MODELING VISION IN PATIENTS WITH AGE RELATED MACULAR DEGENERATION

MODELING VISION IN PATIENTS WITH AGE RELATED MACULAR DEGENERATION

By

DAVID HUTCHINSON, M. Math. (University of Waterloo)

A Project
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Science

McMaster University

December 1993

MASTER OF SCIENCE (1994) MCMASTER UNIVERSITY

(Statistics) Hamilton, Ontario

TITLE: MODELING VISION IN PATIENTS WITH AGE RELATED MACULAR DEGENERATION

AUTHOR: David Hutchinson

M. Math. (University of Waterloo)

SUPERVISOR: Dr. A. Willan

NUMBER OF PAGES: vi, 56

Abstract

The purpose of this project is to find a mathematical model to describe the vision profile of patients after treatment for choroidal neovascularization. In this model the dependent variable is the level of vision which will be predicted by time after treatment and a number of other variables measured before treatment. A standard multiple regression analysis is used to find significant predictor variables, to investigate interactions and an appropriate transformation. To take the correlation of observations on the same patient into account a linear mixed effects model is fitted. Finally the usefulness of a nonlinear mixed effects model is investigated.

Acknowledgements

I would like to thank Dr. Andy Willan for being my supervisor. I would like to thank Dr. Roman Viveros and Dr. Wayne Taylor for reading my project.

Contents

Chapter 1

Description of Problem

1.1 Introduction

The leading cause of new blindness in Canada and the United States is age-related macular degeneration. One form of this disease that is often responsible for severe vision loss is choroidal neovascularization (CNV) [7]. The cause of vision loss with this condition is the presence of lessions in the macular with resultant bleeding under and scarring of the retina.

It has been shown that the argon blue-green laser provides an effective treatment for CNV [11]. The goal of treatment is to completely obliterate the lession without damaging the foveal avascular zone at the center of the eye. Choroidal neovascular lessions tend to recur after laser treatment. However in one , three and five years of follow up after treatment there is considerably less vision loss in laser treated eyes than in nontreated eyes [11,14,15].

The Canadian Opthomology Study Group (COSG) was formed in 1985 to design a clinical trial to determine whether krypton red laser or argon green laser is superior for the treatment of CNV. The resulting trial commenced in July 1986 at 16 Canadian centers. Patients enrolled in the study had to

meet certain eligibility requirements [5]. The requirements included:

- 50 years of age or older
- best corrected visual acuity score of 35 (number of letters read) or better on the ETDRS chart (measurements taken as recommended in the COSG Manual of Proceedures [5])
- an angiographically proven CNV whose edge was 200-2500 μm from the edge of the foveal avascular zone (FAZ).

The treatment was done according to standard COSG protocol [5]. Visual acuity was measured at the time of treatment and recorded as the number of letters read from a ETDRS chart . Visual accuity was also measured at follow-up visits occuring at 2 weeks, 6 weeks, 3 months, 6 months and at 6 month intervals thereafter up to 3 years.

1.2 Outline of Project

The purpose of this project is to find a suitable mathematical model to describe the visual acuity profile as a function of time after treatment for CNV and the following variables measured at or just before treatment.

- distance from the edge of the lession to the edge of the foveal avascular zone (FAZ)
- age of the patient
- type of treatment,either argon green laser or krypton red laser
- size of the CNV

- time between symptoms and treatment

The purpose of such a model will be to identify and describe the effect of these independent variables on the patients vision profile and to predict follow up vision. The data used consists of 1817 observations from 190 patients from the previously mentioned clinical trial.

In Chapter 2 the method of linear regression is used to model vision with the independent variables. Significant independent variables are determined and transformations of the time after treatment variable are investigated. In Chapter 3 linear mixed effects models are discussed and fitted to the data using the independent variables deemed significant in Chapter 2. Nonlinear mixed effects models are described and used in Chapter 4.

Chapter 2

Fitting a Regression Model

2.1 Introduction

In this chapter the results from fitting a standard regression model to the data will be presented. This of course implies that the errors are assumed to be independently distributed normal random variables with mean zero and the same variance. First the independent variables that have a significant effect on the response variable will be identified. Interaction terms and some other functional forms will then be investigated in order to improve the model. The adequacy of the fitted model will be discussed.

The following abreviations will be used.

VIS the level of vision measured in number of letters read

TIME time after treatment measured in years

TRT treatment group, coded 0 for the krypton red laser and 1 for the argon green laser

AGE age of the patient in years

SIZE size of the CNV measured in μm

DIST distance between the edge of the CNV and the edge of the FAZ measured in μm

PRE time between symptoms and treatment measured in days

2.2 Significant Independent Variables

The purpose of this section is to identify which of the independent variables previously discussed have a significant effect on the response variable. The following model was fit to the data.

$$VIS = \beta_0 + \beta_1 TIME + \beta_2 TRT + \beta_3 AGE + \beta_4 SIZE + \beta_5 DIST + \beta_6 PRE$$

Using SAS the following results were obtained.

TABLE 2.1 Regression Results

(a)Regression Coefficients

Coefficient	Value	Std. Error	p-value
β_0	88.75225	5.0772349	.0001
β_1	-7.80131	.5248592	.0001
β_2	-.11117	1.5248592	.9168
β_3	-.55227	.0697365	.0001
β_4	.00072	.0009389	.4454
β_5	.00776	.0011398	.0001
β_6	.00681	.0072096	.3452

(b)Analysis of Variance

Source	SS	df	MS	p-value
Regression	162789.69	6	27131.61	.0001
Error	923560.97	1810	510.25	
Total	1086350.65	1816		

The nonsignificant T ratios for β_2, β_4 and β_6 indicate that there is little evidence that the corresponding independent variables individually affect vision. To test the hypothesis

$H_0 : \beta_2 = \beta_4 = \beta_6 = 0$

H_1 : not all of $\beta_2, \beta_4, \beta_6$ are equal zero

the following F test will be used.

$$F^* = \frac{MSR(TRT, SIZE, PRE \mid AGE, DIST, TIME)}{MSE}$$

where $MSR(TRT, SIZE, PRE \mid AGE, DIST, TIME)$ is the extra sum of squares obtained from adding TRT,SIZE and PRE to the model when AGE,DIST and TIME are already present, divided by the appropriate degrees of freedom. For details see page 271 of [17].

If H_0 holds then $F^* \sim F(3, 1810)$. Thus large values of F^* will lead to rejection of H_0 in favour of H_1.

In this case $F^* = ((162789.69 - 162019.55)/3) \div 510.25 = 0.5032$ which yields a p-value of .6801. Thus it is concluded that the variables TRT,SIZE and PRE do not have a significant effect on VIS and can be dropped from the model.

2.3 Interaction Effects

In this section the interactions between the three variables AGE,DIST and TIME will be analyzed to determine if any of them would significantly improve the model. The regression results of the following model are shown in Table 2.2.

$$VIS = \beta_0 + \beta_1 AGE + \beta_2 DIST + \beta_3 TIME + \beta_4 AGE \times DIST + \beta_5 AGE \times TIME + \beta_6 DIST \times TIME + \beta_7 AGE \times DIST \times TIME$$

TABLE 2.2 Model With Interactions

(a)Regression Coefficients

Coefficient	Value	Std. Error	p-value
β_0	90.82167	10.79835	.0001
β_1	- .56802	.15310	.0002
β_2	-.00327	.01164	.7790
β_3	-2.09285	7.51885	.7808
β_4	.00016	.00017	.3377
β_5	-.08173	.10707	.4453
β_6	.00146	.00785	.8526
β_7	-.00002	.00011	.8562

(b)Analysis of Variance

Source	SS	df	MS	p-value
Regression	163746.43	7	23392.35	.0001
Error	922604.22	1809	510.01	
Total	1086350.65	1816		

The t^* values indicate that any one of the interactions could be dropped from the model individually. An F test based on extra sums of squares as in the last section yields $F^* = .846$. Under the null hypothesis of all coefficients of interaction effects equal to zero the test statistic is distributed $F(4, 1809)$. A p-value of .4959 results and it is concluded that none of these interaction effects need be included in the model.

2.4 Quadratic and Exponential Terms

It is possible that curvature exists in the relation of some predictor variable with VIS. The following model was analyzed to check for significant quadratic terms and the results are in Table 2.3.

$$VIS = \beta_0 + \beta_1 AGE + \beta_2 DIST + \beta_3 TIME + \beta_4 AGE^2 + \beta_5 DIST^2 + \beta_6 TIME^2$$

TABLE 2.3 Model With Quadratic Terms

(a)Regression Coefficients

Coefficient	Value	Std. Error	p-value
β_0	133.84180	33.92127	.0001
β_1	-1.83435	.99526	.0655
β_2	.01226	.00410	.0029
β_3	-15.60817	1.80846	.0001
β_4	.00931	.00721	.1972
β_5	.0000021	.0000018	.2427
β_6	2.79733	.61880	.0001

(b)Analysis of Variance

Source	SS	df	MS	p-value
Regression	173772.79	6	28962.13	.0001
Error	912577.86	1810	504.19	
Total	1086350.65	1816		

These results provide evidence that the only quadratic term that is important is the $TIME^2$ term. It can be verified that the $DIST^2$ and AGE^2 terms can be dropped from the model with an extra sum of squares F test

as before. In this case a p-value of .5543 is the result.

With the $DIST^2$ and the AGE^2 terms dropped the model is as follows.

$$\begin{aligned} VIS = \; & 91.66 - .5492AGE + .00769DIST - 15.66TIME \\ & +2.815TIME^2 \end{aligned}$$

All coefficients are significant with p-values of .0001. Note that the negative coefficient for AGE indicates a decrease in the expected VIS the older the patient gets. The positive coefficient for DIST indicates an increase in expected VIS the further the edge of the lession from the edge of the FAZ. The negative coefficient for TIME indicates the expected VIS decreases as time after treatment increases. The positive coefficient of the $TIME^2$ term implies the rate of expected VIS decrease slows as time increases.

This model has a coefficient of determination, R^2 value, of .1588. It is not uncommon for medical data to yield low R^2 values because of high between and within patient variability.

However it still may be possible to improve the fit. Instead of a quadratic relationship between TIME and VIS an inverse exponential relationship may prove more descriptive. This is in because very few patients vision actually decreased to zero as in the quadratic relationship and e^{-TIME} approaches zero only when time gets very large. The results of the following model can be found in Table 2.4. For this and future models AGE, TIME and DIST will be centered around their most favourable values.

$$VIS = \beta_0 + \beta_1(AGE - 50) + \beta_2(DIST - 2500) + \beta_3 e^{-TIME}$$

TABLE 2.4 Model With Exponential Term

(a)Regression Coefficients

Coefficient	Value	Std. Error	p-value
β_0	61.01080	4.93306	.0001
β_1	-.55098	.06853	.0001
β_2	.00778	.00112	.0001
β_3	23.74847	1.49518	.0001

(b)Analysis of Variance

Source	SS	df	MS	p-value
Regression	176308.38	3	58769.46	.0001
Error	910042.28	1813	501.95	
Total	1086350.65	1816		

The resulting model shows a marginal improvement in the R^2 value to .1609. The coeeficients indicate an expected decrease of 5.5 letters read for every 10 years of AGE and an expected increase of 7.8 letters read for every millimetre of DIST between the edge of the lession and the edge of the FAZ. The coefficient of e^{-TIME} indicates that the total expected decrease in VIS will be 23.7 letters read.

2.5 Model Adequacy

To evaluate the adequacy of the fitted model the following scatter plots were prepared.

- predicted values vs. residuals (Figure 2.1)
- AGE vs. residuals (Figure 2.2)
- DIST vs. residuals (Figure 2.3)
- TIME vs. residuals (Figure 2.4)
- e^{-TIME} vs. residuals (Figure 2.5)
- normal probability plot of the residuals (Figure 2.6)

The adequacy of the assumption of normal errors is demonstrated in Figure 2.6 as the correlation between the residuals and the quantiles of the standard normal is .9934.

Figure 2.1

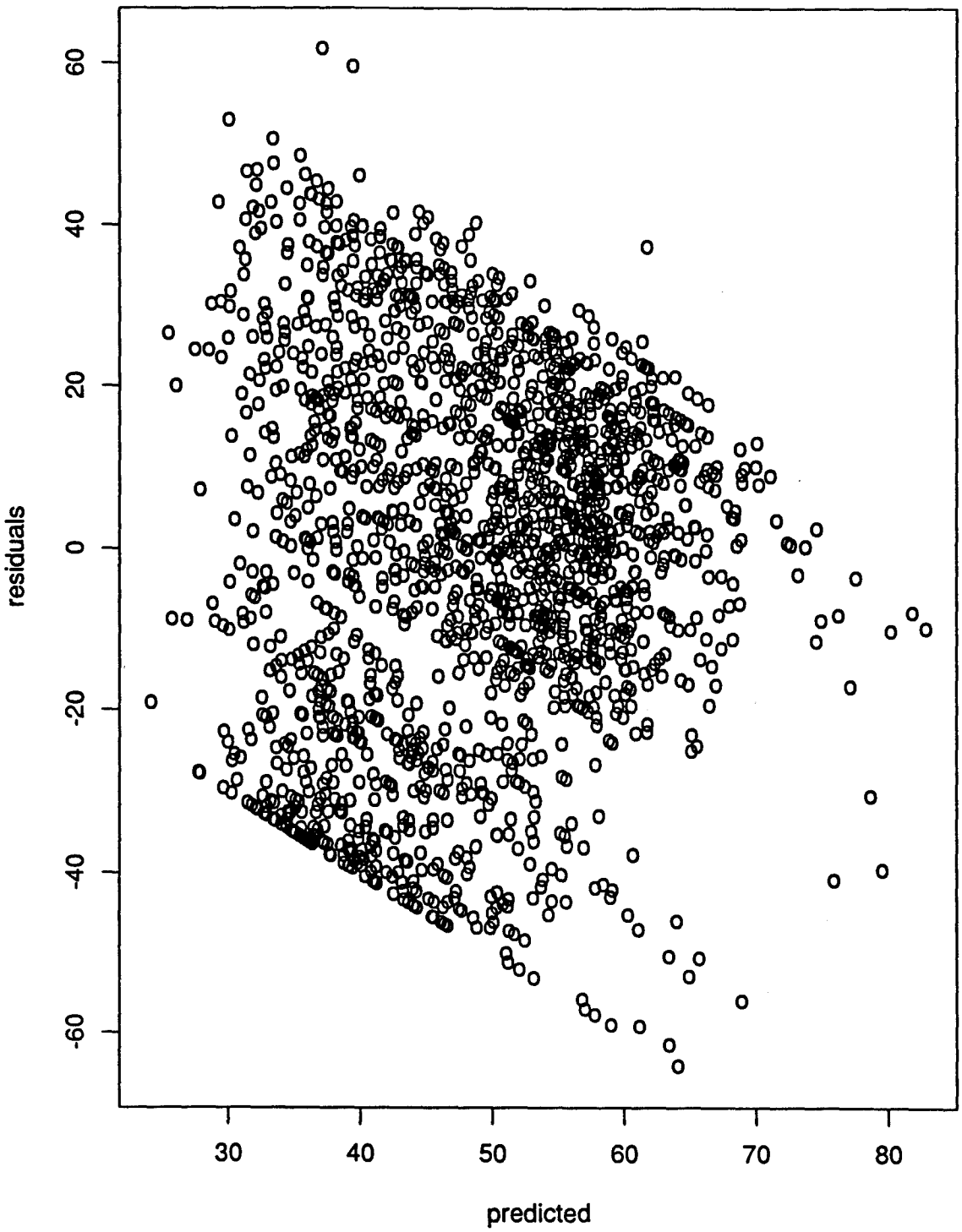


Figure 2.2

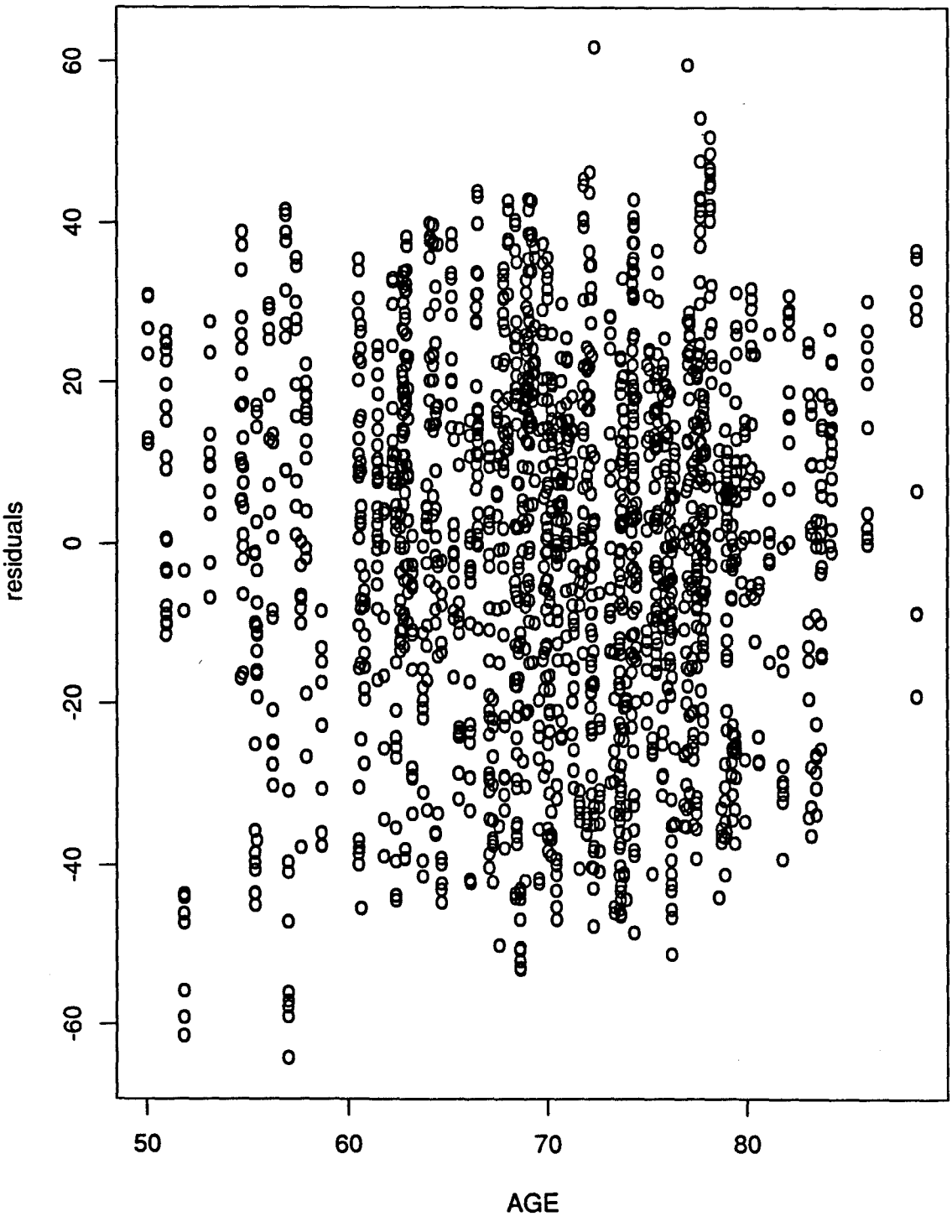


Figure 2.3

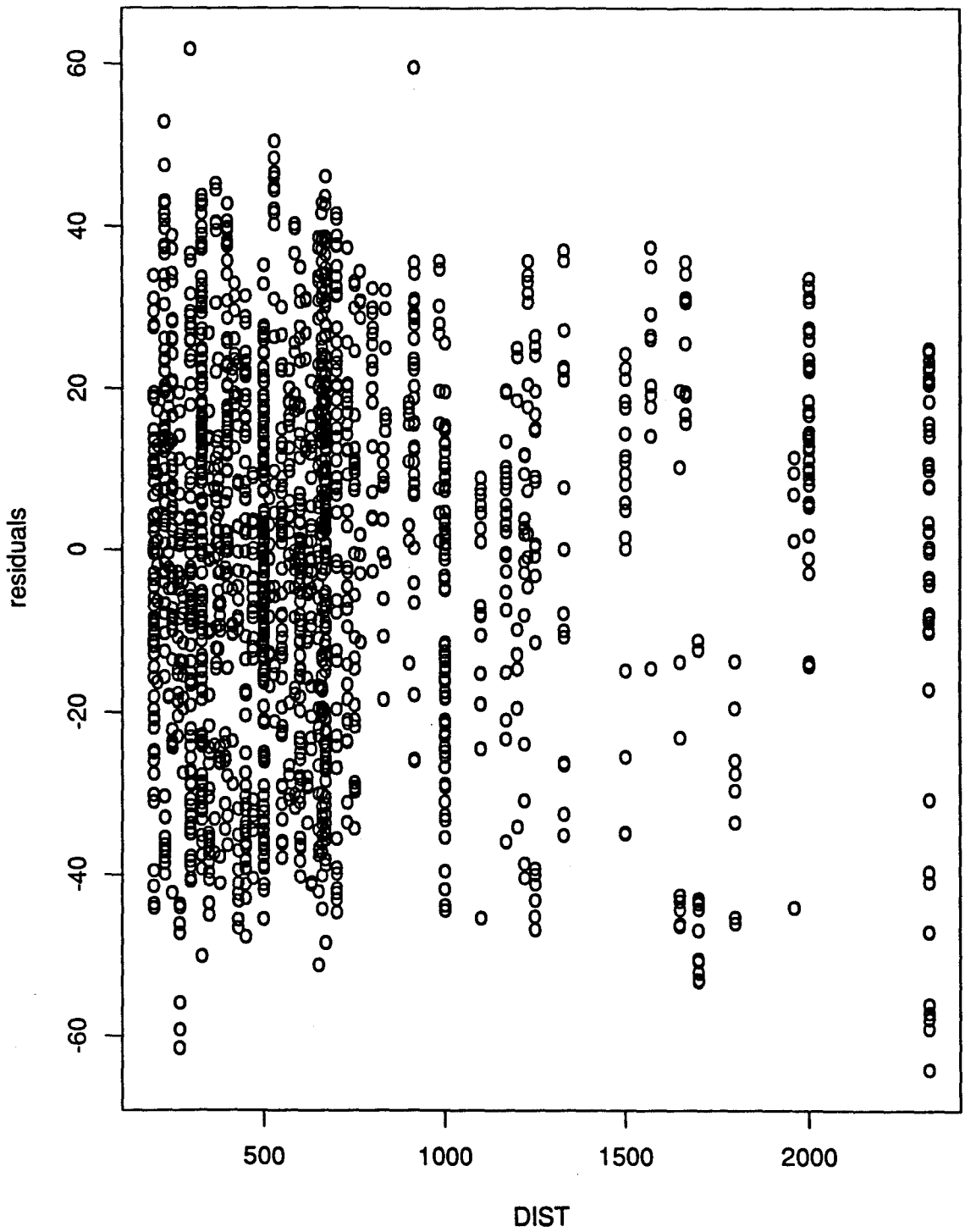


Figure 2.4

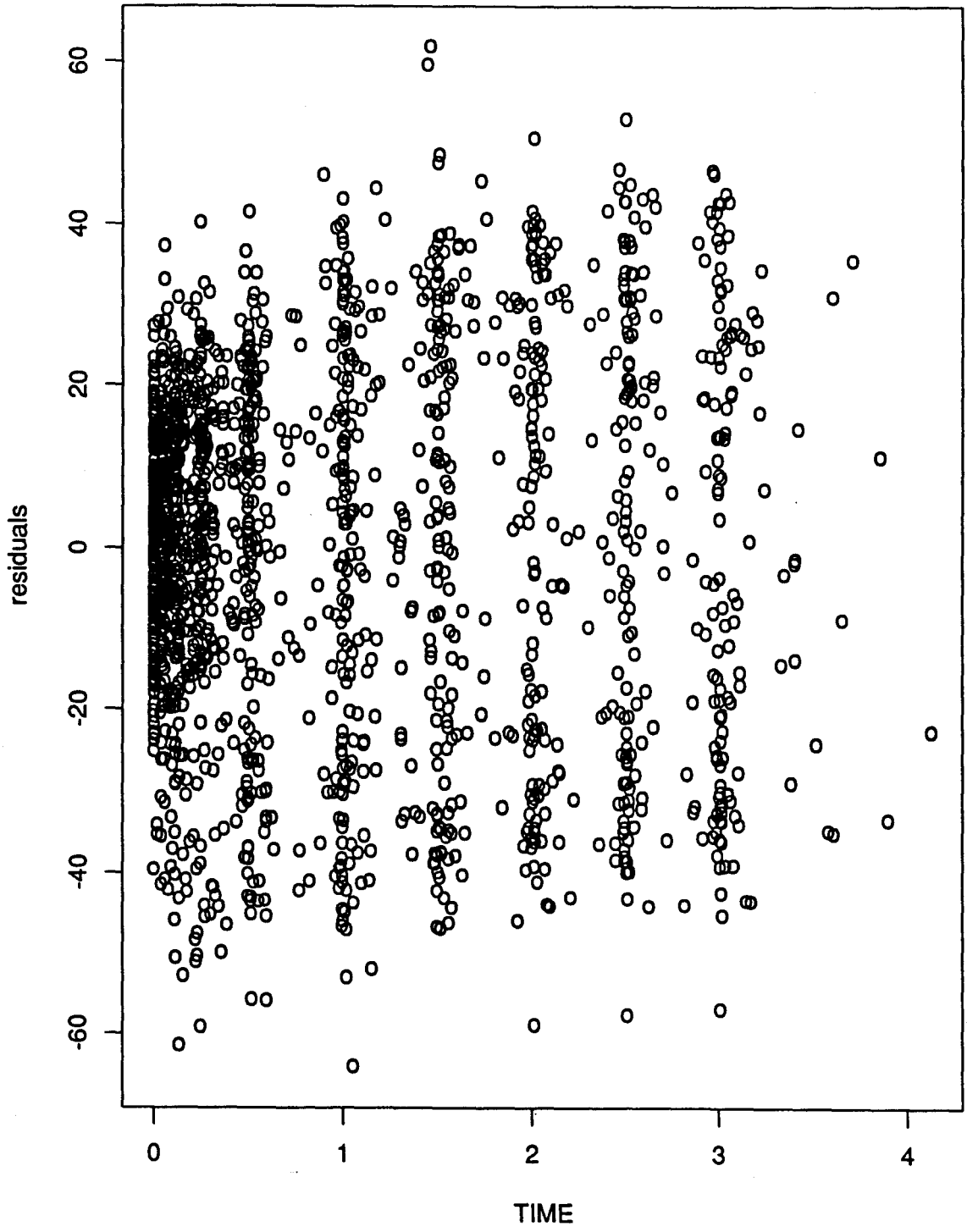


Figure 2.5

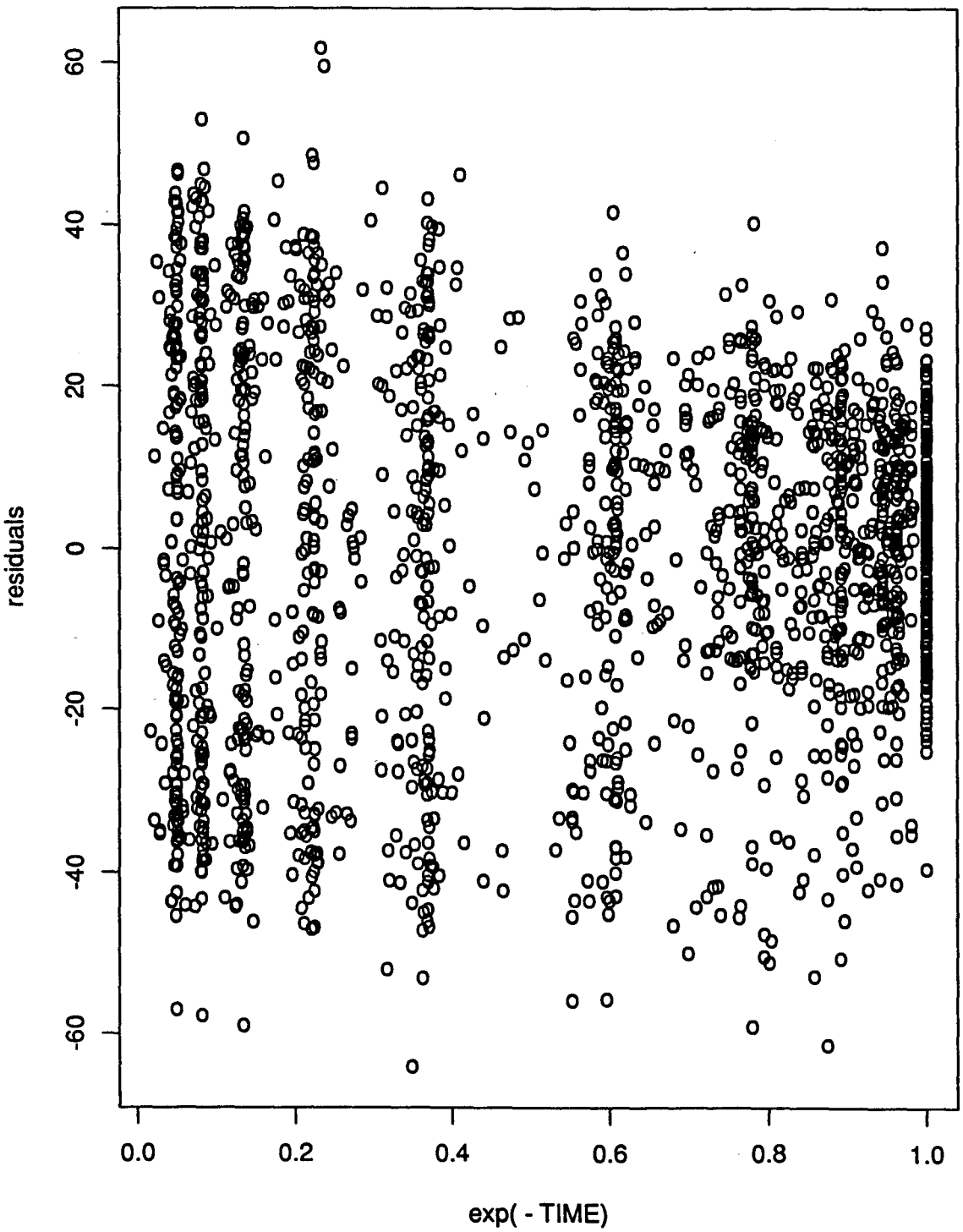
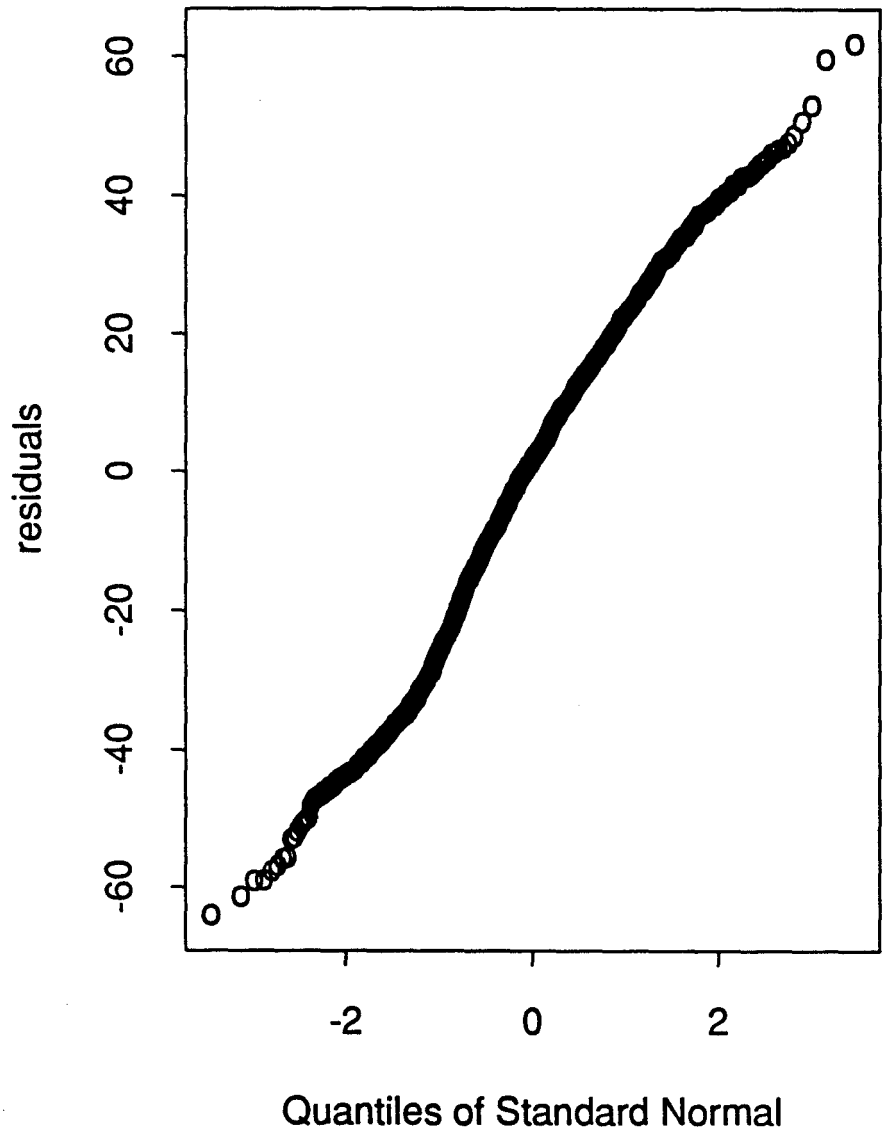


Figure 2.6


residuals
60
40
20
0
-20
-40
-60
-2
0
2
Quantiles of Standard Normal

None of the plots of AGE, DIST, TIME and e^{-TIME} against the residuals indicate a serious lack of fit problem. The plot of predicted values against residuals does indicate some irregularities. Although the variance appears constant there is a definite trend for a more frequent underfit for lower predicted values. There is also a greater number of severe cases of overfit for higher predicted values. This effect appears more pronounced visually as it is partially caused by the zero vision cut off point creating a sharp downward trend at the bottom of the plot. However this situation is a usual symptom of lack of fit. Therfore although this model seems to be a reasonable predictor of the mean vision level there is some degree of a lack of fit problem.

2.6 Adding The Initial Vision Variable

In order to decrease the error variance of the model, increase its R^2 value and thus improve its use for prediction another variable will be considered. The patients initial vision observation will now be excluded from the data as an observation and instead will be used as a prediction variable denoted STVIS. It is expected that the inverse exponential relationship will more adequately describe the effect of time in this case since often there was an increase in VIS between the initial reading, before treatment, and the second reading, which was after treatment. In the previous model this increase contributed to error as the exponential function is decreasing. Hence the following model and regression results.

$$\begin{aligned} VIS = \ & \beta_0 + \beta_1(AGE - 50) + \beta_2(DIST - 2500) + \beta_3 e^{-TIME} \\ & + \beta_4 STVIS \end{aligned}$$

TABLE 2.5 Model With Initial Vision Variable

(a)Regression Coefficients

Coefficient	Value	Std. Error	p-value
β_0	17.80838	4.39043	.0001
β_1	-.35027	.07380	.0001
β_2	.00691	.00118	.0001
β_3	23.78202	1.64401	.0001
β_4	.629904	.04799	.0001

(b)Analysis of Variance

Source	SS	df	MS	p-value
Regression	237018.92	4	59254.73	.0001
Error	798315.14	1622	492.18	
Total	1035334.06	1626		

The result is that STVIS is a significant variable, the R^2 value has increased to .2289 and the estimate of the error standard deviation has decreased to $\sqrt{492.18} = 22.19$. Thus with every additional 10 letters read before treatment the expected VIS will increase by 6.3 letters read.

The model can be further improved with the inclusion of the interactions of STVIS with both AGE and DIST. The resulting model is as follows.

$$\begin{aligned} VIS = \; & 59.6 + 1.53(AGE - 50) - 0.0151(DIST - 2500) + 23.6e^{-TIME} \\ & + (1.91 - 0.0318(AGE - 50) + 0.000367(DIST - 2500))STVIS \end{aligned}$$

All coefficients are significant with p-value of .0001 except β_2 which has a p-value of .01 and β_6 with .0002. The R^2 value is .2473 and the estimate of

the error standard deviation is 21.93. The comments made concerning the adaquacy of the model without STVIS are similar for this model.

Chapter 3

Linear Mixed Effects Models

3.1 Introduction

Repeated measures data refers to data where several observations are taken on a sample of individuals. The set of observations taken on any one individual is called a cluster. Mixed effects models recognize the relationship between observations taken within a cluster. Unlike multivariate models, mixed effects models can model data that is unbalanced in the sense that different experimental designs are used on different individuals. The vision loss data reported herein is unbalanced in this way as differing numbers of observations are taken on different individuals. In this chapter the linear mixed effects model will be described and applied to the vision loss data.

3.2 Description of Model

Let $\mathbf{y_i}$ be a vector of length n_i of observations on individual i for $i = 1, 2, \ldots, M$. Let $\mathbf{X_i}$ and $\mathbf{Z_i}$ be design matrices of sizes $n_i \times p$ and $n_i \times k$ respectively. Let β denote a vector of length p of fixed effect coefficients and

let $\mathbf{b_i}$ denote the vector of length k of random effects. For individual i,

$$\mathbf{y_i} = \mathbf{X_i}\beta + \mathbf{Z_i}\mathbf{b_i} + \mathbf{e_i},$$

where $\mathbf{e_i}$ are independently distributed multivariate normal with mean 0 and covariance matrix $\sigma^2\mathbf{R_i}$. The $\mathbf{b_i}$ are distributed multivariate normal with mean 0 and covariance matrix $\sigma^2\mathbf{D}$ independently of each other and the $\mathbf{e_i}$. Marginally, $\mathbf{y_i}$ is distributed multivariate normal with mean $\mathbf{X_i}\beta$ and covariance matrix $\sigma^2(\mathbf{R_i} + \mathbf{Z_i}\mathbf{D}\mathbf{Z_i^T})$.

The combined model for all the data can be written in matrix form by letting

$$\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \\ \vdots \\ \mathbf{X_M} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_M} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \mathbf{b_1} \\ \mathbf{b_2} \\ \vdots \\ \mathbf{b_M} \end{bmatrix}, \mathbf{e} = \begin{bmatrix} \mathbf{e_1} \\ \mathbf{e_2} \\ \vdots \\ \mathbf{e_M} \end{bmatrix}$$

$\overline{\mathbf{D}}$ =diag$(\mathbf{D}, \mathbf{D}, \ldots, \mathbf{D})$, $\mathbf{Z}$ =diag$(\mathbf{Z_1}, \mathbf{Z_2}, \ldots, \mathbf{Z_M})$ and $\mathbf{R}$ =diag$(\mathbf{R_1}, \mathbf{R_2}, \ldots, \mathbf{R_M})$. Thus

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \mathbf{e}$$

where $\mathbf{e}$ is distributed multivariate normal with mean 0 and covariance matrix $\sigma^2\mathbf{R}$ and $\mathbf{b}$ is distributed multivariate normal with mean 0 and covariance matrix $\sigma^2\overline{\mathbf{D}}$. Therefore $\mathbf{y}$ is distributed multivariate normal with mean $\mathbf{X}\beta$ and covariance matrix $\sigma^2(\mathbf{R} + \mathbf{Z}\overline{\mathbf{D}}\mathbf{Z^T})$.

3.3 Inference of Unknown Parameters

Let θ be a vector that contains the unique elements of $\mathbf{R}$ and $\bar{\mathbf{D}}$ and let $\mathbf{V} = \mathbf{R} + \mathbf{Z}\bar{\mathbf{D}}\mathbf{Z^T}$. If θ is known the standard estimator for β is the generalized least square estimator

$$\hat{\beta}(\theta) = (\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{y}.$$

To estimate $\mathbf{b}$ the empirical Bayes estimate is used. Thus

$$\hat{\mathbf{b}}(\theta) = \overline{\mathbf{D}}\mathbf{Z}^{\mathbf{T}}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}(\theta))$$

which is the posterior mean, $E(\mathbf{b} \mid \mathbf{y}, \hat{\beta}(\theta), \theta)$. This was shown to be the best unbiased estimator by Harville[9].

The standard errors of $\hat{\beta}(\theta)$ and $\hat{\mathbf{b}}(\theta)$ are estimated by

$$\text{var}(\hat{\beta}(\theta)) = (\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}$$

and

$$\text{var}(\hat{\mathbf{b}}(\theta)) = \overline{\mathbf{D}}\mathbf{Z}^{\mathbf{T}}(\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{V}^{-1})\mathbf{Z}\overline{\mathbf{D}}.$$

If θ is unknown it has to be estimated to use in the above formulas. If $\hat{\theta}$ is an estimate of θ the estimates of β and $\mathbf{b}$ will be $\hat{\beta}(\hat{\theta})$ and $\hat{\mathbf{b}}(\hat{\theta})$. To estimate the variance components θ and σ two methods are used. First consider maximum likeihood estimation (ML). The maximum likelihood estimators of θ and σ maximize

$$\begin{aligned} l_F(\beta, \theta, \sigma \mid y) = \; & -\tfrac{1}{2}\log \mid \sigma^2\mathbf{V} \mid \\ & -\tfrac{1}{2}\sigma^{-2}(\mathbf{y} - \mathbf{X}\beta)^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta) \end{aligned}$$

as does the estimate of β, $\hat{\beta}(\theta)$.

The ML estimates of θ and σ are biased downward since they fail to take into account the loss of degrees of freedom from the estimation of β.

The second method is restricted maximum likelihood (RML). The RML estimates are not biased. The RML estimates of θ and σ are obtained by

maximizing the likelihood of θ and σ based not on $\mathbf{y}$ but on $N - p$ ($N = \sum_{i=1}^{M} n_i$) linearly independent error contrasts, $\mathbf{u^T y}$, chosen so that

$$E(\mathbf{u^T y}) = \mathbf{0}.$$

This log likelihood, derived by Harville[8], is

$$l_R(\widehat{\beta}(\theta), \sigma, \theta \mid \mathbf{y}) = \frac{-1}{2} \log \mid \sigma^{-2}\mathbf{X^T V^{-1} X} \mid + l_F(\widehat{\beta}(\theta), \sigma, \theta \mid \mathbf{y}).$$

Estimates of the standard errors can be obtained by substituting $\widehat{\theta}$ into the appropriate formula i.e. $\text{var}(\widehat{\beta}(\widehat{\theta})), \text{var}(\widehat{\mathbf{b}}(\widehat{\theta}))$. Methods for adjusting these formulas for the uncertainty of the estimation of θ are not available at present.

3.4 Fitting a Linear Mixed Effects Model to the Vision Data

The following model will be fitted with the vision loss data.

$$VIS_{i,j} = (\beta_0 + b_{0i}) + \beta_1(AGE_i - 50) + \beta_2(DIST_i - 2500) + (\beta_3 + b_{3i})e^{-TIME_{i,j}} + e_{i,j}$$

where $i = 1, \ldots, M$ with M the number of patients and $j = 1, \ldots, n_i$ with n_i the number of observations on patient i. $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ are the fixed effects and $\mathbf{b_i} = (b_{0i}, b_{3i})^T$ are the random effects specific to patient i. The $e_{i,j}$ are assumed to be independently distributed normal random variables with mean 0 and variance σ_2 Thus the matrix $\mathbf{R}$ is the identity matrix in this case.

To estimate β and the b_i the methods of the previous section were employed. The calculations were done using a function written for the S statistical software package ([4], [6]) by Bates, Lindstrom and Pinherio[3]. This

function was written utilizing the ability of S to interphase with C language programming. The maximum likelihood equations are solved using the Newton-Ralphson method to yield the parameter estimates. Estimates of the standard errors and correlations of the fixed effects are also provided. The estimation results using maximum likelihood estimation follow. The estimates of the random effects b_{0i}, b_{3i} for $i = 1, \ldots, 190$ are not shown. The estimate of the residual variance (σ^2) is 116.45.

TABLE 3.1 Linear Mixed Effects using ML

(a)Fixed Effects Estimates

	Value	Std Error	Z Ratio
β_0	52.78381	5.19533	10.160
β_1	-0.42781	0.12461	-3.433
β_2	0.00653	0.00208	3.137
β_3	28.11356	2.28402	12.309

(b)Correlations of Fixed Effects Estimates

	β_0	β_1	β_2
β_1	-0.5039		
β_2	0.7392	-0.0117	
β_3	-0.4195	0.0034	0.0011

(c)Covariance/Correlation Matrix of b

	b_1	b_3
b_1	986.21	-841.08
b_3	-0.92	842.09

The model was also fitted using restricted maximum likelihood with results shown in table 3.2. The estimate of σ_2 in this case is 116.44.

TABLE 3.2 Linear Mixed Effects using RML

(a)Fixed Effects Estimates

	Value	Std Error	Z Ratio
$\beta_{0)}$	52.77831	5.23267	10.086
β_1	-0.42777	0.12562	-3.405
β_2	0.00653	0.00210	3.115
β_3	28.11718	2.29079	12.274

(b)Correlations of Fixed Effects Estimates

	β_0	β_1	β_2
β_1	-0.5043		
β_2	0.7398	-0.0117	
β_3	-0.4177	0.0033	0.0011

(c)Covariance/Correlation Matrix of b

	b_0	b_3
b_0	994.26	-846.52
b_3	-0.92	847.80

Notice that the variance component estimates for the covariance matrix of random effects are slightly lower using maximum likelihood estimation. This is as expected due to the biasing downward as previously discussed. The fixed effects estimates are very similar in both fits in fact they are equal to three decimal places.

The conclusion from this model is that with every additional 10 years

of age the expected vision reading of the patient will decrease by 4.28 letters and with every additional mm that separates the CNV and the FAZ the vision reading will increase by 6.53 letters read. The effect of TIME after treatment will cause the expected vision to decrease with an inverse exponential relationship as time increases. With AGE and DIST at their most favourable values the expected level of VIS will be 52.78+28.11=80.89. As TIME gets large the expected level of VIS will approach 52.78. These coefficient estimates are somewhat different than the regression model without random effects and they can be considered more reliable as the within patient correlation is taken into account.

Note how much greater the variance of the random effects is than the error variance (986 and 842 to 116). This indicates that a much larger portion of the error not accounted for by the fixed effect part of the model can be attributed to between patient variance than to within cluster random error.

3.5 Evaluation of Model

The discussion of the fitted model in this section will refer to the model estimated by maximum likelihood. These comments are also applicable to the model fitted with the restricted maximum likelihood method as it differs only slightly.

The method of estimation of the random effects is based on the normality assumption for the error terms and the random effects. By this assumption the estimator for the random effects will be normaly distributed as will the estimates of the errors. Normal probability plots of the random effect and error term estimates were prepared and are shown in Figures 3.1, 3.2, and 3.3. For the random effect estimates the normal probability plots show a fairly linear trend with correlation coefficients of .977 and .975 for $\mathbf{b_0}$ and $\mathbf{b_3}$ respectively. However in both plots there is some curvature in the high and low extremes. This indicates the tails of the distributions of the estimators of the random effects are shorter than the tails of a normal distribution. Thus there is some evidence that the normality assumptions in the model may not be adequate for this data. The normal probability plot concerning the error estimates is also fairly linear with a correlation coefficient of .978.

To check the assumption of a constant variance of the error terms Figure 3.4 was prepared. The boxplots of the error estimates for each cluster seem to give some evidence against this assumption as there is large differences in dispersion over the clusters. However this can be partially explained by the large number of observations and small relative size of the clusters.

The high estimated variances of both random effects indicate that it is reasonable to assume that their respective variables' coefficient has a random component that varies over the clusters.

Figure 3.1

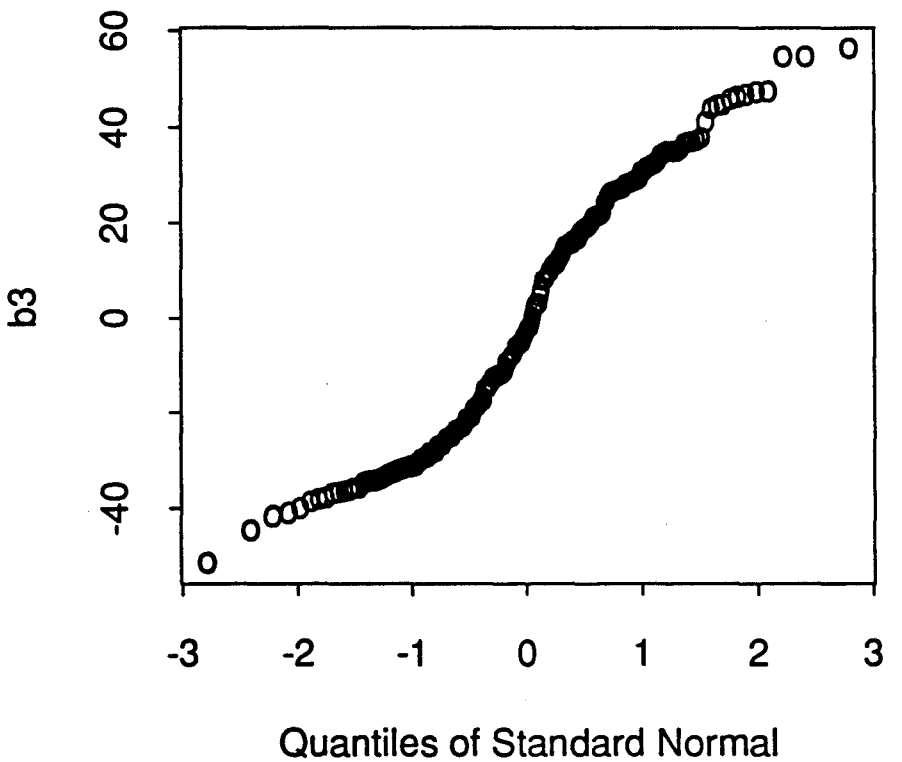


Figure 3.2

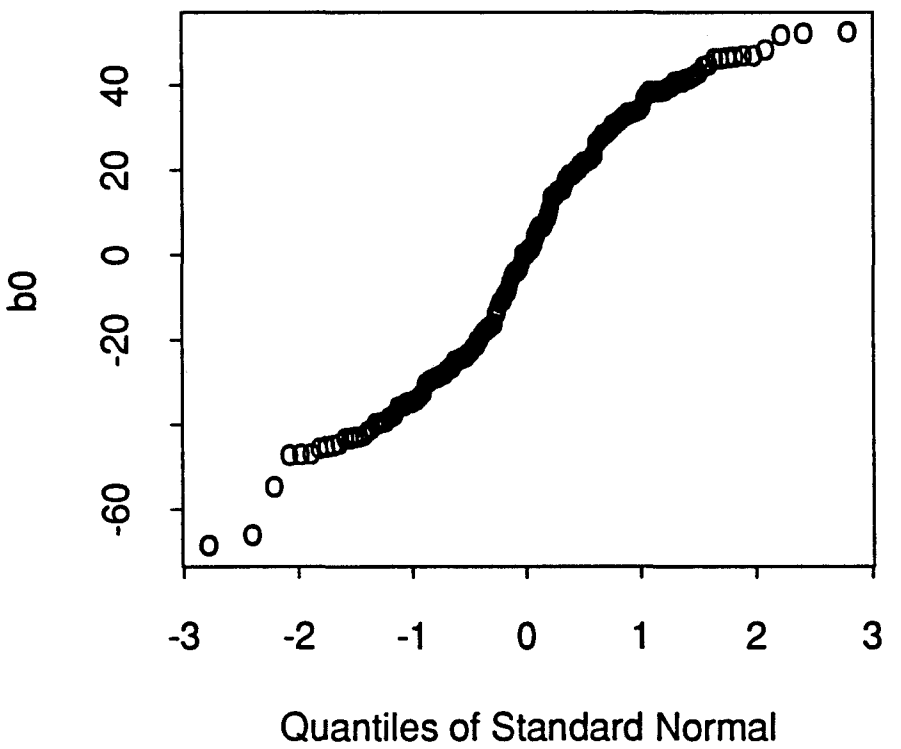


Figure 3.3

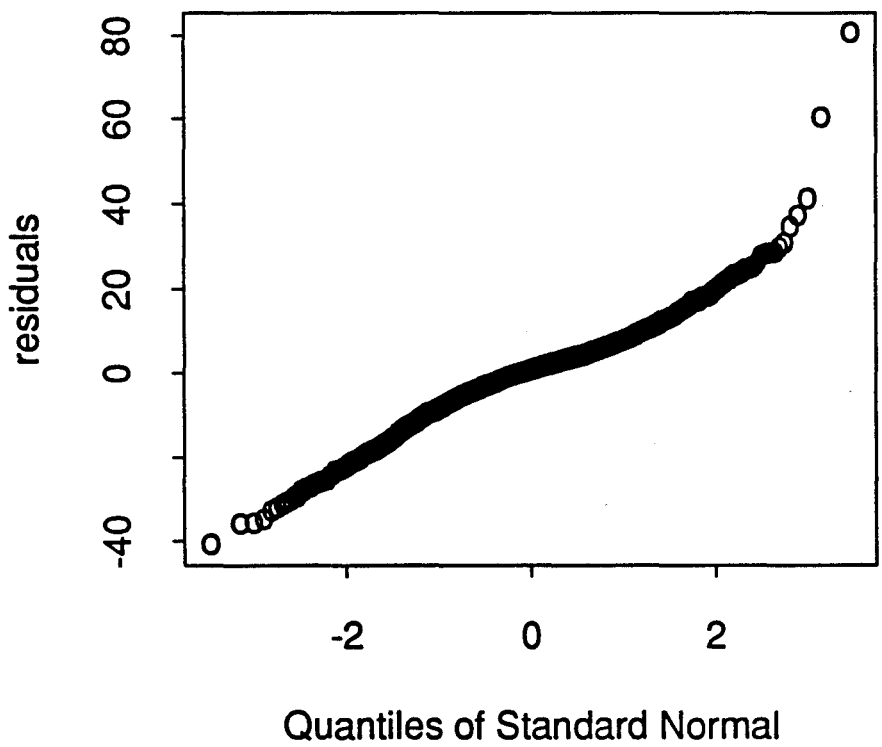


Figure 3.4

Boxplots of Residuals by cluster

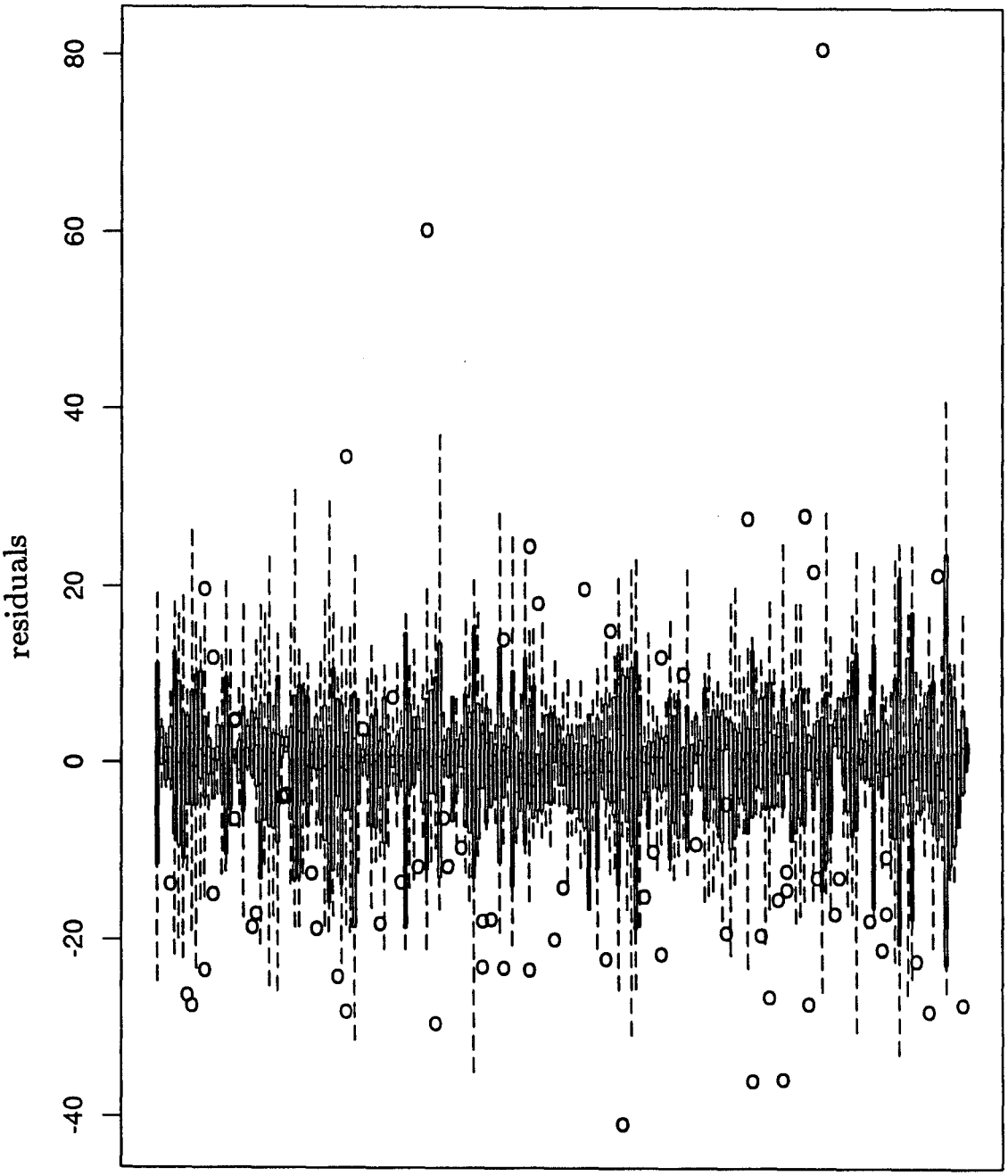


Figure 3.5

Predicted Values vs. VIS for Linear Model

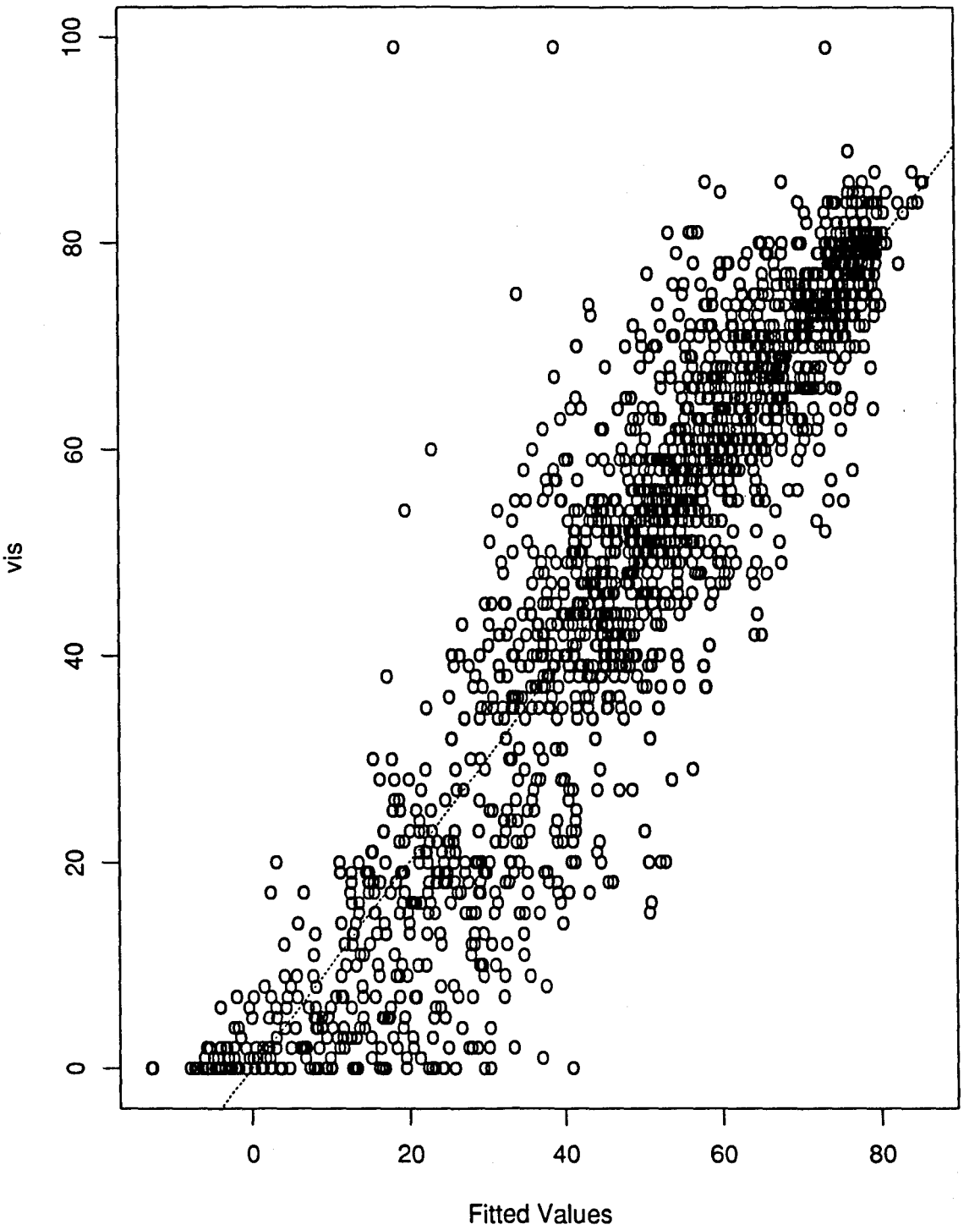
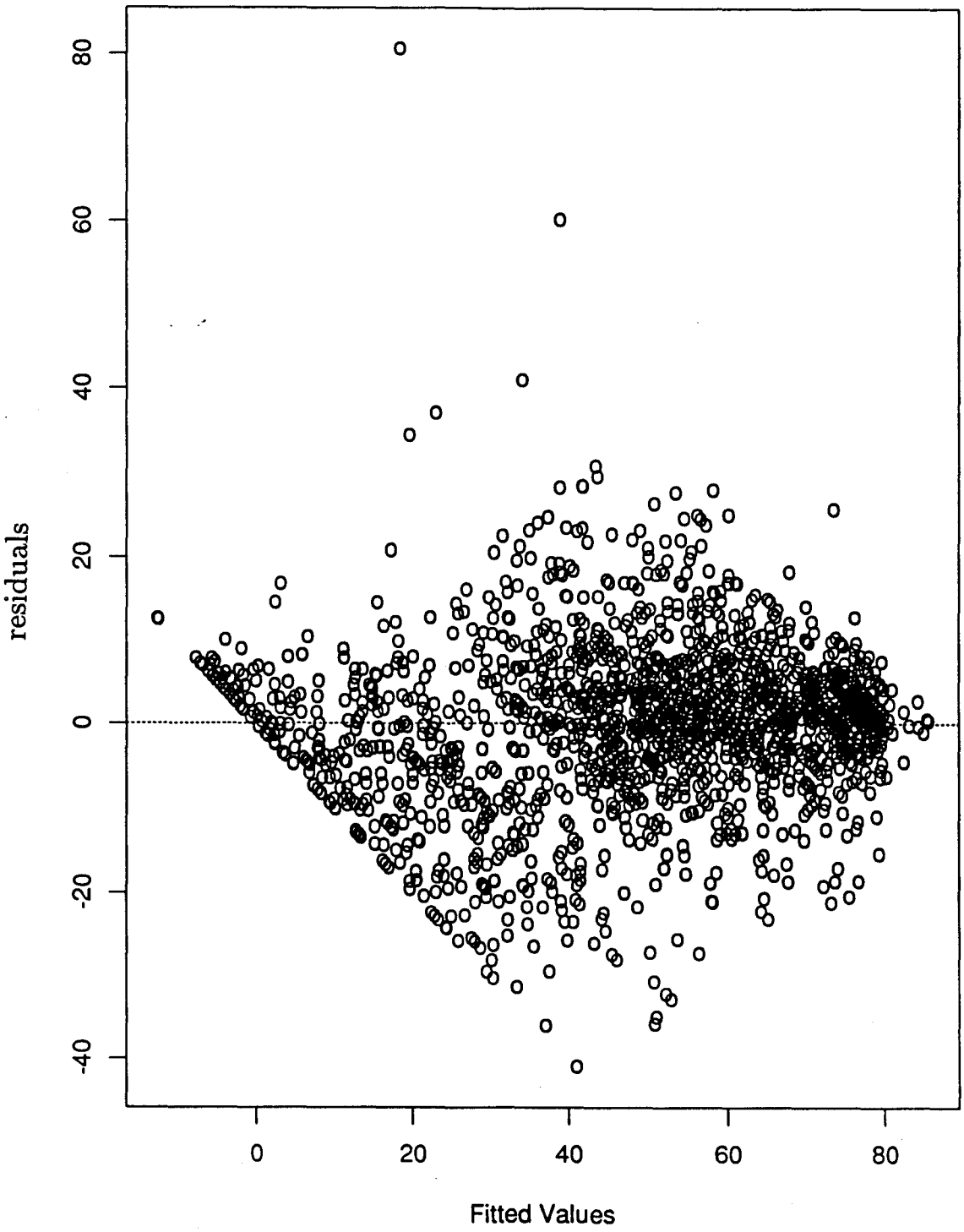


Figure 3.6

Predicted Values vs. Residuals for Linear Model



The plots of residuals and actual values against fitted values (Figures 3.5, 3.6) indicate some overfit in the low predicted values and underfit in the high range. These plots are tightened up considerably compared to the corresponding standard regression model with no random effects ie Var(**b**) = **0**. The error variance estimate is down to 116.45 compared to 501.95 in the standard regression model.

Although this model does fit the data better this is due to the random effect estimates. When using the model to predict for patients not used as clusters in the model the random effects must be considered zero. To assess and compare this models usefulness for prediction an R^2 value can be calculated by finding the fraction of the total sum of squares that is explained by the model with random effect estimates taken as zero. This R^2 value turns out to be .1860 which is an improvement over the value of .1623 for the corresponding standard regression model.

3.6 Linear Mixed Effects Model With Initial Vision Variable

The model with the initial vision variable is again considered. The model is the same as the one from the previous section except the initial vision readings are used as a predictor variable (STVIS) and are excluded as data points.

$$\begin{aligned} VIS_{ij} = \; & (\beta_0 + b_{0,i}) + \beta_1(AGE_i - 50) + \beta_2(DIST_i - 2500) \\ & + (\beta_3 + b_{3,i})e^{-TIME_{ij}} + \beta_4 STVIS_i + \epsilon_{ij} \end{aligned}$$

The initial vision is constant across clusters and thus is fitted with only a fixed effect. The estimation results follow. The estimate of σ^2 is 110.78.

TABLE 3.3 Linear Model With Initial Vision Variable

(a)Fixed Effects Estimates

	Value	Std Error	Z Ratio
β_0	13.05141	8.04782	1.622
β_1	-0.22061	0.13304	-1.658
β_2	0.00787	0.00210	3.740
β_3	28.42298	2.31833	7.389
β_4	0.63936	0.08653	7.389

(b)Correlations of Fixed Effects Estimates

	β_0	β_1	β_2	β_3
β_1	-0.5147			
β_2	0.5412	-0.0403		
β_3	-0.2619	0.0029	-0.0028	
β_4	-0.7595	0.2434	-0.0796	0.0052

(c)Covariance/Correlation Matrix of b

	b_0	b_3
b_0	913.76	-802.05
b_3	-0.91	847.18

The coeeficient of STVIS indicates the expected vision will increase by .639 for every unit of increase in STVIS. Figures 3.7 and 3.8 show further improvement of the fit although the same problems exist.

Figure 3.7

Predicted Values vs. VIS for Linear Model
With Initial Vision Variable

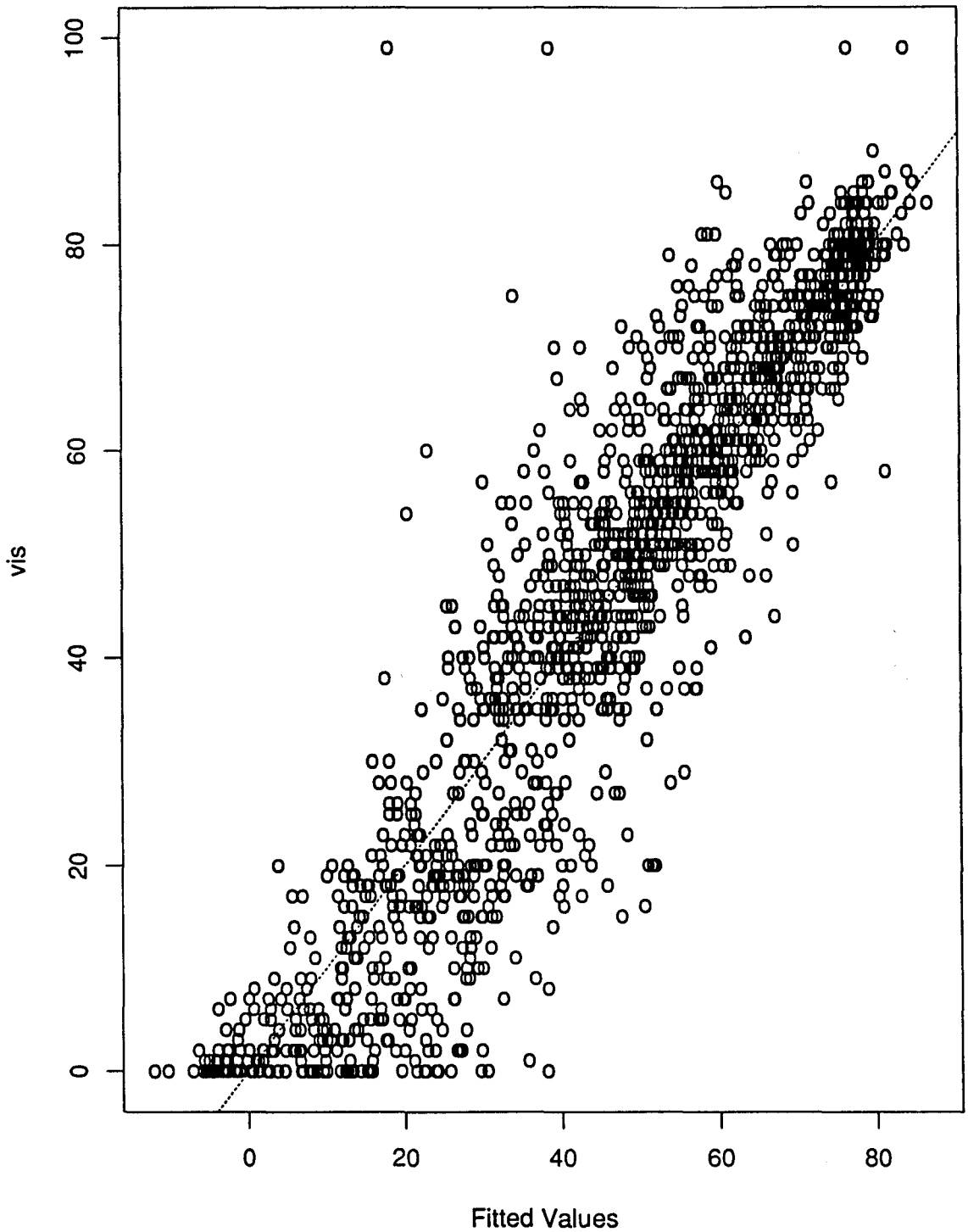
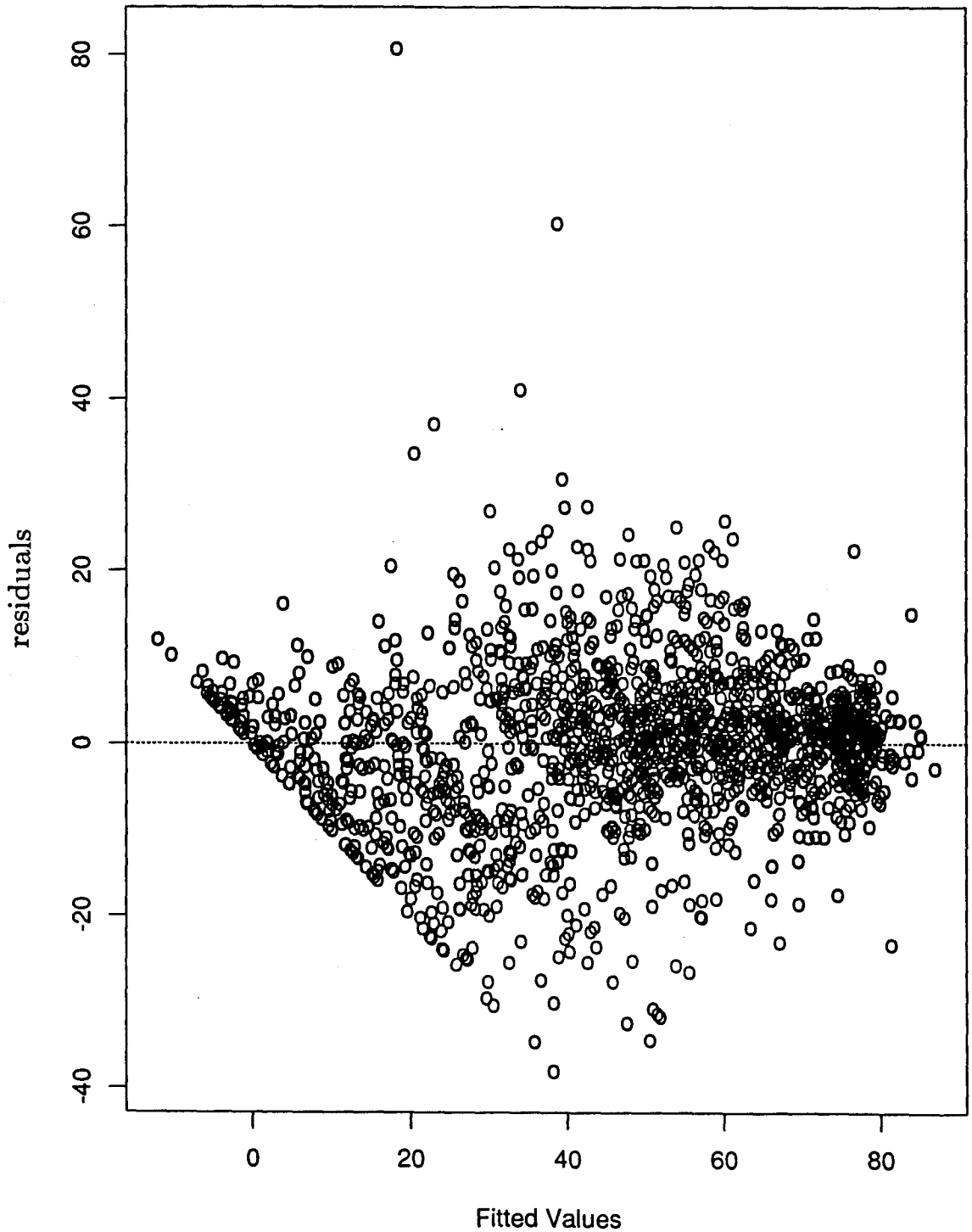


Figure 3.8

Predicted Values vs. Residuals for Linear Model
With Initial Vision Variable



Chapter 4

NonLinear Mixed Effects Models

4.1 Description of Model

A general nonlinear mixed effects model is defined as follows.

$$\mathbf{y} = \mathbf{f}(\mathbf{A}\beta + \mathbf{B}\mathbf{b}, \mathbf{X}) + \mathbf{e}$$

- $\mathbf{y}$ is the vector of responses of length $N = \sum_{i=1}^{M} n_i$ where n_i is the number of observations on cluster i and M is the number of clusters.
- $\mathbf{X}$ is the matrix of prediction vectors of length N.
- $\mathbf{e}$ is a normally distributed error vector with mean $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{R}$

- $\mathbf{f}(\mathbf{A}\beta + \mathbf{B}\mathbf{b}, \mathbf{X}) = \begin{bmatrix} f(\mathbf{A}_1\beta + \mathbf{B}_1\mathbf{b}_1, \mathbf{x}_{11}) \\ f(\mathbf{A}_1\beta + \mathbf{B}_1\mathbf{b}_1, \mathbf{x}_{12}) \\ \vdots \\ f(\mathbf{A}_1\beta + \mathbf{B}_1\mathbf{b}_1, \mathbf{x}_{1n_1}) \\ f(\mathbf{A}_2\beta + \mathbf{B}_2\mathbf{b}_2, \mathbf{x}_{21}) \\ \vdots \\ f(\mathbf{A}_2\beta + \mathbf{B}_2\mathbf{b}_2, \mathbf{x}_{2n_2}) \\ \vdots \\ f(\mathbf{A}_M\beta + \mathbf{B}_M\mathbf{b}_M, \mathbf{x}_{Mn_M}) \end{bmatrix}$ where β is a vector of length p of fixed population parameters, $\mathbf{b_i} \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{D})$ and is a vector of length q of random effects associated with individual i, $\mathbf{A_i}$ and $\mathbf{B_i}$ are design matrices of size $r \times p$ and $r \times q$ respectively. $\mathbf{x_{ij}}$ is the row of $\mathbf{X}$ that corresponds to the jth observation in cluster i, $f(\mathbf{A_i}\beta + \mathbf{B_i}\mathbf{b_i}, \mathbf{x_{ij}})$ is the expectation function of the component of $\mathbf{y}$ corresponding to the jth observation on cluster i,

  $\mathbf{b} = (\mathbf{b_1^T}, \mathbf{b_2^T}, \ldots, \mathbf{b_M^T})^\mathbf{T}$, $\mathbf{A} = (\mathbf{A_1^T}, \mathbf{A_2^T}, \ldots, \mathbf{A_M^T})^\mathbf{T}$ and

  $\mathbf{B}$ =diag$(\mathbf{B_1}, \mathbf{B_2}, \ldots, \mathbf{B_M})$.

The design matrices $\mathbf{A}$ and $\mathbf{B}$ allow situations to be considered where not all of the parameters have a random component or different fixed effects are needed for grouped data. The conditional distribution $\mathbf{y} \mid \mathbf{b} \sim N(\mathbf{f}(\mathbf{A}\beta + \mathbf{B}\mathbf{b}, \mathbf{X}), \sigma^2\mathbf{R})$.

4.2 Estimation of Parameters

The estimation procedure proposed by Bates and Lindstrom in [2] is a two stage alternating algorithm. Let θ be the vector of unknown variance components. In the first stage of the algorithm estimates of β and $\mathbf{b}$ are obtained

by maximizing the function

$$g(\mathbf{b} \mid \mathbf{y}, \beta, \theta, \sigma) = -\tfrac{1}{2}\sigma^{-2}(\mathbf{y} - \mathbf{f}(\mathbf{A}\beta + \mathbf{B}\mathbf{b}))^{\mathbf{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{A}\beta + \mathbf{B}\mathbf{b})) - \tfrac{1}{2}\sigma^{-2}\mathbf{b}^{\mathbf{T}}\bar{\mathbf{D}}^{-1}\mathbf{b}.$$

Given β, θ and σ it can be shown that the posterior mean of $\mathbf{b}$ maximizes g. In the first iteration of the algorithm starting values of β, σ and θ are used and estimates of $\mathbf{b}$ are obtained.

Because the expectation function $\mathbf{f}$ is nonlinear in $\mathbf{b}$ there is no closed form expression for the marginal density of $\mathbf{y}$. Thus maximum likelihood estimators of θ, σ^2 and β with respect to the marginal density of $\mathbf{y}$ are not calculated. Instead the conditional distribution of $\mathbf{y}$ given $\mathbf{b}$ is approximated. The expectation $\mathbf{f}(\mathbf{A}\beta + \mathbf{B}\mathbf{b})$ is approximated by a first order Taylor series expansion about a current estimate of $\mathbf{b}$ say $\hat{\mathbf{b}}$. Thus

$$\mathbf{f}(\mathbf{A}\beta + \mathbf{B}\mathbf{b}) \simeq \mathbf{f}(\mathbf{A}\beta + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Z}}(\mathbf{b} - \hat{\mathbf{b}})$$

where

$$\hat{\mathbf{Z}} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{b}}\right|_{\hat{\mathbf{b}}}.$$

The approximate conditional distribution of $\mathbf{y}$ given $\mathbf{b}$ is

$$N(\mathbf{f}(\mathbf{A}\beta + \mathbf{B}\hat{\mathbf{b}}) - \hat{\mathbf{Z}}\hat{\mathbf{b}} + \hat{\mathbf{Z}}\mathbf{b}, \sigma^2\mathbf{R})$$

Therefore the marginal distribution of $\mathbf{y}$ can be approximated as

$$N(\mathbf{f}(\mathbf{A}\beta + \mathbf{B}\hat{\mathbf{b}}) - \hat{\mathbf{Z}}\hat{\mathbf{b}}, \sigma^2\widehat{\mathbf{V}})$$

where $\widehat{\mathbf{V}} = \mathbf{R} + \hat{\mathbf{Z}}\bar{\mathbf{D}}\hat{\mathbf{Z}}^{\mathbf{T}}$.

The log likelihood corresponding to this approximate distribution is

$$l_F(\beta, \sigma, \theta \mid \mathbf{y}) = -\tfrac{1}{2}\log \mid \sigma^2\widehat{\mathbf{V}} \mid - \tfrac{1}{2}\sigma^{-2}(\mathbf{y} - \mathbf{f}(\mathbf{A}\beta + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Z}}\hat{\mathbf{b}})^{\mathbf{T}}\widehat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{A}\beta + \mathbf{B}\hat{\mathbf{b}}) + \hat{\mathbf{Z}}\hat{\mathbf{b}}).$$

In the second stage of the algorithm l_F is maximized, with the estimates of $\mathbf{b}$ from stage 1, to obtain estimates of β, σ and θ. These estimates are then used in stage 1 in the next iteration of the algorithm.

Standard errors and correlations of the fixed effects are obtained from the likelihood function used in stage 2 of the last iteration of the estimation algorithm. They tend to underestimate the actual standard error and correlation as their estimates are conditional on θ as in the linear mixed effect estimation.

4.3 Fitting a Nonlinear Mixed Effects Model To the Vision Data

With the algorithm for estimating coefficients for a nonlinear mixed effects model the coefficient for the TIME variable, that previously was taken to be -1, can be estimated. An attempt was made to fit the following model with the vision loss data. The calculations were again carried out by an S function written by Bates, Lindstrom and Pinheiro [3].

$$\begin{aligned} VIS_{ij} = \; & (\beta_0 + b_{0,i}) + \beta_1(AGE_i - 50) \\ & + \beta_2(DIST_i - 2500) + (\beta_3 + b_{3,i})e^{(\beta_4 + b_{4,i})TIME_{ij}} + \epsilon_{ij} \end{aligned}$$

The subscript on the variables denotes the jth observation on the ith cluster. The ϵ_{ij} are independently distributed normal random variables with mean 0 and variance σ^2. Thus the R matrix in this case is the identity. $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ and $\mathbf{b_i} = (b_{0i}, b_{3i})^T$ are the fixed and random effects respectively.

The algorithm would not converge with this model. Estimates of the covariance matrix of $\mathbf{b}$ in step 2 of the algorithm become larger in absolute value with each successive iteration. This indicates that the model may be overparameterized in terms of random effects. It was decided that the random effect would be dropped from the coefficient of TIME. This parameter effects the shape of the exponential curve and it is reasonable to have the shape parameter constant between clusters. Therefore a fit of the following model was done.

$$VIS_{ij} = (\beta_0 + b_{0,i}) + \beta_1(AGE_i - 50) + \beta_2(DIST_i - 2500) + (\beta_3 + b_{3,i})e^{\beta_4 TIME_{ij}} + e_{ij}$$

The estimation results follow. The estimates of the random effects, $b_{0,i}, b_{3,i}$ for $i = 1, 2, \ldots, 190$ are not shown.

TABLE 4.1 Nonlinear Model

(a)Fixed Effects Estimates

	Value	Std Error	Z Ratio
β_0	55.32297	5.06697	10.918
β_1	-0.43307	0.12283	-3.526
β_2	0.00644	0.00205	3.138
β_3	26.29027	2.16294	12.155
β_4	-1.35471	0.08689	-15.591

(b)Correlations of Fixed Effects Estimates

	β_0	β_1	β_2	β_3
β_1	-0.5076			
β_2	0.7454	-0.0120		
β_3	-0.3973	0.0004	0.0017	
β_4	-0.0721	-0.0140	0.0179	0.1278

(c)Covariance/Correlation Matrix of b

	b_0	b_3
b_0	848.93	-727.16
b_3	-0.91	746.27

The fixed effects estimates, their estimated standard errors and the random effect estimates (not shown) are all very similar to their corresponding estimates in the linear mixed effects model. The variance of both random effects is lower (849 to 986 for b_0, 746 to 842 for b_3) while the residual variance (estimate of $\sigma^2 = 116.89$) is slightly higher. It seems that the estimation of the coefficient for TIME has caused more of the random variation to be attributed to within cluster error than to between cluster error.

The use of the approximate log likelihood accounts for a slight decrease in the prediction usefulness of the model as measured by an R^2 value calculated with random effects taken as zero (18.60 to 18.19). The standard errors of the fixed effects are all slightly lower than in the linear mixed effect model. Although in both models they are expected to be biased downward, these estimates are even more of an approximation in this model as they are obtained from an approximate log likelihood.

Figure 4.1 shows the fitted values plotted against the VIS variable to assess the fit. It is similar to the corresponding plot for the linear model.

4.4 Adding the Initial Vision Variable

A nonlinear mixed effects model with the first observation on each patient ommitted and the vision component used as a predictor variable will now be discussed. Everthing in the following model corresponds to the model of the previous section except for the addition of the STVIS variable and the coefficient β_5.

$$\begin{aligned} VIS_{ij} = \; & (\beta_0 + b_{0,i}) + \beta_1(AGE_i - 50) + \beta_2(DIST_i - 2500) \\ & + (\beta_3 + b_{3,i})e^{\beta_4 TIME_{ij}} + \beta_5 STVIS_i + \epsilon_{ij} \end{aligned}$$

Figure 4.1

Predicted Values vs. VIS for Nonlinear Model

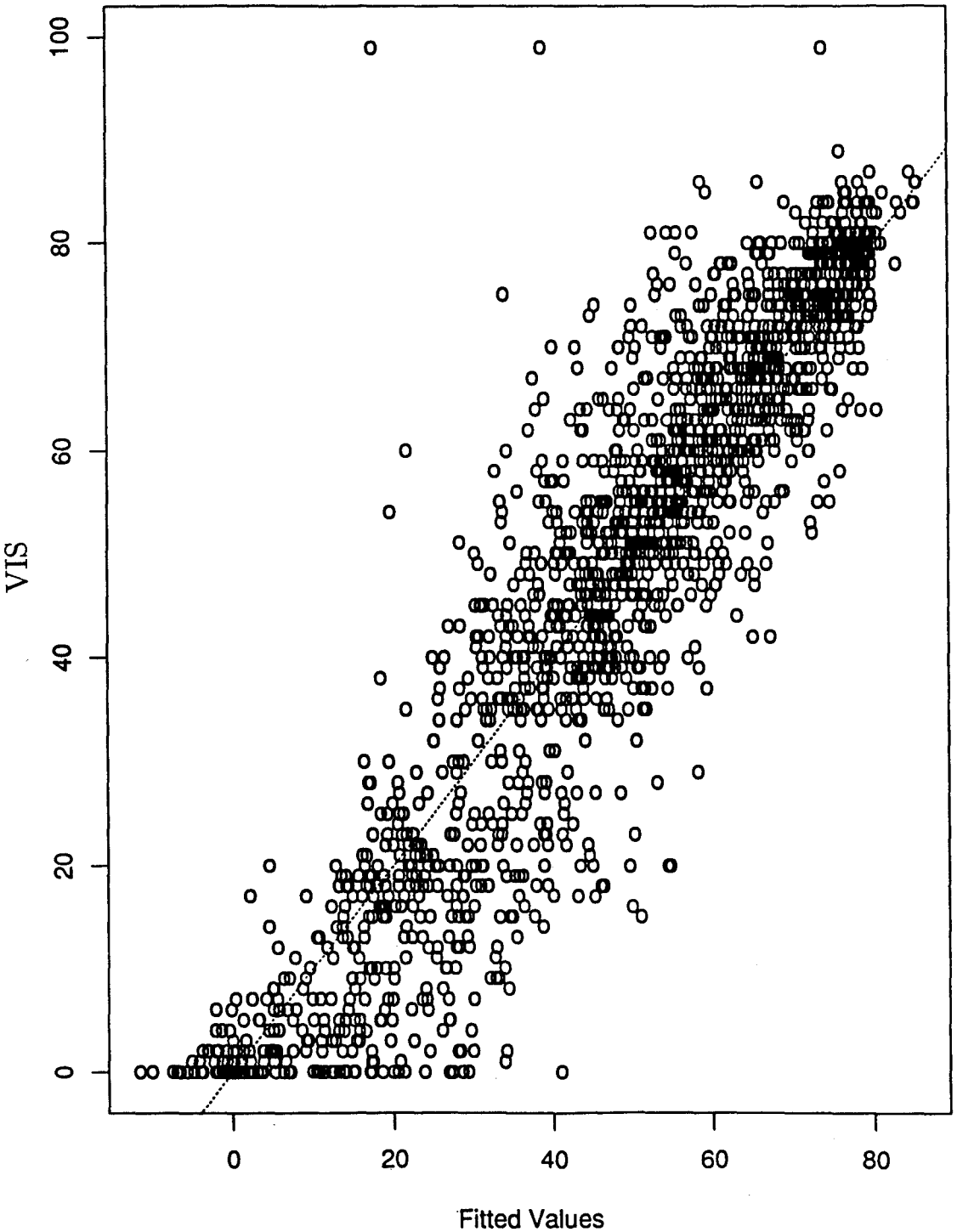


TABLE 4.2 Nonlinear Model With Initial Vision Variable

(a)Fixed Effects Estimates

	Value	Std Error	Z Ratio
β_0	15.65356	7.91746	1.977
β_1	-0.22858	0.13146	-1.739
β_2	0.00784	0.00207	3.773
β_3	27.22975	2.23136	12.203
β_4	-1.31777	0.08725	-15.102
β_5	0.63540	0.08547	7.433

(b)Correlations of Fixed Effects Estimates

	β_0	β_1	β_2	β_3	β_4
β_1	-0.5155				
β_2	0.5431	-0.0407			
β_3	-0.2460	-0.0004	-0.0015		
β_4	-0.0333	-0.0242	0.0159	0.0804	
β_5	-0.7619	0.2416	-0.0801	0.0041	-0.0161

(c)Covariance/Correlation Matrix of b

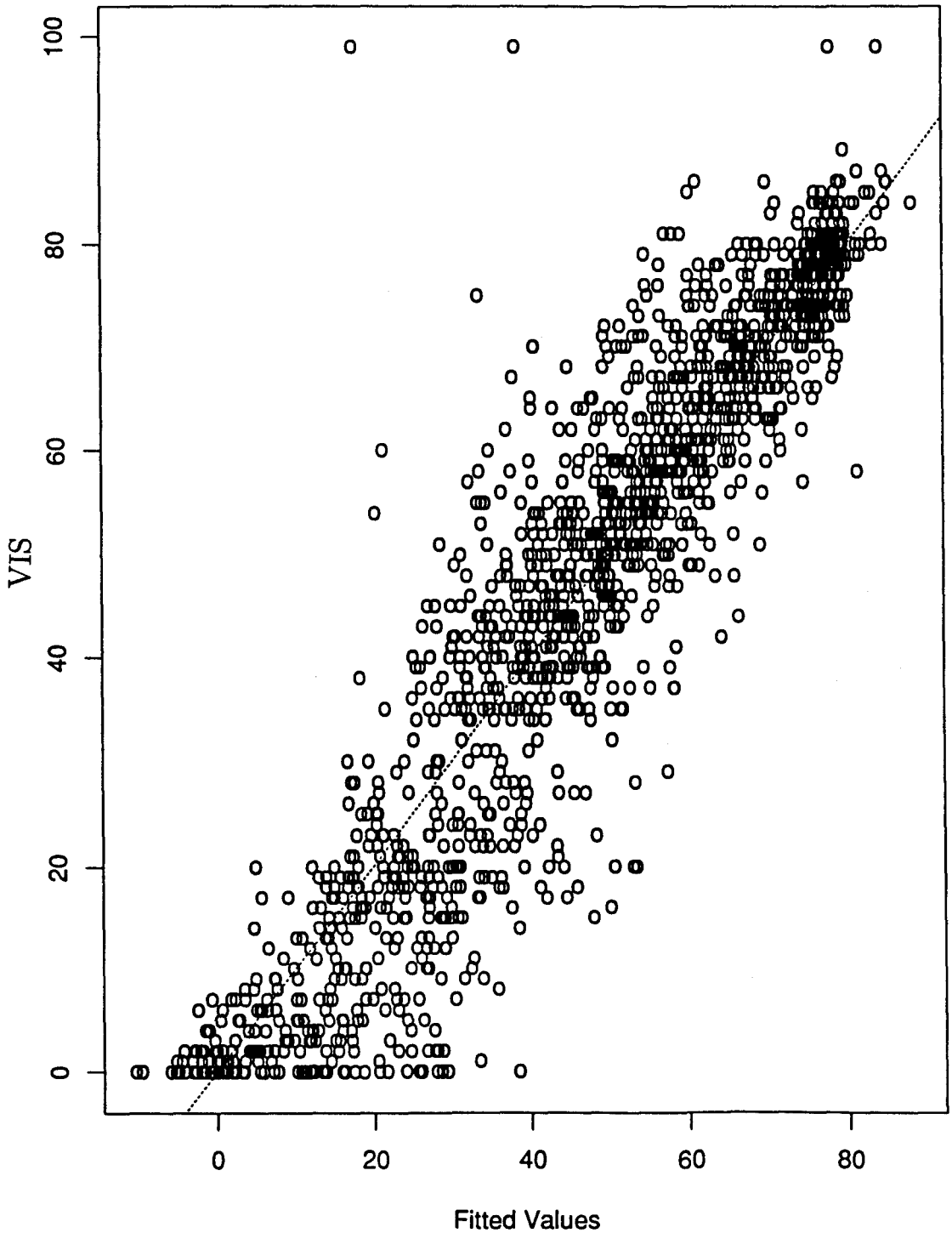
	b_0	b_3
b_0	796.73	-712.19
b_3	-0.90	783.32

The comparison between this model and the corresponding linear mixed effects model is the same as the comparison for the models without the STVIS variable. The fixed effects estimates and standard deviations are similar to the estimates for the linear model and the estimate of σ^2 is up slightly to 111.14.

The plot of fitted values against the VIS variable is shown in Figure 4.2. In the nonlinear mixed effect models as with the others, the model with the initial vision variable added is preferable when used for prediction however it may not be preferable when estimating the effect of AGE on VIS as the standard error for β_1 is quite high.

Figure 4.2

Predicted Values vs. VIS for Nonlinear Model
With Initial Vision Variable



Chapter 5

Conclusion

It can be concluded from the estimated coefficients that increases in DIST and STVIS will result in a increase in expected vision while an increase in AGE will result in a decrease in expected vision. As TIME after treatment increases the expected vision decrease can be described with the inverse exponential relationship.

Table 5.1 shows a comparison of the estimated coefficients and estimated error and random effect variances for the models fit with the effect of time after treatment described by an inverse exponential relationship.

Table 5.1 Summary of Results

Model		Linear Fixed Effects Only	Linear Mixed Effects	Nonlinear Mixed Effects
All	AGE	-5.51	-4.28	-4.33
Observations	DIST	7.78	6.53	6.44
	$e^{\beta TIME}$	23.8	28.1	26.3
	TIME	-1	-1	-1.35
	σ^2	502	116	117
	Var(b_0)	-	986	849
	Var(b_3)	-	842	746
Post Treatment	AGE	-3.50	-2.21	-2.29
Observations	DIST	6.91	7.87	7.84
	$e^{\beta TIME}$	23.8	28.4	27.2
	TIME	-1	-1	-1.32
	STVIS	.630	.639	.635
	σ^2	492	111	114
	Var(b_0)	-	914	797
	Var(b_3)	-	847	783

In all 3 model types the coefficient of AGE decreased with the addition of the STVIS variable. The obvious correlation between these two variables accounts for this. The variance of the random effect associated with the intercept (b_0) was reduced in the models with the start vision variable. This indicates, as expected, that the addition of STVIS reduces the between subject variability.

The error variance (σ^2) decreased greatly in the mixed effects models. This is because the random effects account for the within patient variability.

The effect of AGE and DIST are less in most cases in the mixed effect models. Thus some of their effect has been accounted for by within patient variability.

There is no improvement in the model when estimating a coefficient for TIME using the nonlinear mixed effect algoithm in fact there is a slight decrease in prediction usefulness. It turns out in this case that the estimate of the coefficient for time is very close to the -1 that had been used in the linear model and that it is not possible to use a random effect with this coefficient. Thus the advantages in this approach are very slight while the disadvantage of using the approximate log likelihood for estimation outweighs them.

However, based on the fact that the estimation results are in fact very close for the two methods it is suggested that the nonlinear mixed effects algorithm is an efficient method to fit a model that has no linear alternative.

To show how the model is used for prediction assume that a patient is 65 years old, the distance from the FAZ is 1 mm and the number of letters read before treatment is 70. Suppose a prediction of vision after 2 years is needed. If the linear model with the initial vision variable is used the values for AGE, DIST, TIME and STVIS are substituted into

$$VIS = 13.05 - .221(AGE - 50) + .00787(DIST - 2500) + .639STVIS + 28.4e^{-TIME}.$$

Notice that the random effects are taken as there expected value of 0. Thus the expected vision after 2 years will be 46.54 letters read.

Bibliography

[1] Bates DM, Lindstrom MJ, Newton-Raphson and EM algorithms for linear mixed effects model for repeated measures data, *JASA* **83** (1988) 1014-1022.

[2] Bates DM, Lindstrom MJ, Nonlinear mixed effects models for repeated measures data, *Biometrics*, **46** (1990) 673-687.

[3] Bates DM, Lindstom MJ, Pinheiro JC, Nonlinear mixed effects classes and methods for S, preprint (1993)

[4] Becker RA, Chambers JM, Wilks AR, *The New S Language: A Programming Environment for Data Analysis and Graphics*, Wadsworth, Belmont, CA, (1988).

[5] The Canadian Ophthalmology Study Group, Argon green versus krypton red laser photocoagulation for extrafoveal choroidal neovascularization: One year results in age-related macular degeneration, *Arch. Ophthalmol.* **111** (1993) 181-185.

[6] Chambers JM, Hastie TJ (eds), *Statistical Models in S*, Wadsworth, Belmont, CA, (1991).

[7] Ferris FL, Fine SL, Hyman L, Age related macular degeneration and

blindness due to neovascular maculopathy, *Arch. Ophthalmol.* **102** (1984), 1640-1642.

[8] Harville DA, Bayesian inferences for variance components using only error contrasts, *Biometrika*, **61** (1974) 383-385.

[9] Harville DA, Extension of the Gauss-Markov theorem to include the estimation of random effects, *The Annals of Statistics*, 4 (1976) 384-395.

[10] Harville DA, Maximum likelihood approaches to variance component estimation and to related problems, *JASA*, **72** (1977) 320-340.

[11] Macular Photocoagulation Study Group, Argon laser photocoagulation for senile macular degeneration: results of a randomized clinical trial, *Arch. Ophthalmol.* **100** (1982) 912-918.

[12] Macular Photocoagulation Study Group, Recurrent choroidal neovascularization after argon laser photocoagulation for neovascular macuopathy, *Arch. Ophthalmol.* **104** (1986) 503-512.

[13] Macular Photocoagulation Study Group, Persistent and recurrent neovascularization after krypton laser photocoagulation for neovascular lessions of age-related macular degeneration , *Arch. Ophthalmol.* **108** (1990) 825-831.

[14] Macular Photocoagulation Study Group, Argon laser photocoagulation for neovascular macuopathy: three year results from randomized clinical trials, *Arch. Ophthalmol.* **104** (1986) 694-701.

[15] Macular Photocoagulation Study Group, Argon laser photocoagulation for neovascular macuopathy: five year results from randomized clinical

trials, *Arch. Ophthalmol.* **109** (1991) 1109-1114.

[16] Laird NM, Ware JH, Random effects models for longitudinal data, *Biometrics* **38** (1982) 963-974.

[17] Neter J, Wasserman W, Kutner MH, *Applied Linear Regression Models, 2nd Ed.*, Richard D. Irwin, (1989).