kill . . . . .	15
1.2.6	Pricey lettuce . . . . .	16
1.2.7	Apportionment of Representatives among the states . . . . .	17
1.2.8	Super-drowsy drug . . . . .	18
1.3	Measures of dispersion . . . . .	20
1.3.1	Texas reapportionment . . . . .	25
1.3.2	Damages for pain and suffering . . . . .	27
1.3.3	Ancient trial of the Pyx . . . . .	28
1.3.4	<i>Bush v. Gore</i> : The U.S. Supreme Court's election 2000 decision . . . . .	30
1.4	A measure of correlation . . . . .	30
1.4.1	Dangerous eggs . . . . .	34
1.4.2	Public school finance in Texas . . . . .	36
1.5	Measuring the disparity between two proportions . . . . .	38
1.5.1	Proficiency test with a disparate impact . . . . .	41
1.5.2	Bail and bench warrants . . . . .	42
1.5.3	Non-intoxicating beer . . . . .	44
1.5.4	SIDS tragedies . . . . .	45
<b>2</b>	<b>How to Count</b> . . . . .	47
2.1	Permutations and combinations . . . . .	47
2.1.1	DNA profiling . . . . .	49
2.1.2	Weighted voting . . . . .	53
2.1.3	Was the bidding rigged? . . . . .	54

2.1.4	A cluster of leukemia . . . . .	55
2.1.5	Measuring market concentration . . . . .	55
2.2	Fluctuation theory . . . . .	57
2.2.1	Tracing funds for constructive trusts . . . . .	59
<b>3</b>	<b>Elements of Probability . . . . .</b>	<b>61</b>
3.1	Some fundamentals of probability calculation . . . . .	61
3.1.1	Interracial couple in yellow car . . . . .	66
3.1.2	Independence assumption in DNA profiles . . . . .	70
3.1.3	Telltale fibers . . . . .	72
3.1.4	Telltale hairs . . . . .	73
3.1.5	Randomized response technique . . . . .	76
3.2	Selection effect . . . . .	77
3.2.1	L'affaire Dreyfus . . . . .	77
3.2.2	Searching DNA databases . . . . .	78
3.2.3	Trawling in DNA Databases . . . . .	79
3.3	Bayes's theorem . . . . .	80
3.3.1	Rogue bus . . . . .	82
3.3.2	Bayesian proof of paternity . . . . .	83
3.4	Screening devices and diagnostic tests . . . . .	85
3.4.1	Airport screening device . . . . .	88
3.4.2	Polygraph evidence . . . . .	90
3.5	Monte Carlo methods . . . . .	92
3.5.1	Sentencing a heroin swallower . . . . .	93
3.5.2	Cheating on multiple-choice tests . . . . .	95
3.6	Foundations of probability . . . . .	98
3.6.1	Relevant evidence defined . . . . .	100
<b>4</b>	<b>Some Probability Distributions . . . . .</b>	<b>101</b>
4.1	Introduction to probability distributions . . . . .	101
4.2	Binomial distribution . . . . .	105
4.2.1	Discrimination in jury selection . . . . .	108
4.2.2	Educational nominating panel . . . . .	108
4.2.3	Small and nonunanimous juries in criminal cases . . . . .	110
4.2.4	Cross-section requirement for federal jury lists . . . . .	113
4.2.5	Cat litter imbroglio . . . . .	114
4.3	Normal distribution and a central limit theorem . . . . .	116
4.3.1	<i>Alexander</i> : Culling the jury list . . . . .	120
4.3.2	<i>Castaneda</i> : Measuring disparities . . . . .	121
4.4	Testing statistical hypotheses . . . . .	123
4.4.1	Hiring teachers . . . . .	125
4.5	Hypergeometric distribution . . . . .	126
4.5.1	Were the accountants negligent? . . . . .	130
4.5.2	Challenged election . . . . .	131
4.5.3	Election 2000: Who won Florida? . . . . .	131
4.5.4	Taking the stand . . . . .	138

---

4.6	Tests of normality . . . . .	139
4.6.1	Heights of French conscripts . . . . .	142
4.6.2	Silver “butterfly” straddles . . . . .	143
4.7	Poisson distribution . . . . .	145
4.7.1	Sulphur in the air . . . . .	148
4.7.2	Vaccinations . . . . .	149
4.7.3	Is the cult dangerous? . . . . .	149
4.7.4	Incentive for good drivers . . . . .	150
4.7.5	Epidemic of cardiac arrests . . . . .	150
4.8	Geometric and exponential distributions . . . . .	152
4.8.1	Marine transportation of liquefied natural gas . . . . .	154
4.8.2	Network affiliation contracts . . . . .	154
4.8.3	Dr. Branion’s case . . . . .	155
4.9	Gamma distribution . . . . .	156
<b>5</b>	<b>Statistical Inference for Two Proportions . . . . .</b>	<b>159</b>
5.1	Fisher’s exact test of equality for two proportions . . . . .	159
5.1.1	Nursing examination . . . . .	161
5.2	The chi-squared and z-score tests for the equality of two proportions . . . . .	162
5.2.1	Suspected specialists . . . . .	168
5.2.2	Reallocating commodity trades . . . . .	169
5.2.3	Police examination . . . . .	169
5.2.4	Promotions at a bank . . . . .	170
5.3	Confidence intervals for proportions . . . . .	171
5.3.1	Confounders and confidence intervals . . . . .	180
5.3.2	Paucity of Crossets . . . . .	180
5.3.3	Purloined notices . . . . .	181
5.3.4	Commodity exchange reports . . . . .	182
5.3.5	Discharge for dishonest acts . . . . .	183
5.3.6	Confidence interval for promotion test data . . . . .	184
5.3.7	Complications in vascular surgery . . . . .	184
5.3.8	Torture, disappearance, and summary execution in the Philippines . . . . .	185
5.4	Statistical power in hypothesis testing . . . . .	187
5.4.1	Death penalty for rape . . . . .	189
5.4.2	Is Bendectin a teratogen? . . . . .	191
5.4.3	Automobile emissions and the Clean Air Act . . . . .	192
5.5	Legal and statistical significance . . . . .	193
5.5.1	Port Authority promotions . . . . .	195
5.5.2	State trooper literacy exam . . . . .	197
5.6	Maximum likelihood estimation . . . . .	198
5.6.1	Purloined notices revisited . . . . .	202
5.6.2	Do microwaves cause cancer? . . . . .	202
5.6.3	Peremptory challenges of prospective jurors . . . . .	203

<b>6</b>	<b>Comparing Multiple Proportions . . . . .</b>	<b>207</b>
6.1	Using chi-squared to test goodness of fit . . . . .	207
6.1.1	Death-qualified jurors . . . . .	209
6.1.2	Spock jurors . . . . .	211
6.1.3	Grand jury selection revisited . . . . .	212
6.1.4	Howland Will contest . . . . .	213
6.1.5	Imanishi-Kari's case . . . . .	215
6.2	Bonferroni's inequality and multiple comparisons . . . . .	217
6.2.1	Wage additives and the four-fifths rule . . . . .	220
6.2.2	Discretionary parole . . . . .	221
6.2.3	Cheating on standardized multiple-choice tests revisited . . . . .	222
6.2.4	InterMune's Actimmune . . . . .	222
6.3	More measures of association: Phi-squared ( $\phi_2$ ) and tau B ( $\tau_B$ ) . . . . .	224
6.3.1	Preventive detention . . . . .	226
<b>7</b>	<b>Comparing Means . . . . .</b>	<b>231</b>
7.1	Student's $t$ -test: Hypothesis testing and confidence intervals . . . . .	231
7.1.1	Automobile emissions and the Clean Air Act revisited . . . . .	235
7.1.2	Voir dire of prospective trial jurors . . . . .	236
7.1.3	Ballot position . . . . .	238
7.1.4	Backdating stock options . . . . .	238
7.1.5	Bullet lead identification . . . . .	240
7.2	Analysis of variance for comparing several means . . . . .	241
7.2.1	Fiddling debt collector . . . . .	245
<b>8</b>	<b>Combining Evidence Across Independent Strata . . . . .</b>	<b>247</b>
8.1	Mantel-Haenszel and Fisher methods for combining the evidence . . . . .	247
8.1.1	Hiring lawyers . . . . .	257
8.1.2	Age discrimination in employment terminations . . . . .	258
8.1.3	Rheumatoid arthritis drug . . . . .	260
8.2	Meta-analysis . . . . .	262
8.2.1	Bendectin revisited . . . . .	267
8.2.2	Avandia and heart attacks . . . . .	268
<b>9</b>	<b>Sampling Issues . . . . .</b>	<b>271</b>
9.1	The theory of random sampling . . . . .	271
9.1.1	Selective Service draft lotteries . . . . .	277
9.1.2	Uninsured motor vehicles . . . . .	280
9.1.3	Mail order survey . . . . .	281
9.1.4	Domino effect . . . . .	283
9.1.5	NutraTaste versus NutraSweet . . . . .	284

---

9.1.6	Cocaine by the bag . . . . .	286
9.1.7	ASCAP sampling plan . . . . .	287
9.1.8	Current Population Survey . . . . .	289
9.1.9	Insurance payments to doctors . . . . .	290
9.1.10	Wage and hour class action . . . . .	292
9.2	Capture/recapture . . . . .	293
9.2.1	Adjusting the census . . . . .	294
<b>10</b>	<b>Epidemiology . . . . .</b>	<b>299</b>
10.1	Introduction . . . . .	299
10.2	Attributable risk . . . . .	301
10.2.1	Atomic weapons tests . . . . .	304
10.3	Epidemiologic principles of causation . . . . .	308
10.3.1	Dalkon Shield . . . . .	312
10.3.2	Radioactive “cocktails” for pregnant women . . . . .	314
10.3.3	Preconception paternal irradiation and leukemia . . . . .	318
10.3.4	Swine flu vaccine and Guillain-Barré Syndrome . . . . .	323
10.3.5	Silicone breast implants . . . . .	326
<b>11</b>	<b>Survival Analysis . . . . .</b>	<b>333</b>
11.1	Death-density, survival, and hazard functions . . . . .	333
11.1.1	Valuing charitable remainders . . . . .	337
11.1.2	Defective house sidings . . . . .	338
11.1.3	“Lifing” deposit accounts . . . . .	338
11.1.4	Exonerations in death-sentence cases . . . . .	340
11.2	The proportional hazards model . . . . .	342
11.2.1	Age discrimination in employment terminations revisited . . . . .	346
11.2.2	Contaminated wells in Woburn . . . . .	346
11.3	Quantitative risk assessment . . . . .	350
11.3.1	Ureaformaldehyde foam insulation . . . . .	353
11.3.2	Ethylene oxide . . . . .	355
<b>12</b>	<b>Nonparametric Methods . . . . .</b>	<b>357</b>
12.1	The sign test . . . . .	357
12.1.1	Supervisory examinations . . . . .	359
12.2	Wilcoxon signed-rank test . . . . .	360
12.2.1	Voir dire of prospective trial jurors revisited . . . . .	361
12.3	Wilcoxon rank-sum test . . . . .	361
12.3.1	Sex discrimination in time to promotion . . . . .	362
12.3.2	Selection for employment from a list . . . . .	363
12.3.3	Sentencing by federal judges . . . . .	364
12.4	Spearman’s rank correlation coefficient . . . . .	366
12.4.1	Draft lottery revisited . . . . .	367

<b>13</b>	<b>Regression Models</b> . . . . .	369
13.1	Introduction to multiple regression models . . . . .	369
13.1.1	Head Start programs . . . . .	374
13.2	Estimating and interpreting coefficients of the regression equation . . . . .	376
13.2.1	Western Union's cost of equity . . . . .	380
13.2.2	Tariffs for North Slope oil . . . . .	381
13.2.3	Ecological regression in vote-dilution cases . . . . .	382
13.2.4	Projecting airline costs . . . . .	385
13.3	Measures of indeterminacy for the regression equation . . . . .	387
13.3.1	Sex discrimination in academia . . . . .	391
13.4	Statistical significance of the regression coefficients . . . . .	394
13.4.1	Race discrimination at Muscle Shoals . . . . .	397
13.5	Explanatory factors for a regression equation . . . . .	398
13.6	Reading multiple-regression computer printout . . . . .	404
13.6.1	Pay discrimination in an agricultural extension service . . . . .	405
13.6.2	Public school financing in the State of Washington . . . . .	413
13.6.3	Public school financing in Pennsylvania . . . . .	416
13.7	Confidence and prediction intervals . . . . .	418
13.7.1	Projecting fuel costs . . . . .	419
13.7.2	Severance pay dispute . . . . .	420
13.7.3	Challenged absentee ballots . . . . .	422
13.7.4	Projecting airline costs revisited . . . . .	423
13.8	Assumptions of the regression model . . . . .	424
13.9	Transformations of variables . . . . .	430
13.9.1	Western Union's cost of equity revisited . . . . .	436
13.9.2	Sex- and race-coefficient models for Republic National Bank . . . . .	436
<b>14</b>	<b>More Complex Regression Models</b> . . . . .	439
14.1	Time series . . . . .	439
14.1.1	Corrugated container price-fixing . . . . .	441
14.1.2	Puppy Chow versus Chewy Morsels . . . . .	442
14.1.3	Losses from infringing sales . . . . .	444
14.1.4	OTC market manipulation . . . . .	445
14.1.5	Fraud-on-the-market damages . . . . .	448
14.1.6	Effects of capital gains tax reductions . . . . .	450
14.1.7	Projecting airline costs re-revisited . . . . .	455
14.2	Interactive models . . . . .	456
14.2.1	House values in the shadow of a uranium plant . . . . .	459
14.3	Alternative models in employment discrimination cases . . . . .	462
14.4	Locally weighted regression . . . . .	466
14.4.1	Urn models for Harris Bank . . . . .	468

---

14.5	Underadjustment bias in employment regressions . . . . .	469
14.5.1	Underadjustment in medical school . . . . .	473
14.6	Systems of equations . . . . .	475
14.6.1	Death penalty: Does it deter murder? . . . . .	477
14.6.2	Neurontin off-label marketing . . . . .	479
14.7	Logit and probit regression . . . . .	480
14.7.1	Mortgage lending discrimination . . . . .	485
14.7.2	Death penalty in Georgia . . . . .	486
14.7.3	Deterring teenage smoking . . . . .	495
14.8	Poisson regression . . . . .	496
14.8.1	Challenger disaster . . . . .	497
14.9	Negative binomial regression . . . . .	499
14.9.1	New York City police stops . . . . .	501
14.10	Jackknife, cross-validation, and bootstrap . . . . .	503
14.10.1	Georgia death penalty revisited . . . . .	505
<b>Erratum . . . . .</b>		<b>E1</b>
<b>Appendix I: Calculations and Comments on the Cases . . . . .</b>		<b>509</b>
<b>Appendix II: Tables . . . . .</b>		<b>599</b>
<b>Glossary of Symbols . . . . .</b>		<b>627</b>
<b>List of Cases . . . . .</b>		<b>629</b>
<b>Bibliography . . . . .</b>		<b>633</b>
<b>Index . . . . .</b>		<b>641</b>





# List of Figures

Fig. 1.2a	Random variable $X$ with a symmetrical probability distribution .....	4
Fig. 1.2b	Random variable $X$ with a right-skewed probability distribution .....	5
Fig. 1.2.1a	Revenue per meter per day for two 10-month periods .....	10
Fig. 1.2.1b	Revenue per collection day, May 1978–March 1981. ....	11
Fig. 1.2.2	Histogram for distribution of assessment/sales ratios .....	12
Fig. 1.2.8	Percentage reduction in sleep latency vs. baseline latency ...	20
Fig. 1.4	(a) Correlation $\approx 0$ ; (b) Correlation $\approx 0.5$ ; and (c) Correlation $\approx 1$ .....	32
Fig. 1.4d	Correlation depends on the range of the variables .....	33
Fig. 1.4.1	Ischemic heart disease mortality per 100,000 vs. egg consumption .....	36
Fig. 2.1.1	An autoradiograph from an actual case illustrating RFLP at the D1S7 locus .....	52
Fig. 2.1.4	Residences of childhood leukemia patients at time of diagnosis, Woburn, Massachusetts, 1969–1979 .....	55
Fig. 2.2a	Record of 10,000 tosses of an ideal coin .....	57
Fig. 2.2b	The André Reflection Principle .....	58
Fig. 3.1	Venn diagrams illustrating conjunction (“and”) and inclusive disjunction (“and/or”) .....	64
Fig. 3.4.	ROC curves for good and bad screening tests .....	88
Fig. 3.5.1	Cumulative frequency distribution of smuggled heroin .....	94
Fig. 4.1a	Probabilities of dice outcomes 2, ..., 12 .....	102
Fig. 4.1b	Probabilities as areas under the standard normal relative frequency curve, $f(x) = (2\pi)^{-0.5} \exp(-x^2/2)$ .....	104
Fig. 4.1c	Standard normal cumulative distribution function .....	104
Fig. 4.3	One-half continuity correction to approximate $P[S \geq 7]$ , where $S \sim \text{Bin}(10, 0.5)$ .....	120

Fig. 4.6	Normal probability paper .....	140
Fig. 4.6.2	Silver “butterfly” straddles output .....	141
Fig. 4.8	Probability density function for the standard exponential distribution .....	153
Fig. 5.2	Members of the family of chi-squared distributions .....	163
Fig. 5.3a	Determining those values of $P$ that lie below or above the one-sided upper 95% confidence limit .....	172
Fig. 5.3b	Determining those values of $P$ that lie below or above the one-sided lower 95% confidence limit .....	172
Fig. 5.3c	Exact two-sided 95% confidence limits for a binomial proportion .....	174
Fig. 5.3d	Exact two-sided 99% confidence limits for a binomial proportion .....	175
Fig. 5.3e	A 95% confidence interval covers the true $p$ in 19 out of 20 replications .....	177
Fig. 6.1.4	Undisputed and challenged signatures of Sylvia Ann Howland .....	214
Fig. 9.1.1	The 1970 Draft: Random selection sequence vs. birth date .....	280
Fig. 9.2.1	Share changes from the proposed adjustment to the census, as of July 15, 1991 .....	297
Fig. 10.2	Attributable risk in the population .....	303
Fig. 10.2.1	Atomic weapons test fallout .....	305
Fig. 10.3.3	The Sellafield area .....	319
Fig. 10.3.4	Panel (a) shows the incidence of GBS among the vaccinated, by week since vaccination. Panel (b) shows the incidence of GBS among the unvaccinated from the start of the vaccination program .....	325
Fig. 10.3.5	Power versus underlying effect measure (including Hennekens) .....	330
Fig. 11.1	Product-limit estimates for two survival functions .....	335
Fig. 13.1a	Galton’s illustration of regression to the mean .....	372
Fig. 13.1b	Heights of sons and midparents, showing two regression lines .....	373
Fig. 13.2	Residuals from a fitted regression model .....	377
Fig. 13.2.3	Turnout rates for Feliciano, the principal Hispanic candidate in the 1982 sheriff primary, Los Angeles County .....	384
Fig. 13.2.4	Plot of military revenues and costs, 1998–2007 with regression line .....	386

---

Fig. 13.3a	Simulated parents-and-sons data illustrating constant error variance .....	388
Fig. 13.3b	Explained and unexplained deviations from the mean .....	390
Fig. 13.3c	Scatterplots with the same correlation coefficient .....	392
Fig. 13.7	Confidence and prediction bands for the regression mean ....	420
Fig. 13.8	A plot of residuals .....	429
Fig. 13.9	Sale prices for houses made more symmetrical by transformation into logs .....	434
Fig. 14.1	Effect of autocorrelation .....	440
Fig. 14.1.1	Actual and projected prices for corrugated containerboard during and after the conspiracy period .....	442
Fig. 14.1.6a	Realized net long-term gains and Gross National Product, 1954–1985 (in billions of dollars) .....	451
Fig. 14.1.6b	Realized net long-term gains and corporate equity of households, 1954–1985 (in billions of dollars) .....	451
Fig. 14.3	Alternative forms of regression analysis in employment discrimination cases .....	467
Fig. 14.5	The mechanism of underadjustment bias .....	470



# List of Tables

Table 1.2.1	Parking meter collection data (revenue in dollars) .....	9
Table 1.2.8	Sleep study data .....	19
Table 1.3.1	Redistricting plan for the Texas House of Representatives .....	26
Table 1.3.2	Normative case awards (\$ in 000s) .....	28
Table 1.4.1	Ischemic heart disease mortality and egg consumption among men aged 55–64 in 40 countries .....	35
Table 1.4.2	Texas school finance sample data .....	37
Table 1.5.1	Department store proficiency test: test performance by race .....	42
Table 1.5.2	Issuance of bench warrants: current issuance by previous issuance .....	43
Table 3.1.2	Numbers of three alleles at locus D2S44 in samples from four populations .....	71
Table 3.5.1	Net weight in grams per internal smuggling trip known to DEA agents at Kennedy Airport, September 1, 1990, to December 10, 1991 .....	94
Table 3.5.2	Wrong answers and matching wrong answers in promotion examination .....	96
Table 4.2.2	Educational Nominating Panels. Philadelphia, 1965–1971 .....	109
Table 4.2.3	Jury voting .....	110
Table 4.5.3	Counted votes and undervotes in Florida counties in the 2000 presidential election .....	134
Table 4.5.4	Jury knowledge of prior criminal record and conviction .....	138
Table 4.6.1	Heights of French conscripts—1817 data .....	142
Table 4.6.2	Silver butterfly analysis: August 1973–July 1979 .....	144
Table 4.7.4	Distribution of accidents for a cohort of drivers .....	150
Table 4.7.5	Deaths in a pediatric ICU during shifts of selected personnel .....	151
Table 4.8.2	Contract terminations for 250 CBS and NBC affiliates ....	154

Table 5.2	Calculation of chi-squared for the <i>Teal</i> data .....	164
Table 5.2.1	Options exchange transactions .....	169
Table 5.2.4	Promotions by race: Federal Reserve Bank of Richmond, 1974–1977 .....	170
Table 5.3.8a	Summary of special master’s damage recommendations for valid claims .....	185
Table 5.3.8b	Special master’s damage recommendations .....	186
Table 5.4.1a	Race of defendant by sentence .....	190
Table 5.4.1b	Race of victim by sentence .....	190
Table 5.4.1c	Race of defendant by prior record .....	190
Table 5.4.1d	Prior record by sentence .....	190
Table 5.4.1e	Race of defendant by type of entry .....	190
Table 5.4.1f	Type of entry by sentence .....	190
Table 6.1.1	Verdict choices of death-qualified and excludable jurors .....	210
Table 6.1.2	Sex composition of jury panels .....	212
Table 6.1.4	Coincidence of 30 downstrokes in 861 signature pairs . . .	215
Table 6.1.5	Frequencies of insignificant cpm digits in questioned and control data .....	217
Table 6.2.1	Non-technical education additives .....	220
Table 6.2.2	Parole by race and ethnicity .....	221
Table 6.3.1	Defendants’ failure to appear and/or rearrest by risk level .....	226
Table 7.1.1	Emission data for NO <sub>x</sub> .....	236
Table 7.1.2	Test judges’ voir dire impaneling times .....	237
Table 8.1	Hiring practices, University of Texas, 1976–1977 .....	249
Table 8.1.1	Attorney applicants and hires at the office of the Attorney General from 1970 to 1983 .....	257
Table 8.1.2a	Ages of copywriters by dates of involuntary termination .....	258
Table 8.1.2b	Ages of art directors by dates of involuntary termination .....	259
Table 8.1.2c	Combining the evidence: copywriters and art directors . . .	259
Table 8.1.3a	R788 results for RA patients (pooled data) .....	260
Table 8.1.3b	R788 results for RA patients (by country) .....	261
Table 8.2.1a	Cohort studies of Bendectin and congenital malformations .....	267
Table 8.2.1b	Findings of Bendectin cohort studies .....	268
Table 8.2.2	Myocardial infarctions in Avandia studies .....	269
Table 9.1.1a	1970 Draft: Random selection sequence, by month and day .....	278
Table 9.1.1b	1971 Draft: Random selection sequence, by month and day .....	279

Table 9.1.3a	Percentage distribution of orders by size (by number not value) .....	281
Table 9.1.3b	Percentage distribution of orders by shipping address .....	282
Table 9.2.1	Estimated net census undercounts, by race and sex, from the Post-Enumeration Survey and from demographic analysis .....	296
Table 10.2.1a	Distribution of childhood leukemia deaths .....	306
Table 10.2.1b	Distribution of leukemia deaths and person-years for southern and northern Utah, eastern Oregon, and Iowa .....	306
Table 10.3.1	Method of contraception for PID cases and controls .....	313
Table 10.3.3	Leukemia in young persons (age under 25 years) born and diagnosed in West Cumbria, 1950–1985, by father’s occupational exposure to ionizing radiation . . .	320
Table 10.3.5a	Sum of “Definite CTDs” (rheumatoid arthritis, systemic lupus, scleroderma, Sjögren’s syndrome, dermatomyositis/polymyositis, and those so designated by author) by study .....	329
Table 10.3.5b	Silicone breast implants: weighted least squares meta-analysis using studies with adjusted odds ratios .....	330
Table 11.1.2	Failure data for laurel green paint .....	339
Table 11.1.3	Observed closures in 1983: regular savings accounts .....	339
Table 11.2.2	Observed and expected exposures to wells G and H for twenty childhood leukemia cases .....	349
Table 11.3.1	Chemical Industry Institute of Toxicology rat formaldehyde study .....	354
Table 11.3.2a	Pathologies in rats after exposure to ethylene oxide .....	355
Table 11.3.2b	Excess lifetime risk of cancer per 10,000 workers .....	356
Table 12.3.1	Time, in months, from hire to promotion .....	362
Table 12.3.2	Selection for employment: Minority list .....	364
Table 12.3.3	Ranks of sentences of individual judges in thirteen model cases .....	365
Table 13.1.1a	Means on post-test according to pre-test score .....	374
Table 13.1.1b	Means on pre-test according to post-test score .....	375
Table 13.2.1	Cost of equity and variability of earnings for selected regulated industries .....	380
Table 13.2.2	Persian Gulf crudes .....	382
Table 13.2.4	ATA Airlines revenue and costs (rounded data) .....	386
Table 13.6.1	Multiple-regression computer output from <i>Bazemore</i> <i>v. Friday</i> .....	407
Table 13.6.2	Regression for 1980 4th-grade language scores .....	414
Table 13.6.3	1995 PSSA test model for 496 Pennsylvania school districts .....	417
Table 13.7.2	A regression model for weeks of severance pay .....	421

Table 13.7.3	Machine votes and absentee ballots in 21 Pennsylvania senatorial elections .....	422
Table 13.9.2a	Plaintiffs' multiple-regression models for Republic National Bank salary levels in 1978 .....	437
Table 13.9.2b	Defendant's multiple-regression models for Republic National Bank 1978 year-end salary .....	437
Table 14.1.1	Price of corrugated containerboard .....	441
Table 14.1.3	Sales data .....	445
Table 14.1.4a	Regression of Olympic Brewing Company stock return on Daily Value Weighted Return (DVWR) .....	447
Table 14.1.4b	Daily return data on Olympic stock and the market for selected days in January and March, 1977 .....	447
Table 14.1.5	Regression of Geri-med closing stock prices on nursing home company stock index .....	449
Table 14.1.6a	Capital gains realizations time series data .....	453
Table 14.1.6b	Time series linear regression models for capital gains realizations .....	453
Table 14.1.6c	A semi-logarithmic regression model for capital gains realizations .....	454
Table 14.1.7	Airline costs regressed on revenue and trend (\$ in millions) .....	455
Table 14.2.1a	Rosen model results .....	460
Table 14.2.1b	Gartside model results .....	461
Table 14.2.1c	Gartside regression model average price (\$ 000's) .....	461
Table 14.2.1d	Gartside regression model—average price changes from 1984 .....	462
Table 14.6.1	Estimated effect of execution risk on criminal homicide rate .....	479
Table 14.6.2	Rosenthal regression output .....	481
Table 14.7.2a	Variable codes for 100 death-eligible cases .....	488
Table 14.7.2b	Logistic regression results for the Procedural Reform Study data .....	488
Table 14.7.2c	100 death-eligible cases from the middle range .....	489
Table 14.7.3	Tax coefficients in probit regression of teenage smoking .....	496
Table 14.8.1	Number of damaged O-rings (Row B) by ambient temperature in degrees Fahrenheit (Row A) at launch for 24 shuttle flights prior to Challenger flight 51-L .....	498
Table 14.9.1	Negative binomial regression of total marijuana enforcement by precinct, demography, and socioeconomic conditions .....	501
Table 14.10.1a	Death sentences and predicted probability of death (construction sample) .....	505
Table 14.10.1b	Death sentences and predicted probability of death (test sample) .....	506