Springer Series in Statistics

Advisors:
P. Diggle, S. Fienberg, K. Krickeberg,
I. Olkin, N. Wermuth

Springer Series in Statistics

Andersen/Borgan/Gill/Keiding: Statistical Models Based on Counting Processes.
Andrews/Herzberg: Data: A Collection of Problems from Many Fields for the Student and Research Worker.
Anscombe: Computing in Statistical Science through APL.
Berger: Statistical Decision Theory and Bayesian Analysis, 2nd edition.
Bolfarine/Zacks: Prediction Theory for Finite Populations.
Brémaud: Point Processes and Queues: Martingale Dynamics.
Brockwell/Davis: Time Series: Theory and Methods, 2nd edition.
Daley/Vere-Jones: An Introduction to the Theory of Point Processes.
Dzhaparidze: Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.
Fahrmeir/Tutz: Multivariate Statistical Modelling Based on Generalized Linear Models.
Farrell: Multivariate Calculation.
Federer: Statistical Design and Analysis for Intercropping Experiments.
Fienberg/Hoaglin/Kruskal/Tanur (Eds.): A Statistical Model: Frederick Mosteller's Contributions to Statistics, Science and Public Policy.
Fisher/Sen: The Collected Works of Wassily Hoeffding.
Good: Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses.
Goodman/Kruskal: Measures of Association for Cross Classifications.
Grandell: Aspects of Risk Theory.
Hall: The Bootstrap and Edgeworth Expansion.
Härdle: Smoothing Techniques: With Implementation in S.
Hartigan: Bayes Theory.
Heyer: Theory of Statistical Experiments.
Jolliffe: Principal Component Analysis.
Kotz/Johnson (Eds.): Breakthroughs in Statistics Volume I.
Kotz/Johnson (Eds.): Breakthroughs in Statistics Volume II.
Kres: Statistical Tables for Multivariate Analysis.
Le Cam: Asymptotic Methods in Statistical Decision Theory.
Le Cam/Yang: Asymptotics in Statistics: Some Basic Concepts.
Longford: Models for Uncertainty in Educational Testing.
Manoukian: Modern Concepts and Theorems of Mathematical Statistics.
Miller, Jr.: Simultaneous Statistical Inference, 2nd edition.
Mosteller/Wallace: Applied Bayesian and Classical Inference: The Case of *The Federalist Papers.*
Pollard: Convergence of Stochastic Processes.
Pratt/Gibbons: Concepts of Nonparametric Theory.
Read/Cressie: Goodness-of-Fit Statistics for Discrete Multivariate Data.

(continued after index)

Nicholas T. Longford

Models for Uncertainty in Educational Testing

With 31 Figures

Springer-Verlag
New York Berlin Heidelberg London Paris
Tokyo Hong Kong Barcelona Budapest

Nicholas T. Longford
Research Division
15-T Educational Testing Service
Rosedale Road
Princeton, NJ 08541 USA

Library of Congress Cataloging-in-Publication Data
Longford, Nicholas T., 1955–
Models for uncertainty in educational testing / Nicholas T. Longford.
p. cm. – (Springer series in statistics)
Includes bibliographical references and index.
ISBN-13: 978-1-4613-8465-6 e-ISBN-13: 978-1-4613-8463-2
DOI: 10.1007/978-1-4613-8463-2
1. Educational tests and measurements. 2. Examinations—Validity. 3. Examinations—Interpretation. 4. Examinations—Design and construction. 5. Examinations—Scoring. 6. Educational Testing Service. I. Title. II. Series.
LB3051.L625 1995
371.2'6'013—dc20 95-8145

Printed on acid-free paper.

© 1995 Springer-Verlag New York, Inc.
Softcover reprint of the hardcover 1st edition 1995

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Production managed by Hal Henglein; manufacturing supervised by Jeffrey Taub.
Photocomposed pages prepared from the author's LaTeX file.

9 8 7 6 5 4 3 2 1

ISBN-13: 978-1-4613-8465-6

To Elyse and Kalman,
The children of the 'lesser' Dad

Preface

> *Empirical Bayes and James-Stein estimation. This closely related pair of ideas is the biggest theoretical success on my list and the biggest underachiever in number of practical applications. The irony here is that these methods potentially offer the biggest practical gains over their predecessors.*
>
> Bradley Efron, *The statistical century*[1]

The idea for this book came about with my growing realization that the Educational Testing Service (ETS) and its databanks are a genuine treasure trove of interesting statistical problems that are also encountered in many other fields. That in itself is a rather unoriginal statement, which can be found in one form or another in many other statistical texts motivated by educational testing issues. However, recent changes in the educational testing industry, especially the growing emphasis on constructed-response items, have brought about new statistical challenges and novel areas of application. This book is intended to be a tasting, rather than a complete and integral enumeration of these challenges, which are likely to be in a state of flux for a few years to come.

The title of the book, arrived at after umpteen revisions, is meant to have no other than statistical connotation. The term 'uncertainty' may

---

[1]Bradley Efron's agreement and permission of the Royal Statistical Society to reprint this passage from Efron (1995) are acknowledged.

sound somewhat negativist, but I hope that its appearance in the title will contribute to the emancipation of educational research from decades of futile attempts to find near-deterministic descriptions of educational phenomena with the aid of imperfect measurement instruments. I do not see the problem in the lack of perfection (when human subjects and observers are perfect and consistent, they are no longer human); it is rather in the reliance on statistical methods that assume perfect measurement and on cursory attention to problems of generalizability from one context (in which the study is conducted) to another (for which inference is desired). I find the term 'uncertainty' more appealing than 'variation' or *unexplained* variation to describe our inability to find a perfect explanation of an observed process.

The book is not intended as a celebration of ETS, but in part a piece of criticism, hopefully constructive, and principally a collection of case studies from educational testing. ETS has been generous in funding (in a piecemeal fashion but with good intentions) much of the research described in the book. I wish to express my gratitude to the research management for their patience and endurance with my criticism on a number of issues, and for giving me a go-ahead for this project. In brief, I acknowledge the flexible exercise of their authority. Where views are expressed, they are my own, unless stated otherwise, and ETS management has had little input in their formation. The book does not represent any aspects of official ETS policy, and I do not expect that it will do so at any time in the future.

One aspect in which my employer is beyond reproach is that I have always been granted leaves in response to invitations from academic institutions. I have done some of the work on this book while visiting the Zentrum für Umfragen, Methoden und Analysen in Mannheim, Germany (Michael Wiedenbeck), the Department of Statistics, University of Adelaide, Australia (Patty Solomon) and the Department of Economics, University of Pompeu Fabra in Barcelona, Spain (Albert Satorra). Their hospitality and stimulating environment are acknowledged. Research on the material presented in Chapter 7 was supported by a generous grant from the National Center for Educational Statistics.

The book would not have come about without the encouragement and exemplary cooperation of Martin Gilchrist of Springer-Verlag, who arranged for several detailed, informative, and insightful reviews of earlier drafts of the manuscript. I am grateful to several colleagues at ETS for their contribution to the book, in a variety of forms. Being mean by nature, I name only Eric Bradlow, Charlie Lewis, Gary Marco, Bob Mislevy, and Neal Thomas. I want to give credit to Bill Strawderman for several discussions and the resulting inspiration.

Those who provided the datasets analysed in the book are acknowledged in the text. Mike Wagner and Joe Bezek attended to the well-being of all the files on my Sun workstation. The production staff at Springer-Verlag helped to iron out the problems in wordprocessing and copy-editing.

My mother, Irena Lefkovičová, who introduced me many years ago to the charm and excitement of numbers, has been a staunch supporter throughout.

Princeton, NJ
April 1995

Nicholas T. Longford

Contents