Reviews of the book by Patrick Royston and Willi Sauerbrei

Multivariable Model-building: A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables

book's website, www.imbi.uni-freiburg.de/biom/Royston-Sauerbrei-book

To give a flavour of the book and the positive comments given in the reviews we cite parts of them.

DAVID W. HOSMER,
Division of Biostatistics and Epidemiology, University of Massachusetts, Amherst, Massachusetts, USA

BIOMETRICS (2009) 65: 989–1001

One might be tempted to conclude from the title that the book is a monograph dealing with a rather narrow topic. While modeling with fractional polynomials is the primary focus, **the book is more about model building in general, with an emphasis on techniques for modeling continuous covariates**.

The **reference section is extensive** and includes the pages in the text where each reference is cited. A small but nice feature is that each chapter begins witha concise summary of its contents. All data used in the book may be downloaded from the book's website, www.imbi.uni-freiburg.de/biom/Royston-Sauerbrei-book, in four formats: SAS, Stata, Excel, and ASCII. The website also provides Powerpoint slides from talks given by the authors on most aspects of the material covered in the book.

So Royston and Sauerbrei are to be commended for preparing such an excellent and useful website. It will undoubtedly enhance adoption of the book and use of the methods in general statistical practice.

In summary, I think the book does an excellent job presenting what fractional polynomials are and how to use them to model continuous covariates in regression models. The authors are to be commended for giving a thorough treatment of factors that can have an adverse effect on their application. As I mentioned earlier, the book is more than its title, **it really is a treatise on how to model data** by two experienced and competent analysts. The book could be used as a text in an intermediate to advanced applied course on regression modeling and I highly recommend it to applied statisticians who want to learn about fractional polynomials and how to use them.

MICHAEL SCHEMPER,
Section of Clinical Biometrics, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Vienna, Austria

STATISTICS IN MEDICINE (2009), 28:537–539

This excellent book **fills a gap in the current literature on statistical modelling**. It is the first time that a book is devoted to the whole breadth of application of fractional polynomials (FP). The authors are *the* experts on this useful methodology and had substantial impact on its development in numerous papers.

The style of the book makes reading a pleasure: **as easy as possible for a text in statistical modelling while not avoiding more difficult issues**. Therefore, it is **suitable for readers with a moderate knowledge of theory and practice of regression models**, e.g. for graduate students, as it is interesting for experienced practicing statisticians. As the authors say in the Preface, '. . . the general objective is to provide a readable text giving the rationale of, and practical advice on, a unified approach to multivariable modelling which aims to make such models simpler and more effective'.

The detailed Contents of 9 pages, an itemized summary for (almost) each chapter at its beginning and a discussion at its end, many examples from 23 well-documented data sets (which can all be downloaded from the accompanying

book web site), references giving the page(s) of their citation, and a **carefully prepared subject index** not only testify to the meticulous work of the authors, but also **guarantee that readers can use this book very efficiently** whether for learning or for reference.

There is little to be criticized. Some readers will possibly miss a more detailed treatment of some peripheral issues to the main topic, such as incomplete data, explained variation, dependent data or other topics, but more consideration of such issues would have made the book daunting heavy and less focused. By way of conclusion, this very useful book can be highly recommended to the readers of this journal.

CHRIS METCALFE,
Department of Social Medicine, University of Bristol, Bristol, UK

STATISTICAL METHODS IN MEDICAL RESEARCH (2010), 19: 197–199

This book is a **model of how to present complex material in a way that is accessible to, and will influence the practice of a wide readership of quantitative researchers**. For the reader who already has some expertise in variable selection or fractional polynomials, this book brings together the recent methodological work on advanced topics with plentiful examples, and does this in a way that identifies areas for further research. Both sets of readers will find the book's website a valuable resource, providing the example datasets, example chapters from the book, associated teaching materials and Stata code (R code is available elsewhere) (http://www.imbi.uni-freiburg.de/biom/Royston-Sauerbrei-book/).

I have recommended this book to PhD students in epidemiology and statistics, and have adopted it as further reading for our short course in regression methods for postgraduate medical researchers. This book is an enjoyable read that **both inspires the reader to work harder at building useful and sensible regression models, and removes any excuse for falling back on inefficient old practices**.

WILLIAM D. DUPONT,
Department of Biostatistics, Vanderbilt University School of Medicine, Nashville, Tennessee, USA
william.dupont@vanderbilt.edu

THE STATA JOURNAL (2010), 10(2): 297–302

Royston and Sauerbrei (2008) provide an **excellent introduction to building nonlinear regression models**. Their preferred approach is to use fractional polynomial models, a technique that they have largely developed and on which they are undisputed authorities (Royston and Altman 1994; Sauerbrei and Royston 1999). This technique is applicable to models in which the response variable is continuous or dichotomous, to survival models, and to any generalized linear model.

The text is full of practical examples that Royston and Sauerbrei use to illustrate their model-building approach. They make **extensive use of smoothed residual plots to evaluate their models and to guide model selection**. Their approach is suitable for epidemiological studies in which the number of observations is at least an order of magnitude larger than the number of model covariates. For such data, the problems of multiple comparisons and the overfitting of models are not an overwhelming concern. The problem of model-building for genomic or other studies in which the number ofcovariates greatly exceeds the number of study subjects is not addressed in the text.

The fact that these models all involve linear combinations of the model parameters means that **the extensive theory of generalized linear models applies and that many sophisticated programs for building and evaluating such models are already available.**

This book's **greatest strength is its lucid writing style and practical guidance on how to build complex multivariable models**. It contains many examples with models that are extensively evaluated using smoothed residual plots and partial predictor plots. I was also intrigued by Royston and Sauerbrei's approach to automated model fitting. Many statisticians have a negative attitude toward such methods because of the risk of overfitting and because of multiple-comparisons concerns (Harrell 2001). I found Royston and Sauerbrei's **emphasis on bootstrap analyses to assess model stability to be reassuring. Software to implement their methods is available in Stata**, either as part of Stata 11 or as user-contributed programs that can be downloaded over the Internet. Weaknesses are few.

This is a very well-written book that provides a thoughtful approach to fitting nonlinear models. The authors are very experienced biostatisticians who have worked extensively in observational and experimental medical science. They make a **convincing argument that fractional polynomials can be a valuable tool for building such models in many situations**. Their many examples and excellent illustrations make their book accessible to a broad audience within statistical science. Their software will make this book particularly useful to the Stata community. I highly recommend this text.

REVIEWS ON AMAZON WEBSITES

All the following reviewers on Amazon rated the book with 5 stars – the highest possible rating.

STEPHEN EVANS,
The London School of Hygiene & Tropical Medicine, London, UK
AMAZON.CO.UK
The authors are experts in this field, but their explanations are clear and many detailed examples are given. The techniques explained in the book should be known by every statistician who actually analyses data, and the methods should be taught, certainly at a postgraduate if not at undergraduate level. It would be a good course text for teaching regression methods which are an essential component of statistical analysis.

T. MORRIS,
London, UK
AMAZON.CO.UK
Deriving a satisfactory multivariable model from several predictors can be a daunting task to researchers and less experienced statisticians. This book helps you to achieve satisfactory models through an integrated and principled approach to model-building.
Despite examples being mainly medical, methods are applicable to many other areas.
I have found this book very helpful and clear, and strongly recommend it to anyone planning research involving multivariable models.

HUGO FONTAN KOHLER
Campinas, Brazil
AMAZON.COM

I do believe this is the [best?] book statistical book I've ever read. Although a physician by formation, I started enjoying medical statistics since the end of college graduation. This book allowed me to better understand clinical models and to develop my own. I recommend it to anyone with interest in medical research.

E. NUNEZ
Pennsylvania, USA
AMAZON.COM

[This book] is a textbook appropriate for clinical trialists and researchers in biomedical science, who are engaged in a daily basis in building multivariable prognostic models. While a number of texts exist for these audiences, this textbook was written at a more accessible level for those without any advanced statistical knowledge.

The book begins laying the groundwork for model building as well as underlying the importance of modeling the appropriate functional form of continuous variables in order to minimize any residual confounding. Moreover, the authors have impressive coverage of the different statistical methodologies available for evaluating the functional

form of continuous variables, like fractional polynomials and restrictive cubic splines. The advantages and drawbacks for each of these methods are also provided.

The authors make use of statistical software packages throughout to demonstrate many of statistical procedures it presents. STATA is the primary package used and many of the necessary commands are given for readers who wish to take advantage of software to assist them in analyzing data. In addition to STATA, the authors provided the link for resources in SAS and R. For survival analysis, STATA is the primary package used with details given on how to implement the methods covered in the text.

This is an amazing book! As a physician-epidemiologist, I strongly recommend it. If you want to excel in multivariable model building, you will not be disappointed.