

# Statistics and Computing

*Series Editors:*

J. Chambers

W. Eddy

W. Härdle

S. Sheather

L. Tierney

**Springer**

*New York*

*Berlin*

*Heidelberg*

*Barcelona*

*Hong Kong*

*London*

*Milan*

*Paris*

*Singapore*

*Tokyo*

# Statistics and Computing

---

*Gentle*: Numerical Linear Algebra for Applications in Statistics.

*Gentle*: Random Number Generation and Monte Carlo Methods.

*Härdle/Klinke/Turlach*: XploRe: An Interactive Statistical Computing Environment.

*Krause/Olson*: The Basics of S and S-PLUS, 2nd Edition.

*Lange*: Numerical Analysis for Statisticians.

*Loader*: Local Regression and Likelihood.

*Ó Ruanaidh/Fitzgerald*: Numerical Bayesian Methods Applied to Signal Processing.

*Pannatier*: VARIOWIN: Software for Spatial Data Analysis in 2D.

*Pinheiro/Bates*: Mixed-Effects Models in S and S-PLUS.

*Venables/Ripley*: Modern Applied Statistics with S-PLUS, 3rd Edition.

*Venables/Ripley*: S Programming.

*Wilkinson*: The Grammar of Graphics.

José C. Pinheiro  
Douglas M. Bates

# Mixed-Effects Models in S and S-PLUS

With 172 Illustrations



Springer

José C. Pinheiro  
Bell Laboratories  
Lucent Technologies  
700 Mountain Avenue  
Murray Hill, NJ 07974-0636  
USA  
jcp@bell-labs.com

Douglas M. Bates  
Department of Computer Sciences  
and Statistics  
University of Wisconsin  
Madison, WI 53706-1685  
USA  
bates@stat.wisc.edu

*Series Editors:*

J. Chambers  
Bell Labs, Lucent  
Technologies  
600 Mountain Ave.  
Murray Hill, NJ 07974  
USA

W. Eddy  
Department of Statistics  
Carnegie Mellon University  
Pittsburgh, PA 15213  
USA

W. Härdle  
Institut für Statistik und  
Ökonometrie  
Humboldt-Universität zu Berlin  
Spandauer Str. 1  
D-10178 Berlin  
Germany

S. Sheather  
Australian Graduate School  
of Management  
University of New South  
Wales  
Sydney NSW 2052  
Australia

L. Tierney  
School of Statistics  
University of Minnesota  
Vincent Hall  
Minneapolis, MN 55455  
USA

Library of Congress Cataloging-in-Publication Data

Pinheiro, José C.

Mixed-effects models in S and S-PLUS / José C. Pinheiro, Douglas M. Bates  
p. cm. — (Statistics and computing)

Includes bibliographical references and index.

ISBN 0-387-98957-9 (alk. paper)

I. Bates, Douglas M. II. Title. III. Series.

QA76.73.S15 P56 2000

005.13'3—dc21

99-053566

© 2000 Springer-Verlag New York, Inc.

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.

TO ELISA AND LAURA

TO MARY ELLEN, BARBARA, AND MICHAEL

# Preface

Mixed-effects models provide a flexible and powerful tool for the analysis of grouped data, which arise in many areas as diverse as agriculture, biology, economics, manufacturing, and geophysics. Examples of grouped data include longitudinal data, repeated measures, blocked designs, and multilevel data. The increasing popularity of mixed-effects models is explained by the flexibility they offer in modeling the within-group correlation often present in grouped data, by the handling of balanced and unbalanced data in a unified framework, and by the availability of reliable and efficient software for fitting them.

This book provides an overview of the theory and application of linear and nonlinear mixed-effects models in the analysis of grouped data. A unified model-building strategy for both linear and nonlinear models is presented and applied to the analysis of over 20 real datasets from a wide variety of areas, including pharmacokinetics, agriculture, and manufacturing. A strong emphasis is placed on the use of graphical displays at the various phases of the model-building process, starting with exploratory plots of the data and concluding with diagnostic plots to assess the adequacy of a fitted model. Over 170 figures are included in the book.

The class of mixed-effects models considered in this book assumes that both the random effects and the errors follow Gaussian distributions. These models are intended for grouped data in which the response variable is (at least approximately) continuous. This covers a large number of practical applications of mixed-effects models, but does not include, for example, generalized linear mixed-effects models (Diggle, Liang and Zeger, 1994).

The balanced mix of real data examples, modeling software, and theory makes this book a useful reference for practitioners who use, or intend to use, mixed-effects models in their data analyses. It can also be used as a text for a one-semester graduate-level applied course in mixed-effects models. Researchers in statistical computing will also find this book appealing for its presentation of novel and efficient computational methods for fitting linear and nonlinear mixed-effects models.

The `nlme` library we developed for analyzing mixed-effects models in implementations of the S language, including S-PLUS and R, provides the underlying software for implementing the methods presented in the text, being described and illustrated in detail throughout the book. All analyses included in the book were produced using version 3.1 of `nlme` with S-PLUS 3.4 running on an Iris 5.4 Unix platform. Because of platform dependencies, the analysis results may be expected to vary slightly with different computers or operating systems and with different implementations of S. Furthermore, the current version of the `nlme` library for R does not support the same range of graphics presentations as does the S-PLUS version. The latest version of `nlme` and further information on the NLME project can be obtained at

<http://nlme.stat.wisc.edu> or

<http://cm.bell-labs.com/stat/NLME>.

Errata and updates of the material in the book will be made available on-line at the same sites.

The book is divided into parts. Part I, comprising five chapters, is dedicated to the linear mixed-effects (LME) model and Part II, comprising three chapters, covers the nonlinear mixed-effects (NLME) model. Chapter 1 gives an overview of LME models, introducing some examples of grouped data and the type of analyses that applies to them. The theory and computational methods for LME models are the topics of Chapter 2. Chapter 3 describes the structure of grouped data and the many facilities available in the `nlme` library to display and summarize such data. The model-building approach we propose is described and illustrated in detail in the context of LME models in Chapter 4. Extensions of the basic LME model to include variance functions and correlation structures for the within-group errors are considered in Chapter 5. The second part of the book follows an organization similar to the first. Chapter 6 provides an overview of NLME models and some of the analysis tools available for them in `nlme`. The theory and computational methods for NLME models are described in Chapter 7. The final chapter is dedicated to model building in the context of NLME models and to illustrating in detail the nonlinear modeling facilities available in the `nlme` library.

Even though the material covered in the book is, for the most part, self-contained, we assume that the reader has some familiarity with linear regression models, say at the level of Draper and Smith (1998). Although enough theory is covered in the text to understand the strengths and weak-

nesses of mixed-effects models, we emphasize the applied aspects of these. Readers who desire to learn in more detail the theory of mixed-effects are referred to the excellent book by Davidian and Giltinan (1995). Some knowledge of the S language is definitely desirable, but not a pre-requisite for following the material in the book. For those who are new to, or less familiar with S, we suggest using in conjunction with this book the, by now, *classic* reference Venables and Ripley (1999), which provides an overview of S and an introduction to a wide variety of statistical models to be used with S.

The authors may be contacted via electronic mail at

jcp@research.bell-labs.com  
bates@stat.wisc.edu

and would appreciate being informed of typos, errors, and improvements to the contents of this book.

### *Typographical Conventions:*

The S language objects and commands referenced throughout the book are printed in a monospaced typewriter font `like this`, while the S classes are printed in sans-serif font *like this*. The standard prompt `>` is used for S commands and the prompt `+` is used to indicate continuation lines.

To save space, some of the S output has been edited. Omission of complete lines are usually indicated by

. . .

but some blank lines have been removed without indication. The S output was generated using the options settings

```
> options( width = 68, digits = 5 )
```

The default settings are for 80 and 7, respectively.

### *Acknowledgements:*

This book would not exist without the help and encouragement of many people with whom we have interacted over the years. We are grateful to the people who have read and commented on earlier versions of the book; their many useful suggestions have had a direct impact on the current organization of the book. Our thanks also go to the many people who have tested and made suggestions on the `nlme` library, in particular the beta testers for the current version of the software. It would not be possible to name all these people here, but in particular we would like to thank John Chambers, Yonghua Chen, Bill Cleveland, Saikat DebRoy, Ramón Díaz-Uriarte, David James, Diane Lambert, Renaud Lancelot, David Lansky, Brian Ripley, Elisa Santos, Duncan Temple Lang, Silvia Vega, and Bill



Venables. Finally, we would like to thank our editor, John Kimmel, for his continuous encouragement and support.

José C. Pinheiro  
Douglas M. Bates  
March 2000

# Contents

<b>Preface</b>	<b>vii</b>
<b>I Linear Mixed-Effects Models</b>	<b>1</b>
<b>1 Linear Mixed-Effects Models</b>	<b>3</b>
1.1 A Simple Example of Random Effects . . . . .	4
1.1.1 Fitting the Random-Effects Model With <code>lme</code> . . . .	8
1.1.2 Assessing the Fitted Model . . . . .	11
1.2 A Randomized Block Design . . . . .	12
1.2.1 Choosing Contrasts for Fixed-Effects Terms . . . .	14
1.2.2 Examining the Model . . . . .	19
1.3 Mixed-Effects Models for Replicated, Blocked Designs . . .	21
1.3.1 Fitting Random Interaction Terms . . . . .	23
1.3.2 Unbalanced Data . . . . .	25
1.3.3 More General Models for the Random Interaction Effects . . . . .	27
1.4 An Analysis of Covariance Model . . . . .	30
1.4.1 Modeling Simple Linear Growth Curves . . . . .	30
1.4.2 Predictions of the Response and the Random Effects	37
1.5 Models for Nested Classification Factors . . . . .	40
1.5.1 Model Building for Multilevel Models . . . . .	44
1.6 A Split-Plot Experiment . . . . .	45
1.7 Chapter Summary . . . . .	52
Exercises . . . . .	52

<b>2</b>	<b>Theory and Computational Methods for LME Models</b>	<b>57</b>
2.1	The LME Model Formulation . . . . .	58
2.1.1	Single Level of Grouping . . . . .	58
2.1.2	A Multilevel LME Model . . . . .	60
2.2	Likelihood Estimation for LME Models . . . . .	62
2.2.1	The Single-Level LME Likelihood Function . . . . .	62
2.2.2	Orthogonal-Triangular Decompositions . . . . .	66
2.2.3	Evaluating the Likelihood Through Decompositions . . . . .	68
2.2.4	Components of the Profiled Log-Likelihood . . . . .	71
2.2.5	Restricted Likelihood Estimation . . . . .	75
2.2.6	Multiple Levels of Random Effects . . . . .	77
2.2.7	Parameterizing Relative Precision Factors . . . . .	78
2.2.8	Optimization Algorithms . . . . .	79
2.3	Approximate Distributions . . . . .	81
2.4	Hypothesis Tests and Confidence Intervals . . . . .	82
2.4.1	Likelihood Ratio Tests . . . . .	83
2.4.2	Hypothesis Tests for Fixed-Effects Terms . . . . .	87
2.4.3	Confidence Intervals . . . . .	92
2.5	Fitted Values and Predictions . . . . .	94
2.6	Chapter Summary . . . . .	94
	Exercises . . . . .	96
<b>3</b>	<b>Describing the Structure of Grouped Data</b>	<b>97</b>
3.1	The Display Formula and Its Components . . . . .	97
3.2	Constructing <code>groupedData</code> Objects . . . . .	101
3.2.1	Roles of Other Experimental or Blocking Factors . . . . .	104
3.2.2	Constructors for Balanced Data . . . . .	108
3.3	Controlling Trellis Graphics Presentations of Grouped Data . . . . .	110
3.3.1	Layout of the Trellis Plot . . . . .	110
3.3.2	Modifying the Vertical and Horizontal Scales . . . . .	113
3.3.3	Modifying the Panel Function . . . . .	114
3.3.4	Plots of Multiply-Nested Data . . . . .	116
3.4	Summaries . . . . .	120
3.5	Chapter Summary . . . . .	130
	Exercises . . . . .	130
<b>4</b>	<b>Fitting Linear Mixed-Effects Models</b>	<b>133</b>
4.1	Fitting Linear Models in <code>S</code> with <code>lm</code> and <code>lmList</code> . . . . .	134
4.1.1	The <code>lmList</code> Function . . . . .	139
4.2	Fitting Linear Mixed-Effects Models with <code>lme</code> . . . . .	146
4.2.1	Fitting Single-Level Models . . . . .	146
4.2.2	Patterned Variance–Covariance Matrices for the Random Effects: The <code>pdMat</code> Classes . . . . .	157
4.2.3	Fitting Multilevel Models . . . . .	167
4.3	Examining a Fitted Model . . . . .	174

4.3.1	Assessing Assumptions on the Within-Group Error . . .	174
4.3.2	Assessing Assumptions on the Random Effects . . .	187
4.4	Chapter Summary . . . . .	196
	Exercises . . . . .	197
<b>5</b>	<b>Extending the Basic Linear Mixed-Effects Model</b>	<b>201</b>
5.1	General Formulation of the Extended Model . . . . .	202
5.1.1	Estimation and Computational Methods . . . . .	202
5.1.2	The GLS model . . . . .	203
5.1.3	Decomposing the Within-Group Variance–Covariance Structure . . . . .	205
5.2	Variance Functions for Modeling Heteroscedasticity . . . . .	206
5.2.1	<code>varFunc</code> classes in <code>nlme</code> . . . . .	208
5.2.2	Using <code>varFunc</code> classes with <code>lme</code> . . . . .	214
5.3	Correlation Structures for Modeling Dependence . . . . .	226
5.3.1	Serial Correlation Structures . . . . .	226
5.3.2	Spatial Correlation Structures . . . . .	230
5.3.3	<code>corStruct</code> classes in <code>nlme</code> . . . . .	232
5.3.4	Using <code>corStruct</code> Classes with <code>lme</code> . . . . .	239
5.4	Fitting Extended Linear Models with <code>glS</code> . . . . .	249
5.5	Chapter Summary . . . . .	266
	Exercises . . . . .	267
<b>II</b>	<b>Nonlinear Mixed-Effects Models</b>	<b>271</b>
<b>6</b>	<b>NLME Models: Basic Concepts and Motivating Examples</b>	<b>273</b>
6.1	LME Models vs. NLME Models . . . . .	273
6.2	Indomethicin Kinetics . . . . .	277
6.3	Growth of Soybean Plants . . . . .	287
6.4	Clinical Study of Phenobarbital Kinetics . . . . .	294
6.5	Chapter Summary . . . . .	300
	Exercises . . . . .	301
<b>7</b>	<b>Theory and Computational Methods for NLME Models</b>	<b>305</b>
7.1	The NLME Model Formulation . . . . .	306
7.1.1	Single-Level of Grouping . . . . .	306
7.1.2	Multilevel NLME Models . . . . .	309
7.1.3	Other NLME Models . . . . .	310
7.2	Estimation and Inference in NLME Models . . . . .	312
7.2.1	Likelihood Estimation . . . . .	312
7.2.2	Inference and Predictions . . . . .	322
7.3	Computational Methods . . . . .	324
7.4	Extending the Basic NLME Model . . . . .	328

7.4.1	General model formulation . . . . .	328
7.4.2	Estimation and Computational Methods . . . . .	329
7.5	An Extended Nonlinear Regression Model . . . . .	332
7.5.1	General Model Formulation . . . . .	333
7.5.2	Estimation and Computational Methods . . . . .	334
7.6	Chapter Summary . . . . .	336
<b>8</b>	<b>Fitting Nonlinear Mixed-Effects Models</b>	<b>337</b>
8.1	Fitting Nonlinear Models in S with <code>nls</code> and <code>nlsList</code> . . . . .	338
8.1.1	Using the <code>nls</code> Function . . . . .	338
8.1.2	Self-Starting Nonlinear Model Functions . . . . .	342
8.1.3	Separate Nonlinear Fits by Group: The <code>nlsList</code> Function . . . . .	347
8.2	Fitting Nonlinear Mixed-Effects Models with <code>nlme</code> . . . . .	354
8.2.1	Fitting Single-Level <code>nlme</code> Models . . . . .	354
8.2.2	Using Covariates with <code>nlme</code> . . . . .	365
8.2.3	Fitting Multilevel <code>nlme</code> Models . . . . .	385
8.3	Extending the Basic <code>nlme</code> Model . . . . .	391
8.3.1	Variance Functions in <code>nlme</code> . . . . .	391
8.3.2	Correlation Structures in <code>nlme</code> . . . . .	395
8.3.3	Fitting Extended Nonlinear Regression Models with <code>gnls</code> . . . . .	401
8.4	Chapter Summary . . . . .	409
	Exercises . . . . .	410
	<b>References</b>	<b>415</b>
<b>A</b>	<b>Data Used in Examples and Exercises</b>	<b>423</b>
A.1	Alfalfa—Split-Plot Experiment on Varieties of Alfalfa . . . . .	425
A.2	Assay—Bioassay on Cell Culture Plate . . . . .	425
A.3	BodyWeight—Body Weight Growth in Rats . . . . .	427
A.4	Cefamandole—Pharmacokinetics of Cefamandole . . . . .	427
A.5	CO <sub>2</sub> —Carbon Dioxide Uptake . . . . .	428
A.6	Dialyzer—High-Flux Hemodialyzer . . . . .	429
A.7	DNase—Assay Data for the Protein DNase . . . . .	429
A.8	Earthquake—Earthquake Intensity . . . . .	430
A.9	ergoStool—Ergometrics Experiment with Stool Types . . . . .	431
A.10	Glucose2—Glucose Levels Following Alcohol Ingestion . . . . .	432
A.11	IGF—Radioimmunoassay of IGF-I Protein . . . . .	433
A.12	Indometh—Indomethicin Kinetics . . . . .	433
A.13	Loblolly—Growth of Loblolly Pine Trees . . . . .	434
A.14	Machines—Productivity Scores for Machines and Workers . . . . .	435
A.15	Oats—Split-plot Experiment on Varieties of Oats . . . . .	435
A.16	Orange—Growth of Orange Trees . . . . .	436
A.17	Orthodont—Orthodontic Growth Data . . . . .	436

A.18 Ovary—Counts of Ovarian Follicles . . . . .	437
A.19 Oxboys—Heights of Boys in Oxford . . . . .	437
A.20 Oxide—Variability in Semiconductor Manufacturing . . . . .	437
A.21 PBG—Effect of Phenylbiguanide on Blood Pressure . . . . .	438
A.22 PBIB—A Partially Balanced Incomplete Block Design . . . . .	439
A.23 Phenobarb—Phenobarbitol Kinetics . . . . .	440
A.24 Pixel—Pixel Intensity in Lymphnodes . . . . .	440
A.25 Quinidine—Quinidine Kinetics . . . . .	441
A.26 Rail—Evaluation of Stress in Rails . . . . .	443
A.27 Soybean—Soybean Leaf Weight over Time . . . . .	443
A.28 Spruce—Growth of Spruce Trees . . . . .	444
A.29 Theoph—Theophylline Kinetics . . . . .	444
A.30 Wafer—Modeling of Analog MOS Circuits . . . . .	448
A.31 Wheat2—Wheat Yield Trials . . . . .	448

## **B S Functions and Classes 451**

ACF . . . . .	451
ACF.lme . . . . .	452
anova.lme . . . . .	453
coef.lme . . . . .	455
coef.lmList . . . . .	457
fitted.lme . . . . .	458
fixef . . . . .	459
gapply . . . . .	460
getGroups . . . . .	461
gls . . . . .	462
gnls . . . . .	464
groupedData . . . . .	466
gsummary . . . . .	469
intervals . . . . .	471
intervals.lme . . . . .	471
intervals.lmList . . . . .	473
lme . . . . .	474
lmeControl . . . . .	476
lmList . . . . .	478
logLik . . . . .	479
nlme . . . . .	479
nlmeControl . . . . .	483
nlsList . . . . .	485
pairs.lme . . . . .	486
plot.lme . . . . .	488
plot.nfnGroupedData . . . . .	490
plot.nmGroupedData . . . . .	492
plot.Variogram . . . . .	494
predict.lme . . . . .	495

qqnorm.lme . . . . .	497
ranef . . . . .	498
ranef.lme . . . . .	499
ranef.lmList . . . . .	501
residuals.lme . . . . .	503
selfStart . . . . .	504
selfStart.default . . . . .	505
selfStart.formula . . . . .	506
Variogram . . . . .	507
Variogram.lme . . . . .	508

## C A Collection of Self-Starting Nonlinear Regression

### Models 511

C.1 SSasymp—The Asymptotic Regression Model . . . . .	511
C.1.1 Starting Estimates for SSasymp . . . . .	511
C.2 SSasympOff—Asymptotic Regression with an Offset . . . . .	512
C.2.1 Starting Estimates for SSasympOff . . . . .	512
C.3 SSasympOrig—Asymptotic Regression Through the Origin . . . . .	513
C.3.1 Starting Estimates for SSasympOrig . . . . .	513
C.4 SSbiexp—Biexponential Model . . . . .	514
C.4.1 Starting Estimates for SSbiexp . . . . .	515
C.5 SSfo1—First-Order Compartment Model . . . . .	516
C.5.1 Starting Estimates for SSfo1 . . . . .	516
C.6 SSfpl—Four-Parameter Logistic Model . . . . .	517
C.6.1 Starting Estimates for SSfpl . . . . .	518
C.7 SSlogis—Simple Logistic Model . . . . .	519
C.7.1 Starting Estimates for SSlogis . . . . .	519
C.8 SSmicmen—Michaelis-Menten Model . . . . .	520
C.8.1 Starting Estimates for SSmicmen . . . . .	521

### Index 523