

DOUGLAS M. BATES
DONALD G. WATTS

**NONLINEAR REGRESSION
ANALYSIS AND ITS
APPLICATIONS**

**WILEY SERIES IN PROBABILITY
AND MATHEMATICAL STATISTICS**



Nonlinear Regression Analysis and Its Applications

Douglas M. Bates

Department of Statistics
University of Wisconsin
Madison, Wisconsin

Donald G. Watts

Department of Mathematics
and Statistics
Queen's University
Kingston, Ontario, Canada



WILEY

JOHN WILEY & SONS

New York Chichester Brisbane Toronto Singapore

Copyright © 1988 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Section 107 or 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

Library of Congress Cataloging in Publication Data:

Bates, Douglas M.

Nonlinear regression analysis and its applications / Douglas M.

Bates, Donald G. Watts.

p. cm.—(Wiley series in probability and mathematical statistics. Applied probability and statistics)

Bibliography: p.

Includes index.

ISBN 0-471-81643-4

1. Regression analysis. 2. Linear models (Statistics)

3. Parameter estimation. I. Watts, Donald G. II. Title.

III. Series.

QA278.2.B375 1988

519.5'36--dc19

88-6065

CIP

Printed in the United States of America

10 9 8 7 6 5 4 3 2

To

Mary Ellen
Barbara
Michael

Valery
Lloyd
Megan

PREFACE

"Reading maketh a full man, conference; a ready man, and writing; an exact man."

— Francis Bacon

In this book we have tried to give a balanced presentation of the theory and practice of nonlinear regression.

We expect readers to have a working knowledge of linear regression at about the level of Draper and Smith (1981) or Montgomery and Peck (1982). Nevertheless, to provide background material and to establish notation, we give a summary review of linear least squares in Chapter 1, together with a geometrical development which is helpful in understanding both linear and nonlinear least squares. On the practical side, we discuss linear least squares in the context of modern computing methods and present useful material for checking the assumptions which are involved in regression and for modifying and improving fitted models. In Chapter 2 we discuss how nonlinear models can arise, and show how linear regression methods can be used iteratively to estimate the parameters. We also show how linear methods can be used to make approximate inferences about parameters and nonlinear model functions: again, the geometry is emphasized. The practical aspects of nonlinear estimation are discussed fully in Chapter 3, including such topics as getting starting values, transforming parameters, derivative-free methods, dealing with correlated residuals and with accumulated data, and comparing models.

In Chapter 4 we cover special methods for dealing with multiresponse data, and in Chapter 5, special techniques for compartment models in which the response function is specified as the solution to a set of linear differential equations.

In Chapter 6 we discuss improved methods for presenting the inferential results of a nonlinear analysis, using likelihood profile traces and profile t plots. Finally, in Chapter 7 we present material concerned with measuring how badly nonlinear a particular model–data set situation is. This chapter is helpful in understanding and appreciating the geometry of nonlinear least squares—and indeed, of linear least squares.

Extensive displays of geometrical constructs have been used to facilitate understanding. We have also used continuing examples so that readers can follow the development of ideas in manageable steps within familiar contexts.

All of the data sets used in this book are *real*, that is, the data were obtained from genuine physical, chemical, and biological experiments. We are grateful to the many authors, researchers, and publishers who gave permission to quote their data. In particular we would like to thank Don deBethizy, Rick Elliott, Steve Havriliak, Nico Linssen, Dave Pierson, Rob Stiratelli, Marg Treloar, and Eric Ziegel. We also acknowledge helpful comments made by participants in courses at Dalhousie University, Queen's University, and the University of Wisconsin, where the book was tested in class.

We are grateful to David Hamilton for stimulating discussions and collaborations, to Gunseog Kang for exemplary service in proofreading, and to Steve Czarniak, Mary Lindstrom and Dennis Wolf who also helped in proofreading.

The book was composed electronically using the *troff* text formatting language on the Statistics research computer at the University of Wisconsin–Madison. The figures were produced using the **S** language for statistics and graphics and both text and graphics were typeset on a Linotronic L300 using the PostScript language. We are appreciative of the good work done by Bea Shube and her colleagues at Wiley and by Bill Kasdorf of Impressions.

Considerable research was involved in developing the material in this book, and we thank the Natural Sciences and Engineering Research Council of Canada and the United States National Science Foundation for support.

Finally, we thank our wives for their continued love and encouragement.

June, 1988

Douglas M. Bates
Donald G. Watts

CONTENTS

| | |
|--|-----------|
| 1. Review of Linear Regression | 1 |
| 1.1 The Linear Regression Model / 1 | |
| 1.1.1 The Least Squares Estimates / 4 | |
| 1.1.2 Sampling Theory Inference Results / 5 | |
| 1.1.3 Likelihood Inference Results / 6 | |
| 1.1.4 Bayesian Inference Results / 7 | |
| 1.1.5 Comments / 7 | |
| 1.2 The Geometry of Linear Least Squares / 9 | |
| 1.2.1 The Expectation Surface / 9 | |
| 1.2.2 Determining the Least Squares Estimates / 12 | |
| 1.2.3 Parameter Inference Regions / 15 | |
| 1.2.3.1 The Geometry of Sampling Theory Results / 15 | |
| 1.2.3.2 Marginal Confidence Intervals / 19 | |
| 1.2.3.3 The Geometry of Likelihood Results / 23 | |
| 1.3 Assumptions and Model Assessment / 23 | |
| 1.3.1 Assumptions and Their Implications / 24 | |
| 1.3.2 Model Assessment / 26 | |
| 1.3.2.1 Plotting Residuals / 27 | |
| 1.3.2.2 Stabilizing Variance / 28 | |
| 1.3.2.3 Lack of Fit / 29 | |
| 2. Nonlinear Regression: Iterative Estimation and Linear Approximations | 32 |
| 2.1 The Nonlinear Regression Model / 32 | |
| 2.1.1 Transformably Linear Models / 34 | |
| 2.1.2 Conditionally Linear Parameters / 36 | |
| 2.1.3 The Geometry of the Expectation Surface / 36 | |
| 2.2 Determining the Least Squares Estimates / 39 | |

- 2.2.1 The Gauss–Newton Method / 40
 - 2.2.1.1 Step Factor / 42
- 2.2.2 The Geometry of Nonlinear Least Squares / 43
- 2.2.3 Convergence / 47
- 2.3 Nonlinear Regression Inference Using the Linear Approximation / 52
 - 2.3.1 Approximate Inference Regions for Parameters / 52
 - 2.3.2 Approximate Inference Bands for the Expected Response / 58
- 2.4 Nonlinear Least Squares via Sums of Squares / 60
 - 2.4.1 The Linear Approximation / 61
 - 2.4.2 Overshoot / 64
- 2.5 Use of the Linear Approximation / 64

3. Practical Considerations in Nonlinear Regression

67

- 3.1 Model Specification / 67
 - 3.1.1 The Expectation Function / 67
 - 3.1.2 The Disturbance Term / 69
- 3.2 Preliminary Analysis / 70
- 3.3 Starting Values / 72
 - 3.3.1 Interpreting the Expectation Function Behavior / 72
 - 3.3.2 Interpreting Derivatives of the Expectation Function / 73
 - 3.3.3 Transforming the Expectation Function / 73
 - 3.3.4 Reducing Dimensions / 76
 - 3.3.5 Conditional Linearity / 76
- 3.4 Parameter Transformations / 77
 - 3.4.1 Constrained Parameters / 77
 - 3.4.2 Facilitating Convergence / 78
- 3.5 Other Iterative Techniques / 79
 - 3.5.1 A Newton–Raphson Method / 79
 - 3.5.2 The Levenberg–Marquardt Compromise / 80
 - 3.5.3 Numerical Derivatives / 82
 - 3.5.4 Derivative-free Methods / 82
 - 3.5.5 Removing Conditionally Linear Parameters / 85
- 3.6 Obtaining Convergence / 86
- 3.7 Assessing the Fit and Modifying the Model / 90
- 3.8 Correlated Residuals / 92
- 3.9 Accumulated Data / 96
 - 3.9.1 Estimating the Parameters by Direct Integration / 98
- 3.10 Comparing Models / 103
 - 3.10.1 Nested Models / 103
 - 3.10.2 Incremental Parameters and Indicator Variables / 104
 - 3.10.3 Non-nested Models / 107

- 3.11 Parameters as Functions of Other Variables / 108
- 3.12 Presenting the Results / 109
- 3.13 Nitrite Utilization: A Case Study / 110
 - 3.13.1 Preliminary Analysis / 110
 - 3.13.2 Model Selection / 113
 - 3.13.3 Starting Values / 113
 - 3.13.4 Assessing the Fit / 113
 - 3.13.5 Modifying the Model / 114
 - 3.13.6 Assessing the Fit / 116
 - 3.13.7 Reducing the Model / 116
 - 3.13.8 Assessing the Fit / 116
 - 3.13.9 Comparing the Models / 120
 - 3.13.10 Reporting the Results / 121
- 3.14 Experimental Design / 121
 - 3.14.1 General Considerations / 122
 - 3.14.2 The Determinant Criterion / 124
 - 3.14.3 Starting Designs / 125
 - 3.14.4 Sequential designs / 127
 - 3.14.5 Subset designs / 129
 - 3.14.6 Conditionally Linear Models / 129
 - 3.14.7 Other design criteria / 131

4. Multiresponse Parameter Estimation

134

- 4.1 The Multiresponse Model / 134
 - 4.1.1 The Determinant Criterion / 138
 - 4.1.2 Inferences for Multiresponse Estimation / 139
 - 4.1.3 Dimensional Considerations in Multiresponse Estimation / 140
- 4.2 A Generalized Gauss-Newton Method / 141
 - 4.2.1 The Gradient and Hessian of the Determinant / 142
 - 4.2.2 An Approximate Hessian / 143
 - 4.2.3 Calculations for Each Iteration / 144
 - 4.2.4 A Multiresponse Convergence Criterion / 145
- 4.3 Practical Considerations / 146
 - 4.3.1 Obtaining Starting Values / 146
 - 4.3.1.1 Starting Estimates for Multiresponse Models Described by Systems of Differential Equations / 147
 - 4.3.2 Assessing the Fit / 149
 - 4.3.3 Dependencies Among Responses / 154
 - 4.3.4 Linear Combinations of Responses / 158
 - 4.3.5 Comparing Models / 162
- 4.4 Missing Data / 164

| | |
|---|------------|
| 5. Models Defined by Systems of Differential Equations | 168 |
| 5.1 Compartment Models and System Diagrams / 168 | |
| 5.2 Estimating Parameters in Compartment Models / 172 | |
| 5.2.1 Solving Systems of Linear Differential Equations / 173 | |
| 5.2.1.1 Dead Time / 175 | |
| 5.2.1.2 Cessation of Infusion / 177 | |
| 5.2.2 Derivatives of the Expectation Function / 178 | |
| 5.3 Practical Considerations / 179 | |
| 5.3.1 Parameter Transformations / 179 | |
| 5.3.2 Identifiability / 180 | |
| 5.3.3 Starting Values / 182 | |
| 5.4 Lipoproteins: A Case Study / 182 | |
| 5.4.1 Preliminary Analysis / 182 | |
| 5.4.2 One Compartment / 182 | |
| 5.4.3 Two Compartments / 183 | |
| 5.4.4 Three Compartments / 184 | |
| 5.4.5 Three Compartments, Common Parameters / 187 | |
| 5.4.6 Conclusions / 187 | |
| 5.5 Oil shale: A Case Study / 188 | |
| 5.5.1 Preliminary Analysis / 188 | |
| 5.5.2 Starting Values for the 673 K Data / 190 | |
| 5.5.3 Fitting the Individual Temperature Data / 191 | |
| 5.5.4 Starting Estimates for the Process Parameters / 193 | |
| 5.5.5 Fitting the Complete Data Set / 194 | |
| 5.5.6 Conclusions / 198 | |
| 6. Graphical Summaries of Nonlinear Inference Regions | 200 |
| 6.1 Likelihood Regions / 200 | |
| 6.1.1 Joint Parameter Likelihood Regions / 200 | |
| 6.1.2 Profile t Plots, Profile Traces, and Profile Pair Sketches / 205 | |
| 6.1.2.1 Profile Traces / 207 | |
| 6.1.2.2 Profile Pair Sketches / 209 | |
| 6.1.3 Comments / 214 | |
| 6.2 Bayes Regions / 216 | |
| 6.2.1 Choice of Bayes Prior on the Parameters / 216 | |
| 6.2.2 Joint HPD Regions / 219 | |
| 6.3 Exact Sampling Theory Confidence Regions / 223 | |
| 6.4 Comparison of the Likelihood, Bayes, and Sampling Theory Approaches / 229 | |

7. Curvature Measures of Nonlinearity **232**

- 7.1 Velocity and Acceleration Vectors / 233
 - 7.1.1 Tangential and Normal Accelerations / 234
 - 7.1.2 The Acceleration in an Arbitrary Direction / 239
- 7.2 Relative Curvatures / 241
 - 7.2.1 Interpreting Terms in the Curvature Arrays / 245
 - 7.2.1.1 Intrinsic Curvature / 245
 - 7.2.1.2 Parameter Effects Curvatures / 246
 - 7.2.2 Reparametrization / 248
- 7.3 RMS Curvatures / 253
 - 7.3.1 Calculating RMS Curvatures / 254
 - 7.3.2 RMS Curvatures in Practice / 256
- 7.4 Direct Assessment of the Effects of Intrinsic Nonlinearity / 256

Appendix 1. Data Sets Used in Examples **267**

- 1.1 PCB / 267
- 1.2 Rumford / 268
- 1.3 Puromycin / 269
- 1.4 BOD / 270
- 1.5 Isomerization / 271
- 1.6 α -Pinene / 272
- 1.7 Sulfisoxazole / 273
- 1.8 Lubricant / 274
- 1.9 Chloride / 274
- 1.10 Ethyl Acrylate / 274
- 1.11 Saccharin / 277
- 1.12 Nitrite Utilization / 278
- 1.13 s-PMMA / 280
- 1.14 Tetracycline / 281
- 1.15 Oil Shale / 282
- 1.16 Lipoproteins / 282

Appendix 2. QR Decompositions Using Householder Transformations **286**

Appendix 3. Pseudocode for Computing Algorithms **290**

- 3.1 Nonlinear Least Squares / 290
 - 3.1.1 Implementation in S / 291
 - 3.1.2 Implementation in GAUSS / 293

| | |
|-------|---|
| 3.1.3 | Implementation in SAS/IML / 296 |
| 3.2 | Linear Summaries and Studentized Residuals / 298 |
| 3.3 | Multiresponse Estimation / 299 |
| 3.4 | Linear Systems of Differential Equations / 300 |
| 3.5 | Profile Calculations / 302 |
| 3.5.1 | Generation of τ and the Profile Traces / 302 |
| 3.5.2 | Profile Pair Plots / 303 |

Appendix 4. Data Sets Used in Problems 305

| | |
|-----|-----------------------------------|
| 4.1 | BOD Data Set 2 / 305 |
| 4.2 | Nitrendipene / 306 |
| 4.3 | Saccharin Data Set 2 / 306 |
| 4.4 | Steady State Adsorption / 306 |
| 4.5 | Leaves / 310 |
| 4.6 | α -Pinene Data Set 2 / 310 |
| 4.7 | Coal Liquefaction / 311 |
| 4.8 | Haloperidol / 315 |

Appendix 5. Evaluating Matrix Exponentials and Convolutions 316

| | |
|-----|----------------------------------|
| 5.1 | Diagonalizable A / 317 |
| 5.2 | Nondiagonalizable A / 319 |
| 5.3 | Complex Eigenvalues / 320 |

Appendix 6. Interpolating Profile Pair Contours 323

Appendix 7. Key to Data Sets 329

References 335

Bibliography 343

Author Index 355

Subject Index 359

CHAPTER 1.

Review of Linear Regression

“Non sunt multiplicanda entia praeter necessitatem.”
(Entities are not to be multiplied beyond necessity.)

– William of Ockham

We begin with a brief review of linear regression, because a thorough grounding in linear regression is fundamental to understanding nonlinear regression. For a more complete presentation of linear regression see, for example, Draper and Smith (1981), Montgomery and Peck (1982), or Seber (1977). Detailed discussion of regression diagnostics is given in Belsley, Kuh, and Welsch (1980) and Cook and Weisberg (1982), and the Bayesian approach is discussed in Box and Tiao (1973).

Two topics which we emphasize are modern numerical methods and the geometry of linear least squares. As will be seen, attention to efficient computing methods increases understanding of linear regression, while the geometric approach provides insight into the methods of linear least squares and the analysis of variance, and subsequently into nonlinear regression.

1.1 The Linear Regression Model

Linear regression provides estimates and other inferential results for the *parameters* $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_P)^T$ in the model

$$\begin{aligned} Y_n &= \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_P x_{nP} + Z_n \\ &= (x_{n1}, \dots, x_{nP}) \boldsymbol{\beta} + Z_n \end{aligned} \tag{1.1}$$

In this model, the random variable Y_n , which represents the *response* for case n , $n = 1, 2, \dots, N$, has a *deterministic* part and a *stochastic* part. The deterministic part, $(x_{n1}, \dots, x_{nP}) \boldsymbol{\beta}$, depends upon the parameters $\boldsymbol{\beta}$ and upon the *predictor*

or regressor variables x_{np} , $p = 1, 2, \dots, P$. The stochastic part, represented by the random variable Z_n , is a *disturbance* which perturbs the response for that case. The superscript T denotes the transpose of a matrix.

The model for N cases can be written

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z} \quad (1.2)$$

where \mathbf{Y} is the vector of random variables representing the data we may get, \mathbf{X} is the $N \times P$ matrix of regressor variables,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1P} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2P} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{NP} \end{bmatrix}$$

and \mathbf{Z} is the vector of random variables representing the disturbances. (We will use bold face italic letters for vectors of random variables.)

The deterministic part, $\mathbf{X}\boldsymbol{\beta}$, a function of the parameters and the regressor variables, gives the mathematical model or the model function for the responses. Since a nonzero mean for Z_n can be incorporated into the model function, we assume that

$$E[\mathbf{Z}] = \mathbf{0} \quad (1.3)$$

or, equivalently,

$$E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$$

We therefore call $\mathbf{X}\boldsymbol{\beta}$ the *expectation function* for the regression model. The matrix \mathbf{X} is called the *derivative matrix*, since the (n,p) th term is the derivative of the n th row of the expectation function with respect to the p th parameter.

Note that for linear models, *derivatives with respect to any of the parameters are independent of all the parameters.*

If we further assume that \mathbf{Z} is normally distributed with

$$\text{Var}[\mathbf{Z}] = E[\mathbf{Z}\mathbf{Z}^T] = \sigma^2 \mathbf{I} \quad (1.4)$$

where \mathbf{I} is an $N \times N$ identity matrix, then the joint probability density function for \mathbf{Y} , given $\boldsymbol{\beta}$ and the variance σ^2 , is

$$\begin{aligned} p(\mathbf{y} | \boldsymbol{\beta}, \sigma^2) &= (2\pi\sigma^2)^{-N/2} \exp \left[\frac{-(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2} \right] \\ &= (2\pi\sigma^2)^{-N/2} \exp \left[\frac{-\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2}{2\sigma^2} \right] \end{aligned} \quad (1.5)$$

where the double vertical bars denote the length of a vector. When provided

with a derivative matrix \mathbf{X} and a vector of observed data \mathbf{y} , we wish to make inferences about σ^2 and the P parameters β .

Example: PCB 1

As a simple example of a linear regression model, we consider the concentration of polychlorinated biphenyls (PCBs) in Lake Cayuga trout as a function of age (Bache et al., 1972). The data set is described in Appendix 1, Section A1.1. A plot of the PCB concentration versus age, Figure 1.1, reveals a curved relationship between PCB concentration and age. Furthermore, there is increasing variance in the PCB concentration as the concentration increases. Since the assumption (1.4) requires that the variance of the disturbances be constant, we seek a transformation of the PCB concentration which will stabilize the variance (see Section 1.3.2). Plotting the PCB concentration on a logarithmic scale, as in Figure 1.2a, nicely stabilizes the variance and produces a more nearly linear relationship. Thus, a linear expectation function of the form

$$\ln(\text{PCB}) = \beta_1 + \beta_2 \text{ age}$$

could be considered appropriate, where \ln denotes the natural logarithm (logarithm to the base e). Transforming the regressor variable (Box and Tidwell, 1962) can produce an even straighter plot, as shown in Figure 1.2b, where we use the cube root of age. Thus a simple expectation function to be fitted is

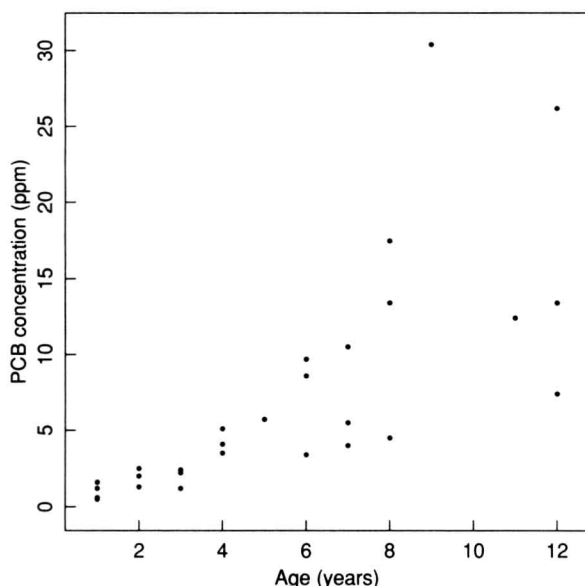


Figure 1.1 Plot of PCB concentration versus age for lake trout.

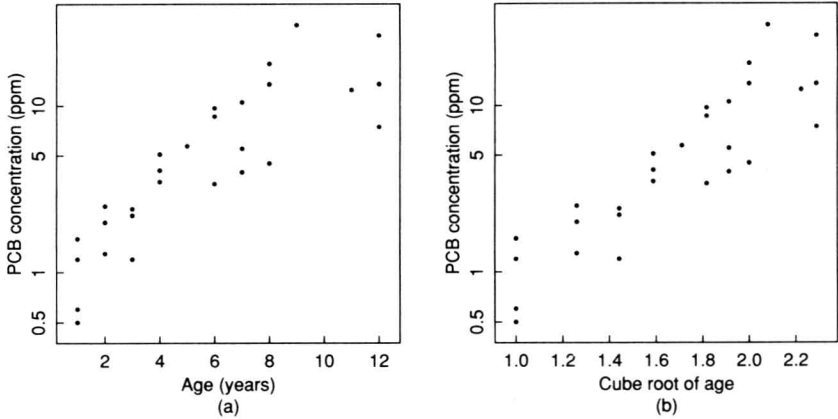


Figure 1.2 Plot of PCB concentration versus age for lake trout. The concentration, on a logarithmic scale, is plotted versus age in part *a* and versus $\sqrt[3]{\text{age}}$ in part *b*.

$$\ln(\text{PCB}) = \beta_1 + \beta_2 \sqrt[3]{\text{age}}$$

(Note that the methods of Chapter 2 can be used to fit models of the form

$$f(\mathbf{x}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = \beta_0 + \beta_1 x_1^{\alpha_1} + \beta_2 x_2^{\alpha_2} + \cdots + \beta_p x_p^{\alpha_p}$$

by simultaneously estimating the conditionally linear parameters $\boldsymbol{\beta}$ and the transformation parameters $\boldsymbol{\alpha}$. The powers $\alpha_1, \dots, \alpha_p$ are used to transform the factors so that a simple linear model in $x_1^{\alpha_1}, \dots, x_p^{\alpha_p}$ is appropriate. In this book we use the power $\alpha = 0.33$ for the age variable even though, for the PCB data, the optimal value is 0.20.) ■

1.1.1 The Least Squares Estimates

The *likelihood function*, or more simply, the *likelihood*, $l(\boldsymbol{\beta}, \sigma | \mathbf{y})$, for $\boldsymbol{\beta}$ and σ is identical in form to the joint probability density (1.5) except that $l(\boldsymbol{\beta}, \sigma | \mathbf{y})$ is regarded as a function of the parameters conditional on the observed data, rather than as a function of the responses conditional on the values of the parameters. Suppressing the constant $(2\pi)^{-N/2}$, we write

$$l(\boldsymbol{\beta}, \sigma | \mathbf{y}) \propto \sigma^{-N} \exp \left[\frac{-\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2}{2\sigma^2} \right] \quad (1.6)$$

The likelihood is maximized with respect to $\boldsymbol{\beta}$ when the *residual sum of squares*

$$\begin{aligned} S(\boldsymbol{\beta}) &= \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \\ &= \sum_{n=1}^N \left[y_n - \left(\sum_{p=1}^P x_{np} \beta_p \right) \right]^2 \end{aligned} \quad (1.7)$$

is a minimum. Thus the *maximum likelihood estimate* $\hat{\beta}$ is the value of β which minimizes $S(\beta)$. This $\hat{\beta}$ is called the *least squares estimate* and can be written

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (1.8)$$

Least squares estimates can also be derived by using sampling theory, since the least squares estimator is the minimum variance unbiased estimator for β , or by using a Bayesian approach with a noninformative prior density on β and σ . In the Bayesian approach, $\hat{\beta}$ is the mode of the marginal posterior density function for β .

All three of these methods of inference, the likelihood approach, the sampling theory approach, and the Bayesian approach, produce the same point estimates for β . As we will see shortly, they also produce similar regions of “reasonable” parameter values. First, however, it is important to realize that the least squares estimates are only appropriate when the model (1.2) and the assumptions on the disturbance term, (1.3) and (1.4), are valid. Expressed in another way, in using the least squares estimates we assume:

- (1) The expectation function is correct.
- (2) The response is expectation function plus disturbance.
- (3) The disturbance is independent of the expectation function.
- (4) Each disturbance has a normal distribution.
- (5) Each disturbance has zero mean.
- (6) The disturbances have equal variances.
- (7) The disturbances are independently distributed.

When these assumptions appear reasonable and have been checked using diagnostic plots such as those described in Section 1.3.2, we can go on to make further inferences about the regression model.

Looking in detail at each of the three methods of statistical inference, we can characterize some of the properties of the least squares estimates.

1.1.2 Sampling Theory Inference Results

The least squares estimator has a number of desirable properties as shown, for example, in Seber (1977):

(1) The least squares estimator $\hat{\beta}$ is normally distributed. This follows because the estimator is a linear function of Y , which in turn is a linear function of Z . Since Z is assumed to be normally distributed, $\hat{\beta}$ is normally distributed.

(2) $E[\hat{\beta}] = \beta$: the least squares estimator is unbiased.

(3) $\text{Var}[\hat{\beta}] = \sigma^2(X^T X)^{-1}$: the covariance matrix of the least squares estimator depends on the variance of the disturbances and on the derivative matrix X .