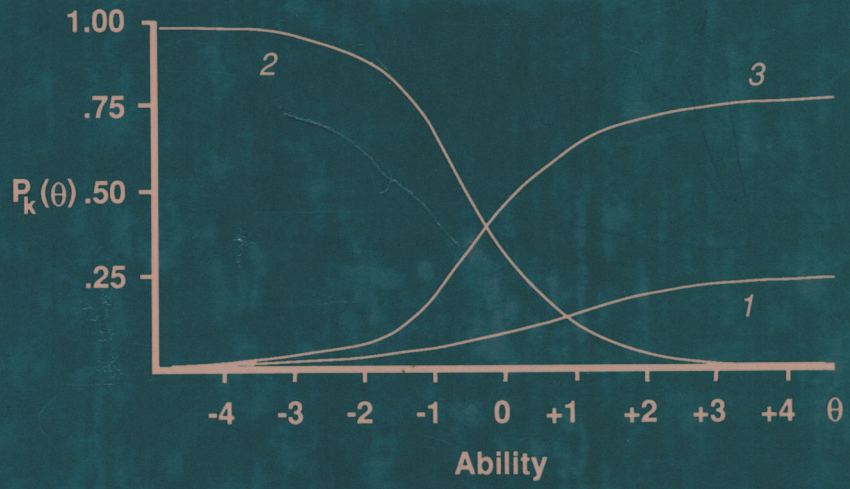
STATISTICS: textbooks and monographs volume 129

ITEM RESPONSE THEORY

Parameter Estimation Techniques



FRANK B. BAKER

ITEM RESPONSE THEORY

Parameter Estimation Techniques

FRANK B. BAKER
University of Wisconsin-Madison
Madison, Wisconsin

Marcel Dekker, Inc. New York • Basel • Hong Kong

Library of Congress Cataloging-in-Publication Data

Baker, Frank B.
Item response theory : parameter estimation techniques / Frank B. Baker.
p. cm. -- (Statistics, textbooks and monographs ; v. 129)
Includes bibliographical references and index.
ISBN 0-8247-8636-X (acid-free paper)
1. Parameter estimation. I. Title. II. Series.
QA276.8.B35 1992
519.5'44--dc20 92-3397
CIP

This book is printed on acid-free paper.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

ITEM RESPONSE THEORY

STATISTICS: Textbooks and Monographs

A Series Edited by

1. The Generalized Jackknife Statistic, *H. L. Gray and W. R. Schucany*
2. Multivariate Analysis, *Anant M. Kshirsagar*
3. Statistics and Society, *Walter T. Federer*
4. Multivariate Analysis: A Selected and Abstracted Bibliography, 1957–1972, *Kocherlakota Subrahmaniam and Kathleen Subrahmaniam*
5. Design of Experiments: A Realistic Approach, *Virgil L. Anderson and Robert A. McLean*
6. Statistical and Mathematical Aspects of Pollution Problems, *John W. Pratt*
7. Introduction to Probability and Statistics (in two parts), Part I: Probability; Part II: Statistics, *Narayan C. Giri*
8. Statistical Theory of the Analysis of Experimental Designs, *J. Ogawa*
9. Statistical Techniques in Simulation (in two parts), *Jack P. C. Kleijnen*
10. Data Quality Control and Editing, *Joseph I. Naus*
11. Cost of Living Index Numbers: Practice, Precision, and Theory, *Kali S. Banerjee*
12. Weighing Designs: For Chemistry, Medicine, Economics, Operations Research, Statistics, *Kali S. Banerjee*
13. The Search for Oil: Some Statistical Methods and Techniques, *edited by D. B. Owen*
14. Sample Size Choice: Charts for Experiments with Linear Models, *Robert E. Odeh and Martin Fox*
15. Statistical Methods for Engineers and Scientists, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
16. Statistical Quality Control Methods, *Irving W. Burr*

17. On the History of Statistics and Probability, *edited by D. B. Owen*
18. Econometrics, *Peter Schmidt*
19. Sufficient Statistics: Selected Contributions, *Vasant S. Huzurbazar (edited by Anant M. Kshirsagar)*
20. Handbook of Statistical Distributions, *Jagdish K. Patel, C. H. Kapadia, and D. B. Owen*
21. Case Studies in Sample Design, *A. C. Rosander*
22. Pocket Book of Statistical Tables, *compiled by R. E. Odeh, D. B. Owen, Z. W. Birnbaum, and L. Fisher*
23. The Information in Contingency Tables, *D. V. Gokhale and Solomon Kullback*
24. Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods, *Lee J. Bain*
25. Elementary Statistical Quality Control, *Irving W. Burr*
26. An Introduction to Probability and Statistics Using BASIC, *Richard A. Groeneveld*
27. Basic Applied Statistics, *B. L. Raktoe and J. J. Hubert*
28. A Primer in Probability, *Kathleen Subrahmaniam*
29. Random Processes: A First Look, *R. Syski*
30. Regression Methods: A Tool for Data Analysis, *Rudolf J. Freund and Paul D. Minton*
31. Randomization Tests, *Eugene S. Edgington*
32. Tables for Normal Tolerance Limits, Sampling Plans and Screening, *Robert E. Odeh and D. B. Owen*
33. Statistical Computing, *William J. Kennedy, Jr., and James E. Gentle*
34. Regression Analysis and Its Application: A Data-Oriented Approach, *Richard F. Gunst and Robert L. Mason*
35. Scientific Strategies to Save Your Life, *I. D. J. Bross*
36. Statistics in the Pharmaceutical Industry, *edited by C. Ralph Buncher and Jia-Yeong Tsay*
37. Sampling from a Finite Population, *J. Hajek*
38. Statistical Modeling Techniques, *S. S. Shapiro*
39. Statistical Theory and Inference in Research, *T. A. Bancroft and C.-P. Han*
40. Handbook of the Normal Distribution, *Jagdish K. Patel and Campbell B. Read*
41. Recent Advances in Regression Methods, *Hrishikesh D. Vinod and Aman Ullah*
42. Acceptance Sampling in Quality Control, *Edward G. Schilling*
43. The Randomized Clinical Trial and Therapeutic Decisions, *edited by Niels Tygstrup, John M Lachin, and Erik Juhl*
44. Regression Analysis of Survival Data in Cancer Chemotherapy, *Walter H. Carter, Jr., Galen L. Wampler, and Donald M. Stablein*
45. A Course in Linear Models, *Anant M. Kshirsagar*
46. Clinical Trials: Issues and Approaches, *edited by Stanley H. Shapiro and Thomas H. Louis*
47. Statistical Analysis of DNA Sequence Data, *edited by B. S. Weir*
48. Nonlinear Regression Modeling: A Unified Practical Approach, *David A. Ratkowsky*
49. Attribute Sampling Plans, Tables of Tests and Confidence Limits for Proportions, *Robert E. Odeh and D. B. Owen*
50. Experimental Design, Statistical Models, and Genetic Statistics, *edited by Klaus Hinkelmann*

51. Statistical Methods for Cancer Studies, *edited by Richard G. Cornell*
52. Practical Statistical Sampling for Auditors, *Arthur J. Wilburn*
53. Statistical Methods for Cancer Studies, *edited by Edward J. Wegman and James G. Smith*
54. Self-Organizing Methods in Modeling: GMDH Type Algorithms, *edited by Stanley J. Farlow*
55. Applied Factorial and Fractional Designs, *Robert A. McLean and Virgil L. Anderson*
56. Design of Experiments: Ranking and Selection, *edited by Thomas J. Santner and Ajit C. Tamhane*
57. Statistical Methods for Engineers and Scientists: Second Edition, Revised and Expanded, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*
58. Ensemble Modeling: Inference from Small-Scale Properties to Large-Scale Systems, *Alan E. Gelfand and Crayton C. Walker*
59. Computer Modeling for Business and Industry, *Bruce L. Bowerman and Richard T. O'Connell*
60. Bayesian Analysis of Linear Models, *Lyle D. Broemeling*
61. Methodological Issues for Health Care Surveys, *Brenda Cox and Steven Cohen*
62. Applied Regression Analysis and Experimental Design, *Richard J. Brook and Gregory C. Arnold*
63. Statpal: A Statistical Package for Microcomputers—PC-DOS Version for the IBM PC and Compatibles, *Bruce J. Chalmer and David G. Whitmore*
64. Statpal: A Statistical Package for Microcomputers—Apple Version for the II, II+, and IIe, *David G. Whitmore and Bruce J. Chalmer*
65. Nonparametric Statistical Inference: Second Edition, Revised and Expanded, *Jean Dickinson Gibbons*
66. Design and Analysis of Experiments, *Roger G. Petersen*
67. Statistical Methods for Pharmaceutical Research Planning, *Sten W. Bergman and John C. Gittins*
68. Goodness-of-Fit Techniques, *edited by Ralph B. D'Agostino and Michael A. Stephens*
69. Statistical Methods in Discrimination Litigation, *edited by D. H. Kaye and Mikel Aickin*
70. Truncated and Censored Samples from Normal Populations, *Helmut Schneider*
71. Robust Inference, *M. L. Tiku, W. Y. Tan, and N. Balakrishnan*
72. Statistical Image Processing and Graphics, *edited by Edward J. Wegman and Douglas J. DePriest*
73. Assignment Methods in Combinatorial Data Analysis, *Lawrence J. Hubert*
74. Econometrics and Structural Change, *Lyle D. Broemeling and Hiroki Tsurumi*
75. Multivariate Interpretation of Clinical Laboratory Data, *Adelin Albert and Eugene K. Harris*
76. Statistical Tools for Simulation Practitioners, *Jack P. C. Kleijnen*
77. Randomization Tests: Second Edition, *Eugene S. Edgington*
78. A Folio of Distributions: A Collection of Theoretical Quantile-Quantile Plots, *Edward B. Fowlkes*
79. Applied Categorical Data Analysis, *Daniel H. Freeman, Jr.*

80. Seemingly Unrelated Regression Equations Models: Estimation and Inference, *Virendra K. Srivastava and David E. A. Giles*
81. Response Surfaces: Designs and Analyses, *Andre I. Khuri and John A. Cornell*
82. Nonlinear Parameter Estimation: An Integrated System in BASIC, *John C. Nash and Mary Walker-Smith*
83. Cancer Modeling, *edited by James R. Thompson and Barry W. Brown*
84. Mixture Models: Inference and Applications to Clustering, *Geoffrey J. McLachlan and Kaye E. Basford*
85. Randomized Response: Theory and Techniques, *Arijit Chaudhuri and Rahul Mukerjee*
86. Biopharmaceutical Statistics for Drug Development, *edited by Karl E. Peace*
87. Parts per Million Values for Estimating Quality Levels, *Robert E. Odeh and D. B. Owen*
88. Lognormal Distributions: Theory and Applications, *edited by Edwin L. Crow and Kunio Shimizu*
89. Properties of Estimators for the Gamma Distribution, *K. O. Bowman and L. R. Shenton*
90. Spline Smoothing and Nonparametric Regression, *Randall L. Eubank*
91. Linear Least Squares Computations, *R. W. Farebrother*
92. Exploring Statistics, *Damaraju Raghavarao*
93. Applied Time Series Analysis for Business and Economic Forecasting, *Sufi M. Nazem*
94. Bayesian Analysis of Time Series and Dynamic Models, *edited by James C. Spall*
95. The Inverse Gaussian Distribution: Theory, Methodology, and Applications, *Raj S. Chhikara and J. Leroy Folks*
96. Parameter Estimation in Reliability and Life Span Models, *A. Clifford Cohen and Betty Jones Whitten*
97. Pooled Cross-Sectional and Time Series Data Analysis, *Terry E. Dielman*
98. Random Processes: A First Look, Second Edition, Revised and Expanded, *R. Syski*
99. Generalized Poisson Distributions: Properties and Applications, *P. C. Consul*
100. Nonlinear L_p-Norm Estimation, *Rene Gonin and Arthur H. Money*
101. Model Discrimination for Nonlinear Regression Models, *Dale S. Borowiak*
102. Applied Regression Analysis in Econometrics, *Howard E. Doran*
103. Continued Fractions in Statistical Applications, *K. O. Bowman and L. R. Shenton*
104. Statistical Methodology in the Pharmaceutical Sciences, *Donald A. Berry*
105. Experimental Design in Biotechnology, *Perry D. Haaland*
106. Statistical Issues in Drug Research and Development, *edited by Karl E. Peace*
107. Handbook of Nonlinear Regression Models, *David A. Ratkowsky*
108. Robust Regression: Analysis and Applications, *edited by Kenneth D. Lawrence and Jeffrey L. Arthur*
109. Statistical Design and Analysis of Industrial Experiments, *edited by Subir Ghosh*
110. *U*-Statistics: Theory and Practice, *A. J. Lee*
111. A Primer in Probability: Second Edition, Revised and Expanded, *Kathleen Subrahmaniam*

112. Data Quality Control: Theory and Pragmatics, *edited by Gunar E. Liepins and V. R. R. Uppuluri*
113. Engineering Quality by Design: Interpreting the Taguchi Approach, *Thomas B. Barker*
114. Survivorship Analysis for Clinical Studies, *Eugene K. Harris and Adelin Albert*
115. Statistical Analysis of Reliability and Life-Testing Models: Second Edition, *Lee J. Bain and Max Engelhardt*
116. Stochastic Models of Carcinogenesis, *Wai-Yuan Tan*
117. Statistics and Society: Data Collection and Interpretation: Second Edition, Revised and Expanded, *Walter T. Federer*
118. Handbook of Sequential Analysis, *B. K. Ghosh and P. K. Sen*
119. Truncated and Censored Samples: Theory and Applications, *A. Clifford Cohen*
120. Survey Sampling Principles, *E. K. Foreman*
121. Applied Engineering Statistics, *Robert M. Bethea and R. Russell Rhinehart*
122. Sample Size Choice: Charts for Experiments with Linear Models: Second Edition, *Robert E. Odeh and Martin Fox*
123. Handbook of the Logistic Distribution, *edited by N. Balakrishnan*
124. Fundamentals of Biostatistical Inference, *Chap T. Le*
125. Correspondence Analysis Handbook, *J.-P. Benzécri*
126. Quadratic Forms in Random Variables: Theory and Applications, *A. M. Mathai and Serge B. Provost*
127. Confidence Intervals on Variance Components, *Richard K. Burdick and Franklin A. Graybill*
128. Biopharmaceutical Sequential Statistical Applications, *edited by Karl E. Peace*
129. Item Response Theory: Parameter Estimation Techniques, *Frank B. Baker*

Additional Volumes in Preparation

Nonparametric Statistical Inference: Third Edition, Revised and Expanded, *by Jean Dickinson Gibbons and Subhabrata Chakraborti*

Design and Analysis of Bioavailability and Bioequivalence Studies, *by Shein-Chung Chow and Jen-Pei Liu*

Survey Sampling: Theory and Methods, *by Arijit Chaudhuri and Horst Stenger*

Bivariate Discrete Distribution, *by Subrahmaniam Kocherlakota and Kathleen Kocherlakota*

To Dr. Frederic M. Lord

who, over a thirty-year period, shepherded item response theory from its infancy to its acceptance as the modern test theory

Preface

The process by which a field of inquiry acquires its name is an interesting one. Often a label is offered, such as Criterion Referenced Testing, which is universally accepted within a short time. In most cases, a field develops slowly and many different labels are employed before consensus is reached as to an acceptable name. Such is the case for the modern test theory that is the focus of this book. The basic concepts of the field have been developed across three-fourths of a century, and a diverse group of persons have made contributions. For many years there was little agreement as to what label should be employed to identify the field. Much of the literature in the 1950s through the 1970s used the term latent trait theory, as it reflected the use of an underlying hypothetical variable. The latent trait terminology is due to Lazarsfeld (1954) and carries with it connotations of psychological scaling procedures. For a short period, Lord (1977) referred to the field as item characteristic curve theory in recognition of the major role of such curves. While item characteristic curve theory is an appropriate label, it tends to be associated with those concepts and procedures based upon dichotomously scored items. The use of this label becomes a bit strained when items are scored on a graded or nominal basis as well as when used within the context of other recent developments. In recent years, the label item response theory, also due to Lord (1980), has gained acceptance as the name of the field since it reflects the dependence of the theory upon an examinee's responses to items. This label has sufficient

generality to encompass the many facets of the field as well as to reflect the basic concepts involved. It is also true that the degree to which a set of concepts and procedures are accepted in practice is often determined by their label and this particular title has meaning to both the researcher and the practitioner.

One of the salient features of item response theory is that it is at once very old and very new even though it is often referred to as the "modern" test theory. The chapters by Allen Birnbaum in Lord and Novick's classic *Statistical Theories of Mental Test Scores* (1968) made item response theory visible. Although the basic theory has been available for a generation, only within the past ten years has a significant level of publication been achieved. A large number of articles dealing with the theory and application of item response theory have appeared in measurement-oriented journals. A key publication was the September 1977 issue of the *Journal of Educational Measurement* under the editorship of Lori Shepard. This issue spanned the range from theoretical concerns to applications. Many basic papers have appeared in *Psychometrika* and much of the research exploring and expanding the theory has appeared in the journal *Applied Psychological Measurement*. In addition to articles, a number of books have appeared: *Item Response Theory: Applications to Psychological Measurement* (Hulein, Draskow, and Parsons, 1983), *The Proceedings of the Montreaux Symposium* (De Gruijter and Vander Kamp, 1976), and *Einführung in die Theorie Psychologisher Tests* (Fischer, 1974). Much of the interest in item response theory is due to the efforts of Benjamin Wright. The American Educational Research Association presessions conducted by Dr. Wright over a 20-year period and the book *Best Test Design* (Wright and Stone, 1979) has made the theory accessible to practitioners. A significant event in the history of item response theory was the publication of Frederic Lord's book *Practical Applications of Item Response Theory* (1980). This book collects into a single source the contents of his numerous Educational Testing Service Research memoranda and papers that have appeared in a variety of places. Because of Lord's pivotal role within the field, this book has had a considerable impact on the use of the theory in applied settings. The book *Applications of Item Response Theory* (Hambelton, 1983) also deals with the use of the theory in practice. Another important book is *Item Response Theory* (Hambelton and Swaminathan, 1985), which provides broad coverage of the whole of the theory. It relates classical test theory and item response theory under the assumption that the reader is familiar with both. The book by Baker (1985) provides an introduction to the basic concepts of item response theory and is unique in that it has an accompanying computer program that allows one to explore the facets of the theory. Since there are a variety of sources for learning about IRT as a

test theory, the present book assumes the reader is familiar with the basic constructs.

One result of the available books and articles is that as a field of inquiry item response theory has become very large. This in turn makes it difficult for a single book to provide in-depth coverage of the whole field. As a result, there is a need for more specialized books dealing with specific aspects of the theory. Because of the mathematical sophistication and the data processing demands of item response theory test analysis procedures, computer program implementation is necessary. Thus, from a practical point of view, the history of item response theory is inextricably intertwined with the available computer programs such as LOGIST, BICAL, BILOG, and MULTILOG. Proper specification of the analyses to be performed and the interpretation of the outputs of these programs depends upon an intimate understanding of the parameter estimation procedures they implement. Thus I believed that a book dealing only with these procedures would be a valuable resource for both those who use item response theory in practice and those who need a basic understanding of these procedures.

The implementation of item response theory rests on the statistical techniques for estimating the parameters of test items and of examinee ability. However, in striving to present the full panorama of the theory, the existing books have necessarily provided only the outlines of these estimation procedures. Thus, presentation of the underlying logic and the mathematical details of these estimation procedures would complement the existing broader coverage. A unification of the existing literature would also be achieved by presenting the mathematics of these estimation procedures in a consistent manner. Finally, providing the full mathematical detail of these estimation procedures would make them accessible to a wider audience.

To meet these goals, this book has been organized in a systematic manner. In the first chapter, the underlying logic of the item characteristic curve concept and the several models for these curves are presented. Chapters 2 and 3 develop the two building blocks from which several other estimation paradigms are constructed. Chapter 2 shows the mathematics of the maximum likelihood procedures for estimating the parameters of the item characteristic curve under the various models. Chapter 3 presents the estimation of an examinee's ability score under these models. In Chapter 4, these two procedures are put together in the Birnbaum paradigm for the joint maximum likelihood estimation of item and ability parameters that is the basis for several widely employed computer programs. Owing to its importance, a separate chapter was devoted to the Rasch model. His original work is presented in a notation consistent with the rest of the book. The estimation procedures developed by Wright and his co-workers are shown, as is the extension of the Rasch model in the form of the linear

logistic test model. Chapter 6 provides the mathematical details of the Bock and Aitkin marginal maximum likelihood/EM approach to the estimation of item and examinee parameters. Chapter 7 presents the marginalized Bayesian approach to estimating item parameters and two Bayesian approaches to estimating an examinee's ability. Chapters 8 and 9 are concerned with the maximum likelihood estimation procedures for parameter estimation when graded and nominally scored items are employed. The BASIC computer programs in the appendixes illustrate how the various estimation procedures are implemented. I hope that the approach taken provides a gradual increase in sophistication of the material such that the reader can progress easily from the basics to the complex estimation procedures. In addition, it should provide a context within which to place further developments.

The lack of a standardized notation within the item response theory literature posed a particularly difficult problem in writing this book. Only a few symbols have been used consistently and the literature has been very lax in differentiating between parameters and their sample estimators. In the early chapters, an attempt has been made to present the mathematics using a symbol system that corresponds to normal statistical usage in this regard. In the later chapters, the symbol system is more closely linked to that used in the item response theory literature. This was done to provide better access to the mathematics of the original articles. In addition, it was occasionally necessary to attach different meanings to the same symbols used in the early and later chapters. I trust that by later chapters the reader will have developed some notational flexibility. Nonetheless, a serious effort was made to make the notation as consistent as possible.

The writing of the first draft of this book was supported by a grant from the Research Committee of the Graduate School, University of Wisconsin. The typing from my atrocious handwriting was done by the secretaries of the Laboratory of Experimental Design: Connie Schlehammer, Beth Brown, Cathy Tobin, Chris Kringle, and Cheryl Houge. The difficult task of converting the manuscript from the NBI word processor to WordPerfect on the IBM PC computer was accomplished with great skill by Lauri Koch, who also typed the final version of the manuscript. Without the skill and diligence of these persons, the book would not have been possible. The contributions of several generations of graduate students are also acknowledged. Their probing questions revealed many places where the manuscript was unclear and in some cases incorrect. They contributed much to the final version. Chapters 6 and 7 were the result of a collaboration with Dr. Micheal Harwell of the University of Pittsburgh in which the mysteries of marginal maximum likelihood and Bayesian estimation were unraveled.

Frank B. Baker

Contents