

Statistics for Social and Behavioral Sciences

Michael O. Finkelstein
Bruce Levin

Statistics for Lawyers

Statistics for Social and Behavioral
Science

Fourth Edition

 Springer

Statistics for Social and Behavioral Sciences

Volume 189

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.

Michael O. Finkelstein • Bruce Levin

Statistics for Lawyers

Statistics for Social and Behavioral Science

Fourth Edition

With 59 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 978-1-0716-4155-2 ISBN 978-1-0716-4156-9 (eBook)
<https://doi.org/10.1007/978-1-0716-4156-9>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 1990, 2001, 2015, 2024

This work is subject to copyright. All rights are solely and exclusively licensed 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, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.

The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

If disposing of this product, please recycle the paper.

DISSERTATIO INAUGURALIS
MATHEMATICO JURIDICA

DE

USU ARTIS
CONJECTANDI
IN JURE,

Quam

DIVINA JUVANTE GRATIA

Auctoritate & Jussu

*Magnifici & Amplissimi Actorum 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.

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

Preface to the Fourth Edition

For a reader who is 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 14 new sections, 4 inserts to the statistical text, and 6 new answers, including the following: §3.3.2 (Insert on priors); §3.6 (Insert on the history of the word probability); §4.5.5 Lipitor and diabetes; §6.1.6 Harvard’s affirmative action practices in admissions (with answers); §7.2.2 New York City garbage trucks (with answers); §8.1 (Insert on tests of odds ratio homogeneity); §8.1.4 Disparate impact of a pre-employment exam on minority applicants (with answers); §8.2.3 Liraglutide and pancreatic cancer; §9.1 (Insert on representative samples); §9.1.11 Reversals in death-penalty cases; §9.2.2 Technology assisted review in e-discovery (with answers); §10.3.6 Asbestos and colon cancer; §13.2.5 Guilty pleas in the federal courts and the IRNC (implicit rate of non-conviction, with answers); §14.1.8 The “financing secured” event study (with answers); and §14.11 Average marginal effects. As in the last edition, 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 statisticians involved in the cases or studies. We thank them all for their generosity in assisting us. In particular, we wish to thank Nathan A. Schachtman, William B. Fairley, Joseph L. Gastwirth, Maura R. Grossman, and Michael Hartzmark for their very helpful contributions to this edition.

We especially thank our wives, Vivian Berger and Betty Levin, for their invaluable help and support in preparing this edition.

New York, NY, USA
April 2024

Michael O. Finkelstein
Bruce Levin

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*¹ 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

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

“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 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 owe 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

We are grateful to the following entities for granting permission to reprint these items.

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 13.2.5. Reprinted by permission from Michael O. Finkelstein & Bruce Levin, *Revisiting the Implicit Rate of Non-Conviction for Marginal Guilty Plea Defendants*, in 59 *Jurimetrics J.* 439 (2019). Copyright 2019 by the American Bar Association.

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

1	Descriptive Statistics	1
1.1	Introduction to descriptive statistics	1
1.2	Measures of central location	3
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	19
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	Bush v. Gore: The U.S. Supreme Court's election 2000 decision	30
1.4	A measure of correlation	31
1.4.1	Dangerous eggs	34
1.4.2	Public school finance in Texas	37
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
2	How to Count	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 54
- 2.1.5 Measuring market concentration 55
- 2.2 Fluctuation theory 57
 - 2.2.1 Tracing funds for constructive trusts 59
- 3 Elements of Probability 61**
 - 3.1 Some fundamentals of probability calculation 61
 - 3.1.1 Interracial couple in yellow car 67
 - 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 80
 - 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 86
 - 3.4.1 Airport screening device 89
 - 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 101
- 4 Some Probability Distributions 103**
 - 4.1 Introduction to probability distributions 103
 - 4.2 Binomial distribution 107
 - 4.2.1 Discrimination in jury selection 110
 - 4.2.2 Educational nominating panel 110
 - 4.2.3 Small and nonunanimous juries in criminal cases 112
 - 4.2.4 Cross-section requirement for federal jury lists 115
 - 4.2.5 Cat litter imbroglio 116
 - 4.3 Normal distribution and a central limit theorem 118
 - 4.3.1 Alexander: Culling the jury list 121
 - 4.3.2 Castaneda: Measuring disparities 123
 - 4.4 Testing statistical hypotheses 124
 - 4.4.1 Hiring teachers 127
 - 4.5 Hypergeometric distribution 128
 - 4.5.1 Were the accountants negligent? 131
 - 4.5.2 Challenged election 132

4.5.3	Election 2000: Who won Florida?	133
4.5.4	Taking the stand	139
4.5.5	Lipitor and diabetes	140
4.6	Tests of normality	142
4.6.1	Heights of French conscripts	143
4.6.2	Silver “butterfly” straddles	146
4.7	Poisson distribution	148
4.7.1	Sulphur in the air	151
4.7.2	Vaccinations	151
4.7.3	Is the cult dangerous?	152
4.7.4	Incentive for good drivers	152
4.7.5	Epidemic of cardiac arrests	153
4.8	Geometric and exponential distributions	155
4.8.1	Marine transportation of liquefied natural gas	156
4.8.2	Network affiliation contracts	157
4.8.3	Dr. Branion’s case	158
4.9	Gamma distribution	159
5	Statistical Inference for Two Proportions	161
5.1	Fisher’s exact test of equality for two proportions	161
5.1.1	Nursing examination	163
5.2	The chi-squared and z-score tests for the equality of two proportions	164
5.2.1	Suspected specialists	170
5.2.2	Reallocating commodity trades	171
5.2.3	Police examination	171
5.2.4	Promotions at a bank	171
5.3	Confidence intervals for proportions	173
5.3.1	Confounders and confidence intervals	181
5.3.2	Paucity of Crossets	182
5.3.3	Purloined notices	183
5.3.4	Commodity exchange reports	184
5.3.5	Discharge for dishonest acts	184
5.3.6	Confidence interval for promotion test data	185
5.3.7	Complications in vascular surgery	185
5.3.8	Torture, disappearance, and summary execution in the Philippines	186
5.4	Statistical power in hypothesis testing	188
5.4.1	Death penalty for rape	190
5.4.2	Is Bendectin a teratogen?	192
5.4.3	Automobile emissions and the Clean Air Act	193
5.5	Legal and statistical significance	194
5.5.1	Port Authority promotions	197
5.5.2	State trooper literacy exam	198

5.6	Maximum likelihood estimation	199
5.6.1	Purloined notices revisited	203
5.6.2	Do microwaves cause cancer?	203
5.6.3	Peremptory challenges of prospective jurors	204
6	Comparing Multiple Proportions	207
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
6.1.6	Harvard's affirmative action practices in admissions	217
6.2	Bonferroni's inequality and multiple comparisons	221
6.2.1	Wage additives and the four-fifths rule	223
6.2.2	Discretionary parole	224
6.2.3	Cheating on standardized multiple-choice tests revisited	225
6.2.4	InterMune's Actimmune	226
6.3	More measures of association: Phi-squared (ϕ^2) and tau B (τ_B)	228
6.3.1	Preventive detention	230
7	Comparing Means	235
7.1	Student's t -test: Hypothesis testing and confidence intervals	235
7.1.1	Automobile emissions and the Clean Air Act revisited	239
7.1.2	Voir dire of prospective trial jurors	241
7.1.3	Ballot position	242
7.1.4	Backdating stock options	242
7.1.5	Bullet lead identification	243
7.2	Analysis of variance for comparing several means	245
7.2.1	Fiddling debt collector	249
7.2.2	New York City Garbage Trucks	250
8	Combining Evidence Across Independent Strata	255
8.1	Mantel-Haenszel and Fisher methods for combining the evidence	255
8.1.1	Hiring lawyers	266
8.1.2	Age discrimination in employment terminations	267
8.1.3	Rheumatoid arthritis drug	267
8.1.4	Disparate Impact of a Pre-employment Exam on Minority Applicants	272